Elucidating the molecular mechanism of TET2 function in hematopoiesis

Wenjun Huang

Magdalen College

Nuffield Department of Medicine

University of Oxford

United Kingdom

A thesis presented for the degree of

Doctor of Philosophy in Clinical Medicine

Hilary 2021

Elucidating the molecular mechanism of TET2 function in hematopoiesis

Wenjun Huang
Magdalen College

A thesis presented for the degree of
Doctor of Philosophy in Clinical Medicine
Hilary 2021

Abstract

DNA modifications are epigenetic marks deposited throughout the mammalian genome and play an important role in gene regulation in normal and malignant biological processes. Biological DNA modifications are introduced by DNA methyltransferases (DNMT) and dioxygenases of ten-eleven translocation (TET) family. DNMT proteins can convert cytosine into 5-methylcytosine. TET proteins catalyze the oxidation of 5mC to 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. These oxidized forms of cytosine stay in the DNA or result in DNA demethylation. TET2, a member of TET family proteins, is one of the most frequently mutated genes in hematological malignancies. While TET2 has roles in normal hematopoiesis, including stem cell self-renewal and specific lineage commitment, the exact molecular mechanism of its function still remains unknown. The aims of the project are to identify the role of TET2 in regulating hematopoietic gene expression and to identify the molecular partners of TET2 in the hematopoietic-associated context, thus to arrive at the mechanism of action of TET2 in normal hematopoiesis.

To study the Tet2 function in a hematopoietic context, I chose the mouse hematopoietic progenitor cell line, HPC-7, as a model system. I constructed the *Tet2*-knock-out (KO) HPC-7 cell and characterized the cell differentiation phenotype caused by Tet2 depletion. To understand the role of Tet2 in gene regulation, I firstly identified genes affected by Tet2 depletion. Then I mapped the Tet2 binding profile in HPC-7 and discovered genes directly targeted by Tet2. Finally, I assessed the contribution of DNA demethylation in Tet2-mediated gene regulation. The results revealed that Tet2 deficiency caused a block of cell differentiation in the Colony-forming-unit assay. Genes involved in hematopoiesis were differentially expressed upon Tet2 deficiency. Surprisingly, Tet2 binding is associated with both gene activation and gene repression. Tet2 binds to the hematopoietic-enhancers together with many hematopoietic transcription factors, and leads to DNA demethylation at enhancers, thus activates the expression of essential hematopoietic genes, such as *csf1R*, *Cebpa*, *Cebpe*. The mechanism by which Tet2 negatively regulates its targeted gene expression needs further investigation.

To identify Tet2 molecular partners in the hematopoietic-associated context, I mapped the TET2 interactome in HPC-7 cells by immunoprecipitation coupled to mass spectrometry. Our data suggests that polycomb repressive complex 2 (PRC2) physically interacts with Tet2 via the C-terminus of Tet2 protein. The functional importance of Tet2-PRC2 interaction in gene regulation and cell differentiation remain further explored.

Taken together, this study systematically investigated the Tet2-mediated gene regulation and identified new Tet2 binding partner, PRC2 complex, suggesting a novel mechanism of Tet2 function in hematopoiesis.

This thesis is the result of my own work unless otherwise stated and does not constitute part of any other thesis. The work herein described was carried out at the Ludwig Institute for Cancer Research, under the supervision of Dr. Skirmantas Kriaucionis. The work presented in Chapter 4.2.5 was a shared work between myself, Dr. Sophie Kirschner and Olena Yavorska, where I performed the cell line construction, Sophie carried out the methylation mapping using TAPS-β, Olena analyzed the TAPS-β data. Protein mass spectrometry was done by Georgina Berridge. Nucleoside mass spectrometry was done by Paolo Spingardi.

Word count: approx. 48503 (including citations and tables)

Wenjun Huang
Magdalen College

Acknowledgements

First of all, I would like to thank my supervisor Dr. Skirmantas Kriaucionis for hosting me in his laboratory and guiding me throughout the past four years. I'm particularly grateful for his support and trust to allow me to drive this project. I also appreciate his patience and understanding for me to get over the difficult period during the DPhil study. I would like to thank Cancer Research UK and Ludwig Institute for Cancer Research for the funding support.

Next, I would like to thank people who have ever helped me with experiments or supported me during my pursuit in science. I would like to thank Dr. Sarah De Val for her encouragements and advice in supporting me throughout the difficulties during the study. I would like to thank Dr.Thomas Milne for allowing me to present my work and join the scientific discussion in his group. I would like to thank Dr. Lynn Quek for her supervision in the colony-forming-unit assay. I'm also quite grateful to be a member of Ludwig and have received numerous help from people within the institute. I would like to thank Dr. Pakavarin Louphrasitthiphol, Sizhu Lu, Diogo Dias, Dr. Pavel Savitsky for providing me the PiggyBac cell expression system, which allows me to overcome the Tet2 expression issue in the project. I would like to thank those who share the DPhil adventure with me at Oxford: Payal, Sizhu, Elphire, Sam and Max.

I also owe a debt of gratitude to my colleagues within the group: Dr. Anandhakumar Chandran, Ashvina Segaran, Dr. Hiromi Tagoh, Dr. Marketa Tomkova, Dr. Martin Cusack, Dr. Michael McClellan, Olena Yavorska, Paolo Spingardi, and Dr. Sophie Kirschner, for their accompany and enthusiasm for science. I'm especially thankful to Dr. Hiromi Tagoh for sharing her depth of knowledge and scientific suggestions. I'm thankful to Dr. Martin Cusack for the mentorship in genomics data analysis. I'm thankful to Olena Yavorska, Paolo Spingardi, Dr. Sophie Kirschner for the participation in this project.

Last but not least, I would like to thank my parents for their support and encouragement despite being so far from home. I would like to thank my boyfriend Yaqiu for his understanding and support that accompany me during the scientific journey.

Contents

List of Figures

List of Abbreviations

2-HG 2-hydroxyglutarate
2-OG 2-oxoglutarate
5mC 5-hydroxymethylcytosine
5fC 5-formylcytosine
5caC 5-carboxylcytosine
A . . . . . Adenine
AITL . . . . . Angioimmunoblastic T-cell lymphoma(AITL)
AML . . . . . Acute myeloid leukemia
AP-MS . . . . . affinity-purification coupled to massspectrometry
APEX2 . . . . . Engineered plant ascorbate peroxidase
BAM . . . . . Binary sequence alignment/map (file type)
BED . . . . . Browser extensible data (file type)
BER . . . . . Base excision repair
BFU-E . . . . . Burst-forming unit-erythroid
BirA* . . . . . *Escherichia coli* biotin-ligase R118G mutant
BP . . . . . Biotin phenol
c . . . . . Cytosine
C . . . . . C-terminus
CD . . . . . Catalytic domain
CFU . . . . . Colony-forming unit assay
CFU-GEM . . . . . Colony-forming unit-granulocyte, erythroid, macrophage
CFU-GM . . . . . Colony-forming unit-granulocyte, macrophage
CFU-M . . . . . Colony-forming unit-macrophage
CGI . . . . . CpG Island
ChIP . . . . . Chromatin Immunoprecipitation
CIMP . . . . . CpG island methylator phenotype
CMML . . . . . Chronic myelomonocytic leukaemia
CpG . . . . . CpG dinucleotides
CTCF . . . . . CCCTC-binding factor
Cys-rich . . . . . Cysteine-rich
DKO . . . . . Double-knockout

DMR . . . . . Differentially methylated region
DNMT . . . . . DNA methyltransferases
DNA . . . . . Deoxyribonucleic acid
Dox . . . . . Doxycycline
DSBH . . . . . Double-stranded beta-helix
DSG . . . . . Disuccinimidyl glutarate
EGFP . . . . . Enhanced green fluorescence protein
ESC . . . . . Embryonic stem cell
FDR . . . . . False discovery rate
FH-Tet2 . . . . . Flag-HA tagged Tet2
FL . . . . . Full length
Flag . . . . . DYKDDDDK peptide
FPKM . . . . . Fragments per kilobase per million
gag . . . . . Viral packaging gene
G . . . . . Guanine
GOI . . . . . Gene of interest
HA . . . . . Human influenza hemagglutinin tag
H3K4me1 . . . . . Histone H3 lysine 4 monomethylation
H3K4me3 . . . . . Histone H3 lysine 4 trimethylation
H3K27me3 . . . . . Histone H3 lysine 27 trimethylation
H3K27ac . . . . . Histone H3 lysine 27 acetylation
H_2O_2 . . . . . Hydrogen peroxide
HPC . . . . . Hematopoietic precursor/progenitor cell
HSC . . . . . Hematopoietic stem cell
N . . . . . N-terminus
IAP . . . . . Intracisternal A particle (IAP)
IDH . . . . . Isocitrate dehydrogenase
Igf2 . . . . . Insulin-like growth factor type 2
IP . . . . . Immunoprecipitation
IP-MS . . . . . Immunoprecipitation coupled mass spectrometry
IRES . . . . . Internal ribosomal entry site
ITRs . . . . . Inverted terminal repeat sequences
KO . . . . . Knock-out
lncRNA . . . . . long non-coding RNA
LSK . . . . . Lineage negative, Sca1 positive, c-kit negative cells
LTR . . . . . Long terminal repeats
MDS . . . . . Myelodysplastic syndrome (MDS)
min . . . . . Minute
ml . . . . . Milliliter

mM . . . . . Millimolar
MPN . . . . . Myeloproliferative neoplasms (MPN)
MSCV . . . . . Murine stem cell virus
MSCVψ . . . . . Viral packaging signal
NLS . . . . . Nuclear localization signal
OGT . . . . . O-GlcNAc transferase
PBS . . . . . Phosphate buffered saline solution
PCA . . . . . Principal component analysis
PCR . . . . . Polymerase chain reaction
PGC . . . . . Primordial germ cells
PRC2 . . . . . Polycomb Repressive Complex 2
PPI . . . . . Protein-protein interaction
PS . . . . . Phospholipid phosphatidylserine
qPCR . . . . . Quantitative Polymerase Chain Reaction
RM . . . . . Restriction-modification
RNA . . . . . Ribonucleic Acid
RNA-seq . . . . . RNA sequencing
RRBS . . . . . Reduced representation bisulfite sequencing
RPKM . . . . . Reads per kilobase per million
rtTA . . . . . Reverse tetracycline-controlled transactivator
s . . . . . Seconds
SCF . . . . . Stem cell factor
SDS-PAGE . . . . . Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sgRNA . . . . . Single-guide RNA
Strep-HRP . . . . . Streptavidin-horseradish peroxidase
T . . . . . Thymine
TAB-seq . . . . . TET-assisted bisulfite-sequencing
TAPS . . . . . TET-assisted pyridine borane sequencing
TBST . . . . . Tris-buffered saline supplemented with Tween-20
TBP . . . . . TATA-box binding protein
TDG . . . . . Thymine-DNA-glycosylase
TET . . . . . Ten-eleven-translocation dioxygenase
TF . . . . . Transcription Factor
TKO . . . . . Triple knockout
TSS . . . . . Transcription start sites
TTS . . . . . Transcription termination sites
μL . . . . . Microliter
UHRF1 . . . . . Ubiquitin-like, containing PHD and RING finger domains 1
WGBS . . . . . Whole-genome bisulfite sequencing

WT . . . . . Wildtype

XCI . . . . . X chromosomeinactivation

Xist . . . . . X inactive specific transcript

Y2H . . . . . Yeast two-hybrid screening

1 Introduction

Contents

This introduction provides a background about TET2 function in hematopoiesis and its molecular mechanism in gene regulation. TET2 is a mammalian DNA modifying enzyme that oxidizes 5mC into 5hmC, resulting in active DNA demethylation. TET2 is a known tumor suppressor gene in human hematological disorders. Current studies about TET2 mainly focus on its enzymatic-activity dependent gene regulation mechanism and its function in normal and malignant hematopoiesis. I will first start the introduction by a brief description of the known DNA modifications in the vertebrate genome (chapter 1.1) and explain the basic properties and biological significance of 5mC and 5hmC (chapter 1.2 and chapter 1.3). Next, I will focus on the key molecule studied in this thesis, TET2 protein, by a description of TET2 function in human hematopoietic disorders and mouse hematopoiesis (chapter 1.4). Then I will introduce the current progress in understanding of the molecular mechanisms of TET2-associated gene regulation. In the discussion section, I will discuss the open questions in the field and explain how my interest in understanding TET2 gene regulation process and in identifying hematopoietic-relevant TET2 interactome are developed. And the final section will characterize the aims of this study.

1.1 DNA modifications in vertebrates

DNA is the genetic material in all known cellular organisms. The information in DNA is stored as a code made up of four types of nucleobases: guanine (G), adenine (A), cytosine (C) and thymine (T). A number of chemically modified nucleobases have been observed to occur naturally in the DNA in the living organisms (Figure 1.1). Methyl group can be added to either the carbon-5 (Hotchkiss, 1948) or nitrogen-4 positions of cytosine (Janulaitis et al., 1983), or the nitrogen-6 positions of adenine (Dunn and Smith, 1955). 5-methylcytosine (5mC) is the most common type of DNA modifications identified naturally in the genome, whereas the other two modified bases were found absent

or in low abundance in the eukaryotic genome (Liu et al., 2016; Ratel et al., 2006; Xiao et al., 2018). In the past decade, other forms of cytosine modifications have been identified in the mammalian genome including 5-hydroxymethylcytosine (5hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (He et al., 2011; Ito et al., 2011). Those novel cytosine modified bases are derived from the oxidation of 5mC and serve as a mechanism for DNA cytosine demethylation.

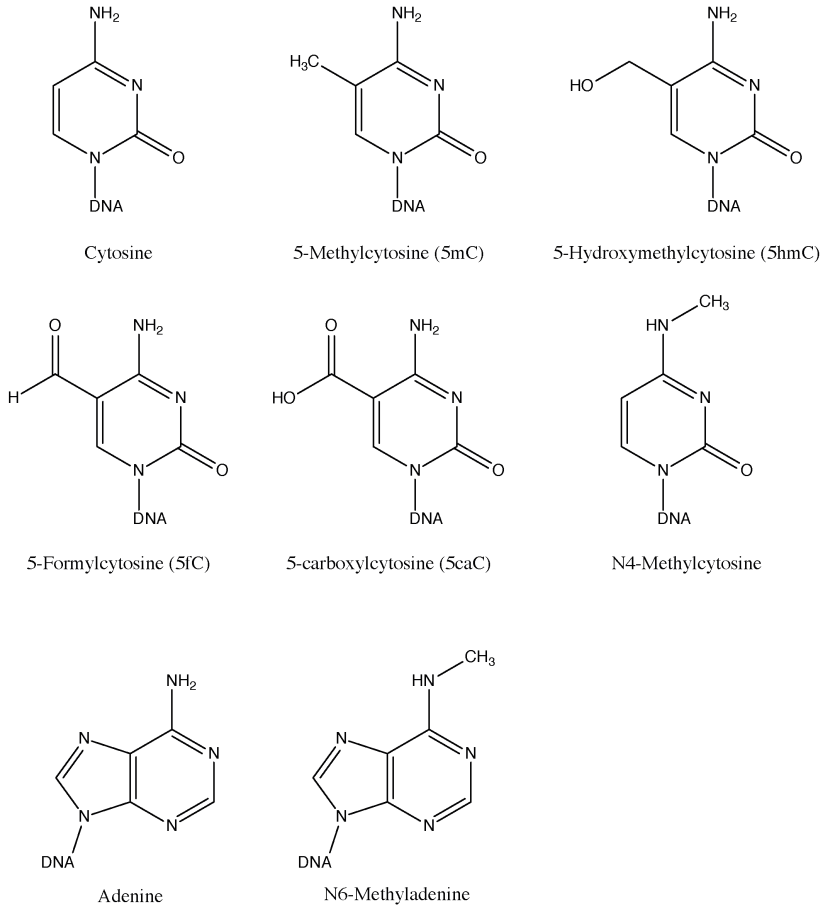


Figure 1.1: DNA modifications identified in living organisms. All modified bases with an exception of N4-methylcytosine have been identified in the mammalian genomes.

Cytosine methylation is the most common type of DNA modifications and it has been shown to play a variety of functions in different living organism. In bacteria, the modifications are deposited at specific sites by diverse types of methyltransferases and serve as a component of the restriction-modification (RM) system, which is used to

defend the bacteria DNA against the phage invasion (Wilson and Murray, 1991). In the RM system, one methyltransferase is usually paired with an associated restriction endonuclease which cleaves the DNA at a specific sequence, only if it is not methylated. The methylated bases in the bacteria genome can distinguish the host DNA from the foreign DNA and prevent the host genome from restriction enzyme digestion. However, eukaryotic organisms don't have the RM system as those observed in bacteria for host defense. DNA methylation in prokaryotes has also been proposed to play functions in other biological processes including DNA repair, DNA replication and transcriptional regulation (Palmer and Marinus, 1994).

In eukaryotic organisms, 5mC is the most abundant and well-studied DNA modification and it is also the first eukaryotic modified base to be identified. In 1948, several years before the discovery of DNA double helix structure, Hotchkiss observed a small and distinct peak in addition to the four canonical nucleobases (A, G, C and T) in the calf thymus DNA during the paper chromatography experiment. This material was originally referred to as 'epicytosine' since it has similar migration rate and absorption characteristics as cytosine (Hotchkiss, 1948). 5mC is widely observed in most eukaryotic species including fungi, plants and animals, despite it has been lost in some common model organisms such as fission yeasts *Schizosaccharomyces pombe*, budding yeasts *Saccharomyces cerevisiae*, worm *Caenorhabditis elegans*, and fruit fly *Drosophila melanogaster* (Raddatz et al., 2013; Zemach and Zilberman, 2010). 5mC is predominantly found at the symmetrical CpG dinucleotides in vertebrate genomes except for CpG islands which are GC-rich regions and are usually found in promoter regions of transcribed genes. However, in invertebrate, plants and fungi show distribution of methylation in non-CG context including CHH methylation (H = A, C, or T) and CHG methylation (Feng et al., 2010; Suzuki and Bird, 2008; Zemach et al., 2010). Interestingly, vertebrate organisms with CpG methylation undergo a global CpG depletion, where CpG is occurred only around 20% of the expected frequency in the genome (Lander et al., 2001). This comes as a cost of 5mC mutability (Cooper and Krawczak, 1989; Holliday and Grigg, 1993). After spontaneous deamination, unmethylated cytosine is converted into uracil and gets removed by the DNA repair machinery. Whereas 5mC gets deaminated to thymine, which

leads to a C to T transition. This might explain the reduced CpG dinucleotides level in the vertebrate genome.

The DNA modifications focused in this study are 5mC and its oxidized derivatives 5hmC, 5fC and 5caC. The following sections will explain the basics about those modifications mainly in mammlian system.

1.2 5mC, the 5th base in the genome

5mC is the first modified base identified in eukaryotic organisms, known as the 5th base in the genome in addition to the four standard nucleobases. 5mC is the most abundant DNA modifications in mammalian genome. The establishment and maintenance of methylation at carbon-5 cytosine is catalyzed by DNA methyltransferases (DNMT) family proteins. 5mC is usually deposited at CpG dinucleotides in genome and shows specific distribution pattern among different tissues. 5mC is tightly related to gene regulation in many physiological activities including normal developmental and malignant processes. In this section, I will introduce the DNMTs-mediated 5mC formation and the distribution pattern of 5mC in the genome and finally describe the known gene regulation activities of 5mC in mammalian system.

1.2.1 DNMT family enzymes establish and maintain the DNA methylation

There are six DNA methyltransferases (DNMT) family proteins identified in mammals, which are grouped into three families encoded by different genes: DNMT1, DNMT2, DNMT3(A/B/C/L). DNMT1, DNMT3A and DNMT3B are the major functional DNA methyltransferases and ubiquitously expressed in cells and tissues (Bestor et al., 1988; Okano et al., 1998). All three DNMT proteins have a similar methyltransferase domain in the C-terminus. The catalytic domain harbors 10 sequence motifs which are conserved within the bacteria cytosine-5 methyltransferases, and add the methyl group from S-adenosyl-L-methionine (SAM) to the carbon-5 position in the cytosine (Goll and Bestor, 2005; Lauster et al., 1989; Pósfai et al., 1989). The N-terminus of the three DNMT proteins is less conserved, which leads to the divergence of their methylation activity.

DNMT3C is a newly identified active methyltransferase only in rodent organisms. DNMT3C is derived from DNMT3B gene duplication and is responsible for methylating the young retrotransposons in male germ line (Barau et al., 2016). DNMT3L is a homologue to DNMT3A and DNMT3B, but it lacks the methyltransferase domain. DNMT3L has restricted function in germ cells and serve as a DNMT3A/3B cofactor (Bourc'his et al., 2001). Although DNMT2 is the most widely-distributed DNMT enzymes in living organisms and has an intact catalytic domain, little evidence has been shown on its methylation activity on DNA (Goll and Bestor, 2005). Instead, the Bestor group showed that DNMT2 enzyme is actually a RNA methyltransferase which methylates the aspartic acid transfer RNA (tRNAAsp) (Goll et al., 2006).

DNMT1 is the first eukaryotic DNMT protein to be identified (Bestor et al., 1988). DNMT1 is known as a maintenance DNA methyltransferase to enable the inheritability of DNA modifications upon DNA replication. This depends on the DNMT1 substrate preference for hemimethylated DNA and its protein interactions with DNA-replication related factors (Yoder et al., 1997a). Structure analysis on DNMT1-DNA complex revealed an autoinhibitory regulation of DNMT1 at unmethylated CpG sites on DNA, DNMT1 CXXC domain specifically binds the unmethylated DNA and exclude the unmodified DNA from the DNMT1 active site to inhibit the *de novo* DNA methylation (Song et al., 2011b). Hemimethylated DNA is usually formed during DNA replication where the parental strand is methylated and the daughter strand is not. UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1, also known as NP95 in mouse) preferentially binds to the hemimethylated DNA by its SRA (SET and RING associated) domain and recruits DNMT1 to replication fork to maintain the DNA methylation during DNA replication (Bostick et al., 2007). *Dnmt1* depletion in mice is embryonic lethal. *Dnmt1* homozygous mutant mouse embryos fail to develop at mid-gestation stage. (Li et al., 1992).

DNMT3A and DNMT3B are *de novo* DNA methyltransferases which catalyze the methylation at unmethylated cytosines. *De novo* methylation activities were examined by introducing the retroviral DNA into the embryonic stem (ES) cells. Retrovirus DNA

was methylated in either *Dnmt3a* or *Dnmt3b* single mutant cells, but not methylated in double mutant cells (Okano et al., 1999). Re-expression of Dnmt3a in *Dnmt3a* or *Dnmt3b* null cells restore the wildtype methylation level, which further confirmed the role of DNMT3 proteins in *de novo* DNA methylation (Chen et al., 2003a). *Dnmt3a* null mice are normal at birth but die around 4 weeks after birth. *Dnmt3b* null mice are not viable and die of developmental defects (Okano et al., 1999). These results indicate the important functions of *de novo* DNA methylation in normal developmental processes and the divergences of the two DNMT3 proteins in regulating cellular activities.

1.2.2 Genomic distribution of DNA methylation

The exploration in understanding of the genomic distributions of 5mC starts from the early studies using methylation-sensitive restriction enzyme HpaII (recognition site CCGG and can not cut methylated CpG) to digestion genomic DNA (Bird et al., 1985; Cooper et al., 1983). A discrete fraction of small DNA fragments were remained after HpaII digestion, suggesting the existence of genomic regions with high density of unmethylated CpG sites (Cooper et al., 1983). These unmethylated CpG-rich genomic regions accounts for less than 2% of the genome and are named as CpG islands (Bird et al., 1985). CpG islands are around 1 kb in length and overlap with promoter regions of 60-70% of all human genes including nearly all ubiquitously active genes (housekeeping genes) and 40% of the tissue-restricted expressing genes (Illingworth and Bird, 2009; Landolin et al., 2010). CpG islands are rarely methylated and are usually marked with histone H3 lysine 4 trimethylation (H3K4me3) in actively transcribed gene promoters, where H3K4me3 prevents the binding of DNMT3A and DNMT3B to chromatin and thus affect the *de novo* DNA methylation (Mikkelsen et al., 2007; Otani et al., 2009). CpG island promoters of non-transcribed genes are usually repressed by Polycomb proteins and characterized by the colocalization of histone H3 lysine 27 trimethylation (H3K27me3) (Mikkelsen et al., 2007; Taberlay et al., 2011). CpG island methylation at gene promoters is usually required for the long-term gene repression such as genes involved in X chromosome inactivation, maternal imprinting and germline-specific genes.

The advances of genome-wide 5mC mapping methods have greatly facilitated our understanding towards the distribution and function of 5mC in the genome. Current gold standard of 5mC mapping method is whole-genome bisulfite sequencing (WGBS) which provides the single base-resolution locations of the modified base in the genome (Lister et al., 2009). This method utilizes sodium bisulphite to convert unmethylated cytosines to uracils, whereas 5mC remains unchanged. After polymerase chain reaction (PCR) amplification, unmethylated cytosine is read as thymine by sequencing, 5mC is amplified as cytosine. Reduced Representation Bisulfite Sequencing (RRBS) is another bisulfite conversion-based method to map DNA methylation level but with less sequencing cost (Meissner et al., 2005). RRBS includes the specific restriction digestion and size selection steps to enrich the CpG dinucleotide-containing DNA fragments before bisulfite conversion which greatly saves the numbers of reads required for sequencing and reduces the cost dramatically. Although RRBS has the limitation in examining the whole genome-wide methylation level, it can capture nearly 80% of human CpG islands and 50% of gene promoters (Gu et al., 2011a). However, the bisulfite conversion-based methods have some drawbacks which might have to be considered during the application. Bisulfite conversion can not distinguish 5mC and 5hmC since they are both unconverted and are read as cytosine by sequencing. Bisulfite treatment itself is a harsh reaction and can degrade the DNA samples, which might limit its use on low-input material. On the other hand, 5mC is relatively low-abundance (5% of total C) in the genome, so that the converted DNA has reduced sequence complexity since most of the cytosine (unmodified) are converted to uracils and only few cytosines (5mC and 5hmC) remain in the genome, this could cause issues in sequencing quality and genome coverage. A recently developed whole genome-wide 5mC mapping method, TET-assisted pyridine borane sequencing (TAPS), has been designed to overcome those problems by replacing sodium bisulfite with a milder chemical, pyridine borane (Liu et al., 2019b). In this method, only the modified cytosine bases were converted to thymine, and the unmodified cytosine remain unchanged.

Facilitated by the WGBS approach, unexpectedly, Lister and colleagues observed the abundant DNA methylation in non-CG contexts (mCHG and mCHH, H = A, C or T) in

human embryonic stem cells H1 cells and this consists of nearly 25% of all methylated cytosines (Lister et al., 2009). The non-CG methylation is suggested to be present as a characteristic of the embryonic stem cell. mCG showed less distribution at CpG islands and TSS, followed by an increase in the 5' UTR and exhibited a similar level in exons, introns and the 3' UTR. Similarly, CHG and CHH methylation also showed the decreased density at the TSS and returned to the same level as the 2 Kb upstream the TSS at the end of the 5' UTR. However, the non-CG methylation density at exons, introns and 3' UTRs were two fold higher than the 2 Kb upstream. Gene ontology analysis revealed genes enriched with non-CG methylation were associated with RNA processing, RNA splicing, and RNA metabolic processes, suggesting a specific role of non-CG methylation in RNA-related biological processes.

In addition to the gene repression function of DNA methylation at CG-promoter, CG methylation in gene exons could result in C (cytosine) to T (thymidine) transition mutations, which promote cancer-causing mutation formation (Rideout et al., 1990). DNA methylation at gene body has been linked to active transcription. This is supported by the observation of the gene body–specific DNA methylation on the active X chromosome (Hellman and Chess, 2007; Wolf et al., 1984). DNA methylation at enhancer is beginning to gain more focus. The relationship between DNA methylation and enhancer was initially studied in the mouse methylomes mapping in stem cells and neuronal progenitors by Stadler and colleagues. (Stadler et al., 2011). Enhancers were found enriched in the low-methylated regions (LMRs) identified in the study and exhibited a dynamic state during cell differentiation. DNA methylation at enhancer regions was suggested to impair the enhancer activity, this is supported by the study on the binding of the glucocorticoid receptor (GR) to distal regulatory elements. (Wiench et al., 2011). GR-bound distal elements were characterized by low CpG density, and showed rapid DNA methylation changes upon glucocorticoids induction. More importantly, DNA methylation was shown to negatively affect GR-DNA interactions *in vitro*.

1.2.3 5mC as a repressive mark

5mC is known to be involved in gene silencing of the following cellular processes : X-chromosome inactivation, gene imprinting, and transposons regulation.

X-chromosome inactivation: X chromosome inactivation (XCI) is the process by which female mammalian cells randomly inactivate one of the two X chromosomes, to balance the gene dosage between XX females and XY males (Lyon, 1961). The establishment and maintenance of gene silencing is under control of the master regulator, X inactivate specific transcript (Xist) (Brockdorff et al., 1991; Brown et al., 1991). Xist is a long non-coding RNA (lncRNA) that coats the inactivated X chromosome and recruits the Polycomb protein complex for gene silencing (Brockdorff, 2011). DNA methylation is thought to be involved in the late stage of the XCI and provides a locking mechanism for the long-term gene repression (Gendrel et al., 2012; Lock et al., 1987; Sado et al., 2004). Evidence supporting this comes from a previous study on mouse *Hprt* gene in which the DNA methylation is shown to occur specifically at gene loci on inactivated X chromosome after XCI (Lock et al., 1987). Studies on *Dnmt1*-deficient mouse embryonic stem cells (ESCs) and mouse embryos suggest the maintenance DNA methylation is dispensable for the establishment of XCI but is required for the stable repression of Xist promoter on active X chromosome in differentiated cells (Beard et al., 1995; Panning and Jaenisch, 1996). The contribution of *de novo* DNA methylation was assessed by a recent study on *Dnmt*-deficient mice (Gendrel et al., 2012). DNMT3b was identified as the key *de novo* DNA methyltransferase involved in the XCI CpG island methylation, since little changes in XCI DNA methylation was observed in neither *Dnmt3a*-null nor *Dnmt3L*-null XX embryos .

Genomic imprinting: Genomic imprinting is a mammalian phenomenon that a subset of genes (referred to as imprinted genes) in the diploid cells restrict its expression to only one parental copy and silence the inactivated copy by DNA methylation. This mono-allelic and parental-specific gene expression which is established in gamete formations behave stable and inheritable after fertilization where there is a global DNA methylation reprogramming event. Imprinted genes are mostly distributed in clusters where 3-12

imprinted genes are under control of the same genomic region called imprinting control region (ICR). Notably, 5mC is a heritable epigenetic mark and can be transmitted after embryogenesis by maintenance enzymes (Barlow and Bartolomei, 2014)

A typical example of DNA methylation in genomic imprinting is the Igf2(Insulin-like growth factor type 2)-H19 cluster. *Igf2* is a paternally expressed imprinted gene (DeChiara et al., 1991; Ferguson-Smith et al., 1991) whereas *H19* gene encodes a long non-coding RNA (lncRNA) and is a maternally expressed imprinted gene (Bartolomei et al., 1991). Igf2 is located at the upstream of H19 and utilizes the enhancer downstream of H19. The ICR which sits between Igf2 and H19 is bound by CCCTC-binding Factor (CTCF) only if the ICR is not methylated. On the maternal allele, CTCF binds to the unmethylated ICR and blocks the activation by H19 enhancer on Igf2. H19 is expressed as normal. On the paternal allele, the blocking activity of CTCF binding is abolished by ICR methylation, so that Igf2 is activated by H19 enhancer, H19 expression is turned off due to the spread of methylation from the ICR. (Bell and Felsenfeld, 2000; Hark et al., 2000). Furthermore, the methyl CpG-binding domain (MBD) protein, MeCP2, is shown to be recruited to the H19 ICR and silences the paternal *H19* allele together with histone deacetylase activity (Drewell, 2002).

Retrotransposons elements repression: Transposable elements, which take up $\sim$70% of the human genome (de Koning et al., 2011), are the main targets of DNA methylation. DNA methylation at gene promoters provides an important mechanism for transposable element silencing and reduces the mobility of those movable elements. Intracisternal A particle (IAP) retroviruses are the most active transposons in mouse genome. IAP retroviruses are usually transcriptionally silenced and heavily methylated in somatic tissue. IAP was derepressed and highly active in *Dnmt1* knock-out mouse embryos (Walsh et al., 1998). Mouse studies revealed that DNMT3L and DNMT3C were both required for the *de novo* methylation of IAP in male germ cells (Barau et al., 2016; Bourc'his et al., 2001).

1.3 5hmC, the 6th base in the genome

5hmC is another stable modified base identified in the mammalian genome. The discovery of TET enzymes link the biogenesis of 5hmC to 5mC oxidation. TET-mediated 5mC oxidation offers a new mechanism for the active DNA demethylation. In this section, I will firstly introduce the basics about TETs protein and its 5mC oxidation activity. Second, I will explain the current understanding of active and passive DNA demethylation. Then I will summarize the current methods for genome-wide 5hmC mapping. Finally, I will discuss the role of 5mC oxidation derivatives beyond as the DNA demethylation intermediates.

1.3.1 TET family enzymes catalyze the 5mC oxidation

5hmC was first discovered in T-even bacteriophages in 1953 (Wyatt and Cohen, 1953), the proof of its existence in mammalian genome came in nearly five decades later. In 2009, two independent groups observed the natural occurrence of 5hmC in mouse Purkinje neurons (Kriaucionis and Heintz, 2009) and in mouse embryonic stem cells (Tahiliani et al., 2009) respectively. In the meantime, ten-eleven translocation (TET) faimly proteins were identified as the enzymes responsible for the 5hmC formation through 5mC oxidation (Tahiliani et al., 2009).

The name of TET proteins is derived from its initial discovery as a MLL gene fusion partner in acute myeloid leukemia containing the t(10;11)(q22;q23) translocation (Lorsbach et al., 2003; Ono et al., 2002). The identification of TET enzymatic activity comes from the computational screens to search for the mammalian homologues of Trypanosome base J binding proteins (JBP) (Iyer et al., 2009). JBP proteins oxidize the methyl group of thymine to 5-hydroxymethyluracil, an intermediate for the production of base J (Borst and Sabatini, 2008). TET proteins were found to utilize 5mC as the oxidation substrate to form 5hmC. The oxidation reaction requires 2-oxoglutarate and Fe(II) as cofactors. Later studies revealed that TET proteins can further oxidize 5hmC to 5fC and 5caC *in vivo* (He et al., 2011; Ito et al., 2011) (Figure 1.2).

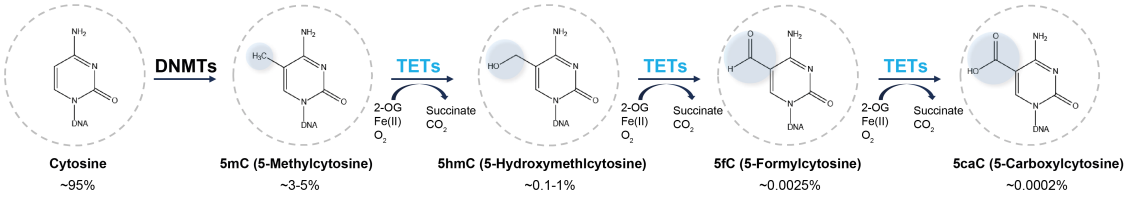


Figure 1.2: TET-catalyzed 5mC oxidation. DNMT, DNA methyltransferase. TET, ten-eleven translocation proteins. Numbers indicate the DNA modifications abundance relative to total amount of cytosine in mouse embryonic stem cells.

The three mammalian TET proteins, TET1, TET2 and TET3, belong to the 2-oxoglutarate- and Fe(II)-dependent oxygenase superfamily (Tahiliani et al., 2009). All members of the TET proteins have been shown to have the 5mC oxidation abilities (Ito et al., 2011). All TET proteins share a conserved structure in their C-terminus catalytic domain (Figure 1.3), which consists of a conserved double-stranded beta-helix (DSBH) domain, a cysteine-rich (Cys-rich) domain and the binding sites for its cofactors Fe(II) and 2-oxoglutarate (2-OG) (Pastor et al., 2013). Both TET full-length and C-terminus catalytic domain (CD) truncated proteins can catalyze the 5mC oxidation but it is suggested that the full-length protein is more enzymatically active than the CD protein (He et al., 2011; Hu et al., 2013). The unique cysteine-rich domain distinguishes TET proteins from the rest of TET-JBP dioxygenases proteins, and the Cys-rich domain together with DSBH domain are required for the integrity of catalytic activity of TET proteins. Crystal structure analysis shows that TET2 catalytic domain preferentially binds the methylated cytosine in a CpG context (Hu et al., 2013) and TET2 is less active on 5hmC-DNA and 5fC-DNA than on 5mC-DNA substrates (Hu et al., 2015), which suggests that 5hmC has the potential to be a stable epigenetic mark on the genome. Structure analysis also shows that the methyl-group of 5mC is not involved in DNA-TET interaction which provides an insight for the observation that TET proteins could further oxidize 5hmC into 5fC and 5caC (Hu et al., 2013).

TET1 and TET3 both have a N-terminus CXXC DNA binding domain which is responsible for the the recruitment of TET proteins to specific genomic regions. CXXC domain is present in many chromatin modifiers such as DNMT1, CFP1 (CXXC finger protein 1), MLL (mixed lineage leukaemia protein) and KDMA (Lys-specific demethylase)

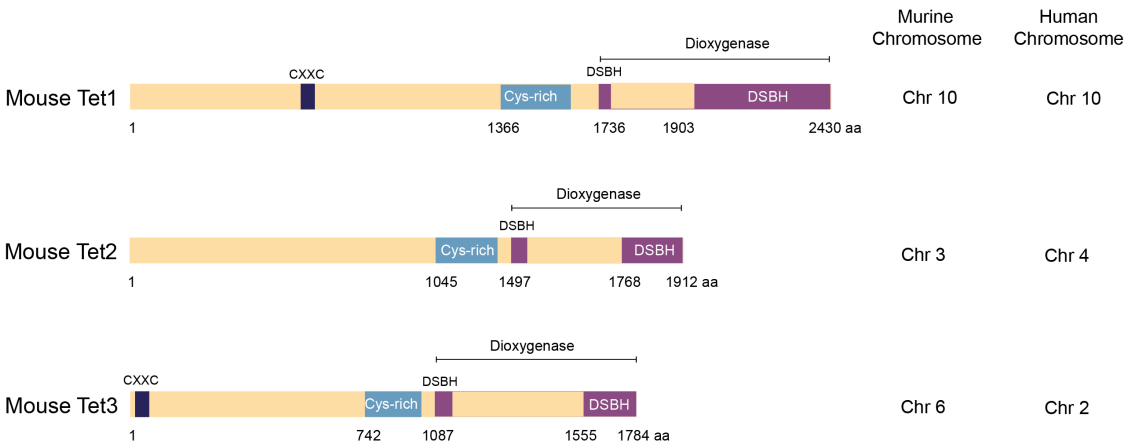


Figure 1.3: Domain structure of Tet proteins. Schematic representation of murine Tet family proteins. The C-terminal catalytic domain conserved in all Tet proteins consists of a cysteine-rich (Cys-rich) domain, the double-stranded beta-helix (DSBH) domain, and the binding sites for Tets cofactors, the Fe(II) and 2-OG. Tet1 and Tet3 have the CXXC domain at their N terminus. The numbers indicate the location of the functional domain.

family proteins and these CXXC domains specifically bind the unmethylated CpG. The CXXC domain of TET1 and TET3 have a truncated linker region and a differnt DNA-binding loop, which differ them from other known CXXC domains (Long et al., 2013). *In vitro* assays show different conclusions on TET1 CXXC DNA binding ability. TET1 CXXC domain is reported to have no DNA binding activity (Frauer et al., 2011) or it prefers to bind CpG DNA without preference for unmodified or modified DNA substrates (Xu et al., 2011b; Zhang et al., 2010). Genome-wide analysis revealed that TET1 prefer to bind CpG-rich regions and localize at transcription start sites (TSS) of CpG-rich promoters (Williams et al., 2011; Xu et al., 2011b). TET3 CXXC domain is observed to bind both non-CpG (CpA, CpT, CpC) and CpG DNA with a slight preference for CpG DNA in the *in vitro* assay. Genome-wide analysis indicates that the TET3 CXXC domain can bind the unmodified C regardless of the consideration if the following base is guanine (Xu et al., 2012). TET2 CXXC domain was separated from *TET2* gene which underwent chromosomal inversion during the evolution, and become an independent gene called *IDAX (CXX4)* (Ko et al., 2013). IDAX preferentially binds unmethylated CpG *in vitro*, and localizes at CpG islands and CpG-rich promoters in genomic DNA. IDAX physically interacts with the catalytic domain of TET2 protein and recruits TET2 to DNA. Interestingly, The recruitment of TET2 by IDAX leads to a caspase-dependent

TET2 protein degradation and IDAX is a negative regulator of TET2 enzymatic activity. The negative regulation of CXXC domain on TET activity is also observed in TET3 protein but not TET1. TET3 CXXC activates the caspase pathway after TET3 protein gets recruited to DNA by CXXC domain, resulting in the down-regulation of protein level. However, IDAX and TET2 are not always expressed in the same cell, such as the 293 T cells which is used for the original IDAX-TET2 protein interaction identification. So that the mechanisms for TET2 genomic recruitment remain still unclear.

1.3.2 Active and passive DNA demethylation

Although DNA methylation is a stable and inheritable epigenetic mark, 5mC can undergo genome-wide dynamic changes during early developmental stages. Global loss of 5mC is observed in pre-implantation embryos and developing primordial germ cells (PGCs), and global gain of 5mC occurs in post-implantation embryonic stage (Hajkova et al., 2002; Hill et al., 2014; Luo et al., 2018; Mayer et al., 2000; Oswald et al., 2000). Global DNA demethylation is important for the establishment of pluripotency and gene imprinting in developing embryos and for removing parental-inherited gene imprints in developing PGCs (Luo et al., 2018). Two types of mechanisms, passive DNA demethylation (Figure 1.4A) and active DNA demethylation (Figure 1.4B), are proposed to be responsible for the removal of methyl group in cytosine.

Passive DNA demethylation refers to the replication-dependent 5mC dilution upon the lack of the maintenance of the DNA methylation (Figure 1.4A). During DNA replication, the maintenance DNA methyltransferase DNMT1 in co-operation with UHRF1 establishes cytosine methylation at newly synthesized DNA strand to restore the symmetrical CpG DNA methylation pattern (Bostick et al., 2007). An example of passive DNA demethylation is the gradual loss of 5mC in the zygotic maternal genome after several cleavage divisions in the early preimplantation embryogenesis (Mayer et al., 2000; Oswald et al., 2000). Later studies revealed an inefficient maintenance of methylation in the maternal DNA due the absence of Dnmt1 (Howell et al., 2001). Instead, mouse oocytes and pre-implantation embryos express a DNMT1 splicing variant, DNMT1o. Unlike DNMT1 which is exclusively nuclear in the cell, DNMTo is largely localized in

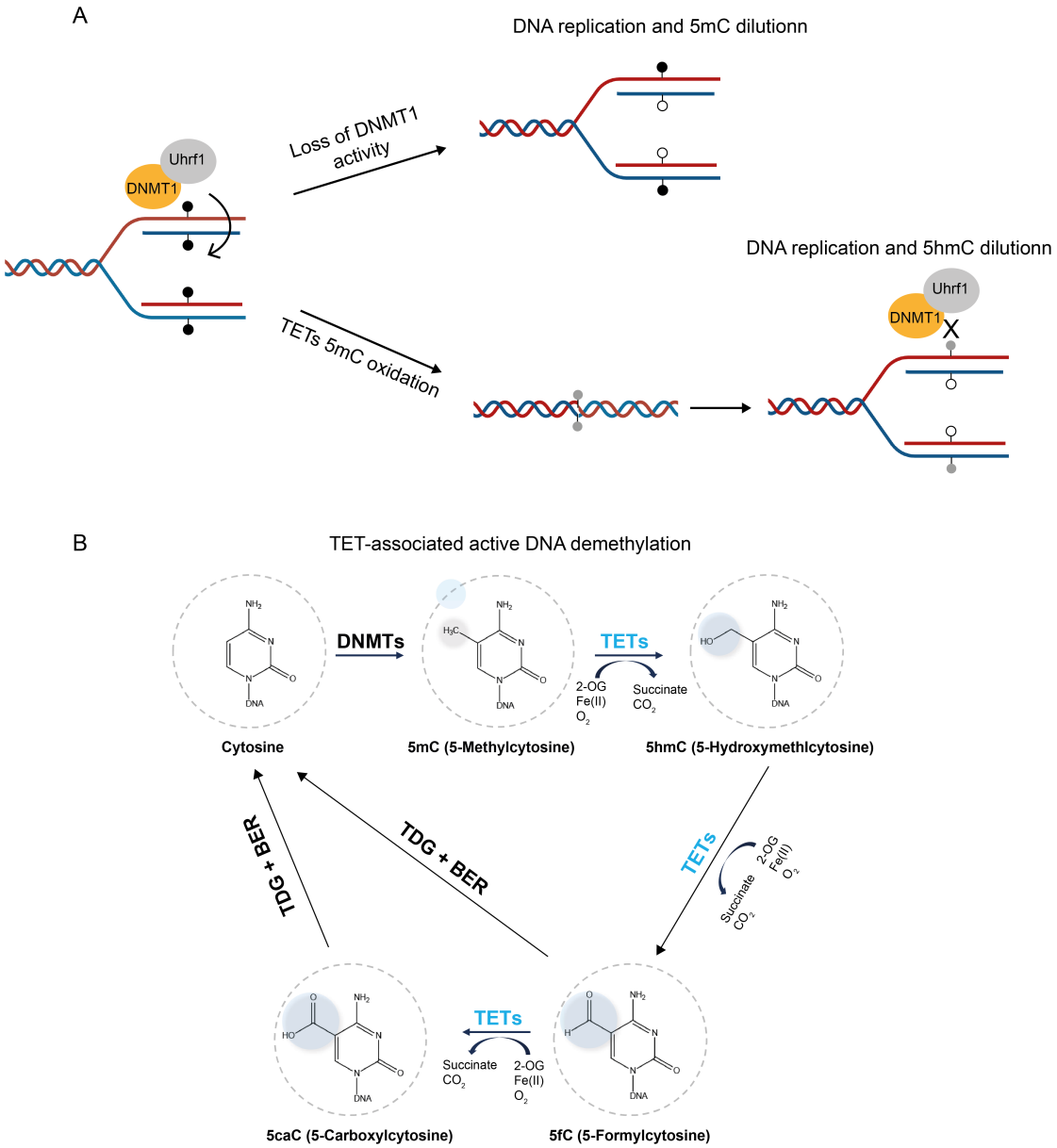


Figure 1.4: Active and passive DNA demethylation. (A) Two types of passive DNA demethylation pathways. DNMT1 is recruited by Uhrf1 to the hemi-methylated DNA and is responsible for the methylation maintenance during DNA replication. Loss of DNMT1 activity results in 5mC dilution during DNA replication , eventually leading to DNA demethylation. On the other hand, TETs oxidize the 5mC into 5hmC. 5hmC inhibits the methylation maintenance, thus results in 5hmC dilution during DNA replication, eventually leading to DNA demethylation. Black dot, 5mC; gray dot, 5hmC; white dot: C. (B) Cytosine is methylated by DNMTs. The methylated cytosine (5mC) can be oxidized by TETs into 5hmC, and further converted to 5fC and 5caC. The latter two modified bases are removed from the DNA via TDG glycosylase-BER repair pathway.

cytoplasma in the oocyte and pre-implantation embryos. As a result, 5mC on the maternal genome is diluted out through the passive replication-dependent manner. Moreover,

5hmC which is derived from TET-mediated 5mC oxidation can interfere the passive DNA demethylation by inhibiting the DNMT1 enzymatic activity. *In vitro* assay shows that DNMT1 has less efficiency in methylating the 5hmC-containing DNA (5hmCG:GC) than hemi-methylated DNA (5mCG:GC), which results in a replication-dependent 5hmC dilution and facilitates the passive DNA demethylation (Valinluck and Sowers, 2007). However, in mammals, the loss of 5mC in paternal genome of the early zygote is a rapid process and occur just a few hours after fertilisation, before the onset of the DNA replication (Mayer et al., 2000; Oswald et al., 2000) Similarly, PGCs in both male and female embryos undergo a rapid erasure of DNA methylation following their entry into the genital ridge (Hajkova et al., 2002). An replication-independent DNA demethylation mechanism is proposed to exist with the consideration of the kinetics of the rapid global 5mC loss in both developmental processes.

The discovery of TET proteins, 5hmC and its oxidation derivatives uncovered the process of enzyme-involved active DNA demethylation in mammalian genome (He et al., 2011; Ito et al., 2011; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009) (Figure 1.4B). All three TET proteins, TET1, TET2 and TET3 can catalyze the oxidation of 5mC into 5hmC, and further oxidize 5hmC into 5fC and 5caC, the latter two modified bases can be recognized and get excised by TDG (thymine-DNA-glycosylase). The resultant abasic site is replaced by unmodified cytosine through the BER (base excision repair) pathway (He et al., 2011; Maiti and Drohat, 2011; Weber et al., 2016), thus eventually leading to a replication-independent active DNA demethylation. Tet1 and Tet2 are responsible for the germline DNA methylation reprogramming during embryogenesis (Hackett et al., 2013; Vincent et al., 2013. Tet3-mediated 5mC oxidation is shown to be responsible for the rapid 5mC loss in paternal genome in the early pre-implanted embryos (Gu et al., 2011b), whereas, maternal genome is protected from the active DNA demethylation by Stella (also known as Dppa3) protein (Nakamura et al., 2007). Both passive and active DNA methylation can exist and function together in the same physiological scenario, such as, in mouse early developing embryos, although passive demethylation is thought to be the dominant mechanism for the maternal genome DNA reprogramming, recent studies observed 5hmC, 5fC and 5caC in both paternal and maternal DNA in the mouse zygote

just following the fertilisation, which suggests that Tet3-related active demethylation could also contribute to the global 5mC changes in the maternal genome (Guo et al., 2014a; Shen et al., 2014; Wang et al., 2014).

1.3.3 Genomic distribution of 5hmC and its derivatives

To gain a better understanding of the biological functions of 5hmC and its oxidized derivatives, many studies have been focused on the characterization of the tissue distribution, genomic localization and dynamic changes of those oxidized bases.

Global abundance of 5hmC

In contrast with 5mC, 5hmC shows a tissue-specific distribution pattern and is present in a variety of tissues (1%–5% of 5mC) with the highest levels identified in brain neuronal cells (10%–40% of 5mC) (Globisch et al., 2010; Kriaucionis and Heintz, 2009). 5hmC is widely observed in all brain regions with a range of distribution at 0.3%-0.7% of all dC (5mC: 3%-5% of all dC). The highest 5hmC levels are found in hippocampus and cortex, which are the brain areas for higher cognitive functions (Kriaucionis and Heintz, 2009; Münzel et al., 2010). Notably, 5hmC is nearly 40% of 5mC in Purkinje cell DNA (Kriaucionis and Heintz, 2009). 5hmC distribution is thought to be determined by cell proliferation rate (Bachman et al., 2014). 5hmC is usually abundant in differentiated, non-proliferating cells such as brain neuronal cells, and is low in high-proliferative tissues, such as spleen and thymus (Bachman et al., 2014; Globisch et al., 2010). Although 5hmC abundance is just around 5% of 5mC levels in mouse ESCs, considering the cell fast-dividing speed, 5hmC is still highly produced in mESCs (Ito et al., 2011; Tahiliani et al., 2009). Quantitative mass spectrometry analysis shows that 5fC and 5caC can exist in a variety of tissues in mice with the highest level observed in brain tissue. 5fC also shows a tissue-specific distribution pattern and is more abundant in slow-proliferating tissue (Bachman et al., 2015; Ito et al., 2011). In mouse ESCs, 5fC is $\sim 0.03\%$ of 5mC and 5caC is even more rarely observed and is around $\sim 0.01\%$ of 5mC (Ito et al., 2011; Pfaffeneder et al., 2011). *In vitro* biochemical and biophysical analysis indicate that TET proteins prefer to utilize 5mC-containg DNA as oxidation substrates than to oxidize

5hmC or 5fC (Hu et al., 2015; Ito et al., 2011), which might explain why there are more 5hmC deposited in the genome. Moreover, 5fC and 5caC are recognized and removed by TDG-BER pathway to achieve active DNA demethylation, which further explain their low frequency observed in cells. There is no clear correlation among the levels of 5hmC, 5fC and 5caC in cells since although mouse ESCs has less 5hmC (6-folds) than brain cortex, 5fC is more enriched in mESCs (Ito et al., 2011).

Genome-wide distribution of 5hmC and its derivatives

Genome-wide mapping methods: There are mainly two types of modified base mapping methods frequently used in most studies. One is based on the enrichment of modified bases by either using the modification-specific antibody (hMeDIP-seq, hydroxymethylated DNA immunoprecipitation-sequencing) or chemical conversion to facilitate the enrichment (hMe-Seal, CMS-seq) (Ficz et al., 2011; Pastor et al., 2011; Song et al., 2011a; Wu et al., 2011a). However, a recent study has revealed a problem while using the immunoprecipitation based method. Lentini and colleagues observed the enrichment of short tandem repeats (STRs) by the the antibody IgG non-specific binding during the immunoprecipitation, and surprisingly, STRs could account for nearly 50%-99%of enriched regions in 5mC DIP studies. (Lentini et al., 2018). In addition, DIP-seq methods also have limitations in identifying the exact location of the modified sites and the abundance of the DNA modifications. The development of bisulfite-sequencing (BS-seq)-based method provides a single base-resolution analysis of the modified sites in genome. One of the drawback of BS-seq is that it can not distinguish 5mC and 5hmC since both are read as cytosine during the sequencing. TAB-seq (TET-assisted BS-seq) was developed to specifically map only 5hmC by using the recombinant TET enzymes to oxidize all 5mC into 5caC, and 5hmC is protected by b-glucosyltransferase (bGT) from TET oxidation, thus after bisulfite conversion, 5hmC remain unchanged and is read as C by sequencing (Yu et al., 2012b). The other BS-seq-based 5hmC mapping method, oxBS-seq (oxidative bisulfite sequencing), oxidizes 5hmC to 5fC using potassium perruthenate, so that after bisulfite conversion, only 5mC remain unchanged and is read as C. 5hmC

signal can be obtained by doing a subtraction between the oxBS-seq and traditional BS-seq (Booth et al., 2012).

5hmC genomic distribution in mouse ESCs: A lot of studies have been published by either using the antibody-enrichment methods (hMeDIP-seq) or BS-seq based methods to capture the distribution features of 5hmC in mouse ESCs since it has relatively high abundances of all oxidized bases. 5hmC is found in promoter regions with low to medium levels of CpG density and with H3K4me3 enrichment (Ficz et al., 2011; Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011a; Yu et al., 2012b). Moreover, 5hmC is relatively associated with promoters of lowly expressed genes (Pastor et al., 2011; Wilson and Murray, 1991; Wu et al., 2011a; Xu et al., 2011b). 5hmC is also found located at the bivalent gene promoters which are characterized with both H3K27me3 and H3K4me3. (Pastor et al., 2013; Wu et al., 2011a; Yu et al., 2012b). Most of 5hmC peaks are reported to be found in gene bodies with a gradual increase from transcription start sites (TSSs) toward transcription termination sites (TTSs) and a quick drop around the TTSs (Ficz et al., 2011; Wu et al., 2011a). 5hmC is relatively enriched in enhancers and the distribution of 5hmC at distal regulatory regions shows a bias towards low CpG density (Pastor et al., 2011; Yu et al., 2012b). Poised (H3K4me1 only) and active enhancers (H3K4me3 and H3K27ac) are suggested to have more 5hmC distribution. 5hmC is actually depleted at the expected transcription factor-binding sites in the distal regulatory elements and found enriched at the adjacent regions (Yu et al., 2012b).

1.3.4 Role of 5hmC and derivatives beyond as the demethylation intermediates

5hmC and 5fC are shown to be stable in post-mitotic tissues, such as neuronal cells, by stable isotope labeling method in mice (Bachman et al., 2014, 2015). These results suggest that 5hmC and its derivatives have a function on DNA that goes beyond as the DNA demethylation intermediates. Moreover, mass spectrometry experiments have identified a lot of transcription-related and chromatin-associated proteins that can interact and recognize the 5mC oxidized bases *in vivo*, suggesting a role of 5hmC and its oxidation derivatives in transcription regulation and DNA repair processes (Iurlaro et al., 2013;

Spruijt et al., 2013). However, there is no clear or direct correlation between gene expression changes and 5hmC alteration in mouse ES cells and hematopoietic stem/progenitor cells (Ficz et al., 2011; Pastor et al., 2011; Williams et al., 2011; Zhao et al., 2015). In the DNA pulled-down experiments searching for proteins interacting with modified bases, 5mC and 5hmC are shown to recruit different proteins in mESC and the interaction between modified bases and proteins are cell-type specific. Neuronal progenitor cells (NPCs) show different protein binding profile of 5mC and 5hmC compared with mESCs, Uhrf2 was identified as a 5hmC-specific reader in NPCs. However, there is no any specific 5hmC reader identified so far since most of the identified 5hmC reader can also bind 5mC with lower affinity (Spruijt et al., 2013). Global run-on sequencing (GRO-seq) experiments indicate that 5fC and 5caC in the gene body can slow down the RNA polymerase II (Pol II) during transcription, which cause retarded transcription elongation. Crystal structure analysis finds that Pol II can sense the modified bases by its epi-DNA recognition loop. 5fC and 5caC, but not 5mC and 5hmC in the DNA could compromise the Pol II GTP incorporation ability (Kellinger et al., 2012; Wang et al., 2015a).

1.4 DNA modifications in normal and malignant physiological processes

DNA modified bases in the mammalian genome provide an additional layer of information in regulating the gene expression pattern in different tissues and cells. DNA methylation shows dramatic changes during epigenetic reprogramming processes involved in early embryogenesis and germline cell development. DNMT and TET family proteins are reported to play an important role in basic developmental processes including embryogeneis and neurological development. DNMT3A and TET2 mutations are frequently observed in patients with hematological malignancies, indicating a role of DNA modifications in tumorigenesis. Importantly, cancer cells usually exhibit aberrant DNA methylation pattern and CpG island methylator phenotype (CIMP) is a hallmark for a subtype of cancers. In this section, I will introduce the current understanding of the function of DNA modifications in normal and malignant physiological processes.

1.4.1 DNMT proteins in mouse embroyo development and human diseases

DNMT proteins in mouse embryogenesis and PGC

DNMT family proteins, DNMT1, DNMT3A/3B, are the main mammalian methyltransferases responsible for the maintenance and *do novo* establishment of cytosine methylation respectively. Mice carrying each individual *Dnmt* gene mutation all show developmental defects in early embryogenesis and fail to survive before birth or die soon after birth. *Dnmt1* depletion in mice is embryonic lethal (Li et al., 1992). *Dnmt1* homozygous mutant mouse embryos display severe developmental delay, and fail to develop at mid-gestation stage. Histology analysis on the mutant embyros show that major organs such as heart, brain, and in some cases forelimb buds, were present but less well developed. Intracisternal A particle (IAP) retrovirus is one of the most transpositionally active transposon elements and is normally transcriptionally silent in cells. IAP transcripts levels are found upregulated 50-100 fold in *Dnmt1-/-* mouse embryos, which suggest a role of DNMT1 in restraining transposon activity (Walsh et al., 1998). *Dnmt3a*-mutant mESCs and *Dnmt3b*-mutant mESCs both behave similar as the wildtype (WT) mESCs. *Dnmt3a*-/- mice show normal development at birth but die 4 weeks after birth. In contrast, *Dnmt3b* depletion in mice is embryonic lethal. *Dnmt3b*-/- mouse embryos have severe developmental defects including growth impairments and rostral neural tube defects. *Dnmt3a* and *Dnmt3b* double mutant homozygous mouse embryos are not viable and show more severe developmental defects than those of either single mutant mice, which suggests that *Dnmt3a* and *Dnmt3b* have functional overlaps in mouse early embryogenesis. The two *de novo* cytosine methyltransferases also show non-overlapping functions in development, Dnmt3b is specifically required for the methylation of centromeric minor satellite repeats (Okano et al., 1999). Dnmt triple knockout (TKO) undifferentiated mESCs show comparable cell growth with undifferentiated WT cells but TKO cells have impaired growth upon embryoid body differentiation, this suggest CpG DNA methylation is not important for the self-renewal ability of undifferentiated cells (Tsumura et al., 2006). DNA methylation is also involved in the gene imprinting regulation of maternal and paternal genomes during gametogenesis. Early study shows that Dnmt3L, a Dnmt3

homologous gene, is required for the gene imprinting establishment in maternal genome (Hata et al., 2002). Dnmt3L binds and cooperates with Dnmt3A and DnmtB to establish maternal imprints in mice.

DNMT proteins in human diseases

DNMT1 mutations are observed in patients with neurological disorders, such as hereditary sensory and autonomic neuropathy type 1E (HSAN1E) which is an inherited autosomal disorder. HSAN1E patients are usually characterized with central and peripheral nervous systems disorders including sensory impairments of the lower legs and feet, hearing loss and dementia. HSAN1E is caused by DNMT1 mutation which affects the enzymatic activity, resulting in the neuronal system impairment (Winkelmann et al., 2012). DNMT3A mutations are associated with growth disorders and hematological disorders. The growth phenotypes of the DNMT3A mutated patients dependents on the specific types of mutation. Heterozygous gain-of-function mutations of DNMT3A PWWP domain have been found in patients with microcephalic dwarfism which is specifically characterized with a reduced brain size. The gain-of-function mutations at the PWWP domain are suggested to abolish DNMT3A binding to H3K36me3, and affect the DNA methylation at the Polycomb-regulated regions (Heyn et al., 2019). Heterozygous DNMT3A mutations are also observed in an overgrowth disorder, the Tatton-Brown–Rahman syndrome (TBRS), which is characterized by tall stature and moderate intellectual disability. TBRS-related mutations are found in the three main functional domains of DNMT3A, PWWP, ADD and methyltransferase domain, and the details of the effect of DNMT3A mutations on disease development remain further investigated (Tatton-Brown et al., 2014). Somatic DNMT3A mutations are frequently observed in patients with myeloid-lineage hematological disorders and in healthy individuals with clonal haematopoiesis. DNMT3A mutations are found in around 30% of acute myeloid leukemia (AML) cases and around 57% of the individuals with clonal hematopoiesis (Buscarlet et al., 2017; Russler-Germain et al., 2014). The arginine-to-histidine substitution mutation at the catalytic domain residue R882, R882H, seems to be the dominant mutation in DNMT3A-mutated patients or healthy individuals with clonal

hematopoiesis. AML patients with R882H mutation exhibit a focal hypomethylation in the genome. (Russler-Germain et al., 2014). DNMT3B mutations are found in individuals with immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome which is characterized by hypomethylation of centrometric satellites II and III and centromeric heterochromatin instability (Okano et al., 1999), this is consistent with the mice studies about the function of Dnmt3b on methylating centromeric minor satellite repeats.

1.4.2 TET proteins in mouse embryo development and human diseases

TET proteins in mouse ESCs and embryogenesis

In mouse ESCs, Tet1 is shown to be involved in the inner cell mass-lineage specification and knockdown or knockout of *Tet1* in mESCs leads to decreased 5hmC resulting in cell differentiation bias towards trophectoderm lineage *in vitro* (Dawlaty et al., 2011; Ito et al., 2010; Koh et al., 2011). But these effects are not observed *in vivo* (Dawlaty et al., 2011). *Tet1* –/– mice with a mixed mouse strain background 129/sv x C57BL/6 are viable, fertile and have normal embryogenesis and postnatal development. Some healthy mutant mice show lower body weight at birth and during postnatal development, which suggest Tet1 loss could cause a developmental delay during embryogenesis. While in another study, the homozygous Tet1-mutant gene-trap mice with a background 129P2/Ola show embryonic lethality but the Tet1-mutant C57BL/6 mice are born normally (Yamaguchi et al., 2012). The female Tet1-mutant mice have smaller ovaries and show decreased fertility efficiency and small litter size. Besides Tet1, Tet2 is also expressed in mESCs. Most of the homozygous Tet1 and Tet2 double-knockout (DKO) mice develop to term and die perinatally with a variety of malformations (Dawlaty et al., 2013). Surviving double mutant mice are fertile but female mice have impaired fertility. Unlike Tet1 single KO mice, loss of both Tet1 and Tet2 lead to a global DNA hypermethylation and particularly affect the DNA methylation at certain gene imprinting loci. Tet1/2/3 triple-knockout (TKO) mESC(Tet1–/–|Tet2–/–|Tet3+/–) maintain the normal ESCs morphology and embryoid body (EB) differentiation ability, TKO EBs show a complete depletion of 5hmC and a

subtle increase in global 5mC levels (Dawlaty et al., 2014). TKO ESCs could develop most germ layers but with some defects in endodermal and selected mesodermal structures which suggests all three Tet proteins cooperate together in regulating ESCs differentiation potential during embryogenesis. Consistent with the finding in mESCs, Tet-TKO mice also show gastrulation defects including smaller body size, developmental delay in mid-streak stage, and impaired maturation of mesoderm specification (Dai et al., 2016), these further support the important function of DNA modification in embryogenesis. Tet3 function is associated with DNA methylation reprogramming of the paternal genome in zygotes. Tet3–/– mice undergo normal embryonic development but die perinatally. Mice with Tet3 depletion in oocytes exhibit impaired 5mC oxidation in the paternal genome and cause developmental defects in the embryos (Gu et al., 2011b).

TET proteins in human diseases

TET protein functions are closely related to human malignant transformations. Mutations of TET proteins, mostly TET1 and TET2 are found in a variety of cancer diseases, particularly related to hematopoietic malignancies. Moreover, global loss of 5hmC and reduced expression level of TET proteins are observed in patients with breast, liver, lung, glioblastoma, melanoma and blood cancers (Kudo et al., 2012; Liu et al., 2013; Turcan et al., 2012; Yang et al., 2013). These suggest an important role of TET proteins in tumorigenesis.

The initial evidence supporting the correlation of TET protein with cancer development comes from the discovery of TET family protein in which TET1 (originally referred as LCX, leukemia-associated protein with a CXXC domain) is identified as a fusion partner of the MLL in patients with acute myeloid leukemia (AML) (Lorsbach et al., 2003; Ono et al., 2002). MLL-TET1 translocation is also identified in other types of blood cancers including B-cell precursor acute lymphoblastic leukemia (BCP-ALL), T-cell lymphoblastic lymphoma (T-LBL), and non-Hodgkin B cell lymphoma (B-NHL) (Burmeister et al., 2009; Cimmino et al., 2015; Ittel et al., 2013). But the frequency of MLL-TET1 translocation in blood cancer is relatively low in contrast to other MLL translocations (Burmeister et al., 2009). Although Tet1 deficiency in mice shows no effect

in animal viability and normal development, both Tet1 heterozygous and homozygous mice develop and die of B cell malignancies in spleen and lymph nodes between 1-2 years, which indicates a role of TET1 as a tumor suppressor in B cell lineage (Cimmino et al., 2015). Tet1-deficient mice tumors show similar mutation spectrum as those in human B-NHL. However, TET1 is not mutated in human B-NHL, instead TET1 gene is silenced by promoter DNA hypermethylation. In general, TET1 mutation in human hematopoietic malignancies are reported in low freqeuncy and the contribution of TET1 deficiency in tumorigenesis still need to be investigated.

TET2 is frequently mutated in a wide range of hematopoietic malignancies including both myeloid and lymphoid lineages (Scourzic et al., 2015). TET2 mutations are firstly reported in a variety of myeloid cancers including acute myeloid leukemia (AML), myelodysplastic syndrome (MDS) and myeloproliferative neoplasms (MPN), chronic myelomonocytic leukaemia (CMML), TET2 mutation is also observed in a number of lymphoid cancers with a general low frequency but enriched in angioimmunoblastic T-cell lymphoma(AITL) (33-83%). Tet2-/- knock-out or hematopoietic-lineage conditional knock-out mice all show increased hematopoietic stem cell population and myeloid differentiation bias in the spleen, which further provide the evidence for TET2 function in blood lineage development (Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011). The details about TET2 function in blood malignancies and in mouse hematopoiesis would be discussed in chapter 1.5.1 and chapter 1.5.2 respectively.

In addition to hematological malignancies, TET protein mutations are also found in solid tumors but with much lower frequency compared with those observed in blood malignancies (Scourzic et al., 2015). All three TETs can be mutated rarely in colorectal cancers (CRCs) (Muzny et al., 2012; Seshagiri et al., 2012). TET1 is also mutated in low frequency in lung cancer (Imielinski et al., 2012; Seo et al., 2012) and bladder cancer (Gui et al., 2011). TET2 mutations are found in lung cancer (Imielinski et al., 2012; Seo et al., 2012) and melanoma (Rasmussen and Helin, 2016). Importantly, decreased TET expression level and reduced 5hmC level are often associated with many solid tumors including glioma (Turcan et al., 2012), CRCs (Kudo et al., 2012), prostate, breast cancer

(Hsu et al., 2012; Yang et al., 2013), liver, lung cancer (Yang et al., 2013) and melanoma (Lian et al., 2012). TET1 expression and 5hmC level were found decreased in patients with prostate, breast cancer and nealry 50% CRCs (Hsu et al., 2012; Kudo et al., 2012). Although TET2 is not frequently mutated in human brain cancers, reduced expression of TET2 led by DNA hypermethylation is found in many glioblastoma patients and ectopic overexpression of TET2 could impair the glioblastoma cells *in vitro* (García et al., 2018). In addition, TET2 enzymatic activity can be inhibited by 2-hydroxyglutarate (2-HG) which is produced by IDH1/2 mutant in human brain tumors (Losman and Kaelin, 2013; Xu et al., 2011a). In all, TET protein could serve as tumor suppressor genes as a consequence of the relatively high frequency of loss-of-function mutations identified in human cancers. The down-regulation of TET expression and 5hmC identified in cancers also suggest another mechanism by which TET protein get involved in tumor development by impairment of its enzymatic activities.

1.4.3 Altered DNA modification pattern is a landmark for cancers

Cancer is a genetic disease caused by deregulated gene expression which leads to aberrant activation of oncogenes that promote cell proliferation or inhibit cell death, or which leads to the inactivation of tumor-suppressor genes that are supposed to interrupt those processes (Baylin and Jones, 2016). Epigenetic regulation plays an important role in gene expression during cancer development (Baylin and Jones, 2011). Promoter DNA hypermethylation has been found in several tumor suppressor genes, which results in gene silencing in human cancers (Chen et al., 2003b; Suzuki et al., 2004). Cancer cells show distinguished DNA methylation landscape which is characterized by global DNA hypomethylation and focal hypermethylation at CpG promoter (CIMP) (Berman, 2012; Hansen et al., 2011). CpG promoter hypermethylation at certain disease-related genes could be used as a biomarker for early cancer diagnosis in lung cancers or colorectal cancers (Brock et al., 2009; Coolen et al., 2010; Lofton-Day et al., 2008). The discovery of TET proteins and 5hmC in the genome provide a new mechanism by which DNA modifications get involved in cancer development. Reduced levels of 5hmC and decreased expression of all three TET proteins have been found in human breast, liver, lung,

pancreatic and prostate cancers, which suggest an important role of DNA modification in tumorigenesis (Kudo et al., 2012; Yang et al., 2013). In summary, DNA modification is highly associated with cancer progression and development. The identification of cancer-related mutations of DNA modifying machineries and dysregulated TET enzymatic activities suggests both direct and indirect ways in which 5mC and 5hmC are involved in tumoriogenesis.

1.5 TET2 in normal and malignant haematopoiesis

Tet2 is the most highly expressed Tet genes in mouse blood lineages, and is widely expressed in a variety of blood cells including hematopoietic stem/progenitor cells (HSPCs), myeloid lineages and lymphoid lineages (Ko et al., 2010). TET2 is suggested be a tumor suppressor gene specifically in human hematological malignancies since the high frequency of TET2 mutations identified in a variety of human blood cancers including both myeloid and lymphoid lineages. Tet2 deficiency in mice leads to altered hematopoietic stem cell function and increased myelopoiesis in spleen. In this section, I will introduce the mutation features of TET2 in human hematopoietic malignancies, and describe the mice phenotypes caused by Tet2 deficiency, and finally discuss the current understanding of the contribution of TET2 mutations to carcinogenesis.

1.5.1 TET2 as a tumor suppressor in human hematopoietic malignancies

TET2 is expressed in a wide range of different human tissues and is particularly enriched in hematological cells with the highest expression in granulocytes (Langemeijer et al., 2009). TET2 mutations in hematological malignancies were initially reported in myeloid-lineage hematopoietic disorders (Delhommeau et al., 2009; Jankowska et al., 2009; Langemeijer et al., 2009; Tefferi et al., 2009). TET2 mutations are frequently identified in 6-26% MDS (myelodysplastic syndrome), 20-58% CMML (chronic myelomonocytic leukemia), 12-27% *de novo* adult AML (acute myeloid leukemia), and 17-32% secondary AML. TET2 mutations are also observed in 33-83% AITL (angioimmunoblastic T cell lymphoma) and rare in B-cell lymphoma (Scourzic et al., 2015).

An initial study observed the TET2 somatic mutation caused by chromosome abnormalities including uniparental disomy (UPD) and deletion of chromosome 4q24 in patients with MDS and MDS/MPN syndromes (Jankowska et al., 2009). Later on, TET2 point mutations such as missense, nonsense, insertion, deletion and frameshift mutations are also observed in patients with hematopoietic disorders (Delhommeau et al., 2009; Jankowska et al., 2009; Langemeijer et al., 2009; Tefferi et al., 2009; Weissmann et al., 2012). Patients with TET2 homozygous/ hemizygous/ heterozygous/ compound heterozygous mutations have similar clinical phenotypes which suggests TET2 single allele mutation is enough to promote tumorigenesis (Jankowska et al., 2009). Patients with myeloid malignancies could have more than one TET2 mutations and in a recent study on CMML, nearly 50% of the TET2-mutant patients have two TET2 mutations, but in general, a single mutation is more commonly observed in patients than multiple mutations (Coltro et al., 2019; Hirsch et al., 2018). TET2 mutations spread across all the exons of TET2 gene, and there is no specific mutation enriched in any cancer type. Nonsense and frameshift mutation can lead to the formation of pre-mature stop codons in the N-terminal of the protein which result in the loss of function of TET2 protein (Delhommeau et al., 2009; Jankowska et al., 2009; Langemeijer et al., 2009). The missense mutations which can cause in-frame deletion and amino acid substitution seem to preferentially locate in the two conserved regions in the C-terminal of the protein (Langemeijer et al., 2009).

TET2 mutations might occur as an early event in leukemogenesis since it is identified in most of the bone marrow cells (median 96%) with a variety of cell lineages, particularly, the mutations identified in CD34+ progenitor cells provide an evidence for TET2 function in early disease development (Langemeijer et al., 2009; Smith et al., 2010). Another evidence to support this is the high frequency of TET2 mutations observed in healthy individual with clonal hematopoiesis. The age-related clonal hematopoiesis is characterized by clonal expansion of the hematopoietic stem cell with preferential differentiation towards myeloid lineage. TET2 mutation is the first somatic mutation reported in healthy individual with clonal hematopoiesis (Busque et al., 2012). Similar to the mutation spectrum in myeloid neoplams, TET2 mutations in clonal hematopoiesis individuals are observed across the whole gene, including missense mutations, truncating

mutations (frame shift, and nonsense), in-frame deletions and insertions (Buscarlet et al., 2017). TET2 mutations in clonal hematopoiesis are suggested to exist as an ancestral event which can predispose to the formation of other cancer-related somatic mutations required for leukemogenesis, thus increase the risk of hematologic malignancies (Busque et al., 2012; Hirsch et al., 2018; Kishtagari et al., 2019). A typical example is that TET2 mutation is frequently found before the occurrence of JAK2V617F mutation in patients with MPN (Abdel-Wahab et al., 2010). In all, TET2 mutations in hematopoietic cells serve as an ancestral hit to favor the occurrence of other leukemogenic mutations for the disease development, rather than being a leukemic driver.

TET2 mutations have different co-occurring mutations in different tumor phenotype. In all myeloid neoplasms, the most frequent TET2 co-occurring mutations are another mutation in TET2 (43%), and mutation in ASXL1 (21%), SRSF2 (18%), and NPM1 (13%) (Hirsch et al., 2018). And by disease types, in MDS patients, TET2 mutation often co-occurs with another TET2 mutation, ASXL1, SF3B1, and SRSF2 mutation; in MDS/MPN patients, TET2 mutation is often in concert with another TET2, SRSF2, ASXL1, RUNX1, and CBL mutation; in MPN, the most frequent TET2 co-occurring mutations are JAK2, ASXL1, and SRSF2; in primary AML, co-occurrence of NPM1, DNMT3A, or FLT3-ITD mutation are more often observed (Hirsch et al., 2018). Although TET2 and DNMT3A are functional counterparts in DNA methylation and demethylation, the occurrence of these two mutations is often observed in myeloid malignancies and clonal hematopoiesis in which both two mutations account for nearly 90% of mutated genes in age-related clonal hematopoiesis (Abdel-Wahab et al., 2009; Buscarlet et al., 2017; Cancer Genome Atlas Research Network et al., 2013; Delhommeau et al., 2009; Robertson et al., 2019). Mice studies suggest that DNMT3A and TET2 cooperate together and affect HSC differentiation by repressing the expression of some lineage-specific transcription factors such as Klf1 to inhibit the HSC differentiation and thus lead to transformation (Zhang et al., 2016). Interestingly, TET2 mutations and certain mutations such as the single amino acid missense mutation at the arginine 132 (R132) of Isocitrate dehydrogenase 1 (IDH1) and IDH2, and Wilms tumor 1 (WT1) mutations are mutually exclusive in multiple cancers due to their overlaps in molecular function. IDH1 and IDH2

are mitochondrial enzymes that can catalyze the formation of 2-oxoglutarate (2-OG), an necessary cofactor for TET enzymatic activity. Gain of function of IDH1 and IDH2 mutations are frequently observed in 80-90% low-grade glioma/secondary GBM (glioblastoma multiform) and 10-30% secondary AML (Losman and Kaelin, 2013). IDH1/2 mutations lead to the loss-of-activities in 2-OG production and gain-of-activities in 2-hydroxyglutarate (2-HG) production (Dang et al., 2009; Figueroa et al., 2010; Ward et al., 2010). 2-HG is an 2-OG antagonist and it can inhibit the enzymatic activity of TET and the other 2-OG-dependent dioxygenases including histone demethylases involved in the demethylation of H3K4, H3K9, H3K27, and H3K79 (Figueroa et al., 2010; Xu et al., 2011a). Consistent with the inhibition role of 2-HG on TET2 activity, IDH-mutant AML and brain tumors have increase DNA methylation levels and show DNA hypermethylation at specific loci (Figueroa et al., 2010; Turcan et al., 2012). Importantly, the expression of mutant IDH1/2 or depletion of Tet2 in mice bone marrow cells could both impair the cell differentiation ability and increased stem/progenitor cell marker, c-Kit, expression, which indicate their consistent roles in regulating normal and malignant hematopoiesis (Figueroa et al., 2010). WT1 mutations are found in AML and is correlated with poor prognosis in AML patients (Hou et al., 2010). WT1 is identified as a TET2 interacting protein which is responsible for the recruitment of TET2 to its targeted regions in the genome, the functional coordination might explain for the mutational exclusivity of the two genes in AML (Wang et al., 2015b).

TET2 mutations in myeloid cancers are associated with the decreased 5hmC and increased 5mC (Figueroa et al., 2010; Ko et al., 2010). TET2 mutations in healthy individual with clonal hematopoiesis are associated with significant focal DNA hypermethylation and transcriptional silencing at specific loci which are hypermethylated in AML patients with TET2 mutation (Busque et al., 2012). However, there is no consistent conclusion of the prognosis value of TET2 mutations in hematological malignancies. TET2 mutated patients are usually associated with older age and have increased white blood cell counts and monocytosis, but TET2 mutation is suggested not have to any impact on the patient survival rates (Chou et al., 2011; Jankowska et al., 2009; Smith et al., 2010; Tefferi et al., 2009). However in a study in AML, patients with TET2 mutation have significantly

shorter event-free survival compared those with WT TET2 (median: 6.7 vs 18.7 months, P=0.009), and higher chances to have relapse, thus TET2 is thought be a prognostic biomarker in AML (Weissmann et al., 2012). In contrast, in a study in CMML, patients with TET2 mutations seem to have better overall survival than those without TET2 mutation (Coltro et al., 2019). In conclusion, TET2 mutations in human blood cancers lead to aberrant DNA methylation changes and affect the gene expression. And the functional importance of TET2-mediated gene regulation in inhibiting tumorigenesis need to be further explored. Given the lack of consistency in correlation between TET2 mutations and patient survival status, the significance of TET2 mutations in clinical prognosis remain to be investigated.

1.5.2 Tet2 in mouse hematopoiesis

To have a better understanding of the role of TET2 in hematopoiesis and dissect the molecular mechanisms behind the process, several groups have generated Tet2-deficient mice by using different strategies including knock-out, blood lineage conditional knock-out and gene-trap method and by targeting at different regions of Tet2 gene (Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011; Shide et al., 2012). In general, those *Tet2*-KO mice show similar defects in blood development and could develop into CMML-like phenotypes at an average time of 12-20 weeks, which supports the finding observed in human myeloid malignancies and confirms the role of TET2 in hematopoietic development.

There are several *Tet2*-KO model mice available which differ in the way of the generation of Tet2 depletion but they do have some similar hematopoietic phenotypes. The Levine group created a hematopoietic-specific Tet2-depleted mice by targeting the first coding Tet2 exon, exon 3, which is one of the most frequently mutated exons in myeloid malignancies (Moran-Crusio et al., 2011. The Bernard group generated an interferon inducible Tet2-depleted mice to achieve the acute Tet2 inactivation in adult stage and exon 11, which is the second frequently mutated exon, were removed to deplete the dioxygenase domain of Tet2 protein (Quivoron et al., 2011). Another group also tried to deplete exon 9 which encodes a motif required for Tet2 enzymatic activity (Ko et al.,

2011). Disruption of the endogenous ATG at exon 3 was also used to deplete Tet2 expression (Li et al., 2011b).

All Tet2-deficient mice are born with the expected Mendelian ratio, are fertile and appear normal as wild-type mice. Tet2 deficiency in blood lineage is not compensated by increased expression of other Tet genes Tet1 and Tet3. Loss of Tet2 leads to a decreased 5hmC level in bone marrow and spleen in mice (Ko et al., 2011). Tet2 deficiency in mice leads to the enlarged hematopoietic stem cell (HSC) compartment and increases the self-renewal ability of HSC *in vivo* which provides the mutant HSC a growth advantage over the wild-type HSC in the competitive transplantation assays. More strikingly, Tet2-depleted animals have a significant enlarged spleen which undergoes elevated extramedullary hematopoiesis around 12-20 weeks of age (Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011). *In vitro* cell culture of the LSK progenitor cells from Tet2-/- mice show premature myeloid differentiation towards monocyte/macrophage cells (Ko et al., 2011). Moreover, heterozygous mutant mice (Tet2+/-) also show increased HSC self-renewal and extramedullary hematopoiesis in spleen, which suggests that monoallelic Tet2 deficiency could contribute to the progression of the myeloid malignancies (Moran-Crusio et al., 2011; Quivoron et al., 2011). Tet2 deficiency in mice has no big impact on lymphocytes lineage differentiation (Ko et al., 2011; Moran-Crusio et al., 2011). However, one group observed the altered hematopoiesis of both myeloid and lymphoid lineages including monocytic, erythroid, T cell and B cell lineage in their *Tet2*-KO mice which the last coding exon of Tet2 was depleted in all cells (Quivoron et al., 2011). In those mice, there were increased immature double-negative CD4-CD8- T cell progenitors in thymus, decreased B cell lineages and increased erythroid precursors in bone marrow and extramedullary erythropoiesis in spleen. However the mice finally still develop a human CMML-like myeloid malignancy which is consistent with other Tet2-depleted mice models. All Tet2-depleted mice models could develop a condition similar to human CMML with age, which is characterized by increased monocytosis and granulocytosis in peripheral blood and spleen, splenomegaly, and extramedullary hematopoiesis. Rarely Tet2-depleted mice could die of disease and the median survival age is around 12-18 months, although in one study some Tet2-/- mice

show poor survial after 1 year and die with myeloid malignancies phenotypes (Li et al., 2011b). Two groups have successfully generated Tet2-deleted AML mice models in combination with the expression of other oncogenic genes, such as AML1-ETO or FLT-ITD (Hatlen et al., 2016; Rasmussen et al., 2015; Shih et al., 2015). scRNA-seq analysis on HSPCs isolated from Tet2 KO mice (Mx1-cre) revealed that Tet2 depletion did not result in any new cell cluster, but led to an increased quiescent HSC cluster, decreased Ki67+ long-term HSCs (LT-HSCs) (Lin-, Sca-1+, c-Kit+, CD150+, CD48-), an increase in some myelomonocytic progenitor cluster and decreased erythroid progenitor frequencies (Izzo et al., 2020). In conclusion, Tet2 is shown to be involved in regulating the expansion and function of HSCs and negatively regulates the HSC differentiation towards myeloid lineage. Moreover, Tet2 seems to inhibit immature myeloid differentiation towards the monocyte/macrophage in the progenitor cells. The long latency and low penetrance of malignant transformation in Tet2-depleted mice could indicate other oncogenic mutations might be required for the leukemogenesis.

TET2 mutations often co-occur with additional mutations in myeloid malignancies, and it is not quite known how the co-occurring mutations cooperate together and lead to the disease development. Several groups have also generated mouse models to explore the hematopoietic phenotypes caused by those mutations. Addition of Sex Combs Like 1 (ASXL1) mutations is a TET2 co-occurring mutation in patient with MDS. Depletion of Asxl1 in mice reduces the HSC self-renewal ability and gives a significant growth disadvantage to the KO HSC/HSPC in the competitive transplantation assay. The additional depletion of Tet2 in the *Asxl1*-KO mice can compensate the self-renewal defect caused by Asxl1 deficiency. Asxl1/Tet2 compound mutant mice develop a disease condition similar to human MDS phenotype, and have poorer survival compared with single KO mice (Abdel-Wahab et al., 2013). Another example is the co-occurrence of Jak2V617F expression and Tet2 mutation in MPN. The combination with JAKV617F mutation facilitates the disease progression since Jak2V617F/Tet2 compound mutant mice develop more accelerated MPN phenotype than either single mutant mice. Tet2 deletion might provide the Jak2V617F-mutant HSC a significant competitive advantage and regulates HSC function in a long-term manner (Chen et al., 2015; Kameda et al.,

2015; Shepherd et al., 2018). In conclusion, loss of Tet2 increases the HSC self-renewal capacity and offers the HSC/HSPC a clonal growth advantage, the additional co-occurring mutation promotes the disease progression.

DNMT3A and TET2 mutations are shown to co-occur in AML and T cell lymphoma. Both enzymes are related to DNA modifications on cytosine but function in opposite manner on modification establishment and removal. Single cell RNA-seq and single cell RRBS on HSPCs from mice with the both enzyme mutations reveal that Dnmt3a KO and Tet2 KO have distinct preference in favoring the erythroid and the myelomonocytic lineage differentiation. DNA hypermethylation by Tet2 KO leads to myelomonocytic skews in HSC population, in contrast Dnmt3a KO-induced hypomethylation leads to a erythroid-lineage bias (Izzo et al., 2020). Although the DNA methylation changes led by the two enzymes show opposite direction which is consistent with their biochemical activity, the clinical hematopoietic phenotypes are quite similar. Whole bone marrow transplantation (BMT) experiments show that mice transplanted with double KO bone marrow cells can develop into CMML-like disease in a faster speed than either single KO cells-transplanted group (Zhang et al., 2016).In conclusion, current study suggests that TET2 and DNMT3A function in a collaborative and competitive way to affect DNA methylation at certain lineage-specific factors, thus affect the normal hematopoietic cell differentiation, finally lead to malignancies transformation.

1.6 Molecular mechanisms of TET2 function in gene regulation

TET2 is firstly known as a 5mC dioxygenase which catalyzes the formation of 5hmC and iterative oxidized products, 5fC and 5caC, resulting in the active DNA demethylation. In parallel, those modified bases might serve as functional epigenetic marks involved in gene regulation. Loss-of-function TET2 mutations are prevalently observed in human blood malignancies and *Tet2*-KO mice exhibit altered hematopoiesis and develop human CMML-like condition, these support the important role of TET2 function in human and mouse hematopoiesis. Moreover, loss of 5hmC and de-regulated Tet2 function have been linked

to many solid tumors. However, little is known about the mechanisms by which TET2 regulates hematopoiesis, the molecular events of TET2 function in normal and malignant hematopoiesis and how defective TET2 contributes to cancer. Current studies focus on the the understanding of transcriptional regulation of TET2 in hematopoiesis, the recruiting mechanisms of TET2 to its targets loci, and the regulation of TET2 enzymatic activity. In this section, I will firstly introduce the enzymatic activity-dependent/ independent TET2 gene regulation activity, then explain the understanding of Tet2 binding events in genome, followed by a description of the known Tet2-interacting proteins. The final section will focus on the protein or molecules involved in TET2 enzymatic activity regulation.

1.6.1 TET2 genomic distribution

In order to identify the molecular events directly contributed by TET2 protein, chromatin immunoprecipitation-sequencing (ChIP-seq) experiments have been performed to map the genome-wide endogenous or exogenous TET2 occupancy in multiple human and mouse genomes (Chen et al., 2013; Deplus et al., 2013; Rasmussen et al., 2019; Uribe-Lewis et al., 2015; Wang et al., 2018; Xiong et al., 2016). Tet2 binding is mainly found either in active gene promoters enriched with H3K4me3 (Chen et al., 2013; Deplus et al., 2013) or enhancer regions (Rasmussen et al., 2019; Uribe-Lewis et al., 2015; Wang et al., 2018; Xiong et al., 2016). However, there are some controversies about Tet2 binding at gene promoters since one study observed few Tet2 binding events at promoters in mESCs and most of Tet2 binding sites were located in CpG-sparse region (Rasmussen et al., 2019) which contradicts with previous finding about Tet2 targeting at high density CpG promoters in mESCs (Chen et al., 2013). The inconsistency from different study might come from the use of different antibody. The specificity of either the commercially produced or home-made Tet2 antibodies need to be further examined. Tet2 binding at gene promoters was firstly reported by two studies where O-GlcNAc transferase (OGT) was identified as the Tet2 interacting partners in mouse ESCs and 293T cells (Chen et al., 2013; Deplus et al., 2013). By ChIP either the endogenous Tet2 protein in mES or using the Halo-tag antibody to ChIP the exogenously-expressed tagged protein in 293 T cells, they both observed that TET2 could recruit OGT on chromatin at promoter regions

which are enriched with H3K4me3 and with high-density CpG sites. The co-occurrence of the TET2 and OGT at active gene promoter regions is thought be associated with transcriptional activation. The binding of TET2 at CpG-rich region and TSS, and the co-localization of TET2 and OGT were also observed in mouse bone marrow cells (Deplus et al., 2013). TET2 binding at CpG-rich gene promoters was also found in human colon cancer cells and human leukemic cell line, HL-60 (Uribe-Lewis et al., 2015; Wang et al., 2015b).

Tet2 activity at enhancer regions is gaining more and more focus since several studies observed the enrichment of DNA hypermethylation caused by Tet2 depletion on enhancer regions (Hon et al., 2014; Lio et al., 2019a; Lu et al., 2014; Rasmussen et al., 2015). Tet2 binding at enhancers region has been recently reported in several studies which provides an evidence for Tet2 direct regulation on enhancer activity. One example of Tet2 binding at enhancer regions comes from a mouse ESCs study where a novel 5hmC binder, SALL4A, facilitate 5hmC oxidation at enhancer region by recruiting TET2 protein (Xiong et al., 2016). The occupancy of TET2 protein at active enhancers regions have been shown to facilitate the recruitment of estrogen receptor α (ERα) in response to estrogen signaling (Wang et al., 2018). Tet2 binding at enhancer regions is also observed in mESCs, immortalized mouse myeloid cells(AML1-ETO) and mouse AML cells (Npm1cA; Flt3-ITD) using a novel Tet2-N antibody (Rasmussen et al., 2019). $\sim$33% of Tet2 binding sites in mouse myeloid cells were found at enhancer regions. Tet2 ChIP-seq in mESCs showed a strong enrichment of Tet2 binding at enhancer regions which undergo DNA hypermethylation in Tet2-deficient cells, whereas little binding signal was observed at promoters and CpG islands. However, the overlaps between differentially methylated regions and Tet2 binding sites is quite small, which might indicate that Tet2 binding to DNA is not always required for its enzymatic activity. Moreover, Tet2 binding sites in mouse myeloid cells and AML cells show different gene ontology term enrichments and TF motif enrichments as those observed from mESCs, indicating a cell-type specific binding profile of Tet2. Although clear ChIP-seq peaks are observed in the identified Tet2 binding sites in the WT cells in this paper, the ChIP profiles between Tet2 WT cells and Tet2 KO cells are quite similar since a lot of the called Tet2 peaks would be also found in

the *Tet2*-KO cells ChIP-seq signal and exhibit a reduced peak height, thus the specificity of this Tet2-N antibody need more careful evaluations.

1.6.2 Enzyme activity-dependent gene regulation

TET2 function is closely related to 5mC oxdiation and the consequent active DNA demethylation process. Altered DNA methylation and hydroxymethylation landscapes have been observed in mouse ESCs, genetically modified mouse models and blood cancer patients with TET2 mutation or with de-regulated TET2 enzymatic activity (Cimmino et al., 2017; Figueroa et al., 2010; Ko et al., 2011, 2010; Moran-Crusio et al., 2011; Quivoron et al., 2011). Moreover, restoration of TET2 expression or treatment with ascorbate, a co-factor for TET2 enzymatic activity could rescue the aberrant mouse HSC self-renewal phenotype and inhibit human leukemic cell progression, these suggesting that TET2 function in physiological hematopoiesis and its dysregulation in pathological hematopoiesis might be dependent on its 5mC oxidation activity. Although two recent studies reported the non-catalytic role of TET2 function in negative regulation of mast cell proliferation (Montagner et al., 2016) and in contribution to lymphopoiesis (Ito et al., 2019). TET2 function in myelopoiesis is more dependent on its catalytic activity since TET2 catalytic mutant itself could cause mice to develop myeloid malignancies whereas mice with complete Tet2 depletion develop both myeloid and lymphoid disorders (Ito et al., 2019). Given the widely-observed TET2 mutations more in human myeloid disorders and the therapeutic values in targeting its enzymatic activity by its cofactors ascorbate and DNA methylation inhibitor, 5-Aza-2'-deoxycytidine (decitabine) in AML, CMML, MDS diseases (Agathocleous et al., 2017; Cimmino et al., 2017; García et al., 2018; Issa et al., 2004; Lainey et al., 2013), here I focus on the TET2 catalytic activity-dependent cellular processes, and I will also emphasize the known understanding of TET2 activity on enhancer regulation.

Tet2 function in mouse embryonic stem cells

In mouse embryonic stem cells (mESCs), Tet1 and Tet2 are highly expressed, Tet3 is expressed at minimal level and undetectable by Western blot (Koh et al., 2011; Lu et al.,

2014). Tet2 mRNA is expressed about 5-fold less abundant than Tet1, Tet2 depletion leads to more significant 5hmC loss in mESCs compared with Tet1 depleted cells, so that Tet2 is thought to be the dominant dioxygenase that catalyzes the 5mC oxidation in mES cells (Hon et al., 2014; Huang et al., 2014), but Tet2 depletion causes less pronounced gene expression changes and is dispensable for the embryonic development (Koh et al., 2011). Instead, Tet1 is the main Tet proteins involved in gene regulation for cell pluripotency since Tet1 depletion causes bigger gene expression changes than Tet2 deficiency and majority of Tet2-affected genes overlap with Tet1-affected group (Ficz et al., 2011; Huang et al., 2014).

Tet1-mediated 5mC oxidation occurs at gene promoter regions and is negatively correlated with gene expression, whereas, Tet2 is responsible for the 5hmC deposition in the gene body which is positively correlated with gene transcription (Huang et al., 2014). However, there is no clear correlation between the gene/exon expression changes and the 5hmC changes at the promoter/gene body caused by Tet1 and Tet2 deletion. Later study identified the role of Tet2-mediated 5mC oxidation at enhancer regions since hypermethylated DMRs (hyper-DMRs) in Tet2-/- are enriched in both active and poised enhancer regions. Although enhancer hypermethylation caused little direct effects on its histone modification levels, enhancer with more than average hypermethylation showed loss of histone acetylation and reduced gene expression in Tet2-/- cells. (Hon et al., 2014; Rasmussen et al., 2019). Chromatin immunoprecipitation experiments suggests that Tet2 preferentially binds to the open chromatin regions and has strong enrichment at DNase I hypersensitivity with enhancer features (EP300, H3K27ac, H3K4me1) which undergo Tet2-dependent 5mC oxidation (Rasmussen et al., 2019). Tet2-mediated 5mC oxidation at enhancer region is involved in chromatin accessibility construction and facilitates the specific groups of transcription factor binding. In all, Tet2 is suggested to be a positive regulator of enhancer activity.

TET2 function in human blood malignancies

TET2 is frequently mutated in patients with myeloid malignancies and healthy individuals with clonal hematopoiesis. TET2 activity in human myeloid neoplasm is closely associ-

ated with its enzymatic activity since decrease of 5hmC is observed in bone marrow DNA from patients with myeloid cancers (Figueroa et al., 2010; Ko et al., 2010; Pronier et al., 2011). Expression of TET2 mutants from patients in HEK 293 cells also cause a lack of gain of 5hmC (Ko et al., 2010). Moreover, clonal hematopoiesis individuals with TET2 mutations show significant focal DNA hypermethylation and transcriptional silencing at specific loci which are hypermethylated in AML patients with TET2 mutation (Busque et al., 2012). However, there is no consistent conclusion about the genome-wide DNA modification changes in cancer genome since different studies gives rise to different observations. Figueroa et al reported the DNA hypermethylation at gene promoter regions in AML patients with TET2 mutation (Figueroa et al., 2010), whereas, Ko et al found in TET2-mutant patients with myeloid neoplasm (MDS, MDS/MPN, primary and secondary AMLs), there was a widespread DNA hypomethylation instead only two hypermethylated genes were identified (Ko et al., 2010). More recent study observed a specific enhancer hypermethylation phenotype in AML patients with TET2 mutations (Rasmussen et al., 2015), which is in line with the previous observation about TET2-mediated 5mC oxidation on enhancer in mouse ESCs. In conclusion, although altered DNA modification levels were observed in patients with TET2 mutations, the detailed mechanisms by which TET2-mediated 5mC oxidation gets involved in gene regulation and regulates the malignancies transformation remain further explored.

Tet2 function in mouse hematopoiesis

Tet2 depletion in mice leads to a significant loss of 5hmC in bone marrow DNA and no significant changes in 5mC level is observed (Ko et al., 2011; Li et al., 2011b; Quivoron et al., 2011). In order to understand the contribution of Tet2 enzymatic activity to gene regulation in normal hematopoietic development, several groups have examined the gene expression changes and mapped the genome-wide DNA modification alteration in blood cells derived from Tet2 WT or mutant mice (Izzo et al., 2020; Lio et al., 2019a; López-Moyado et al., 2019; Rasmussen et al., 2015).

scRNA-seq and scRRBS analysis on HSPCs (lineage-negative hematopoietic progenitor) isolated from Tet2 KO mice (Mx1-cre) found that Tet2-mediated DNA demethylation

is involved in HSC cell-lineage commitment by affecting the cell-lineage transcription factor binding to DNA since different lineage-defining transcription factors have distinct CpG contents which lead to varied sensitives to the methylation changes led by Tet2 deficiency(Izzo et al., 2020). Tet2 deficiency in HSC caused the myelomonocytic differentiation skews over the erythroid lineage. In all, Tet2 regulates HSC cell differentiation by its DNA demethylation activity through affecting the DNA binding events of lineage-specific transcription factors and facilitate those lineage transcription factor with more CpG enrichment.

To understand the role of TET2 in leukemia progression, Rasmussen et al have generated a novel Tet2-deficient human AML-like model which combines Tet2 deficiency with the expression of an oncofusion protein, AML1-ETO (AE) (Rasmussen et al., 2015). Tet2-deficient AML mice developed a disease condition similar to TET2 mutated human AML. Tet2-depleted mouse leukemic cells lost the expression of CD45 and exhibited similar expression profile as the human AML cells. Moreover, the Tet2-deficient mice showed poor viability and usually died of disease within 200 days. *In vitro* cultured GMP cells (Lin- cKit+ Sca1- CD16/32+ CD150-) isolated from Tet2-deficient AML mice showed decreased 5hmC level and altered gene expression related to stem cell function, leukemogenesis, and cancer, particularly, several putative tumor suppressor genes and oncogenes are down-regulated or up-regulated respectively. 5hmC mapping by 5hmC-DIP-seq revealed a marked loss of 5hmC in the genome and DNA hypermethylation was found in 25% of all enhancers in Tet2-/- AE cells by eRRBS. Interestingly, CpG island and CpG gene promoters showed the cell passage-dependent DNA hypermethylation in Tet2 depleted AE cells. DNA hypermethylation at enhancer regions is associated with loss of H3K27ac which is an indicator for enhancer activity. Enhancer DNA hypermethylation with >30% methylation increase is negatively associated with gene expression of nearby genes, and there is no direct or significant correlation between enhancer-associated gene regulation and the absolute 5mC/5hmC level. Moreover, there is no clear correlation between the loss of 5hmC at enhancer and gene expression or enhancer activity, suggesting Tet2 regulates enhancer activity and its associated gene expression mainly by its DNA demethylation activity. Tet2 depletion-led DNA hypermethylation on enhancers occurs

both during hematopoiesis and malignant transformation since the increase of 5mC is found in both primary GMP cells with Tet2 mutation alone and primary mouse leukemic cell. Moreover, motifs of several hematopoietic transcription factors including PU.1 and Runx1 were found enriched in hypermethylated enhancer DMR. In all, Tet2 is suggested to be responsible for the DNA demethylation at enhancer regions in mouse blood cells and leukemic cells and the increased DNA methylation caused by Tet2 depletion is negatively associated with enhancer activity and its gene expression regulation on nearby genes.

In addition to myeloid malignancies, TET2 mutations are also found in diffuse large B cell lymphoma (Schmitz et al., 2018) and peripheral T-cell lymphomas (Lemonnier et al., 2012). Lio et al and Lopez-moyado et al examined the Tet2 activity in B cells lineage (Lio et al., 2019a) and T cells/NKT-cell lymphoma cells from Tet2/3 T cell-lineage conditional KO mice (López-Moyado et al., 2019) respectively. Aicda encodes the activation-induced cytidine deaminase (AID), a CSR-essential enzyme. Tet2 and Tet3 were found both involved in the mouse B cell class switch recombination (CSR) by its DNA demethylation activity at the enhancer region of Aicda and positively regulated the expression level of Aicda. Tet-deposited 5hmC was highly enriched at active (H3K4me1+ H3K27Ac+) relative to poised (H3K4me1+ H3K27Ac-) enhancers and associated with accessible chromatin defined by ATAC-seq (assay for transposase-accessible chromatin using sequencing). Deposition of 5hmC followed by B cell activation at Aicda locus caused increased accessibility and promoted the bZIP transcription factor BATF binding at the Aicda enhancers, resulting in an up-expression of Aicda. This is consistent with the finding observed in mouse myeloid cells where Tet2-mediated DNA demethylation affects the transcription factors binding at enhancer regions. Despite the well-known function of Tet2 regulation at enhancer regions, Lopez-moyado et al identified the role of Tet activity in DNA methylation dynamic at heterochromatin since there was DNA hypomethylation in heterochromatin compartments in several mouse Tet-deficient genomes including Tet2/3 DKO T cells/NKT-cell lymphom cells, HSPCs from Dnmt3a-Tet2 DKO Mice, and all Tet-mutant mESCs (López-Moyado et al., 2019). DNA hypomethylation in heterochromatin region in Tet2/3 DKO NKT cells were found associated with a large increase in DNA double-strand breaks and transposable elements activation. And the

detailed mechanism for Tet function in heterochromatin activity and its functional importance for cancer transformation remain further illustrated.

Tet2 function at different genomic regions

Gene expression regulation in mammalian genome is shaped by the coordination of a collection of transcription factors and cofactors on the *c-cis* DNA regulatory elements and the chromatin environment containing these regulatory elements (Lelli et al., 2012). CpG DNA methylation is highly associated with gene repression at CpG-enriched gene promoters and enhancer regions (Greenberg and Bourc'his, 2019), whereas the correlation between 5hmC and gene regulation is less clear. Current understandings of Tet2 function in gene regulation focus on its DNA demethylation activity at enhancer regions since DNA hypermethylation at enhancer regions caused by Tet2 deficiency is widely observed in many cell contexts including human AML patients, mouse ESCs and mouse blood cells and during the normal murine B cell activation process (Hon et al., 2014; Lio et al., 2019a; Lu et al., 2014; Rampal et al., 2014; Rasmussen et al., 2015).

In *Tet*-TKO mESCs, DNA hypermethylation is observed in both active and bivalent enhancers and it is negatively correlated with H3K27ac levels which is an indicator for enhancer activity. Transcriptionally active genes show decreased expression levels when their associated enhancers are hypermethylated (Lu et al., 2014). However, in general, hypermethylation in enhancer regions leads to a small change in gene expression level in TKO mESCs. Tet2 depletion in mESCs dose not lead to a global active and repressive chromatin marks change at enhancer regions. Moreover, enhancer hypermethylation does not lead to dramatic gene expression changes of its associated genes, but enhancer with more significant gain of 5mC is more prone to active chromatin marks loss. Rasmussen et al also observed the enhancer DNA hypermethylation in human AML petients with TET2 mutations (Rasmussen et al., 2015). And in the Tet2-/- human AML mice model generated in the same study, DNA hypermethylation is confirmed with correlation with loss of H3K27ac. Tet2-mediated DNA demethylation at enhancer regions is suggested to affect the chromatin accessibility and regulate the lineage-specific transcription factor binding (Rasmussen et al., 2019; Wang et al., 2018). Wang et al firstly observed the

impaired recruitment of estrogen receptor α (ERα) to active enhancers following Tet2 KO in MCF cells, a ERα positice human breast cancer cell line. Assay for Transposase Accessible Chromatin and sequencing (ATAC-seq) in myeloid progenitor cells from Tet2 KO mice revealed that most of ATAC peaks with loss of DNA accessibility in myeloid progenitor cells have increased DNA methylation and partially overlap with annotated enhancers in blood lineages (Rasmussen et al., 2019). Loss of Tet2 lead to both increased and decreased activities of different subsets of transcription factors which are clustered by their similarities in DNA-binding motifs. Among the changes, basic helix–loop–helix (bHLH) family TFs (such as MYC and ITF2) show strong loss of activity in Tet2 KO myeloid cells concomitantly with the downregulation of their targeted genes in HSC. Moreover, ITF2 binding sites in HSC with Tet2 depletion show increased DNA methylation which further suggest that Tet2 plays a role in regulating the transcription factor binding by affecting the DNA methylation level at the binding motifs. The role of Tet2-mediated DNA demethylation in chromatin accessibility regulation is also observed in the study about the mouse B cell activation, where Tet2/3-mediated DNA demethylation at two enhancer regions of Aicda is important for transcriptionally activation of Aicda through the maintenance of accessible chromatin structure thus contributing to the B cell class switch recombination process (Lio et al., 2019a).

In addition to the well-accepted DNA demethylation activity at enhancer regions, Tet2-mediated gene promoter DNA demethylation is also associated with the upregulation of SRY gene which is a mammalian key regulator of male sex determination (Okashita et al., 2019). Unexpectedly, a recent study observed the accumulation of DNA hypomethylation in heterochromatin regions and reactivation of the repeat elements in multiple mouse genomes with Tet deficiency including developmental cells such as mESCs and HSPC, T cells and also primary murine T cell lymphoma cells (López-Moyado et al., 2019). In mESCs, loss of Tet1 leads to the re-localization of Dnmt3a from heterochromatin to euchromatic compartment, which might explain the DNA hypomethylation at heterochromatin and increased DNA methylation at euchromatic regions.

Recently, DNA methylation canyon or DNA methylation valley has been identified as a novel genomic feature in human/ mouse embryonic stem cells and mouse hematopoietic stem cells (Jeong et al., 2014; Xie et al., 2013). Canyons are defined as large DNA fragments that are usually around 3.5kb-25kb in length, remain low DNA methylation level (<10%) even during differentiation process, and have multiple CpG islands (at least 5 CpG per kilobase to satisfy the 5% permutation-based false discovery rate (FDR)). Canyons are quite stable among species and cell types since majority of the mouse ESC canyons overlap with those identified in human ESCs and mouse HSCs (Jeong et al., 2014; Xie et al., 2013). Canyons-associated genes exhibit a bimodal distribution of the active and repressive histone marks, which half of the genes show high levels of H3K4me3 marks and the rest covered with high levels of H3K27me3 marks and canyon boundaries are marked with 5hmC peaks in wild-type HSCs (Jeong et al., 2014). Some of the HSC canyon-associated genes are observed dysregulated in human leukemias which suggest a potential role of the DNA methylation regulation at canyon in normal hematopoiesis and leukemogenesis. In a recent study, Wiehle et al revealed the role of Tet1/2-mediated DNA demethylation in protecting canyon from methylation and regulating the canyon-associated gene expression in mouse embryonic fibroblasts (MEF) (Wiehle et al., 2016).

1.6.3 TET2 interacting proteins

Unlike TET1 and TET3 which have a clear DNA binding domain, CXXC domain, responsible for the recruitment of the TET proteins to their target sites, TET2 lost its CXXC domain during evolution (Ko et al., 2013). Therefore, the recruiting mechanisms for Tet2 targeting at its binding loci remain unknown. The current hypothesis is that TET2 gets recruited to the targets by interacting with other DNA binding proteins so that a lot of TET2 interactome screenings have been done in multiple cell types including mouse ESCs, 293 T cells, human leukemic cells (HL-60) and murine erythroleukemia cells (MEL) by doing affinity purification-coupled with mass spectrometry in either endogenous or over-expressed cell context (Chen et al., 2013; Deplus et al., 2013; Guallar et al., 2018; Rampal et al., 2014; Wang et al., 2015c). In addition to the proposed recruiting mechanism for TET2 DNA binding, some of identified Tet2 interacting partners have

been also involved in post-translational regulation of TET2 enzymatic activity and protein stability control, or the interactors get recruited by TET2 protein and play a role in TET2-involved gene regulation. In this section, I will introduce the known Tet2 interacting partners and the functional importance of its interaction with TET2.

Interactors for TET2 chromatin recruiting:

1. IDAX (CXXC4):

TET2 gene underwent chromosomal inversion during evolution and lost its CXXC domain which become an independent gene called *IDAX* (also known as CXXC4) (Iyer et al., 2009). *IDAX* (4q22-q24) is localized around 650 kb upstream of the *TET2* gene (4q24) in the human genome and transcribed in opposite direction. IDAX binds unmethylated CpG-containing DNA *in vitro* and is localized at CpG island and CpG-rich promoters *in vivo*. The physical interaction between IDAX and TET2 was confirmed by co-immunoprecipitation experiments in 293 T cells with the over-expression of the two proteins (Ko et al., 2013). Specifically, IDAX interacts with the catalytic domain of TET2 and recruits TET2 to DNA. Interestingly, the recruitment by IDAX leads to a caspase-dependent TET2 protein degradation and similar phenotype is also observed on TET3 protein where the CXXC domain recruits TET3 protein to DNA followed by the activation of caspase pathway and results in the down-regulation of TET3 protein level, which suggests a general auto-regulation mechanism for TET protein level by its DNA binding domain.

2. WT1 (Wilms tumor 1):

WT1 gene encodes a transcription factor with a proline/glutamine-rich DNA-binding domain at the N-terminus and zinc finger motifs at the C-terminus. Its original function is associated with the pediatric kidney cancer, Wilm's tumor in which nearly 15% patients with Wilms'tumor have the WT1 mutation. WT1 is shown to be involved in development and tissue homeostasis in many organs including kidney, gonads, heart and nervous system(Hastie, 2017). WT1 mutations in AML lead to a global decreased 5hmC level and is mutually exclusive with TET2 and IDH1/2 mutations, suggesting WT1 might

participate in AML inhibition in a similar manner as the DNA modifying enzymes (Rampal et al., 2014; Wang et al., 2015b). The endogenous interaction between TET2 and WT1 is observed in multiple genomes including mouse ESCs, mouse bone marrow cells, and several human leukemic cells such as HL-60 cell, human erythroleukemia cell (HEL cells) and acute myeloid leukemia cell (Nomo-1 cells) (Rampal et al., 2014; Wang et al., 2015b). The zinc-finger domain of WT1 interacts with the catalytic domain of TET2 and WT1 also interacts with TET3. WT1 recruits TET2 to its targeted genes where most of them are located in the CpG-rich promoter regions and have altered gene expression level dependent on TET2 enzymatic activity. The cooperation between TET2 and WT1 plays an inhibitory role in leukemia cell proliferation and colony formation (Wang et al., 2015b). Importantly, some AML-related TET2 mutations abolish its interaction with WT1 suggesting a function relevance of TET2 chromatin recruiting by WT1 in suppressing tumorigenesis.

3. PML (promyelocytic leukemia):

PML is a known tumor suppressor gene in acute promyelocytic leukemia where the PML mutations cause the disrupted subnuclear structures, PML nuclear bodies (PML-NBs). PML plays important roles in cell cycle regulation, cell survival and apoptosis. Inactivation or down-regulation of PML is frequently found in cancer cells (Guan and Kao, 2015). The endogenous interaction between PML and Tet2 is identified in 293 T cells by co-IP method. PML recruits TET2 to PML-NBs in response to th chemotherapeutic agent, doxorubicin, to facilitate the 5hmC formation and re-activates the gene transcription in PML-NBs (Song et al., 2018). The identification of interaction between PML and TET2 reveals a role of TET2-mediated 5mC oxidation in response to chemotherapeutic agents.

4. MBD3L2 (Methyl-CpG-binding domain protein 3-like 2)

MBD3L2 is a homolog of methyl-CpG-binding domain protein 3 (MBD3) but lack the MBD domain. The physical protein interaction and cellular co-localization of TET2 and MBD3L2 was identified in 293 T cells in the transient over-expression system and the interaction occurs at the cysteine-rich dioxygenase domain of TET2 (Peng et al., 2016). Importantly, MBD3L2 modulates both the enzymatic activity and DNA binding affinity

of TET2 protein. Both *in vitro* TET2 enzymatic reaction with purified protein and *in vivo* co-expression assay in 293 T cells show that additional expression of MBD3L2 can enhance the TET2-mediated 5hmC formation and genome-wide DNA modified base mapping by both hMeDIP-seq and RRBS reveal the increase of 5hmC and decrease of 5mC in genomic regions co-bound by TET2 and MBD3L2. Electrophoretic mobility shift assay (EMSA) experiments show that MBD3L2 can specifically enhance the binding affinity of TET2 to the methylated DNA, and ChIP-qPCR also shows the increased TET2 binding to its targeted regions by co-expression of MBD3L2 in cells. In conclusion, the discovery of the interaction between MBD3L2 and TET2 reveals a novel function of interactor to module TET2 activity in addition to the DNA recruiting mechanism.

5. SNIP1 (The SMAD nuclear interacting protein 1)

SNIP1 was identified as a TET2 interactor via a mammalian two-hybrid screen assay in searching for any human transcription factors or putative DNA-binding proteins that can interact with TET2 and activate the luciferase reporter gene expression (Chen et al., 2018). SNIP1 encodes a protein containing a forkhead-associated (FHA) domain and nuclear localization signal (NLS). It is originally identified as the nuclear interactor of the Smad family proteins and CBP/p300 in which SNIP1 inhibits the function and level of CBP/p300 in nucleus (Kim et al., 2000). SNIP1 also physically interacts with c-Myc and stabilizes the protein against proteosomal degradation and bridge the interaction between c-Myc and p300 complex (Fujii et al., 2006). The endogenous interaction between TET2 and SNIP1 was confirmed in U2OS cells. SNIP1 interacts with TET2 protein via the N-terminal domain of both proteins. SNIP1 can not bind to the full-length TET1 or TET3 proteins. However, SNIP1 itself has no direct DNA binding ability and usually interacts with the other DNA binding proteins such as c-MYC, and further experiments confirm the interaction between c-MYC and TET2 which is in SNIP1-dependent manner. TET2 is recruited by SNIP1 and promotes the DNA demethylation activity at the promoter regions of c-MYC targeted genes which are associated with DNA damage response and cell viability, suggesting a role of TET2-mediated DNA demethylation in maintaining genome stability.

6. PSPC1 (Paraspeckle Component 1)

Pspc1 was identified as a Tet2 interactor in mESCs and can recruit Tet2 to DNA in a RNA-dependent manner (Guallar et al., 2018). Pspc1 itself is a RNA-binding protein and binds to retrotransposon elements (RE). Interestingly, the recruitments of Tet2 to Pspc1-bound RE transcripts such as mouse retroelement MERVL RNAs lead to the formation of 5hmC in RNA and negatively regulate the RNA stability, resulting in their degradation in mESCs. In all, Pspc1 is responsible for the recruitment of Tet2 to RNA and the interaction offers a new role of Tet2 in affecting gene expression by its catalytic activity on RNA stability control.

7. Transcription factors related to hematopoiesis:

Runx1 (runt-related transcription factor 1):

RUNX1 (also known as acute myelogenous leukemia-1 (AML1)) is originally identified in the chromosome translocation events in acute myelogenous leukemia (Ichikawa et al., 2013). RUNX1 is a transcription factor that binds to consensus DNA sequence TGTGGT or TGCGGT and regulates the gene expression of hematopoietic-related genes, and it is essential for the embryonic hematopoiesis since mice with Runx1 depletion show embryonic lethality and have hematopoiesis defects (Okuda et al., 1996), whereas, RUNX1 negatively regulates the cell proliferation of HSCs and myeloid progenitors in adult hematopoiesis. The interaction between Tet2 and Runx1 was identified by flag affinity purification in murine erythroleukemia (MEL) cells with the stable expression of Flag-Tet2 proteins (Chu et al., 2018). Runx1 DNA binding motifs were enriched in Tet2 ChIP-seq peaks in MEL cells. Luciferase reporter assay shows that Tet2 is involved in the transcriptional regulation at Runx1-bound genes. However, the functional importance of TET2 interaction with RUNX1 in TET2-associated hematopoietic processes remain unknown.

8. Transcription factors related to cell reprogramming:

The interactions between TET2 and several transcription factors including C/EBPa, Klf4 and Tfcp211 have been recently identified in a study where they observed the function of

TET2-mediated DNA demethylation in cell-fate regulation (Sardina et al., 2018). TET2-dependent DNA demethylation is involved in the several cell reprogramming processes including C/EBPa-enhanced B cell reprogramming into induced pluripotent cells (iPSCs), and MEF cell reprogramming into iPSCs, and TET2 activity is also vital for the normal hematopoietic cell differentiation such as B cell development and HSC differentiation into myeloid lineage. The interaction between TET2 and C/EBPa was identified in Bα' cells where C/EBPa recruits TET2 protein to highly methylated enhancers and leads to the active DNA demethylation at pluripotency genes involved in cell fate transition of B cell lineage. The interactions between the transcription factors Klf4, Tfcp211 and Tet2 were discovered in mESC and the interactions are important for the gene regulation activity of Tet2 on pluripotency genes in somatic cell reprogramming process. Klf4 recruits Tet2 to closed chromatin regions and induces the DNA demethylation at the targeted genes. The recruitment of Tet2 to DNA by Tfcp211 is responsible for the DNA demethylation at enhancer region of Nanog, which is a key pluripotency TF for successful iPSC reprogramming. In all, the interaction between TET2 and transcription factors at enhancer regions reveals a recruiting mechanism for understanding of TET2-mediated DNA demethylation activity at enhancer regions.

Interactors that mediate post-translational modifications of TET2 protein:

1. CRL4VprBP (VprBP-DDB1-CUL4-ROC1 E3 ubiquitin ligase)

VprBP, also known as DCAF1(DDB1 and CUL4-associated factors 1), is involved in the ubiquitination substrate recruiting for cullin-RING ligases 4 (CRL4) which belongs to the eukaryotic E3 ubiquitin ligases family (Schabla et al., 2019). The endogenous interaction between TET2 and VprBP was identified in mouse embryonic fibroblasts (MEFs), mESCs, and human monocytes (Nakagawa et al., 2015). VprBP interacts with catalytic domain (CD) of all three TETs. However the catalytic activity is not involved in the interaction between VprBP and TET2 since the the catalytic activity-dead CD mutant could still bind to VprBP. VprBP together with CUL4A and CUL4B catalyze the monoubiquitination at the CD of all three TET and promote the TET2 binding to unmethylated DNA, but with no effects on TET enzymatic activity. Importantly, several

TET2 loss-of-function mutations from leukemia patients target the ubiquitination site of TET2 and disrupt the interaction with VprBP, suggesting a function of VprBP-mediated TET2 ubiquitination in inhibiting tumor growth.

2. P300

P300, the histone acetyltransferase (HAT), physically interacts with endogenous TET2 in human ovarian cancer cells and acetylates the lysine 110/111 in the N-terminal of TET2, resulting in an increased TET2 enzymatic activity and enhanced TET2 protein stability (Zhang et al., 2017). Importantly, the histone deacetylases HDAC1/2 which physically interact with TET2 and deacetylate TET2 are often over-expressed in cancers (Glozak and Seto, 2007). However, the interaction between HADC1/2 and TET2 is not observed in other study (Sun et al., 2018), this might be explained by the differences of cells used for interaction identification, where the other paper identified SIRT1 as a unique TET2 deacetylase in human leukemic cells, MDS-L cells.

3. SIRT1 (Sirtuin 1)

TET2 was identified as an acetylated protein that could be deacetylated by SIRT1 in leukemia cells derived from MDS patients (Sun et al., 2018). SIRT1 physically interacts with the CD domain of TET2 protein. The deacetylation at TET2 lysine 1468/ 1472/ 1473/ 1478 by SIRT1 results in enhanced TET2 catalytic activity suggesting a role of acetylation in negative regulation of TET2 enzymatic activity. Importantly, SIRT1 is found down-regulated in hematopoietic stem/progenitor cells (HSPCs) from MDS patients, which might explain the observed decrease of 5hmC in leukemia patients with wild-type TET2 proteins.

Other interactors for TET2-involved gene regulation

1. OGT (O-linked b-N-acetylglucosamine (O-GlcNAc) transferase):

The interaction between OGT and TET2 has been identified in many cell types (Chen et al., 2013; Deplus et al., 2013; Vella et al., 2013; Zhang et al., 2014). OGT catalyzes the O-GlcNAcylation to the hydroxyl group of serine or threonine residues of many nuclear and cytoplasmic proteins (Kreppel et al., 1997). Histone H2B GlcNAcylation mediated

by OGT promotes gene activation by facilitating the H2BK120 monoubiquitination (Fujiki et al., 2011). The interaction between TET2 and OGT was originally identified by IP-MS in 293 T cells with the over-expression of the tagged TET2 proteins (Chen et al., 2013; Deplus et al., 2013). The endogenous interaction between TET2 and OGT was confirmed in mouse ESCs (Chen et al., 2013). The N-terminal of OGT interacts with the catalytic domain of both TET2 and TET3 but not TET1 in 293 T cells (Chen et al., 2013; Deplus et al., 2013). However, other study observed the interaction between Tet1 and Ogt in mouse ESCs and showed Tet1 was involved in the recruiting of Ogt to gene promoter regions (Vella et al., 2013). More importantly, knock-down of TET2 impairs the binding of OGT to its target (Chen et al., 2013). The interaction between TET2 and OGT provides a recruiting mechanism for OGT binding to DNA by TET2 and leads to a TET2-dependent histone H2B O-GlcNAcylation, This suggests a new role of TET2 in gene activation. TET3 can be directly modified by OGT and the O-GlcNAcylation leads to the cytoplasmic localization of TET3 proteins and impairs the normal TET3 enzymatic activity (Zhang et al., 2014).

2. EBF1 (early B-cell factor 1)

EBF1 was identified as a TET2 interactor by searching for transcription factor binding motifs enriched in hypermethylated regions in four types of human IDH-mutant containing cancers including acute myeloid leukaemia (AML), low-grade glioma, cholangiocarcinoma (CC) and chondrosarcoma (CS) (Guilhamon et al., 2013). The endogenous interaction between TET2 and EBF1 was validated in human chondrosarcoma cell line, SW 1353. However the physical interaction mechanism and molecular function of TET2 and EBF1 in regulating TET2 DNA recruiting, enzymatic activity or DNA binding affinity need to be further characterized and the authenticity of the interaction should be carefully examined in other cell background and the biological relevance of interaction in understanding of TET2 function in cancer development need to be explored.

1.6.4 TET2 enzymatic activity regulators

Loss of 5hmC has been observed in many cancers including breast, liver, lung, glioblastoma, melanoma and blood cancers (Kudo et al., 2012; Turcan et al., 2012; Yang et al., 2013). Although TET2 mutations are widely observed in many blood disorders, some AML patients with wild-type TET2 also show similar decrease of 5hmC compared to AML patients with TET2 mutation (Ko et al., 2013). Moreover, TET2 mutation is rarely identified in solid tumors. These observations suggest there are some mechanisms in cancer cells for regulating the TET2 enzymatic activity. A lot of factors including microRNA, small molecules and proteins have been identified for TET2 enzymatic activity regulation through various mechanisms.

TET2 gene expression regulators:

MicroRNAs (miRNAs) are small non-coding RNAs that downregulate the targeted gene expression by affecting the mRNA stability and repressing the translation process. Several miRNAs including miR-125b, miR-29b, miR-29c, miR-101, and miR-7, miR-22 have been identified that down-regulate TET2 expression and cause a cellular decrease of 5hmC (Cheng et al., 2013; Song et al., 2013). Importantly, these TET2-associated miRNAs have been found over-expressed in AML or MDS patients with wild-type TET2 protein which might explain the observed decrease of 5hmC in those patients. The expression of TET2-associated miRNAs in mice also mimic the HSC expansion and skewed myeloid differentiation phenotypes observed in Tet2-deficient mice. Mice with the expression of miR-22 could develop a condition similar to human MDS (Song et al., 2013). Retinoic acid (RA) or retinol (vitamin A) were reported that could enhance TET2/3 gene expression through RA receptor (RAR) binding to TET2 gene, and promote the iPSCs cell reprogramming (Hore et al., 2016).

TET2 protein stability regulator:

The stability of TET2 proteins can be regulated by either the proteolytic systems or by post-translational modifications. The degradation of TET2 protein by caspase was observed during the characterization of the protein interaction between IDAX and TET2

(Ko et al., 2013). Surprisingly, the expression of IDAX in cells leads to the degradation of TET2 protein which requires the activation of caspase 3 and caspase 8 activities. Protease systems inhibitor treatment experiments revealed the calpains-dependent protein degradation for all three Tet proteins in mouse ESCs (Wang and Zhang, 2014). Calpain 1 is suggested to regulate the protein turnover of Tet1 and Tet2 in mESCs since calpain1 depletion leads to an increased Tet1 and Tet2 protein levels in mESCs, whereas, calpain 2 is responsible for the Tet3 protein control in differentiated cells as a result of increased Tet3 observed in calpain2 knockout embryonic body. However the calpains or caspase-mediated protein degradation was not observed in a recent study where the ubiquitin-proteasome pathway was responsible for the TET2 protein degradation in human ovarian carcinoma cells (A2780) and human colorectal carcinoma cells (HCT116) (Zhang et al., 2017). The observation of different protein degradation pathways in regulating TET2 protein level in different cells might suggests a cell-type specific mechanism in controlling TET2 protein level. Several studies have recently reported the positive regulation of TET2 protein stability by AMP-activated kinase (AMPK)-mediated TET2 phosphorylation (human TET2 serine 99/ mouse Tet2 serine 97), which lead to an increase 5hmC level in cells (Wu et al., 2018; Zhang et al., 2019). 14-3-3 phosphorylated serine/threonine-binding proteins bind to the phosphorylated TET2 and further enhance the protein stability by protecting it from the dephosphorylation by protein phosphatase 2A at human TET2 serine 99 (Kundu et al., 2020; Zhang et al., 2019).

Tet2 enzymatic activity co-factors or inhibitors 2-oxoglutarate (2-OG), a TET2 enzymatic activity cofactor, is produced via the IDH enzymes-mediated oxidative decarboxylation of isocitrate. Gain-of-function IDH1/2 mutants abrogate the 2-OG formation, instead, mutant IDH produces 2-hydroxyglutarate (2-HG) (Tommasini-Ghelfi et al., 2019). 2-HG is a structure analog of 2-OG. 2-HG competes with 2-OG and inhibits the enzymatic activity of TET2 and other 2-OG-dependent dioxygenases such histone demethylases (Xu et al., 2011a). Human gliomas and leukemia cells with mutant IDH show an increased global 5mC level (Figueroa et al., 2010; Tommasini-Ghelfi et al., 2019; Turcan et al., 2012). The expression of 2-HG-producing IDH1/2 mutant or the depletion of Tet2 in mouse primary bone marrow cells result in similar hematopoietic phenotypes including

the increased c-Kit+ population and impaired myeloid differentiation (Figueroa et al., 2010).

Ascorbate (Vitamin C) functions as a co-factor of Fe2+/ 2-OG-dependent dioxygenases (Loenarz and Schofield, 2008), to enhance the TET enzymatic activity and lead to an increased global 5hmC in cells (Blaschke et al., 2013). However, ascorbate can not induce the DNA modification changes in regions resistant to DNA demethylation changes such as imprinted regions and retro-elements. Importantly, ascorbate treatments enhance AML cell death by increasing its sensitivity to Poly ADP ribose polymerase (PARP) inhibition which is developed to increase tumor sensitivity to DNA damage induced by chemotherapy (Cimmino et al., 2017). This provides a therapeutic possibility of ascorbate in cancer treatment.

TET2 enzymatic activity can also be regulated by post-translational modifications. P300-catalyzed acetylation on TET2 lysine 110/111 enhance the enzymatic activity of TET2 (Zhang et al., 2017). Interestingly, the SIRT1-mediated TET2 deacetylation on lysine 1468/1472/1473/478 can also promote the enzymatic activity of TET2 (Sun et al., 2018). These results might suggest acetylation on the N-terminal and C-terminal of TET2 protein lead to different consequence in TET2 enzymatic activity control.

1.7 Discussion

Current studies about TET2 focus on its function particularly in hematopoietic-relevant processes since the important role of TET2 activity in inhibiting human hematopoietic malignancies and affecting mouse hematopoietic stem cell self-renewal and myeloid lineage differentiation. Elucidating the mechanisms by which TET2 affects the gene expression in physiological hematopoietic cells hold the key to our understanding of the disease mechanisms after TET2 loss of function in human hematopoietic disorders. Many studies have linked TET2-mediated DNA demethylation activity to its gene regulation function. Particularly, mouse hematopoietic/non-hematopoietic cells with Tet2 depletion and human leukemia cells with TET2 mutations exhibit a DNA hypermethylation phenotype at enhancer regions (Hon et al., 2014; Lu et al., 2014; Rampal et al., 2014;

Rasmussen et al., 2015). Several Tet2-depleted mouse models have been generated and used for the study of Tet2-mediated gene regulation activity in hematopoietic system. Recent studies by Rasmussen and colleagues reported that Tet2 loss leads to the DNA hypermethylation at enhancers in mouse primary hematopoietic cells (GMP cells), thus alters the transcription factors binding at those distal regulatory elements by affecting the chromatin accessibility via the DNA demethylation activity (Rasmussen et al., 2015, 2019). However, the cells used in their study are derived from a Tet2-deficient human AML mouse model where additional expression of human oncogene AML1-ETO is used to facilitate the disease progression. As is indicated in this study, the expression of AML1-ETO could also contribute to the gene expression changes and DNA methylation alterations in the GMP cells, which might affect the accuracy in interpretation of TET2 alone-mediated gene regulation in hematopoiesis.

Another open question regarding to TET2 is its DNA recruiting mechanism and since the lack of clear DNA binding domain, TET2 is suggested to bind its targeted sites via the interactions with other DNA binding partners. Although several TET2 interactors have been identified, little is known about the function relevance of TET2 interactions in hematopoietic background since most of the interaction characterization was done in hematopoietic-irelavant cells. Currently there is no TET2 interactome screening done in physiological hematopoietic cells as a result of the limited availability of the primary hematopoietic stem cells/progenitor cells (HSC/HSPC). Further discovery of hematopoietic-relevant TET2 interactors would be useful to facilitate our understanding of TET2-involved cellular processes and the mechanism by which TET2 mutation leads to tumorigenesis.

1.8 Aims of the study

The overall aim of this study is to elucidate the molecular mechanism by which Tet2 exerts its function in gene regulation in a hematopoietic context. The first goal was to identify Tet2 targets which its expression was affected by Tet2 activity and exhibited direct Tet2 binding. The second goal was to explore the possible mechanisms of Tet2

function in gene regulation via dissecting the role of Tet2 binding in gene regulation and understanding the contribution of Tet2-mediated DNA demethylation in gene regulation. The third goal was to identify molecular partners of Tet2 in the hematopoietic context, thus give us some better insights in understanding the way by which Tet2 gets involved in cellular processes, especially those related to hematopoiesis.

To study Tet2 function in a hematopoietic context, I utilized the mouse hematopoietic progenitor cells, HPC-7, as the cell line model to systematically characterize Tet2 function in gene regulation. *Tet2*-KO HPC-7 cells were generated using CRISPR-Cas9 and the cell differentiation phenotype of the *Tet2*-KO cells was characterized in Chapter 3. To identify genes targeted by Tet2 activity, transcriptomics profiling of *Tet2*-KO cells versus WT cells was performed to discover genes affected Tet2 deficiency (Chapter 4.2.1), Tet2 genomic binding mapping was combined in order to identify genes whose expression were directly regulated by Tet2 binding (Chapter 4.2.3 and 4.2.4). To assess the function of Tet2-mediated DNA demethylation in Tet2-associated gene regulation, TAPS-β analysis was used to reveal the DNA methylation changes caused by Tet2 deficiency (Chapter 4.2.5). The relationship between DNA demethylation and gene regulation would be discussed in Chapter 4.2.7. In addition, I also characterized the genomic features of Tet2 binding sites in HPC-7 cells in Chapter 4.2.3 and 4.2.7.

Prior to the mapping of Tet2 molecular partners in the hematopoietic context (HPC-7), I did a few attempts in optimisations of the protein-protein interaction mapping method (Chapter 5.2.1) and stable gene expression methods in HPC-7 cells (Chapter 5.2.2). Finally, the classical anti-HA IP-MS approach was performed to discover the novel Tet2 interactors in HPC-7 context (Chapter 5.2.3 and 5.2.4). The validation of the newly identified Tet2 interacting partners would be introduced in Chapter 6.

2

Materials and methods

Contents

2.1 Materials

2.1.1 Plasmids

The following plasmids were used in this study. (Table2.1).

Table 2.1: Plasmids used in this study.

Plasmid	Backbone	Insert	Source
pSpCas9(BB)-2A-GFP	PX458	SpCas9-2A-GFP	Addgene 48138
pSpCas9-Tet2-sgRNA1-2A-GFP	PX458	Tet2 sgRNA-1	This study
pSpCas9-Tet2-sgRNA2-2A-GFP	PX458	Tet2 sgRNA-2	This study
pcDNA3-Flag-Tet2	pcDNA3	mouse Flag-Tet2 FL	Addgene 60939
pEF-FH-Tet2	pEF1a	mouse Flag-HA Tet2 FL	Addgene 41710
pcDNA3-Flag-Tet2-corrected	pcDNA3-Flag-Tet2 (Addgene 60939)	mouse Tet2 FL	This study
pcDNA3-Flag-Tet2(N1040)	pcDNA3	mouse Tet2 first 1040 aa	This study
pcDNA3-Flag-Tet2(CD)	pcDNA3	mouse Tet2 catalytic domain	This study
pcDNA3-BirA (R118G)	pcDNA3	BirA R118G mutant	Gyrd-Hansen lab

Continuation of Table 2.1

pcDNA3-Flag-APEX2-NES	pcDNA3	APEX2	Gyrd-Hansen lab
pcDNA3-Flag-mTet2-BirA	pcDNA3	mouse Tet2-BirA	This study
pcDNA3-BirA-Flag-mTet2	pcDNA3Flag-Tet2(Addgene60939)	BirA-mouse Tet2	This study
pcDNA3-Flag-mTet2-APEX2	pcDNA3	mouse Tet2-APEX2	This study
pcDNA3-APEX2-Flag-mTet2	pcDNA3Flag-Tet2(Addgene60939)	APEX2-mouse Tet2	This study
pCMV-Gag-Pol	pCMV	Gag-Pol	Kriaucionis lab
MSCV-IRES-GFP	MSCV	GFP	Addgene 20672
MSCV-Flag-HA-Tet2-IRES-GFP	MSCV-IRES-GFP	Flag-HA mouse Tet2 FL	This study
MSCV-APEX2-Flag-Tet2-IRES-GFP	MSCV-Flag-HA-Tet2-IRES-GFP	APEX2-mouse Tet2 FL	This study
MSCV-APEX2-Flag-3xNLS-IRES-GFP	MSCV	APEX2-Flag-3xNLS	This study
pPB CAG-rtTA-IRES-Neo	Piggybac Transposon	reverse tetracycline transactivator (rtTA)	Goding lab
pPB transposase	Piggybac Transposon	transposase	Goding lab
pPB-TetON-MCS-3xHA	Piggybac Transposon	tetracycline response element	Goding lab
pDONR221	pDONR221		Pavel Savitsky (Oxford)
pPB-CMV-TetON-3XHA-DEST	Piggybac Transposon	tetracycline response element	Pavel Savitsky (Oxford)
pENTR221-Tet2	pDONR221	mouse Tet2 FL	This study
pPB-TetON-Flag-HA-Tet2	pPB-TetON-MCS-3xHA	mouse Tet2 FL	This study
pPB-TetON-3xHA-Tet2	pPB-CMV-TetON-3XHA-DEST	mouse Tet2 FL	This study
Suz12 (NM-199196) Mouse ORF Clone	pCMV6-Entry	Suz12	Origene MR210406
Ezh2 (NM-007971) Mouse ORF Clone	pCMV6-Entry	Ezh2	Origene MR210436
Rbbp7 (NM-009031) Mouse ORF Clone	pCMV6-Entry	Rbbp7	Origene MR206775
Ppp1r9b (NM-172261) Mouse ORF Clone	pCMV6-Entry	Ppp1r9b	Origene MR222264
Ifi204 (NM-008329) Mouse ORF Clone	pCMV6-Entry	Ifi204	Origene MR222527
Alyref (NM-011568) Mouse ORF Clone	pCMV6-Entry	Alyref	Origene MR220236
Noc2l (NM-021303) Mouse ORF Clone	pCMV6-Entry	Noc2l	Origene MR217828
pCMV-Noc2l-3HA	pCMV	Noc2l-3HA	This study
pCMV-Alyref-3HA	pCMV	Alyref-3HA	This study
pCMV-Suz12-3HA	pCMV	Suz12-3HA	This study
pCMV-Ezh2-3HA	pCMV	Ezh2-3HA	This study
pcDNA3-3HA-Ppp1r9b	pcDNA3	3HA-Ppp1r9b	This study
pcDNA3-3HA-Rbbp7	pcDNA3	3HA-Rbbp7	This study

2.1.2 Antibodies

The following antibodies were purchased and used in this study (Table2.2).

Table 2.2: Primary and secondary antibodies

Primary antibodies				
Antibody	Supplier	Catalog number	Application	Dilution
anti-Tet2, clone 9F7	Sigma-Aldrich	MABE1132	WB, ChIP, IF, (used for all the murine Tet2 protein detection unless with specific indication)	WB:1:1000; IF:1:500
anti-Tet2, clone hT2H 21F11	Sigma-Aldrich	MABE462	ChIP	
anti-Tet2 (D9K3E)	Cell Signalling	92529S	WB, ChIP	WB: 1:1000
anti-Tet2	Abcam	abcam 94580	WB, ChIP	
anti-Tet2	Abcam	abcam 124297	WB, ChIP	
anti-TET2	Bethyl	A304-247A	IP	
anti-TET2	Active motif	61389	WB	1:1000
anti-HA (C29F4)	Cell signalling	3724S	WB, IP, ChIP	WB: 1:1000
anti-Flag M2	Sigma-Aldrich	F3165	WB	1:1000
anti-Flag M2	Sigma-Aldrich	F1804	WB, IF	WB:1:1000; IF:1:2000
Rabbit (DA1E) IgG	Cell signalling	3900S	IP	
Rabbit IgG	Abcam	ab27478	IP	
Normal mouse IgG	Sigma-Aldrich	12-371	IP	
anti-OGT	Abcam	96718	WB	1:500
anti-WT1	Santa-cruz	sc-192	WB	1:100
anti-TBP, 3TF1-3G3	Active Motif	61329	WB	1:1000
anti-CTCF	Millipore	07-729	ChIP	
anti-Ppp1r9b	Cell signalling	14136S	WB, IP	WB:1:1000
anti-Coro1c	Proteintech	14749-1-AP	WB	1:1000
anti-Noc2l	abcam	ab75047	WB	1:1000
anti-Fyb1	Sigma-Aldrich	07-546	WB, IP	WB: 1:1000
anti-Suz12	Cell signalling	3737	WB, IP	WB: 1:1000
anti-Ezh2	Cell signalling	5246	WB, IP	WB: 1:3000
anti-Rbbp7	Santa Cruz	sc-377197	WB	1:1000
anti-Rbbp7	Proteintech	20365-1-AP	IP	
anti-Ddx50	Abcam	ab109515	WB	1:1000
anti-Rae1	Abcam	ab36139	WB	1:1000
anti-Rfc5	Santa Cruz	sc-376528	WB	1:1000
anti-Lima1	Santa Cruz	sc-136399	WB	1:1000
anti-Vrk1	Santa Cruz	sc-271061	WB	1:1000
anti-Ifi204	Novus Biologicals	NBP2-27153SS	WB	1:500
anti-Alyref	Abcam	ab202894	WB, IP	WB: 1:10000
anti-Osbpl8	Novus Biologicals	NBP1-59815	WB	1:1000
anti-Esf	Proteintech	123496-1-AP	WB	1:1000
Streptavidin-HRP	Invitrogen	S911	WB	1:2000 RT for 2 h
Streptavidin, Alexa Fluor® 488 Conjugate	Invitrogen	S32354	IF	1:1000
Flow cytometry antibodies				
Antibody	Supplier	Catalog number	Application	Dilution
anti-mouse CD34, FITC (RAM34)	eBioscience	11-0341-81	Flow Cyt	1:1000
anti-mouse CD44, APC (IM7)	eBioscience	17-0441-81	Flow Cyt	1:1000
anti-mouse CD29, PE (HMb1-1)	eBioscience	12-0291-81	Flow Cyt	1:1000

Continuation of Table 2.2

Antibody	Supplier	Catalog number	Application	Dilution
anti-mouse CD117 (c-Kit), PE-Cyanine7 (2B8)	eBioscience	25-1171-81	Flow Cyt	1:30000
anti-mouse CD3e, PE	eBioscience	12-0031-81	Flow Cyt	1:500
anti-Mouse CD45R/B220, PE	BD Biosciences	561878	Flow Cyt	1:5000
anti-Mouse TER-119, PE	BD Biosciences	561071	Flow Cyt	1:1000
anti-mouse CD19, PE (MB19-1)	eBioscience	12-0191-81	Flow Cyt	1:5000
anti-mouse CD11b, PerCP/Cy5.5	Biolegend	101227	Flow Cyt	1:1000
anti-mouse F4/80, APC	Biolegend	123115	Flow Cyt	1:1000
anti-mouse Ly-6G/Ly-6C (Gr-1), Brilliant Violet 421	BioLegend	108433	Flow Cyt	1:300
Purified Rat Anti-Mouse CD16/CD32 (Mouse BD Fc Block)	BD Biosciences	553141	Flow Cyt	
Secondary antibodies				
Antibody	Supplier	Catalog number	Application	Dilution
Alexa Fluor 546 Donkey anti-mouse IgG (H+L)	Invitrogen	A11003	IF	1:1000
Alexa Fluor 647 Goat anti-Rat IgG (H+L)	Invitrogen	A21247	IF	1:1000
Goat Anti-Rabbit IgG (H+L)-HRP Conjugate	Bio Rad	1706515	WB	1:10000
Goat Anti-Mouse IgG (H+L)-HRP Conjugate	Bio Rad	1706516	WB	1:10000
Goat anti Rat IgG (H/L)-HRP	Bio Rad	5204-2504	WB	1:10000

2.1.3 Cell lines

The following cell lines were used in this study(Table2.3)

Table 2.3: Cell line used in this study.

Cell line	Source
293T	Kriaucionis lab
ES-J1	Kriaucionis lab
Plat-E (Platinum-E)	Kriaucionis lab
DF-1 (Doug Foster strain 1)	Alison Turner, Jenner Institute(Oxford)
CHO-K3	Kriaucionis lab
HPC-7	Kriaucionis lab
HPC-7 Tet2-KO clone 1-3	This study
HPC-7 Tet2-KO clone 1-8	This study
HPC-7 Tet2-KO clone 2-1	This study
HPC-7 Tet2-KO clone 2-7	This study
HPC-7 Tet2-KO clone 2-9	This study
HPC7 Flag-HA-Tet2 FL	This study
HPC7 APEX2-Flag-Tet2 FL	This study
HPC7 APEX2-3 x NLS	This study
HPC7 Tet2-KO 2-1/ Tet-On Flag-HA-Tet2 FL	This study
HPC7 Tet2-KO 2-1/ Tet-On 3x HA-Tet2 FL	This study
HPC7 Tet2-KO 2-7/ Tet-On Flag-HA-Tet2 FL	This study
HPC7 Tet2-KO 2-7/ Tet-On 3x HA-Tet2 FL	This study

Cell lines constructed in this study and its construction method were listed in (Table2.4)

Table 2.4: Cell line constructed in this study.

Cell line	Construction method	Vector	Parental	Stable selection
HPC-7 Tet2-KO clone 1-3	CRISPR-Cas9	pSpCas9-Tet2-sgRNA1-2A-GFP	HPC-7	
HPC-7 Tet2-KO clone 1-8	CRISPR-Cas9	pSpCas9-Tet2-sgRNA1-2A-GFP	HPC-7	
HPC-7 Tet2-KO clone 2-1	CRISPR-Cas9	pSpCas9-Tet2-sgRNA2-2A-GFP	HPC-7	
HPC-7 Tet2-KO clone 2-7	CRISPR-Cas9	pSpCas9-Tet2-sgRNA2-2A-GFP	HPC-7	
HPC-7 Tet2-KO clone 2-9	CRISPR-Cas9	pSpCas9-Tet2-sgRNA2-2A-GFP	HPC-7	
HPC7 Flag-HA-Tet2 FL	Retrovirus	MSCV-Flag-HA-Tet2-IRES-GFP	HPC-7	GFP
HPC7 APEX2-Flag-Tet2 FL	Retrovirus	MSCV-APEX2-Flag-Tet2-IRES-GFP	HPC-7	GFP
HPC7 APEX2-3 x NLS	Retrovirus	MSCV-APEX2-Flag-3xNLS-IRES-GFP	HPC-7	GFP
HPC7 Tet2-KO 2-1/ Tet-On Flag-HA-Tet2 FL	PiggyBac Transposon	pPB-TetON-Flag-HA-Tet2	HPC-7	Puromycin (0.5 μg/mL)
HPC7 Tet2-KO 2-1/ Tet-On 3x HA-Tet2 FL	PiggyBac Transposon	pPB-TetON-3xHA-Tet2	HPC-7	Puromycin (0.5 μg/mL)
HPC7 Tet2-KO 2-7/ Tet-On Flag-HA-Tet2 FL	PiggyBac Transposon	pPB-TetON-Flag-HA-Tet2	HPC-7	Puromycin (0.5 μg/mL)
HPC7 Tet2-KO 2-7/ Tet-On 3x HA-Tet2 FL	PiggyBac Transposon	pPB-TetON-3xHA-Tet2	HPC-7	Puromycin (0.5 μg/mL)

2.1.4 Buffers

The following buffers were made and used in this study(Table2.5)

Table 2.5: Homemade buffers used in this study.

Buffer	Composition
TAE buffer	40 mM Tris base, 2 mM EDTA, 20 mM acetic acid, pH 8.5
SDS running buffer	25 mM Tris base, 192 mM Glycine, 0.1% SDS, pH 8.3
Transfer buffer	25 mM Tris, 192 mM glycine, pH 8.3, 20% methanol (vol/vol)
TBST buffer	20 mM Tris, 150 mM NaCl, 0.05% Tween 20. Adjust pH with HCl to pH 7.4
2 x SDS loading buffer	125 mM Tris-HCl, pH 6.8, 4% (g/mL) SDS, 0.05% (g/mL) bromophenol blue, 10% (mL/mL) ß-mercaptoethanol, 20% (mL/mL) glycerol
5 x SDS loading buffer	200 mM Tris-HCl pH 6.8, 400mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol
2x HBS	280 mM NaCl, 50 mM HEPES, 1.5 mM Na_2HPO_4 12 mM Glucose, 10 mM KCl, pH 6.96
Quencher solution	DPBS containing 10 mM sodium ascorbate, 10 mM sodium azide, and 5 mM Trolox

Continuation of Table 2.5

Buffer	Composition
Whole cell lysis buffer	150 mM NaCl,1 mM EDTA, 10 mM Tris-HCl pH 8.0, 0.5% NP-40, 1x Roche protease inhibitor cocktails EDTA-free, 1 mM PMSF, 100 Units Benzonase (SGC home-made)
Sucrose cell lysis buffer	320 mM sucrose, 10 mM Tris pH 8.0, 50 mM KCl, 20 mM NaCl, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.2 % NP-40, 1 mM Dithiothreitol(DTT), 0.15 mM spermine, 0.5 mM spermidine, 1x Roche protease inhibitor
Low salt buffer	10 mM Hepes pH 7.9, 20% glycerol, 20 mM KCl, 2 mM $MgCl_2$, 0.5 mM DTT, 1x Roche protease inhibitor, 1x Roche PhosSTOP phosphatase inhibitor
High salt buffer	10 mM Hepes pH 7.9, 20% glycerol, 700 mM KCl, 2 mM $MgCl_2$, 0.5 mM DTT, 1x Roche protease inhibitor, 1x Roche PhosSTOP phosphatase inhibitor
Streptavidin pull-down wash buffer	20 mM Tris pH 8, 250 mM KCl, 2 mM $MgCl_2$, 10% glycerol, 1 mM DTT, 1 x PhosSTOP phosphatase inhibitor, 1 x protease inhibitor cocktail
mTet1 CD oxidation buffer	167 mM HEPES pH8.0, 333 mM NaCl, 3.3 mM α-Ketoglutaric acid, 6.67 mM L-Ascorbic acid, 4 mM ATP, 8.33 mM DTT
Buffer LB1	50 mM HEPES KOH pH 7.9, 140mM NaCl, 1 mM EDTA, 10% Glycerol, 0.25% NP40, 0.25% Triton X-100
Buffer LB2	10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA
Buffer LB3	10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N- lauroylsarcosine
Chromatin Dilution Buffer	20 mM Tris, pH 8.0, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA
ChIP-Low Salt Buffer	20 mM Tris pH 8.0, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA
CHIP-High Salt Buffer	50 mM HEPES pH7.5, 500 mM NaCl, 0.1% deoxycholic acid, 1% Triton X-100 and 1 mM EDTA
Lithium Chloride Buffer	10 mM Tris pH 8.0, 250 mM LiCl, 1% deoxycholic acid, 1% NP-40 and 1 mM EDTA
TE Buffer	10 mM Tris-HCl pH 8.0, 1 mM EDTA
ChIP-Elution Buffer	1% SDS and 0.1 M $NaHCO_3$

2.2 Molecular cloning

The following plasmids were cloned by using the strategies illustrated in the table (2.6).

Table 2.6: Plasmids cloned in this study.

Plasmid	Backbone	Insert	Cloning method
pSpCas9-Tet2-sgRNA1-2A-GFP	PX458	Tet2 sgRNA-1	Restriction enzyme ligation
pSpCas9-Tet2-sgRNA2-2A-GFP	PX458	Tet2 sgRNA-2	Restriction enzyme ligation
pcDNA3-Flag-Tet2-corrected	pcDNA3 Flag-Tet2 (Addgene 60939)	mouse Tet2 FL	Restriction enzyme ligation
pcDNA3-Flag-Tet2(N1040)	pcDNA3	mouse Tet2 first 1040 aa	Restriction enzyme ligation
pcDNA3-Flag-Tet2(CD)	pcDNA3	mouse Tet2 catalytic domain	Restriction enzyme ligation

Continuation of Table 2.6

Plasmid	Backbone	Insert	Cloning method
pcDNA3-Flag-mTet2-BirA	pcDNA3	mouse Tet2-BirA	Restriction enzyme ligation
pcDNA3-BirA-Flag-mTet2	pcDNA3 Flag-Tet2(Addgene60939)	BirA-mouse Tet2	Gibson assembly
pcDNA3-Flag-mTet2-APEX2	pcDNA3	mouse Tet2-APEX2	Restriction enzyme ligation
pcDNA3-APEX2-Flag-mTet2	pcDNA3 Flag-Tet2(Addgene60939)	APEX2-mouse Tet2	Gibson assembly
MSCV-Flag-HA-Tet2-IRES-GFP	MSCV-IRES-GFP	Flag-HA mouse Tet2 FL	Restriction enzyme ligation
MSCV-APEX2-Flag-Tet2-IRES-GFP	MSCV-Flag-HA-Tet2-IRES-GFP	APEX2-mouse Tet2 FL	Gibson assembly
MSCV-APEX2-Flag-3xNLS-IRES-GFP	MSCV-IRES-GFP	APEX2-Flag-3xNLS	Restriction enzyme ligation
pENTR221-Tet2	pDONR221	mouse Tet2 FL	Gateway cloning
pPB-TetON-Flag-HA-Tet2	pPB-TetON-MCS-3xHA	mouse Tet2 FL	Restriction enzyme ligation
pPB-TetON-3xHA-Tet2	pPB-CMV-TetON-3XHA-DEST	mouse Tet2 FL	Gateway cloning
pCMV-Noc2l-3HA	pCMV-Noc2l	3xHA	Gibson assembly
pCMV-Alyref-3HA	pCMV-Alyref	3xHA	Gibson assembly
pCMV-Suz12-3HA	pCMV-Suz12	3xHA	Gibson assembly
pCMV-Ezh2-3HA	pCMV-Ezh2	3xHA	Gibson assembly
pcDNA3-3HA-Ppp1r9b	pcDNA3	3HA-Ppp1r9b	Restriction enzyme ligation
pcDNA3-3HA-Rbbp7	pcDNA3	3HA-Rbbp7	Restriction enzyme ligation

2.2.1 Polymerase chain reaction

Phusion High-Fidelity DNA Polymerase (NEB, M0530L) was used for the amplification of most of the inserts for cloning according to the manufacturer's instructions. 50 μL PCR reactions were set up on ice as follows: 10 ng of plasmid template DNA, 2.5 μL 10 μM forward primer, 2.5 μL 10 μM reverse primer, 1 μL 10 mM dNTPs, 10 μL 5X Phusion HF, 0.5 μL Phusion DNA Polymerase and nuclease-free H_2O. Primers used for the construction of vectors based on restriction enzyme ligation method were shown in table (2.7). Samples were run in Veriti Thermal Cycler (4375786,Applied Biosystems) with following programs: initial denaturation at 98^{o}C for 30 s, denaturation at 98^{o}C for 10 s, annealing reaction for 30 s, extension at 72^{o}C for 30 s/kb, final extension at 72^{o}C for 10 min, 30 cycles were used. Successful amplification was examined by loading 5 μL PCR

products on 1% agarose gel and running in TAE buffer in the horizontal electrophoresis system (1704401, Biorad) at 120 V for 40 min. Gel was visualized using Chemidoc (Bio-Rad, XRS). PCR products were purified using NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel). Purified PCR products were used for following restriction enzyme digestion or stored at -20^{o}C.

Table 2.7: Primers used in restriction enzyme ligation cloning.

Primer pair	Sequence	Target	Template	Plasmid cloned	RE
SK4420 SK4421	GGTGGTTCCGGACTAAGTGATAATGG GGTGGTGCGGCCGCTCATTCACCAGGAGTATCTGATT	Tet2 first 1040 aa	pEF-FH-Tet2	pcDNA3-Flag-Tet2(N1040)	BspEI NotI
SK4422 SK4423	GGTGGTGAATTCGCCCTTTACAAAGTCAGAATGGCAAATGT GGTGGTGAATTCTATCATACAAATGTGTTGTAAGG	Tet2 catalytic domain	pEF-FH-Tet2	pcDNA3-Flag-Tet2(CD)	EcoRI
SK1900 SK1901	GGTGGTCTCGAGCAATTGGGCGGTGGAGGTATGAGCAGAGCCGCCATCAACAAG GGTGGTTCTAGATCAAGCCTTCTCTGCGCTTCTCAGG	BirA	pcDNA3-BirA (R118G)	pcDNA3-Flag-mTet2-BirA	XhoI XbaI
SK1898 SK1899	GGTGGTCTCGAGCAATTGGGCGGTGGAGGTGGAAAGTCTTACCCAACTGTGAGTG GGTGGTTCTAGATCAGGCATCAGCAAACCCAAGCTCG	APEX2	pcDNA3-Flag-APEX2-NES	pcDNA3-Flag-mTet2-APEX2	XhoI XbaI
SK3037 SK3038	GGTGGTGAATTCGCCACCATGGGAAAGTCTTACCCAAC GGTGGTCTCGAGTCACACCTTCCTCTTCTTCTTGGGGTCCACCTTCCTCTTCTTCTTGGGGTCCACCTTCTCTTCTTCTTGGGGCCGTTACTAGTGGATCC	APEX2	pcDNA3-Flag-mTet2-APEX2	MSCV-APEX2-Flag-3xNLS-IRES-GFP	EcoRI XhoI
-	-	Flag-HA-Tet2	pEF-FH-Tet2	pPB-TetON-Flag-HA-Tet2	SpeI NotI
-	-	Flag-HA-Tet2	pEF-FH-Tet2	MSCV-Flag-HA-Tet2-IRES-GFP	EcoRI
SK4426 SK4427	GGTGGTGGATCCATGGCGAGTAAAGAGATG GGTGGTGCGGCCGCTCAAGATCCTTGCCCCTCCAGTT	Rbbp7	pCMV-Rbbp7	pcDNA3-3xHA-Rbbp	BamHI NotI
SK4430 SK4431	GGTGGTGAATTCGGATCCATGATGAAGACGGAGCCTCG GGTGGTGAATTCTCAAGTAGAATTGGAGTTCCTCA	Rbbp7	pCMV-Ppp1r9b	pcDNA3-3xHA-Ppp1r9b	EcoRI

2.2.2 Restriction enzyme digestion and dephosphorylation

1 μg purified PCR products or 1 μg cloning vector were digested with desired restriction enzymes (New England Biolabs) at 37^{o}C for 1h according to the manufacturer's instructions. For single enzyme digestion, digested vector was dephosphorylated using calf intestinal phosphatase (CIP) (M0290, New England Biolabs) at 37°C for 30 min before purification. Digested PCR products were purified using NucleoSpin Gel and PCR Clean-up kit. Digested vector was run on an agarose gel and the desired vector fragment was purified using NucleoSpin Gel and PCR Clean-up kit.

2.2.3 Ligation with T4 DNA ligase

100 ng linearized vector DNA was ligated to the PCR products using a molar ratio of 1:5 (vector to insert DNA). Ligation reaction was set up using 1 μL T4 DNA ligase (M0202, New England Biolabs) in a final 20 μL volume and incubated at room temperature for 1 h or 16^{o}C overnight.

2.2.4 Bacterial transformation

Ligation product was transformed to into the DH5α competent bacterial cells as follows: Competent cells were thawed on ice for 10 min before use. For each transformation, 5 μL ligation products were added into 100 μL competent cells and incubated on ice for 30 min, followed by heat-shock at 42^{o}C for 30 s, and ice incubation for 2 min. Transformed bacterial cells were added with 700 μL S.O.C media (15544034, Invitrogen) and grown at 37^{o}C for 1 h. 200 μL bacteria culture was plated on a pre-warmed LB agar plate supplemented with antibiotics. 100 μg/mL were used for Ampicillin plates and 50 μg/mL were for Kanamycin plates. Bacteria were grown at 37 ^{o}C overnight. Bacteria colonies were grown in 2 mL LB media supplemented with antibiotics at at 37^{o}C overnight, followed by plasmid purification using QIAprep Spin Miniprep Kit (27104, Qiagen). Plasmids were sent for sanger sequencing in Eurofins Genomics.

2.2.5 Gibson assembly

Vector was linearized by restriction enzyme digestion and linearized vector was assembled with corresponding PCR amplified-DNA fragments using the NEBuilder HiFi DNA Assembly cloning kit (E5520S, New England Biolabs) according to the manufacturer's instructions. Information about DNA fragments and restriction enzymes were listed in table(2.8). Briefly, 100 ng linearized vector was used in a 20 μL reaction and assembled with 1-2 DNA fragments using a molar ratio of 1:2 (vector to insert DNA). A molar ratio of 1:1 was used for 4-6 fragments assembly. 10 μL 2x HiFi DNA Assembly Master Mix was added and incubated at 50^{o}C for 1 h. 2 μL assembly products were transformed to the NEB 5-alpha Competent *E. coli* or stored at -20^{o}C.

Table 2.8: Primers used in Gibson assembly DNA fragments amplification .

Primer pair	Sequence	Template	Target	Plasmid cloned	RE to linearize vector
SK1931	AACTTCAGGGTGAGTTTGGGGAC	pcDNA3Flag-Tet2(Addgene60939)	BirA-Tet2 frag1	pcDNA3-BirA-Flag-Tet2	PpuMI BspEI
SK1937	CTCTGCTCATCATAAGCTTGGGGTCGAC				
SK1938	CAAGCTTATGATGAGCAGAGCCGCCATC	pcDNA3-BirA (R118G)	BirA-Tet2 frag2		
SK1939	CTTTGTAGTCACCTCCACCGCCCAATTGAGCCTTCTCTGCGCTTCTC				
SK1935	CGGTGGAGGTGACTACAAAGACGATGAC	pcDNA3-Flag-Tet2(Addgene60939)	BirA-Tet2 frag3		
SK1936	GCTCTCCATTATCACTTAGTC				
SK1931	AACTTCAGGGTGAGTTTGGGGAC	pcDNA3Flag-Tet2(Addgene60939)	APEX2-Tet2 frag1	pcDNA3-APEX2-Flag-mTet2	PpuMI BspEI
SK1937	CTCTGCTCATCATAAGCTTGGGGTCGAC				
SK1933	CAAGCTTATGATGGGAAAGTCTTACCCAACTG	pcDNA3-Flag-APEX2-NES	APEX2-Tet2 frag2		
SK1934	CTTTGTAGTCACCTCCACCGCCCAATTGGGCATCAGCAAACCCAAGCTC				
SK1935	CGGTGGAGGTGACTACAAAGACGATGAC	pcDNA3Flag-Tet2(Addgene60939)	APEX2-Tet2 frag3		
SK1936	GCTCTCCATTATCACTTAGTC				
SK2332	CCCTGGACCTATGGGCAAGCCCATCCCCAACCCCCTGCTGGGCCTGGACAGCACCGGAAAGTCTTACC-CAACTGTGA	pcDNA3-APEX-TET2	APEX2		
SK2333	TGTCCTGTTCACCTCCACCGCCCAATTGGGCATCAGC				

Continuation of Table 2.8

Primer pair	Sequence	Template	Target	Plasmid cloned	RE
SK2334	CGGTGGAGGTGAACAGGACAGAACCACCC	HPC-7 G-DNA	Tet2 exon3 ATG 3'arm		
SK2335	ATTCGAGCTCGGTACCCGGGTTCTTCAGCAAATAGCTCCATG				
SK2332	CCCTGGACCTATGGGCAAGCCCATCCCCAACCCCCTGCTGGGCCTGGACAGCACCGGAAAGTCTTACCCAACTGTGA	pcDNA3-APEX-TET2	APEX2		
SK2333	TGTCCTGTTCACCTCCACCGCCCAATTGGGCATCAGC				
SK2334	CGGTGGAGGTGAACAGGACAGAACCACCC	HPC-7 G-DNA	Tet2 exon3 ATG 3'arm		
SK2335	ATTCGAGCTCGGTACCCGGGTTCTTCAGCAAATAGCTCCATG				
SK3026	TGTTTGCGCCTGCGTCTGTAC	MSCV-Flag-HA-Tet2-IRES-GFP	APEX2-Tet2 frag1	MSCV-APEX2-Flag-Tet2-IRES-GFP	SpeI BspEI
SK3027	CCATGGTGGCATCGAGCGGAAGATCTAAG				
SK3036	TCCGCTCGATGCCACCATGGGAAAGTCTTACCCAACTG	pcDNA3-APEX2-Tet2	APEX2-Tet2 frag2		
SK3029	GCTCTCCATTATCACTTAGTCCGGAGACATTTGGCTGAC				
SK4572	CAGAAGAGGACAGCGGACCGACGCGTACGCGGCCGCTCTACCCATACGATGTTCCAGATTACGCTGGCTATCCCTATGACGTCCCGGACTATGCAGGATCCTATCCATATGACGTTCCAGATTACGCTTAAGATATCCTGGATTACAAGGATGACGACGATAAG			pCMV-Noc2l-3xHA	NotI EcoRV
SK4573	CTGGATGCTTACAATGCAAGGATGGACACCAGCACGCGTACGCGGCCGCTCTACCCATACGATGTTCCAGATTACGCTGGCTATCCCTATGACGTCCCGGACTATGCAGGATCCTATCCATATGACGTTCCAGATTACGCTTAAGATATCCTGGATTACAAGGATGACGACGATAAGG			pCMV-Alyref-3xHA	MluI EcoRV
SK4574	AGAGCAAGAAACAAAAACTCACGCGTACGCGGCCGCTCTACCCATACGATGTTCCAGATTACGCTGGCTATCCCTATGACGTCCCGGACTATGCAGGATCCTATCCATATGACGTTCCAGATTACGCTTAAGATATCCTGGATTACAAGGATGACG			pCMV-Suz12-3xHA	MluI EcoRV

Continuation of Table 2.8

Primer pair	Sequence	Template	Target	Plasmid cloned	RE
SK4575	CGAGAAATGGAAATCCCTACGCG TACGCGGCCGCTCTACCCAT-ACGATGTTCCAGATTACGCTG-GCTATCCCTATGACGTCC-CGGACTATGCAGGATCCTATCCATAT GACGTTCCAGATTACGCTTAA-GATATCCTGGATTACAAGGAT-GACGACGATAAG			pCMV-Ezh2-3xHA	NotI EcoRV

2.2.6 Gateway cloning

PiggyBac TetOn-3xHA-Tet2 expression vectors were constructed using Gateway cloning method as follows: attB recombination sites were added to the 5' and 3' end of Tet2 cDNA by PCR on pEF-FH-Tet2 vector using primers (SK3682: GGGACAAGTTTGTA-CAAAAAAGCAGGCACCgaacaggacagaaccacccatgctg; SK3683: GGGACCACTTTG-TACAAGAAAGCTGGtcatacaaatgtgttgtaaggccct). attB-Tet2-attB DNA was inserted into pDONR221 vector to get pENTRY Tet2 vector by BP Clonase-mediated recommbination between attB and attP sites. 100 ng attB-Tet2-attB, 100 ng pDONR221 vector, 2 μL BP Clonase II enzyme mix (11789, Invitrogen) and TE, pH 8.0 buffer were used to assemble a 10 μL BP reaction, and incubated at RT for 4 h. Reaction was terminated by adding 1 μL Proteinase K (AM2546, Invitrogen) and incubation at 37 ^{o}C for 10 min. 1 μL BP reaction product was transformed to bacterial competent cells. Colonies were cultured in LB media supplemented with Kanamycin. pENTRY Tet2 plasmids were extracted and sent for Sanger sequencing. Tet2 was transferred from pENTRY vector to pPB-TetON-3xHA-DEST vector by LR clonase-mediated recombination between attL and attR sites. 100 ng pENTRY Tet2, 100 ng pPB-TetON-3xHA-DEST, 2 μL LR Clonase II enzyme mix (11791, Invitrogen) and TE, pH 8.0 buffer were used to assemble a 10 μL LR reaction, and incubated at RT for 2 h. Reaction was terminated by adding 1 μL Proteinase K (AM2546, Invitrogen) and incubation at 37 ^{o}C for 10 min. 1 μL LR reaction product was transformed to bacterial competent cells . Colonies were cultured in LB media supplemented with Ampicillin. Plasmids were purified and sent for Sanger sequencing.

2.3 Genomic DNA extraction

Suspension cells were harvested by centrifugation at 300 g for 5 min. Cells were washed once with cold PBS. Genomic DNA was extracted using the Gene Jet Genomic DNA Extraction Kit (K0721, Thermo Scientific) according to the manufacturer's instructions.

2.4 DNA hydrolysis for LC/MS-MS analysis

500 ng of DNA in 20 μL of nuclease-free water was added to 200 μL of hydrolysis solution (100 mM NaCl, 20 mM $MgCl_2$, 20 mM Tris pH 7.9, 1000 U/mL Benzonase (E8263, Sigma-Aldrich), 600 mU/mL Phosphodiesterase I, 80 U/mL Alkaline phosphatase, 36 μg/mL EHNA hydrochloride, 2.7 mM deferoxamine). The mixture was incubated for two hours and then lyophylized by SpeedVac. The lyophylisate was resuspended in 1000 μL of buffer A (10 mM ammonium acetate, pH 6) and 300 μL was transferred into an LC-MS vial for analysis. The HPLC–MS mass spectrometry analysis was performed by Paolo Spingardi.

2.5 RNA extraction

Suspension cells were harvested by centrifugation at 300 g for 5 min. Cells were washed once with cold PBS. RNA was extracted using the RNeasy Mini kit (74104, Qiagen) according to the manufacturer's instructions. Briefly, cells were lysed in buffer RLT and homogenized by passing them through the QIAshredder (79656, Qiagen). On-column DNase digestion was performed with RNase-free DNase Set (79256, Qiagen) according to the manufacturer's instructions. RNA concentrations were measured by Nanodrop 2000 (Thermo Scientific) and adjusted to 100 ng/μL with nuclease-free water.

2.6 RT-PCR and data analysis

1 μg RNA was used for cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kits (4368814, Applied Biosystem) according to the manufacturer's instructions. Briefly, for each reaction, 10 μL 2x RT master mix was firstly prepared as follows: 2

μL 10x RT buffer, 0.8 μL 25 x dNTP mix (100 mM), 2 μL 10x RT random primers, 1 μL MultiScribeTM reverse transcriptase, 0.5 μL RiboLock RNase Inhibitor (EO0381, Thermo Scientific), and 3.7 μL nuclease-free H_2O. 10 μL 100 ng/μL RNA was mixed with the 2x RT master mix, and the resultant mixture was incubated in a thermomixer (5382000031, Eppendorf) at 25 ^{o}C for 10 min, 37 ^{o}C for 120 min, and 85 ^{o}C for 5 min. Synthesized cDNA was diluted 1:50 using nuclease-free H_2O and stored at -20 ^{o}C until use.

PCR reaction was prepared on ice as follows: 12.5 μL 2x QuantiFast SYBR green mix (204056, Qiagen), 0.25 μL 100 μM forward primer, 0.25 μL 100 μM reverse primer, 10 μL diluted cDNA, and 2 μL nuclease-free water. Primers used for each gene were listed in table. Reactions were run in a 96-well plate format in a StepOnePlus Thermal Cycler (4376600, Applied Biosystems). Samples were heated at 95 ^{o}C for 5min, followed by 40 cycles at 95 ^{o}C for 10s and 60 ^{o}C for 30s after which a melt curve analysis was performed by ramping the temperature from 50 ^{o}C to 95 ^{o}C (5s per ^{o}C). All measurements were done using three biological replicates and three technical duplicate for each biological replicate. Relative gene expression was determined by comparative CT method (Schmittgen and Livak 2008), and GADPH was used as the housekeeping gene for normalization.

2.7 Cell culture methods

All cell lines used in this study were cultured in a 37 ^{o}C incubator with a 5 % CO_2 atmosphere. And cells were checked for Mycoplasma contamination monthly using the MycoAlertTM mycoplasma detection kit (LT07-318, Lonza)

2.7.1 HPC-7 cell culture

HPC-7 cells were obtained as a gift from Dr Lief Carlsson (Umea, Sweden). Frozen cells were thawed and cultured in Stem Pro-34 media (10639011, Gibco) supplemented with 10% Stem Pro supplements, 150 μM MTG (M6145, Sigma-Aldrich), 2 mM L-glutamine (25030081, Life Technologies), 1x penicillin- streptomycin (15140122, Life Technologies) and 100 ng/mL murine stem cell factor (250-03-100, PeproTech) until it

achieves normal growth speed. Then cells were grown in IMDM medium (12440061, Gibco) supplemented with 5% fetal bovine serum (FBS, Biosera), 150 μM MTG, 1% Penicillin/Streptomycin, 2 mM L-glutamine and 10% SCF-conditioned media. Cells were subcultured every 2 days. Cell density was maintained between 2 x 10^5-1.5 x 10^6 cells/mL. When freezing cell, 10^7 cells were pelleted at 300 g for 5 min at RT, and resuspended in 1 mL Stem Pro-34 media supplemented with 10% Stem Pro supplements, 150 μM MTG, 2 mM L-glutamine, 1x penicillin- streptomycin, 40% FBS, and 10 % DMSO (D2650, Sigma -Aldrich). Cell suspension was then transferred into 1.8 mL cryogenic tube (377267, Thermo Scientific) and tubes were placed in Corning CoolCell Cell Freezing Container (CLS432001, Sigma-Aldrich) at -80^{o}C for 24 h. For long-term storage, frozen cells were transferred to liquid-nitrogen.

SCF-conditioned media were produced by CHO-K3 cell line. Briefly, CHO-K3 cells were grown in DMEM media supplemented with 10 % FBS until 100 % confluency and subsequently grown in IMDM culture media (supplemented with 5% FBS, 150 μM MTG, 1% Penicillin/Streptomycin, and 2 mM L-glutamine) for 1 day, then cells were grown in fresh IMDM culture media for other 2 days. SCF-conditioned media were harvested and filtered through 0.45 μM filter. For longer storage, conditioned media were aliquoted and stored at -80^{o}C.

2.7.2 Adherent cell culture

293 T cells and Plat-E cells were grown in DMEM media supplemented with 10 % FBS, 1% Penicillin/Streptomycin and 2 mM L-glutamine. Cells were subcultured every 2 days. ES-J1 cells were grown in 0.1 % gelatin (G1890-100G, Sigma-Aldrich)-coated dishes in DMEM (BE12-604F, Lonza) supplemented with 10 % FBS (FB-1001G/500, lot-1148 4, Biosera), 1% Penicillin/Streptomycin and 2 mM L-glutamine, 1x non-essential amino acids (11140050, Life Technologies), leukemia inhibitory factor (LIF, produced in-house following protocol by Tomala et al., 2010). Cells were subcultured every 2-3 days.

When subculturing, cells were washed with PBS once and treated with 0.05 % Trypsin-EDTA (25300054, Life Technolofies) at 37 ^{o}C until they were detached. Trypsin

was inactivated by adding equal volume of cell culture media. Detached cells were resuspended into cell culture media, split at 10-20 % confluency and seeded onto appropriate dishes. When freezing cells, cells were grown to 90 % confluency in a 10 cm dish. Cells were washed with PBS, detached by trypsinisation and pelleted at 300 g for 5 min at RT. After pelleting, cells were resuspended in 1 mL DMEM growth media supplemented with 20 % FBS and 10 % DMSO. Cells suspension was then transferred into cryogenic tube and stored in CoolCell Cell Freezing Container at -80^{o}C for 24 h. Frozen cells were transferred to liquid-nitrogen for long-term storage.

2.7.3 Transfection

For transfection, all plasmid DNA were prepared using NucleoBond Xtra Midi plus kit (740412.50, Macherey-Nagel).

For polymer-mediated transfection, 293 T cells were seeded at 40 % confluency on a 10 cm dish prior to transfection. The next day, cells were transfected using jetPRIME transfection reagent (114-15, Polyplus) according to the manufacturer's instructions. For each reaction, 5 μg plasmid DNA, 500 μL jetPRIME buffer and 15 μL jetPRIME were mixed and incubated at RT for 10 min. Transfection mixture was added dropwise onto the cells. 4 h after transfection, media were replaced by fresh growth media. 24 h after transfection, cells were harvested for Western blot analysis, or used for other applications.

For calcium phosphate-mediated transfection, Plat-E cells were seeded at 50% confluency in 9 ml media on a 10 cm dish, grown for 8 h, and then transfected by calcium-phosphate precipitation. For each transfection, 500 μL sterile water containing 20 μg MSCV expression vector, 10 μg helper plasmid (pCMV-Gag-Pol), and 62.5 μL 2M $CaCl_2$ was added dropwise with gentle agitation into equal volume of 2x HBS buffer (pH 6.96). The mixture was incubated at RT for 15 min and added dropwise onto Plat-E cells in the presence of 25 μM Chloroquine (C6628, Sigma-Aldrich).

2.7.4 Electroporation

HPC-7 cells were electroporated with plasmid DNA using Neon transfection system (MPK5000, Invitrogen) according to the manufacturer's instructions. For electroporation

on a 6-well plate per well, 2 x 10^6 cells were washed once with PBS and pelleted by centrifugation at 300 g for 5 min. Cells were resuspended in 100 μL Resuspension buffer R at a final density 2 x 10^7/mL, 10 μg plasmid DNA was mixed into the cell suspension. Cell/DNA mixture was transferred into a 100 μL Neon tip and electroporated under the following condition: 1650 V, 20 ms,1 pulse. Electroporated cells were immediately transferred to a 6-well plate containing prewarmed 4 mL Stem Pro growth media per well. 24 h post electroporation, cells were analyzed for transient fluorescent protein expression by flow cytometry. 72 h post electroporation, antibiotics was added into media for stable cell line selection.

2.7.5 Generation of knock-out cell line by CRISPR-Cas9 method

For the generation of HPC-7 Tet2 Knock-out (KO) clones, two pairs of guide RNA (gRNAs) were designed and used for targeting mouse Tet2 (Table (2.9)). gRNAs were ordered from IDT Technologies and ligated into pSpCas9(BB)-2A-GFP (PX458, Addgene 48138) according to the protocol from Ran et al., 2013.

Table 2.9: Guide RNAs used to knock out Tet2.

Primer pair	Sequence	Target	gRNA source	Plasmid cloned
SK2252	CACCGCGAAGCAAGCCTGATGGAAC	Tet2 exon3 first ATG	This study	pSpCas9-Tet2-sgRNA1-2A-GFP
SK2253	AAACGTTCCATCAGGCTTGCTTCGC			
SK2772	CACCGAAAGTGCCAACAGATATCC	Tet2 exon3	H Wang et al,. Cell. 2013	pSpCas9-Tet2-sgRNA2-2A-GFP
SK2773	AAACGGATATCTGTTGGCACTTTC			

HPC-7 cells were electroplated with two different gRNA-containing plasmids as previously described in section 2.7.4, 1 plasmid per well. 24 h after electroporation, GFP-positive single cell was sorted by FACS on SONY SH800Z sorter at Wellcome Centre for Human Genetics and sorted cells were plated into 96-well plates in 150 μL Stem Pro growth media supplemented with 100 ng/mL SCF and 50% conditioned media per well. Conditioned media were harvested from HPC-7 cells culture (culture for 48 hr) by centrifugation at 1000 g for 10 min, and media were filtered through 0.22μM filter. Cell colonies could be seen under microscope after two or three-week expansion. 50 μL

Stem Pro growth media with 400 ng/mL SCF were added into plate per well during one or two-week expansion. Cell clones were transferred into a 24-well plate and further expanded for genomic DNA extraction, Western blot analysis and cryopreservation.

Cell clones were selected for Tet2 mutations by PCR amplification of the targeted genomic region and Sanger sequencing. Briefly, genomic DNA from each clone was extracted using Gene Jet Genomic DNA Extraction Kit according to the manufacturer's instructions. PCR reaction was setup as follows: 15 ng genomic DNA, 1.5 μL DMSO, 1μL 10 mM dNTP (10297018, Invitrogen), 1 μL 10 μM forward primer, 1 μL 10 μM reverse primer, 0.4 μL Maximo Taq Polymerase (S101, GeneOn) and 44 μL Maximo Tap PCR buffer(home-made). Samples were run in Veriti Thermal Cycler (4375786, Applied Biosystems) using following program: initial denaturation at 94 ^{o}C for 4 min, denaturation at 94 ^{o}C for 15 s, annealing reaction for 15 s, extension at 72 ^{o}C for 1 min/kb, final extension at 72 ^{o}C for 5 min, 35 cycles were used. PCR products were run on 1% agarose gel to check for fragment size, and purified using QIAquick PCR Purification Kit (28104, Qiagen). Mutations were identified by Sanger sequencing in Eurofins Genomics. sgRNA2-targeting locus contain a EcoRV recognition site, PCR products from sgRNA2-electroporated group were digested with EcoRV (R3195S, New England Biolabs). Genomic DNA of identified mutated clones were hydrolyzed and measured for DNA modifications using LC/MS-MS by Paolo Spingardi at Target Discovery Institute, TDI, Oxford. Cell clones with a clear decrease in 5hmC are pelleted and boiled in appropriate volume of 2x SDS loading buffer at 95 ^{o}C for 10 min. SDS samples were analyzed by SDS-PAGE Western blot. Tet2 protein was detected using anti-Tet2, clone 9F7 antibody. Only clones containing frameshift indels at Tet2 locus, showing decreased 5hmC level and having no Tet2 protein were recognized as true Tet2 knock-out clones. Oligonucleotides used for PCR amplification and Sanger sequencing were listed in table (2.10)

Table 2.10: Primers used for Tet2 mutation detection.

Primer pair	Sequence	Target	Template
SK2271	CTGCCAGAGCCAATCAAGTATGG	~500 bp fragment around sgRNA1 targeting site	Genomic DNA of sgRNA1 targeting clones
SK2272	CTTGTTTGGATGGAGCCCAGAG		

Continuation of Table 2.10

Primer pair	Sequence	Target	Template
SK2774	CAGATGCTTAGGCCAATCAAG	~500 bp fragment around sgRNA2 targeting site	Genomic DNA of sgRNA2 targeting clones
SK2775	AGAAGCAACACACATGAAGATG		
Sanger sequencing primers			
Primer pair	Sequence	Sequencing targets	
SK2813	GAGTTACTGGGGAGGTATAAAC	PCR products of SK2271 + SK2272 amplification	
SK2814	CTCTGGCTGCCTTTCATGTG		
SK2774	CAGATGCTTAGGCCAATCAAG	PCR prodcuts of SK2774 + SK2775 amplification	
SK2775	AGAAGCAACACACATGAAGATG		

2.7.6 Retrovirus-based stable cell line construction

A retroviral system was used to introduce the stable expression of Flag-HA Tet2 and APEX2-Tet2 in HPC-7 cells as previously described (Serrano et al., 1997) with some modifications. In this study, target cells were infected four times. Chloroquine (25 μM) was used for higher transfection efficiency. Polybrene (4 μg/mL, TR-1003-G, Sigma-Aldrich) was used for higher infection efficiency.

For the production of retrovirus, Plat-E cells were seeded on a 10 cm dish and transfected by calcium-phosphate precipitation with MSCV expression plasmid and helper plasmid as previously described in section 2.7.3. MSCV expression plasmids used for cell line construction were listed in table (2.4). 12 h post transfection, culture media were replaced by 10 mL fresh DMEM growth media. 12 h post 1stmedia replacement, media were replaced by 5 mL HPC-7 IMDM growth media (supplemented with 5% FBS, 150 μM MTG, 1% Penicillin/Streptomycin, and 2 mM L-glutamine). After 12 h, the virus-containing media were collected, filtered (0.45 μM filter unit) and supplemented with 4 μg/mL polybrene and 100 ng/mL SCF (1stsupernatant). The 2ndand 3rdsupernatant were collected every 5 h as before. 4thsupernatant was collected 12 h after the 3rdcollection.

For infection, one 6-well plate of HPC-7 cells (5 x 10^5 cells in 1 mL media per well) were infected by virus supernatant (~800 μL per well) with centrifugation at 1100 g for 15 min at RT in a benchtop centrifuge, and infected cells were incubated at 37^{o}C until next infection. The infection process was repeated using the 2nd, 3rd, and 4thsupernatant.

4h post last infection, media were replaced by fresh HPC7 IMDM comple growth media (100 ng/mL SCF). 16 h later, infection efficiency was assessed by flow cytometry and GFP-positive cell population were sorted in bulk on SONY SH800Z sorter. Sorted cells were expanded for Western blot analysis, RNA extraction and cryopreservation.

2.7.7 PiggyBac transposon-based stable cell line construction

PiggyBac Transposon and Tetracycline-controlled Tet-On systems (PiggyBac Tet-On) were combined to produce an inducible expression of exogenous Tet2 gene (3xHA-Tet2 or Flag-HA Tet2) in Tet2 knock-out HPC7 cells. Cells were electroporated with TetON-Flag-HA-Tet2 or TetON-3xHA-Tet2 vector together with CAG-rtTA-IRES-Neo vector and PiggyBac transposase plasmid as previously described in section 2.7.4. For each electroporation on a 6-well plate, 8 μg TetON expression vector, 1 μg CAG-rtTA-IRES-Neo vector and 1 μg PiggyBac transposase vector were used. One well of no-DNA electroporated control was set up for each experiment. 72 hr post electroporation, cells were selected with 0.5 μg/mL Puromycin until cells in no-DNA electroporated control were all died. Cells were maintained in antibiotics-supplemented media and expanded for cryopreservation and tested for Tet2 expression. To induce the expression, cells were seeded at a density of 5 x 10^5 cells/mL and grown in the media supplemented with Doxycycline (Dox) at a wide range of concentration (10-1000 ng/mL) for 24 h. Induced cells were harvested for genomic DNA extraction, and Western blot analysis.

2.8 Immunofluorescence staining

293 T cells were seeded at 40% confluency on a fibronectin(341635, Sigma-Aldrich)-coated 2-well cell culture chamber (94.6170.202, Sarstedt) overnight. Next day, cells were transfected with Flag-Tet2-APEX2 or APEX2-Flag-Tet2 vector by jetPRIME reagents as previously described in section 2.7.3. For each transfection, 0.6 μg DNA, 75 μL jetPRIME buffer and 1.8 μL jetPRIME were used. 24 h post transfection, cells were treated with appropriate reagents to initiate the APEX2-mediated *in vivo* biotinylation (detailed method see section 2.14). One well of DNA-transfected but non-biotinylation induction control was set up for experiments. Cells were washed 3x PBS and fixed

in 4% paraformaldehyde/DPBS (16% solutiom, 15710, Electron Microscopy Sciences) at RT for 15 min. Cells were washed three time in PBS and permeabilized in 0.1% Triton X-100/PBS at RT for 7 min. Cells were washed three time in PBS and blocked in 5 % (w/v) BSA/PBS at RT for 30 min. Cells were washed 3x PBS and stained with anti-Tet2 (MABE1132) together with anti-Flag (F1804) antibodies at RT for 1 h in the dark. Cells were washed three time in PBS and incubated with anti-mouse and anti-rat fluorophore conjugated secondary antibodies together with Alexa 488-conjugated streptavidin antibody at RT for 1 h in the dark. Cells were washed three time in PBS and coverslips were mounted with VECTASHIELD HardSet Mounting Medium with DAPI (H-1500, Vector Laboratories). Slides were left at RT in the dark for 1 h and sealed by nail polish and stored at 4^{o}C. Images were obtained by a confocal microscope (Zeiss 710). Information about antibodies and dilution was listed in table(2.2).

2.9 Flow cytometry

Cells were washed once with cold PBS and resuspended in cold FACS buffer (PBS containing 2% FBS and 2 mM EDTA) at a final cell density 5 x 10^6-1x10^7/mL, For each staining, 100 μL cell suspension (0.5-1x 10^6 cells) was blocked with 2.5 μg Mouse BD Fc Block (553141, BD Biosciences) at RT for 5 min. Cells were incubated with 100 μL fluorophore conjugated antibody dilution in FACS buffer at 4^{o}C for 1 h in the dark. Cells were washed once with cold FACS buffer and resuspended in FACS buffer, followed by flow cytometry on BD LSR Fortessa™ X20 at the Ludwig Institute. Data was analyzed using FlowJo software.

For analysis of wildtype and *Tet2*-KO HPC-7 cell features, cells were stained with antibodies against following: CD34, c-Kit, CD44, CD29, CD3e, B220, CD19, TER119, CD11b, F4/80 and Gr-1. For analysis of HPC-7 *in vitro* myeloid differentiation assay, cells were stained with antibodies against following: CD34, CD11b, F4/80. Information about antibodies and dilution was provided in table(2.2).

2.10 *In vitro* myeloid differentiation

To induce myeloid differentiation of HPC-7 cells and its derivatives, cells were seeded on a 6-well plate at a cell density 5 x 10^5 cells/mL in Stem Pro-34 media supplemented with 10% Stem Pro supplements, 150 μM MTG, 2 mM L-glutamine, 1x penicillin-streptomycin and the following murine cytokines: 20 ng/mL G-CSF, 10 ng/mL GM-CSF, 10 ng/mL IL-3 and 10 ng/mL IL-6. All the cytokines were purchased from Peprotech. Cells were grown in differentiation media for 8 days and media were replaced every 2 days. Cells were analyzed by flow cytometry as previously described in section 2.9.

2.11 Colony-forming-unit assay

Colony-forming-unit assay was performed under the instruction from Dr. Lynn Quek at Weatherall Institute of Molecular Medicine (WIMM), Oxford.

HPC-7 cells and its derivatives were grown in methylcellulose-based media with recombinant cytokines including SCF (Recombinant mouse stem cell factor), IL-3 (Recombinant mouse interleukin 3), IL-6 (Recombinant human interleukin 6), and EPO (Recombinant human erythropoietin) (M3434, STEMCELL Technologies) according to the manufacturer's instructions. Briefly, cells were resuspended in IMDM media supplemented with 2% FBS at a cell density 4000 cells/mL. 400 cells were seeded in 1 mL methylcellulose media on a 35 cm dish and cultured at 37^{o}C. Three technical replicates were prepared for each experiment and three biological replicates were prepared for each cell group.

After 14 days, colonies were counted under Zeiss Axiovert 40C inverted microscope (2X objective) and data were analyzed using GraphPad Prism software. Colonies were imaged under the digital inverted microscope (EVOS X1, AMG) at 2X, 10X, 20X and 40X magnification at WIMM. For further analysis of cell differentiation, cells from a single colony were resuspended in 200 μL IMDM media supplemented with 2% FBS and transferred to slides by centrifugation at 1000 rpm for 5 min in the Cytospin™ 4 centrifuge (Thermo Scientific). Slides were stained by modified Wright's Staining (Hematek 2000, Bayer Healthcare) according to the manufacturer's instructions. Coverslips were mounted

to the slides using DPX mounting medium (06522-100ML, Sigma). Slides were imaged by NanoZoomer S210 Digital slide scanner at 20X resolution at Ludwig Institute. Images were analyzed using NDP Viewer2 software.

2.12 Cell apoptosis assay

Cell apoptosis of HPC-7 and its derivatives was measured by APC Annexin V (550474, BD Biosciences)/ 7-AAD staining (559925, BD Biosciences) according to the manufacturer's instructions. Cells were washed twice with cold PBS and resuspended in 1x binding buffer (556454, BD Biosciences) at a concentration of 10^7 cells/mL. For each stain, 10^6 cells were stained with 0.5 μL APC Annexin V and 5 μL 7-AAD at RT for 15 min in the dark. Cells wer added with 400 μL 1x binding buffer and analyzed by flow cytometry on BD LSR Fortessa™ X20 within 1 h. Unstained cells and cells stained with APC Annexin V or 7-AAD alone were used to set up compensation and gating.

2.13 *In vivo* BirA labelling assay

293 T cells were transfected with BirA-Flag-Tet2 or Flag-Tet2-BirA plasmid using jetPRIME reagents as previously described in section 2.7.3. 4 h post transfection, media were replaced by fresh growth media supplemented with 50 μM Biotin (B4501, sigma). After additional 24 h, cells were harvested for nuclear protein extraction (see details in section 2.15.2). The concentration of nuclear extracts was determined by Bradford assay (see details in section 2.15.3). 5x SDS loading buffer was added to the nuclear extracts and samples were heated at 95 ^{o}C for 5 min. Samples were analyzed by SDS-PAGE Western blot. The expression of Tet2-BirA fusion protein was detected using anti-Tet2 antibody (61389, Active Motif) and anti-Flag antibody (F3165, Sigma). Biotinylation was detected using Streptavidin-HRP antibody. Biotinylated protein was enriched by incubating nuclear extracts with streptavidin beads (see details in section 2.15.4) and analyzed by SDS-PAGE Western blot.

2.14 *In vivo* APEX2 labelling assay

APEX2 labelling assay was used to biotinylate Tet2 interacting partners *in vivo* as previously described in (Hung et al., 2016). For assay on 293 T cells, cells were transfected with APEX2-Flag-Tet2 or Flag-Tet2-APEX2 plasmid using jetPRIME reagents as previously described in section 2.7.3. 24 h post transfection, cells were incubated with 500 μM biotin-phenol (LS-3500.0250, Iris Biotech GmbH) in DMEM media at 37°C under 5% CO_2 for 30 min. H_2O_2 was added to the culture for a final concentration 1 mM and incubated at room temperature for 1 min to initiate biotinylation. To quench the reaction, cells were washed 3 times with quencher solution. Cells were fixed in paraformaldehyde for imaging experiments (see details in section 2.8) or harvested for genomic DNA extraction or nuclear protein extraction.

For assay on HPC-7 and its derivatives, cells were incubated with biotin-phenol followed by centrifugation at 300 g for 5 min. Cells were resuspended in media containing 1 mM H_2O_2 at a concentration of 5 x 10^7 cells/mL and incubated at RT for 1 min. Reaction was quenched by adding 100 volume of quencher solution. cells were spun down and washed two more times with quencher solution prior to nuclear protein extraction.

2.15 Protein method

2.15.1 Whole cell lysate extraction

Cells were washed twice with cold PBS and lysed in cold whole cell lysis buffer in a pre-chilled 1.5 mL eppendorf tube and incubated at 4^{o}C for 30 min with rotation. Lysates were cleared by centrifugation at 21,130 g at 4^{o}C for 30 min. Supernatant was collected and supplemented with 5 x SDS loading buffer, and incubated at RT for 5 min before boiling at 95 ^{o}C for 5 min. For extraction on HPC-7 cells, 4 x 10^6 cells was lysed in 100 μL lysis buffer and 20 μL 5 X SDS loading buffer was added, 10 μL sample was loaded for SDS-PAGE Western blot.

2.15.2 Nuclear extraction from mammalian cells

Cells were washed twice with cold PBS and resuspended in sucrose lysis buffer without NP-40 (50 μL per 10^7 cells). Equal volume of sucrose lysis buffer containing NP-40 was added. Cytoplasmic and nuclear compartment were separated by centrifugation at 500 g for 10 min at 4^{o}C. Nuclei were washed once with sucrose lysis buffer without NP-40 and pelleted by centrifugation at 500 g for 5 min. Nuclei were resuspended in 5 volume low salt buffer and 5 volume high salt buffer was added dropwise with gentle agitation. Nuclei suspension was supplemented with Benzonase (71206-3, Millipore) at a concentration of 50 units/10^7 cells and incubated at 4^{o}C for 1 h and cleared by centrifugation at 16,000 xg, for 15 minutes at 4C. Supernatant was taken as nuclear extract and transferred to a pre-chilled fresh tube. Nuclear extract was either used immediately or snap frozen in liquid nitrogen and stored at -80^{o}C.

2.15.3 Protein quantification by Bradford assay

Protein concentration was determined using Bradford reagent (B6916, Sigma-Aldrich) by comparison to a protein standard curve according to the manufacturer's instructions. Cell lysate was diluted 1:5 or 1:10 in lysis buffer before measurement. To prepare a protein standard curve, BSA standard (P0914, Sigma-Aldrich) was diluted in H_2O at a concentration of 0-1 mg/mL. Briefly, 5 μL protein sample was added to separate well on a 96-well flat bottom plate (3650, Costar) and 5 μL sample buffer was added to the well as blank control. Bradford reagent was equilibrated to RT before use. 200 μL reagent was added to each well and the plate was incubated at RT for 10 min. Absorbance at 595 nm was measured using FLUOstar Omega microplate-reader. All measurements were prepared in triplicates. Data was analyzed using Excel.

2.15.4 Streptavidin pulled-down assay

Pierce streptavidin magnetic beads (88817, Thermo Scientific) were washed twice with PBS and twice with nuclear extract buffer (low salt buffer with 350 mM KCl). 600 μg nuclear extract was incubated with 35 μL streptavidin beads with rotation at 4 ^{o}C overnight. Beads were washed in following buffers: two times in wash buffer (20 mM

Tris pH 8, 250 mM KCl, 2 mM $MgCl_2$, 10% glycerol), one times in 1 M KCl, one times in 0.1 M Na_2CO_3, one times in 10 mM Tris-HCl supplemented with 2 M urea, two times in wash buffer. For each wash, beads were resuspended in 1 mL buffer with rotation at 4 ^{o}C for 5 min. Biotinylated proteins were eluted by incubation beads with 40 μL PBS containing 8M urea, 2%SDS and 10 mM biotin (B4501, Sigma-Aldrich) at room temperature for 20 min, followed by heating the beads at 60 ^{o}C for 20 min. Beads were removed using a magnetic stand and eluate was supplemented with 5 x SDS loading buffer and boiled at 95^{o}C for 5 min for Western blot.

2.15.5 SDS PAGE and Western blot

Protein sample in 1 X SDS loading buffer was heated at 95^{o}C for 5 min prior to loading on a 4-15% Criterion TGX Gel (Bio-Rad). 20 μg nuclear extract or sample made from 2-3 x 10^5 HPC-7 cells was loaded into each lane. PageRuler Plus Prestained Protein Ladder (26619, Thermo Scientific) was loaded to indicate the protein size. Gel was run in the Criterion Cell (1656001 Bio-Rad) in 1x SDS running buffer at 120 V for 1.5 h until the bromophenol blue run to the bottom of the gel.

After running, gel was assembled into the transfer sandwich as follows: cathode, foam pad, 2x filter paper, gel, nitrocellulose membrane, 2x filter paper, foam pad, anode. Nitrocellulose membrane (10600002, GE Healthcare), filter papers and foam pads were soaked in 1x transfer buffer before use. Transfer was run in a Criterion Blotter (1704070, Bio-Rad) at 50 V for 1.5 h at 4^{o}C. Complete transfer was examined by staining the membrane in Ponceau S solution (P7170, Sigma-Aldrich) and destained in 1x TBST.

Membrane was blocked in 5% (w/v) milk/TBST (fat-free Milk powder, Marvel) at RT for 1 h with gentle shaking. Antibody was diluted in 5% milk/TBST and stored at -20^{o}C. Information about antibodies and dilution was listed in table(2.2). Membrane was incubated in primary antibody dilution at 4^{o}C overnight with gentle shaking. For streptavidin-blot, membrane was stained with streptavidin-HRP at RT for 2 h instead of overnight incubation. Membrane was washed 3 times in 1x TBST (10 min/per wash) at RT and incubated with secondary antibody at RT for 1 h with gentle shaking. Membrane

was washed 3 times in 1x TBST (10 min/per wash) at RT and developed using ECL Select Western Blotting Detection Reagent (RPN2235, GE Healthcare) according to the manufacturer's instructions. Chemiluminescent signals were captured using ChemiDoc Imaging system (Bio-Rad) or X-ray films (Fujifilm).

For re-probing the membrane with an antibody of the same species as the first primary antibody, membrane was stripped using Restore Western Blot Stripping Buffer (21059, Thermo Scientific) and incubated at RT for 15 min with gentle shaking followed by 3 times 1x TBST wash. Membrane was re-blocked in 5 % milk/TBST at RT for 1 h before incubation with new primary antibody overnight. For re-probing the membrane with an antibody of different species, membrane was stripped in 5% (w/v) sodium azide in 1x TBST at 4^{o}C overnight in the dark with gentle shaking. Membrane was washed 3 times with 1x TBST and incubated with the new primary antibody overnight.

2.15.6 Immunoprecipitation

Immunoprecipitation (IP) was performed using the Nuclear Complex CO-IP kit (54001, Active Motif) according to the manufacturer's instructions. For assay on HPC-7 cells, 5x 10^7 cells were washed in 50 mL PBS containing phosphatase inhibitors and lysed in 500 μL 1x hypotonic buffer and incubated on ice for 10 min. Nuclei was collected by centrifugation at 14000 g for 30 s at 4^{o}C. Nuclei was resuspended in 250 μL digestion buffer supplemented with 1.25 μL PMSF, 2.5 μL protease inhibitor and 1.25 μL enzymatic shearing cocktail. Nuclei suspension was incubated at 4^{o}C for 1.5 h with gently vortex every 30 min incubation, and cleared by centrifugation at 14000 g for 30 min at 4^{o}C. Supernatant was taken as nuclear extracts and transferred to a fresh pre-chilled tube. Protein concentration was determined using Bradford reagents as previously described in section 2.15.3.

Dynabeads (Invitrogen) or Anti-FLAG® M2 Magnetic Beads (Sigma, M8823) were washed 3 times with PBS and blocked in 10 volume 2 mg/mL BSA/PBS at 4^{o}C overnight with rotation. Blocked beads were washed 3 times with PBS and 1 time with 1x IP high buffer. Nuclear lysates were diluted at a final concentration of 1 μg/μL with 1x IP high

buffer supplemented with protease inhibitor. 500 μg nuclear lysates was pre-cleared with 10 μL Dynabeads at 4^{o}C for 1 h with rotation. For anti-Flag IP, protein lysates were pre-cleared with 20 μL Mouse IgG agarose beads (Sigma, A0919). After pre-clearing, beads were discarded and lysates were transferred to a fresh pre-chilled 1.5 eppendorf tube. 25 μL lysates were taken as 5% IP input and mixed with 5x SDS loading buffer, boiled at 95^{o}C for 5 min. The rest cleared nulear lysates were incubated with 1-2 μg antibody at 4^{o}C overnight with rotation. The next day, 15 μL blocked Dynabeads was added in the protein/antibody mixture and incubated at 4^{o}C for 3 h with rotation. For anti-Flag IP, protein lysates were incubated with 15 μL blocked anti-FLAG M2 magnetic beads at 4^{o}C for 2 h with rotation. IP wash buffer was prepared by adjusting the salt concentration in the IP high buffer at final 150 mM and buffer was supplemented with protease inhibitor before use. Beads were collected using a magnetic stand and resuspended in 1 mL IP wash buffer supplemented with 1 mg/mL BSA. Beads suspension were transferred to a fresh tube and rotated at 4^{o}C for 5 min. Beads were washed 2 more times with wash buffer supplemented with BSA and 3 times with wash buffer. Beads were collected and boiled in 30 μL 2x SDS loading buffer at 95 ^{o}C for 15 min. SDS sample was removed from the beads, followed by Western blot analysis.

2.15.7 Immunoprecipitation coupled to mass spectrometry (IP–MS)

IP sample preparation:

3 x 10^8 HPC7 *Tet2*-KO 2-7/ Tet-On 3x HA-Tet2 cells were grown in 600 mL media supplemented with 250 ng/mL dox for 24 h. HPC7 *Tet2*-KO 2-7 cells were grown in parallel as control. Four biological replicates were prepared for each cell type. Nuclear extraction and IP were performed as previously described in section 2.15.6. Briefly, 5 mg nuclear lysates, 120 μL Dynabeads Protein A (Clearing: 30 μL; IP: 90 μL), and 25 μL anti-HA antibody (Cell Signalling, 3724) were used for each IP. 50 μg nuclear lysates were taken as 1% input control. After washing, beads were boiled in 250 μL 2x SDS loading buffer at 95 ^{o}C for 15 min. Beads were discarded and 200 μL SDS samples were submitted for LC-MS/MS analysis at Mass Spec facility, TDI. 10 μL sample was loaded for SDS-PAGE Western blot and 5 μL was used for a silver stain. Silver stain

was performed using Pierce Silver Stain for Mass Spectrometry reagents (24600, Thermo Scientific) according to the manufacturer's instructions.

Mass-spec experiments including Mass-spec sample preparation and running, and data analysis were all performed by Georgina Berridge at TDI.

Mass-spec sample preparation:

100 μL IP eluate in 2x SDS sample was alkylated by incubating with 2 μL of 50mM iodoactamide for 30 min at RT. Samples were acidified by addition of 1.2 μL of 12% phosphoric acid, and mixed with 165 μL of S-Trap buffer (90% MeOH, 100mM TEAB pH 7.55). Samples was loaded to the S-Trap micro spin column (Profiti) in a 1.7 mL centrifuge tube and centrifuged for 4000 xg for 30 s. The trapped protein was washed 3 times with 150 μL S-Trap buffer. The S-Trap column was transferred to a clean tube. 20 μL of 50 mM TEAB,pH 8 containing 1:10 wt:wt trypsin was added onto the top of the column. Slight positive pressure was applied to the top of the column to ensure no air bubbles between the column matrix and digestion solution. The columns were incubated for 90 min at 47^{o}C. Digested peptides were eluted with 40 μL each of 50 mM TEAB and then 0.2% Formic acid and finally 50% Acetonitrile containing 0.2% Formic acid. Elution was achieved by centrifugation at 4000 g. Pooled peptides were dried in a rotary evaporator. Dried peptides were resuspended into 100 μL of 2%ACN containing 1% Formic acid, then desalted by applying to C18 Sola cartridges (60109-001, Thermo Scientific), washed with 1 mL 2% acetonitrile in water (containing 0.1% formic acid) and eluted with 2 x 60 μL 65% acetonitrile: water (65:35). Eluted peptides digest was concentrated and acetonitrile was removed by evaporation using a rotary evaporator (30 ^{o}C for 4 to 6 hours until dry). Dried peptides were resuspended into 20 μL 2% acetonitrile: 98% water containing 0.1% formic acid.

LC-MS/MS setting:

1 μL peptides were subjected to C18 column chromatography and analysed by data-dependant MS/MS on a ThermoFisher Q-Exactive mass spectrometer. Samples were

desalted online (PepMAP C18, 300 μm x 5 mm, 5 μm particle, Thermo) for 1 minute at a flow rate of 20 μL/min, and separated on a nEASY column (PepMAP C18, 75 μm x 500 mm, 2 μm particle, Thermo) over 60 minutes using a gradient of 2-35% acetonitrile in 5% DMSO/0.1% formic acid at 250 nl/min. Survey scans were acquired at a resolution of 70,000 at 200 m/z and the 15 most abundant precursors were selected for HCD fragmentation.

Data analysis:

Acquired data was loaded into Progenesis QI(Nonlinear Dynamics) and automatically aligned for retention time and peak picking. The MS/MS spectra were exported into an .mgf file and searched using the denovo search algorithm PEAKS (V8.5, Bioinformatics Solutions). The results were exported at a false discovery rate of 1%, mass deviation of 10 ppm for MS1 and 0.05 Da for MS2 spectra. Database searches (Swissprot; mouse) were performed using MS+/- 10ppm, MSMS +/-0.05Da, 2 missed cleavages (M) deamidation (N) as variable modifications, iodacetamide (C) as a fixed modification. The data was evaluated in Perseus 1.5.2.4 (Tyanova et al., 2016) using the PCA plots and filtering for only present proteins. The raw intensities were exported and analysed in Perseus. Essentially, the data was log2 transformed, normalized by subtracting the median from the columns (each sample) and then missing values were imputed. Samples were analysed for fold change difference using two-samples test comparing KO and Tet2 samples.

2.16 RNA-sequencing

RNA-seq on wildtype and *Tet2*-KO HPC-7 cells were prepared with exogenous calibration. Two *Tet2-KO* clones (clone 2-1 and clone 2-7) and WT HPC-7 were sequenced. Three biological replicates were prepared for each cell group.

2.16.1 Sample preparation

Chicken DF-1 cells were spiked-in HPC-7 cells on a per-cell basis before RNA extraction using AllPrep® DNA/RNA Kits (80204, QIAGEN). Briefly, DF-1 cells were grown on a 10 cm dish in DMEM media supplememnted with 10% FBS, 1% Penicillin/Streptomycin,

2 mM L-glutamine and 0.15% sodium bicarbonate at 37^{o}C, 10% CO_2. Cells were harvested and lysed in buffer RLT Plus at a concentration of 1 x 10^7 cells/mL. DF-1 cell lysates were aliquoted and stored at -80^{o}C. 5 x 10^6 HPC-7 cells were lysed in buffer RLT Plus and mixed with 50 μL spike-in cell lysates (5 x 10^5), homogenized by QIAshredder. Homogenized spiked-in cell lysates were loaded on a AllPrep DNA spin column and centrifuged at 8000 g for 30 s. Flow-though was used for RNA purification, followed by on-column DNaseI digestion. RNA was eluted in 50 μL RNase-free water. Genomic DNA was purified from the AllPrep DNA spin column according to the manufacturer's instructions, followed by a second-round purification by Gene Jet Genomic DNA extraction. DNA was eluted in 100 μL elution buffer. RNA and DNA concentration were measured by Qubit 3.0 fluorometer (Thermo Scientific). RNA quality was examined by Bioanalyzer 2100 (Agilent) using the RNA 6000 Nano Kit (Agilent). 1 μg RNA and 1.5 μg DNA were submitted to Oxford Genomics Center for library preparation and sequencing. Ribo zero-derived RNA library and DNA library were multiplexed and sequenced on a HiSeq 4000 instrument to generate 75bp paired-end reads.

2.16.2 Data analysis

Quality control and alignment:

Sequencing quality was examined by FastQC (v0.11.7) (Andrews). Low quality RNA-seq reads were removed using Trim Galore (v0.5.0) (Felix) with parameters: –paired –quality 35 –fastqc. Before alignment, mm10 and galGal5 genome sequences were combined by Linux cat command. GTF files of two genomes were also combined. Genome index was generated with the combined genome and GTF files using STAR (v2.6.1d) (Dobin et al., 2013) with the option: –runMode genomeGenerate. RNA-seq reads were aligned to the combined mm10 and galGal5 genome using STAR with option: –outSAMtype BAM SortedByCoordinate. mm10 unique mapped reads and galGal5 unique mapped reads were extracted using Samtools (v1.9) view command with parameters: -q 255, and separated by linux grep command. DNA-seq reads were aligned to the combined mm10 and galGal5

genome using Bowtie2 (v2.2.5) (Langmead and Salzberg, 2012) with parameters: –no-mixed –no-discordant. Multiple mapped reads were filtered out using grep command with parameters: -v XS: - and unmapped reads were excluded using Samtools (v1.3) (Li et al., 2009) view command with parameters: -F 4. mm10 unique mapped reads and galGal5 unique mapped reads were separated using grep command.

Differential gene expression analysis:

Differential gene expression with spike-in calibration were performed under the instructions from Dr. Martin Cusack, a former member of the group.

mm10 and galGal5 unique mapped RNA-seq reads were assigned to genomic features (gene exons) and counted using featureCounts (Subread v1.6.2) (Liao et al., 2014) with specification of following parameters: -p -t exon -g gene_id -s 0. Differential gene expression between WT and *Tet2*-KO HPC-7 cells was determined using R package DESeq2 with raw read counts as input (Love et al., 2014). Two types of count normalization size factors with spike-in calibration were calculated to scale the expression of mm10 genes relative to those of galGal5 genes. In the calibrated and adjusted method, before size factor calculation, galGal5 RNA-seq reads counts in each sample were pre-normalized using the galGal5/mm10 read ratio of the corresponding genomic-DNA seq samples to minimize the variation in spike-in cell mixing. In the calibrated method, galGal5 RNA-seq read counts were not adjusted by the galGal5/mm10 DNA-seq count ratio before quantification. Count normalization size factor without spike-in calibration was calculated using the default median of ratios method to adjust for sequencing depth and RNA composition. Genes were considered as significantly differentially expressed if they exhibited a p-adj < 0.05. Log2 fold change values were visualized in MA plots, heatmaps and barplot using R and ggplot2 (Wickham, 2009). R package, pheatmap (Kolde, 2012), was used to plot heatmaps of log2 fold gene expression changes.

Functional analysis on differentially expressed genes:

David v6.8 (Huang et al., 2009) was used to perform gene ontology (GO) analysis on differentially expressed genes. Mouse hemopoiesis gene set (GO:0030097), cell

proliferation gene set (GO:0006915) and cell apoptosis gene set (GO:0006915) were obtained from MGI (Mouse Genome Informatics) (Smith and Eppig, 2009). Gene set enrichment analysis was performed by R package fgsea (Huang et al., 2009) using transcription factors-targeted gene sets as permutation type, 1,000 permutations and log2 ratio of fold change value as metric for gene ranking. Transcription factors-targeted gene sets, ChEA-2016 (Lachmann et al., 2010) were obtained from Enrich website (Kuleshov et al., 2016).

2.17 ChIP-sequencing

2.17.1 Chromatin crosslinking and sonication

5 x 10^7 HPC-7 PiggyBac 3xHA Tet2 expressing cells were cultured in 100 mL SCF growth medium containing 100 ng/mL dox for 24 hr. Next day, $\sim$ 1 x 10^8 cells were pelleted, washed with PBS. For 2-step crosslinking protocol, cells were resuspended in 20 mL PBS supplemented with disuccinimidyl glutarate (DSG, 20593, Thermo Scientific) to a final concentration of 2 mM and cells were crosslinking at room temperature for 1 hr with gentle rotation. 1.25 mL 16% methanol-free formaldehyde (28908, Thermo Scientific) was added to a final concentration of 1% and cells were crosslinking for 10min at room temperature for 10 min with gentle rotation.

The crosslinking reaction was quenched by adding 2 mL 2M glycine, followed by incubation at room temperature for 3 min. Crosslinked cells were pelleted by centrifugation at 211 x g for 4min at 4^{o}C. The following steps were all carried out at 4^{o}C. Chromatin was prepared by firstly rotating the cells in 20 mL buffer LB1 (50mM HEPES KOH pH 7.9, 140mM NaCl, 1mM EDTA, 10% Glycerol, 0.25% NP40, 0.25% Triton x-100, freshly complemented with 1x cOmpleteTM EDTA-free Protease Inhibitor Cocktail (05056489001, Roche)) for 10 min, pelleted and resuspended in 20 mL buffer LB2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, freshly complemented with 1x cOmpleteTM EDTA-free Protease Inhibitor Cocktail) for 5 min. Nuclei was pelleted, resuspended in 2 mL buffer LB3 (10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium deoxycholate, 0.5% N- lauroylsarcosine, freshly

complemented with 1x cOmpleteTM EDTA-free Protease Inhibitor Cocktail). and split into two vials of 1 mL milliTUBE with AFA Fiber (520130, Covaris) for sonication. Chromatin was sheared using a Covaris S220 AFA ultrasonicator (settinng: 140W Peak Incident Power, 8% Duty Factor, 200 Cycles/Burst) for 18min. After sonication, the soluble fraction was collected by centrifugation at 15,000 x g for 20min. The 2 aliquots of the sheared chromatin were merged.

To prepare chromatin from mouse ESCs, mESCs were grown to 70-90% confluency in 15 cm cell culture dishes. Cells were washed with PBS, detached by trypsinisation, pelleted, resuspended in certain volume of PBS and counted to get $\sim$ 1 x 10^8 for crosslinking.

2.17.2 Chromatin immunoprecipitation

The following steps were all carried out at 4^{o}C. For immunoprecipitation (IP), the sheared chromatin was diluted 10 fold in Chromatin Dilution Buffer (20mM Tris, pH 8.0, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, freshly complemented with 1x cOmpleteTM EDTA-free Protease Inhibitor Cocktail) in a 15 mL falcon tube. Chromatin from $\sim$ 1 x 10^8 *Tet2*-KO cells or 5 x 10^7 3xHA Tet2 expressing cells was used to prepare one ChIP-seq replicate. Prior to IP, 10 mL diluted chromatin (5 x 10^7 cells) was pre-cleared with 30 μL blocked Dynabeads Protein A slurry (10002D, Invitrogen) for 1 hr at 4^{o}C with rotation. Before use, Dynabeads Protein A were washed with 0.5% BSA/PBS twice, blocked by incubation in 10 volume 0.5% BSA/PBS containing 1 mg/mL yeast tRNA (10109517001, Roche) for at least 1 hr at 4^{o}C with rotation. Blocked beads were washed with 0.5% BSA/PBS and resuspended in Chromatin Dilution Buffer to achieve the original slurry volume prior to use.

Pre-cleared chromatin was removed from the beads and 100 μL (1%) pre-cleared chromatin was taken as an input control. The rest pre-cleared chromatin was incubated with 30 μL anti-HA antibody (3724, Cell Signalling) at 4^{o}C with rotation overnight. Immune complex was captured by incubation with 50 μL blocked Dynabeads Protein A at 4^{o}C for 3 hr with rotation. Beads were collected and washed with following buffers: once with Low Salt Buffer (20mM Tris pH 8.0, 150mM NaCl, 0.1% SDS, 1% Triton

X-100, 2mM EDTA), once with High Salt Buffer (50mM HEPES pH7.5, 500mM NaCl, 0.1% deoxycholic acid, 1% Triton X-100 and 1mM EDTA), once with Lithium Chloride Buffer (10mM Tris pH 8.0, 250mM LiCl, 1% deoxycholic acid, 1% NP-40 and 1mM EDTA) and twice with TE Buffer (10mM Tris-HCl pH 8.0, 1mM EDTA). For each wash, the beads were resuspended in 1 mL wash buffer in the 1.5 mL Eppendorf tube, rotated for 5 min and collected by placing samples onto the DynaMag™-2 Magnet (12321D, Invitrogen). Samples were transferred into fresh tube after adding the 1st wash buffer and prior to elution step. Captured crosslinked material was eluted from the beads in 200 μL Elution Buffer (1% SDS and 0.1M $NaHCO_3$) by vortexing and shaking vigorously for 30min at room temperature. The eluate was transferred to a new 1.5 mL Eppendorf tube. Input samples were added up to 200 μL with Elution Buffer. 13 μL 5M NaCl was added to both IP and input samples and reverse-crosslinking was done at 65 ^{o}C overnight. Subsequently, samples were treated with 1 μL 10 mg/mL RNase A (EN0531, Thermo Scientific) at 37^o for 1 hr, followed by incubation with 1 μL of 20 mg/mL Proteinase K (AM2546, Invitrogen) at 55^o for 1 hr. DNA was purified using the ChIP DNA Clean Concentrator kit (D5205, Zymo Research) following the manufacturer's instructions. DNA was eluted in 30 μL elution buffer.

For each IP reaction for Tet2 ChIP-qPCR application, soluble chromatin from 5 x 10^6 mESCs was pre-cleared using 15 μL blocked Dynabeads Protein A or Protein G slurry and incubated with 5 μL Tet2 antibody. The immune complex was captured by 15 μL blocked Dynabeads Protein A or Protein G slurry. ChIP DNA was eluted in 120 μL elution buffer and purified using the QIAquick PCR Purification kit (28104, Qiagen) following the manufacturer's instructions. Prior to column binding step, sample pH was adjusted by addition of 10 μL 3M NaAc pH5.2. DNA was eluted in 100 μL nuclease-free water.

2.17.3 Real-time PCR analysis of ChIP DNA

Real-time PCR reaction was setup as follows: 0.15 μL forward primer, 0.15 μL reverse primer, 7.5 μL 2x QuantiFast SYBR Green PCR Master Mix (204056, Qiagen), and 4 μL DNA in 0.1 mL Capped Strip Tubes (981106, Qiagen). Samples were run in Qiagen

Rotor-Gene Q platform. Samples were heated at 95^o for 5min, followed by 40 cycles at 95^o for 10s and 60^o for 30s. A melt curve analysis was performed by ramping the temperature from 50^o to 95^o (5s per °C). All measurements were done using two technical duplicate for each biological replicate. Crossing threshold (Ct) values were averaged between technical duplicates. Table 2.11 summarizes the primers used for ChIP-qPCR. The percentage enrichment over input (% input) was calculated to represent the difference in Ct values between ChIP DNA and input sample.

Table 2.11: Primers used for ChIP-qPCR.

Locus Name	Use	Primer pair	Sequence	mm10 coordinates
Asah3l	putative mESCs Tet2 BS	SK3043	GTATCCGGAGGCTGTCCCG	chr4:86,520,282-86,520,377
		SK3044	TCGGAGCAGCTTCTCTGCA	
MLL1	putative mESCs Tet2 BS	SK3106	CCTGCCCTGCAATCTGCATA	chr9:44,689,521-44,689,595
		SK3107	GGGCTCGGCATAGAGCAG	
Zmym6	putative mESCs Tet2 BS	SK3110	AGGCGCCGCAGGAAGTTCTG	chr4:126,754,563-126,754,652
		SK3111	CCGCCGGTCACGTGATGAG	
Ostm1	putative mESCs Tet2 BS	SK3116	CTCGTGACTCCACTCAGGGC	chr10:42,398,696-42,398,787
		SK3117	CCGGGCCGGACTGTAACAGA	
Ptafr	putative HPC-7 Tet2 BS	SK4166	CTTGCTCTGCAGGAAACCTC	chr4:132,564,044-132,564,117
		SK4167	CCAGATGCCACAGGAACCTC	
Prtn3	putative HPC-7 Tet2 BS	SK4168	GCAGCCGTTGTGTTGTGTAA	chr10:79,877,934-79,878,023
		SK4169	TTGAAATGAAGCCTTGCCGC	
Kit	putative HPC-7 Tet2 BS	SK4170	AGGCAACTCTACTACAACCGC	chr5:75,382,335-75,382,419
		SK4171	CTCTTGAGTGCCCCAGGATT	
Plec	putative HPC-7 Tet2 BS	SK4172	CAGCCCACCAGAAACAGGAA	chr15:76,195,842-76,195,921
		SK4173	TCCCCCATCGCTATACCCAG	
Slc2a3	putative HPC-7 Tet2 BS	SK4176	TCTGATCTTCACCGCTCCAC	chr6:122,742,629-122,742,722
		SK4177	ACCACGAGGAGGATGTGGTA	
NC-1	no Tet2 binding in ESCs	SK3055	CCTGGTCGTGTAGGTGATCC	chr7:28,518,819-28,518,900
		SK3056	CACCCACATGGTGGTTCACA	
NC-2	no Tet2 binding in ESCs and HPC-7	SK2457	AAGTGAGTATTGTCAACTGGCT	chr2:107497162-107497233
		SK2458	TACATGGTGTATCCCAGAAGGG	

2.17.4 ChIP-seq sample preparation

For sequencing, HA ChIP DNA from 3xHA Tet2 cells and HA ChIP DNA from *Tet2*-KO cells were prepared in triplicates. $\sim$ 2 ng ChIP DNA or input DNA which was quantified by Qubit™ dsDNA HS Assay Kit (Q32851, Invitrogen) was used for ChIP-seq libraries for illumina sequencing. Library was prepared using the NEB Ultra™ II DNA Library Prep Kit (E7645L, NEB) and NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Set 1) following manufacturer's instructions. Amplified library was double-sided size-selected using AMPure XP beads (Beckman Coulter) at the ratio 0.5x-0.8x to remove adaptor sequence. Library was quantified using NEBNext® Library Quant Kit for Illumina (E7630S, New England Biolabs). Equal molar amount of samples with compatible indexes were pooled and sequenced on Illumina NovaSeq 6000 (150 bp, pair-end).

2.17.5 Data analysis

In addition to our generated HPC-7 Tet2 ChIP-seq data, HPC-7 histone modification ChIP-seq and transcription factor ChIP-seq were obtained from publicly available data. H3K4me3 ChIP-seq was downloaded from Xu et al., 2018. H3K27ac and H3K4me1 ChIP-seq were downloaded from Org et al., 2015. H3K27me3 and H3K36me3 ChIP-seq were downloaded from Wilson et al., 2016. HPC-7 transcription factors ChIP-seq were downloaded from CODEX database (Sánchez-Castillo et al., 2015).

Quality control and alignment:

Sequencing quality was examined by FastQC (v0.11.7) (Andrews). Low quality ChIP-seq reads and adaptor sequence were removed using Trim Galore (v0.5.0) (Felix) with parameters: –stringency 3 –quality 30 –paired –illumina. ChIP-seq reads were aligned to mm10 genome using Bowtie2 (v2.3.5) (Langmead and Salzberg, 2012) using default parameters. Samtools (v1.3) (Li et al., 2009) was used to convert the alignment file from .SAM to .BAM format, sort the .BAM file according to genomic coordinates, remove duplicates, and index .BAM files. BEDTools (v2.2) (Quinlan and Hall, 2010) was used to filter aligned reads in blacklist region. mm10 blacklist region was obtained from

Amemiya et al., 2019. Unique mapped reads were extracted using grep command by specifying 'XS tag:- '.

BigWig files were generated to visualize ChIP-seq in UCSC Genome Browser. In order to generate bigWig files, aligned reads obtained from biological replicate were merged using Samtools 'merge' command, converted into bigWig files using the deeptools (v2.4.2) (Ramírez et al., 2016) function 'bamCoverage' with following options: -e. -bs 2, – normalizeUsingRPKM.

Peak calling:

Regions occupied by Tet2 or enriched in histone modifications were identified by MACS2 (v2.0.10) (Zhang et al., 2008) function 'callpeak' with q-value cutoff (-q) 0.05. Peak calling was performed independently using unique mapped reads from each biological replicate and unique mapped reads from input DNA was used as control to call peaks. High-confidence Tet2 binding sites were defined by differential binding analysis between 3xHA Tet2 cells and *Tet2*-KO cells using R package Diffbind (v3.0.7) (Stark and Brown). Statistically significant differentially bound regions (FDR < 0.5 and log2 fold change <0) identified in 3xHA Tet2 cells were referred as high-confidence Tet2 peaks (=21883) in HPC-7 cells.

Peak Annotation

ChIP-seq peaks were annotated to the nearest gene TSS and assigned to genomic features (Promoter, 5' UTR, 3' UTR, Exon, Intron and Downstream Intergenic) using R package ChIPseeker (v1.26.0) 'annotatePeak' (yu). Gene annotation files were downloaded from TxDb.Mmusculus.UCSC.mm10.knownGene (v3.10.0). Average Profile of ChIP peaks binding to TSS region (+/- 3Kb to TSS) was plotted using 'plotAvgProf'. Genomic annotation was visualized by 'plotAnnoPie' or 'plotAnnoBar'. Functional profiles comparison of ChIP-seq peaks were using R package clusterProfiler (v3.18.0) (Yu et al., 2012a).

Transcription factor binding motif analysis

ChIP-seq peaks-associated DNA sequence was obtained using BEDtools (v2.2) 'getfasta' (Quinlan and Hall, 2010) and analyzed for motif discovery, motif enrichment analysis using MEME-ChIP (v 5.3.0) (Machanick and Bailey, 2011). Input motif was from HOCOMOCO Mouse (v11 FULL).

Downstream analyses and visualisation

ChIP-seq aligned reads and peaks were visualized in average signal plot, heatmap using EaSeq (v1.05) (Lerdrup et al., 2016). Random matched control regions for Tet2 peak sets (with respect to gene bodies) were generated using the 'Matched controls' tool in EASeq with default settings. Tet2 peaks were uploaded to Cistrome Data Browser (Zheng et al., 2019) using the 'ToolKit' function to search for factors which have a significant binding overlap with Tet2 peak set.

2.18 TAPS-β:

DNA methylation changes of HPC-7 and its derivatives were examined using TAPS-β (TET-assisted pyridine borane sequencing-β) by Dr. Sophie Kirshner in the group as previously described in Liu et al., 2019b. Genomic DNA from each of following cell group were prepared for experiments: WT HPC-7, *Tet2*-KO clone 2-1, *Tet2*-KO clone 2-7 and HPC-7 KO 2-7/PiggyBac 3xHA Tet2 cells (induced by 100 ng/mL dox for 24 hr). Data analysis was performed by Olena Yavorska.

2.18.1 Sample prepration

Genomic DNA sonication and 5hmC blocking:

Genomic DNA was prepared as previously described in section2.3. 100 ng genomic DNA was mixed with 0.5 ng M.SssI methylated λ DNA in 130 μL elution buffer (19086, Qiagen) and fragmented to around 300 bp using Covaris S220 ultrasonicator with following program: Peak Incident Power-175 watt, Duty Factor-10%, Cycles/Burst-200, Duration-180 s. DNA was size-selected to 200-400 bp using double-sided 0.55X – 1X Agencourt AMPure XP beads (A63881, Beckman Coulter). DNA was eluted in 21 μL elution buffer and 1 μL DNA was run on Bioanalyzer 2100 to check for size distribution. DNA

was additionally spiked-in with 250 bp v1 spike-in oligo (home-made) before 5hmC blocking. For a 50 μL 5hmC blocking reaction, 2.5 μL 1 M HEPES, pH 8.0 (H3375, Sigma), 1.25 μL 1 M $MgCl_2$, 5 μL 50x UDP-Glc (2mM, New England Biolabs), 1 μL T4-βGT (M0357S, New England Biolabs), 0.4 μL 250 bp v1 spike-in (1 ng/μL), 20 μL gDNA (100 ng) and 19.85 μL nuclease-free H_2O (10977035, Invitrogen) were mixed and incubated at 37^{o}C for 1 h. DNA was purified twice using Micro Bio-Spin P6 column (In tris buffer, 7326221, Bio-Rad), followed by purification with 1.8X AMPure XP beads, and eluted in 25 μL elution buffer.

End-repair/dA-tailing and adapter ligations:

Sequencing library was prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645, New England Biolabs) according to the manufacturer's instructions. End-repair/ dA-tailing reaction was assembled as follows: 1.5 μL NEBNext Ultra II End Prep Enzyme Mix, 3.5 μL NEBNext Ultra II End Prep Reaction Buffer, 25 µl 5hmC blocked fragmented DNA. Reaction was placed in the Veriti Thermal Cycler with the heated lid set to 75^{o}C, and run the following program: 20^{o}C for 30 min, 65 ^{o}C for 30 min, and hold at 4 ^{o}C. 1.75 µl home-made Adaptor (15 μM), 15 μL NEBNext Ultra II Ligation Master Mix, 0.5 μL NEBNext Ligation Enhancer were added into the End Prep Reaction Mixture and incubated at 20 ^{o}C for 30 min in the thermo cycler with the heated lid off. adapter-ligated DNA was purified with 0.9X AMPure XP beads and eluted in 20 μL nuclease-free water.

mTet1 CD oxidation:

Oxidation reaction was set up in a 50 μL volume on ice as follows: 20 μL adapter-ligated DNA, 15 μL mTet1CD oxidation buffer, 3.33 μL 1.5 mM $Fe(NH_4)_2(SO_4)_2$, 9 μL 20 μM moue Tet1 CD (home-made) and nuclease-free H_2O. The reaction was incubated at 37^{o}C for 80 min and stopped by adding 1 μL Proteinase K (P8107S, New England Biolabs) and incubated at 50^{o}C for 1 h. Oxidized DNA was purified using Micro-Spin P30 columns (in Tris buffer, 7326250, Bio-Rad), followed by purification with 1.8 x AMPure XP beads, and eluted in 35 μL nuclease-free H_2O.

Pyridine borane reduction:

Oxidized DNA in 35 μL H_2O was reduced by adding 10 μL of 3 M sodium acetate solution, pH = 4.3 and 5 μL of 10 M pyridine borane (L13178.09, Alfa Aesar). Reaction was incubated at 37^{o}C for 16 h in a ThermoMixer C (Eppendorf) at 850 rpm. Converted DNA was purified using Zymo-Spin IC columns (C1004-50, Zymo) and eluted in 20 μL H_2O.

Validation of 5mC conversion and 5hmC protection using TaqαI assay:

TaqαI digestion assay was used to confirm the conversion of 5mC to T in the TAPS-β library. A 194 bp DNA containing a single TaqαI restriction site (TCGA) was PCR amplified from the methylated λ DNA. A 25 μL PCR reaction was assembled as follows: 2 ng converted DNA, 1x Taq buffer, 0.2 mM dNTPs, 0.2 μM forward primer (SK4182: GCTGGGGAACTACAGGCT), 0.2 μM reverse primer (SK4183: AGAACCAGAACTCAAACTGTAC), 0.125 μL Taq DNA polymerase (M0273, New England Biolabs), nuclease-free H_2O. PCR was run in a thermo cycler using following program: initial denaturation at 95 ^{o}C for 30 s, denaturation at 95 ^{o}C for 30 s, annealing at 48 ^{o}C for 30 s, extension at 68 ^{o}C for 1 min, final extension at 68 ^{o}C for 5 min, hold at 4 ^{o}C, 35 cycles were used. PCR products were run on a 2% agarose gel to check the amplicon size and purified using the NucleoSpin PCR clean-up kit. PCR products were digested in a 10 μL reaction with 1x CutSmart buffer and 0.2 μL TaqαI (R0149S, New England Biolabs). Reaction was incubated at 65 ^{o}C for 30 min. Non-digested and digested DNA were run together in a 2% agarose gel to check for the digestion. 250 bp v1 spike-in DNA contains a single TaqαI restriction site with symmetrical 5hmC site. 5hmC was protected by βGT-mediated blocking from C to T conversion and 5hmC was kept intact as C in the TAPS-β library. Spike-in DNA was amplified from the converted TAPS-β DNA as previously described with following primer pairs: annealing at 44 ^{o}C. PCR products were digested with TaqαI enzyme and run on the agarose gel to check for the digestion.

Library amplification and sample pooling:

The final sequencing library was amplified using KAPA HiFi HotStart Uracil+ ReadyMix PCR Kit (KK2801, Roche) according to the manufacturer's instructions. Briefly, converted/ adapter-ligated library DNA was mixed with 25 μL 2x KAPA HiFi HotStart Uracil+ReadyMix, NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Set 1) in a final 50 μL reaction and library was amplified in a thermo cycler for 5 cycles and purified on 1X Ampure XP beads. Pooled library was sequenced in a Novaseq 6000 to generate 150 bp paired-end reads.

2.18.2 Data analysis

Quality control and alignment

Sequencing data quality was examined by FastQC (v0.11.9). Low quality reads were trimmed using Trim Galore (v0.6.4) with parameters: -q 20 –length 20 –paired. Reads were mapped to a genome including spike-in sequences (lambda DNA, and 250 bp oligo and the mm10 mouse genome using asTair (v3.2.6) with default parameters. PCR duplicates were removed with Picard MarkDuplicates (v2.22.8) from the lambda DNA and mm10 reads.

Spike-in analysis

DNA modifications for each spike-in DNA were called by asTair call (v3.2.6) with following parameters: –context all –minimum_base_quality 13 –method mCtoT –N_threads 1 –max_depth 4000000.

Genome analysis

DNA modifications were called on each chromosome separately by asTair call (v3.2.6) with following parameters: astair call –context CpG –minimum_base_quality 13 –method mCtoT –N_threads 5 -chr $CHR_NAME. DNA modifications from each chromosome was combined to get one modification file per sample. Data was filtered by awk (v4.0.2) to contain positions with 5' reads. DNA modification changes on following comparison groups: KO1 and WT, KO7 and WT, KO7 and HAT, were analyzed by bedtools intersect (v2.25.0)

with these parameters: intersectBed -wb -sorted -a $INPUTFILENAME1 -b $INPUT-FILENAME2 | awk -v FS='\t|_' 'printf "%s\t%s\t%s\t%s\t%s\t%s \t%s\t%s\n", $1, $2, $3, $4, $5, $9, $10 ' -> $RESULT_DIR\$INPUTFILENAME1_short_overlap_$INPUTFILENAME2_short.txt. Modification level was compared using a custom R script using a prop.test() to obtain a p-value of the significance of modification changes. DNA modification level for each sample group was visualized in bigWig format at ucsc genome browser.

Genomic annotation

CpG islands, shelf, and shore annotation were directly extracted from the annotatr package (v1.16.0) (Cavalcante and Sartor, 2017). Promoters were defined by the transcription start sites of the mm10 genes and HPC-7 H3K4me3 ChIP-seq data. Four types of promoter regions were defined as follows: 1.H3K4me3 peaks; 2. H3K4me3 proximal regions: 1 kb either side of the H3K4me3 peaks; 3. Gene-associated H3K4me3 peaks: H3K4me3 peaks overlap with a known start-codon within 1 kb; 4. Promoter proximal regions, i.e. 4 kb upstream of gene-associated H3K4me3 peaks. Enhancers were defined as regions positive for H3K4me1 and have no H3K4me3 signal. DNA modification level on the annotated genomic regions were calculated using bedtools map (v2.25.0) with following parameters: bedtools map -a $REF_FILE -b $INPUTFILENAME -c 4 -null 'NA' -o mean,count > $RESULT_DIR\$INPUTFILENAME_short.

3

The effect of Tet2 deficiency in HPC-7 cells

Contents

3.1 Introduction

The importance of Tet2 function in normal and malignant hematopoiesis has been supported by the high frequency of TET2 mutations identified in patients with hematopietic

malignancies and abnormal myelopoiesis caused by *Tet2* depletion in mouse. Current understanding of the molecular mechanism of Tet2 function in hematopoiesis has been hampered by the limited availability of the primary hematopoietic stem cells/progenitor cells (HSC/HSPC). To overcome the shortage of the studying material and to have a comprehensive understanding of Tet2 function in a physiological hematopoietic context, I used an established mouse multipotent HSCs-like cell line, hematopoietic progenitor cell (HPC)-7 cells, as a model to study hematopoiesis. In this chapter, I firstly examined the characteristics of HPC-7 cells as a multipotent HSPC cell line, then I generated a *Tet2*-KO HPC-7 cell line by CRISPR-Cas9 technique and explored the cell growth and cell differentiation phenotype of the KO cells.

3.2 Results

3.2.1 HPC-7 shows similar surface marker expression as the hematopoietic progenitor cells

HPC-7 cells are immortalized murine HPC (hematopoietic precursor cell) cells generated by differentiating the LH2 (LIM-homeobox gene)-transduced mouse ES (embryonic stem) cell (Pinto do O, 1998). HPC-7 require stem cell factor (SCF) for growth. HPC-7 have been used as a cell line model for multipotent myeloid progenitor cell (Hardison et al., 2020). To test whether the HPC-7 cells in culture maintain the HPC features, I first examined the HPC-7 surface marker expression by flow cytometry analysis (Figure 3.1). The result suggests that most of the HPC-7 express the hematopoietic progenitor cells marker, CD34. And the entire population of HPC-7 cells exhibit the expression of c-Kit, CD44, and CD29. HPC-7 have no significant expression of CD3e (T cell marker), TER119 (erythroid marker), CD11b/ Mac-1 (macrophage/ monocyte marker), F4/80 (macrophage/ monocytes marker), and Gr-1 (granulocyte marker). A small proportion of HPC-7 express B220 (B cell marker), but have no expression of CD19 (B cell marker). Flow cytometry analysis with CD34/ B220 double staining revealed this population was $CD34^{+}B220^{+}$ (Data not shown). Collectively, the surface marker expression profile of HPC-7 cells in culture is consistent with the finding in Pinto do O, 1998. The expression

of c-Kit and CD34 and lack of expression of lineage marker support that HPC-7 display features of multipotent hematopoietic progenitor cells.

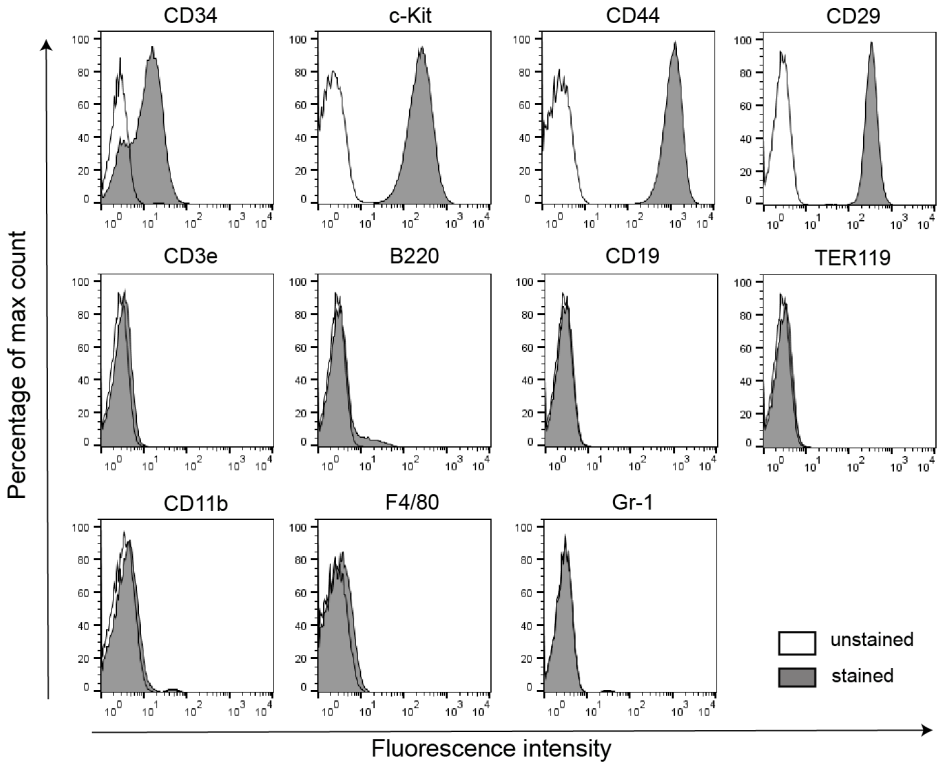


Figure 3.1: FACS analysis of surface marker expression of HPC-7 cells. Cells were cultured in SCF (100 ng/mL) before staining. Antibody stainings are as indicated. Hematopoietic progenitor cell marker: CD34 and c-Kit, T cell marker: CD3e, B cell marker: B220 and CD19, erythroid marker: TER119, macrophage/ monocyte marker: CD11b/ Mac-1 and F4/80, granulocyte marker: Gr-1.

3.2.2 HPC-7 exhibits hematopoietic differentiation potential *in vitro*

To test if the HPC-7 in culture exhibit differentiation ability, I performed colony-forming unit (CFU) assay by culturing the HPC-7 cells in the methylcellulose medium (M3434, STEMCELL Technologies) supplemented with appropriate cytokines (SCF, IL-3, IL-6 and EPO). After 14 days of culture, four types of colonies were derived from HPC-7 (Figure 3.2A):

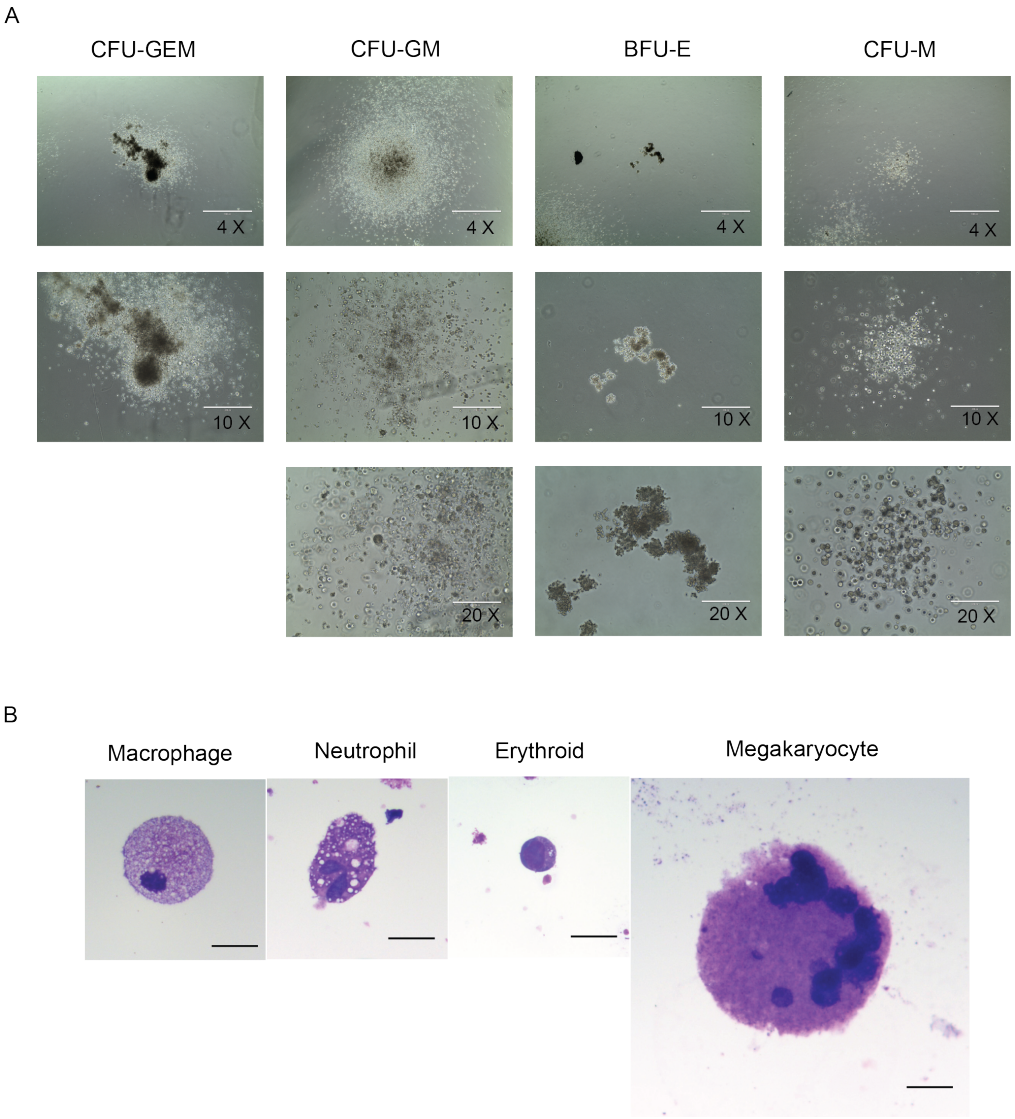


Figure 3.2: HPC-7 differentiates into various blood cell types *in vitro*. Cells were cultured in methylcellulose CFU medium with recombinant cytokines(SCF, IL-3, IL-6, EPO) for 14 days. (A) Graphs shows the formed colonies from HPC-7 cells, including CFU-GEM (granulocyte, erythrocyte, macrophage), CFU-GM (granulocyte/macrophage), BFU-E (Burst-forming unit-erythroid) and CFU-M (macrophage). (B) Giemsa-Wright staining of cells from HPC-7 derived colonies. Cytospins images are shown. Scale bar indicates 20 μM.

CFU-GEM: Colony-forming unit-granulocyte, erythroid, macrophage. CFU-GEM is smaller than the CFU-GM in size. CFU-GEM has a highly dense core with a border between the center and peripheral cells.

CFU-GM: Colony-forming unit-granulocyte, macrophage. CFU-GM is a large colony of multiple cell clusters with a dense core. The larger cells are monocytic cells, the relatively smaller round cells are granulocytes.

BFU-E: Burst-forming unit-erythroid. BFU-E is tiny and made of erythroid clusters.

CFU-M: Colony-forming unit-macrophage. CFU-M only contain macrophage lineage.

Morphology analysis on cells from the CFU colonies indicated that HPC-7 could differentiate into macrophage, neutrophil, erythroid, and megakaryocyte (Figure 3.2B). In summary, the existence of four types of colonies in the CFU assay indicated the different hematopoietic potential within the HPC-7 cell population. Some HPC-7 cells exhibit multipotent progenitor cell capacity and could differentiate into tri-lineage colony (CFU-GEM), whereas some of the cells are more restricted into one or two lineages (CFU-M/CFU-GM,BFU-E). Furthermore, nearly 90% of the colonies formed in the CFU assays were CFU-GM, suggesting most of the HPC-7 cells have a myeloid lineage differentiation preference *in vitro*.

3.2.3 Generation of Tet2 knock-out HPC-7 cell line by CRISPR-Cas9

To investigate the Tet2 function in HPC-7 cells, I generated the *Tet2* knock-out (KO) HPC-7 cell lines using CRISPR-Cas9 technique to identify cellular and molecular changes led by Tet2 deficiency. To knock-out *Tet2* gene in HPC-7, I tested and compared two Single-guide RNA (sgRNA). I designed sgRNA1 using Benchling software (https://benchling.com), to target the start codon of Tet2 first coding exon, exon3 (Figure 3.3A). sgRNA2 with the EcoRV recognition site within the target locus was published by Wang et al., 2013. Two sgRNA were cloned into the pSpCas9(BB)-2A-GFP plasmid (PX458) respectively and delivered into the HPC-7 cells by electroporation. GFP-positive single cell was sorted after electroporation and grown for clonal cell line isolation.

To identify true *Tet2*-KO clone, I firstly screened the clonal cell lines to identify clone which exhibits decreased 5hmC (Figure 3.4A). In addition to the detection of Tet2 enzymatic activity, to identify true *Tet2*-KO clone from sgRNA2-targeting group, I also utilized the restriction enzyme digestion site included in the sgRNA sequence. A 500 bp PCR product was amplified around the sgRNA2-targeted site and digested with EcoRV (Figure 3.3B). A correctly targeted allele will lose the restriction enzyme digestion site,

thus the PCR fragments can not be digested by EcoRV. All 9 sgRNA2-targeting KO clones (1-9) had mutations on both alleles of *Tet2* gene. KO clone 10 had a smaller PCR product compared with the wild-type cells, indicating a big deletion on both alleles of *Tet2* gene. In summary, all 10 clones from sgRNA2-targeting group carried mutations at both alleles of *Tet2* gene.

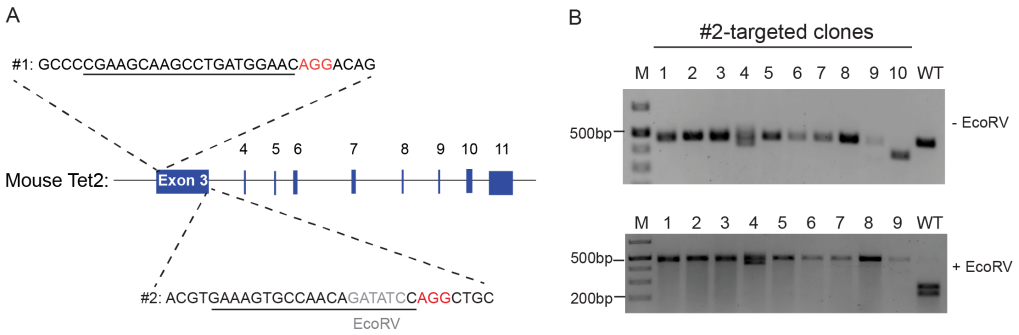


Figure 3.3: Construction of *Tet2*-KO HPC-7 cell lines by CRISPR-Cas9. (A) Schematic representation of CRISPR gRNA-targeting sites in *Tet2* gene. Two different gRNAs were designed and used for targeting *Tet2* in HPC-7 cells separately. The gRNA-targeting site is underlined, followed by PAM sequence (Red), the restriction digestion site is labeled in gray. (B) Genotyping analysis of #2 gRNA-targeted HPC-7 clones. A 500 bp PCR product was amplified around the sgRNA2-targeted site and digested with EcoRV.

Two clones (KO 1-3 and 1-8) out of ten examined clones from sgRNA1-targeting cells and three clones (KO 2-1, KO 2-7 and KO 2-9) out of ten examined clones from sgRNA2-targeting cells showed decreased 5hmC level (Figure 3.4A Right), suggesting a loss of Tet2 5mC oxidation activity. The residual 5hmC in *Tet2*-KO clones might reflect the Tet3 activity since HPC-7 cells express both Tet2 and Tet3 (Figure 3.5). Complete loss of Tet2 full-length protein was confirmed in three clones (KO 2-1, KO 2-7 and KO 2-9) from sgRNA2-targeting cells but not in clones from sgRNA1-targeting cells (Figure 3.4B). A truncated Tet2 protein which had a smaller molecular weight than the WT protein was still detected in the KO clones (Figure 3.4B). The protein truncation could be generated due to the mutation led by sgRNA1-targeting which leads to a pre-mature stop codon in the reading frame, or Tet2 might have more than one ATG to initiate translation, thus, the disruption of the predominant ATG does not lead to a complete loss of protein. Sanger sequencing revealed that both alleles of *Tet2* gene were mutated and showed different mutations in *Tet2*-KO 2-1 and KO 2-7 clone (Figure 3.4C).

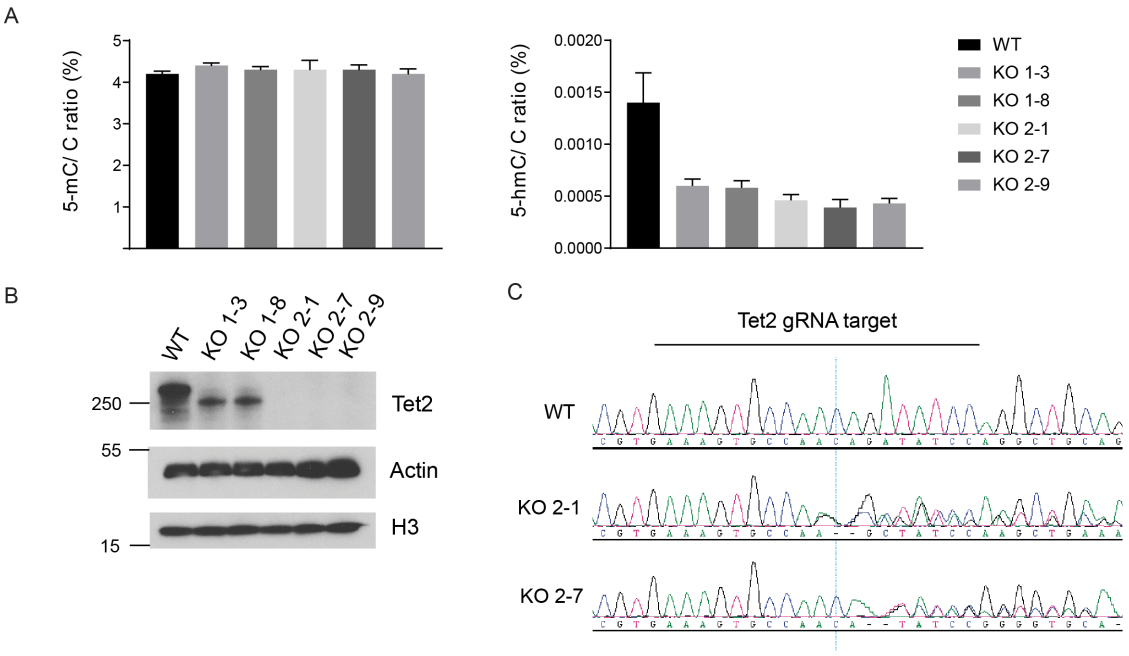


Figure 3.4: Characterization of *Tet2*-KO HPC-7 cell lines. (A) LC-MS/MS measurements of genomic 5-mC level (left) and 5-hmC level (right) in *Tet2*-KO HPC-7 clones. Mean values from three biological replicates are plotted. Error bars indicate standard deviation. LC-MS/MS measurement was done by Paolo Spingardi. (B) Western blot analysis of Tet2 protein in whole cell lysate of *Tet2*-KO HPC-7 clones. Actin and total histone H3 serve as loading controls. (C) Sanger sequence of sgRNA2-targeting sites in *Tet2*-KO clones.

To examine whether *Tet2* depletion in HPC-7 cells could lead to any compensation activity from other Tet family proteins, I compared the mRNA expression of all three *Tet* genes in *Tet2*-KO versus WT cells (Figure 3.5). *Tet2* and *Tet3* are expressed in HPC-7, whereas *Tet1* is lowly expressed in HPC-7. *Tet2*-KO clones had no significant changes in *Tet1* and *Tet3* gene expression. Interestingly, *Tet2* mRNA was still detected in *Tet2*-KO clones. In summary, two *Tet2*-KO HPC-7 cell lines, KO 2-1 and KO 2-7, were generated using CRISPR-Cas9 technique. The two *Tet2*-KO cell lines were characterized with different mutations on both alleles of *Tet2* gene, complete loss of Tet2 protein and decreased 5hmC level. In the rest of this chapter and the next chapter, I will introduce the cellular and molecular changes led by Tet2 deficiency which were shown in the proposed experiments listed in the Figure 3.6.

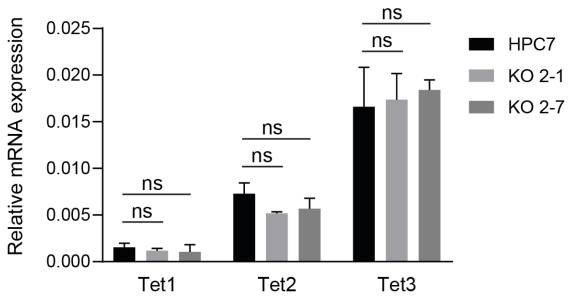


Figure 3.5: Tet2 Knock-out in HPC-7 cells does not lead to significant changes of other Tets gene expression. Quantitative RT-qPCR analysis of *Tet1, Tet2, and Tet3* in *Tet2*-KO HPC-7 clones. Data are normalized to Gapdh. Mean values from three biological replicates are plotted. Error bars indicate standard deviation. The difference between samples was compared using 2-way ANOVA and Tukey's multiple comparisons test. ns, not significant.

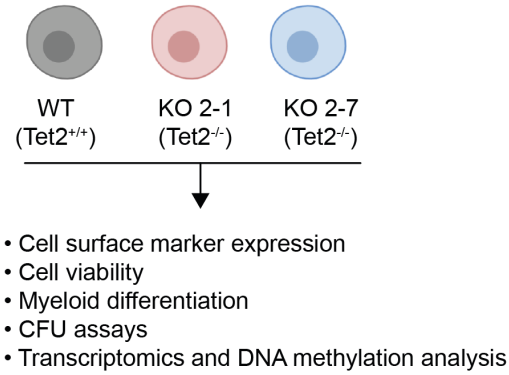


Figure 3.6: Flowchart of experiments on *Tet2*-KO HPC-7 clone 2-1 and 2-7.

3.2.4 *Tet2* depletion causes no significant changes in HPC-7 surface marker expression

Tet2 depletion in mice leads to increased hematopoietic stem cell and promotes the cell differentiation towards myeloid lineage (Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011). To understand if Tet2 deficiency alters the progenitor cell features of HPC-7 and leads to spontaneous cell differentiation under SCF-containing medium culture, I examined the surface marker expression of *Tet2*-KO and WT cells maintained in medium supplemented with SCF for expansion. In general, *Tet2*-KO clones showed similar surface marker expression as the WT cells (Figure 3.7). Particularly, *Tet2*-KO cells still showed expression of the two hematopoietic progenitor cell markers CD34 and c-Kit, and maintained the HPC-like properties without the expression of any hematopoietic lineage markers. Intriguingly, compared with WT cells, both *Tet2*-KO cells

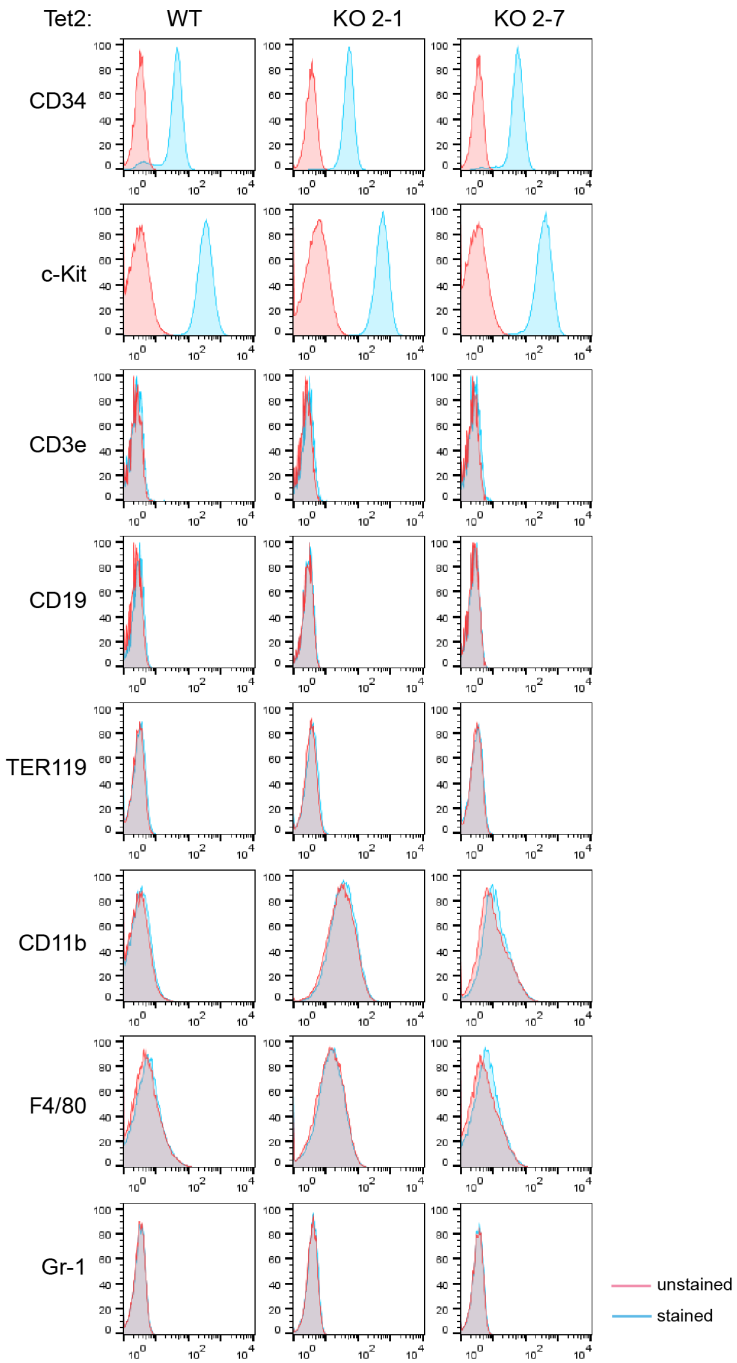


Figure 3.7: FACS analysis of surface marker expression of *Tet2*-KO HPC-7 cell lines. Cells were cultured in SCF (100 ng/mL). Antibody stainings are as indicated. Hematopoietic progenitor cell marker: CD34 and c-Kit, T cell marker: CD3e, B cell marker: B220 and CD19, erythroid marker: TER119, macrophage/ monocyte marker: CD11b/ Mac-1 and F4/80, granulocyte marker: Gr-1. WT: wildtype HPC-7, KO 2-1: HPC-7 *Tet2* knock-out clone 2-1, KO 2-7: HPC-7 *Tet2* knock-out clone 2-7.

showed reduced $CD34^-$ population, reflecting that Tet2 might play a role in regulating CD34 expression. But RNA-seq does not reveal there is a significant change of *CD34*

mRNA level in *Tet2*-KO cells compared with WT. And another possible explanation for the reduced $CD34^-$ population is that this could be a consequence of clonal cell isolation where the initial single *Tet2*-KO cell comes from the $CD34^+$ WT cells.

3.2.5 *Tet2* depletion reduces HPC-7 cell viability at low-SCF culture condition

Previous studies have suggested that TET2 might play a role in cell proliferation or cell viability. Single-cell clonogenic assays showed that LSK cells isolated from $Tet2^{-/-}$ mice developed more colonies and generated greater numbers of total cells (Li et al., 2011b), suggesting a role for TET2 in repressing cell proliferation. In support of this idea, overexpression of TET2 in human AML HL-60 leukemic cells was found to reduce the cell proliferation and clonogenicity, indicating a role for TET2 in inhibiting the leukemia cell proliferation (Wang et al., 2015b). To investigate the effect of *Tet2* depletion on cell growth in HPC-7, I checked the cell viability of *Tet2*-KO cells and WT cells grown in medium with a range of SCF concentrations for 2 days. Cell viability was examined by Annexin V and 7-AAD staining. Annexin V binds to phospholipid phosphatidylserine (PS) exposed by the apoptotic cells. 7-AAD would be excluded from the viable cells with intact membrane, whereas, dead and damaged cells lose its membrane integrity and are permeabilized to 7-AAD. Viable cells which are characterized by Annexin V^- and 7-AAD^-) were quantified by flow cytometry analysis (Figure 3.8). Both WT and *Tet2*-KO HPC-7 cells showed SCF-dependent cell proliferation and less extent of cell apoptosis was observed in higher SCF-containing medium. Moreover, *Tet2*-KO clone 2-7 exhibited significantly reduced cell viability at low SCF concentration (10 and 20 ng/mL). These data suggest that Tet2 seems to promote cell viability at low SCF concentration medium, which contradict with known function of TET2 in negatively affecting cell proliferation. The observed cell viability defects in *Tet2*-KO HPC-7 cells might be a cell type-specific phenotype.

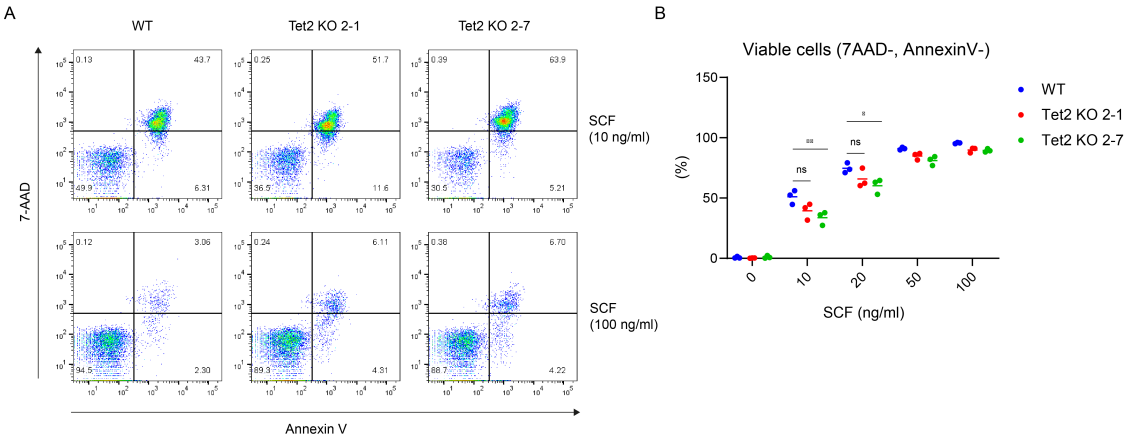


Figure 3.8: Depletion of *Tet2* in HPC-7 leads to decreased cell viability in low SCF concentration. Cells were seeded in equal cell density, cultured in media with indicated SCF concentration for 2 days, and stained with 7-AAD and Annexin V. (A) A representative example of 7-AAD and Annexin V staining of cells cultured in 10 ng/ml or 100 ng/ml SCF media. (B) The percentage of viable cells (Annexin V^-, 7-AAD $^-$) in cell culture with different SCF media. Mean values from three biological replicates are plotted. The difference between samples was compared using 2-way ANOVA and Tukey's multiple comparisons test. ns, not significant, *, $p<0.05$, **, $p<0.01$.

3.2.6 *Tet2* depletion in HPC-7 results in decreased cell viability and increased myeloid differentiation *in vitro*

Tet2 depletion in mice skew the cell differentiation towards myeloid lineage (Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011). To test the role of *Tet2* deficiency in HPC-7 cell differentiation, I cultured the *Tet2*-KO and WT cells in the presence of G-CSF (granulocyte colony-stimulating factor), GM-CSF (granulocyte-macrophage colony-stimulating factor), IL-3 (Interleukin-3) and IL-6 (Interleukin-6), a cytokine combination designed for myeloid differentiation (Dassé et al., 2012). After 8 days, almost all the HPC-7 cells showed decreased expression of the stem cell marker CD34 and some of the HPC-7 cells expressed the myeloid marker CD11b and macrophage marker F4/80 (Figure 3.9A and B). Compared to WT cells, *Tet2*-KO HPC-7 cells had a larger population of F4/80-expressing cells and CD11b-expressing cells (Figure 3.9C), suggesting that the absence of Tet2 could promote HPC differentiation into myeloid lineage. This differentiation phenotype in *Tet2*-KO HPC-7 cells is consistent with the finding in *Tet2*-deficient mouse models, where Tet2 deficiency leads to an elevated extramedullary myelopoiesis in the significantly enlarged spleen of the animals (Ko et al.,

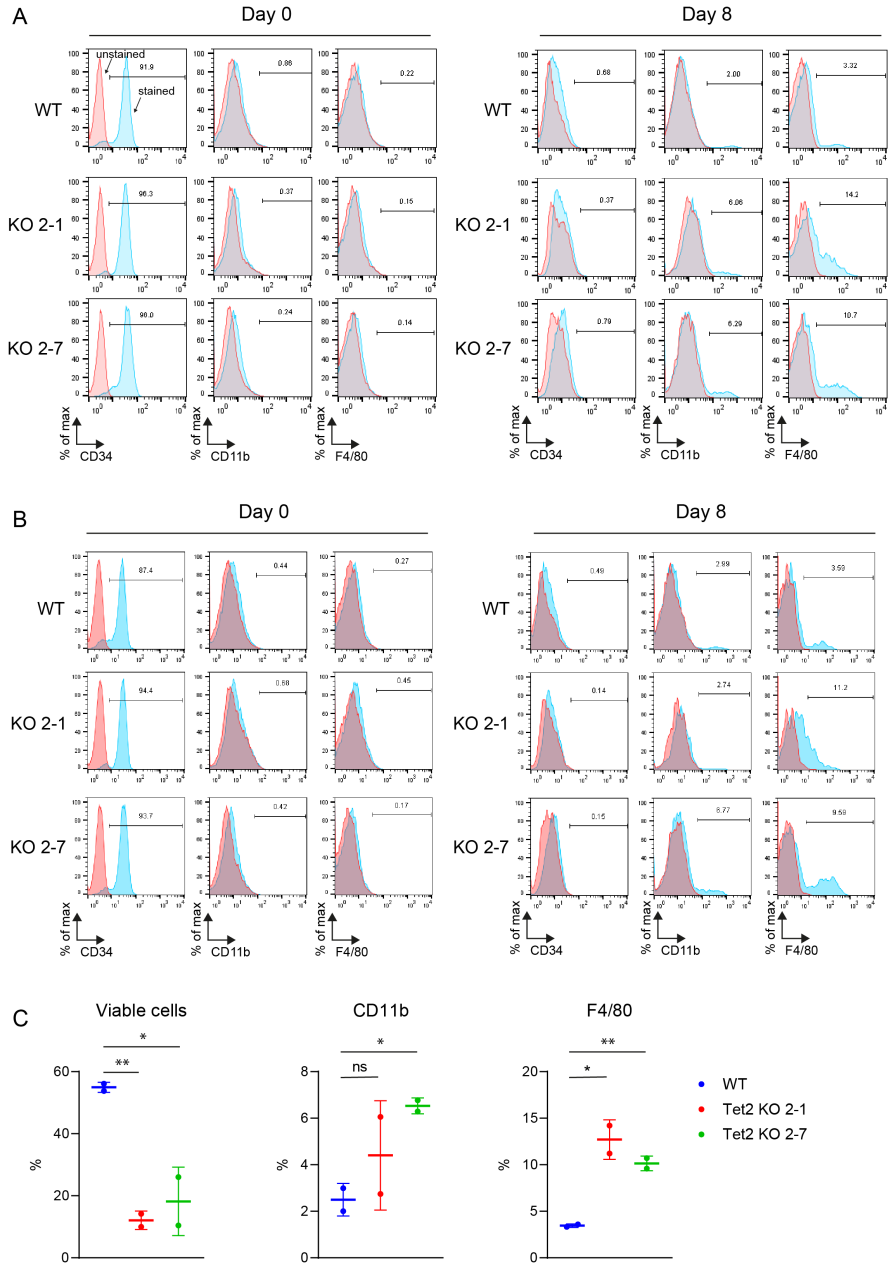


Figure 3.9: Depletion of *Tet2* in HPC-7 results in decreased cell viability and increased myeloid differentiation *in vitro*. Cells were cultured in Stem Pro-34 media supplemented with G-CSF, GM-CSF, IL-3 and IL-6 for 8 days. Cells were stained with murine HPC marker, CD34, and myeloid lineage markers, CD11b and F4/80, and analyzed by flow cytometry. (A) and (B) Flow cytometry analysis of surface marker expression of *Tet2*-KO cells during myeloid differentiation. (A) and (B) were results from two biological replicates. (C) The percentage of viable cells (7-AAD$^-$), CD11b$^+$, and F4/80$^+$ cells in *Tet2*-KO cells cultured in differentiation media for 8 days. Mean values from two biological replicates are plotted. Error bars indicate standard deviation. The difference between samples was compared using unpaired t-test. ns, not significant, *, $p<0.05$, **, $p<0.01$

2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011). A similar differentiation phenotype was also observed in an *in vitro* study where knock-down of *Tet2* in mouse bone marrow cells led to an expansion of CD11b and F4/80 cells when culturing with G-CSF and GM-CSF (Ko et al., 2010). Another interesting finding of the *in vitro* myeloid differentiation assay was the impaired viability in *Tet2*-KO cells (Figure 3.9C left). The detailed mechanism for the cell viability defects can not be addressed without further experiments.

3.2.7 *Tet2* depletion in HPC-7 causes reduced colony numbers and immature cell differentiation in CFU assay

To examine if *Tet2* depletion had an effect on the progenitor cell potential, I performed the CFU assay on the two *Tet2*-KO clones to determine the cell proliferation and differentiation ability of individual cells upon Tet2 deficiency. After 14 days of culture,

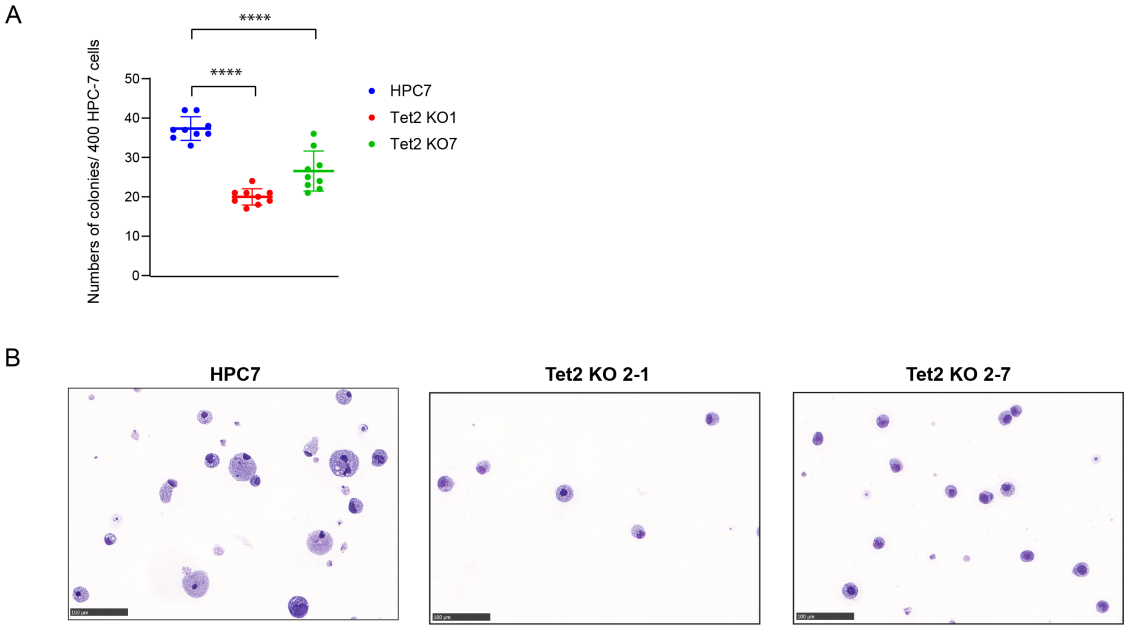


Figure 3.10: Depletion of *Tet2* in HPC-7 causes reduced colony numbers and immature cell differentiation in CFU assay.Cells were cultured in methylcellulose CFU medium with recombinant cytokines(SCF, IL-3, IL-6, EPO) for 14 days. (A) Absolute number of colonies is shown. Mean values from three biological replicates (three technical replicates for each independent experiment) are plotted. Error bars indicate standard deviation. The difference between samples was compared using unpaired t-test. ****, <0.0001. (B) Giemsa-Wright staining of cells from *Tet2*-KO cell-derived colonies. Cytospins images are shown. Scale bar indicates 100 μM.

both *Tet2*-KO clones had the significantly reduced numbers of colonies than the WT cells (Figure 3.10A), which is in line with the observation of poor viability identified in *Tet2*-KO cells in SCF culture assay (Figure 3.8) and myeloid differentiation (Figure 3.9B. Left).

Strikingly, most colonies derived from *Tet2*-KO cells were CFU-M and showed similar shape and structure under microscope. Giemsa staining revealed that the differentiated cells in *Tet2*-KO exhibited an uniformed cell morphology compared to cells developed from WT cells (Figure 3.10B), reflecting a cell differentiation block phenotype caused by Tet2 deficiency. And this cell differentiation defect phenotype was also supported by the previous CFU assay on c-Kit$^+$ progenitor cells from Tet2-deficient mouse (Quivoron et al., 2011), suggesting a role of Tet2 in restricting the immature cell differentiation.

3.3 Discussion

In this chapter, I chose the mouse hematopoietic progenitor cell line, HPC-7 cells, as a model to study the molecular mechanism of Tet2-mediated gene regulation in a hematopoietic context. The mulitpotency of HPC-7 cells was confirmed by the CFU assay (Figure 3.2). To study the Tet2 function in HPC-7 cells, a *Tet2*-KO HPC-7 cell line was generated using CRISPR-Cas9 technique. *Tet2*-KO clone 2-1, and *Tet2*-KO clone 2-7 were identified as true KO clone which is characterized with complete loss of Tet2 full-length protein and decreased 5hmC level (Figure 3.4).

Tet2-KO cells exhibited similar cell growth rate as the WT cells when maintained in the SCF medium at the concentration of 100 ng/mL, whereas in low SCF concentration, KO cells showed slightly impaired cell viability (Figure 3.8), and this viability defect was exacerbated in liquid myeloid differentiation (Figure 3.9) and CFU assays (Figure 3.10). Lack of SCF in the differentiation medium might contribute to the observed poor viability in the liquid myeloid differentiation since a great number of cells died at the myeloid differentiation Day2 when cells showed no sign of differentiation. But this could not explain the reduced numbers of colonies derived from the *Tet2*-KO cells in the CFU assay since the CFU medium contains SCF. The reduced colonies numbers in *Tet2*-KO

cells might reflect: 1. The proportion of progenitor cell population within HPC-7 cells is decreased in *Tet2*-KO group since CFU assay is to examine the cell proliferation and differentiation ability of each individual cells within the population; 2. *Tet2* depletion might provide a cell proliferation disadvantage and the cell viability defect become more significant when cells were under certain stimuli, such as, low SCF or differentiation pressure. Furthermore, the cell viability defect led by *Tet2* depletion in HPC-7 seems to contradict the HSC growth advantages phenotype observed in *Tet2*-deficient mice and increased *in vitro* replating ability (Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011). Given that HPC-7 is a hematopoietic progenitor cell line originally derived from mouse embryonic stem cells and might not be able to capture all the characteristics of the *in vivo* bone marrow-resided HSC/HPC cells, the cell growth defect caused *Tet2* depletion in HPC-7 could be a cell-type dependent phenotype specific to HPC-7 itself. Furthermore, *in vitro* culture system oversimplifies the HSC/HPC growth environment in the mice, which makes it difficult to directly compare this HPC-7 cell growth defect to the HSC growth advantages reported in the *Tet2*-/- mice, so that it might be interesting to do *in vivo* competitive repopulating assay to transplant both the WT and *Tet2*-KO HPC-7 cells into the same recipient mouse to compare their growth ability.

In addition to the cell viability difference, *Tet2*-KO HPC-7 cells showed increased differentiated cells marked with the expression of CD11b and F4/80 when culturing in IL-3, IL-6, G-CSF and GM-CSF (Figure 3.9). Strikingly, in the CFU assay, differentiated *Tet2*-KO cells exhibited similar cell morphology, whereas WT group had more diverse cell types after differentiation (Figure 3.10). The uniformed differentiated cells might reflect a differentiation arrest caused by Tet2 deficiency. And it would be interesting to identify the stage of cell differentiation in the *Tet2*-KO cells. Finally, the molecular changes caused by *Tet2* deficiency and the possible mechanism for Tet2-mediated gene regulation would be introduced in the Chapter 4.

4

The molecular role of Tet2 function in HPC7 cells

Contents

4.1 Introduction

In this chapter, I focus on the understanding of the molecular changes upon Tet2 deficiency in HPC-7 cells and the possible mechanism of Tet2 function in gene regulation. I will firstly introduce the gene expression changes measured by RNA-seq in *Tet2*-KO HPC-7 cells (Chapter 4.2.1). In order to identify Tet2-direct gene targets, endogenous ChIP experiment was initially attempted and failed to continue because of high background of ChIP signal in control cells (*Tet2*-KO cells). Tet2 binding sites (BS) were finally determined by HA ChIP-seq on *Tet2*-KO HPC-7 with the re-expression of 3xHA Tet2 (Chapter 4.2.2). The distribution of genomic features, functional pathway and transcription factor binding sites enriched in Tet2 BS will be explained in Chapter 4.2.3. The effect of Tet2 binding in gene regulation will be discussed in Chapter 4.2.4. 5mC Changes in *Tet2*-KO versus WT cells were quantified by TAPS-β (Chapter 4.2.5) and the contribution of Tet2-involved DNA demethylation to gene expression will be discussed in Chapter 4.2.6 and 4.2.7. Finally, I will introduce the observed co-localization of Tet2 and several key hematopoietic transcription factors in HPC-7 genome (Chapter 4.2.8). The TAPS-β experiment was performed by Dr. Sophie Kirschner and the data was analyzed by Olena Yavorska.

4.2 Results

4.2.1 Differential gene expression analysis in *Tet2*-KO HPC-7 cells

To uncover the molecular changes led by Tet2-deficiency, I performed calibrated RNA-sequencing using mRNA isolated from *Tet2*-KO (clone 2-1 and clone 2-7) and WT HPC-7 cells with chicken embryo fibroblast DF-1 cell spike-in (Figure 4.1). The typical gene expression analysis is based on the assumption that the total number of transcripts per cell is almost constant across all conditions (Taruttis et al.), but this assumption does not always hold (e.g. P493-6 cells expressing higher levels of *c-Myc* were found to produce 2-3-fold higher levels of RNA found in cells with low level of *c-Myc* (Lin et al.)). The

use of spike-in cells would allow the normalization to cell number and permit correction for differences in RNA yields among conditions (Taruttis et al.).

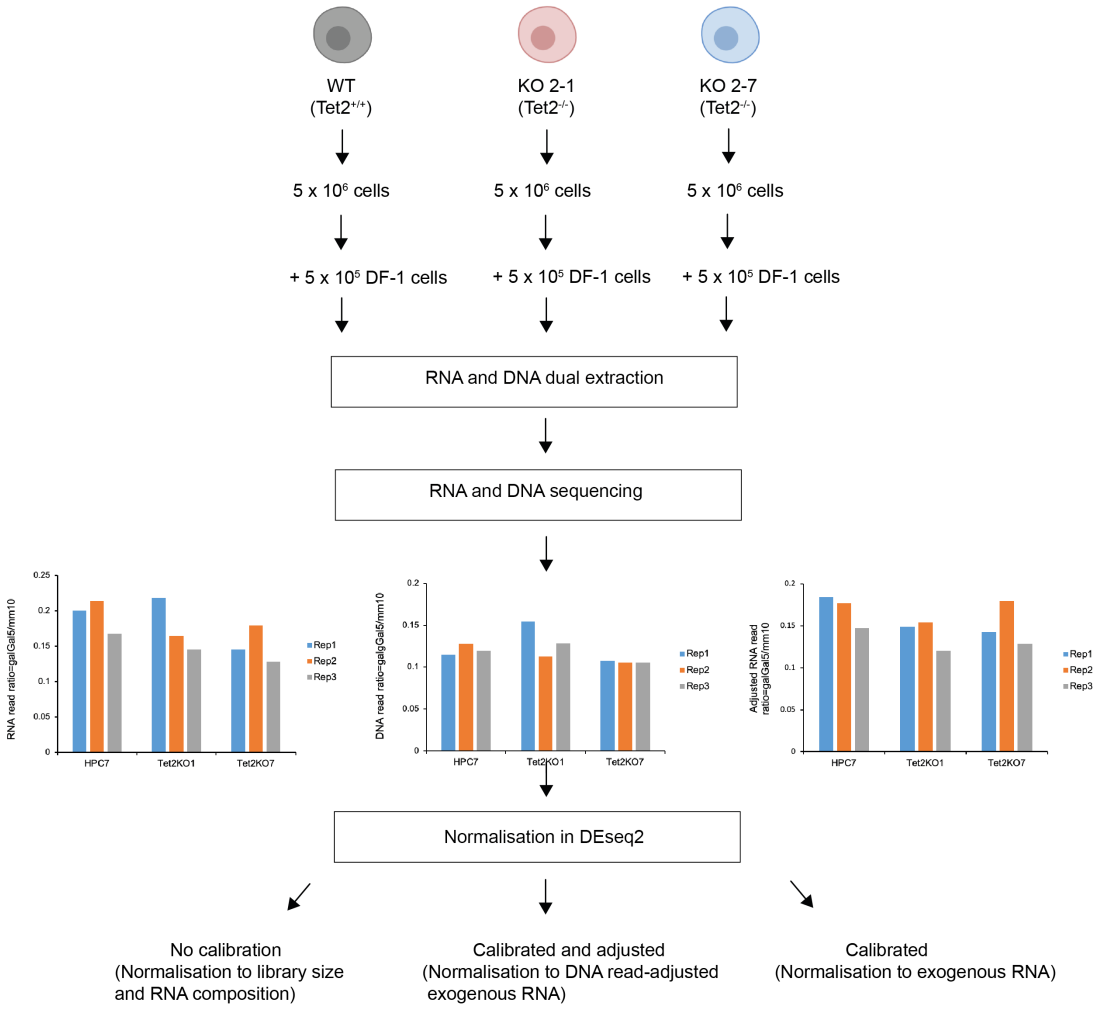


Figure 4.1: Flowchart of RNA-sequencing experiments on *Tet2*-KO HPC-7 cells. Chicken DF-1 cells were spiked-in HPC-7, *Tet2*-KO clone 2-1, and *Tet2*-KO clone 2-7 cells on a per-cell basis followed by RNA/ DNA dual extraction and genomic sequencing. Three biological replicates were prepared. RNA-seq (bar chart on the left) and DNA-seq (bar chart in the middle) reads ratio between mm10 and galGal5 genome were calculated. Adjusted RNA-seq reads ratio (bar chart on the right) between mm10 and galGal5 genome was calculated based on the DNA-seq read ratio. Three normalization methods were compared in DESeq2 analysis.

Briefly, fixed numbers (5 x 10^7) of HPC-7 cells were mixed with fixed numbers (5 x 10^6) of DF-1 cells. RNA and genomic DNA were extracted and sequenced on the same sample. After sequencing, reads aligned to mouse and chicken genome were separated and counted. Spike-in calibration was incorporated into the differential gene expression analysis by DESeq2 (Love et al., 2014). Two types of count normalization size factors

with spike-in calibration were calculated to scale the expression of mm10 genes relative to those of galGal5 genes. In the calibrated and adjusted method, before size factor calculation, galGal5 RNA-seq reads counts in each sample were pre-normalized using the galGal5/mm10 read ratio of the corresponding genomic-DNA seq samples to minimize the variation in spike-in cell mixing. In the calibrated method, galGal5 RNA-seq read counts were not adjusted by the galGal5/mm10 DNA-seq count ratio before quantification. Count normalization size factor without spike-in calibration was calculated using the default median of ratios method to adjust for sequencing depth and RNA composition. Read counts and alignment rates were provided in Table 4.1. galGal5 DNA-seq reads were $\sim$10% of mm10 DNA-seq reads in each sample, indicating the equal mixing of the spike-in cells.

Table 4.1: Total read counts and alignment metrics for RNA-seq experiments in *Tet2*-KO HPC-7 cells.

	DNA-sequencing					RNA-sequencing					Adjusted		
Sample name	Total paired-end reads	% Uniquely mapped reads	Number of mm10 uniquely mapped reads	Number of galGal5 uniquely mapped reads	DNA read ratio=galGal5/mm10	Total paired-end reads	% Uniquely mapped reads	Number of mm10 uniquely mapped reads	Number of galGal5 uniquely mapped reads	RNA read ratio=galGal5/ mm10	Factor to normalise to minimum DNA read galGal5/mm10 ratio	Downsampled RNA galGal5 fragment count (expected)	RNA read ratio= downsampled galGal5/ mm10 (expected)
HPC7 WT rep1	4060860	69.60	4875573	558781	0.11	72288887	81.50	45050959	9023797	0.20	0.92	8290924	0.18
HPC7 WT rep2	3953485	72.16	4882236	623151	0.13	72130789	85.00	42922438	9184721	0.21	0.83	7577414	0.18
HPC7 WT rep3	4191089	71.89	5213608	622677	0.12	61035805	85.90	41158404	6891374	0.17	0.88	6075904	0.15
Tet2 KO1 rep1	3979602	70.22	4672958	720651	0.15	66754173	83.20	45236706	9860978	0.22	0.68	6733121	0.15
Tet2 KO1 rep2	4808404	72.34	6007635	677288	0.11	71631326	85.40	47996306	7911959	0.16	0.93	7389988	0.15
Tet2 KO1 rep3	4130224	73.41	5179715	663295	0.13	62323317	84.90	48995629	7140073	0.15	0.82	5871261	0.12
Tet2 KO7 rep1	4831345	72.36	6071725	652682	0.11	61470653	85.50	41988531	6104575	0.15	0.98	5979921	0.14
Tet2 KO7 rep2	3998092	71.59	4983683	525004	0.11	72039073	85.30	41127286	7392561	0.18	1.00	7389453	0.18
Tet2 KO7 rep3	5136352	70.59	6328862	666431	0.11	60400600	85.20	42031258	5392224	0.13	1.00	5392224	0.13

The relationship between replicates and conditions were examined by Principal component analysis (PCA) and hierarchical cluster analysis. Unexpectedly, biological replicates within the same group were more separated in the PCA plot when spike-in calibration was considered for the differential gene expression analysis (Figure 4.2A). Pre-normalization of the spike-in RNA-seq count to the DNA-seq counts seem to be necessary for the calibrated RNA-seq analysis since sample clustering was less affected in the calibrated/adjusted group compared to calibrated group (Figure 4.2A Middle and

Right). Similarly, replicates from each sample group were not clustered in the same block when samples were normalized with spike-in calibration (Figure 4.2B). Taken together, the altered sample clustering of replicates within the same group suggest that spike-in calibration should not be considered for read count normalization for these experiments. Importantly, *Tet2*-KO clone 1 and clone 7 were clustered together, suggesting a similarity of gene expression changes by Tet2 deficiency.

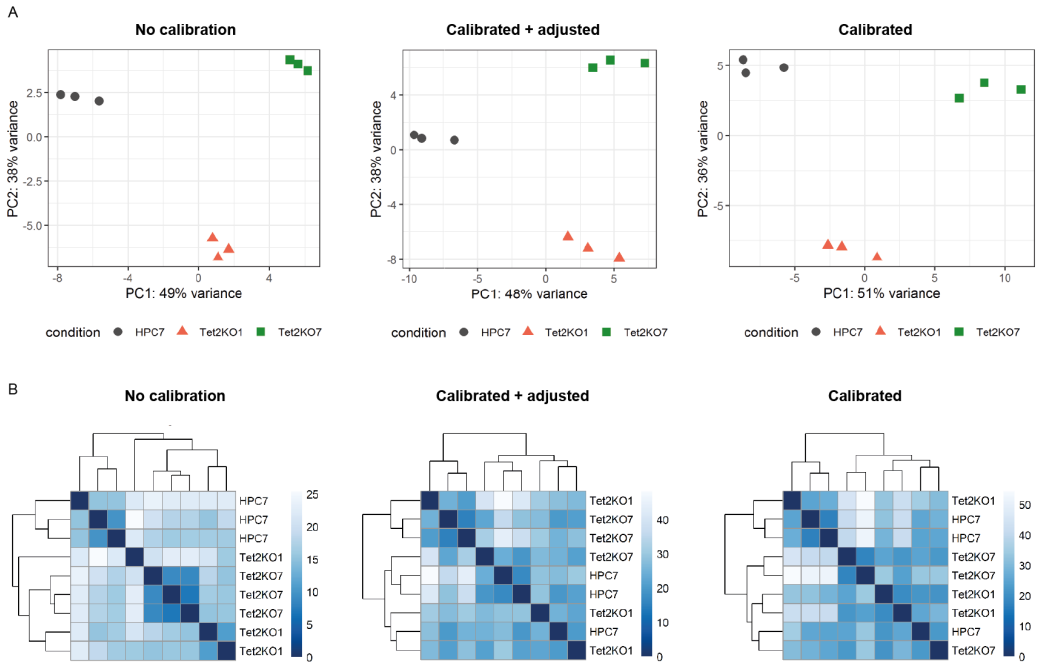


Figure 4.2: Exogenous Chicken DF-1 cell calibration in DESeq2 analysis affects the sample relationships of biological replicates within the same sample group. (A) Principal component analysis (PCA) of samples from RNA-seq experiments on *Tet2*-KO HPC-7 cells. (B) Heatmap of sample distance matrix analysis of samples from RNA-seq experiments on *Tet2*-KO HPC-7 cells. No calibration: normalization to RNA-seq sample library size in DESeq2 analysis; Calibrated and adjusted: normalization to adjusted exogenous RNA-read ratio in DESeq2 analysis; Calibrated: normalization to exogenous RNA-read ratio in DESeq2 analysis.

To identify differentially expressed genes, I finally did the read count normalization by the median of ratios method to account for sequencing depth and RNA composition (Anders and Huber, 2010). Differentially expressed genes in *Tet2*-KO cells were identified by pairwise comparison between each *Tet2*-KO clone and WT HPC-7 cells (Figure 4.3 A). Consistently, there are slightly more up-regulated genes (Tet2-*KO* 2-1: 625; Tet2-*KO* 2-7: 819) than down-regulated genes (Tet2-*KO* 2-1: 491; Tet2-*KO* 2-7: 552) observed in both *Tet2*-KO clones, but the down-regulated genes showed larger $\log_2$ fold

change. Moderately and highly expressed genes were more found to change upon Tet2 deficiency. *Tet2*-KO clone 2-7 has slightly more differentially expressed genes than KO clone 2-1 (1371 versus 1116). And *Tet2*-KO clone 2-1 and *Tet2*-KO clone 2-7 share some common differentially expressed genes (Figure 4.3B). Nearly 50% of the differentially expressed genes were identified in both *Tet2*-KO clones.

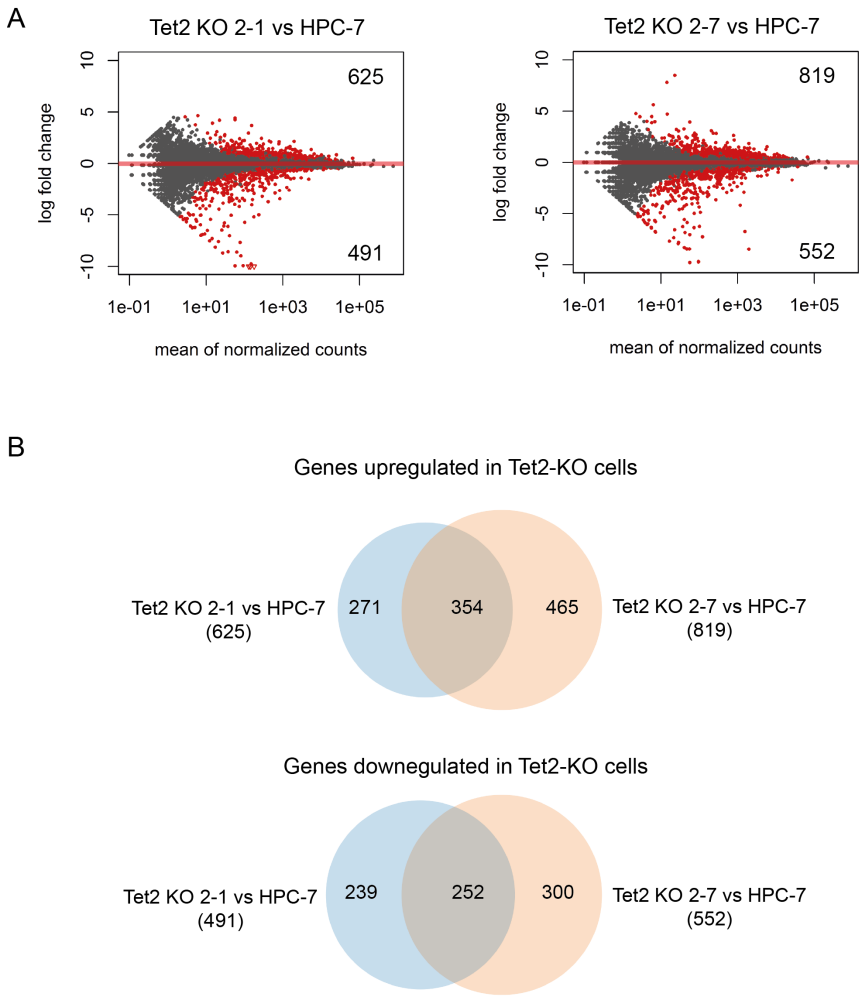


Figure 4.3: RNA-seq analysis identifies dysregulated genes in *Tet2*-KO clones relative to wildtype HPC-7 cells. (A) Left: MA plot displaying genes differentially expressed in *Tet2*-KO clone 2-1 relative to wildtype HPC-7 cells. Right: MA plot displaying genes differentially expressed in *Tet2*-KO clone 2-7 relative to wildtype HPC-7 cells. MA plot showing the mean of the normalized counts versus the $\log_2$ fold changes for all genes. Differential gene expression analysis was performed by using DESeq2 package. The genes that are significantly expressed (FDR < 0.05) are colored in red. The number of differentially expressed genes in each group is indicated. (B) Venn diagram shows overlap of significantly differentially expressed genes between *Tet2*-KO 2-1 vs wildtype HPC-7 (blue) and Tet2 KO 2-7 vs wildtype HPC-7 (orange). The number of genes in each group is indicated.

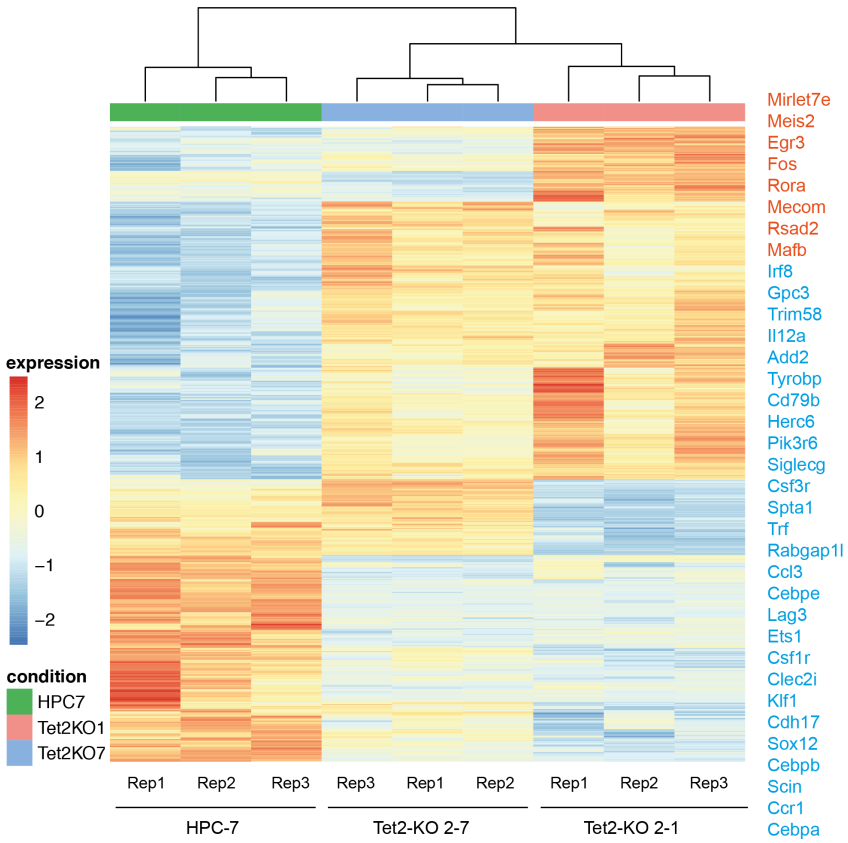


Figure 4.4: Heatmap showing the gene expression changes in *Tet2*-KO versus WT HPC-7 cells. Genes listed in the graph were hematopoietic-related genes showing significantly differential expression (p.adjust<0.05 and LogFC>1.5) in *Tet2*-KO cells. Genes upregulated in KO cells were in red, down-regulated genes were in blue.

More importantly, genes related to hematopoietic processes showed differential expression upon Tet2 deficiency in HPC-7 cells (Figure 4.4). 125 genes in mouse hemopoiesis gene set (n =980, GO:0030097, MGI) were identified as significantly differentially expressed genes (p.adjust<0.05) in *Tet2*-KO clone 2-7 versus WT cells (Table 8.1). Consistent with the published RNA-seq experiments on blood cells isolated from *Tet2*-/- mouse (Izzo et al., 2020; Moran-Crusio et al., 2011), *Tet2*-KO HPC-7 cells showed significant upregulation of stem cell marker, *c-Kit*, HSC self-renewal factors, *Hlf* and *Sox4* and exhibited decreased expression of myeloid-specific factors (e.g. *Cebpa, Cebpe, Csf1r*). *Cebpa* (CCAAT/enhancer binding protein α) is important for the transition from common myeloid progenitor (CMP) to the granulocyte/monocyte progenitor (GMP) (Zhang et al., 2004). *Cebpe* (CCAAT/enhancer binding protein ϵ) is required for normal granulocytic

differentiation (Yamanaka et al., 1997). Macrophage colony-stimulating factor (M-CSF) receptor *CSF1r* (colony-stimulating factor-1 receptor) gene, also known as *c-fms*, is critical for myelopoiesis and macrophage differentiation (Tagoh et al., 2002). The altered expression of myeloid-specific factors in *Tet2*-KO cells might account for the immature cell differentiation phenotype observed in the CFU assay. Tet2 deficiency in HPC-7 cells also impaired the expression of transcription factors involved in erythrocyte differentiation such as, *Klf1, Gata1*, and *Sox6* (Table 8.1). A recent study has revealed decreased peripheral red blood cells in *Tet2*-KO mice and reduced ability of *Tet2*-KO bone marrow cells to generate erythroid colonies, suggesting a role of Tet2 in regulating erythroid cell fate (Izzo et al., 2020). The reduced erythroid-lineage TFs might provide a way to explain the function of Tet2 in erythrocyte differentiation, and the mechanism of Tet2 regulation on TF gene expression remain further explored.

Among the upregulated genes, I found nearly all members of Egr (early growth response gene) family, including *Egr3, Egr2* and *Egr1*. Egr proteins contain highly conserved zinc-finger DNA-binding domains and can bind to same cognate GC-rich consensus DNA binding motif, the Egr response element (ERE) (Beckmann and Wilce, 1997; Swirnoff and Milbrandt, 1995). Egr proteins are involved in T cell positive selection and thymocyte differentiation (Bettini et al., 2002; Lauritsen et al., 2008). Egr-1 is also a positive modulator of macrophage differentiation and negatively regulates granulocytic differentiation (Krishnaraju et al., 2001). More importantly, over-expression of Egr2 and Egr3 significantly impaired the growth of LSK cells in SCF-containing culture medium (Aoyama et al., 2018), which recapitulate the growth defect phenotype observed in the *Tet2*-KO cells.

Gene ontology (GO) analysis and pathway analysis of the common differentially expressed genes identified in both *Tet2*-KO clones revealed that many essential cellular processes of hematopoiesis were significantly enriched (Figure 4.5). Upregulated genes were enriched in genes related to positive regulation of transcription such as *NCoA1*, *NCoA3* and *Sox4* (Figure 4.5A Left). Downregulated genes showed more immune-related biological processes, while upregulated genes exhibited more general cellular functions

(Figure 4.5A Right). In summary, *Tet2*-KO in HPC-7 cells lead to reduced myeloid-specific factors and TFs important for erythrocyte lineage, and in the meantime, more genes, such as, *Egr* family, were up-regulated in *Tet2*-KO cells, suggesting a repressive role of Tet2-mediated gene regulation.

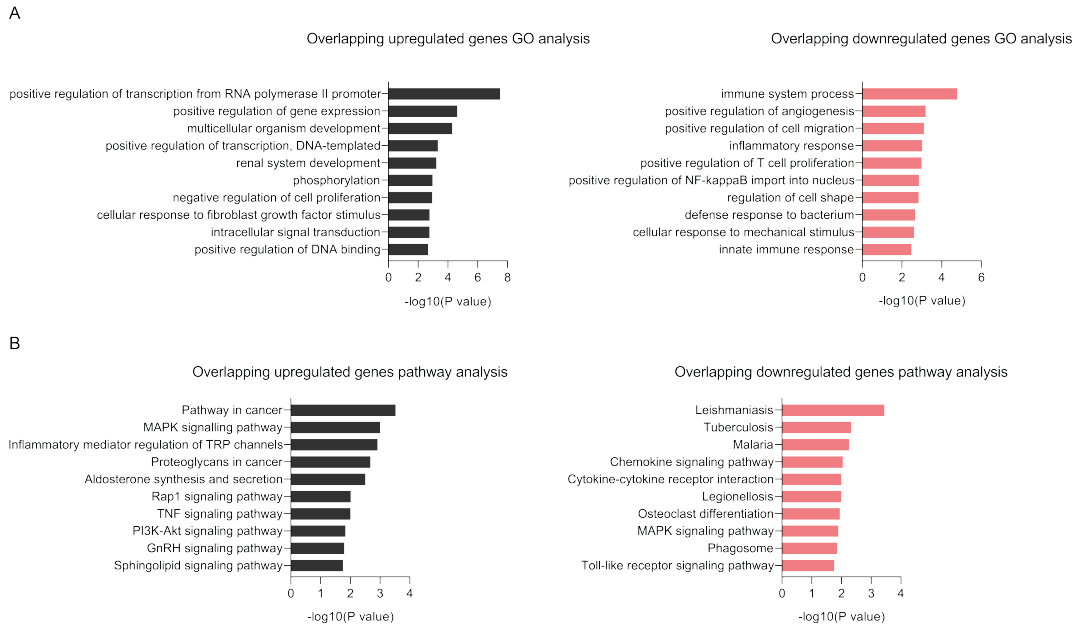


Figure 4.5: RNA-seq analysis reveals the enriched functional pathways in *Tet2*-KO HPC-7 cells. (A) Gene Ontology (GO) analysis of genes significantly upregulated (Left, n=354) and downregulated (Right, n=252) in both *Tet2*-KO HPC-7 clones. (B) Pathway analysis of genes significantly upregulated and downregulated in both *Tet2*-KO HPC-7 clones.

4.2.2 Identification of Tet2 binding sites in HPC7 cells

To identify molecular events directly regulated by Tet2, I performed ChIP-seq (Chromatin immunoprecipitation-sequencing) experiments to map the Tet2 binding sites in HPC-7 cells. Optimisation of ChIP experiments on endogenous Tet2 protein and on ectopic 3xHA Tet2 protein were both attempted in this study.

Optimisation of Tet2 ChIP-qPCR on WT HPC-7 cells

Tet2 is relatively a large protein (MW >200 KDa) and has no clear DNA binding domain suggesting an indirect DNA binding ability of Tet2. These represent the difficulties in mapping the endogenous protein chromatin binding landscape. A few Tet2 ChIP-seq experiments have been carried out using either Tet2 protein antibody or protein tag

antibody to recognize the tagged-version Tet2 protein (Chen et al., 2013; Deplus et al., 2013; Lio et al., 2016; Pan et al., 2017; Rasmussen et al., 2019; Wang et al., 2018). When I started the project, there were no Tet2 ChIP-seq experiments performed in normal hematopoietic cells without any extra gene mutations. A recent study from Rasmussen and colleagues generated a Tet2 antibody raised against the N-terminus of the protein and used it to ChIP endogenous Tet2 in immortalized mouse myeloid cells (AML1-ETO) (Rasmussen et al., 2019). I will discuss their findings in comparison with our results in Chapter 4.2.3.

In this project, I firstly tested several commercial Tet2 ChIP-grade antibodies by doing ChIP-qPCR in mESCs since there are reported mESCs Tet2 binding sites (Chen et al., 2013). Among the tested antibodies (Figure 4.6), the Tet2 antibody from Abcam has been used for ChIP-qPCR/ChIP-seq experiments in published literature (Lio et al., 2016). In addition to the antibody varieties, I also tried a two-step cross-linking protocol where protein-protein interactions are stabilized by Disuccinimidyl glutarate (DSG) cross-linking, followed by formaldehyde DNA-protein crosslinking (Tian et al., 2012). ChIP-qPCR results suggest that the two-step cross-linking protocol is necessary for Tet2 ChIP since DSG treated groups showed clear Tet2 binding at the published Tet2 binding sites (Promoter region of gene: *Asah3l*, *Mll1* and *Zmym6*). The TET2 antibody from CST (Cell signalling Technology) showed the best enrichment of ChIP signal over the input at the known Tet2 ESCs binding sites compared to other Tet2 antibody. The two Millipore antibodies did not show the enrichment of Tet2 binding at the reported Tet2 binding sites compared to the ChIP-qPCR signal at the Tet2 un-bound regions.

Tet2 antibody from CST was chosen to ChIP endogenous Tet2 in HPC-7 cells (Figure 4.7). To estimate the antibody non-specific binding, ChIP using rabbit IgG antibody was included as the control. I utilized the constructed *Tet2*-KO HPC-7 cells and did the ChIP experiments using the Tet2 CST antibody as another control for antibody binding background. Interestingly, Tet2 ChIP-qPCR experiments in HPC-7 showed enrichment of Tet2 binding at the putative Tet2 binding sites and no Tet2 binding was observed at the Tet2-unbound region (NC) (Figure 4.7A). Surprisingly, Tet2 ChIP enrichment was

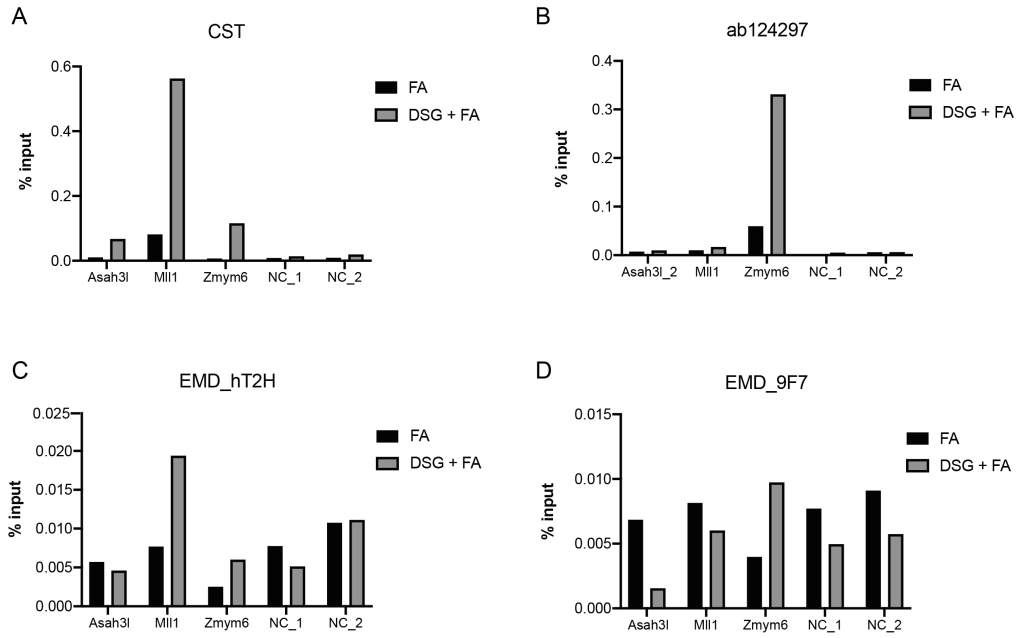


Figure 4.6: RT-PCR analysis of Tet2 ChIP in mouse ES with different Tet2 antibodies and cross-linking conditions. Mouse ESCs were crosslinked with DSG + FA or FA alone. Crosslinked chromatin was immunoprecipitated with Tet2 antibody purchased from different supplier: (A) CST (Cell Signaling Technology); (B) Abcam,124297; (C) EMD Millipore, clone hT2H; (D) EMD Millipore, clone 9F7. ChIP DNA was analyzed by RT-PCR with primer pairs specific for the indicated genes. The percentage enrichment over input (% input) was calculated to represent the difference in Ct values between qPCR on ChIP DNA and input sample. NC: negative control, no Tet2 binding. One biological replicate was prepared. DSG: disuccinimidyl glutarate; FA: formaldehyde.

also observed in *Tet2*-KO cells group. Comparatively, IgG control group showed no Tet2 binding at the putative binding sites. The contradicting results from the IgG control and *Tet2*-KO cell control raise concerns about the specificity of the Tet2 antibody. Most of the Tet2 antibodies were designed to recognize the C-terminus of the protein where all Tet family proteins share a high similarity in the C-terminal catalytic domain. The observed ChIP-qPCR signal in the KO cells group might have resulted from the cross-reactivity of other Tet proteins.

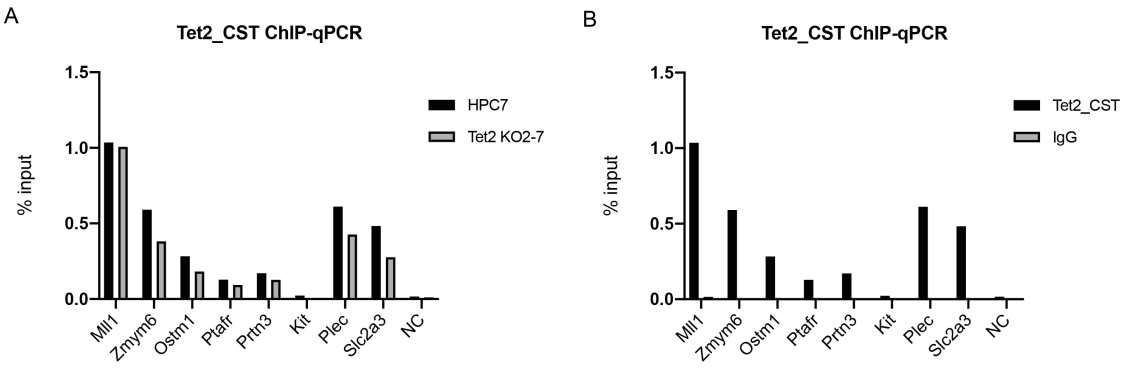


Figure 4.7: RT-PCR analysis of Tet2 ChIP in HPC-7 with antibody from CST. (A) Wildtype HPC-7 cells or *Tet2* KO 2-7 cells were crosslinked with DSG + FA. Crosslinked chromatin was immunoprecipitated with Tet2 antibody purchased from CST (Cell Signaling Technology). (B) Wildtype HPC-7 cells were crosslinked with DSG + FA. Crosslinked chromatin was immunoprecipitated with IgG antibody or Tet2 antibody purchased from CST. ChIP DNA was analyzed by RT-PCR with primer pairs specific for the indicated genes. The percentage enrichment over input (% input) was calculated to represent the difference in Ct values between qPCR on ChIP DNA and input sample. NC: negative control, no Tet2 binding. One biological replicate was prepared.

Identification of differential Tet2 binding sites in PiggBac 3xHA Tet2 versus *Tet2*-KO cells

To avoid the non-specific ChIP background from the Tet2 antibody, I utilized a constructed HPC-7 cell line which ectopically expressed the 3xHA Tet2 protein and performed anti-HA ChIP-seq to identify Tet2 genomic binding activities. Details about the 3xHA Tet2 HPC-7 cell line would be introduced in Chapter 5.2.3. Briefly, 3xHA-Tet2 was re-introduced into the HPC-7 *Tet2*-KO clone 2-7 cells and the expression of 3xHA Tet2 protein was tuned to be similar to the endogenous level by modulating dox concentration. To find potential Tet2 binding sites in HPC-7 cells, which I could use for validation of HA ChIP, I overlapped the genes differentially expressed in HPC-7 *Tet2*-KO cells with reported Tet2 peaks in mouse myeloid cells (Rasmussen et al., 2019) (Figure 4.8). Differentially expressed genes with putative Tet2 binding at their promoter regions were considered as HPC-7 Tet2-bound genes. Figure 4.8 showed two examples of HPC-7 *Tet2*-KO DE genes with Tet2 binding in mouse myeloid cells. Interestingly, the published endogenous mouse Tet2 ChIP-seq data also exhibited background signal at the KO cells ChIP-seq track. And this background issue is frequently observed in the identified Tet2 peaks in this study.

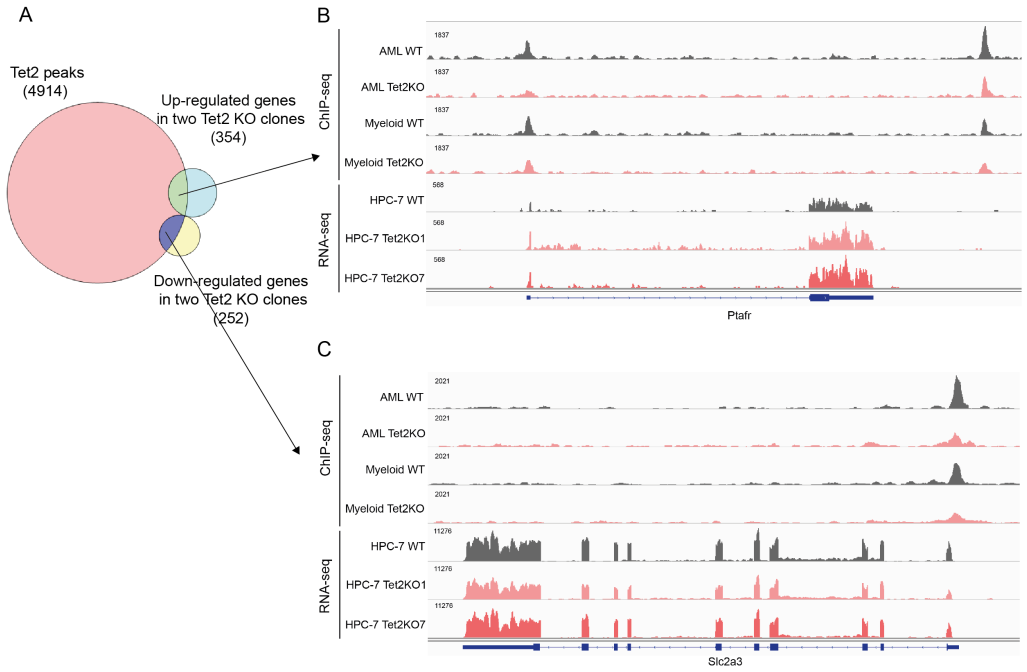


Figure 4.8: Schematic representation of the qPCR primers designed for HA-Tet2 ChIP experiments in HPC-7 cells.(A) Venn diagram shows overlap of genes significantly deregulated in two *Tet2*-KO HPC-7 clones with genes bound by Tet2 in mouse AML and myeloid cells. Tet2 ChIP-seq in mouse AML and myeloid cells were obtained from Rasmussen et al., 2019. The numbers of genes in each group is indicated. (B) Genome browser track shows a representative example of gene upregulated in *Tet2*-KO HPC-7 cells and bound by Tet2 at the gene promoter region in mouse AML and myeloid cells. (C) Genome browser track shows a representative example of gene downregulated in *Tet2*-KO HPC-7 cells and bound by Tet2 at the gene promoter region in mouse AML and myeloid cells.

Finally, I performed HA ChIP-seq experiments on 3xHA-Tet2 HPC-7 cells and prepared ChIP DNA obtained from HA-ChIP on *Tet2*-KO cells as control for antibody non-specific background. Biological triplicates were prepared for each group. Prior to sequencing, qPCR on ChIP DNA was performed to check if there was any enrichment on the putative HPC-7 Tet2 binding sites (Figure 4.9). In general, there was a consistent ChIP enrichment at the tested gene locus among biological replicates and no ChIP background was observed in the ChIP DNA from KO cells group.

I prepared the DNA library of the ChIP DNA and one input DNA from each sample group (*Tet2*-KO cells and HA-Tet2 cells). The average DNA fragment size in the prepared ChIP-seq library is around 400-450 bp, indicating the average ChIP DNA fragments is approximately 300 bp. Table 4.2 summarized the ChIP-seq read counts, read alignment

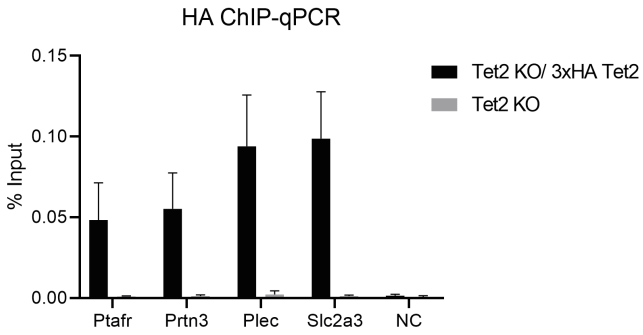


Figure 4.9: RT-PCR analysis of HA-Tet2 ChIP-seq DNA. HPC-7 *Tet2*-KO 2-7/ Tet-on 3xHA Tet2 cells were treated with 100 ng/mL dox for 24 h prior to DSG + FA crosslinking. Crosslinked and sonicated chromatin was immunoprecipitated with HA antibody. ChIP on chromatin from *Tet2*-KO HPC-7 cells was prepared as control for antibody non-specific binding. ChIP DNA from three biological replicates were analyzed by RT-qPCR with primer pairs specific for 4 putative Tet2 binding sites (Ptafr/Prtn3/Plec/Slc2a3) and 1 background region (NC).

rate and MACS2-called peak numbers. ChIP-seq reads were trimmed to remove adaptor sequence contamination, aligned to mm10 genome by Bowtie2 (Langmead and Salzberg, 2012), and ChIP-seq peaks were identified by MACS2(Zhang et al., 2008) using input sample as control. MACS2 identified around 20,000 peaks in 3xHA-Tet2 cells whereas replicates of *Tet2*-KO cells showed more variation of numbers of peaks identified by MACS2.

Table 4.2: Total read counts and alignment metrics for HA-Tet2 ChIP-seq experiments.

Sample Name	Total paired-end reads	% overall alignment rate	% uniquely mapped reads	Number of uniquely mapped reads after filtering	Numbers of peaks called MACS2 (q<0.05)
Input_Tet2 KO	55566962	94.59	70.32	31521507.5	
Input_HA Tet2	55818835	94.78	70.85	31961682	
ChIP on Tet2KO_rep1	60718114	92.09	64.32	28334258	796
ChIPon Tet2KO_rep2	66492165	90.72	65.11	32070349.5	290
ChIPon Tet2KO_rep3	55166302	93.3	68.08	29363645	179
ChIPon HATet2_rep1	76749597	92.16	68.76	39495377	19689
ChIPon HATet2_rep2	76029155	91.65	67.8	39440944.5	27140
ChIPon HATet2_rep3	66226878	93.92	70.76	36241206	21777

Tet2 binding sites (BS) were defined by enrichment over non-specific ChIP peaks identified in *Tet2*-KO cells by Diffbind analysis (Stark and Brown). Sample clustering

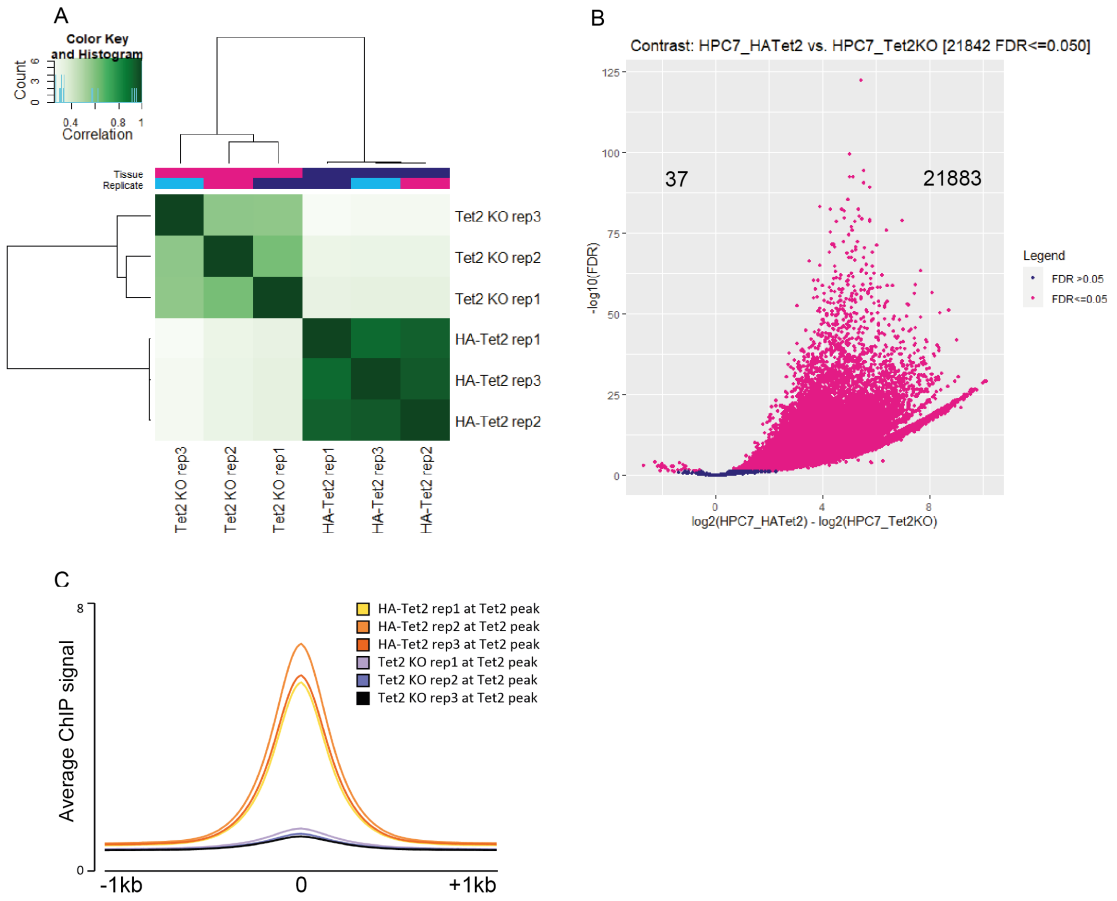


Figure 4.10: Identification of Tet2 binding sites by differential analysis of HA-ChIP in HA-Tet2 versus *Tet2*-KO HPC-7 cells.(A) Correlation heatmap generated by Diffbind (Stark and Brown) comparing the ChIP-seq signal from biological replicates at identified Tet2 binding sites. (B) Volcano plot showing the false discovery rate (FDR) and the fold enrichment values of ChIP-seq signal analyzed by differential binding analysis in HA-Tet2 versus *Tet2*-KO cells. ChIP-seq peaks showing significant differential binding (FDR<0.05) and numbers of region in significant changes were indicated in red. (C) Metaplot showing the average HA-Tet2 ChIP-seq signal in a 2 Kb region surrounding the center of Tet2 peaks (n=21883). Coverage of each ChIP-seq samples was normalized to sequencing depth.

analysis showed replicates from each sample group were clustered together (Figure 4.10A). 21883 peaks exhibited significantly enrichment (FDR<0.05) in ChIP on 3xHA-Tet2 cells versus ChIP on KO cells (Figure 4.10B). The 21883 peaks were defined as true Tet2 BS in HPC-7 cells . HA-Tet2 ChIP-seq signal was enriched in the center of the identified Tet2 BS (n=21883) and no ChIP enrichment was observed in the KO cells control group (Figure 4.10C).

4.2.3 Characterization of HPC-7 Tet2 binding sites

To understand the distribution characteristics of the Tet2 BS, I annotated the Tet2 BS (n=21883) to the nearest gene and explored the genomic feature, histone modification distribution, functional pathway enrichment and transcription factor binding motif enrichment on the Tet2 BS. A comparison between our Tet2 BS in HPC-7 cells and the recently published mouse myeloid cells Tet2 peaks will be introduced in this section.

Genomic feature distribution of Tet2 binding sites

Around 37% of the Tet2 BS were inside gene promoter and almost 30% of the Tet2 BS located within 1 Kb of the gene promoter (Figure 4.11B). The gene promoter was defined as +/- 3 Kb away from the gene TSS (transcription start site). Around 31 % of Tet2 BS was localized in intron regions (1st intron + other intron) and 25 % of the Tet2 BS overlap with distal intergenic regions. Taken together, Tet2 showed binding at the gene promoters and distal regulatory regions in HPC-7 genome.

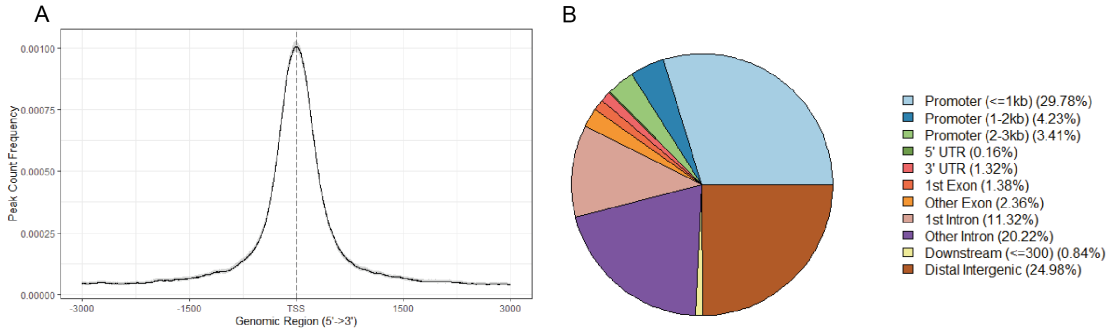


Figure 4.11: Genomic feature distribution of Tet2 binding sites in HPC-7 cells (A) Frequency plot showing the distance of HA-Tet2 ChIP-seq peaks (n=21883) to the nearest transcription start site (TSS). (B) Pie chart showing overlap of HA-Tet2 ChIP-seq peaks with genomic annotations. Downstream is defined as the region downstream of gene within 3 Kb.

Next, I explored the histone modification levels on the Tet2 BS using the publicly available HPC-7 H3K4me3 (Xu et al., 2018), H3K4me1, H3K27ac (Org et al., 2015), H3K27me3 and H3K36me3 (Wilson et al., 2016) ChIP-seq data. Histone modifications have been widely used to identify cis-regulatory elements including promoters and enhancers (Ernst and Kellis, 2010). Trimethylation of the 4th lysine of histone 3 (H3K4me3) and acetylation of histone H3 Lys27 (H3K27ac) have been associated with transcriptionally

active gene promoters (Calo and Wysocka, 2013; Heintzman et al., 2009). H3K27me3 is known as a transcription repressive mark, whereas, H3K36me3 is enriched in the gene body of transcribed genes (Ringrose and Paro, 2004; Vakoc et al., 2006). H3K4me1 shows enrichment at enhancer regions. (Barski et al., 2007; Heintzman et al., 2009; Wang et al., 2008). Enhancers could be classified into three categories based on the histone modification marks combination: 1. Active enhancer: have both H3K4me1 and H3K27ac, and can increase transcription of its regulated gene; 2. Primed enhancers: are marked with H3K4me1 alone and require additional TF binding to activate its enhancer activity; 3. Poised enhancers: are similar to primed enhancer and contain additionally H3K27me3, a repressive mark (Calo and Wysocka, 2013; Creyghton et al., 2010; Ernst and Kellis, 2010; Heinz et al., 2015).

Tet2 BS displayed enrichment of H3K4me3/H3K27ac/H3K4me1 ChIP-seq signals at the center of Tet2 peaks (Figure 4.12A), whereas, no enrichment of H3K27me3 and H3K36me3 was observed at Tet2 BS. Heatmap analysis revealed the co-localization of H3K4me3 and H3K27ac, and depletion of H3K4me1 at the TSS-proximal Tet2 BS suggesting Tet2 binds to genes with high transcription activity (Figure 4.12B). H3K4me1 located predominantly at Tet2 BS distant from gene TSS, and co-localized with H3K27ac, suggesting that Tet2 binds to the distal enhancer elements in the HPC-7 genome. This support the current focus of Tet2 function in enhancer regions (Hon et al., 2014; Lio et al., 2019a; Lu et al., 2014; Rasmussen et al., 2015, 2019; Wang et al., 2018). H3K27me3 showed a tiny enrichment at the TSS-proximal Tet2 BS which suggests a possibility of Tet2 function in PRC2 complex-mediated gene repression. Tet2 BS is depleted of H3K36me3 signal.

Rasmussen and colleagues reported that Tet2 genomic binding displayed a cell type-specific manner since little overlap of Tet2 binding sites was found between ChIP in mouse ESCs and ChIP in mouse hematopoietic cells. Furthermore, Tet2 promoter-distal (1.5 kb/+500 bp from TSS) BS contained different TF binding motifs enrichment. Here, I compared our identified HPC-7 Tet2 BS to the published Tet2 peaks identified in mouse hematopoietic cells. In general, HPC-7 Tet2 BS showed some overlap with the published

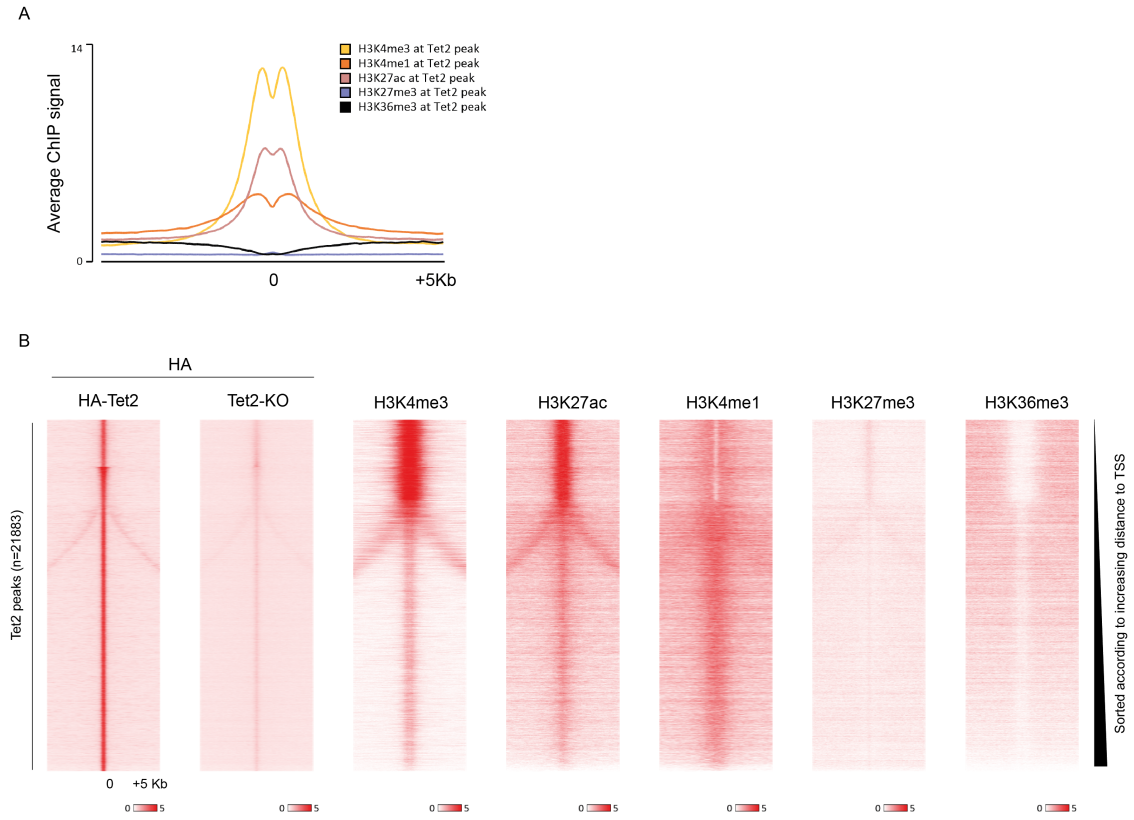


Figure 4.12: Histone modifications distribution at Tet2 binding sites in HPC-7 cells. (A) Metaplot showing the average histone modification ChIP-seq signal in a 5 Kb region surrounding the center of Tet2 peaks (n=21883) (B) Heatmap showing the publicly available HPC-7 histone modification ChIP-seq signal in a 5 Kb region surrounding the center of HA-Tet2 peaks (n=21883). The vertical axis includes all HA-Tet2 binding sites sorted by increasing distance to the TSS of the nearest gene. H3K4me3 ChIP-seq was downloaded from Xu et al., 2018. H3K27ac and H3K4me1 ChIP-seq were downloaded from Org et al., 2015. H3K27me3 and H3K36me3 ChIP-seq were downloaded from Wilson et al., 2016.

hematopoietic cell Tet2 peaks. $\sim$ 35% of HPC-7 Tet2 BS were also found in published mouse myeloid cell Tet2 BS (Figure 4.13A). And all three Tet2 BS data suggest the binding of Tet2 at the distal regulatory elements (Figure 4.13B).

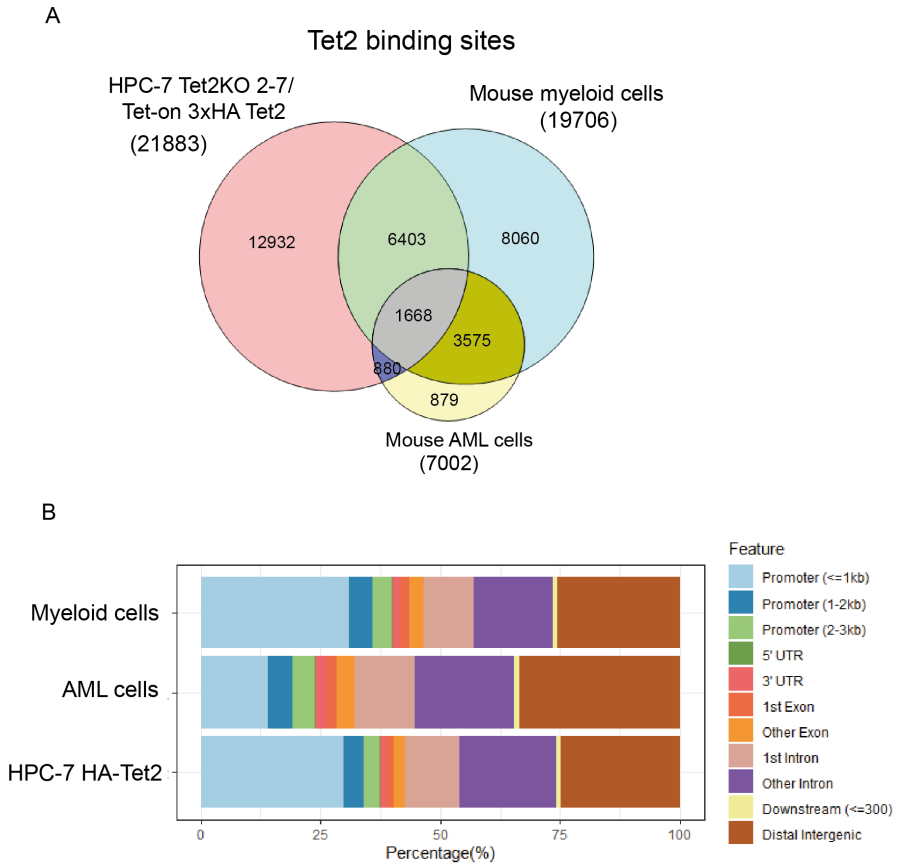


Figure 4.13: Comparison to the publicly available Tet2 ChIP-seq data in mouse myeloid cells and acute myeloid leukemia cells. (A) Venn diagram showing the overlap of HA-Tet2 ChIP-seq peaks in HPC-7 cells with Tet2 ChIP-seq peaks in mouse myeloid cells and acute leukemia cells. Tet2 ChIP-seq data in mouse myeloid cells and acute leukemia cells were downloaded from Rasmussen et al., 2019. (B) Percentage of Tet2 peaks in mouse myeloid cells, AML cells and HPC-7 cells that localize with different genomic features. Downstream is defined as the region downstream of gene within 3 Kb.

Functional enrichment on Tet2 binding sites

Genes associated with HPC-7 Tet2 BS and genes associated with published mouse myeloid cell Tet2 BS share similarities in the pathway enrichment analysis (Figure 4.14). In both cells, Tet2-bound genes were mainly associated with pathways related to hematopoietic processes, such as cytokine signalling (e.g. Kit, Cxcr4, Csf1), platelet activation, signaling and aggregation (e.g. Itga2b, Gata1), interleukins signaling (Il1b, Nod2, Egr1), suggesting the role of Tet2 in hematopoiesis.

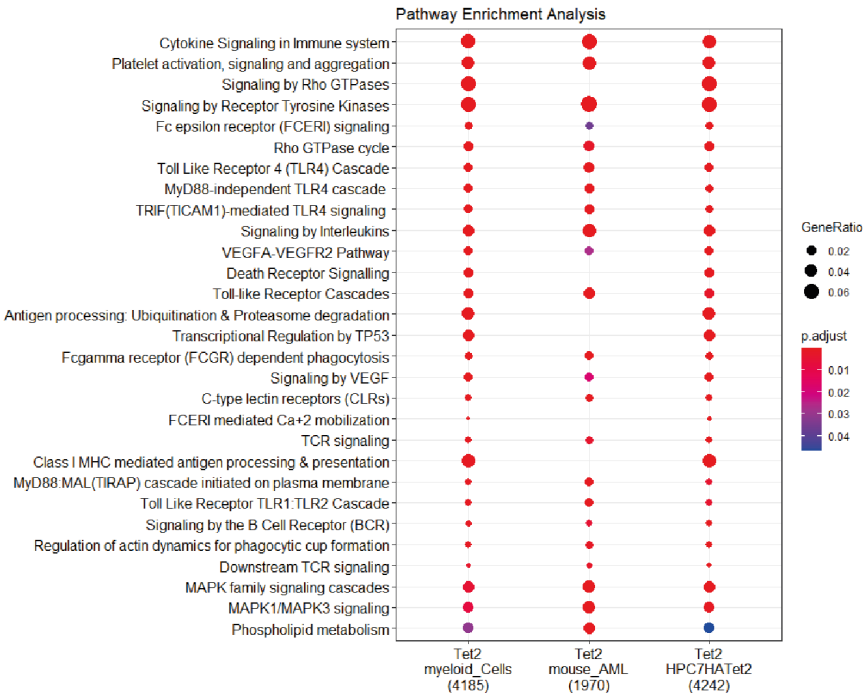


Figure 4.14: Pathway enrichment comparison of HA-Tet2 ChIP-seq with publicly available Tet2 ChIP-seq data. The y-axis represents the name of pathway, the x-axis represents each Tet2 ChIP-seq data set. Number in parentheses represents the number of gene clusters in each data set. Dot size indicates the gene ratio. Gene ratio is the number of genes in the Tet2 ChIP-seq peak overlapping with indicated pathway gene set divided by the total number of genes in the Tet2 ChIP-seq peaks. Color indicates the q-value. P.adjust is p value adjusted by Benjamini Hochberg method.

Motif enrichment on Tet2 binding sites

The lack of a known DNA-binding domain in Tet2 protein leads to the hypothesis that Tet2 might interact with DNA-binding partners, such as transcription factors, to be recruited to its targeted sites. Thus, I performed MEME-ChIP analysis ((Machanick and Bailey, 2011)) to discover transcription factor binding sites enriched in HPC-7 Tet2-bound regions. A previous study revealed the cell-type specific binding pattern of Tet2 where Tet2 BS showed enrichment of ESC master TFs binding, including Sox2, Klf4, Oct4, and Nanog, in ES cells, and the top enriched motifs in mouse myeloid cells were Erg, Runx1, Cebpa and Gata1 (Rasmussen et al., 2019). In HPC-7 cells, I observed the similar TF motifs enrichment as what was shown in mouse myeloid cells, Tet2 BS was enriched with key hematopoietic TF BS, such as Etv6, Runx1, Fli1, and Gata1 in HPC-7 cell (Figure 4.15).

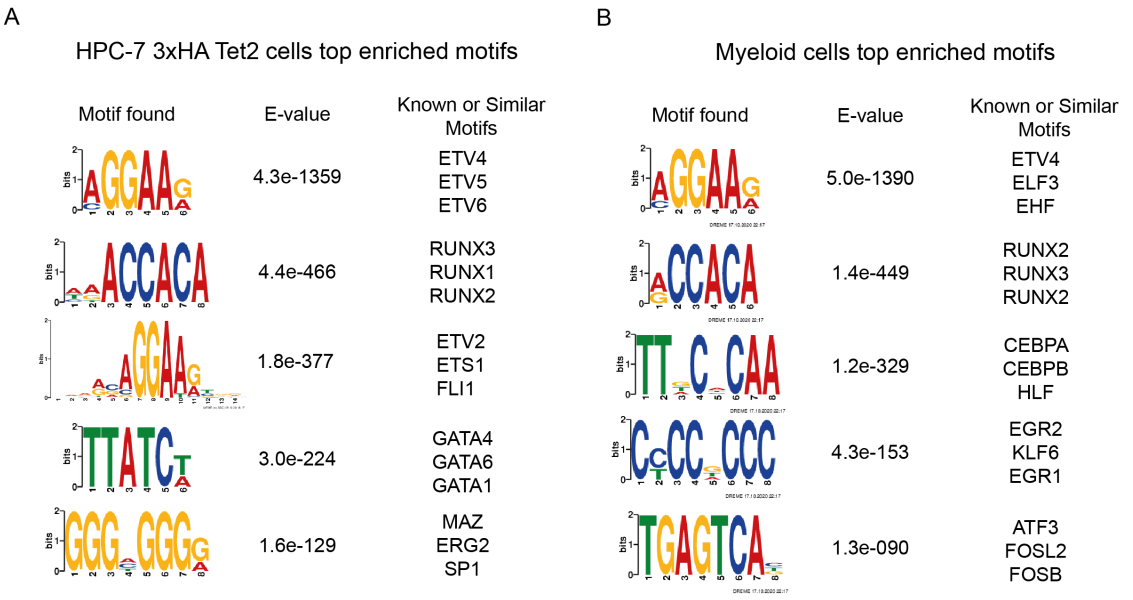


Figure 4.15: Motif enrichment comparison of HA-Tet2 ChIP-seq with publicly available Tet2 ChIP-seq data. List of top 5 enriched transcription factor binding motifs identified in Tet2 binding regions in HPC-7 cells (A) and mouse myeloid cells (B).

4.2.4 The relationship between Tet2 genomic occupancy and gene expression

To understand the function of Tet2 binding in gene regulation, I checked the overlap between Tet2-bound genes and differentially expressed genes upon Tet2 depletion (Figure 4.16). Tet2-bound genes were defined by gene annotation on Tet2 binding sites using the ChIPseeker package. In general, only a small fraction of Tet2 binding events will affect its associated gene expression since ∼1000 out of 9714 Tet2-bound genes showed differential expression in *Tet2*-KO cells versus WT HPC-7 cells (Figure 4.16A and B). Tet2-bound regions correlated to both up-regulated and down-regulated gene identified in *Tet2*-KO cells, suggesting a dual function of Tet2 binding in gene activation and repression. However, more up-regulated genes were found overlapping with Tet2-bound genes. Next, I focused on Tet2 function at gene promoter region since promoter activity is more well-studied and associated with gene expression regulation. ∼65% Tet2-bound genes displayed Tet2 binding at its promoter region (Distance to TSS <3Kb). Tet2 binding at the gene promoter were associated with both gene repression (Figure 4.16C) and activating

(Figure 4.16D). Taken together, in HPC-7 cells, Tet2 could positively and negatively regulate its associated gene expression by directly binding to the targeted genes.

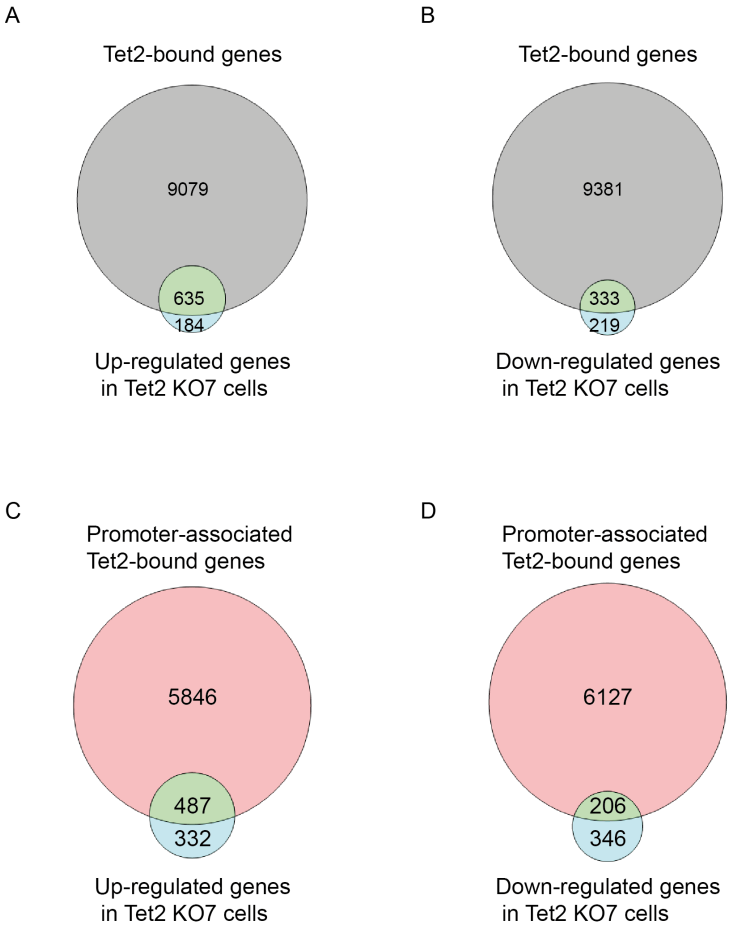


Figure 4.16: Tet2 binding is associated with both gene activating and repressing function. (A) Venn diagram shows overlap of Tet2-bound genes with genes upregulated in *Tet2*-KO clone 7. (B) Venn diagram shows overlap of Tet2-bound genes with genes downregulated in *Tet2*-KO clone 7. (C) Venn diagram shows overlap of genes with Tet2 binding at the promoter region with genes upregulated in *Tet2*-KO clone 7. Gene promoter was defined as +/- 3 Kb of the TSS. (D)) Venn diagram shows overlap of genes with Tet2 binding at the promoter region with genes downregulated in *Tet2*-KO clone 7. Gene promoter was defined as +/- 3 Kb of the TSS.

To explore the possible mechanism for Tet2 function in gene regulation, I explored the TF binding motifs enriched in Tet2 BS associated with gene activating (Figure 4.17A) and gene repression (Figure 4.17B), to check if Tet2 get involved in gene regulation through different DNA-binding partners. However, no obvious difference was observed in TF motif enrichment analysis of Tet2 BS linked to gene activation and repression.

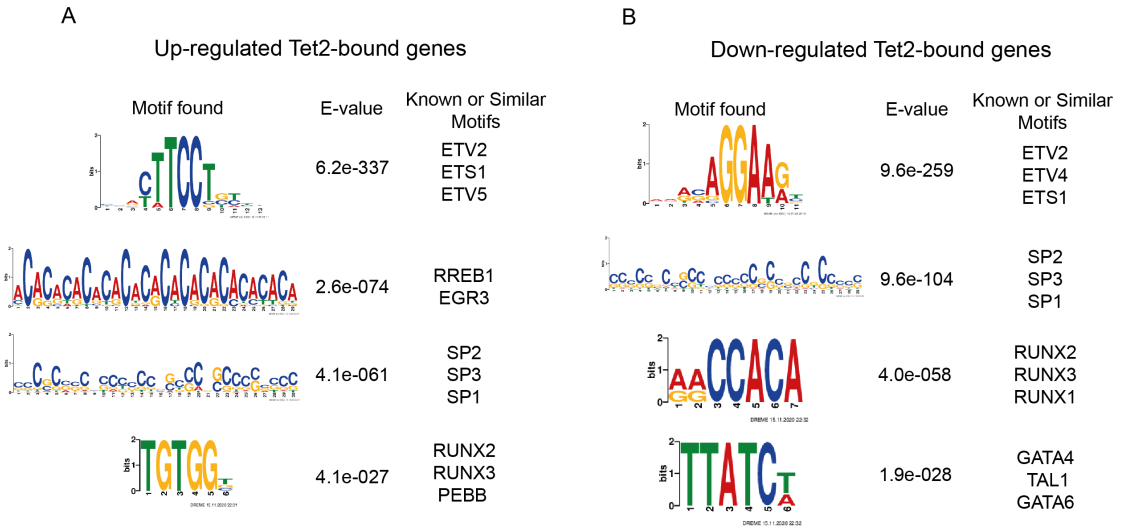


Figure 4.17: Motif enrichment in Tet2-regulated genes. List of top 3 enriched transcription factor binding motifs identified in genes upregulated by Tet2 binding (A) and genes downregulated by Tet2 binding (B).

4.2.5 Differential DNA methylation analysis in *Tet2*-KO HPC-7 cells

Tet2 belongs to the Tet protein family which is responsible for the 5mC oxidation. Tet-mediated 5mC oxidation facilitates an active DNA demethylation process. Tet2 function in hematopoiesis is closely related to its demethylase activity since altered DNA modification landscapes have been observed widely in genetically modified mouse models and blood cancer patients with TET2 mutation or with de-regulated TET2 enzymatic activity (Cimmino et al., 2017). Moreover, TET2 catalytic mutant itself could cause mice to develop myeloid malignancies (Ito et al., 2019). In order to understand the contribution of Tet2-mediated DNA demethylation in regulating the hematopoietic-related gene expression, we performed the TAPS-β experiment to map the 5mC distribution and to quantify the 5mC level in *Tet2*-KO cells and WT HPC-7 cells. In addition, we also mapped the 5mC distribution in the constructed HPC-7 PiggyBac cell line which has no endogenous Tet2 expression but expresses an ectopic 3xHA-Tet2 protein. Briefly, before harvest for genomic DNA extraction, HPC-7 PiggyBac cells were treated with 100 ng/mL dox for 24 hr to induce the expression of 3xHA Tet2 protein. Increased 5hmC in the Dox-treated cells was confirmed by LC/MS-MS quantification (Figure 5.15).

In general, there is no global gain or loss of DNA methylation across all the CpG sites in

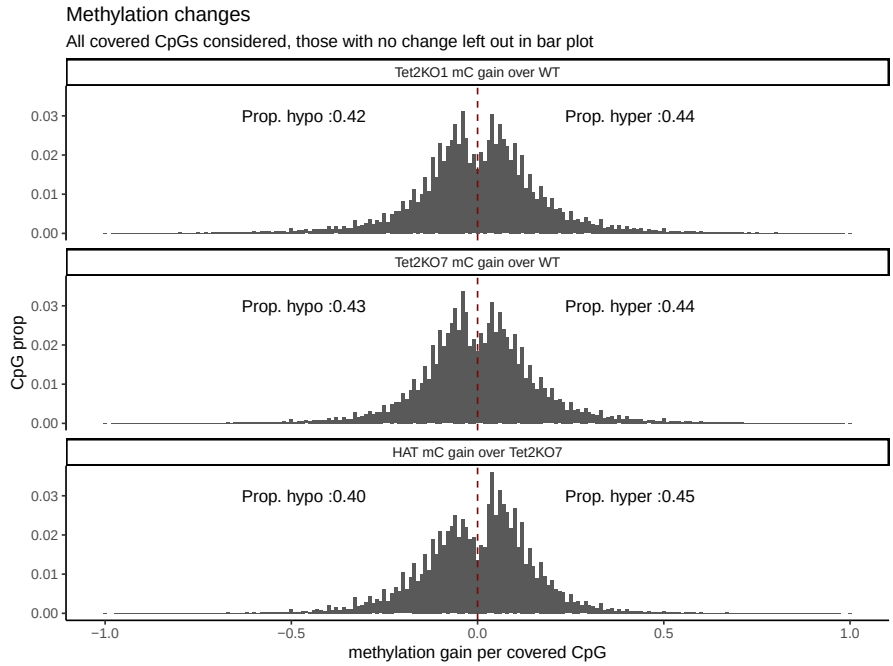


Figure 4.18: Tet2 depletion in HPC-7 leads to both DNA hypermethylation and hypomethylation changes. For each CpG on the 5' reads of samples in each comparison group, the difference of 5mC level was calculated and plotted as the x axis. The proportion of CpGs falling into each bin of change was shown in the y axis. If the x axis value is greater than 0, this indicates a methylation gain. The proportions of hyper(5mC change >0) /hypo(5mC change <0) CpGs were indicated in the plot.

both *Tet2*-KO clones compared with the WT cells (Figure 4.18). ∼44% of total CpG in *Tet2*-KO cells had increased 5mC level and ∼42-43% of total CpG exhibited decreased 5mC. ∼13-14% of total CpG showed no change of DNA methylation upon Tet2 deficiency. Interestingly, in the HA-Tet2 cells, we observed there was slightly more increase of DNA methylation than loss of DNA methylation. Next, to examine the methylation changes on different genomic features, we calculated the methylation level on the genomic features and compared the methylation level difference between *Tet2*-KO cells and WT cells. The comparison between HA-Tet2 group and *Tet2*-KO group was used to discover the DNA demethylation events restored by the re-expression of Tet2 protein. HPC-7 cell enhancers were defined by regions enriched with H3K4me1 and have no signal of H3K4me3. The results revealed that enhancers underwent the biggest gain of methylation in *Tet2*-KO cells (Figure 4.19). More importantly, restoration of the 3xHA Tet2 expression in *Tet2*-KO cells reversed the increase of 5mC at enhancer region in KO cells. In addition to enhancers, DNA hypermethylation was also found in H3K4me3-promoter regions in *Tet2*-KO cells

but the changes were not as evident as those observed in enhancers. The expression of HA-Tet2 could also reverse the gain of methylation phenotype in *Tet2*-KO cells. Tet2 depletion causes no significant changes of DNA methylation at CpG island but leads to a slight increase of methylation at CpG island shore which are regions 2 kb upstream and downstream of CpG island. In summary, our data suggests that Tet2-mediated DNA demethylation predominantly affects the enhancer regions.

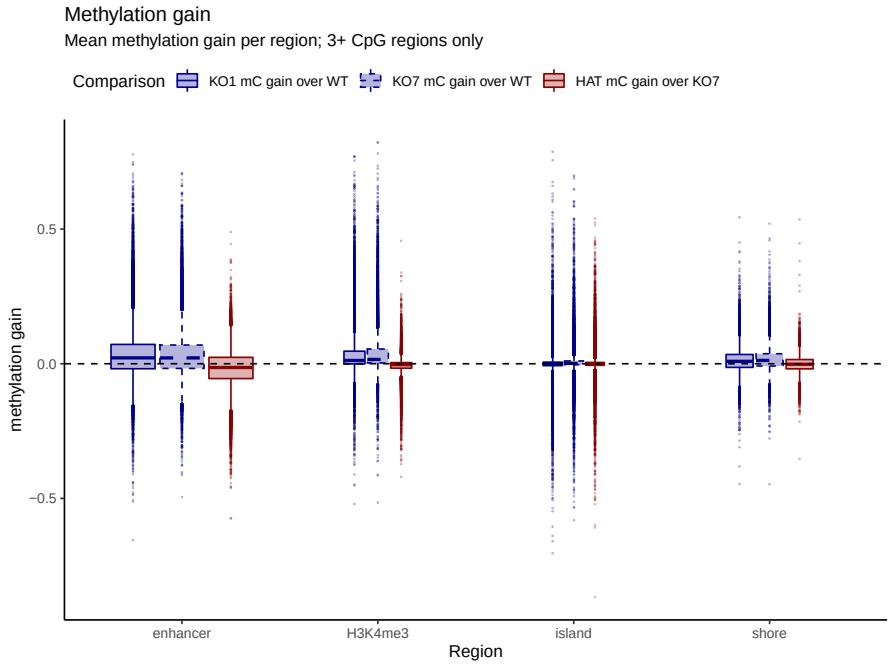


Figure 4.19: Tet2 is responsible for the DNA demethylation at enhancer regions. For each region type, the mean modification change across all CpGs in the region (only those with 3+ covered CpGs considered) was calculated using bedtools map (v2.25.0). Enhancers were defined as regions positive for H3K4me1 and have no H3K4me3 signal. Promoters were defined by the transcription start sites of the mm10 genes and HPC-7 H3K4me3 ChIP-seq data. H3K4me3+ promoters were defined as the transcription start sites of the mm10 genes and HPC-7 H3K4me3 ChIP-seq data. CpG islands, shelf, and shore annotation were directly extracted from the annotatr package (v1.16.0) (Cavalcante and Sartor, 2017).

4.2.6 The relationship between Tet2-involved DNA demethylation and Tet2 binding

To examine whether Tet2 binding is associated with its DNA demethylase activity at its binding site, we calculated the DNA methylation changes between *Tet2*-KO cells and WT HPC-7 cells on the Tet2 binding sites (Figure 4.20). In order to confirm the demethylation events directly regulated by Tet2, we also looked at the methylation level on Tet2 binding

sites between HA-Tet2 cells and *Tet2*-KO cells. We observed more increase of DNA methylation than decrease of methylation on Tet2 binding sites upon Tet2 depletion. Furthermore, the re-expression of Tet2 in the *Tet2*-KO cells leads to the decrease of DNA methylation on Tet2 binding sites. These results suggest that Tet2 gets involved in DNA demethylation at some of the Tet2 binding sites.

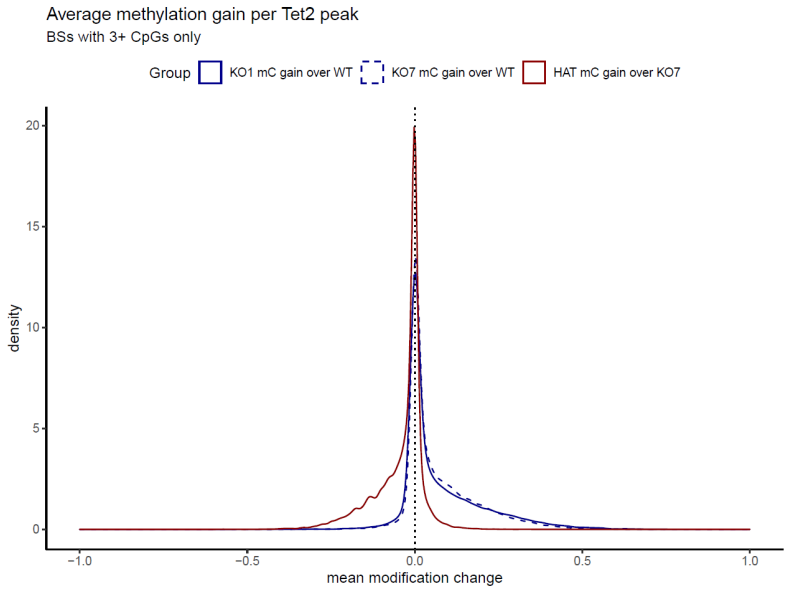


Figure 4.20: Tet2 is responsible for the DNA demethylation at its binding sites. For each Tet2 binding sites, the mean modification change across all CpGs in the region (only those with 3+ covered CpGs considered) was calculated using bedtools map (v2.25.0). The methylation changes of all Tet2 binding sites were plotted. The x axis represents the methylation level changes. The y axis represents the number of Tet2 binding sites showing the indicated methylation changes on the x axis.

To assess the DNA demethylation activity of Tet2 on its binding sites across different genomic features, we grouped the Tet2 binding sites based on their overlap with enhancer, H3K4me3+promoter, CpG island. And we examined the DNA methylation changes caused by Tet2 depletion and Tet2 restoration (Figure 4.21). First of all, Tet2 depletion leads to an increase of DNA methylation on all Tet2 binding sites, and the restoration of Tet2 reverses the gain of methylation on Tet2 binding sites. The gain of DNA methylation on Tet2 binding sites in *Tet2*-KO cells is more significant on enhancer-associated Tet2 binding events, whereas Tet2 binding at H3K4me3+promoter and CpG island has no correlation with DNA demethylation activity. In conclusion, the data suggests that Tet2 binding at enhancer region is responsible for the DNA demethylation activity.

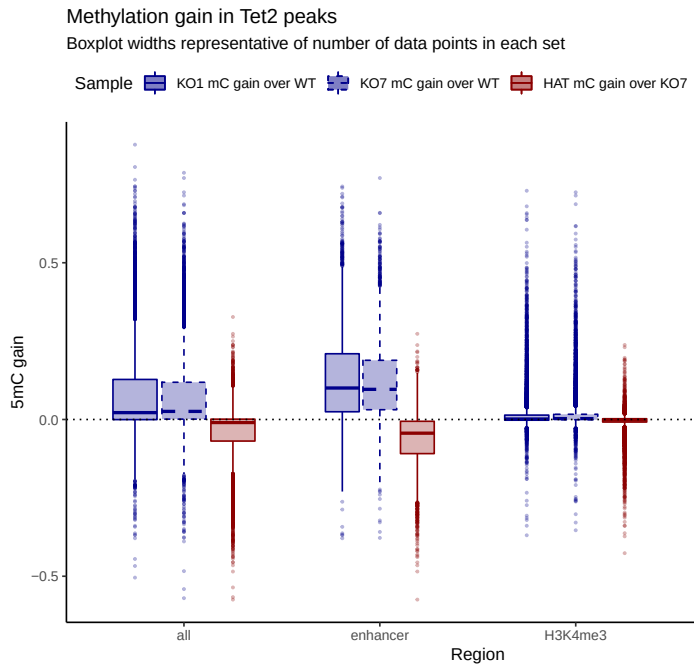


Figure 4.21: Tet2 is responsible for the DNA demethylation at its binding sites at enhancer region. For each Tet2 binding sites, the mean modification change across all CpGs in the region (only those with 3+ covered CpGs considered) was calculated using bedtools map (v2.25.0). Tet2 binding sites were grouped according to overlaps with enhancers, active promoters (H3K4me3), and CpG islands.

4.2.7 The relationship between Tet2-involved DNA demethylation and gene expression

To determine the transcriptional effect of Tet2-mediated DNA demethylation, we quantified the 5mC level of Tet2-bound genes showing differential expression upon Tet2 deficiency. In total, 487 out of 819 significantly up-regulated genes exhibited Tet2 binding at gene promoter region whereas 206 out of 552 genes significantly down-regulated genes contained Tet2 binding at gene promoter. Interestingly, both up-regulated and down-regulated genes were associated with increased 5mC (Figure 4.22), and down-regulated genes showed more gain of 5mC compared with up-regulated genes, suggesting that Tet2 might be involved in gene activation through its DNA demethylaton activity. And the contribution of Tet2-mediated DNA demethylation to the gene repression remain unknown.

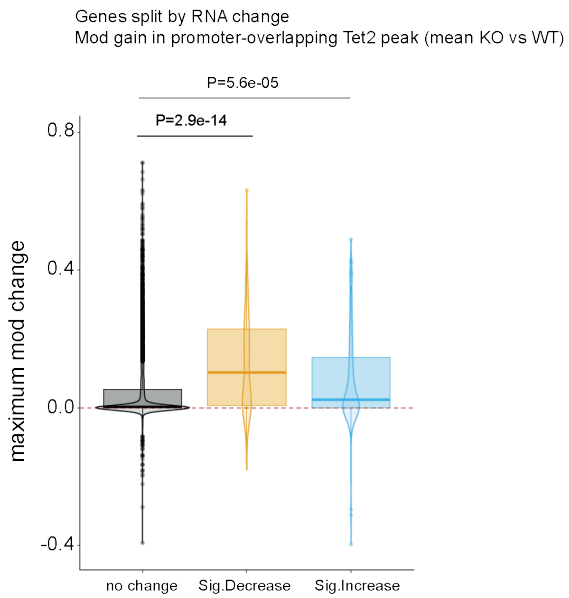
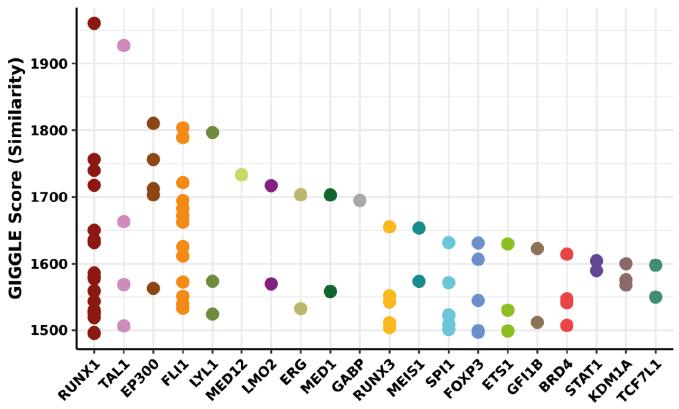


Figure 4.22: Depletion of *Tet2* leads to a DNA methylation change on the Tet2-targeted genes. For each gene, the maximum modification change across all the Tet2 binding sites associated with the gene was calculated using bedtools map (v2.25.0). And gene was further classified based on the RNA-Seq signal. Sig.Increase group was defined as genes with log fold change > 0 and p-adj < 0.05 in both *Tet2*-KO clones compared to WT cells. Sig.Decrease group was defined as genes with log fold change < 0 and p-adj < 0.05 in both *Tet2*-KO clones compared to WT cells. All other genes fall into the No. Change group.

4.2.8 Co-localization of Tet2 and transcription factors in HPC-7 cells

Motif enrichment analysis of Tet2 BS discovered the binding motifs of many essential hematopoietic transcription factors such as Etv6, Runx1, Fli1 (Figure 4.15A), which indicate a potential co-operation between Tet2 and transcription factors in gene regulation. To further identify Tet2 genomic co-factors, I uploaded the HPC-Tet2 BS (n=21883) to the Cistrome Data Browser (Zheng et al., 2019) and searched for the publicly available mouse transcription factor ChIP-seq to discover the mouse factors which will have the most significant peak overlap with the Tet2 BS identified in HPC-7 cells (Figure 4.23A). The GIGGLE score is used to measure the similarities between the user uploaded peak sets and Cistrome samples (Layer et al., 2018; Zheng et al., 2019). The factor with higher GIGGLE score shares more overlap with the input peak set. The top factors showing high similarity with Tet2 peaks sets were ChIP-seq on hematopoieitc TFs in HPC-7 cells and hematopoietic-relevant cell lines, suggesting a specfic role of Tet2 in hematopoiesis. Among the enriched factors, Runx1 has been found to physically interact with Tet2 in

A



B

GSM_ID	Factor	Biosource	GIGGLE_score
GSM552241	RUNX1	HPC-7;Haematopoietic Progenitor Cell	1960.36
GSM552242	TAL1	HPC-7;Haematopoietic Progenitor Cell	1926.92
GSM1973452	EP300	HPC-7;Hematopoietic Cell	1810.32
GSM552233	FLI1	HPC-7;Haematopoietic Progenitor Cell	1803.84
GSM1708654	LYL1	416B;Myeloid Progenitor Cell;Blood	1796.69
GSM1692850	FLI1	Hematopoietic progenitors	1789.62
GSM989023	FLI1	Hematopoietic Cell	1788.99
GSM1708657	RUNX1	416B;Myeloid Progenitor Cell;Blood	1756.38
GSM1973458	EP300	HPC-7;Hematopoietic Cell	1756.29
GSM1692856	RUNX1	Hematopoietic progenitors	1739.89
GSM1973455	MED12	HPC-7;Hematopoietic Cell	1733.31
GSM654873	FLI1	T Lymphocyte;Blood	1721.58
GSM1103355	RUNX1	Megakaryocyte	1717.71
GSM552237	LMO2	HPC-7;Haematopoietic Progenitor Cell	1717.17
GSM1973456	EP300	HPC-7;Hematopoietic Cell	1712.48
GSM1708650	ERG	416B;Myeloid Progenitor Cell;Blood	1703.42
GSM1973457	EP300	HPC-7;Hematopoietic Cell	1703.32
GSM1184314	MED1	Thymocyte;Blood	1703.09
GSM2719134	GABP	T Lymphocyte	1694.85
GSM1692860	FLI1	Macrophage	1694.41

Figure 4.23: Overlap of transcription factors binding at Tet2 occupancy sites. (A) Dot plot showing the top 20 transcription factors in the Cistrome Data Browser that have significant binding overlaps with HPC-7 HA-Tet2 peak set. Y axis represents the GIGGLE score which indicates the similarity between samples and HA-Tet2 peak set. X axis represents different factors. The information of the top 20 enriched transcription factors ChIP-seq data is listed in (B). List is ranked by GIGGLE scores. Analysis was performed using Cistrome DB Toolkits (Zheng et al., 2019).

mouse erythroleukemia (MEL) cells (Chu et al., 2018). PU.1 is also reported to interact with Tet2 (de la Rica et al., 2013). The observed overlap of Tet2 BS with Runx/PU.1 binding sites might indicate a potential interaction between Tet2 and Runx1/PU.1 in

HPC-7 cells.

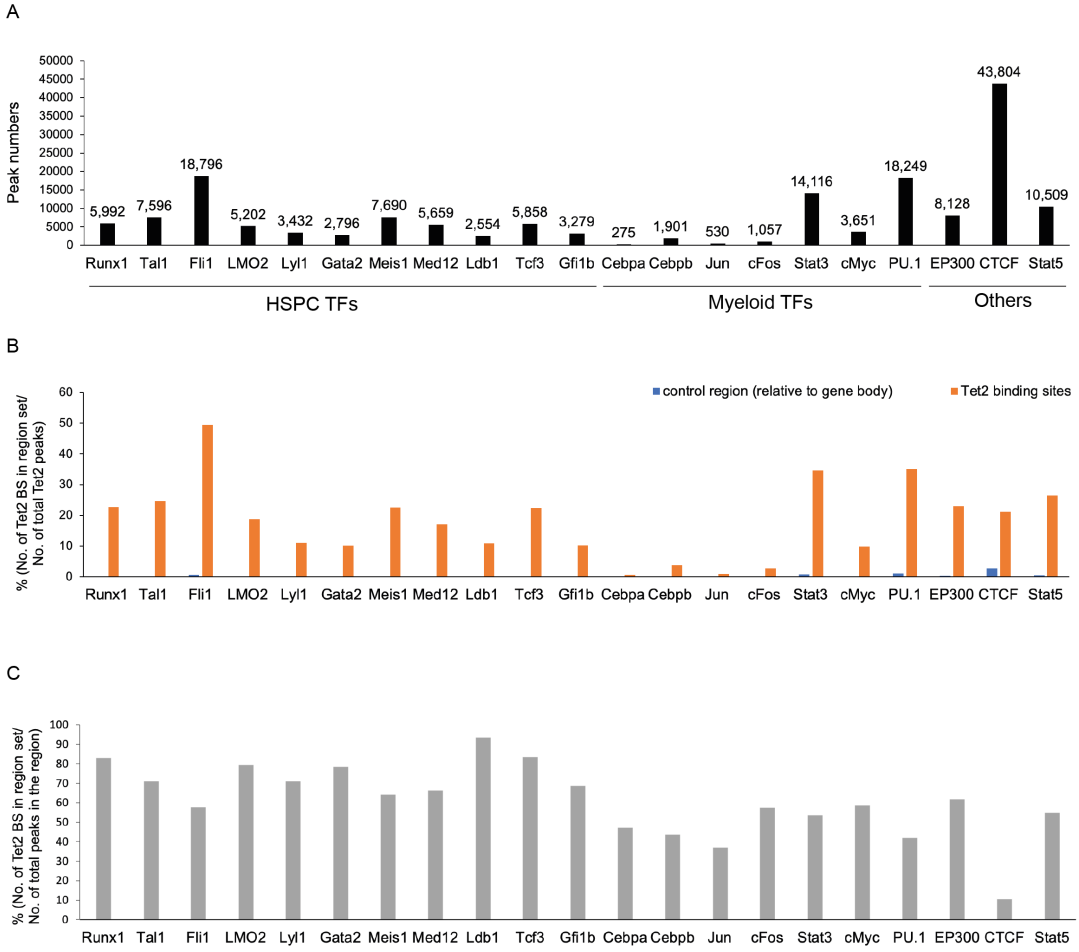


Figure 4.24: Tet2 co-localizes with multiple transcription factors in HPC-7 cells. (A) Histogram showing the peak numbers of publicly available HPC-7 transcription factors (TFs) ChIP-seq data. (B) Histogram showing overlap of HA-Tet2 ChIP-seq binding sites (n=21883) with publicly available HPC-7 TFs binding sites. Random matched control regions (control) was generated with same size, orientation, distance relative to gene bodies by Easeq (Lerdrup et al., 2016). Data shown was determined by dividing the numbers of HA-Tet2 peaks (n=21883) overlapping with the given region by the total numbers of HA-Tet2 peaks (n=21883). (C) Histogram showing overlap of individual TF binding site with Tet2 binding sites. Data shown was determined by dividing the numbers of indicated TF peaks overlapping with the Tet2 binding regions by the total numbers of TF peaks. TF ChIP-seq data was downloaded from Wilson et al., 2016.

Then I focused on the co-localization of Tet2 and hematopoietic TFs in HPC-7 cells by utilizing the publicly available HPC-7 TF ChIP-seq data (Wilson et al., 2016). Tet2 peaks overlapped with the binding regions of a variety of TFs including both HSPC-specific and myeloid-lineage factors (Figure 4.24B), and Tet2 BS shared the most overlap with Fli1 (Friend leukemia integration 1 transcription factor) ChIP-seq peaks. Random matched

control regions (control) was generated with same size, orientation, distance relative to gene bodies based on Tet2 BS by Easeq (Lerdrup et al., 2016).

Heatmap revealed the co-localization of Tet2 and TF at different genomic distribution (Figure 4.25). Runx1, Fli1 and Gata2 prefer to locate at the TSS-distal Tet2 BS suggesting a co-operation of these TF and Tet2 at distal regulatory elements, whereas, Tal1 and PU.1 showed genomic binding throughout all Tet2 BS, and exhibited a slightly more enrichment in TSS-proximal Tet2 peaks, suggesting a role of Tal1/PU.1 at Tet2-bound gene promoter region. Collectively, Tet2 binding showed overlap with many essential hematopoietic transcription factors, such as Runx1, Fli1, Gata2, PU.1 and the co-location of Tet2 with TFs showed a TF-dependent manner, where Tet2 and Runx1/Fli1/Gata2 were more located in enhancers. Tal1 and PU.1 co-localized with Tet2 more in gene promoter.

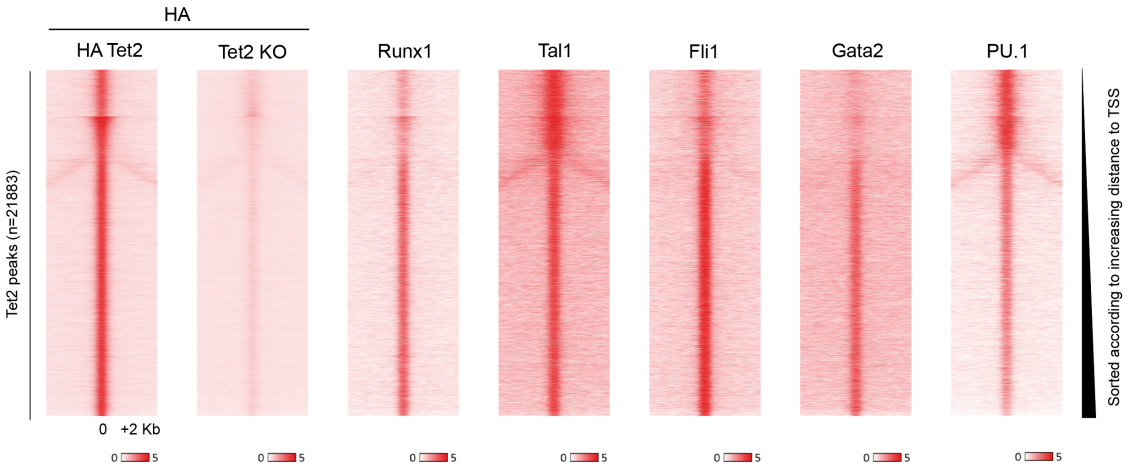


Figure 4.25: Genomic features of Tet2-TF co-bound regions. (A) Heatmap showing the publicly available HPC-7 TF ChIP-seq signal in a 2 Kb region surrounding the center of HA-Tet2 peaks (n=21883). The vertical axis includes all HA-Tet2 binding sites sorted by increasing distance to the TSS of the nearest gene. TF ChIP-seq was downloaded from Wilson et al., 2016.

We quantified the DNA methylation level on Tet2-transcription factor co-binding sites upon Tet2 depletion and restoration. Tet2 depletion caused increased DNA methylation at Tet2-Fli1 co-binding sites and Tet2-Runx1 co-binding sites. More importantly, the restoration of Tet2 expression could reverse the increased DNA methylation at the Tet2-TF co-binding sites.

RNA-seq showed some myeloid-specific genes, e.g. *csf1r*, *Cebpa*, *Cebpe*, were down-regulated upon Tet2 deficiency (Figure 4.4). Tet2 was found directly bound to the known

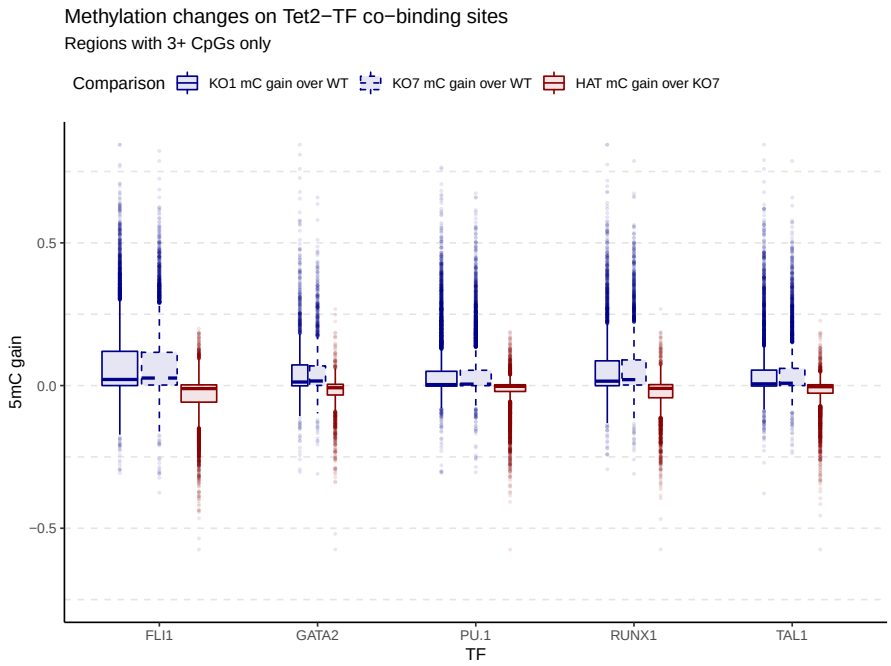


Figure 4.26: DNA methylation changes at transcription factor binding sites. The mean modification change on individual transcription factor binding sites were calculated.

enhancer of the de-regulated myeloid-specific genes and affect the DNA demethylation at the enhancer region (Figure 4.27, 4.28, 4.29). The expression of *csf1R* (c-fms) gene is dependent on the Ets family transcription factor PU.1 which binds to the promoter and the c-fms intronic regulatory element (FIRE enhancer) of the c-fms gene (Krysinska et al., 2007). c-fms promoter is primed by PU.1 binding in progenitor cells stage. In differentiated myeloid cells, the promoter and FIRE enhancer is fully bound by PU.1 which promotes the induction and binding of the secondary transcription factors Egr2, Cebpb, Runx1, leading the high expression of *csf1R*. In HPC-7 cells, Tet2 co-localized with PU.1, Runx1, and Cebpb at FIRE enhancer region of *csf1R* gene and FIRE was hypermethylated in *Tet2*-KO HPC-7 cells, suggesting a role of Tet2-mediated DNA demethylation at enhancer in gene regulation.

Another example showing the effect of Tet2 binding in gene regulation occurred at the *Cebpa* locus. Cebpa is preferentially expressed in myeloid cells and is important for the myeloid lineage differentiation. *Cebpa* gene has a conserved 450 bp enhancer located at +37 Kb. Mutation of the enhancer site leads to more decreased expression of *Cebpa* mRNA than mutation of promoter sequence (Guo et al., 2012). Reporter

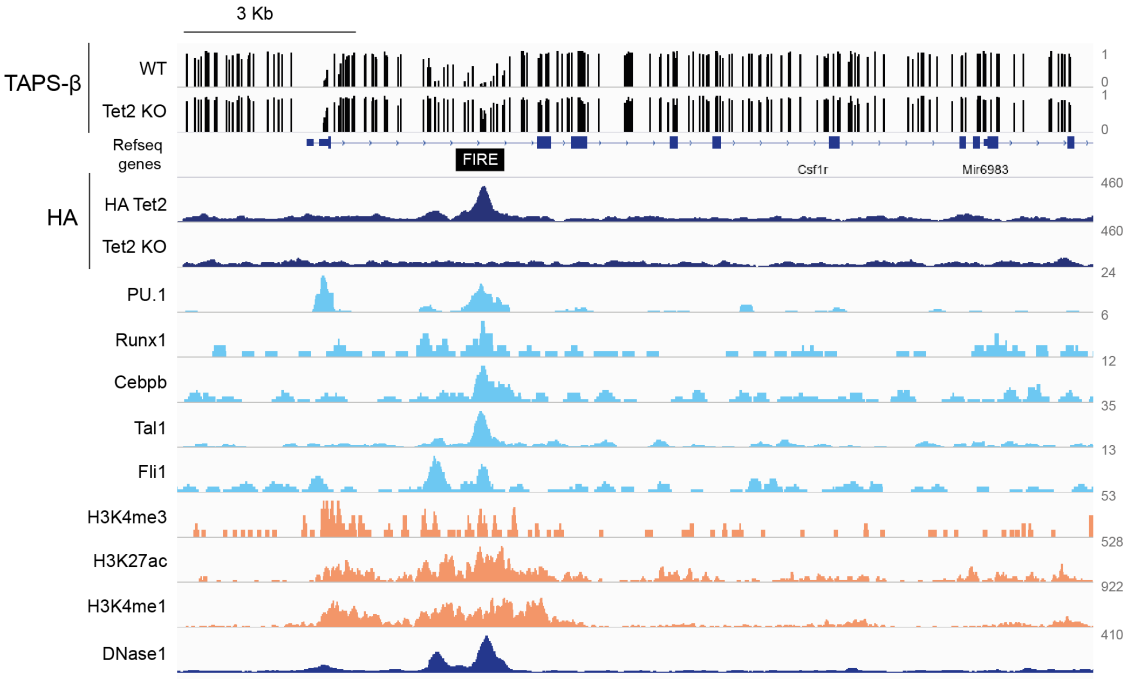


Figure 4.27: Tet2 co-localizes with PU.1 and Runx1 at *csf1R* enhancer. Genome browser tracks shows representative examples of co-localization of Tet2 with PU.1, Runx1 and Tal1 at c-fms intronic regulatory element (FIRE enhancer). 5mC level in wild-type and *Tet2*-KO HPC-7 cells was quantified using TAPS-β method by Dr. Sophie Kirschner and data analysis was performed by Olena Yavorska from our lab. Beta-value ranging from 0 to 1 was used to measure the percentage of methylation. HPC-7 transcription factor ChIP-seq, histone modification ChIP-seq and DNase1-seq data were downloaded from Wilson et al., 2016.

assay confirmed the role of Cebpa +37 Kb enhancer in regulating the myeloid-lineage specific expression (Guo et al., 2014b). An integrated analyses of ChIP-Seq data sets with genome-wide Promoter Capture Hi-C information revealed co-localization of many essential hematopoietic TFs, including Runx1, Gata2, Meis1, PU.1, Fli1, at the +37 Kb enhancer and confirmed the interaction between +37 Kb enhancer and Cebpa promoter (Wilson et al., 2016). Our result showed that Tet2 could directly bind to the Cebpa +37 Kb enhancer and mediate the DNA demethylation activity at its bound region (Figure 4.28).

Tet2 also binds to the recently discovered Cebpe +6 Kb enhancer in HPC-7 cells. Cebpe is essential for granulocytic differentiation and *Cebpe*-knockout mice display defects in terminal differentiation of granulocytes (Yamanaka et al., 1997). The +6 Kb enhancer is important for the expression of Cebpe since the depletion of the enhancer lead to reduced

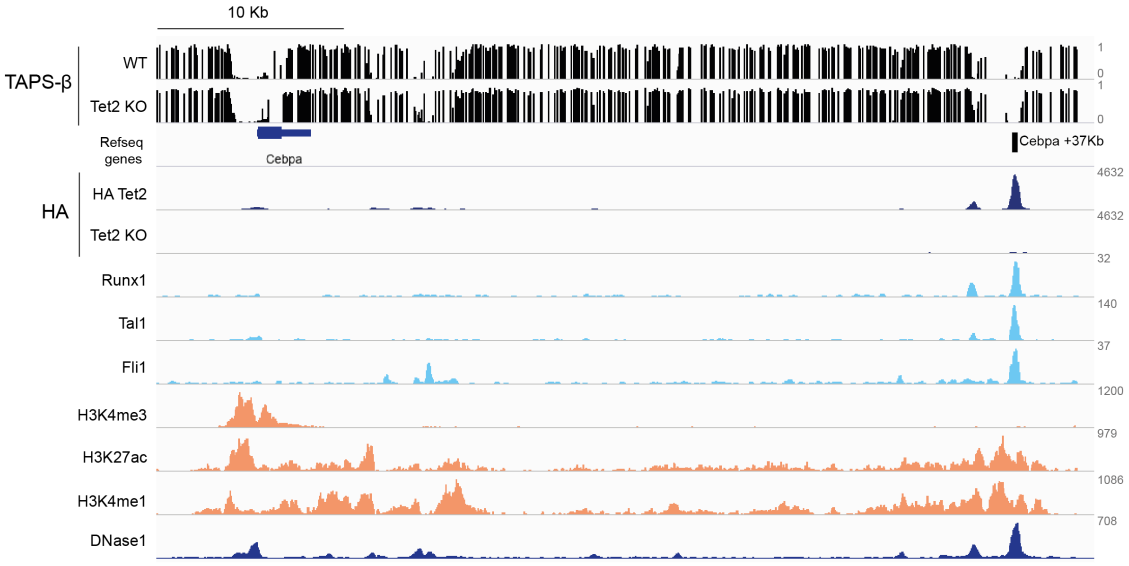


Figure 4.28: Tet2 co-localizes with multiple transcription factors at *Cebpa* +37 Kb enhancer.

Cebpe and cause the block of granulocytic differentiation (Shyamsunder et al., 2019). Tet2 binds to the Cebpe enhancer together with many TFs in HPC-7 cells (Figure 4.29) and affect the DNA methylation of the enhancer-associated region.

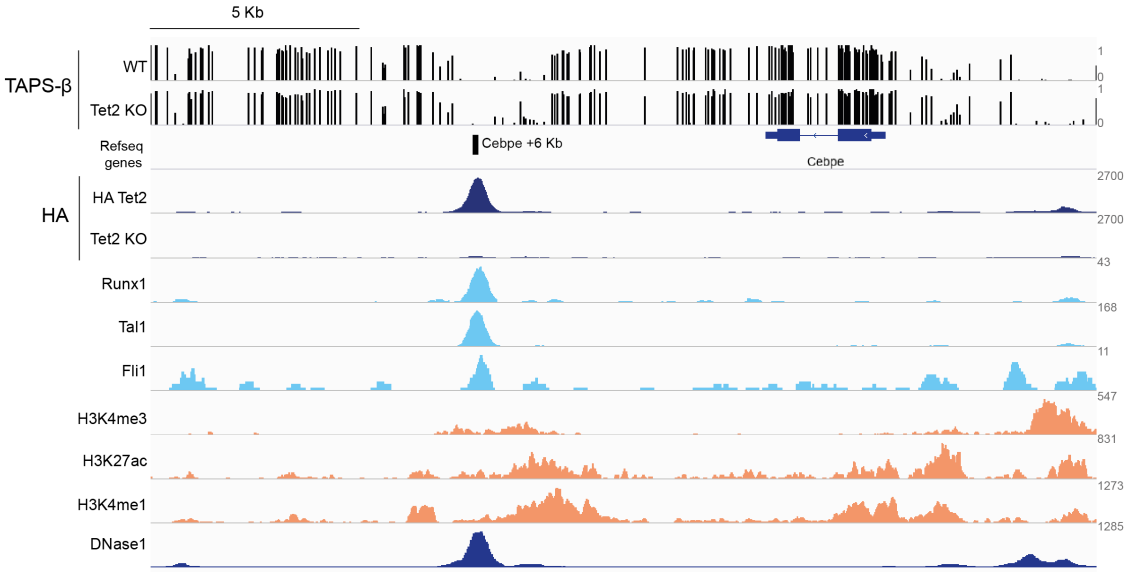


Figure 4.29: Tet2 co-localizes with multiple transcription factors at *Cebpe* +6 Kb enhancer.

4.3 Discussion

In this chapter, I performed multi-omics analysis including RNA-seq, ChIP-seq and TAPS-β analysis to map the molecular changes led by Tet2 deficiency and to dissect the

role of Tet2 in gene expression regulation.

Transcriptomic profiling of *Tet2*-KO HPC-7 cells identified slightly more up-regulated genes than down-regulated genes in both *Tet2*-KO clones compared with the WT HPC-7 cells, whereas down-regulated genes exhibited larger fold changes (Figure 4.3A). Two *Tet2*-KO clones share some common DE genes since ∼50% of identified DE genes in each KO clone could be also found in the other KO clone (Figure 4.3A). More importantly, among the top downregulated genes upon Tet2 deficiency, I observed several myeloid-specific factors such as *csf1R*, *Cebpa*, *Cebpe*. Upregulated genes included stem cell marker, c-Kit, HSC self-renewal factors (e.fg. Hlf, Sox4) which were also reported in other *Tet2*-KO cells isolated from mouse (Izzo et al., 2020). Around 125 genes related to hematopoiesis process were identified as significantly differential expressed genes in *Tet2*-KO cells, suggesting an important role of Tet2 activity in hematopoiesis.

Both *in vitro* myeloid differentiation (Figure 3.9) and CFU (Figure 3.10) assay showed reduced cell viability of *Tet2*-KO cells relative to WT cells under cytokine stimulation . Several cell cycle genes including *Cdkn2a* and *Cdkn2c* (Figure 4.30A) and cell apoptosis genes, such as *c-Fos* and *Casp3* (Figure 4.30B) showed differential gene expression in *Tet2*-KO cells. However, RNA-seq experiments were performed on cells cultured in SCF medium (100 ng/mL) whereas little difference in cell viability was observed at this condition (Figure 3.8). In order to understand the molecular changes on cell viability affected by Tet2 deficiency, transcriptomics profiling on cells under differentiation conditions would be necessary to reveal the answer.

To identify genes directly affected by Tet2 deficiency, I performed ChIP-seq experiments to map Tet2 binding sites and discover Tet2-direct targets. Lack of ChIP-grade Tet2 antibody and high background ChIP signal observed in *Tet2*-KO cells made it difficult to directly map the endogenous Tet2 protein. Thus, I constructed a HPC-7 cell line where the endogenous Tet2 protein was removed and 3xHA-Tet2 was expressed to be a similar level as the endogenous Tet2 protein. HA-ChIP identified 21883 Tet2 BS in HPC-7 cells (Figure 4.10). Tet2 peaks were mainly located in gene promoter enriched with H3K4me3 and H3K27ac, intron region and distal regulatory elements marked with H3K4me1 and

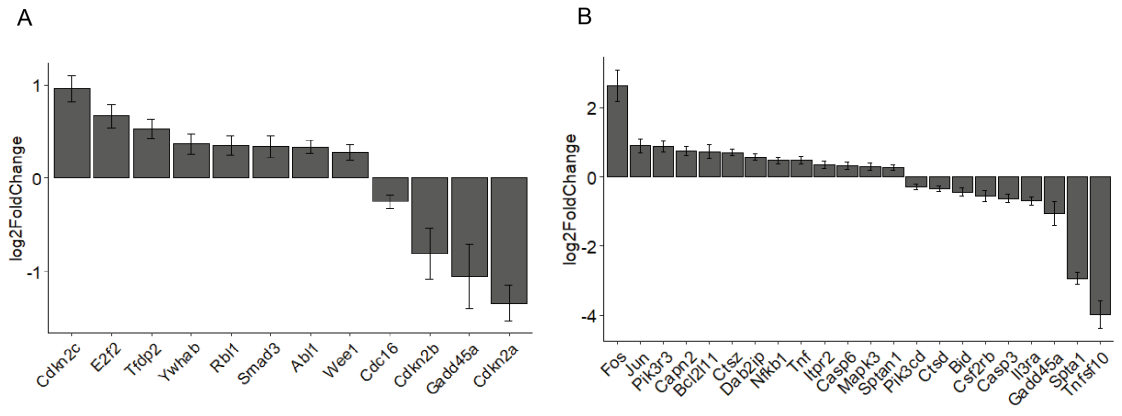


Figure 4.30: Expression changes of genes involved in cell cycle and cell apoptosis in *Tet2*-KO cells versus WT cells. Bar plots showing significantly (FDR<0.05) differential expressed genes involved in cell cycle (A) and cell apoptosis (B). Differential expression analysis was done by DESeq2. Y-axis represents the $\log_2$ fold change of gene expression in *Tet2*-KO clone 7 versus WT HPC-7 cells. Error bars indicate estimated standard error for the log2-scaled fold changes.

H3K27ac (Figure 4.11 and 4.12). Rasmussen and colleagues revealed a cell-type specific manner of Tet2 binding pattern (Rasmussen et al., 2019). Our identified HPC-7 Tet2 BS was enriched with myeloid-specific TF binding motifs and exhibited the similar TF binding pattern as what was reported in the mouse myeloid cell Tet2 peaks (Figure 4.15). More importantly, most of the up/down-regulated DE genes identified in *Tet2*-KO cells were found to have Tet2 genomic binding, indicating the dual function of Tet2 activity in gene activation and repression (Figure 4.16).

Tet2 is known as a cytosine methylation dioxygenase and Tet-mediated 5mC iterative oxidation provides an DNA demethylation mechanism. To explore whether Tet2 binding is closely related to its DNA demethylation activity, Dr. Sophie Kirschner from our group performed TAPS-β experiments to map the 5mC level in two *Tet2*-KO clones, Tet2 WT cells and HA-Tet2 expressing HPC-7 cells. Olena Yavorska, a DPhil student from the group did the data analysis. Although there was no global gain or loss of DNA methylation observed in *Tet2*-KO cells (Figure 4.18), enhancers exhibited DNA hypermethylation upon Tet2 deficiency (Figure 4.19). More importantly, restoration of Tet2 expression in *Tet2*-KO cells reverses the hypermethylation phenotype observed on enhancers. These results support a predominant role of Tet2-mediated DNA demethylation at enhancer regions. And this is consistent with findings from other studies where the enrichment

of DNA hypermethylation caused by Tet2 depletion on enhancer regions was identified in a variety of cells types including mouse ESCs, mouse blood cells and humam AML patients (Hon et al., 2014; Lio et al., 2019a; Lu et al., 2014; Rasmussen et al., 2015). Moreover, Tet2 depletion causes an increase of DNA methylation at some Tet2 binding sites (Figure 4.20) and the increase of DNA methylation is more significant in enhancer-associated Tet2 binding sites, which further supports the important role of Tet2 binding and Tet2-mediated DNA demethylation at enhancer regions (Figure 4.21). For genes having Tet2 binding at the promoter region, the expression could get downregulated and exhibited hypermethylation once Tet2 is depleted, suggesting that Tet2 is involved in gene activation via its DNA demethylase activity. Wherease, some Tet2-bound genes could also get upregulated when Tet2 is not expressed, this indicates a repressive role of Tet2 in gene expression regulation, and the mechanism by which Tet2 negatively affects its targeted gene expression still remain unknown.

In the study, I also found several examples of Tet2 function at hematopoietic enhancer regions and the binding of Tet2 affects the DNA demethylation and thus lead to the expression changes of the enhancer-cognate genes (Figure 4.27,4.28,4.29). *csf1R*, *Cebpa* and *Cebpe* are key hematopoietic factors involved in myeloid lineage specification, and the expression of genes have been found to be regulated by enhancer elements. The three genes were significantly down-regulated in *Tet2*-KO cells and exhibited Tet2 binding at their associated enhancer region. More importantly, DNA hypermethylation was observed at these enhancer regions upon Tet2 deficiency. Enhancers are DNA regulatory elements bound by many DNA-binding transcription factors to activate the transcription of its associated gene to a higher level (Koch et al., 2011; Krivega and Dean, 2012. At the Tet2-bound myeloid-specific enhancers, I also observed the binding of many other TFs, including Runx1, Fli1, Gata2, PU.1, and Cebpb. Further analysis revealed Tet2 co-localizes with many essential hematopoietic TFs in the HPC-7 genome (Figure 4.24). Among the TFs, Runx1 and PU.1 has been reported to recruit Tet2 (Chu et al., 2018; de la Rica et al., 2013) at the TF-targeted genes. So that Runx1 or PU.1 might be responsible for the recruitment of Tet2 to the chromatin, Tet2 might further recruit other factor to the chromatin or affect the TF complex formation at the target region. A

systematic analysis have revealed the impact of cytosine methylation on DNA binding ability of human transcription factors (Yin et al., 2017), where DNA methylation could affect the binding of TFs in both positive and negative way. The analysis showed that bHLH-, bZIP-, and ETS-family TFs were generally inhibited by mCpG (methyl-minus), whereas NFAT (RHD) TFs and many members of the homeodomain family show preference in binding to mCpG-containing sequences (Yin et al., 2017). The binding of Fli1, a member of ETS family protein, was shown to be inhibited by DNA methylation. The predisposition of Tet2 at the enhancer region alters the DNA methylation level and could prevent or promote the further binding of TF, thus affect the enhancer function in regulating gene expression.

5

Tet2 interactome mapping in HPC-7 cells

Contents

5.1 Introduction

In vivo, protein rarely performs as an isolated individual when exerting its function. Over 80% of proteins have been reported to operate within complex (Berggård et al., 2007). The study of protein-protein interaction (PPI) not only helps to understand the function of the protein but also helps to predict a protein's unknown function and related cellular processes. Previous studies have identified several Tet2 interacting partners which have

provided some insights in understanding of Tet2 chromatin recruiting mechanism, Tet2-involved gene regulation or post-translational regulation of Tet2 protein. However, little is known about the interacting partners of Tet2 in hematopoietic context since almost all Tet2 interactors were identified in hematopoietic-irrelevant cell background. Thus, in this study, facilitated by the availability of the HPC-7 cells for biochemical applications, I performed the first Tet2 interactome mapping in a hematopoietic cell context, and hope it could benefit our understanding of Tet2-mediated gene regulation and the cellular phenotypes identified in *Tet2*-KO cells.

In this chapter, I will present the work in identifying Tet2-interacting partners in HPC-7 cells by firstly introducing the exploration of proximity-based methods (BioID and APEX2-labelling) for PPI identification (Chapter 5.2.1), then I will explain the problems and difficulties when applying the proximity-based method in HPC-7 (Chapter 5.2.2). In the end, I will illustrate the construction of an inducible 3xHA Tet2-expressing HPC-7 cell (Chapter 5.2.3) and introduce the Tet2 interactome mapping using anti-HA immunoprecipitation-Mass Spectrometry (IP-MS) (Chapter 5.2.4).

5.2 Results

5.2.1 Optimisation of protein-protein interaction mapping methods

Yeast two-hybrid (Y2H) screening and affinity-purification coupled to mass-spectrometry (AP-MS) approach were two major methods for PPI identification at the large scale (Dunham et al., 2012; Yu et al., 2008b). Despite Y2H is a powerful approach for two proteins interaction mapping, testing the mammalian PPI in a heterologous system such as the yeast might result in false-positive or false-negative results if the PPI is dependent on certain post-translational modifications. In contrast, AP-MS capture the protein complex of interest from the mammalian cell system via enrichment with antibody against protein of interest or epitope tag, which is well-suited for high-throughput experiments. However, lack of a good protein antibody for affinity-purification have brought limits while using the method. Although expressing the recombinant tagged protein and performing IP using the epitope tag antibody could alleviate this issue, the over-expression of the recombinant

protein, or the misfolding of the protein, and the *in vitro* purification steps could lead to increase in false-positive or false-negative results.

Recent years, the proximity-based labelling has been developed and widely used as a powerful method for protein interactome mapping. BioID and APEX2 labelling are the two main proximity-based interactome mapping methods. In both methods, protein of interest and its interacting proteins were labelled *in vivo* through the enzyme-catalyzed biotinylation, and identified by streptavidin purification/mass spectrometry. The *in vivo* labelling-based method could avoid the false-positive interactions formed during *in vitro* purification and false-negative interactions lost during the purification process (Trinkle-Mulcahy, 2019). To test if the approach is suited for Tet2 interaction identification, I examined the strength of the BioID and APEX2 labelling method in identifying the known Tet2 interaction with OGT in 293 T cells which are easy-to-transfect cells and was originally used for identifying TET2-OGT interaction (Chen et al., 2013).

BioID is not able to capture known Tet2 interaction with OGT

BioID utilizes the promiscuous prokaryotic biotin ligase mutant BirA* which generates the reactive biotinyl-AMP (bioAMP) from biotin and ATP (Choi-Rhee et al., 2004; Cronan, 2005; Kim et al., 2014; Lane et al., 1964; Roux et al., 2012). And the free bioAMP will react with adjacent primary amines, resulting in biotinylation of the proteins

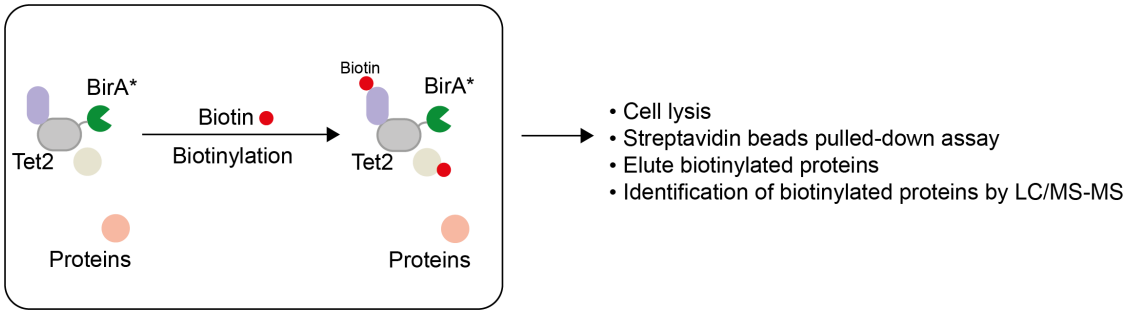


Figure 5.1: Schematic representation of BioID proximity labelling assay. Image is adapted from *Roux et al., 2012 Figure 1*. Cells expressing Tet2-BirA* fusion proteins were treated with 50 μM biotin for 24 h to biotinylate protein in close proximity to the fusion protein. Cells were lysed and biotinylated proteins were captured by streptavidin-conjugated beads for downstream identification.

in close proximity (labelling radius, 10 nm) to the fusion protein (Figure 5.1). BiA*-induced biotinlyated proteins were captured by streptavidin-conjugated beads, followed by Western blot analysis.

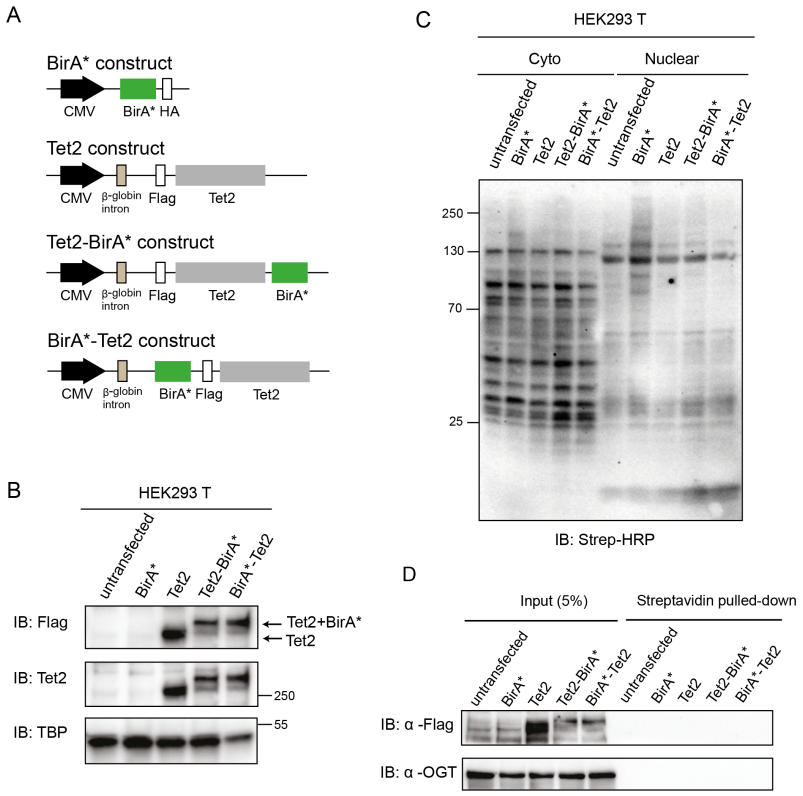


Figure 5.2: Tet2-birA fusion protein is not enzymatically active in biotinylation labelling in 293 T cells (A) Constructs used for testing BioID method in HEK293 T cells. CMV: human cytomegalovirus promoter; BirA*: *Escherichia coli* biotin-ligase R118G mutant; HA: human influenza hemagglutinin tag; β-globin intron: intron from rabbit β-globin gene; Flag: DYKDDDDK peptide; Tet2: mouse Tet2 gene. (B) Western blot analysis of Tet2-BirA* fusion protein in nuclear extracts from HEK293 T cells transiently transfected with indicated constructs. untransfected: No DNA and no transfection reagents. Tet2 only serves as a control and shows the nuclear expression in 293 T cells. TBP is served as the nuclear loading control. (C) Western blot analysis of BioID-mediated proximity labelling by blotting cytoplasmic (Cyto) and nuclear (Nuclear) fractions from transfected cells shown in (B) with streptavidin-horseradish peroxidase (Strep-HRP). IB: immunoblot. untransfected: No DNA and no transfection reagents. BirA* serves as a control and indicates the biotinylation dependent on BirA* expression. (D) Streptavidin pull-down of HEK293 T cells transiently expressing Tet2-BirA* constructs. Eluted biotinlyated proteins were analyzed by Western blotting antibodies against Flag and OGT.

In this study, BirA* was fused to either N-terminus or C-terminus of Tet2 protein (Figure 5.2A) and the fusion constructs were transiently expressed in 293 T cells cultured in biotin supplemented medium for 24 hr. Tet2-BirA* fusion protein showed clear expression indicating the tag does not interfere the protein folding or the antibody recognition

(Figure 5.2B). Western blot probed with streptavidin-HRP did not detect increased protein biotinylation signal in nucleus and cytoplasm fractions of transfected cells (Figure 5.2C). Moreover, streptavidin pull-down assay did not capture any biotinlyated Tet2 or biotinlyated OGT protein (Figure 5.2D), suggesting BirA* tag is not catalytically active when fused with Tet2 protein. In summary, BioID method is not capable in capturing Tet2 interactome because of the lack of activity of BirA* tag.

APEX2 proximity labelling was suitable for Tet2 interactome mapping in 293 T cells

APEX2 labelling assay was also tested in mapping the known interaction between Tet2 and OGT in 293 T cells. APEX2 is an engineered plant ascorbate peroxidase which has the same molecular weight as the GFP protein (27 kDa) (Hung et al., 2016). Figure 5.3 illustrates the process of APEX2-mediated proximity labelling in interactome mapping. Tet2-APEX2 expressing cells were incubated with biotin-phenol for 30 min and then treated with 1 mM Hydrogen Peroxide (H_2O_2) for 1 min to initiate biotinylation labelling. APEX2 uses H_2O_2 as an oxidant to catalyze the oxidation of biotin-phenol (BP) into a biotin-phenoxyl radical. Biotin-phenoxyl radical is short lived (<1ms) and covalently

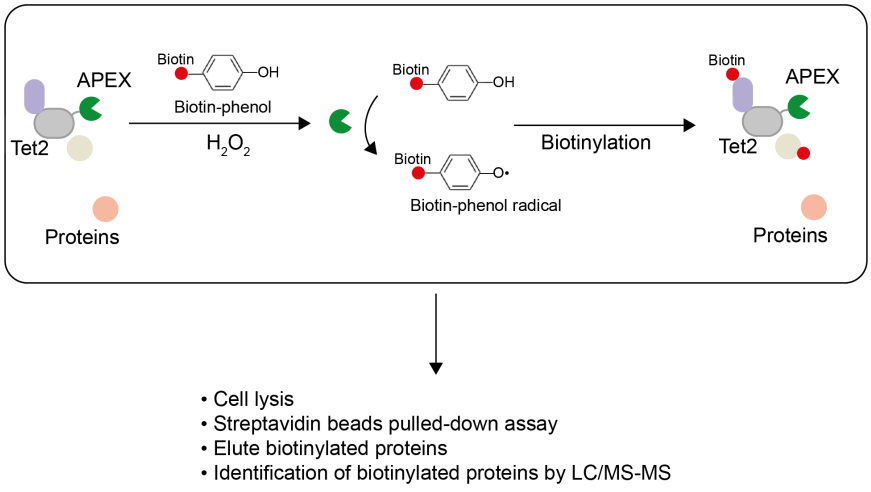


Figure 5.3: Schematic representation of APEX2 proximity labelling assay. Image is adapted from *Hung et al., 2016 Figure 1a*. Cells expressing Tet2-APEX2 fusion proteins were treated with biotin-phenol for 30 min and then incubated with 1 mM H_2O_2 for 1 min to initiate biotinylation labelling of proteins in close proximity to the fusion protein. Cells were lysed and biotinylated proteins were captured by streptavidin-conjugated beads for downstream identification.

tags the proximal neighbouring proteins (Hung et al., 2016). Compared to BioID method, APEX2 method has much faster labeling rate (1 minute instead of 24 h).

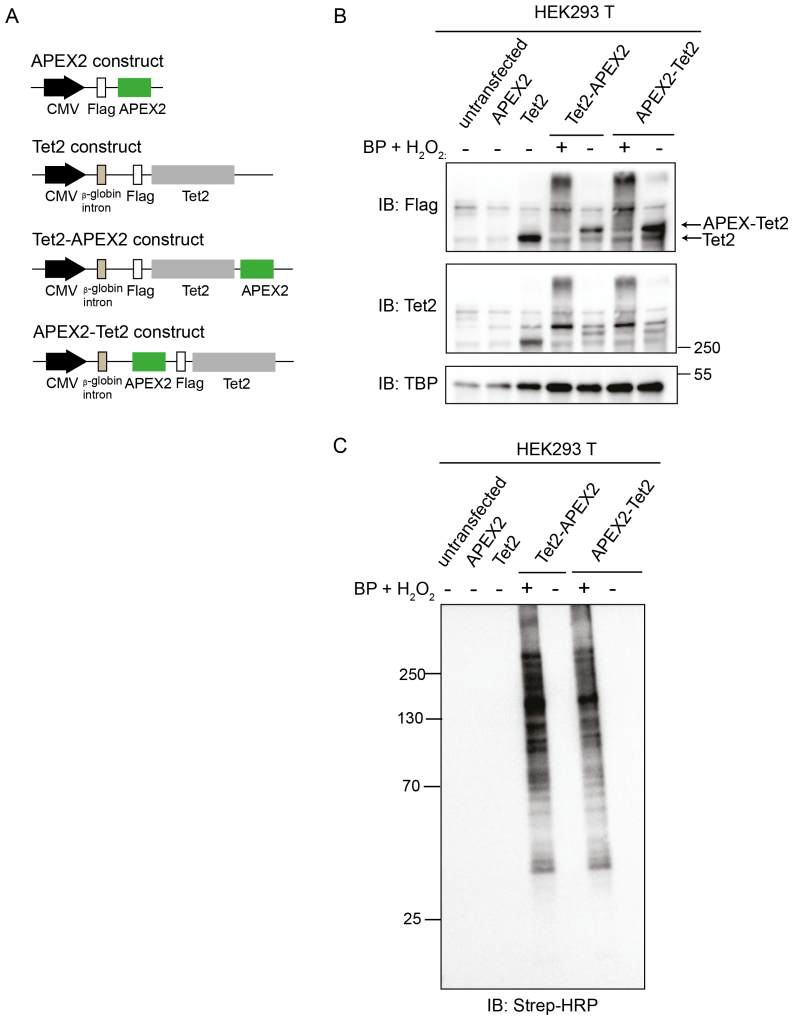


Figure 5.4: Tet2-APEX2 fusion protein is biotinylation labelling-active. (A) Constructs used for testing APEX2 method in HEK293 T cells. CMV: human cytomegalovirus promoter; Flag: DYKDDDDK peptide; APEX2: engineered plant ascorbate peroxidase; β-globin intron: intron from rabbit β-globin gene; Tet2: mouse Tet2 gene. (B) Western blot analysis of Tet2-APEX2 fusion protein in nuclear extracts from HEK293 T cells transiently transfected with indicated constructs. untransfected: No DNA and no transfection reagents. Tet2 only serves as a control and shows the nuclear expression in 293 T cells. Samples in which biotin-phenol and H_2O_2 were omitted serve as negative controls. TBP is used as the nuclear loading control. (C) Western blot analysis of APEX2-mediated proximity labelling by blotting nuclear lysates from transfected cells shown in (B) with streptavidin-horseradish peroxidase (Strep-HRP). Samples in which biotin-phenol and H_2O_2 were omitted serve as negative controls and indicate the endogenous biotinylation level.

APEX2 tag was fused to either N-terminus or C-terminus of Tet2 protein and fusion constructs were transiently expressed in 293 T cells (Figure 5.4)A. BP and H_2O_2 omitted

samples were used as the control for the specificity of biotinylation labelling. Tet2-APEX2 showed clear expression in the transfected cells (Figure 5.4B). Interestingly, the addition of BP and H_2O_2 to the medium induced a band shift of Tet2-APEX2 fusion protein on the blot (Figure 5.4B). It is unclear about what causes the band shift of Tet2-APEX2 fusion protein during the BP/ H_2O_2 treatment. To test if the Tet2-APEX2 fusion protein is biotinylation labelling-active, I checked the protein biotinylation level in the extracted nuclear lysates of the transfected cells (Figure 5.4C). Tet2-APEX2 induced increased biotinylation when cells were treated with BP and H_2O_2 (Figure 5.4C). The labelled protein patterns were similar between N- and C-terminal APEX2-tagged proteins.

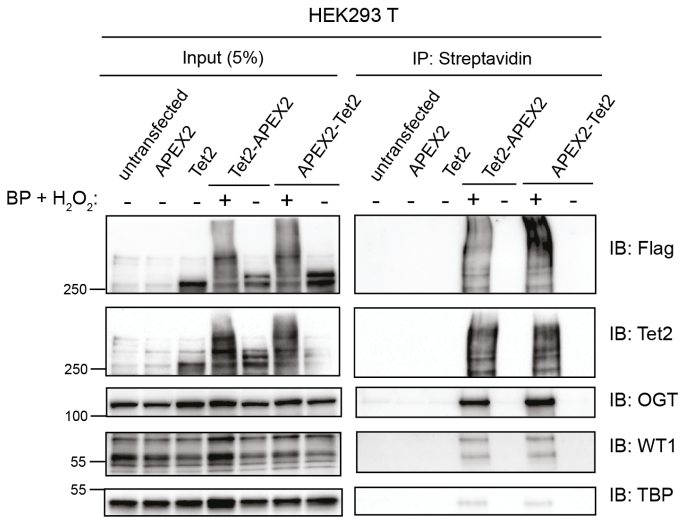


Figure 5.5: Identification of the known Tet2 interactions in 293 T cells by APEX2 proximity labelling assay. Streptavidin pull-down of HEK293 T cells transiently expressing the APEX2 or Tet2-APEX2 fusion proteins. Eluted proteins were separated by SDS-PAGE and analyzed by Western blotting using the indicated antibodies. untransfected: No DNA and no transfection reagents. Samples in which biotin-phenol and H_2O_2 were omitted serve as negative controls and indicate the endogenous biotinylation level. The expression of Tet2 and Tet2 fusion protein are detected by immunoblotting (IB) with anti-Flag and anti-Tet2. The detection of reported Tet2 interactions in Tet2-APEX2 labelling proteome are examined by IB with anti-OGT and anti-WT1. TBP serves as the loading control and negative control for non-specific biotinylation labelling.

Streptavidin pull-down experiments revealed that Tet2-APEX2 fusion protein could self-biotinylate and biotinylate the known Tet2-interactors, OGT and WT1 (Figure 5.5), suggesting the strength of APEX2 labelling as a way to map Tet2 interactome. However, TATA-box binding protein (TBP) which has not been reported to interact with Tet2 was also captured by Streptavidin-conjugated beads when cells were treated with BP and H_2O_2,

indicating TBP was biotinylated by Tet2-APEX2 *in vivo*. The detection of biotinylated TBP might be a consequence of over-expression of Tet2-APEX2 fusion protein and these might suggest a potential problem in having false positive protein-protein interaction when using the APEX2 labelling assay.

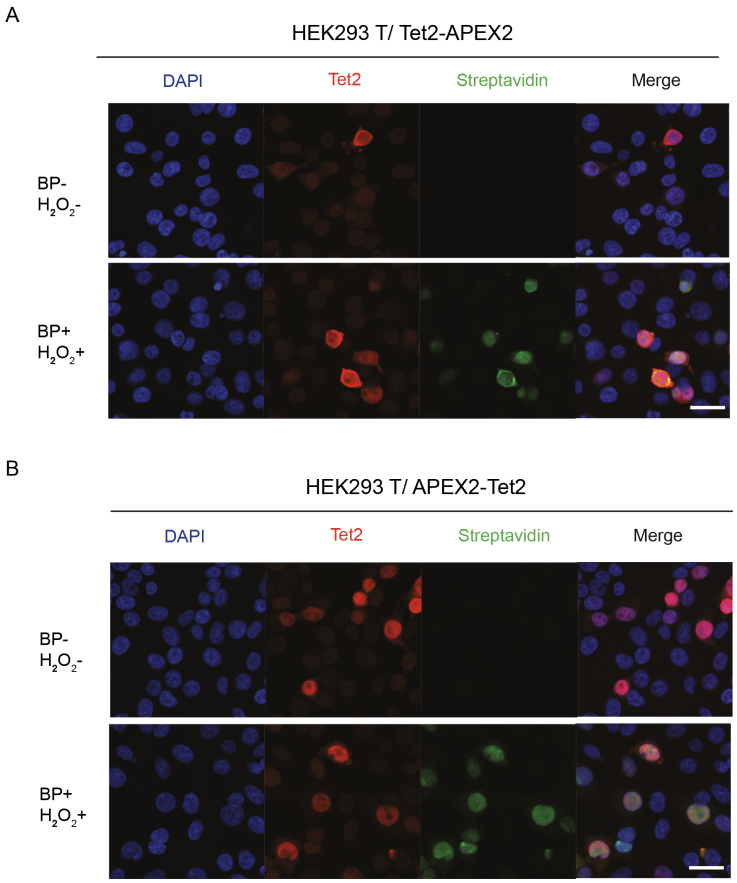


Figure 5.6: Tet2-APEX2 is localized in nucleus. Immunofluorescence imaging of Tet2-APEX2 fusion protein in HEK293 T cells. Cells were transfected with either Tet2-APEX2 (A) or APEX2-Tet2 (B) construct for 24 h, and treated with biotin-phenol and H_2O_2 to initiate biotinylation labelling. Samples in which biotin-phenol and H_2O_2 are omitted serve as a negative control and indicate the effects of chemical treatments on protein localization. DAPI (blue) indicates nuclear staining. Tet2-APEX2 fusion protein is stained with antibody against Tet2 (red). Biotinylation labelling is visulized by Streptavidin, Alexa Fluor® 488 Conjugate (Green). Scale bar= 20 μM.

Immunofluorescence results confirmed the predominantly nuclear localization of Tet2-APEX2 fusion protein and the specificity of the biotinylation reaction (Figure 5.6). Furthermore, the N-terminal fusion protein (APEX2-Tet2) was shown to be expressed more specifically inside the nucleus, indicating the N-terminal tagging has less effect on protein localization, thus it is more suitable for Tet2 tagging (Figure 5.6B). Expression

of WT Tet2 could lead to evidently increased 5hmC and decreased 5mC (Figure5.7). Comparatively, Tet2-APEX2-transfected cells also showed clear induction of 5hmC. Moreover, BP/H_2O_2 treatment seem to have little effect on Tet2 enzymatic activity.

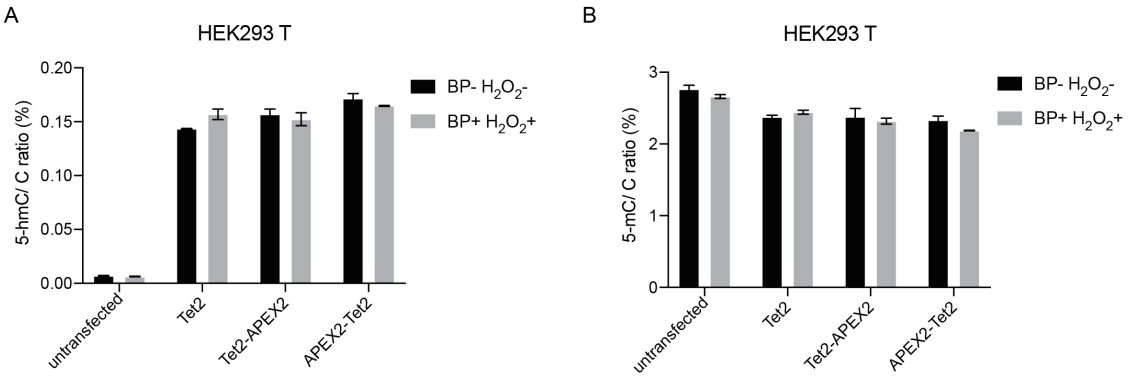


Figure 5.7: Tet2-APEX2 can catalyze the 5-mC oxidation. LC-MS/MS measurements of genomic 5-hmC level (A) and 5-mC level (B) in HEK293 T cells transiently expressing Tet2 or Tet2-APEX2 fusion proteins. untransfected: No DNA and no transfection reagents. Tet2 only serves as a control and indicates the 5mC oxidation activity of wildtype protein. Samples in which biotin-phenol and H_2O_2 are omitted serve as a negative control and indicate the effects of chemical treatments on 5mC oxidation. Mean values from three biological replicates are plotted. Error bars indicate standard deviation.

In summary, Tet2-APEX2 expresses mainly in nucleus and APEX2 tag does not interfere the 5mC oxidation activity of Tet2 protein. Tet2-APEX2 could label the known Tet2-interactor, OGT, through biotinylation, suggesting APEX2 proximity labelling method could be used for Tet2 interactome mapping in HPC-7 cells. However, the unexpected TBP biotinylation observed in the Tet2-APEX2 expressing cells suggests a potential false positive labelling due to the overexpression of the APEX2 fusion protein. To express the APEX2 protein via a weak promoter follow by a stable expression, or to use the inducible expression system might be necessary to avoid any false positive labelling.

5.2.2 Tet2 could not be expressed by Retrovirus expression system

To circumvent problems with protein overexpression, I firstly attempted to use the MSCV retrovirus system to achieve the stable expression of APEX2-Tet2 in HPC-7 cells (Figure 5.8). Murine stem cell virus (MSCV) expression system was selected in this study since MSCV LTR exhibits a strong activity in driving gene expression in myeloid cells and embryonal cells (Weber and Cannon, 2007) and MSCV system has also been widely used

for ectopic gene expression in HPC-7 cells. HPC-7 cells expressing APEX2 protein was generated and used as a control for the specific biotinylation labelling of Tet2-interactome. GFP was used as a marker to facilitate the detection of integration of the fusion protein expression cassette. After viral transduction, GFP$^+$ cells were isolated and maintained in SCF culture medium (Figure 5.8).

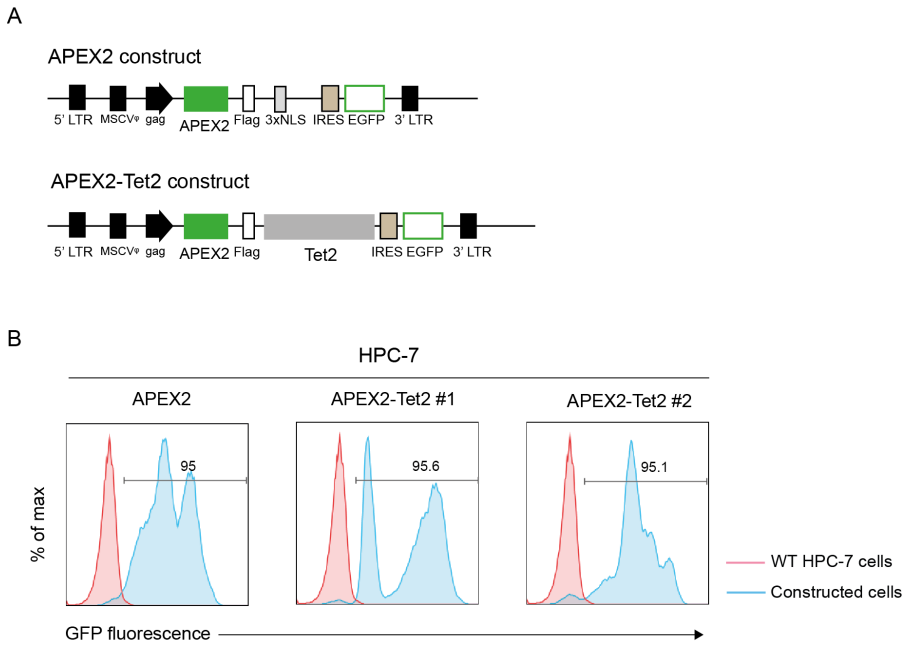


Figure 5.8: Construction of APEX2-tagged Tet2-expressing HPC-7 cells by retrovirus system. (A) Retroviral constructs used for Tet2-APEX2 expression in HPC-7 cells. LTR: long terminal repeats; MSCVψ: viral packaging signal ; gag: viral packaging gene; APEX2: engineered plant ascorbate peroxidase; Flag: DYKDDDDK peptide; Tet2: mouse Tet2 gene; NLS: nuclear localization signal; EGFP: enhanced green fluorescence protein; IRES: internal ribosomal entry site. (B) FACS analysis of HPC-7 cells expressing either APEX2 or APEX2-Tet2 fusion protein. Percentage of eGFP$^+$ cells is indicated. APEX2-Tet2 #1 and APEX2-Tet2 #2 are cells infected at the same time but sorted independently.

Although APEX2-Tet2 HPC-7 cells showed clear expression of GFP (Figure 5.8B), APEX2-Tet2 protein was not detected in Western Blot using antibodies against Tet2 and Flag tag (Figure 5.9). To further examine the expression of APEX2-Tet2 fusion protein, I treated the APEX2-Tet2 cells with BP and H_2O_2 to check if the fusion protein could induce biotinylation labelling *in vivo* (Figure5.10).

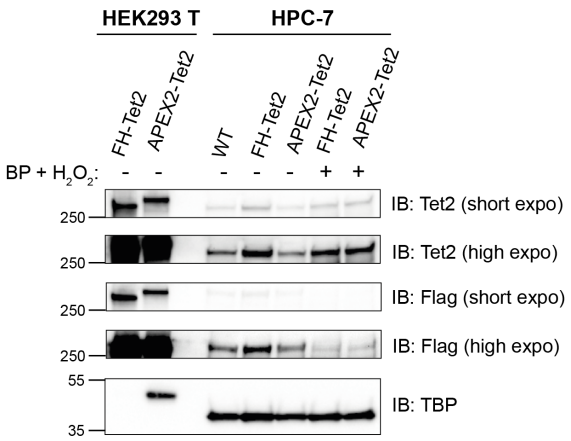


Figure 5.9: APEX2-Tet2 is not expressed in HPC-7 by retrovirus system. Cells were treated with biotin-phenol and H_2O_2 to initiate biotinylation labelling. Nuclear proteins are extracted, separated by SDS-PAGE and analyzed by western blotting. Tet2 fusion protein is detected by antibodies against Tet2 and Flag tag. TBP is served as the nuclear loading control. Tet2 fusion proteins expressed in 293 T cells serve as a control and indicate the right molecular size.

Although APEX2-Tet2 could induce biotinylation labelling in nucleus when cells were treated with BP and H_2O_2 (Figure 5.10A), the increased biotinylation was also observed in treated WT HPC-7 cells. More importantly, the treated WT cells showed similar biotinylation signal as the treated APEX2-Tet2 expressing cells (Figure 5.10A). Further experiments revealed the induced biotinylation in WT cells only occurred when cells were treated with both BP and H_2O_2, indicating the existence of an endogenous peroxidase in WT cells. Streptavidin pull-down experiments revealed the weak bands in blots stained with antibody against Tet2 and its known interactor, OGT (Figure 5.10C). However, little difference was observed in streptavidin pull-down results between WT and APEX2-Tet2 cells. In conclusion, lack of clear expression of APEX2-Tet2 and the unexpected endogenous biotinylation background in WT HPC-7 cells bring us concerns in applying the proximity labelling method for Tet2 interactome mapping.

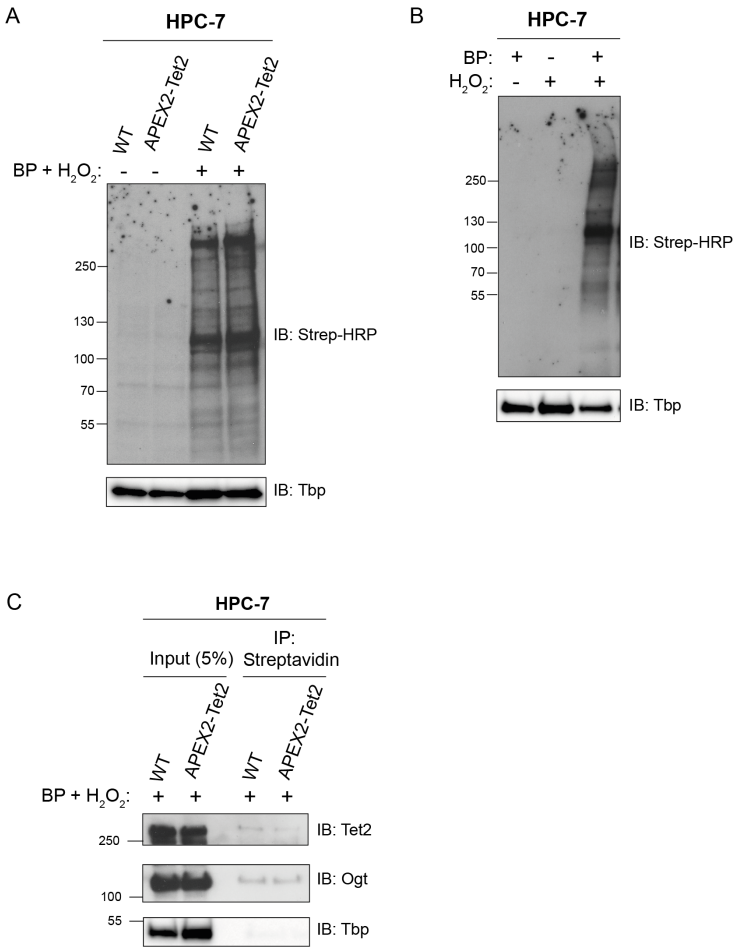


Figure 5.10: HPC-7 cells have endogenous biotinylation background. (A) Western blot analysis of APEX2-mediated proximity labelling by blotting nuclear lysates from HPC-7 cells stably expressing Tet2-APEX2 with streptavidin-horseradish peroxidase (Strep-HRP). Wild-type HPC-7 serves as a control and indicates endogenous biotinylation level. Samples in which biotin-phenol and H_2O_2 were omitted serve as control for the reaction specificity. (B) Western blot analysis of endogenous biotinylation background in nuclear lysates from wild-type HPC-7 cells. (C) Streptavidin pull-down of HPC-7 stably expressing Tet2-APEX2 fusion proteins. Eluted proteins were separated by SDS-PAGE and analyzed by Western blotting using the indicated antibodies. The expression of Tet2 and Tet2 fusion protein are detected by immunoblotting (IB) with anti-Tet2. The detection of reported Tet2 interactions in Tet2-APEX2 labelling proteome are examined by IB with anti-OGT. TBP serves as the loading control and negative control for non-specific biotinylation labelling.

To map Tet2 interactome in HPC-7 cells, I also tried the classic affinity-purification based method in the meantime and generated a HPC-7 cell line with the stable expression of Flag-HA-Tet2 (FH-Tet2) (Figure 5.11A). FH-Tet2 cells showed clear expression of GFP by flow cytometry (Figure 5.11B). Although there was no big change in the total *Tet2* mRNA level in FH-Tet2 cells, *FH-Tet2* mRNA did get expressed in the constructed cells

(Figure 5.11C).

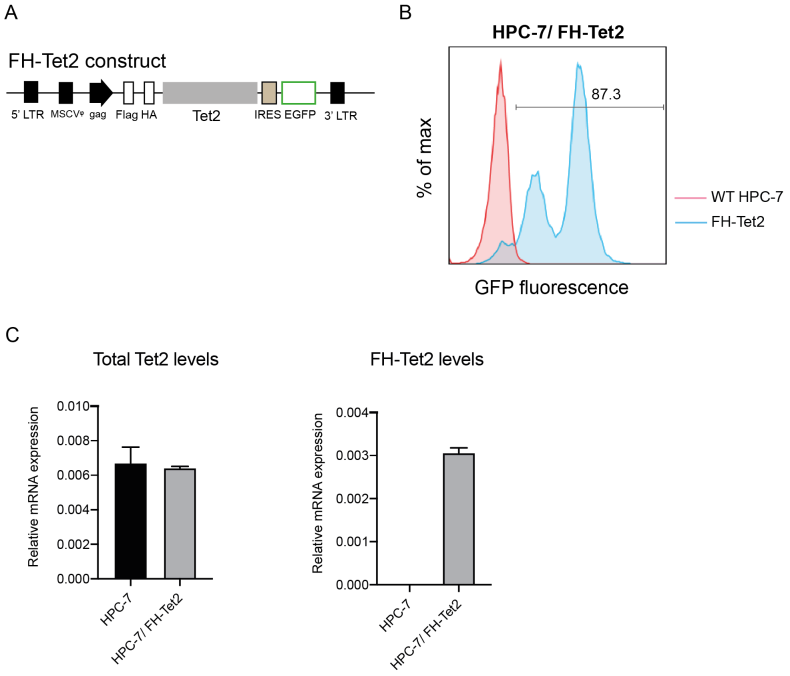


Figure 5.11: Construction of Flag-HA-tagged Tet2-expressing HPC-7 cells by retrovirus system. (A) Retroviral constructs used for Flag-HA tagged Tet2 (FH-Tet2) expression in HPC-7 cells. LTR: long terminal repeats; MSCVψ: viral packaging signal ; gag: viral packaging gene; Flag: DYKDDDDK peptide; HA: human influenza hemagglutinin tag ; Tet2: mouse Tet2 gene; EGFP: enhanced green fluorescence protein; IRES: internal ribosomal entry site. (B) FACS analysis of GFP signal in HPC-7 cells expressing Flag-HA tagged Tet2. Percentage of eGFP$^+$ cells is indicated. (C) Quantitative RT-qPCR analysis of total *Tet2* (Left) and FH-Tet2 (Right) in HPC-7 stably expressing FH-Tet2. Data are normalized to Gapdh. Mean values from three biological replicates are plotted. Error bars indicate standard deviation.

The expression of FH-Tet2 protein was confirmed by Western blot probed with anti-HA antibody (Figure 5.12A). A slightly increased total Tet2 level was observed in FH-Tet2 cells (Figure 5.12B). Problem was encountered while validating the reported interaction between Tet2 and OGT (Figure 5.12C). Although anti-HA IP identified OGT in the IP products from FH-Tet2 cells, OGT was also observed in IP experiments on WT HPC-7 cells. Silver stain did not reveal any difference in IP experiments between WT and FH-Tet2 cells (Figure 5.12D). Moreover, no protein band was detected in IP products during silver stain, which might indicate the low efficiency of IP enrichment or the low expression level of bait protein (FH-Tet2) in cells . In summary, lack of expression of APEX2-Tet2 fusion protein and low expression of FH-Tet2 both indicate the incapability of retroviral expression system when expressing large constructs like Tet2 (>200 kDa).

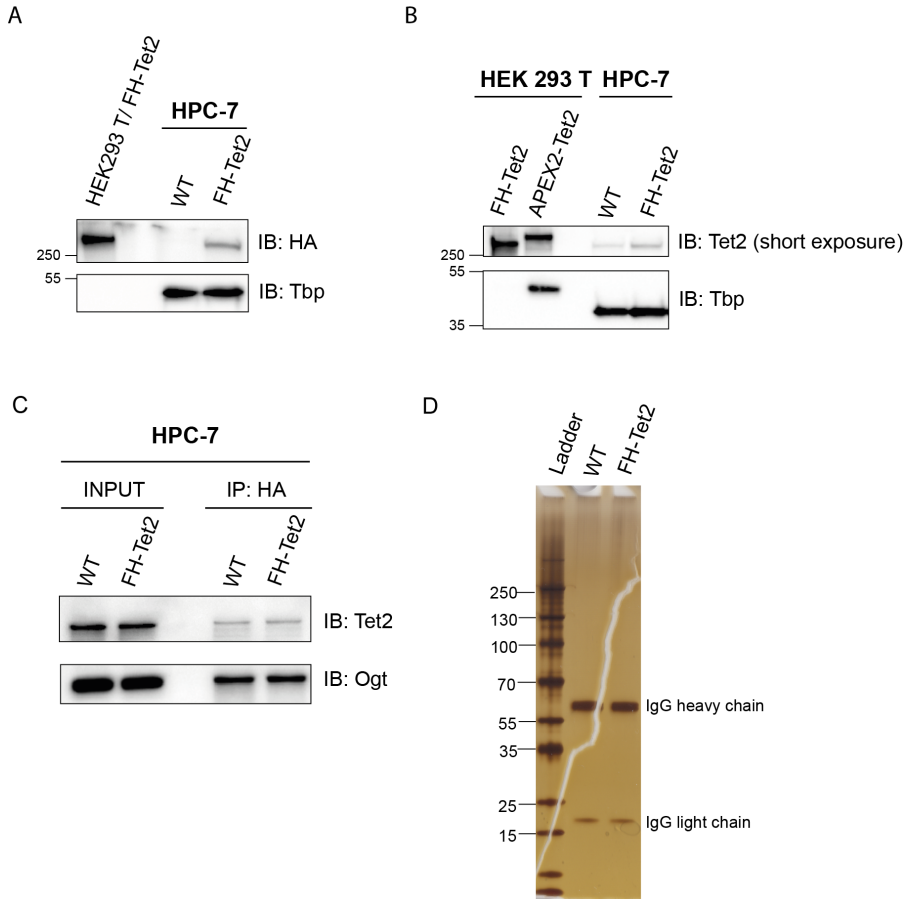


Figure 5.12: Characterization of FH-Tet2 expression in HPC-7. Western blot analysis of FH-Tet2 expression using antibodies against HA tag (A) and Tet2 (B). TBP is served as the nuclear loading control. FH-Tet2 proteins expressed in 293T cells serve as a control and indicate the right molecular size. (C) FH-Tet2 protein was immunoprecipitated with HA antibody, followed by the western blotting analysis to detect the co-immunoprecipitation between Tet2 and Ogt. WT: wild-type HPC-7. (D) silver stain of 10% anti-HA immunoprecipitation purification with WT or FH-Tet2 expressing HPC-7 cells.

5.2.3 Generation of an inducible 3xHA Tet2 expressing HPC-7 cell line

Lack of sufficient expression of tagged-Tet2 with the retrovirus system suggest that the mammalian gene expression system with a stronger promoter activity would be required for the ectopic Tet2 expression. Strong constitutive promoters such as SV40, CMV, and EF1A would be helpful to achieve the strong expression of protein of interest (Qin et al., 2010) but over-expression might occur at the same time. In order to achieve a good expression level of the tagged-Tet2 protein for affinity purification/MS, I adapted the inducible expression system, piggyBac-Tet-on system, from Colin Goding group (Figure

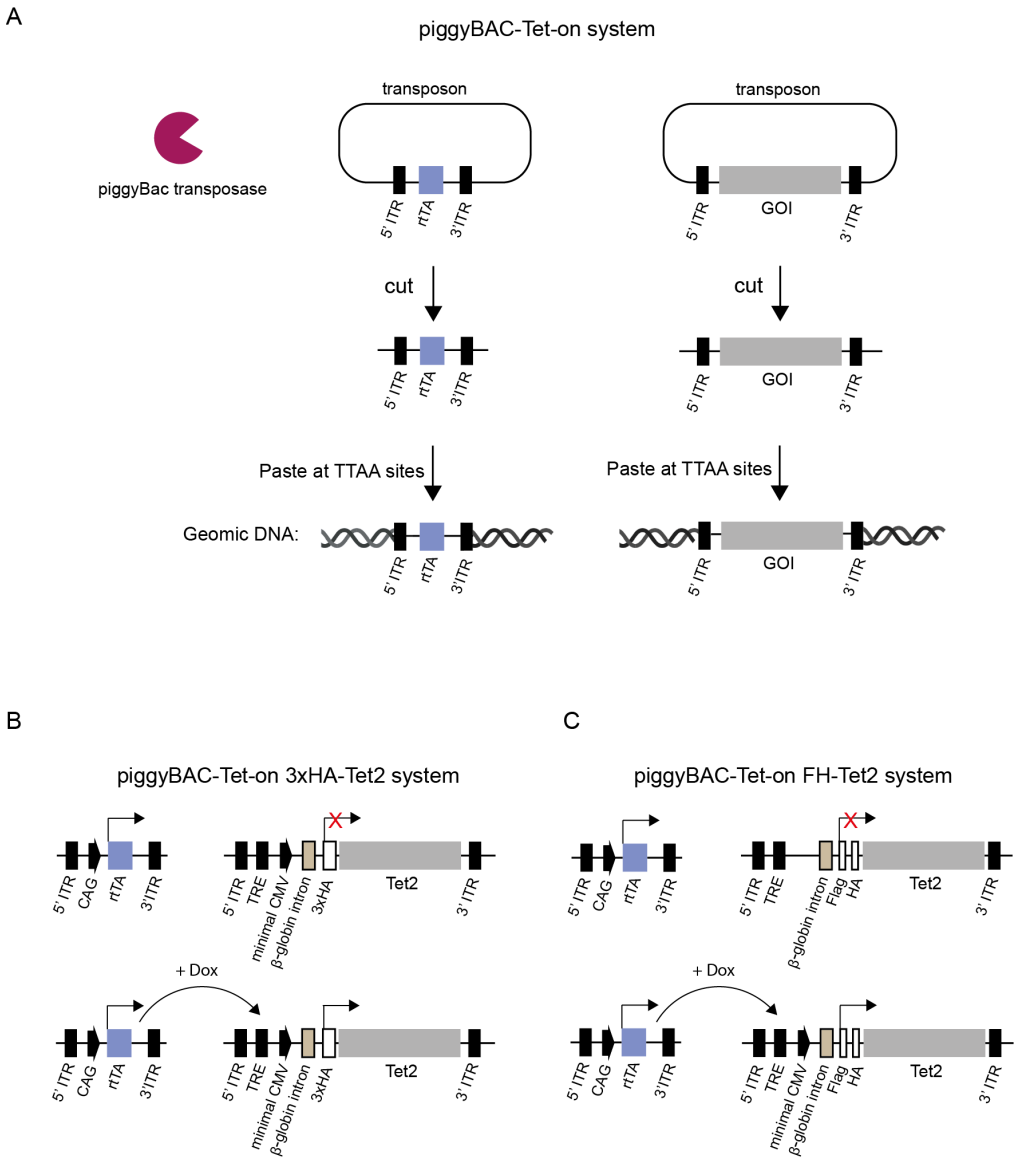


Figure 5.13: Construction of the inducible expression of tagged-Tet2 in HPC-7 *Tet2*-KO cells by piggyBAC-Tet-on system. Schematic representation of the piggyBAC-Tet-on system used in this study. (A) Tet-on system: rtTA and GOI are flanked by ITRs, followed by the piggyBac transposase-mediated excision, integrated at TTAA sites in the HPC-7 genome. (B) Constructs for inducible 3xHA Tet2 expression used in this study. (C) Constructs for inducible FH-Tet2 expression used in this study. rtTA binds to the TRE and activates the expression of mouse *Tet2* gene in the presence of dox. ITR: inverted terminal repeat; rtTA: reverse tetracycline-controlled transactivator; GOI: gene of interest; Dox: doxycycline; CAG: the CMV early enhancer/chicken actin promoter; TRE: tetracycline response element; minimal CMV: minimal cytomegalovirus promoter; β-globin intron: intron from rabbit β-globin gene; HA: human influenza hemagglutinin tag; Flag: DYKDDDDK peptide; Tet2: mouse *Tet2* gene.

5.13A). piggyBac is a DNA transposon, originally isolated from the cabbage looper moth *Trichoplusia ni* genome and has a large cargo size (Cary et al., 1989; Li et al.,

2011a). piggyBac has been used for the efficient transposition of its carried genes into the mammalian genome via a 'cut and paste' mechanism (Ding et al., 2005). Transposition is catalyzed by the transposase which recognizes the inverted terminal repeats (ITRs) of the transposon, and 'cuts' the DNA segment from the transposon vector and 'paste' it into the TTAA sites in the mammalian genome to achieve the stable integration of gene expression cassettes (Figure 5.13A). Tet-on system controls the timing and level of expression of gene of interest. The rtTA (transcriptional transactivator) was developed by the fusion of VP16 transcription activation domain and a mutant Tet repressor (Gossen et al., 1995). In the presence of tetracycline or its analogs (e.g. doxycycline), rtTA can bind to the TRE (tetracycline response element) and activate the expression of downstream gene. Two types of tagged-Tet2 piggBAC-Tet-on expression cassettes, 3xHA-Tet2 and FH-Tet2, were constructed and introduced into the HPC-7 *Tet2*-KO clone 2-1 and KO 2-7 respectively (Figure 5.13B and C).

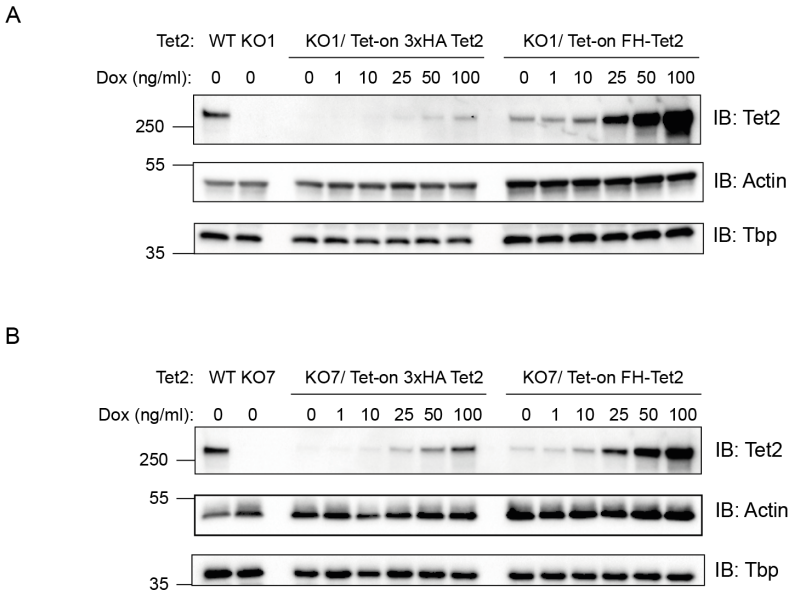


Figure 5.14: Western blot analysis of inducible expression of tagged-Tet2 in whole cell lysates from *Tet2*-KO HPC-7 cells. (A) Inducible expression of 3xHA Tet2 or FH-Tet2 in HPC-7 *Tet2*-KO clone 2-1 background. (B) Inducible expression of 3xHA Tet2 or FH-Tet2 in HPC-7 *Tet2*-KO clone 2-7 background. Cells were treated with indicated dox concentration for 24 h to induce the tagged-Tet2 expression before whole cell lyaste extraction. Tagged-Tet2 is detected by immunoblotting with anti-Tet2. Actin and Tbp serve as loading controls. WT: HPC-7; KO1: HPC-7 *Tet2*-KO clone 2-1; KO7: HPC-7 *Tet2*-KO clone 2-7. Dox: doxycycline.

Cells with stable integration of expression cassette were selected by the antibiotics

selection medium. *Tet2*-KO cells were treated with dox for 24 h to induce the expression of 3xHA Tet2 or FH-Tet2 protein (Figure 5.14). Tet2 was expressed in a dox concentration-dependent manner. One of the advantages of the inducible expression system is to control the expression level of protein by dox concentration. In this experiment, a level of tagged-Tet2 comparable to the endogenous WT protein was able to achieve by the indicated dox concentration (KO1/FH-Tet2: 10 ng/mL; KO7/3xHA-Tet2: 100 ng/mL; KO7/FH-Tet2: 25 ng/mL). Noticeably, 3xHA Tet2 was relatively lowly-expressed in *Tet2*-KO1 cells even if cells were treated with the highest dox dosage, which could be explained by the low amount of rtTA expressed in cells or low copy numbers of the integrated expression cassette (Figure 5.14A). Interestingly, FH-Tet2 started to be expressed even there was no dox present in the medium. This FH-Tet2 leaky expression might come from the minimal promoter activity in the constructs.

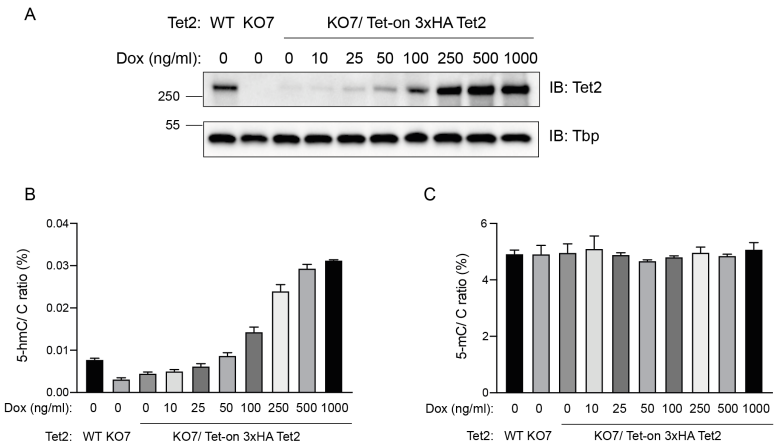


Figure 5.15: Characterization of 3xHA Tet2-expressing HPC-7 *Tet2*-KO 2-7 cells. (A) Western blot analysis of inducible expression of 3xHA Tet2 in whole cell lysates from HPC-7 *Tet2*-KO 2-7 by piggyBAC-Tet-on system. Cells were treated with indicated dox concentration for 24 h to induce the 3xHA Tet2 expression prior to Western blot analysis. 3xHA Tet2 is detected by immunoblotting with anti-Tet2. Tbp serves as the loading control. (B) LC-MS/MS measurements of genomic 5-hmC level (left) and 5-mC level (right) in 3xHA Tet2-expressing HPC-7 *Tet2*-KO 2-7 cells. Cells were treated with indicated dox concentration for 24 h, followed by genomic DNA extraction, hydrolysis and LC-MS/MS quantification. LC-MS/MS quantification was provided by Paolo Spingardi. Mean values from three biological replicates are plotted. Error bars indicate standard deviation.

Re-expression of tagged-Tet2 resulted in increased 5hmC in *Tet2*-KO cells (Figure 5.15B). When cells were treated with 100 ng/mL dox, 3xHA Tet2 was expressed in a level close

to the endogenous protein, however, 5hmC induced at this dox concentration was higher than the level in WT cells.

5.2.4 Anti-HA immunoprecipitation (IP) to identify Tet2 interacting proteins in HPC-7 cells

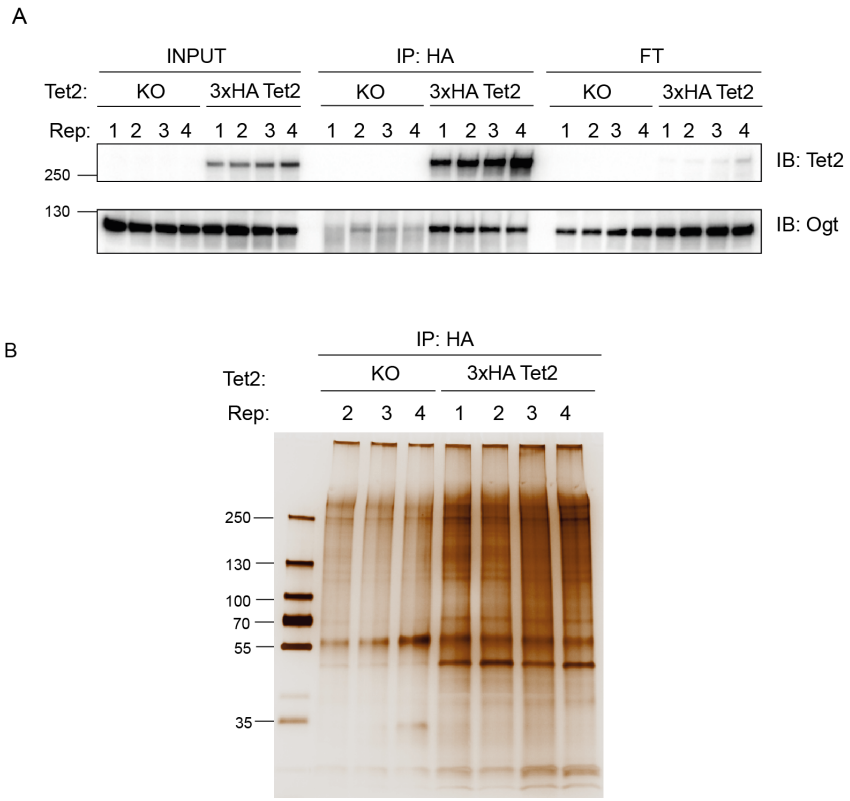


Figure 5.16: Characterization of anti-HA IP-mass spectrometry HPC-7 *Tet2*-KO 2-7 Tet-ON 3x HA Tet2 cells were treated with 250 ng/mL dox for 24 h prior to nuclear protein extraction, anti-HA immunoprecipitation, and LC-MS/MS analysis. Anti-HA IP with nuclear proteins from HPC-7 *Tet2*-KO 2-7 was used as the negative control for Tet2 interactome mapping. Four biological replicates were prepared and submitted for MS analysis. (A) Western blot analysis of anti-HA IP-mass samples. A known Tet2 interactor, Ogt, is tested in the anti-HA IP samples. IP: immunoprecipitation; FT: flowthrough. (D) Silver stain of anti-HA IP-mass spec samples. 5% IP samples were analyzed by SDS-PAGE, followed by silver staining. Replicate 1 in KO group is an outlier and excluded from downstream analysis

To identify Tet2 interacting partners in HPC-7, *Tet2*-KO7/ Tet-on 3xHA Tet2 cells were cultured with 250 ng/mL dox for 24 hr to induce the expression of 3xHA Tet2 near the endogenous level, followed by nuclear protein extraction. To prepare one replicate for IP-MS, 5 mg nuclear lysates were incubated with anti-HA antibody overnight, followed by 3 hr beads incubation, to capture Tet2 and its interacting complex. Interacting complexes

were eluted from the beads in SDS sample buffer, followed by Mass-spec identification. Anti-HA IP on *Tet2*-KO7 cells was prepared as control for antibody non-specific binding and beads matrix binding. 4 biological replicates were prepared for each group. IP-MS samples were analyzed by Western blot before submission for Mass-spectrometry identification (Figure 5.16A). Western blot analysis showed that 3xHA Tet2 was expressed in a similar level among the four biological replicates. Tet2 and its known interacting partner OGT were identified by anti-HA IP, reflecting the strength of affinity purification-based approach for Tet2 interaction identification. Silver stain revealed the different protein pattern present in ant-HA IP samples between 3xHA Tet2 cells and *Tet2*-KO cells (Figure 5.16B).

5.3 Discussion

Tet2 is well-known for its function in the iterative oxidation of cytosine methylation in mammalian cells, this provides an mechanism for active DNA demethylation. Tet2 mutations were frequently identified in patients with hematological malignancies and healthy individuals with clonal hematopoiesis, indicating an important role of TET2 in normal and malignant hematopoiesis. In this chapter, I focused on the understanding of Tet2 interacting partners in the hematopoietic context. Hematopoietic progenitor cell, HPC-7, was chosen for the hematopoietic-relevant Tet2 interactome mapping.

Compared to the classical affinity purification-based interactome mapping approach, proximity labelling-based method has advantages in capturing weak and transient protein-protein interaction. BioID was not functional in our study since Tet2-BirA* fusion protein did not lead to induction of active biotinylation labelling and failed to label the known Tet2 interactor OGT *in vivo* (Figure 5.2). Instead, APEX2 labelling assay was able to validate the reported interaction between Tet2 and OGT in 293 T cells (Figure 5.5). CRISPR-Cas9 mediated APEX2 knock-in at the endogenous *Tet2* locus in HPC-7 was firstly attempted and no true targeted clone was isolated at the end (Data not shown). APEX2-Tet2 fusion protein can not be expressed by MSCV retrovirus expression system and unexpected high level of endogenous biotinylation signal was observed in WT HPC-7

cells during APEX2 assay, which could hamper the strength of proximity labelling for interaction mapping (Figure 5.8). Protein low expression also occurred when using the retroviral system to expression FH-Tet2 (Figure 5.12). This might suggest an inefficiency of the retrovirus system in expressing large construct. In the end, I managed to achieve an inducible tagged-Tet2 expression in the *Tet2*-KO HPC-7 cells by using the piggyBAC-Tet-on system (Figure 5.14). 3xHA-Tet2 was expressed in a level close to the endogenous protein. To identify Tet2 interactome in HPC-7 cells, anti-HA IP was performed in *Tet2*-KO cells with the re-expression of ectopic tagged-Tet2 (250 ng/mL dox), followed by Mass-spec protein identification (Figure 5.16). In the next chapter, I will introduce the IP-MS results of the identified Tet2 interactome in HPC-7 cells.

6

Validation of novel Tet2 interactors

Contents

6.1 Introduction

In this chapter, I will introduce the IP-MS results of Tet2 interactome mapping in HPC-7 cells (Chapter 6.2.1). To validate the IP-MS-identified novel Tet2 interactions, I firstly did the reciprocal co-IP in HPC-7 cells to test if the interactor-IP could capture its interaction with Tet2 protein (Chapter 6.2.2). IP using the antibody against Tet2 protein was attempted to explore the interaction between the endogenous Tet2 protein and identified interactors in HPC-7 cells (Chapter 6.2.2). Finally, I identified the domain of Tet2 responsible for its interaction with novel interactors (Figure 6.2.3).

6.2 Results

6.2.1 Tet2 interactome landscape in HPC-7 cells

To identify Tet2 interactors in HPC-7 cells, four biological replicates of anti-HA IP on 3xHA Tet2-expressing cells (dox: 250 ng/ml, 24h) and IP on *Tet2*-KO cells were submitted for mass-spec analysis. Replicate 1 of IP on *Tet2*-KO cells was identified as an outlier and excluded from the downstream analysis. The reproducibility of biological replicates were assessed by PCA plot in in Perseus (v1.5.2.4) (Tyanova et al., 2016) (Figure 6.1A). Replicates within the same sample group showed good consistency of the protein profile identified by MS since replicate samples were grouped together. To discover proteins significantly identified in IP on 3xHA Tet2 cells over KO cells, the raw protein peptide intensities were exported and analysed in Perseus. The data was log2 transformed, normalized by subtracting the median from the columns (each sample) and then missing values were imputed. The difference (ratio of the normalized protein intensity) of identified proteins between 3xHA Tet2 cells and *Tet2*-KO cells were examined using two-samples test (Figure 6.1B). Importantly, Tet2 itself and the reported Tet2 interactor, OGT (Chen et al., 2013; Deplus et al., 2013), was discovered as a protein significantly enriched (FDR<0.05, Difference<0) in 3xHA Tet2 cells group,

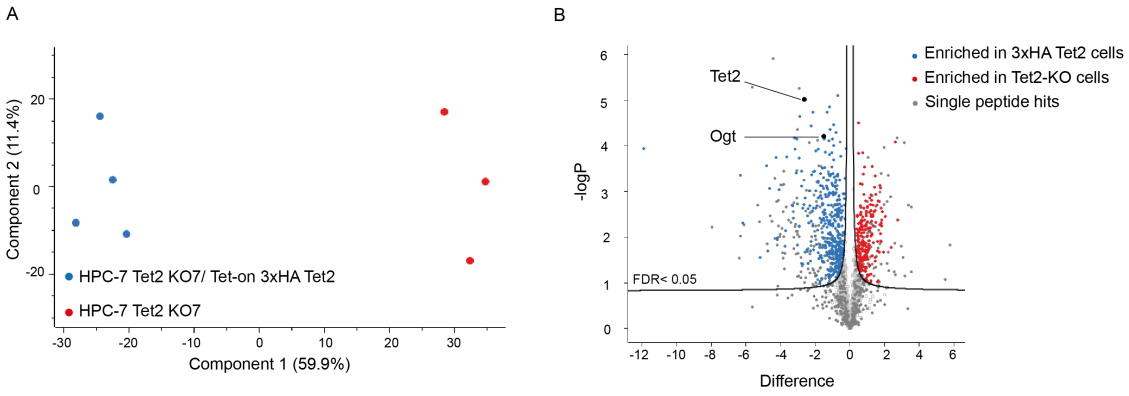


Figure 6.1: Overview of anti-HA IP-MS for Tet2 interactome mapping in HPC-7. (A) Principal component analysis (PCA) plot of anti-HA IP-MS samples. (B) Volcano plot showing the false discovery rate (FDR) and the fold enrichment values of protein abundance quantified by MS. Proteins showing significantly differential enrichment (FDR<0.05) in 3xHA Tet2 group (Blue) and in *Tet2*-KO group (Red) were indicated. The difference between samples was compared using two-sided t-test.

demonstrating that our IP-MS experiments were able to capture the bait of the MS and the known interaction. In total, 380 proteins were shown to be significantly enriched in 3xHA Tet2 IP group.

To identify high confidence Tet2 interacting partners, a few criteria have been applied to filter the IP-MS data (Figure 6.2). Total 1891 proteins were identified by IP-MS. Proteins with two-samples test p value >0.05 and difference value >0 were discarded. Difference represents the ratio of the normalized protein intensity between 3xHA Tet2 and Tet2 KO group. Difference <0 represents enrichment in 3XHA Tet2 group. Protein detected with only one peptide, common MS contaminants e.g. actin, keratin, and ribosomal proteins were removed from the list. The rest proteins were ranked by confidence score which reflects the numbers of peptide identified. In the end, 11 proteins with high confidence score and reported nuclear function were selected for further biochemical validation (Table 6.1).

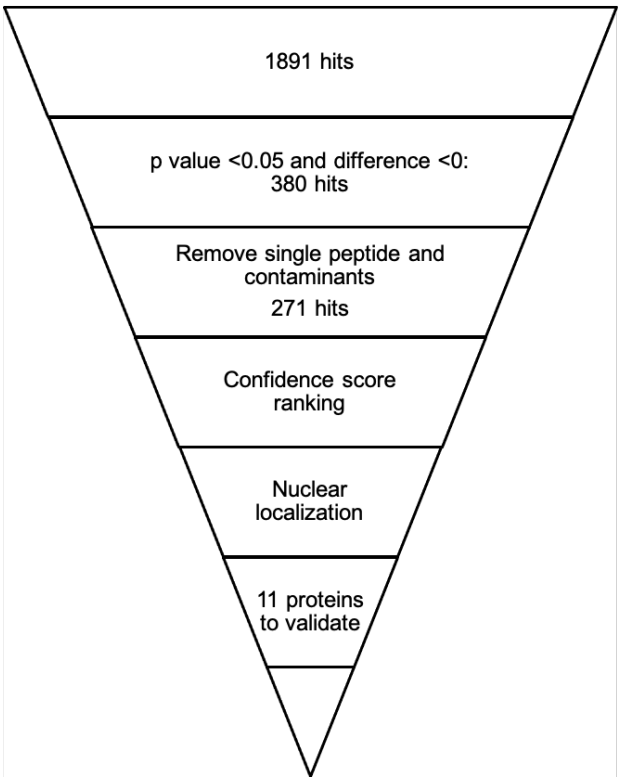


Figure 6.2: LC/MS-MS data filtering strategy for downstream HPC-7 Tet2 interaction validation.

Table 6.1: Tet2 high-confidence interacting proteins in HPC-7 cells.

Gene	Function	Confidence score	Peptide count	Unique peptide	Anova (p) value	Difference
Tet2	Methylcytosine dioxygenase TET2	3455.25	47	34	0.0001	2.67
Ppp1r9b	subunit of phosphatase1a	2995.68	41	35	0.0071	3.04
Noc2l	an inhibitor of histone acetyltransferase activity	1541.5	22	19	0.0055	1.91
OGT	UDP-N-acetylglucosamine–peptide N-acetylglucosaminyltransferase	1495.58	23	16	0.0058	1.53
Mndal	interferon-inducible p200 (IFI-200) family protein	1466.38	21	8	0.0375	1.42
Fyb1	an adapter protein of the FYN and LCP2 signaling cascades in T-cells	1464.89	26	24	0.0228	1.97
Suz12	Polycomb group (PcG) protein	1389.32	19	17	0.0496	0.49
Rae1	mRNA transport	1027.98	14	11	0.0321	1.16
Rfc5	subunit of the replication factor C complex	991.17	13	10	0.0075	1.97
Rbbp7	core histone-binding subunit	870.63	15	7	0.0017	1.84
Ifi204	Ifi-200 family protein	757.03	12	1	0.0437	2.20
Alyref	chaperone for transcription factors containing basic leucine zipper (bZIP) domains	639.21	10	5	0.0018	2.04
Ezh2	Polycomb group (PcG) protein	227.6	4	3	0.0169	3.72

6.2.2 Validation of novel Tet2 interactions in HPC-7 cells

To validate the 11 novel Tet2 interactors discovered by mass-spec. I firstly examined the existence of the 11 novel interactors in the anti-HA IP on 3xHA Tet2 cells by Western blot (Figure 6.3). Except Noc2l and Mndal which don't have a good antibody used for Western blot, all the other identified Tet2 interactors were confirmed to be found in our experiments. Although there was a background signal observed in KO cell control group when detecting Ogt, Ppp1r9b and Fyb1, the absence of MeCP2 which has no reported interaction with Tet2 in the IP sample indicates that the IP condition is able to exclude some non-specific binding. Taken together, I was able to validate the existence of 9 out of the 11 Tet2 interactors. And in the following experiments, I further focused on the Suz12, Ezh2, Rbbp7, Ppp1r9b and Alyref which have antibody available for IP, thus could allow us to do reciprocal IP to validate their interaction with Tet2 in HPC-7 cells.

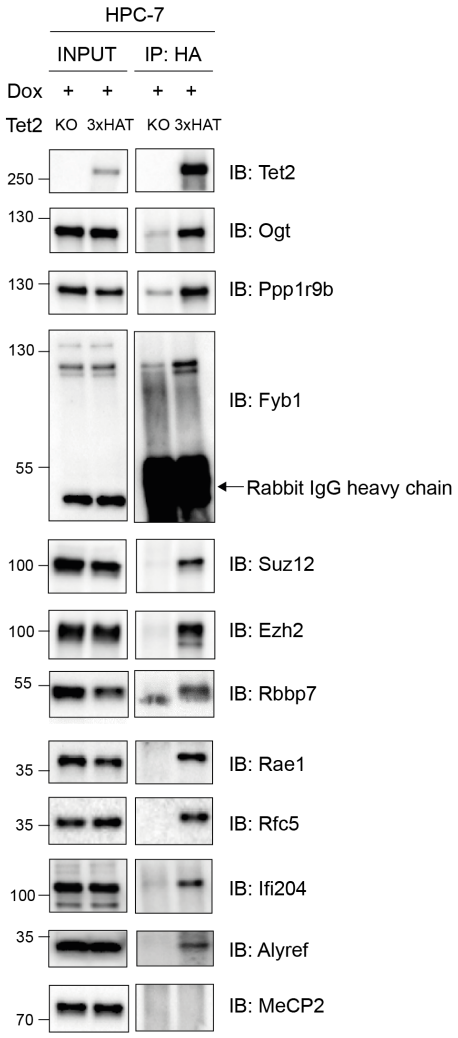


Figure 6.3: Confirmation of the identified Tet2 interactors by anti-HA immunoprecipitation with HPC-7 cells expressing 3xHA Tet2. HPC-7 *Tet2*-KO 2-7/ Tet-on 3xHA Tet2 cells and *Tet2*-KO 2-7 cells were both treated with 250 ng/mL dox for 24 hr, followed by nuclear protein extraction and anti-HA IP purification. IP on *Tet2*-KO cells and anti-MeCP2 detection by IB serve as negative control for antibody non-specific binding.

PRC2 interact with Tet2 in HPC-7 cells

Among the identified HPC-7 Tet2 interactors, Suz12, Ezh2 and Rbbp7 were subunits of core PRC2 (Polycomb Repressive Complex 2) (Hauri et al., 2016). Interaction between PRC2 complex and Tet2 was confirmed by reciprocal co-IP using antibody against the polycomb proteins in HPC-7 cells (Figure 6.4). Here, I performed the reciprocal co-IP in two cell types contexts. Suz12, Ezh2 and Rbbp7 were firstly shown to interact with 3xHA Tet2 in HPC-7 cells which is the original cell type used for anti-HA-Tet2 IP-MS (Figure 6.4A). To exempt the possibility that observed interaction was a result based on the ectopic

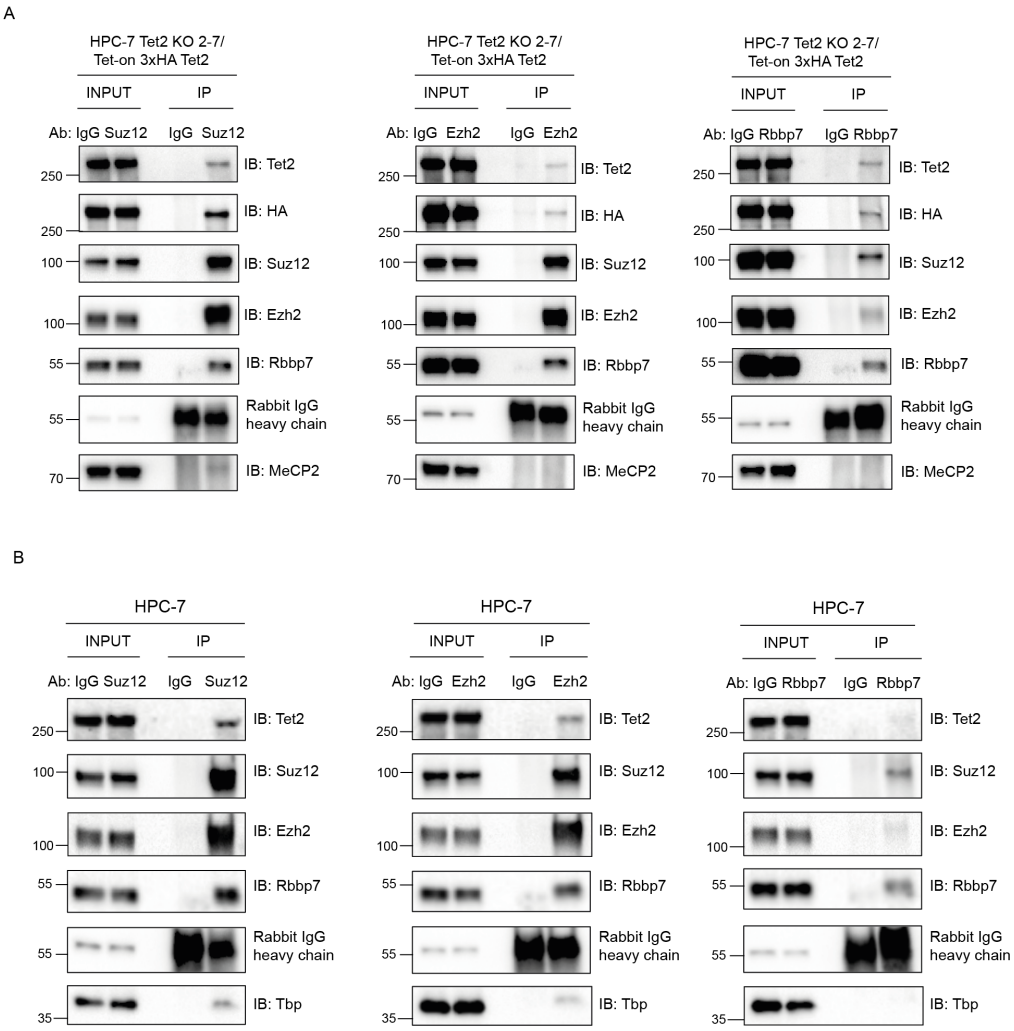


Figure 6.4: PRC2 interact with 3xHA Tet2 in HPC-7 cells.(A) Western blot analysis of reciprocal co-IP with antibody specific for PRC2 complex on nuclear lysates of HPC-7 *Tet2*-KO 2-7/ Tet-on 3xHA Tet2 cells (250 ng/mL dox for 24 h). (B) Western blot analysis of reciprocal co-IP with antibody specific for PRC2 complex on HPC-7 nuclear lysate. IgG, anti-MeCP2 and anti-TBP detection by immunoblotting serve as negative controls for IP.

HA-Tet2 protein, I further performed the reciprocal co-IP in WT HPC-7 cells (Figure 6.4B). Consistently, Suz12 and Ezh2 showed clear interaction with the endogenous Tet2 in HPC-7 cells. In the IP experiments, both MeCP2 and Tbp were considered as control for IP non-specific background since they don't have reported interaction with Tet2. The observation of Tbp in the IP samples might suggest more stringent wash condition would be required to remove non-specific binding. Despite the interaction between endogenous Tet2 and Rbbp7 was not found in WT HPC-7 cells (Figure 6.4B Right), we could not conclude that Rbbp7 does not interact with endogenous Tet2 since the protein antibody

showed poor performance in the enrichment of the IP bait, Rbbp7 itself and could not clearly characterize the interaction between Rbbp7 and the other PRC2 components, such as, Ezh2.

Ppp1r9b and Alyref interact with Tet2 in HPC-7 cells

To validate the interaction between Tet2 and Ppp1r9b, Alyref, I found antibody against Ppp1r9b and Alyref which could be used for IP and performed the reciprocal IP in the WT HPC-7 cells (Figure 6.5). The data suggest that Ppp1r9b and Alyref could interact with endogenous Tet2 in HPC-7 cells respectively.

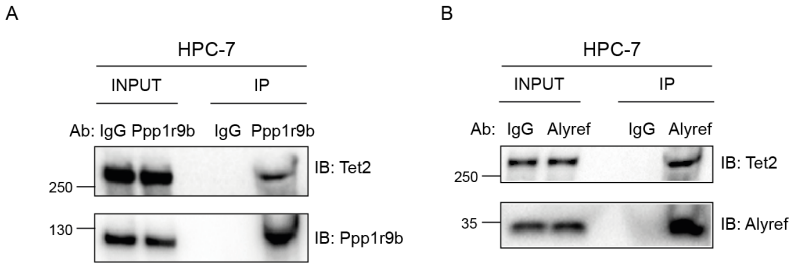


Figure 6.5: Ppp1r9b and Alyref interact with endogenous Tet2 in HPC-7 cells (A) Western blot analysis of reciprocal co-IP from HPC-7 nuclear lysates with antibody specific for Ppp1r9b. (B) Western blot analysis of reciprocal co-IP from HPC-7 nuclear lysates with antibody specific for Ppp1r9b.

Endogenous Tet2 co-immunoprecipitation in HPC-7

Lack of a reliable Tet2 antibody for IP-based applications has brought difficulties in mapping endogenous Tet2 interactome. A recent paper from Helin's group has reported a new antibody raised against the N-terminal of mouse Tet2 protein. ChIP-seq experiments were successfully performed by using this Tet2-N antibody (Rasmussen et al., 2019). We requested some aliquot of this Tet2-N antibody and tested it by co-IP experiments in HPC-7 cells. Endogenous Tet2 protein was able to be pulled down by the Tet2-N antibody (Figure 6.6). The known Tet2 interactor, OGT, was able to be detected in the IP samples. However, high background of IP signal was found in IP on *Tet2*-KO cells but not in IgG control group. This make it difficult to draw conclusions from the experiments about whether Tet2 interacts with identified new interacting proteins. To be able to conclude, I have to include more wash steps to reduce the non-specific binding since Tbp was present

in the IP samples. If the background signal still existed in the KO cell samples but not the IgG control group, this could reflect that non-specificity of the Tet2-N antibody in recognising the Tet2 protein.

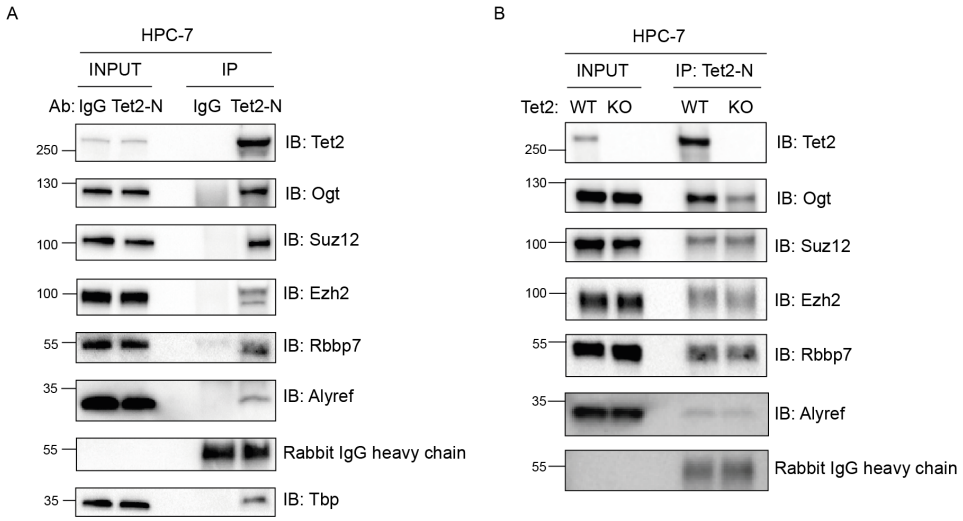


Figure 6.6: Western blot analysis of endogenous Tet2 immunoprecipitation with Tet2-N antibody in HPC-7 cells. (A) Nuclear lysates from HPC-7 cells were immunoprecipitated with the antibody specific to the N-terminal of Tet2 protein (A gift from Helin's group). Immunoprecipitated proteins were detected by Western blotting with indicated antibodies. IgG and anti-Tbp detection by immunoblotting serve as negative controls for IP. (B) Nuclear lysates from HPC-7 or HPC-7 *Tet2*-KO 2-7 were immunoprecipitated with the antibody specific to the N-terminal of Tet2 protein. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies.

6.2.3 PRC2 interact with the C-terminus of Tet2

Previous results from co-IP experiments validated the interaction between PRC2 and Tet2 in HPC-7. However, the observation of Tbp in the IP samples brought us concerns about the specificity of the experiments. Although the IgG isotype control could reflect the non-specific background to some extent, the best control for IP experiments would be cells lack of the expression of the bait protein. The difficult-to-transfect nature of the HPC-7 cells and the time-consuming process to construct KO cells lines made us move our validation experiments into the 293 T cells which are much easier to achieve transient gene expression or gene knock-down. Furthermore, I generated constructs to express tagged version of Tet2 and the identified novel interactors and utilized the co-expression experiments in 293 T cells to facilitate the validation of the interaction without relying

on the availability of the protein antibody. And to gain insight into the domain of Tet2 responsible for the interaction, the N-terminus (residue 1-1041) and C-terminus Tet2 (residue 1041-1912) were co-expressed with 3xHA-interactor (Suz12, Ezh2 and Rbbp7, Ppp1r9b, Alyref) in 293 T cells (Figure 6.7. The C-terminus includes the entire catalytic domain (CD) of the Tet2 protein.

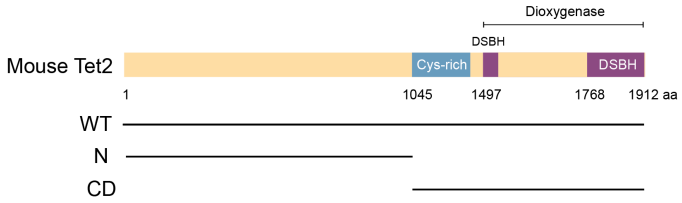


Figure 6.7: Schematic representation of mouse full-length Tet2 Cys-rich: cysteine-rich domain; DSBH: double-stranded beta-helix domain; WT: wildtype; N: N terminus; CD: C-terminal cysteine-rich dioxygenase domain.

The data showed that Suz12 interacted with the C-terminus of Tet2 but not the N-terminus (Figure 6.8A). Interaction between Suz12 and Tet2 were also confirmed by Suz12 reciprocal co-IP (Figure 6.8B). Similarly, the other components of core PRC2 complex, Ezh2, was also found to interact with the C-terminus of Tet2, which was suggested by the Flag-Tet2 co-IP and HA-Ezh2 reciprocal co-IP experiments (Figure 6.9). Taken together, the C-terminus of Tet2 is responsible for the interaction with the PRC2

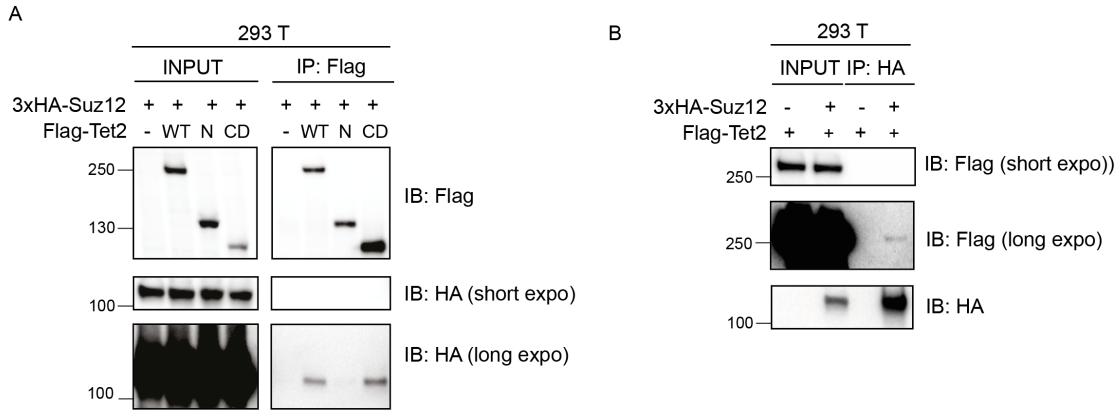


Figure 6.8: Suz12 interacts with the C terminus of Tet2 (A) Flag-tagged Tet2 truncation was transiently co-expressed with 3xHA Suz12 in 293T cells. 24 h post transfection, whole cell lysates were immunoprecipitated with Flag M2 magnetic beads. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies. (B) 3xHA Suz12 was transiently co-expressed with Flag-Tet2 in 293 T cells. 24 h post transfection, whole cell lysates were extracted and immunoprecipitated with antibody specific for HA tag. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies.

complex. Indeed, it is clear that the amounts of PRC2 proteins and Tet2 within the same complex is not that substantial since the amount of proteins in the IP samples was not highly enriched over the input. On one hand, this might suggest that Tet2-PRC2 is not the dominant form of either Tet2 or PRC2 protein complex to present in the 293 T cells, since the existence of other proteins in cells might occupy and compete with the formation of the PRC2 and Tet2 complex. On the other hand, the low IP signal over the input might suggests the interaction between PRC2 proteins and Tet2 might require the comparable expression of other co-factors which are important to stabilize the PRC2-Tet2 complex.

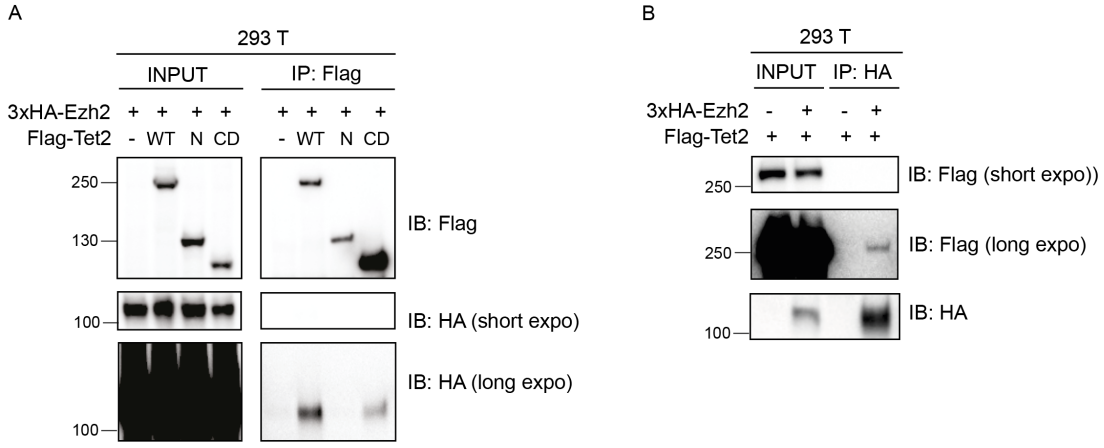


Figure 6.9: Ezh2 interacts with the C terminus of Tet2 (A) Flag-tagged Tet2 truncation was transiently co-expressed with 3xHA Ezh2 in 293T cells. 24 h post transfection, whole cell lysates were immunoprecipitated with Flag M2 magnetic beads. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies. (B) 3xHA Ezh2 was transiently co-expressed with Flag-Tet2 in 293 T cells. 24 h post transfection, whole cell lysates were extracted and immunoprecipitated with antibody specific for HA tag. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies.

Although Rbbp7 seems to be pulled-down by anti-Flag IP on FL Tet2, a background signal was observed in the control group where there was no flag protein expressed but had the expression of HA-Rbbp7 (Figure 6.10A), which made it difficult to distinguish the interaction from the non-specific background. Moreover, little difference of the Rbbp7 IP signal was observed between N-terminus Tet2 and C-terminus Tet2. Finally reciprocal co-IP showed that 3xHA Rbbp7 was not likely to interact with Tet2 in 293 T cells, indicating that Rbbp7 is not the main component in the PRC2 complex to be responsible for the interaction with Tet2 (Figure 6.10B). In the IP experiments, the detection of OGT

by Western blot in Flag-Tet2 co-IP experiment was used as a control since OGT was reported to interact with the C-terminus of Tet2 (Vella et al., 2013).

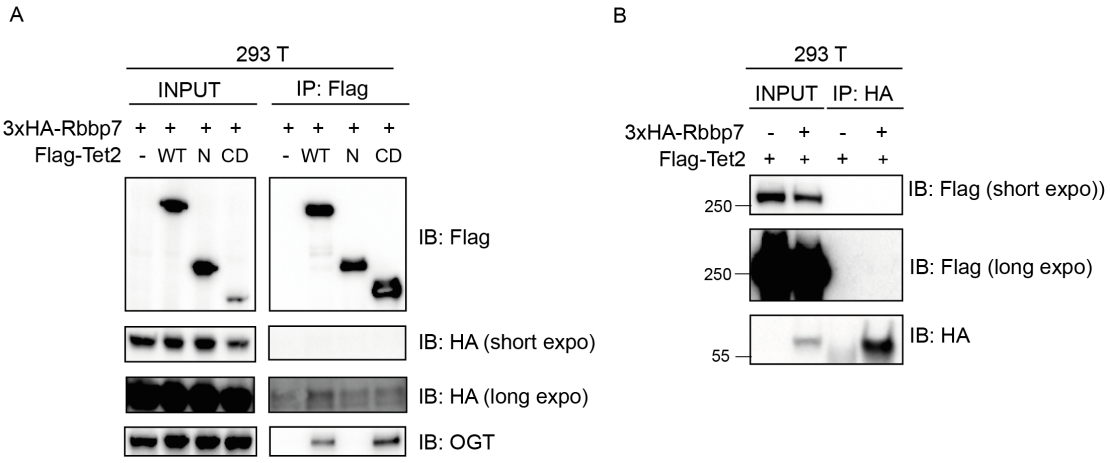


Figure 6.10: Rbbp7 does not interact with Tet2 in 293T cells (A) (A) Flag-tagged Tet2 truncation was transiently co-expressed with 3xHA Rbbp7 in 293T cells. 24 h post transfection, whole cell lysates were immunoprecipitated with Flag M2 magnetic beads. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies. (B) 3xHA Rbbp7 was transiently co-expressed with Flag-Tet2 in 293 T cells. 24 h post transfection, whole cell lysates were extracted and immunoprecipitated with antibody specific for HA tag. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies.

In addition to the PRC2 proteins, I also performed the co-expression experiments followed by co-IP in 293 T cells to examine the interaction between Tet2 and Ppp1r9b and Alyref, which were shown to interact with endogenous Tet2 in HPC-7 cells. However, I was not able to conclude the interaction of Tet2 with Ppp1r9b or Alyref from the experiments since the anti-Flag co-IP and anti-HA co-IP could not lead to a clear conclusion. Alyref seems to strongly bind to the anti-Flag beads even though OGT did not show any background in the cells expressing Alyref alone (Figure 6.11A), which interfere our understanding of the interaction. Reciprocal co-IP did not support the interaction between Tet2 and Alyref, indicating that Alyref was not likely to interact with Tet2 in 293 T cells. Although anti-Flag co-IP was not able to show the interaction between Ppp1r9b and Tet2, a weak Flag-Tet2 band was found in anti-HA IP, which provides a likelihood of interaction between Tet2 and Ppp1r9b.

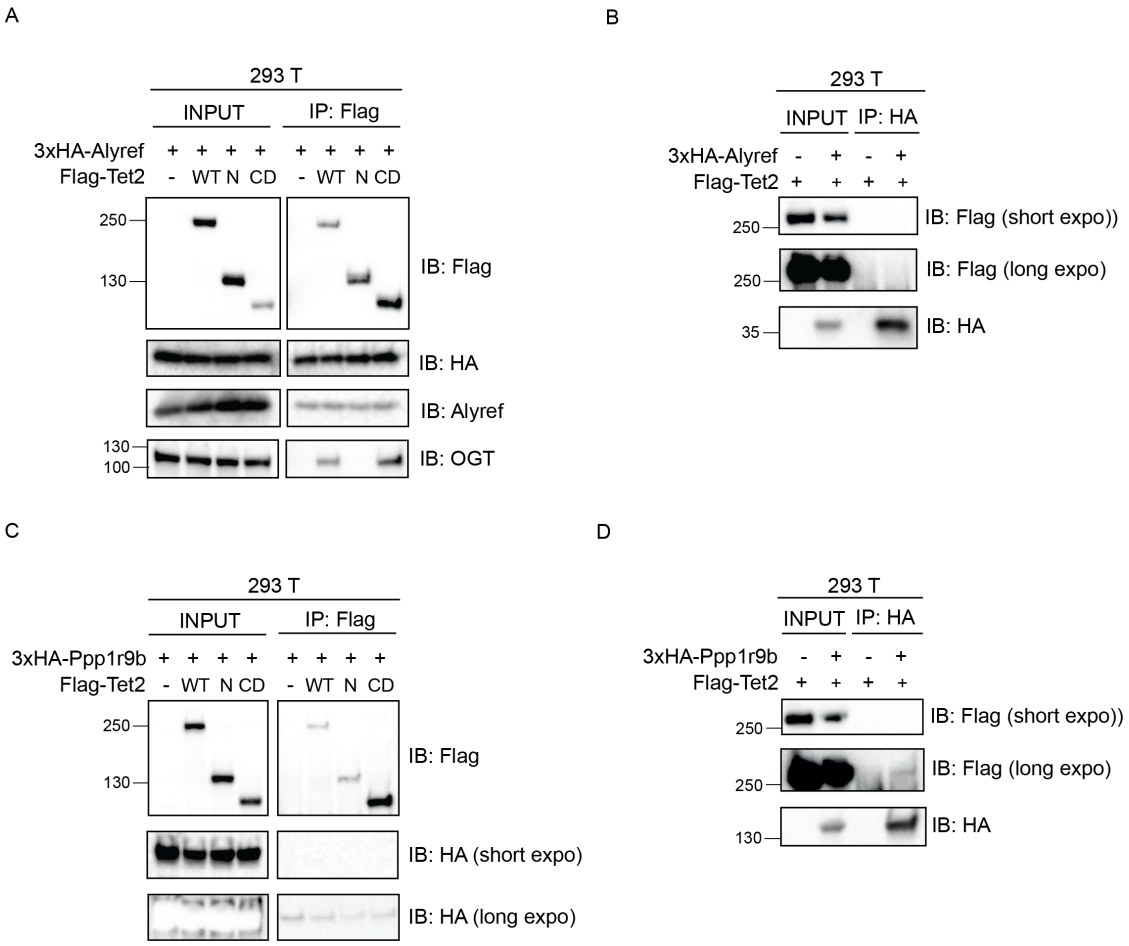


Figure 6.11: Validation of Tet2 interaction with Alyref and Ppp1r9b in 293 T cells. Flag-tagged Tet2 truncation was transiently expressed with 3xHA Alyref (A) or 3xHA Ppp1r9b (C) in 293T cells. 24 h post transfection, whole cell lysates were extracted and immunoprecipitated with Flag M2 magnetic beads. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies. 3xHA Alyref (B) or 3xHA Ppp1r9b (D) was transiently expressed with Flag-tagged full-length Tet2 in 293 T cells. 24 h post transfection, whole cell lysates were extracted and immunoprecipitated with antibody specific for HA tag. Immunoprecipitated proteins were detected by Western blotting with indicated antibodies.

6.3 Discussion

In this chapter, I identified 11 putative Tet2 interactors in HPC-7 cells based on the IP-MS results and performed the interaction validation assay in HPC-7 cells and 293 T cells to discover true Tet2 interators. First, in order to discover proteins significantly identified in IP on 3xHA Tet2 cells over KO cells, the raw protein peptide intensities were exported and analysed in Perseus. The difference (ratio of the normalized protein intensity) of identified proteins between 3xHA Tet2 cells and *Tet2*-KO cells were calculated and examined

using two-samples test (Figure 6.1B). Importantly, Tet2 and its known interactor, OGT (Chen et al., 2013; Deplus et al., 2013), were both recognized as a protein significantly enriched (FDR<0.05, Difference<0) in 3xHA Tet2 cells group, demonstrating the strength of our IP-MS experiments in capture of the bait of MS and the known interaction. To identify high confidence Tet2 interacting partners, a few criteria have been applied to filter the IP-MS data (Figure 6.2). In total 1891 proteins were identified by IP-MS. Among the 380 proteins which showed significant enrichment (two-samples test p value <0.05 and difference value <0) of protein signal in 3XHA Tet2 group, proteins detected with only one peptide, common MS contaminants were further removed from the list. Finally I focused on 11 proteins (Table 6.1) which exhibited high confidence score (related to the numbers of peptide identified) in the Mass-spec results and had known nuclear function. The 11 putative Tet2 interactors were confirmed to be present in the anti-HA IP samples by Western blot (Figure 6.3).

Among the 11 putative Tet2 interactors, Suz1, Ezh2, and Rbbp7 were subunits of core PRC2 (Polycomb Repressive Complex 2) which is associated with gene repression (Hansen et al., 2008). I was able to detect the interaction between PRC2 (Suz12 and Ezh2) and endogenous Tet2 in the HPC-7 cells via PRC2 protein co-IP experiments (Figure 6.4). And the C-terminus of Tet2 was found to be responsible for the interaction with Suz12 and Ezh2 (Figure 6.8,6.9). The interaction between Tet2 and Rbbp7 was not able to conclude since the Rbbp7 antibody does not show a good enrichment of Rbbp7 itself in IP in HPC-7 cells (Figure 6.4), and Rbbp7 is less likely to interact with Tet2 in the co-expressing experiments in 293 T cells (Figure 6.10). Moreover, the interaction between PRC2 proteins and Tet2 was not very abundant in 293 T cells since the amount of proteins in the IP samples was not highly enriched over the input. On one hand, this might suggest that Tet2-PRC2 is not the dominant form of either Tet2 or PRC2 to present in the 293 T cells, since the expression of other proteins in the cells could occupy and compete the formation of the PRC2 and Tet2 complex. On the other hand, the low IP signal over the input might suggest the interaction between PRC2 proteins and Tet2 would require the comparable expression of the co-factors which are important for the stabilization of the PRC2-Tet2 complex.

Previous studies have already linked PRC2 function to Tet protein. Genome-wide study in mouse ESCs has revealed the co-localization between Tet1 and PRC2 (Suz12, Ezh2 and Sin3A) at the promoter region of bivalent genes (H3K4me3 and H3K27me3) (Neri et al., 2013; Williams et al., 2011). Suz12 and EED, subunits of core PRC2, were found to physically interact with Tet1 in mouse embryonic stem cells but not differentiated cells, e.g. MEF and HEK 293T cells (Neri et al., 2013; Williams et al., 2011). The interaction between PRC2 and Tet1 was responsible for the recruitment of Tet1 and its 5hmC deposition activity since Suz12 knock-down led to reduced Tet1 binding and decreased 5hmC at bivalent gene promoter. 5hmC frequency correlates with histone modification H3K27me3 and PRC2 binding in a cell-type specific manner since the genomic co-localization was only identified in mESCs but in MEF cells, brain and liver cells. On the other hand, a number of observations have also pointed out the role of DNA methylation in negative regulation of PRC2 chromatin recruitment (Lindroth et al., 2008; Lynch et al., 2012; Mendenhall et al., 2010). Tet1-catalyzed DNA demethylation was suggested to contribute to gene silencing by recruiting PRC2 complex (Wu et al., 2011b). Moreover, DNA methylation is also required for the proper gene repression mediated by PRC2 (Reddington et al., 2013). The cross-talk between Tet1-involved DNA demethylation and PRC2-mediated gene repression offers a dynamic mechanism for activating or repressing the targeted gene expression.

In hematopoietic progenitor cells, HPC-7 cells, we also observed the cross-talk between Tet family and PRC2 complex supported by the identified physical interaction between Tet2 and PRC2. And the interaction with PRC2 is dependent on Tet2 catalytic domain which is highly conserved among Tet family proteins, this might provide a structural basis for the interaction between Tet proteins and PRC2. Previously published IP-MS experiments in HEK 293 T cells failed to observe the interaction between Tet2 and PRC2 (Chen et al., 2013; Deplus et al., 2013), we also found the PRC2-Tet2 interaction was not substantially present in the 293 T cells. These results might suggest the functional interaction between Tet2 and PRC2 is in a cell-type specific manner and might be mediated by HPC-7-specific factors or dependent on certain post-translation modifications. In terms of the potential function of the Tet2-PRC2 interaction in HPC-7, the previous

RNA-seq on *Tet2*-KO HPC-7 cells and Tet2 ChIP-seq experiments have pointed out the repressive role of Tet2 in gene regulation. The interaction with PRC2 might provide a possible explanation for Tet2-related gene repression. Genome-wide analysis to assess the overlap of Tet2-PRC2 binding in HPC-7 cells would be necessary to understand the global co-operation between Tet2 and PRC2 in gene regulation. More importantly, like TET2, Ezh2 was also found as a tumor suppressor gene in myeloid malignancies. *Ezh2* mutation was reported in patients with myeloid dysplasia (Neri et al., 2013). *Ezh2* depletion in mice resulted in enhanced repopulating capacity of HSCs and promoted cell differentiation towards myeloid lineage which is similar to what observed in *Tet2*-deficient mice (Muto et al., 2013). This might indicate a potential co-operation between Tet2 and PRC2 in the pathological hematopoiesis.

In addition to the PRC2 complex, I also found the interaction between Tet2 and Ppp1r9b or Alyref in HPC-7 cells. Recent study has identified Alyref as a RNA 5mC reader (Yang et al., 2017). Alyref was originally found to interact with LEF-1 and acute myeloid leukemia (AML)-1 transcription factors leading to the transcriptional activation of the T-cell receptor gene (Bruhn et al., 1997). Tet2 was shown to be responsible for the 5mC oxidation on mRNA (Shen et al., 2018). The potential interaction between Tet2 and Alyref might offer a way to explain the binding mechanism of Tet2 to mRNA. However the co-expression co-IP experiments failed to show a clear interaction between Tet2 and Alyref or Ppp1r9b in 293 T cells. The contradicting results might reflect that Tet2 interaction is present in a cell-type specific mechanism. Additional expression of co-factors or specific post-translational modifications would be required for certain proteins to be able to interact with Tet2.

7
Discussion

Contents

Although Tet2 function in hematopoiesis including affecting hematopoietic stem cell self-renewal ability and restricting the immature myeloid differentiation has been proposed, much less is known about the molecular role of Tet2 function in driving the cellular processes involved in hematopoiesis. In my DPhil project, I carried out a systematic study to explore the molecular mechanism about how Tet2 gets involved in gene regulation by using a well-characterized mouse hematopoietic progenitor cell line, HPC-7. Knocking out *Tet2* gene in HPC-7 cells leads to impaired colony growth and aberrant cell differentiation in colony-forming-unit assay. Genes involved in hematopoiesis such as *c-Kit*, *Cebpa*, *Cebpe*, and *Csf1r* have been shown to get de-regulated in *Tet2*-KO cells. To identify the molecular functions of Tet2, I firstly mapped the genomic binding profile of Tet2, then characterized the Tet2-mediated DNA demethylation activity and assessed the contribution of Tet2-mediated DNA demethylation in Tet2-related gene regulation. In the meantime, I also performed an IP-MS experiment to identify novel Tet2 interacting

partners in HPC-7 cells, thus to provide some insights in understanding the molecular role of Tet2 and its chromatin recruitment mechanism.

7.1 The molecular role of Tet2 in gene regulation in HPC-7 cells

To study Tet2 function in HPC-7, I firstly generated the *Tet2*-KO HPC-7 cell line by CRISPR-Cas9 technique. Tet2 deficiency in HPC-7 leads to impaired colony growth and causes immature cell differentiation in colony-forming-unit assay (Figure 3.10). Consistently, both *Tet2*-KO clones show impaired cell growth in the liquid differentiation assay (Figure 3.9), these observations suggest a cell growth defect caused by *Tet2* depletion. However, this finding seems to contradict the observation from the *Tet2*-deficient mouse studies where Tet2 depletion leads to expanded HSC population and promotes the HSC self-renewal ability (Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011; Quivoron et al., 2011). Thus, the inconsistency of Tet2 function in cell growth might reflect a cell type-dependent manner of Tet2 function. Although HPC-7 cells exhibit similar gene expression profile as the multipotent hematopoietic progenitor cells and maintain the cell differentiation potential during *in vitro* culture, HPC-7 cells were originally derived from differentiated mouse embryonic stem cells and immortalized by the stable expression of *LH2* gene, which might offer some phenotypic difference in cell growth and differentiation compared to the primary HSC. Furthermore, *in vitro* cell differentiation assay provides a relatively simplified environment for cell growth and differentiation, thus it might not fully recapitulate the original environment of the primary cells. Thus, *in vivo* culture of the *Tet2*-KO HPC-7 cells by transplanting the cells into mice might be helpful to better understand the cellular function of Tet2 in regulating HSC/HSPC cell proliferation and differentiation.

In order to understand the molecular changes caused by Tet2 deficiency, I performed RNA-seq experiments on *Tet2*-KO HPC-7 cells and identified more genes showing up-regulation than down-regulation in *Tet2*-KO cells compared with WT (Figure 4.3A). The two *Tet2*-KO clones share some common differentially expressed (DE) genes since 50%

of identified DE genes could be found deregulated in both KO clones (Figure 4.3B). More importantly, among the top down-regulated genes identified in both *Tet2*-KO clones, I observed several myeloid lineage-essential genes such as *csf1R*, *Cebpa*, and *Cebpe*. The impaired expression of these genes have been found to disrupt the normal myeloid differentiation *in vivo*, thus the de-regulation of myeloid genes in *Tet2*-KO HPC-7 cells might provide a potential mechanism for the differentiation block phenotype observed in the CFU assay. Upregulated genes identified in *Tet2*-KO HPC-7 cells including stem cell marker (c-Kit) and HSC self-renewal factors (e.g. Hlf, Sox4) were already reported to be up-regulated in the *Tet2*-KO bone marrow cells isolated from mouse (Izzo et al., 2020). In total, around 125 genes related to hematopoiesis processes were found significantly differentially expressed in both *Tet2*-KO clones, reflecting the important role of Tet2 activity in hematopoiesis.

To identify the transcriptional events directly regulated by Tet2 activity, I firstly mapped the genome-wide occupancy of Tet2 protein in HPC-7 cells, and further identified Tet2-bound genes. Although the first hematopoietic cells Tet2 ChIP-seq was published last year by Rasussen and colleagues (Rasmussen et al., 2019), the ChIP-seq signals showing in that paper exhibit a high background and show similar peak pattern between ChIP in *Tet2*-KO cells control group and ChIP in WT cells, which brings us concerns about the specificity of Tet2 antibody used in the ChIP experiments. Moreover, I have tested several commercially available Tet2 antibodies for ChIP-qPCR experiments and many of the antibodies failed to show enrichment of Tet2 binding at the reported Tet2 ESCs binding sites. Although Tet2 antibody from Cell Signaling could capture Tet2 binding events at the reported genomic regions, ChIP background in the KO cells control group was also observed when doing the ChIP-qPCR experiments in HPC-7 cells (Figure 4.7). Lack of a good Tet2 antibody used for ChIP application really limits the mapping of endogenous Tet2 binding activity. Finally, to overcome the background issue brought by the Tet2 antibody non-specific binding, I mapped the Tet2 genomic binding sites in HPC-7 cells by doing ChIP-seq experiments with the HA tag antibody. HA-Tet2 protein was expressed in a level close to the endogenous Tet2 protein in the HPC-7 cells, moreover, the endogenous Tet2 protein was already depleted by CRISPR-Cas9 technique (Figure 5.14). Around

22,000 Tet2 binding sites were defined by the differential binding analysis using the HA ChIP-seq signal from the *Tet2*-KO cells as a control. Functional analysis reveals that Tet2 binds the gene promoters enriched with H3K4me3 and H3K27ac and distal regulatory elements marked with H3K4me1 and H3K27ac (Figure 4.11 and 4.12). Furthermore, Tet2 binding sites display the enrichment of many essential hematopoietic transcription factor binding motifs, such as Etv family proteins, Runx1, Fli1 and Gata family proteins. This observation is consistent with the reported TF motif enrichment analysis on Tet2 ChIP-seq in mouse hematopoietic cells from Rasmussen's study (Rasmussen et al., 2019), these together indicate a cell-type specific genomic binding of Tet2. In line with this observation, pathway analysis on Tet2-bound genes also revealed an enrichment of pathways related to hematopoiesis and immune response (Figure 4.14). Interestingly, Tet2-bound genes can be either down-regulated or up-regulated upon Tet2 deficiency (Figure 4.16), indicating a dual function of Tet2 in gene activation and repression. And the different transcriptional activity caused by Tet2 binding might suggest a cooperation of other chromatin factors involved in Tet2-mediated gene regulation.

Many studies have linked TET2-mediated DNA demethylation activity to its gene regulation function, particularly, TET2 loss leads to a DNA hypermethylation phenotype at enhancer regions in mouse hematopoietic/non-hematopoietic cells with Tet2 depletion and human leukemia cells with TET2 mutations (Hon et al., 2014; Lu et al., 2014; Rampal et al., 2014; Rasmussen et al., 2015). To understand the role of Tet2-mediated DNA demethylation in Tet2-associated gene regulation in HPC-7 cells, we mapped the genome-wide 5mC distribution in Tet2 WT, *Tet2*-KO and HA-Tet2 HPC-7 cells. Our results suggest that Tet2 depletion does not lead to a global gain or loss of DNA methylation (Figure 4.18). However, enhancer regions undergo a clear increase of DNA methylation caused by Tet2 deficiency and the re-expression of Tet2 protein in KO cells could reverse the gain of 5mC on enhancer regions (Figure 4.19), these suggest an important role of Tet2 in DNA demethylation at enhancer regions. Although previous studies do not observe a correlation between Tet2 binding and its DNA demethylation activity, our data suggest that Tet2 binding is positively correlated with Tet2-mediated DNA demethylation activity since Tet2 binding sites were found hypermethylated upon

Tet2 depletion and hypomethylated upon Tet2 restoration (Figure 4.20). Furthermore, Tet2-bound enhancer regions undergo the most significant gain of DNA methylation in the *Tet2*-KO cells (Figure 4.21). These further support the notion that Tet2 binds the enhancer regions and is responsible for the DNA demethylation at the bound enhancers. In our data, Tet2 binding is suggested to be associated with both gene activation and gene repression. So we further looked at the methylation changes on the Tet2-bound differentially expressed genes identified in *Tet2*-KO cells vs WT cells (Figure 4.22). Genes positively regulated by Tet2 binding showed decreased expression level and increased DNA methylation in *Tet2*-KO cells, suggesting that Tet2 is involved in gene activation via its DNA demethylase activity. Interestingly, genes negatively regulated by Tet2 binding also showed a minor increase in DNA methylation suggesting that Tet2 is involved in gene repression partially via its DNA demethylase activity, but the detailed mechanism for Tet2-mediated gene repression still remain unclear. Furthermore, Tet2 co-localizes with many key hematopoietic transcription factors including Runx1, Tal1, Fli1, Gata2, and PU.1 (Figure 4.25) at enhancer region of myeloid-specific genes, such as, *csf1R*,*Cebpa* and *Cebpe*. Previous studies have reported the physical interaction between Tet2 and Runx1/PU.1 (Chu et al., 2018; de la Rica et al., 2013). Thus Runx1 or PU.1 might be responsible for the recruitment of Tet2 to chromatin through direct protein-protein interaction. More importantly, Tet2 binding at hematopoietic enhancer elements leads to DNA methylation loss, which might further promote the binding of TF sensitive to DNA methylation or inhibit the binding of certain TF that favors the 5mC-containing DNA.

7.2 Identification of Tet2 hematopoietic-specific interactors

To study the protein-protein interaction (PPI) not only helps to understand the function of the protein but also helps to predict a protein's unknown function and related cellular processes. Although several Tet2 interactors have been identified to be involved in the Tet2-associated gene regulation or to modulate Tet2 enzymatic activity and protein stability, little is known about the interacting partners of Tet2 in hematopoietic context

since the scarcity of the hematopoietic stem/progenitor cells. In our study, facilitated by the great availability of the HPC-7 cells for biochemical applications, I performed the first Tet2 interactome mapping in the hematopoietic cell context, and hope this could provide any insights in understanding of Tet2-mediated gene regulation and the cellular phenotypes identified in *Tet2*-KO cells. Given the advantages of proximity labelling based method in weak and transient protein-protein interaction identification, I firstly attempted the APEX2 labelling assay which utilizes the APEX2 peroxidase to biotinylate proteins in close proximity to the APEX2-fusion protein. In the preliminary experiments performed in 293T cells, APEX2 labelling assay captures the reported interaction between Tet2 and OGT by biotinylating OGT. More importantly, APEX2 tagging shows no effect on Tet2 enzymatic activity. However, unexpected detection of TBP in the APEX2-Tet2 labelled biotinylated proteins brings us concern about the specificity of the APEX2 labelling which could be a consequence of over-expression of APEX2 fusion protein in cells. Furthermore, the high self-biotinylated background found in WT HPC-7 cells further cause difficulties in applying the APEX2 method for Tet2 interactome mapping in HPC-7 cells (Figure 5.10).

At the end, by using the classical IP-MS approach, I performed the first Tet2 interactors mapping in hematopoietic cells. To achieve an optimal expression level of Tet2 proteins in the HPC-7 cells, I constructed an 3xHA-Tet2 expressing cell line where the endogenous Tet2 protein was depleted by CRISPR-Cas9 technique and the 3xHA-Tet2 protein was expressed in a level close to the WT protein by dox treatment. Anti-HA IP-MS reveals many novel Tet2 interactors together with the known Tet2 interactor, OGT, in HPC-7 cells. Interestingly, components of the core PRC2 complex, such as Suz12, Ezh2, and Rbbp7 were identified as Tet2 interacting partners by the IP-MS experiments. The interaction between PRC2 protein and endogenous Tet2 is confirmed by the PRC2 reciprocal co-IP in HPC-7 cells. Moreover, Suz12 and Ezh2 are shown to interact specifically with the C terminus of Tet2 protein by the 293 T co-expression experiments (Figure 6.8,6.9). Suz12 and Ezh2 are associated with the di/trimethylation of lysine 27 of histone, which is closely related to gene repression (Cao and Zhang, 2004). Our study firstly discovered the biochemical interaction between Tet2 and PRC2 complex in HPC-7 cells which might

indicate a potential cross-talk between Tet2 activity and PRC2-mediated gene repression. However, the interaction between Tet2 and PRC2 were shown to be quite weak by the co-expression assay in the 293 T cells, this might indicate that Tet2 interacts with PRC2 indirectly, or the establishment of PRC2-Tet2 complex might require the expression of additional components to stabilize the protein complex. Although the interaction between Tet2 and other MS-identified Tet2 interactors including Noc2l, Alyref, and Pppr1r9b can be confirmed by the interactor reciprocal co-IP in the HPC-7 cells, I fail to detect their interaction with Tet2 when co-expressing the tagged proteins of Tet2 and the interactors in 293 T cells (Figure 6.11). These might suggest that Tet2 interacts with different molecular partners in a cell context-dependent manner. Lack of expression of co-factors involved in Tet2 interacting complex might interfere the detection of the identified novel interactions in non HPC-7 cells.

7.3 The potential function of Tet2-PRC2 interaction

In our study, multiple lines of evidence have suggested the potential cross-talk between Tet2 and PRC2 complex in gene regulation. In our study, some of the reported mESC Ezh2-targeted genes including *Egr3*, *Egr2*, *Irx3*, and *Irx5* exhibit H3K27me3 signal at their gene promoter region in HPC-7 cells (Figure 7.1B), suggesting these genes could be true Ezh2-targeted genes in HPC-7 cells. More importantly, Tet2 binds to the gene promoter region of those reported Ezh2-targeted genes in HPC-7 cells, and the binding of Tet2 seems to play a repressive role in gene regulation since Tet2 depletion in HPC-7 cells leads to increased gene expression of those Ezh2-targeted genes (Figure 7.1A). Thus, these predicted Ezh2-targeted genes in HPC-7 cells might serve as examples of Tet2 function in PRC2-mediated gene repression. The other evidence from our study suggesting a functional cross-talk between Tet2 and PRC2 complex comes from the identification of protein interaction between Tet2 and Suz12, Ezh2, components of core PRC2 complex. IP-MS experiment and followed interaction domain mapping results provide a direct physical proof for the co-operation between Tet2 and PRC2 complex. In addition, Tet2 and Ezh2 have been shown to share some functional similarities in normal and malignant hematopoiesis. Both Tet2 and Ezh2 were reported as tumour suppressor

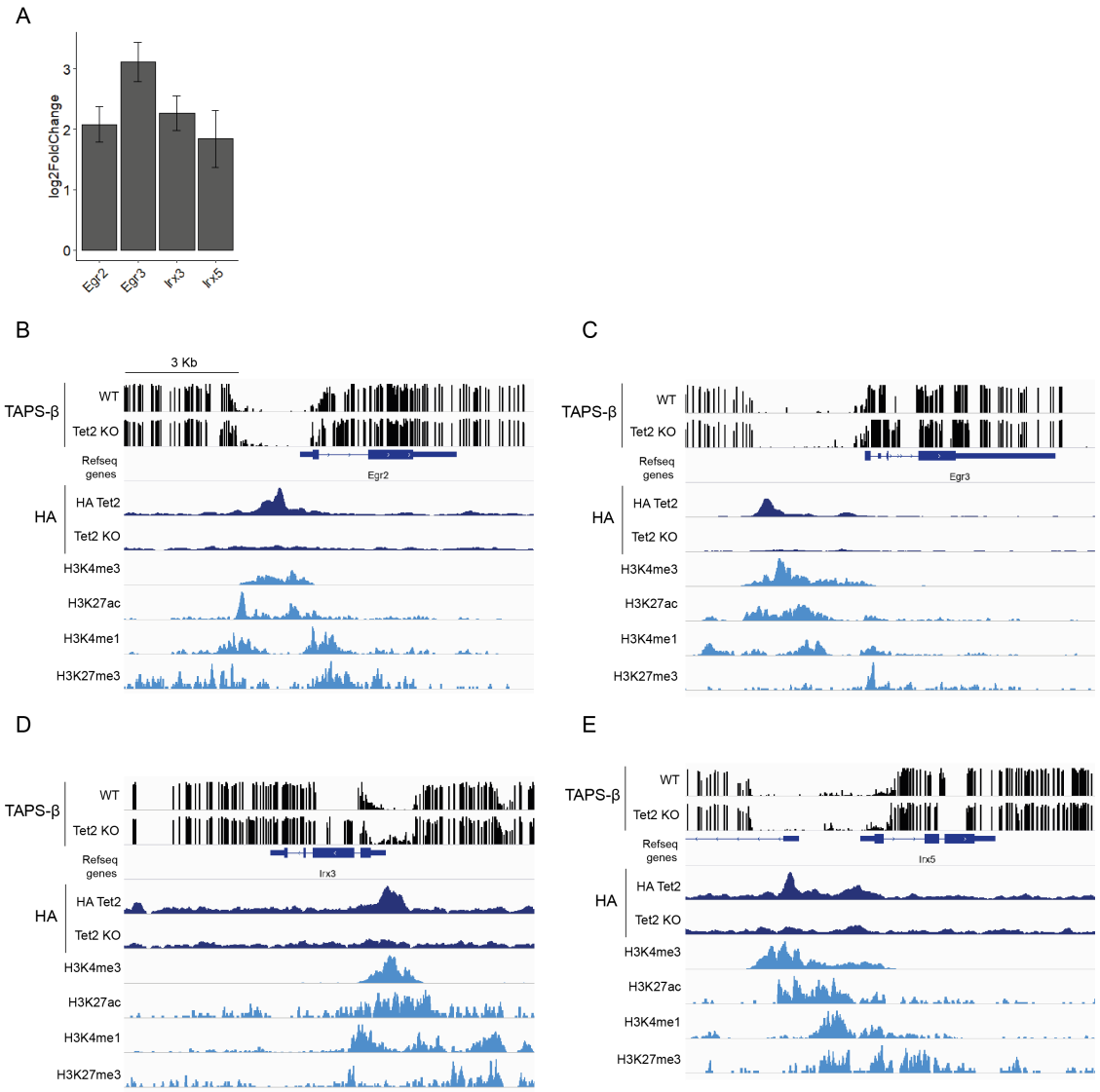


Figure 7.1: Putative Ezh2-targeted genes were found significantly up-regulated in *Tet2*-KO HPC-7 cells. (A) Bar plots showing the gene expression changes of putative Ezh2-targeted genes. Differential expression analysis was done by DESeq2. Y-axis represents the $\log_2$ fold change of gene expression in *Tet2*-KO clone 7 versus WT HPC-7 cells. Error bars indicate estimated standard error for the log2-scaled fold changes. (B-E) Genome browser tracks shows representative examples of putative Ezh2 targeted genes in HPC-7 cells. 5mC level in wild-type and *Tet2*-KO HPC-7 cells was quantified using TAPS-β method by Dr. Sophie Kirschner and data analysis was performed by Olena Yavorska from our lab. Beta-value ranging from 0 to 1 was used to measure the percentage of methylation. HPC-7 H3K4me3 (Xu et al., 2018), H3K4me1, H3K27ac (Org et al., 2015), H3K27me3 and H3K36me3 (Wilson et al., 2016) were retrieved from the respective study.

genes involved in myeloid malignancies. *Ezh2* depletion in mice results in enhanced repopulating capacity of HSCs and promotes the cell differentiation towards myeloid lineage (Muto et al., 2013), these hematopoietic lineage alterations identified in *Ezh2*-

deficient mice are similar to what have been reported in *Tet2*-deficient mouse models. Taken together, our data newly identify that PRC2 complex can physically interact with the C-terminus of Tet2 protein and the interaction between Tet2 and PRC2 might affect the PRC2-mediated gene expression since *Tet2*-KO de-represses the expression of Ezh2-targeted genes, this reveals a new mechanism by which Tet2 gets involved in gene regulation.

Interestingly, in 2011, Wu and colleagues have reported the dual role of the other Tet family protein, Tet1, in transcriptional regulation in mouse embryonic stem cells (Wu et al., 2011b). More importantly, Tet1 binds to the promoters of Polycomb-repressed genes. Tet1 depletion in mESCs leads to the de-repression of Polycomb-repressed genes and impairs the chromatin binding of the PRC2 complex. These findings establish a role of Tet1 in Polycomb-targeted gene repression. It is likely that Tet promotes PRC2 DNA binding via its DNA demethylation activity since DNA methylation is shown to negatively affect PRC2 binding (Holoch and Margueron, 2017; Reddington et al., 2013). However, some studies offer another way of interpretation on Tet2-PRC2 cross-talk. Genomic co-localization of Tet1, 5hmC and PRC2 was also identified by Neri and colleagues (Neri et al., 2013). In their study, Tet1 was found to be recruited to the H3K27me3 positive regions of the genome by its interaction with PRC2 components, Suz12 and Ezh2, since knock-down of Suz12 leads to a reduced binding of Tet1 in the genome. Furthermore, Suz12 depletion cause an increase of DNA methylation in the genome suggesting an important role of PRC2 activity in Tet1-mediated DNA demethylation. Later on, a study from Xie's group also supports this notion that Polycomb promotes the hypomethylation of DNA methylation valley (Li et al., 2018). To summarize, current findings have provided a bi-directional function of Tet-PRC2 cross-talk. On one hand, PRC2 recruits Tet proteins to the genome and promotes DNA demethylation at the PRC2-Tet co-bound DNA regions. On the other hand, Tet proteins facilitate the genomic binding of PRC2 by its DNA demethylation activity.

Although our data reveal the biophysical interaction between Tet2 and PRC2 in mouse hematopoietic progenitor cells, the functional relevance of the Tet2-PRC2 cross-talk

remain unknown. In the future, more experiments would be needed in order to answer the following questions:

1. What is the role of Tet2-PRC2 complex in gene regulation?

First, genome-wide mapping of PRC2 binding sites in HPC-7 cells, e.g Suz12, Ezh2, H3K27me3 ChIP-seq, would be useful for the understanding of PRC2 genomic distribution and the identification of PRC2-targeted genes in HPC-7 cells. Then, check the degree of overlap between Tet2 genomic binding and PRC2 genomic binding. Characterize the genomic features of the Tet2-PRC2 co-bound regions, to see if Tet2-PRC2 prefer to co-localize at gene promoter or enhancer regions. If Tet2-PRC2 co-bound regions are more enriched in the enhancer category, this might suggest a role of Tet2 activity in the establishment or function of bivalent enhancer regions. If promoters regions are more identified in Tet2-PRC2 co-bound regions, this might indicate a potential role of Tet2 in PRC2-mediated gene regulation. To further explore if Tet2 is involved in the repression of PRC2-targeted genes, it would be interesting to check if genes showing increased expression in *Tet2*-KO cells are enriched with PRC2 chromatin binding and H3K27me3 distribution compared to the genes down-regulated in KO cells. The alternative way is to assess the expression changes of Tet2-PRC2 co-bound genes, to see if those genes could get up-regulated if Tet2 is depleted from cells.

2. What is the genomic binding order between Tet2 and PRC2?

Hypothesis 1: Tet2 is responsible for the genomic binding of PRC2 in HPC-7 cells. In order to test this, PRC2 and H3K27me3 ChIP-seq in *Tet2*-KO cells compared to those done in WT cells would be able to tell if Tet2 depletion will cause any disruption or re-distribution of PRC2 chromatin binding. Next, if Tet2-mediated DNA demethylation is important for the Tet2-dependent PRC2 recruitment, increased DNA methylation at regions with loss of PRC2 binding in *Tet2*-KO cells and with binding of both Tet2 and PRC2 would be expected in *Tet2*-KO cells compared to WT cells. More importantly, it would be interesting to see the expression changes of genes showing a loss of PRC2 binding in *Tet2*-KO cells, increased expression of those genes would be expected if Tet2 is truly involved in PRC2-mediated gene repression.

Hypothesis 2: Tet2 is recruited to DNA by interacting with PRC2 in HPC-7 cells. In order to test this, Tet2 ChIP-seq in PRC2-depleted HPC-7 cells would be able to tell if loss of PRC2 will cause any effect in Tet2 genomic binding. Moreover, to compare the 5mC distribution in PRC2-depleted and WT HPC-7 cells would be another way to assess the disruption of Tet2 activity caused by PRC2 deficiency.

3. What is the functional importance of Tet2-PRC2 complex in blood lineage?

To confirm the cross-talk between Tet2 and PRC2 in primary blood cells, endogenous IP in mouse bone marrow cells would be useful to tell if Tet2 interacts with PRC2 endogenously. To identify any leukemia-related Tet2 mutation that can disrupt the physical interaction between Tet2 and PRC2 would be able to provide an evidence for the mechanism by which Tet2 mutation contributes to leukemogenesis through the loss of co-operation with PRC2 complex. If PRC2 recruits Tet2 to the genome, disrupted Tet2 genomic binding, increased DNA methylation or altered 5hmC distribution might occur in leukemic cells with PRC2 mutation. On the other hand, if Tet2 recruits PRC2 to the DNA, alteration in PRC2 genomic binding and H3K27me3 distribution would be expected in leukemic cells with certain types of Tet2 mutation that can abolish the interaction between Tet2 and PRC2.

8

Appendices

8.1 Significantly differential expressed genes involved in hematopoietic-related cell process

Hematopoietic-related gene expression changes in Tet2-KO7 versus WT HPC-7 cells

Geneid	baseMean	log2FoldChange	lfcSE	pvalue	padj	sigchange	GO_term
Abl1	5024.351	0.338	0.074	4.89E-06	1.64E-04	Sig.Increase	alpha-beta T cell differentiation;regulation of T cell differentiation;transitional one stage B cell differentiation
Cd79b	19.752	-3.430	0.588	5.36E-09	3.25E-07	Sig.Decrease	B cell differentiation
Hdac9	1540.452	0.748	0.199	1.75E-04	3.72E-03	Sig.Increase	B cell differentiation
Ptprj	2120.055	0.294	0.099	2.91E-03	3.87E-02	Sig.Increase	B cell differentiation
Sp3	6150.578	0.318	0.094	7.12E-04	1.22E-02	Sig.Increase	B cell differentiation;definitive hemopoiesis;enucleate erythrocyte differentiation;erythrocyte differentiation;granulocyte differentiation;megakaryocyte differentiation;monocyte differentiation;natural killer cell differentiation;T cell differentiation
Kit	66169.008	0.523	0.101	1.96E-07	8.85E-06	Sig.Increase	B cell differentiation;embryonic hemopoiesis;erythrocyte differentiation;hematopoietic progenitor cell differentiation;hemopoiesis;immature B cell differentiation;lymphoid progenitor cell differentiation;mast cell differentiation;megakaryocyte development;myeloid leukocyte differentiation;myeloid progenitor cell differentiation;T cell differentiation
Zbtb7a	1391.317	0.316	0.089	3.75E-04	7.11E-03	Sig.Increase	B cell differentiation;erythrocyte maturation;regulation of osteoclast differentiation
Cdh17	320.051	-2.136	0.418	3.24E-07	1.41E-05	Sig.Decrease	B cell differentiation;germinal center B cell differentiation;marginal zone B cell differentiation
Gimap1	1196.558	0.489	0.118	3.27E-05	8.72E-04	Sig.Increase	B cell differentiation;T cell differentiation
Malt1	1847.795	0.225	0.075	2.93E-03	3.89E-02	Sig.Increase	B-1 B cell differentiation;positive regulation of T-helper 17 cell differentiation
Gata1	634.574	-1.317	0.214	7.32E-10	5.19E-08	Sig.Decrease	basophil differentiation;dendritic cell differentiation;embryonic hemopoiesis;eosinophil differentiation;eosinophil fate commitment;erythrocyte development;erythrocyte differentiation;megakaryocyte differentiation;myeloid cell differentiation;platelet formation;positive regulation of erythrocyte differentiation;regulation of definitive erythrocyte differentiation;regulation of primitive erythrocyte differentiation
Rsad2	79.837	1.582	0.438	3.08E-04	6.02E-03	Sig.Increase	CD4-positive, alpha-beta T cell differentiation
Bcr	2517.463	0.507	0.124	4.07E-05	1.04E-03	Sig.Increase	definitive hemopoiesis
Hoxb4	351.633	0.466	0.134	4.98E-04	8.98E-03	Sig.Increase	definitive hemopoiesis;hematopoietic stem cell differentiation;hemopoiesis
Zfp36l2	19499.984	0.459	0.066	4.91E-12	4.90E-10	Sig.Increase	definitive hemopoiesis;hemopoiesis;regulation of B cell differentiation;T cell differentiation in thymus
Irf8	42.693	-6.871	1.058	8.24E-11	6.73E-09	Sig.Decrease	dendritic cell differentiation;myeloid cell differentiation;plasmacytoid dendritic cell differentiation
Klf1	54.898	-2.171	0.401	6.09E-08	3.03E-06	Sig.Decrease	embryonic hemopoiesis;erythrocyte development;erythrocyte differentiation;erythrocyte maturation
Gata3	286.097	0.773	0.164	2.50E-06	9.16E-05	Sig.Increase	embryonic hemopoiesis;erythrocyte differentiation;mast cell differentiation;positive regulation of T cell differentiation;pro-T cell differentiation;regulation of CD4-positive, alpha-beta T cell differentiation;T cell differentiation;T cell differentiation in thymus;T-helper 2 cell differentiation;thymic T cell selection
Tgfbr2	3332.257	0.362	0.112	1.20E-03	1.88E-02	Sig.Increase	embryonic hemopoiesis;myeloid dendritic cell differentiation;positive regulation of NK T cell differentiation
Bpgm	3807.933	0.682	0.148	4.42E-06	1.50E-04	Sig.Increase	erythrocyte development
Sox6	166.345	-1.266	0.213	3.02E-09	1.92E-07	Sig.Decrease	erythrocyte development;erythrocyte differentiation;hemopoiesis
Casp3	11597.067	-0.623	0.111	2.02E-08	1.09E-06	Sig.Decrease	erythrocyte differentiation
Slc25a38	427.847	-0.521	0.148	4.28E-04	7.93E-03	Sig.Decrease	erythrocyte differentiation
Tmem14c	4512.547	-0.355	0.102	4.72E-04	8.59E-03	Sig.Decrease	erythrocyte differentiation
L3mbtl3	646.857	0.930	0.149	4.27E-10	3.21E-08	Sig.Increase	erythrocyte maturation;granulocyte differentiation;macrophage differentiation;myeloid cell differentiation
Fam210b	443.735	0.636	0.115	3.50E-08	1.82E-06	Sig.Increase	erythrocyte maturation;positive regulation of erythrocyte differentiation
Sox13	344.984	-0.577	0.130	9.40E-06	2.96E-04	Sig.Decrease	gamma-delta T cell differentiation;positive regulation of gamma-delta T cell differentiation;regulation of gamma-delta T cell differentiation
Ada	1957.660	0.421	0.104	5.38E-05	1.33E-03	Sig.Increase	germinal center B cell differentiation;positive regulation of alpha-beta T cell differentiation;positive regulation of T cell differentiation;positive regulation of T cell differentiation in thymus;regulation of T cell differentiation
Cebpa	1261.989	-1.496	0.152	8.74E-23	3.06E-20	Sig.Decrease	granulocyte differentiation;macrophage differentiation;myeloid cell differentiation
Cebpe	25.949	-2.459	0.571	1.66E-05	4.84E-04	Sig.Decrease	granulocyte differentiation;macrophage differentiation;myeloid cell differentiation
Arl11	35.160	-1.303	0.358	2.75E-04	5.49E-03	Sig.Decrease	hematopoietic progenitor cell differentiation
Dock1	1437.887	0.500	0.130	1.26E-04	2.81E-03	Sig.Increase	hematopoietic progenitor cell differentiation
Dock7	2328.940	0.384	0.104	2.14E-04	4.43E-03	Sig.Increase	hematopoietic progenitor cell differentiation
Herc6	178.037	-3.221	0.212	3.67E-52	6.30E-49	Sig.Decrease	hematopoietic progenitor cell differentiation
Siglecg	8.381	-2.991	0.850	4.31E-04	7.96E-03	Sig.Decrease	hematopoietic progenitor cell differentiation
Sirpa	2771.370	0.570	0.102	1.96E-08	1.06E-06	Sig.Increase	hematopoietic progenitor cell differentiation
Slc7a6os	1171.325	-0.289	0.095	2.26E-03	3.15E-02	Sig.Decrease	hematopoietic progenitor cell differentiation
Ston2	1379.410	-1.202	0.141	1.74E-17	3.77E-15	Sig.Decrease	hematopoietic progenitor cell differentiation
Thsd1	1929.067	0.678	0.127	9.01E-08	4.36E-06	Sig.Increase	hematopoietic progenitor cell differentiation
Tmem143	447.329	0.362	0.124	3.51E-03	4.51E-02	Sig.Increase	hematopoietic progenitor cell differentiation
Csf1r	44.532	-2.253	0.389	6.79E-09	4.03E-07	Sig.Decrease	hematopoietic progenitor cell differentiation;hemopoiesis;osteoclast differentiation;positive regulation of osteoclast differentiation
Mirlet7e	2.831	4.449	1.491	2.85E-03	3.80E-02	Sig.Increase	hematopoietic stem cell differentiation
Mecom	99.208	1.611	0.345	2.98E-06	1.07E-04	Sig.Increase	hematopoietic stem cell proliferation
Shb	166.204	0.919	0.157	4.89E-09	3.00E-07	Sig.Increase	hematopoietic stem cell proliferation;hemopoiesis;positive regulation of T-helper cell differentiation
Add2	3.773	-5.849	1.551	1.62E-04	3.48E-03	Sig.Decrease	hemopoiesis
Asxl1	4167.318	0.285	0.066	1.54E-05	4.54E-04	Sig.Increase	hemopoiesis
Pdgfrb	5471.586	0.453	0.098	4.16E-06	1.43E-04	Sig.Increase	hemopoiesis
Rogdi	166.899	-0.630	0.155	5.04E-05	1.25E-03	Sig.Decrease	hemopoiesis
Selplg	1322.475	-0.371	0.100	1.93E-04	4.06E-03	Sig.Decrease	hemopoiesis
Spta1	1462.812	-2.945	0.173	3.47E-65	7.44E-62	Sig.Decrease	hemopoiesis
Tet2	3712.223	-0.655	0.158	3.29E-05	8.74E-04	Sig.Decrease	hemopoiesis;myeloid cell differentiation;myeloid progenitor cell differentiation
Myo1e	152.485	1.205	0.253	1.95E-06	7.43E-05	Sig.Increase	hemopoiesis;post-embryonic hemopoiesis
Runx2	1395.677	0.538	0.087	6.17E-10	4.45E-08	Sig.Increase	hemopoiesis;T cell differentiation
Irf2bp2	8728.787	0.315	0.098	1.39E-03	2.11E-02	Sig.Increase	immature B cell differentiation
Rras	61.830	0.789	0.256	2.04E-03	2.91E-02	Sig.Increase	leukocyte differentiation
Relb	147.296	1.024	0.234	1.19E-05	3.63E-04	Sig.Increase	lymphocyte differentiation;myeloid dendritic cell differentiation;T-helper 1 cell differentiation
Vegfa	762.341	0.499	0.167	2.87E-03	3.82E-02	Sig.Increase	macrophage differentiation;monocyte differentiation;primitive erythrocyte differentiation;regulation of hematopoietic progenitor cell differentiation
Csf1	839.072	1.044	0.330	1.58E-03	2.35E-02	Sig.Increase	macrophage differentiation;osteoclast differentiation;positive regulation of macrophage differentiation;positive regulation of monocyte differentiation;positive regulation of osteoclast differentiation
Bcl3	196.115	-0.527	0.172	2.23E-03	3.13E-02	Sig.Decrease	marginal zone B cell differentiation;T-helper 2 cell differentiation
Prtn3	854.922	-1.042	0.186	2.24E-08	1.20E-06	Sig.Decrease	mature conventional dendritic cell differentiation
Pip4k2a	2643.787	0.296	0.088	7.59E-04	1.28E-02	Sig.Increase	megakaryocyte development
Rabgap1l	49.562	-2.685	0.329	3.21E-16	5.62E-14	Sig.Decrease	megakaryocyte development
Il3ra	490.936	-0.698	0.115	1.31E-09	8.85E-08	Sig.Decrease	monocyte differentiation

Mef2c	80.027	-1.272	0.335	1.45E-04	3.21E-03	Sig.Decrease	monocyte differentiation;platelet formation;regulation of megakaryocyte differentiation
Jun	1394.760	0.905	0.200	5.95E-06	1.95E-04	Sig.Increase	monocyte differentiation;positive regulation of monocyte differentiation
Ccr1	294.847	-1.601	0.469	6.36E-04	1.11E-02	Sig.Decrease	myeloid cell differentiation;positive regulation of osteoclast differentiation
Slamf1	412.051	-0.781	0.197	7.71E-05	1.85E-03	Sig.Decrease	natural killer cell differentiation
Tmem176a	1465.554	0.355	0.092	1.21E-04	2.72E-03	Sig.Increase	negative regulation of dendritic cell differentiation
Tmem176b	3947.302	0.297	0.073	4.21E-05	1.06E-03	Sig.Increase	negative regulation of dendritic cell differentiation
Zfp36	765.702	1.188	0.186	1.83E-10	1.44E-08	Sig.Increase	negative regulation of erythrocyte differentiation;negative regulation of myeloid cell differentiation
Mafb	40.956	1.526	0.354	1.59E-05	4.66E-04	Sig.Increase	negative regulation of erythrocyte differentiation;negative regulation of osteoclast differentiation;regulation of myeloid cell differentiation;T cell differentiation in thymus
Zfp36l1	153.545	0.707	0.229	2.03E-03	2.90E-02	Sig.Increase	negative regulation of erythrocyte differentiation;positive regulation of monocyte differentiation;regulation of B cell differentiation;T cell differentiation in thymus
Rara	402.348	0.561	0.143	8.37E-05	1.98E-03	Sig.Increase	negative regulation of granulocyte differentiation;positive regulation of T-helper 2 cell differentiation;regulation of granulocyte differentiation
Nfe2l2	5039.757	0.465	0.103	7.06E-06	2.29E-04	Sig.Increase	negative regulation of hematopoietic stem cell differentiation
Cdkn2a	784.322	-1.344	0.193	3.52E-12	3.67E-10	Sig.Decrease	negative regulation of immature T cell proliferation in thymus
Fgl2	820.994	-0.494	0.120	3.94E-05	1.01E-03	Sig.Decrease	negative regulation of memory T cell differentiation
Meis2	6.974	3.976	1.056	1.66E-04	3.56E-03	Sig.Increase	negative regulation of myeloid cell differentiation
Clec2i	29.006	-2.199	0.384	1.03E-08	5.88E-07	Sig.Decrease	negative regulation of osteoclast differentiation
Gpr55	48.085	-1.070	0.325	1.01E-03	1.62E-02	Sig.Decrease	negative regulation of osteoclast differentiation
Nf1	5057.015	0.573	0.104	3.61E-08	1.87E-06	Sig.Increase	negative regulation of osteoclast differentiation
Pilrb1	248.877	-0.443	0.140	1.62E-03	2.41E-02	Sig.Decrease	negative regulation of osteoclast differentiation
Hes1	62.629	0.866	0.254	6.62E-04	1.15E-02	Sig.Increase	negative regulation of pro-B cell differentiation
Cd44	7823.645	0.285	0.080	3.92E-04	7.39E-03	Sig.Increase	negative regulation of regulatory T cell differentiation
Lag3	9.131	-2.370	0.692	6.17E-04	1.08E-02	Sig.Decrease	negative regulation of regulatory T cell differentiation
Itk	92.576	-1.217	0.310	8.47E-05	2.00E-03	Sig.Decrease	NK T cell differentiation
Gpc3	1555.095	-6.774	0.210	7.60E-228	1.30E-223	Sig.Decrease	osteoclast differentiation
Junb	2562.634	0.840	0.135	4.67E-10	3.46E-08	Sig.Increase	osteoclast differentiation
Trf	62.437	-2.942	0.338	3.25E-18	7.53E-16	Sig.Decrease	osteoclast differentiation
Tyrobp	62.641	-3.570	0.437	3.30E-16	5.73E-14	Sig.Decrease	osteoclast differentiation;positive regulation of osteoclast development;regulation of osteoclast development
Tnf	1048.064	0.478	0.116	3.90E-05	1.01E-03	Sig.Increase	osteoclast differentiation;positive regulation of osteoclast differentiation;regulation of osteoclast differentiation
Mitf	156.102	0.785	0.263	2.79E-03	3.73E-02	Sig.Increase	osteoclast differentiation;regulation of osteoclast differentiation
Actn1	2142.490	0.319	0.081	8.63E-05	2.03E-03	Sig.Increase	platelet formation
Il2rg	1575.408	0.357	0.091	8.46E-05	2.00E-03	Sig.Increase	positive regulation of B cell differentiation;positive regulation of CD4-positive, CD25-positive, alpha-beta regulatory T cell differentiation;positive regulation of lymphocyte differentiation;positive regulation of T cell differentiation in thymus
Ets1	73.921	-2.330	0.277	4.55E-17	9.30E-15	Sig.Decrease	positive regulation of erythrocyte differentiation
Trim58	72.798	-6.391	0.683	8.58E-21	2.45E-18	Sig.Decrease	positive regulation of erythrocyte enucleation
Sox4	11799.843	0.585	0.121	1.40E-06	5.47E-05	Sig.Increase	positive regulation of gamma-delta T cell differentiation;pro-B cell differentiation;T cell differentiation
Foxc1	486.280	1.028	0.179	8.65E-09	5.00E-07	Sig.Increase	positive regulation of hematopoietic progenitor cell differentiation;positive regulation of hematopoietic stem cell differentiation
Il1b	57.243	-1.385	0.288	1.47E-06	5.72E-05	Sig.Decrease	positive regulation of immature T cell proliferation in thymus
Scin	29.418	-1.647	0.379	1.42E-05	4.26E-04	Sig.Decrease	positive regulation of megakaryocyte differentiation
Jag1	56.275	1.189	0.284	2.83E-05	7.74E-04	Sig.Increase	positive regulation of myeloid cell differentiation
Ccl3	27.389	-2.545	0.589	1.53E-05	4.53E-04	Sig.Decrease	positive regulation of osteoclast differentiation
Fos	781.008	2.636	0.451	5.09E-09	3.11E-07	Sig.Increase	positive regulation of osteoclast differentiation
Gnas	12359.765	0.671	0.091	1.54E-13	1.98E-11	Sig.Increase	positive regulation of osteoclast differentiation
Sox12	47.836	-1.932	0.422	4.71E-06	1.60E-04	Sig.Decrease	positive regulation of regulatory T cell differentiation
Il12a	36.979	-6.293	0.796	2.68E-15	4.08E-13	Sig.Decrease	positive regulation of T cell differentiation
Pik3r6	126.340	-3.036	0.477	1.95E-10	1.52E-08	Sig.Decrease	positive regulation of T cell differentiation
Egr3	205.381	3.106	0.323	6.19E-22	2.00E-19	Sig.Increase	positive regulation of T cell differentiation in thymus;regulation of gamma-delta T cell differentiation
Foxp1	6273.955	0.337	0.102	9.94E-04	1.60E-02	Sig.Increase	pre-B cell differentiation;T follicular helper cell differentiation
Stk4	10933.996	0.478	0.105	5.08E-06	1.69E-04	Sig.Increase	primitive hemopoiesis
Cebpb	25.472	-1.729	0.482	3.31E-04	6.40E-03	Sig.Decrease	regulation of dendritic cell differentiation;regulation of osteoclast differentiation
Pim1	6668.796	-0.283	0.087	1.12E-03	1.77E-02	Sig.Decrease	regulation of hematopoietic stem cell proliferation
Csf3r	520.343	-2.961	0.149	1.05E-87	3.60E-84	Sig.Decrease	regulation of myeloid cell differentiation
Fgfr3	91.915	-0.922	0.221	2.89E-05	7.88E-04	Sig.Decrease	regulation of osteoclast differentiation
Cyp26b1	2064.406	1.402	0.342	4.17E-05	1.06E-03	Sig.Increase	regulation of T cell differentiation
Egr1	1283.481	1.106	0.182	1.22E-09	8.32E-08	Sig.Increase	T cell differentiation
Prex1	1686.500	0.859	0.130	4.68E-11	4.06E-09	Sig.Increase	T cell differentiation
Ptpn22	872.520	0.477	0.115	3.33E-05	8.85E-04	Sig.Increase	T cell differentiation
Fzd5	749.961	0.534	0.112	1.93E-06	7.36E-05	Sig.Increase	T cell differentiation in thymus
Fzd7	1888.726	0.279	0.096	3.51E-03	4.51E-02	Sig.Increase	T cell differentiation in thymus
Mr1	420.132	0.423	0.141	2.63E-03	3.56E-02	Sig.Increase	T cell differentiation in thymus
Clec4d	516.128	-0.538	0.175	2.14E-03	3.03E-02	Sig.Decrease	T cell differentiation involved in immune response
Il18r1	25.448	-1.089	0.360	2.45E-03	3.37E-02	Sig.Decrease	T-helper 1 cell differentiation
Rora	484.369	1.660	0.384	1.56E-05	4.59E-04	Sig.Increase	T-helper 17 cell differentiation
Il6ra	974.432	-0.316	0.103	2.26E-03	3.15E-02	Sig.Decrease	T-helper 17 cell lineage commitment
Ly9	323.366	-1.298	0.201	1.07E-10	8.67E-09	Sig.Decrease	T-helper 17 cell lineage commitment
Atp7a	1322.533	0.464	0.102	5.60E-06	1.84E-04	Sig.Increase	T-helper cell differentiation

References

ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. 31. ISSN 1367-4811. doi: 10.1093/bioinformatics/btv145.

Omar Abdel-Wahab, Ann Mullally, Cyrus Hedvat, Guillermo Garcia-Manero, Jay Patel, Martha Wadleigh, Sebastien Malinge, JinJuan Yao, Outi Kilpivaara, Rukhmi Bhat, Kety Huberman, Sabrena Thomas, Igor Dolgalev, Adriana Heguy, Elisabeth Paietta, Michelle M. Le Beau, Miloslav Beran, Martin S. Tallman, Benjamin L. Ebert, Hagop M. Kantarjian, Richard M. Stone, D. Gary Gilliland, John D. Crispino, and Ross L. Levine. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood*, 114(1):144–147, July 2009. ISSN 0006-4971, 1528-0020. doi: 10.1182/blood-2009-03-210039.

Omar Abdel-Wahab, Taghi Manshouri, Jay Patel, Kelly Harris, Jinjuan Yao, Cyrus Hedvat, Adriana Heguy, Carlos Bueso-Ramos, Hagop Kantarjian, Ross L. Levine, and Srdan Verstovsek. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Research*, 70(2):447–452, January 2010. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-09-3783.

Omar Abdel-Wahab, Jie Gao, Mazhar Adli, Anwesha Dey, Thomas Trimarchi, Young Rock Chung, Cem Kuscu, Todd Hricik, Delphine Ndiaye-Lobry, Lindsay M. LaFave, Richard Koche, Alan H. Shih, Olga A. Guryanova, Eunhee Kim, Sheng Li, Suveg Pandey, Joseph Y. Shin, Leon Telis, Jinfeng Liu, Parva K. Bhatt, Sebastien Monette, Xinyang Zhao, Christopher E. Mason, Christopher Y. Park, Bradley E. Bernstein, Iannis Aifantis, and Ross L. Levine. Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. *Journal of Experimental Medicine*, 210(12): 2641–2659, November 2013. ISSN 0022-1007. doi: 10.1084/jem.20131141.

Michalis Agathocleous, Corbin E. Meacham, Rebecca J. Burgess, Elena Piskounova, Zhiyu Zhao, Genevieve M. Crane, Brianna L. Cowin, Emily Bruner, Malea M. Murphy, Weina Chen, Gerald J. Spangrude, Zeping Hu, Ralph J. DeBerardinis, and Sean J. Morrison. Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature*, 549(7673):476–481, September 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature23876.

Haley M. Amemiya, Anshul Kundaje, and Alan P. Boyle. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports*, 9(1), June 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-45839-z. Number: 1 pages = 9354,.

Jungeun An, Edahí González-Avalos, Ashu Chawla, Mira Jeong, Isaac F. López-Moyado, Wei Li, Margaret A. Goodell, Lukas Chavez, Myunggon Ko, and Anjana Rao. Acute loss of TET function results in aggressive myeloid cancer in mice. *Nature Communications*, 6(1), December 2015. ISSN 2041-1723. doi: 10.1038/ncomms10071.

Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106, October 2010. ISSN 1474-760X. doi: 10.1186/gb-2010-11-10-r106.

Simon Andrews. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data, January .

Kazumasa Aoyama, Motohiko Oshima, Shuhei Koide, Emi Suzuki, Makiko Mochizuki-Kashio, Yuko Kato, Shiro Tara, Daisuke Shinoda, Nobuhiro Hiura, Yaeko Nakajima-Takagi, Goro Sashida, and Atsushi Iwama. Ezh1 Targets Bivalent Genes to Maintain Self-Renewing Stem Cells in Ezh2-Insufficient Myelodysplastic Syndrome. *iScience*, 9:161–174, November 2018. ISSN 2589-0042. doi: 10.1016/j.isci.2018.10.008.

Beatriz Aranda-Orgilles, Ricardo Saldaña-Meyer, Eric Wang, Eirini Trompouki, Anne Fassl, Stephanie Lau, Jasper Mullenders, Pedro P. Rocha, Ramya Raviram, María Guillamot, María Sánchez-Díaz, Kun Wang, Clarisse Kayembe, Nan Zhang, Leonela Amoasii, Avik Choudhuri, Jane A. Skok, Markus Schober, Danny Reinberg, Piotr

Sicinski, Heinrich Schrewe, Aristotelis Tsirigos, Leonard I. Zon, and Iannis Aifantis. MED12 Regulates HSC-Specific Enhancers Independently of Mediator Kinase Activity to Control Hematopoiesis. *Cell Stem Cell*, 19(6):784–799, December 2016. ISSN 1934-5909, 1875-9777. doi: 10.1016/j.stem.2016.08.004.

Martin Bachman, Santiago Uribe-Lewis, Xiaoping Yang, Michael Williams, Adele Murrell, and Shankar Balasubramanian. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature Chemistry*, 6(12):1049–1055, December 2014. ISSN 1755-4349. doi: 10.1038/nchem.2064.

Martin Bachman, Santiago Uribe-Lewis, Xiaoping Yang, Heather E. Burgess, Mario Iurlaro, Wolf Reik, Adele Murrell, and Shankar Balasubramanian. 5-Formylcytosine can be a stable DNA modification in mammals. *Nature Chemical Biology*, 11(8): 555–557, August 2015. ISSN 1552-4469. doi: 10.1038/nchembio.1848.

Joan Barau, Aurélie Teissandier, Natasha Zamudio, Stéphanie Roy, Valérie Nalesso, Yann Hérault, Florian Guillou, and Déborah Bourc'his. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science (New York, N.Y.)*, 354(6314):909–912, 2016. ISSN 1095-9203. doi: 10.1126/science.aah5143.

Denise P. Barlow and Marisa S. Bartolomei. Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*, 6(2):a018382, January 2014. ISSN , 1943-0264. doi: 10.1101/cshperspect.a018382.

Artem Barski, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837, May 2007. ISSN 0092-8674. doi: 10.1016/j.cell.2007.05.009.

M. S. Bartolomei, S. Zemel, and S. M. Tilghman. Parental imprinting of the mouse H19 gene. *Nature*, 351(6322):153–155, May 1991. ISSN 0028-0836. doi: 10.1038/351153a0.

Stephen B. Baylin and Peter A. Jones. A decade of exploring the cancer epigenome — biological and translational implications. *Nature Reviews Cancer*, 11(10):726–734, October 2011. ISSN 1474-175X, 1474-1768. doi: 10.1038/nrc3130.

Stephen B. Baylin and Peter A. Jones. Epigenetic Determinants of Cancer. *Cold Spring Harbor Perspectives in Biology*, 8(9):a019505, September 2016. ISSN 1943-0264. doi: 10.1101/cshperspect.a019505.

C. Beard, E. Li, and R. Jaenisch. Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes & Development*, 9(19):2325–2334, January 1995. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.9.19.2325.

A. M. Beckmann and P. A. Wilce. Egr transcription factors in the nervous system. *Neurochemistry International*, 31(4):477–510; discussion 517–516, October 1997. ISSN 0197-0186. doi: 10.1016/s0197-0186(96)00136-2.

A. C. Bell and G. Felsenfeld. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, 405(6785):482–485, May 2000. ISSN 0028-0836. doi: 10.1038/35013100.

Tord Berggård, Sara Linse, and Peter James. Methods for the detection and analysis of protein–protein interactions. *PROTEOMICS*, 7(16):2833–2842, 2007. ISSN 1615-9861. doi: https://doi.org/10.1002/pmic.200700131.

Benjamin P Berman. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains. *Nature Genetics*, 44(1):9, 2012.

T. Bestor, A. Laudano, R. Mattaliano, and V. Ingram. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology*, 203(4):971–983, October 1988. ISSN 0022-2836. doi: 10.1016/0022-2836(88)90122-2.

T. H. Bestor and V. M. Ingram. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 80(18):5559–5563, September 1983. ISSN 0027-8424. doi: 10.1073/pnas.80.18.5559.

Matthew Bettini, Hongkang Xi, Jeffrey Milbrandt, and Gilbert J. Kersh. Thymocyte Development in Early Growth Response Gene 1-Deficient Mice. *The Journal of Immunology*, 169(4):1713–1720, August 2002. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.169.4.1713.

A. Bird, M. Taggart, M. Frommer, O. J. Miller, and D. Macleod. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1): 91–99, January 1985. ISSN 0092-8674. doi: 10.1016/0092-8674(85)90312-5.

Kathryn Blaschke, Kevin T. Ebata, Mohammad M. Karimi, Jorge A. Zepeda-Martínez, Preeti Goyal, Sahasransu Mahapatra, Angela Tam, Diana J. Laird, Martin Hirst, Anjana Rao, Matthew C. Lorincz, and Miguel Ramalho-Santos. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, 500 (7461):222–226, August 2013. ISSN 1476-4687. doi: 10.1038/nature12362.

Michael J. Booth, Miguel R. Branco, Gabriella Ficz, David Oxley, Felix Krueger, Wolf Reik, and Shankar Balasubramanian. Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science*, 336(6083):934–937, May 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1220671.

Piet Borst and Robert Sabatini. Base J: discovery, biosynthesis, and possible functions. *Annual Review of Microbiology*, 62:235–251, 2008. ISSN 0066-4227. doi: 10.1146/annurev.micro.62.081307.162750.

Magnolia Bostick, Jong Kyong Kim, Pierre-Olivier Estève, Amander Clark, Sriharsa Pradhan, and Steven E. Jacobsen. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, 317(5845):1760–1764, September 2007. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1147939.

D. Bourc'his, G. L. Xu, C. S. Lin, B. Bollman, and T. H. Bestor. Dnmt3L and the establishment of maternal genomic imprints. *Science (New York, N.Y.)*, 294(5551): 2536–2539, December 2001. ISSN 0036-8075. doi: 10.1126/science.1065848.

Robert L. Bowman and Ross L. Levine. TET2 in Normal and Malignant Hematopoiesis. *Cold Spring Harbor Perspectives in Medicine*, 7(8), August 2017. ISSN 2157-1422. doi: 10.1101/cshperspect.a026518.

Malcolm V. Brock, Craig M. Hooker, Emi Ota-Machida, Yu Han, Mingzhou Guo, Stephen Ames, Sabine Glöckner, Steven Piantadosi, Edward Gabrielson, Genevieve Pridham, Kristen Pelosky, Steven A. Belinsky, Stephen C. Yang, Stephen B. Baylin, and James G. Herman. DNA Methylation Markers and Early Recurrence in Stage I Lung Cancer, June 2009.

N. Brockdorff. Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns. *Development*, 138(23):5057–5065, December 2011. ISSN 0950-1991, 1477-9129. doi: 10.1242/dev.065276.

N. Brockdorff, A. Ashworth, G. F. Kay, P. Cooper, S. Smith, V. M. McCabe, D. P. Norris, G. D. Penny, D. Patel, and S. Rastan. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, 351(6324):329–331, May 1991. ISSN 0028-0836. doi: 10.1038/351329a0.

Carolyn J. Brown, Andrea Ballabio, James L. Rupert, Ronald G. Lafreniere, Markus Grompe, Rossana Tonlorenzi, and Huntington F. Willard. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, 349(6304):38–44, January 1991. ISSN 1476-4687. doi: 10.1038/349038a0.

L. Bruhn, A. Munnerlyn, and R. Grosschedl. ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function. *Genes & Development*, 11(5), January 1997. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.11.5.640.

Thomas Burmeister, Claus Meyer, Stefan Schwartz, Julia Hofmann, Mara Molkentin, Eric Kowarz, Björn Schneider, Thorsten Raff, Richard Reinhardt, Nicola Gökbuget,

Dieter Hoelzer, Eckhard Thiel, and Rolf Marschalek. The MLL recombinome of adult CD10-negative B-cell precursor acute lymphoblastic leukemia: results from the GMALL study group. *Blood*, 113(17):4011–4015, April 2009. ISSN 0006-4971. doi: 10.1182/blood-2008-10-183483.

Manuel Buscarlet, Sylvie Provost, Yassamin Feroz Zada, Amina Barhdadi, Vincent Bourgoin, Guylaine Lépine, Luigina Mollica, Natasha Szuber, Marie-Pierre Dubé, and Lambert Busque. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood*, 130(6):753–762, August 2017. ISSN 0006-4971. doi: 10.1182/blood-2017-04-777029.

Lambert Busque, Jay P Patel, Maria E Figueroa, Aparna Vasanthakumar, Sylvie Provost, Zineb Hamilou, Luigina Mollica, Juan Li, Agnes Viale, Adriana Heguy, Maryam Hassimi, Nicholas Socci, Parva K Bhatt, Mithat Gonen, Christopher E Mason, Ari Melnick, Lucy A Godley, Cameron W Brennan, Omar Abdel-Wahab, and Ross L Levine. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nature Genetics*, 44(11):1179–1181, November 2012. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2413.

Dimitrios Cakouros, Sarah Hemming, Kahlia Gronthos, Renjing Liu, Andrew Zannettino, Songtao Shi, and Stan Gronthos. Specific functions of TET1 and TET2 in regulating mesenchymal cell lineage determination. *Epigenetics & Chromatin*, 12(1):3, January 2019. ISSN 1756-8935. doi: 10.1186/s13072-018-0247-4.

Eliezer Calo and Joanna Wysocka. Modification of enhancer chromatin: what, how, and why? *Molecular Cell*, 49(5):825–837, March 2013. ISSN 1097-4164. doi: 10.1016/j.molcel.2013.01.038.

Cancer Genome Atlas Research Network, Timothy J. Ley, Christopher Miller, Li Ding, Benjamin J. Raphael, Andrew J. Mungall, A. Gordon Robertson, Katherine Hoadley, Timothy J. Triche, Peter W. Laird, Jack D. Baty, Lucinda L. Fulton, Robert Fulton, Sharon E. Heath, Joelle Kalicki-Veizer, Cyriac Kandoth, Jeffery M. Klco, Daniel C. Koboldt, Krishna-Latha Kanchi, Shashikant Kulkarni, Tamara L. Lamprecht, David E.

Larson, Ling Lin, Charles Lu, Michael D. McLellan, Joshua F. McMichael, Jacqueline Payton, Heather Schmidt, David H. Spencer, Michael H. Tomasson, John W. Wallis, Lukas D. Wartman, Mark A. Watson, John Welch, Michael C. Wendl, Adrian Ally, Miruna Balasundaram, Inanc Birol, Yaron Butterfield, Readman Chiu, Andy Chu, Eric Chuah, Hye-Jung Chun, Richard Corbett, Noreen Dhalla, Ranabir Guin, An He, Carrie Hirst, Martin Hirst, Robert A. Holt, Steven Jones, Aly Karsan, Darlene Lee, Haiyan I. Li, Marco A. Marra, Michael Mayo, Richard A. Moore, Karen Mungall, Jeremy Parker, Erin Pleasance, Patrick Plettner, Jacquie Schein, Dominik Stoll, Lucas Swanson, Angela Tam, Nina Thiessen, Richard Varhol, Natasja Wye, Yongjun Zhao, Stacey Gabriel, Gad Getz, Carrie Sougnez, Lihua Zou, Mark D. M. Leiserson, Fabio Vandin, Hsin-Ta Wu, Frederick Applebaum, Stephen B. Baylin, Rehan Akbani, Bradley M. Broom, Ken Chen, Thomas C. Motter, Khanh Nguyen, John N. Weinstein, Nianziang Zhang, Martin L. Ferguson, Christopher Adams, Aaron Black, Jay Bowen, Julie Gastier-Foster, Thomas Grossman, Tara Lichtenberg, Lisa Wise, Tanja Davidsen, John A. Demchok, Kenna R. Mills Shaw, Margi Sheth, Heidi J. Sofia, Liming Yang, James R. Downing, and Greg Eley. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England Journal of Medicine*, 368(22): 2059–2074, 2013. ISSN 1533-4406. doi: 10.1056/NEJMoa1301689.

R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science*, 298(5595):1039–1043, November 2002. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1076997.

Ru Cao and Yi Zhang. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular Cell*, 15(1):57–67, July 2004. ISSN 1097-2765. doi: 10.1016/j.molcel.2004.06.020.

Alejandro Carrillo-Jimenez, Özgen Deniz, Maria Victoria Niklison-Chirou, Rocio Ruiz, Karina Bezerra-Salomão, Vassilis Stratoulias, Rachel Amouroux, Ping Kei Yip, Anna Vilalta, Mathilde Cheray, Alexander Michael Scott-Egerton, Eloy Rivas, Khadija Tayara, Irene García-Domínguez, Juan Garcia-Revilla, Juan Carlos Fernandez-Martin,

Ana Maria Espinosa-Oliva, Xianli Shen, Peter St George-Hyslop, Guy Charles Brown, Petra Hajkova, Bertrand Joseph, Jose Luis Venero, Miguel Ramos Branco, and Miguel Angel Burguillos. TET2 Regulates the Neuroinflammatory Response in Microglia. *Cell Reports*, 29(3):697–713.e8, October 2019. ISSN 2211-1247. doi: 10.1016/j.celrep.2019.09.013.

Lynne Csiszar Cary, Michael Goebel, Bartholomew G. Corsaro, Hwei-Gene Wang, Elliot Rosen, and M. J. Fraser. Transposon mutagenesis of baculoviruses: Analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, 172(1):156–169, September 1989. ISSN 0042-6822. doi: 10.1016/0042-6822(89)90117-7.

Raymond G Cavalcante and Maureen A Sartor. annotatr: genomic regions in context. *Bioinformatics*, 33(15):2381–2383, August 2017. ISSN 1367-4803. doi: 10.1093/bioinformatics/btx183.

Edwin Chen, Rebekka K. Schneider, Lawrence J. Breyfogle, Emily A. Rosen, Luke Poveromo, Shannon Elf, Amy Ko, Kristina Brumme, Ross Levine, Benjamin L. Ebert, and Ann Mullally. Distinct effects of concomitant Jak2V617F expression and Tet2 loss in mice promote disease progression in myeloproliferative neoplasms. *Blood*, 125(2):327–335, January 2015. ISSN 0006-4971. doi: 10.1182/blood-2014-04-567024.

Lei-Lei Chen, Huai-Peng Lin, Wen-Jie Zhou, Chen-Xi He, Zhi-Yong Zhang, Zhou-Li Cheng, Jun-Bin Song, Peng Liu, Xin-Yu Chen, Yu-Kun Xia, Xiu-Fei Chen, Ren-Qiang Sun, Jing-Ye Zhang, Yi-Ping Sun, Lei Song, Bing-Jie Liu, Rui-Kai Du, Chen Ding, Fei Lan, Sheng-Lin Huang, Feng Zhou, Suling Liu, Yue Xiong, Dan Ye, and Kun-Liang Guan. SNIP1 Recruits TET2 to Regulate c-MYC Target Genes and Cellular DNA Damage Response. *Cell Reports*, 25(6):1485–1500.e4, 2018. ISSN 2211-1247. doi: 10.1016/j.celrep.2018.10.028.

Qiang Chen, Yibin Chen, Chunjing Bian, Ryoji Fujiki, and Xiaochun Yu. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, 493(7433):561–564, January 2013. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11742.

Taiping Chen, Yoshihide Ueda, Jonathan E. Dodge, Zhenjuan Wang, and En Li. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, 23(16):5594–5605, August 2003a. ISSN 0270-7306. doi: 10.1128/mcb.23.16.5594-5605.2003.

Wen Yong Chen, Xiaobei Zeng, Mark G. Carter, Craig N. Morrell, Ray-Whay Chiu Yen, Manel Esteller, D. Neil Watkins, James G. Herman, Joseph L. Mankowski, and Stephen B. Baylin. Heterozygous disruption of Hic1 predisposes mice to a gender-dependent spectrum of malignant tumors. *Nature Genetics*, 33(2):197–202, February 2003b. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng1077.

Jijun Cheng, Shangqin Guo, Suning Chen, Stephen J. Mastriano, Chaochun Liu, Ana C. D'Alessio, Eriona Hysolli, Yanwen Guo, Hong Yao, Cynthia M. Megyola, Dan Li, Jun Liu, Wen Pan, Christine A. Roden, Xiao-Ling Zhou, Kartoosh Heydari, Jianjun Chen, In-Hyun Park, Ye Ding, Yi Zhang, and Jun Lu. An Extensive Network of TET2-Targeting MicroRNAs Regulates Malignant Hematopoiesis. *Cell Reports*, 5(2): 471–481, October 2013. ISSN 22111247. doi: 10.1016/j.celrep.2013.08.050.

Shigeru Chiba. Dysregulation of TET2 in hematologic malignancies. *International Journal of Hematology*, 105(1):17–22, January 2017. ISSN 1865-3774. doi: 10.1007/s12185-016-2122-z.

Eunjoo Choi-Rhee, Howard Schulman, and John E. Cronan. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Science*, 13(11), 2004. ISSN 1469-896X. doi: 10.1110/ps.04911804.

Wen-Chien Chou, Sheng-Chieh Chou, Chieh-Yu Liu, Chien-Yuan Chen, Hsin-An Hou, Yuan-Yeh Kuo, Ming-Cheng Lee, Bor-Sheng Ko, Jih-Luh Tang, Ming Yao, Woei Tsay, Shang-Ju Wu, Shang-Yi Huang, Szu-Chun Hsu, Yao-Chang Chen, Yi-Chang Chang, Yi-Yi Kuo, Kuan-Ting Kuo, Fen-Yu Lee, Ming-Chi Liu, Chia-Wen Liu, Mei-Hsuan Tseng, Chi-Fei Huang, and Hwei-Fang Tien. TET2 mutation is an

unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood*, 118(14):3803–3810, October 2011. ISSN 0006-4971. doi: 10.1182/blood-2011-02-339747.

Yajing Chu, Zhigang Zhao, David Wayne Sant, Ganqian Zhu, Sarah M. Greenblatt, Lin Liu, Jinhuan Wang, Zeng Cao, Jeanette Cheng Tho, Shi Chen, Xiaochen Liu, Peng Zhang, Jaroslaw P. Maciejewski, Stephen Nimer, Gaofeng Wang, Weiping Yuan, Feng-Chun Yang, and Mingjiang Xu. Tet2 Regulates Osteoclast Differentiation by Interacting with Runx1 and Maintaining Genomic 5-Hydroxymethylcytosine (5hmC). *Genomics, Proteomics & Bioinformatics*, 16(3):172–186, 2018. ISSN 2210-3244. doi: 10.1016/j.gpb.2018.04.005.

Luisa Cimmino, Meelad M Dawlaty, Delphine Ndiaye-Lobry, Yoon Sing Yap, Sofia Bakogianni, Yiting Yu, Sanchari Bhattacharyya, Rita Shaknovich, Huimin Geng, Camille Lobry, Jasper Mullenders, Bryan King, Thomas Trimarchi, Beatriz Aranda-Orgilles, Cynthia Liu, Steven Shen, Amit K Verma, Rudolf Jaenisch, and Iannis Aifantis. TET1 is a tumor suppressor of hematopoietic malignancy. *Nature Immunology*, 16(6): 653–662, June 2015. ISSN 1529-2908, 1529-2916. doi: 10.1038/ni.3148.

Luisa Cimmino, Igor Dolgalev, Yubao Wang, Akihide Yoshimi, Gaëlle H. Martin, Jingjing Wang, Victor Ng, Bo Xia, Matthew T. Witkowski, Marisa Mitchell-Flack, Isabella Grillo, Sofia Bakogianni, Delphine Ndiaye-Lobry, Miguel Torres Martín, Maria Guillamot, Robert S. Banh, Mingjiang Xu, Maria E. Figueroa, Ross A. Dickins, Omar Abdel-Wahab, Christopher Y. Park, Aristotelis Tsirigos, Benjamin G. Neel, and Iannis Aifantis. Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell*, 170(6):1079–1095.e20, September 2017. ISSN 1097-4172. doi: 10.1016/j.cell.2017.07.032.

M. Clamp, B. Fry, M. Kamal, X. Xie, J. Cuff, M. F. Lin, M. Kellis, K. Lindblad-Toh, and E. S. Lander. Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences*, 104(49):19428–19433, December 2007. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0709013104.

Samuel Collombet, Chris van Oevelen, Jose Luis Sardina Ortega, Wassim Abou-Jaoudé, Bruno Di Stefano, Morgane Thomas-Chollier, Thomas Graf, and Denis Thieffry. Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *Proceedings of the National Academy of Sciences*, 114(23):5792–5799, June 2017. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1610622114.

Giacomo Coltro, Abhishek A. Mangaonkar, Terra L. Lasho, Christy M. Finke, Prateek Pophali, Ryan Carr, Naseema Gangat, Moritz Binder, Animesh Pardanani, Martin Fernandez-Zapico, Keith D. Robertson, Alberto Bosi, Nathalie Droin, Alessandro M. Vannucchi, Ayalew Tefferi, Anthony Hunter, Eric Padron, Eric Solary, and Mrinal M. Patnaik. Clinical, molecular, and prognostic correlates of number, type, and functional localization of TET2 mutations in chronic myelomonocytic leukemia (CMML)-a study of 1084 patients. *Leukemia*, December 2019. ISSN 1476-5551. doi: 10.1038/s41375-019-0690-7.

Marcel W. Coolen, Clare Stirzaker, Jenny Z. Song, Aaron L. Statham, Zena Kassir, Carlos S. Moreno, Andrew N. Young, Vijay Varma, Terence P. Speed, Mark Cowley, Paul Lacaze, Warren Kaplan, Mark D. Robinson, and Susan J. Clark. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. *Nature Cell Biology*, 12(3):235–246, March 2010. ISSN 1476-4679. doi: 10.1038/ncb2023.

D N Cooper, M H Taggart, and A P Bird. Unmethylated domains in vertebrate DNA. *Nucleic Acids Research*, 11(3):647–658, February 1983. ISSN 0305-1048.

David N. Cooper and Michael Krawczak. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human Genetics*, 83(2):181–188, September 1989. ISSN 1432-1203. doi: 10.1007/BF00286715.

M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. Histone H3K27ac separates active from poised enhancers and predicts

developmental state. *Proceedings of the National Academy of Sciences*, 107(50):21931–21936, December 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1016071107.

John E. Cronan. Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase. *The Journal of Nutritional Biochemistry*, 16(7):416–418, July 2005. ISSN 0955-2863. doi: 10.1016/j.jnutbio.2005.03.017.

Hai-Qiang Dai, Bang-An Wang, Lu Yang, Jia-Jia Chen, Guo-Chun Zhu, Mei-Ling Sun, Hao Ge, Rui Wang, Deborah L. Chapman, Fuchou Tang, Xin Sun, and Guo-Liang Xu. TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. *Nature*, 538(7626):528–532, October 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature20095.

Lenny Dang, David W. White, Stefan Gross, Bryson D. Bennett, Mark A. Bittinger, Edward M. Driggers, Valeria R. Fantin, Hyun Gyung Jang, Shengfang Jin, Marie C. Keenan, Kevin M. Marks, Robert M. Prins, Patrick S. Ward, Katharine E. Yen, Linda M. Liau, Joshua D. Rabinowitz, Lewis C. Cantley, Craig B. Thompson, Matthew G. Vander Heiden, and Shinsan M. Su. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*, 462(7274):739–744, December 2009. ISSN 1476-4687. doi: 10.1038/nature08617.

E. Dassé, G. Volpe, D. S. Walton, N. Wilson, W. Del Pozzo, L. P. O'Neill, R. K. Slany, J. Frampton, and S. Dumon. Distinct regulation of c-myb gene expression by HoxA9, Meis1 and Pbx proteins in normal hematopoietic progenitors and transformed myeloid cells. *Blood Cancer Journal*, 2(6), June 2012. ISSN 2044-5385. doi: 10.1038/bcj.2012.20.

Meelad M. Dawlaty, Achim Breiling, Thuc Le, M. Inmaculada Barrasa, Günter Raddatz, Qing Gao, Benjamin E. Powell, Albert W. Cheng, Kym F. Faull, Frank Lyko, and Rudolf Jaenisch. Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells. *Developmental Cell*, 29(1):102–111, April 2014. ISSN 1534-5807. doi: 10.1016/j.devcel.2014.03.003.

Meelad M. Dawlaty, Kibibi Ganz, Benjamin E. Powell, Yueh-Chiang Hu, Styliani Markoulaki, Albert W. Cheng, Qing Gao, Jongpil Kim, Sang-Woon Choi, David C. Page, and Rudolf Jaenisch. Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. *Cell Stem Cell*, 9(2): 166–175, August 2011. ISSN 19345909. doi: 10.1016/j.stem.2011.07.010.

Meelad M. Dawlaty, Achim Breiling, Thuc Le, Günter Raddatz, M. Inmaculada Barrasa, Albert W. Cheng, Qing Gao, Benjamin E. Powell, Zhe Li, Mingjiang Xu, Kym F. Faull, Frank Lyko, and Rudolf Jaenisch. Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Developmental Cell*, 24(3):310–323, February 2013. ISSN 15345807. doi: 10.1016/j.devcel.2012.12.015.

A. P. Jason de Koning, Wanjun Gu, Todd A. Castoe, Mark A. Batzer, and David D. Pollock. Repetitive elements may comprise over two-thirds of the human genome. *PLoS genetics*, 7(12):e1002384, December 2011. ISSN 1553-7404. doi: 10.1371/journal.pgen.1002384.

Lorenzo de la Rica, Javier Rodríguez-Ubreva, Mireia García, Abul B. M. M. K. Islam, José M. Urquiza, Henar Hernando, Jesper Christensen, Kristian Helin, Carmen Gómez-Vaquero, and Esteban Ballestar. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biology*, 14(9):R99, 2013. ISSN 1474-760X. doi: 10.1186/gb-2013-14-9-r99.

T. M. DeChiara, E. J. Robertson, and A. Efstratiadis. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, 64(4):849–859, February 1991. ISSN 0092-8674. doi: 10.1016/0092-8674(91)90513-x.

François Delhommeau, Véronique Della Valle, Aline Massé, Jean-Pierre Le Couedic, Yann Lécluse, Christophe Marzac, Catherine Lacombe, Philippe Dessen, Michaela Fontenay, and Olivier A Bernard. Mutation in TET2 in Myeloid Cancers. *The New England Journal of Medicine*, page 13, 2009.

Rachel Deplus, Benjamin Delatte, Marie K Schwinn, Matthieu Defrance, Jacqui Méndez, Nancy Murphy, Mark A Dawson, Michael Volkmar, Pascale Putmans, Emilie Calonne, Alan H Shih, Ross L Levine, Olivier Bernard, Thomas Mercher, Eric Solary, Marjeta Urh, Danette L Daniels, and François Fuks. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO Journal*, 32(5): 645–655, January 2013. ISSN 0261-4189, 1460-2075. doi: 10.1038/emboj.2012.357.

Sheng Ding, Xiaohui Wu, Gang Li, Min Han, Yuan Zhuang, and Tian Xu. Efficient Transposition of the piggyBac (PB) Transposon in Mammalian Cells and Mice. *Cell*, 122(3):473–483, August 2005. ISSN 0092-8674. doi: 10.1016/j.cell.2005.07.013.

Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1):15–21, January 2013. ISSN 1367-4811. doi: 10.1093/bioinformatics/bts635.

R. A. Drewell. Methylation-dependent silencing at the H19 imprinting control region by MeCP2. *Nucleic Acids Research*, 30(5):1139–1144, March 2002. ISSN 13624962. doi: 10.1093/nar/30.5.1139.

Wade H. Dunham, Michael Mullin, and Anne-Claude Gingras. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *PROTEOMICS*, 12(10): 1576–1590, 2012. ISSN 1615-9861. doi: https://doi.org/10.1002/pmic.201100523.

D. B. Dunn and J. D. Smith. Occurrence of a New Base in the Deoxyribonucleic Acid of a Strain of Bacterium Coli. *Nature*, 175(4451):336–337, February 1955. ISSN 1476-4687. doi: 10.1038/175336a0.

Jason Ernst and Manolis Kellis. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology*, 28(8), August 2010. ISSN 1546-1696. doi: 10.1038/nbt.1662. Number: 8 pages = 817–825,.

Celestia Fang, Yu Qiao, Se Hwan Mun, Min Joon Lee, Koichi Murata, Seyeon Bae, Baohong Zhao, Kyung-Hyun Park-Min, and Lionel B. Ivashkiv. Cutting Edge: EZH2

Promotes Osteoclastogenesis by Epigenetic Silencing of the Negative Regulator IRF8. *The Journal of Immunology*, 196(11):4452–4456, June 2016. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1501466.

Krueger Felix. Babraham Bioinformatics - Trim Galore!

S. Feng, S. J. Cokus, X. Zhang, P.-Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan, M. Pellegrini, and S. E. Jacobsen. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, 107(19):8689–8694, May 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1002720107.

Yimei Feng, Xiaoping Li, Kaniel Cassady, Zhongmin Zou, and Xi Zhang. TET2 Function in Hematopoietic Malignancies, Immune Regulation, and DNA Repair. *Frontiers in Oncology*, 9, April 2019. ISSN 2234-943X. doi: 10.3389/fonc.2019.00210.

A. C. Ferguson-Smith, B. M. Cattanach, S. C. Barton, C. V. Beechey, and M. A. Surani. Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature*, 351(6328):667–670, June 1991. ISSN 1476-4687. doi: 10.1038/351667a0.

Gabriella Ficz and John G. Gribben. Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence? *Genomics*, 104(5):352–357, November 2014. ISSN 0888-7543. doi: 10.1016/j.ygeno.2014.08.017.

Gabriella Ficz, Miguel R. Branco, Stefanie Seisenberger, Fátima Santos, Felix Krueger, Timothy A. Hore, C. Joana Marques, Simon Andrews, and Wolf Reik. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, 473(7347):398–402, May 2011. ISSN 1476-4687. doi: 10.1038/nature10008.

Maria E. Figueroa, Omar Abdel-Wahab, Chao Lu, Patrick S. Ward, Jay Patel, Alan Shih, Yushan Li, Neha Bhagwat, Aparna Vasanthakumar, Hugo F. Fernandez, Martin S. Tallman, Zhuoxin Sun, Kristy Wolniak, Justine K. Peeters, Wei Liu, Sung E. Choe, Valeria R. Fantin, Elisabeth Paietta, Bob Löwenberg, Jonathan D. Licht, Lucy A.

Godley, Ruud Delwel, Peter J.M. Valk, Craig B. Thompson, Ross L. Levine, and Ari Melnick. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell*, 18(6):553–567, December 2010. ISSN 15356108. doi: 10.1016/j.ccr.2010.11.015.

Carina Frauer, Andrea Rottach, Daniela Meilinger, Sebastian Bultmann, Karin Fellinger, Stefan Hasenöder, Mengxi Wang, Weihua Qin, Johannes Söding, Fabio Spada, and Heinrich Leonhardt. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PloS One*, 6(2):e16627, February 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0016627.

M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, and C. L. Paul. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, 89(5):1827–1831, March 1992. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.89.5.1827.

Makiko Fujii, Lyudmila A. Lyakh, Cameron P. Bracken, Junya Fukuoka, Morisada Hayakawa, Tadasuke Tsukiyama, Steven J. Soll, Melissa Harris, Sonia Rocha, Kevin C. Roche, Shin-ichi Tominaga, Jin Jen, Neil D. Perkins, Robert J. Lechleider, and Anita B. Roberts. SNIP1 Is a Candidate Modifier of the Transcriptional Activity of c-Myc on E Box-Dependent Target Genes. *Molecular Cell*, 24(5):771–783, December 2006. ISSN 1097-2765. doi: 10.1016/j.molcel.2006.11.006.

Ryoji Fujiki, Waka Hashiba, Hiroki Sekine, Atsushi Yokoyama, Toshihiro Chikanishi, Saya Ito, Yuuki Imai, Jaehoon Kim, Housheng Hansen He, Katsuhide Igarashi, Jun Kanno, Fumiaki Ohtake, Hirochika Kitagawa, Robert G. Roeder, Myles Brown, and Shigeaki Kato. GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature*, 480(7378):557–560, December 2011. ISSN 1476-4687. doi: 10.1038/nature10656.

Guillermo Garcia-Manero and Pierre Fenaux. Hypomethylating Agents and Other Novel Strategies in Myelodysplastic Syndromes. *Journal of Clinical Oncology*, 29(5):516–523, February 2011. ISSN 0732-183X. doi: 10.1200/JCO.2010.31.0854.

María G. García, Antonella Carella, Rocío G. Urdinguio, Gustavo F. Bayón, Virginia Lopez, Juan Ramón Tejedor, Marta I. Sierra, Estela García-Toraño, Pablo Santamarina, Raúl F. Perez, Cristina Mangas, Aurora Astudillo, M. Daniela Corte-Torres, Inés Sáenz-de Santa-María, María-Dolores Chiara, Agustín F. Fernández, and Mario F. Fraga. Epigenetic dysregulation of TET2 in human glioblastoma. *Oncotarget*, 9(40): 25922–25934, May 2018. ISSN 1949-2553. doi: 10.18632/oncotarget.25406.

Anne-Valerie Gendrel, Anwyn Apedaile, Heather Coker, Ausma Termanis, Ilona Zvetkova, Jonathan Godwin, Y. Amy Tang, Derek Huntley, Giovanni Montana, Steven Taylor, Eleni Giannoulatou, Edith Heard, Irina Stancheva, and Neil Brockdorff. Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. *Developmental Cell*, 23 (2):265–279, August 2012. ISSN 1534-5807. doi: 10.1016/j.devcel.2012.06.011.

Daniel Globisch, Martin Münzel, Markus Müller, Stylianos Michalakis, Mirko Wagner, Susanne Koch, Tobias Brückl, Martin Biel, and Thomas Carell. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PloS One*, 5(12):e15367, December 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0015367.

M. A. Glozak and E. Seto. Histone deacetylases and cancer. *Oncogene*, 26(37):5420–5432, August 2007. ISSN 1476-5594. doi: 10.1038/sj.onc.1210610.

Mary Grace Goll and Timothy H. Bestor. Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, 74:481–514, 2005. ISSN 0066-4154. doi: 10.1146/annurev.biochem.74.010904.153721.

Mary Grace Goll, Finn Kirpekar, Keith A. Maggert, Jeffrey A. Yoder, Chih-Lin Hsieh, Xiaoyu Zhang, Kent G. Golic, Steven E. Jacobsen, and Timothy H. Bestor. Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2. *Science*, 311(5759): 395–398, January 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1120976.

M Gossen, S Freundlieb, G Bender, G Muller, W Hillen, and H Bujard. Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268(5218):1766–1769, June 1995. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.7792603.

Maxim V. C. Greenberg and Deborah Bourc'his. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, 20(10):590–607, October 2019. ISSN 1471-0072, 1471-0080. doi: 10.1038/s41580-019-0159-6.

Hongcang Gu, Zachary D. Smith, Christoph Bock, Patrick Boyle, Andreas Gnirke, and Alexander Meissner. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*, 6(4):468–481, April 2011a. ISSN 1750-2799. doi: 10.1038/nprot.2010.190.

Tian-Peng Gu, Fan Guo, Hui Yang, Hai-Ping Wu, Gui-Fang Xu, Wei Liu, Zhi-Guo Xie, Linyu Shi, Xinyi He, Seung-gi Jin, Khursheed Iqbal, Yujiang Geno Shi, Zixin Deng, Piroska E. Szabó, Gerd P. Pfeifer, Jinsong Li, and Guo-Liang Xu. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 477(7366):606–610, September 2011b. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature10443.

Diana Guallar, Xianju Bi, Jose Angel Pardavila, Xin Huang, Carmen Saenz, Xianle Shi, Hongwei Zhou, Francesco Faiola, Junjun Ding, Phensinee Haruehanroengra, Fan Yang, Dan Li, Carlos Sanchez-Priego, Arven Saunders, Feng Pan, Victor Julian Valdes, Kevin Kelley, Miguel G. Blanco, Lingyi Chen, Huayan Wang, Jia Sheng, Mingjiang Xu, Miguel Fidalgo, Xiaohua Shen, and Jianlong Wang. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nature Genetics*, 50(3):443–451, March 2018. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-018-0060-9.

Dongyin Guan and Hung-Ying Kao. The function, regulation and therapeutic implications of the tumor suppressor protein, PML. *Cell & Bioscience*, 5(1):60, November 2015. ISSN 2045-3701. doi: 10.1186/s13578-015-0051-9.

Yaoting Gui, Guangwu Guo, Yi Huang, Xueda Hu, Aifa Tang, Shengjie Gao, Renhua Wu, Chao Chen, Xianxin Li, Liang Zhou, Minghui He, Zesong Li, Xiaojuan Sun, Wenlong Jia, Jinnong Chen, Shangming Yang, Fangjian Zhou, Xiaokun Zhao, Shengqing Wan, Rui Ye, Chaozhao Liang, Zhisheng Liu, Peide Huang, Chunxiao Liu, Hui Jiang, Yong Wang, Hancheng Zheng, Liang Sun, Xingwang Liu, Zhimao Jiang, Dafei Feng, Jing Chen, Song Wu, Jing Zou, Zhongfu Zhang, Ruilin Yang, Jun Zhao, Congjie Xu, Weihua Yin, Zhichen Guan, Jiongxian Ye, Hong Zhang, Jingxiang Li, Karsten Kristiansen, Michael L. Nickerson, Dan Theodorescu, Yingrui Li, Xiuqing Zhang, Songgang Li, Jian Wang, Huanming Yang, Jun Wang, and Zhiming Cai. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nature Genetics*, 43(9):875–878, September 2011. ISSN 1546-1718. doi: 10.1038/ng.907.

Paul Guilhamon, Malihe Eskandarpour, Dina Halai, Gareth A. Wilson, Andrew Feber, Andrew E. Teschendorff, Valenti Gomez, Alexander Hergovich, Roberto Tirabosco, M. Fernanda Amary, Daniel Baumhoer, Gernot Jundt, Mark T. Ross, Adrienne M. Flanagan, and Stephan Beck. Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nature Communications*, 4(1), October 2013. ISSN 2041-1723. doi: 10.1038/ncomms3166.

Fan Guo, Xianlong Li, Dan Liang, Tong Li, Ping Zhu, Hongshan Guo, Xinglong Wu, Lu Wen, Tian-Peng Gu, Boqiang Hu, Colum P. Walsh, Jinsong Li, Fuchou Tang, and Guo-Liang Xu. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell*, 15(4):447–459, October 2014a. ISSN 1934-5909. doi: 10.1016/j.stem.2014.08.003.

Hong Guo, Ou Ma, Nancy A. Speck, and Alan D. Friedman. Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis. *Blood*, 119(19):4408–4418, May 2012. ISSN 0006-4971. doi: 10.1182/blood-2011-12-397091.

Hong Guo, Ou Ma, and Alan D. Friedman. The Cebpa +37-kb enhancer directs transgene expression to myeloid progenitors and to long-term hematopoietic stem cells. *Journal*

of Leukocyte Biology, 96(3):419–426, September 2014b. ISSN 1938-3673. doi: 10.1189/jlb.2AB0314-145R.

Jamie A. Hackett, Roopsha Sengupta, Jan J. Zylicz, Kazuhiro Murakami, Caroline Lee, Thomas A. Down, and M. Azim Surani. Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science*, 339(6118):448–452, January 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1229277.

Petra Hajkova, Sylvia Erhardt, Natasha Lane, Thomas Haaf, Osman El-Maarri, Wolf Reik, Jörn Walter, and M. Azim Surani. Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development*, 117(1):15–23, September 2002. ISSN 0925-4773. doi: 10.1016/S0925-4773(02)00181-8.

Kasper Daniel Hansen, Winston Timp, Héctor Corrada Bravo, Sarven Sabunciyan, Benjamin Langmead, Oliver G. McDonald, Bo Wen, Hao Wu, Yun Liu, Dinh Diep, Eirikur Briem, Kun Zhang, Rafael A. Irizarry, and Andrew P. Feinberg. Increased methylation variation in epigenetic domains across cancer types. *Nature Genetics*, 43 (8):768–775, August 2011. ISSN 1546-1718. doi: 10.1038/ng.865.

Klaus H. Hansen, Adrian P. Bracken, Diego Pasini, Nikolaj Dietrich, Simmi S. Gehani, Astrid Monrad, Juri Rappsilber, Mads Lerdrup, and Kristian Helin. A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biology*, 10(11):1291–1300, November 2008. ISSN 1476-4679. doi: 10.1038/ncb1787.

Ross C. Hardison, Yu Zhang, Cheryl A. Keller, Guanjue Xiang, Elisabeth F. Heuston, Lin An, Jens Lichtenberg, Belinda M. Giardine, David Bodine, Shaun Mahony, Qunhua Li, Feng Yue, Mitchell J. Weiss, Gerd A. Blobel, James Taylor, Jim Hughes, Douglas R. Higgs, and Berthold Göttgens. Systematic integration of GATA transcription factors and epigenomes via IDEAS paints the regulatory landscape of hematopoietic cells. *IUBMB Life*, 72(1):27–38, 2020. ISSN 1521-6551. doi: https://doi.org/10.1002/iub.2195.

A. T. Hark, C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse, and S. M. Tilghman. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, 405(6785):486–489, May 2000. ISSN 0028-0836. doi: 10.1038/35013106.

Hideharu Hashimoto, Yusuf Olatunde Olanrewaju, Yu Zheng, Geoffrey G. Wilson, Xing Zhang, and Xiaodong Cheng. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes & Development*, 28(20):2304–2313, October 2014. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.250746.114.

Nicholas D. Hastie. Wilms' tumour 1 (WT1) in development, homeostasis and disease. *Development*, 144(16):2862–2872, August 2017. ISSN 0950-1991, 1477-9129. doi: 10.1242/dev.153163.

Kenichiro Hata, Masaki Okano, Hong Lei, and En Li. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development (Cambridge, England)*, 129(8):1983–1993, April 2002. ISSN 0950-1991.

Megan A. Hatlen, Kanika Arora, Vladimir Vacic, Ewa A. Grabowska, Willey Liao, Bridget Riley-Gillis, Dayna M. Oschwald, Lan Wang, Jacob E. Joergens, Alan H. Shih, Franck Rapaport, Shengqing Gu, Francesca Voza, Takashi Asai, Benjamin G. Neel, Michael G. Kharas, Mithat Gonen, Ross L. Levine, and Stephen D. Nimer. Integrative genetic analysis of mouse and human AML identifies cooperating disease alleles. *Journal of Experimental Medicine*, 213(1):25–34, January 2016. ISSN 0022-1007. doi: 10.1084/jem.20150524.

Simon Hauri, Federico Comoglio, Makiko Seimiya, Moritz Gerstung, Timo Glatter, Klaus Hansen, Ruedi Aebersold, Renato Paro, Matthias Gstaiger, and Christian Beisel. A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Reports*, 17(2):583–595, October 2016. ISSN 2211-1247. doi: 10.1016/j.celrep.2016.08.096.

Yu-Fei He, Bin-Zhong Li, Zheng Li, Peng Liu, Yang Wang, Qingyu Tang, Jianping Ding, Yingying Jia, Zhangcheng Chen, Lin Li, Yan Sun, Xiuxue Li, Qing Dai, Chun-Xiao Song, Kangling Zhang, Chuan He, and Guo-Liang Xu. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, 333(6047): 1303–1307, September 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1210944.

Nathaniel D. Heintzman, Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, Leonard K. Lee, Rhona K. Stuart, Christina W. Ching, Keith A. Ching, Jessica E. Antosiewicz-Bourget, Hui Liu, Xinmin Zhang, Roland D. Green, Victor V. Lobanenkov, Ron Stewart, James A. Thomson, Gregory E. Crawford, Manolis Kellis, and Bing Ren. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243):108–112, May 2009. ISSN 1476-4687. doi: 10.1038/nature07829.

Sven Heinz, Casey E. Romanoski, Christopher Benner, and Christopher K. Glass. The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology*, 16(3), March 2015. ISSN 1471-0080. doi: 10.1038/nrm3949. Number: 3 pages = 144–154,.

A. Hellman and A. Chess. Gene Body-Specific Methylation on the Active X Chromosome. *Science*, 315(5815):1141–1143, February 2007. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1136352.

Patricia Heyn, Clare V. Logan, Adeline Fluteau, Rachel C. Challis, Tatsiana Auchynnikava, Carol-Anne Martin, Joseph A. Marsh, Francesca Taglini, Fiona Kilanowski, David A. Parry, Valerie Cormier-Daire, Chin-To Fong, Kate Gibson, Vivian Hwa, Lourdes Ibáñez, Stephen P. Robertson, Giorgia Sebastiani, Juri Rappsilber, Robin C. Allshire, Martin A. M. Reijns, Andrew Dauber, Duncan Sproul, and Andrew P. Jackson. Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nature Genetics*, 51(1):96–105, January 2019. ISSN 1546-1718. doi: 10.1038/s41588-018-0274-x.

Peter W. S. Hill, Rachel Amouroux, and Petra Hajkova. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: An emerging complex story. *Genomics*, 104(5):324–333, November 2014. ISSN 0888-7543. doi: 10.1016/j.ygeno.2014.08.012.

Hideyo Hirai, Pu Zhang, Tajhal Dayaram, Christopher J. Hetherington, Shin-ichi Mizuno, Jiro Imanishi, Koichi Akashi, and Daniel G. Tenen. C/EBP is required for 'emergency'

granulopoiesis. *Nature Immunology*, 7(7):732–739, July 2006. ISSN 1529-2916. doi: 10.1038/ni1354.

Cassandra M. Hirsch, Aziz Nazha, Kassy Kneen, Mohamed E. Abazeed, Manja Meggendorfer, Bartlomiej P. Przychodzen, Niroshan Nadarajah, Vera Adema, Yasunobu Nagata, Abhinav Goyal, Hassan Awada, Mohammad Fahad Asad, Valeria Visconte, Yihong Guan, Mikkael A. Sekeres, Ryszard Olinski, Babal Kant Jha, Thomas LaFramboise, Tomas Radivoyevitch, Torsten Haferlach, and Jaroslaw P. Maciejewski. Consequences of mutant TET2 on clonality and subclonal hierarchy. *Leukemia*, 32(8):1751–1761, August 2018. ISSN 1476-5551. doi: 10.1038/s41375-018-0150-9.

Amber Hogart, Jens Lichtenberg, Subramanian S. Ajay, Stacie Anderson, Elliott H. Margulies, and David M. Bodine. Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. *Genome Research*, 22(8):1407–1418, August 2012. ISSN 1088-9051. doi: 10.1101/gr.132878.111.

R. Holliday and G. W. Grigg. DNA methylation and mutation. *Mutation Research*, 285 (1):61–67, January 1993. ISSN 0027-5107. doi: 10.1016/0027-5107(93)90052-h.

Daniel Holoch and Raphaël Margueron. Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends in Biochemical Sciences*, 42(7):531–542, July 2017. ISSN 0968-0004. doi: 10.1016/j.tibs.2017.04.003.

Gary C. Hon, Chun-Xiao Song, Tingting Du, Fulai Jin, Siddarth Selvaraj, Ah Young Lee, Chia-An Yen, Zhen Ye, Shi-Qing Mao, Bang-An Wang, Samantha Kuan, Lee E. Edsall, Boxuan Simen Zhao, Guo-Liang Xu, Chuan He, and Bing Ren. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Molecular Cell*, 56(2):286–297, October 2014. ISSN 1097-4164. doi: 10.1016/j.molcel.2014.08.026.

Timothy Alexander Hore, Ferdinand von Meyenn, Mirunalini Ravichandran, Martin Bachman, Gabriella Ficz, David Oxley, Fátima Santos, Shankar Balasubramanian,

Tomasz P. Jurkowski, and Wolf Reik. Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. *Proceedings of the National Academy of Sciences*, 113(43):12202–12207, October 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1608679113.

R. D. Hotchkiss. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry*, 175(1):315–332, August 1948. ISSN 0021-9258 0021-9258.

Hsin-An Hou, Tai-Chung Huang, Liang-In Lin, Chieh-Yu Liu, Chien-Yuan Chen, Wen-Chien Chou, Jih-Luh Tang, Mei-Hsuan Tseng, Chi-Fei Huang, Ying-Chieh Chiang, Fen-Yu Lee, Ming-Chih Liu, Ming Yao, Shang-Yi Huang, Bor-Sheng Ko, Szu-Chun Hsu, Shang-Ju Wu, Woei Tsay, Yao-Chang Chen, and Hwei-Fang Tien. WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system. *Blood*, 115(25): 5222–5231, June 2010. ISSN 0006-4971. doi: 10.1182/blood-2009-12-259390. Publisher: American Society of Hematology.

Carina Y. Howell, Timothy H. Bestor, Feng Ding, Keith E. Latham, Carmen Mertineit, Jacquetta M. Trasler, and J. Richard Chaillet. Genomic Imprinting Disrupted by a Maternal Effect Mutation in the Dnmt1 Gene. *Cell*, 104(6):829–838, March 2001. ISSN 0092-8674. doi: 10.1016/S0092-8674(01)00280-X.

Chih-Hung Hsu, Kai-Lin Peng, Ming-Lun Kang, Yi-Ren Chen, Yu-Chih Yang, Chin-Hsien Tsai, Chi-Shuen Chu, Yung-Ming Jeng, Yen-Ting Chen, Feng-Mao Lin, Hsien-Da Huang, Yun-Yuh Lu, Yu-Ching Teng, Shinn-Tsuen Lin, Ruo-Kai Lin, Fan-Mei Tang, Sung-Bau Lee, Huan Ming Hsu, Jyh-Cherng Yu, Pei-Wen Hsiao, and Li-Jung Juan. TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. *Cell Reports*, 2(3):568–579, September 2012. ISSN 2211-1247. doi: 10.1016/j.celrep.2012.08.030.

Lulu Hu, Ze Li, Jingdong Cheng, Qinhui Rao, Wei Gong, Mengjie Liu, Yujiang Geno Shi, Jiayu Zhu, Ping Wang, and Yanhui Xu. Crystal structure of TET2-DNA complex:

insight into TET-mediated 5mC oxidation. *Cell*, 155(7):1545–1555, December 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.11.020.

Lulu Hu, Junyan Lu, Jingdong Cheng, Qinhui Rao, Ze Li, Haifeng Hou, Zhiyong Lou, Lei Zhang, Wei Li, Wei Gong, Mengjie Liu, Chang Sun, Xiaotong Yin, Jie Li, Xiangshi Tan, Pengcheng Wang, Yinsheng Wang, Dong Fang, Qiang Cui, Pengyuan Yang, Chuan He, Hualiang Jiang, Cheng Luo, and Yanhui Xu. Structural insight into substrate preference for TET-mediated oxidation. *Nature*, 527(7576):118–122, November 2015. ISSN 1476-4687. doi: 10.1038/nature15713.

Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4 (1):44–57, 2009. ISSN 1750-2799. doi: 10.1038/nprot.2008.211.

Y. Huang, L. Chavez, X. Chang, X. Wang, W. A. Pastor, J. Kang, J. A. Zepeda-Martinez, U. J. Pape, S. E. Jacobsen, B. Peters, and A. Rao. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences*, 111(4):1361–1366, January 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1322921111.

Victoria Hung, Namrata D. Udeshi, Stephanie S. Lam, Ken H. Loh, Kurt J. Cox, Kayvon Pedram, Steven A. Carr, and Alice Y. Ting. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nature Protocols*, 11(3), March 2016. ISSN 1750-2799. doi: 10.1038/nprot.2016.018.

Motoshi Ichikawa, Akihide Yoshimi, Masahiro Nakagawa, Nahoko Nishimoto, Naoko Watanabe-Okochi, and Mineo Kurokawa. A role for RUNX1 in hematopoiesis and myeloid leukemia. *International Journal of Hematology*, 97(6):726–734, June 2013. ISSN 1865-3774. doi: 10.1007/s12185-013-1347-3.

Robert S. Illingworth and Adrian P. Bird. CpG islands–'a rough guide'. *FEBS letters*, 583(11):1713–1720, June 2009. ISSN 1873-3468. doi: 10.1016/j.febslet.2009.04.012.

Marcin Imielinski, Alice H. Berger, Peter S. Hammerman, Bryan Hernandez, Trevor J. Pugh, Eran Hodis, Jeonghee Cho, James Suh, Marzia Capelletti, Andrey Sivachenko, Carrie Sougnez, Daniel Auclair, Michael S. Lawrence, Petar Stojanov, Kristian Cibulskis, Kyusam Choi, Luc de Waal, Tanaz Sharifnia, Angela Brooks, Heidi Greulich, Shantanu Banerji, Thomas Zander, Danila Seidel, Frauke Leenders, Sascha Ansén, Corinna Ludwig, Walburga Engel-Riedel, Erich Stoelben, Jürgen Wolf, Chandra Goparju, Kristin Thompson, Wendy Winckler, David Kwiatkowski, Bruce E. Johnson, Pasi A. Jänne, Vincent A. Miller, William Pao, William D. Travis, Harvey I. Pass, Stacey B. Gabriel, Eric S. Lander, Roman K. Thomas, Levi A. Garraway, Gad Getz, and Matthew Meyerson. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell*, 150(6):1107–1120, September 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.08.029.

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, February 2001. ISSN 0028-0836, 1476-4687. doi: 10.1038/35057062.

Jean-Pierre J. Issa, Guillermo Garcia-Manero, Francis J. Giles, Rajan Mannari, Deborah Thomas, Stefan Faderl, Emel Bayar, John Lyons, Craig S. Rosenfeld, Jorge Cortes, and Hagop M. Kantarjian. Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2-deoxycytidine (decitabine) in hematopoietic malignancies. *Blood*, 103(5):1635–1640, March 2004. ISSN 0006-4971, 1528-0020. doi: 10.1182/blood-2003-03-0687.

Kyoko Ito, Joun Lee, Stephanie Chrysanthou, Yilin Zhao, Katherine Josephs, Hiroyo Sato, Julie Teruya-Feldstein, Deyou Zheng, Meelad M. Dawlaty, and Keisuke Ito. Non-catalytic Roles of Tet2 Are Essential to Regulate Hematopoietic Stem and Progenitor Cell Homeostasis. *Cell Reports*, 28(10):2480–2490.e4, September 2019. ISSN 2211-1247. doi: 10.1016/j.celrep.2019.07.094.

Shinsuke Ito, Ana C. D'Alessio, Olena V. Taranova, Kwonho Hong, Lawrence C. Sowers, and Yi Zhang. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal

and inner cell mass specification. *Nature*, 466(7310):1129–1133, August 2010. ISSN 1476-4687. doi: 10.1038/nature09303.

Shinsuke Ito, Li Shen, Qing Dai, Susan C Wu, Leonard B Collins, James A Swenberg, Chuan He, and Yi Zhang. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, 333(6047):1300–1303, 2011. ISSN 0036-8075. doi: 10.1126/science.1210597.

Antoine Ittel, Eric Jeandidier, Catherine Helias, Nathalie Perrusson, Catherine Humbrecht, Bruno Lioure, Isabelle Mazurier, Caroline Mayeur-Rousse, Amandine Lavaux, Sylvie Thiebault, Felix Lerintiu, Carine Gervais, and Laurent Mauvieux. First description of the t(10;11)(q22;q23)/MLL-TET1 translocation in a T-cell lymphoblastic lymphoma, with subsequent lineage switch to acute myelomonocytic myeloid leukemia. *Haematologica*, 98(12):e166–e168, December 2013. ISSN 0390-6078. doi: 10.3324/haematol.2013.096750.

Mario Iurlaro, Gabriella Ficz, David Oxley, Eun-Ang Raiber, Martin Bachman, Michael J Booth, Simon Andrews, Shankar Balasubramanian, and Wolf Reik. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, 14(10):R119, 2013. ISSN 1465-6906. doi: 10.1186/gb-2013-14-10-r119.

Lakshminarayan M. Iyer, Mamta Tahiliani, Anjana Rao, and L. Aravind. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, 8(11):1698–1710, June 2009. ISSN 1538-4101, 1551-4005. doi: 10.4161/cc.8.11.8580.

Franco Izzo, Stanley C. Lee, Asaf Poran, Ronan Chaligne, Federico Gaiti, Baptiste Gross, Rekha R. Murali, Sunil D. Deochand, Chelston Ang, Philippa Wyndham Jones, Anna S. Nam, Kyu-Tae Kim, Steven Kothen-Hill, Rafael C. Schulman, Michelle Ki, Priscillia Lhoumaud, Jane A. Skok, Aaron D. Viny, Ross L. Levine, Ephraim Kenigsberg, Omar Abdel-Wahab, and Dan A. Landau. DNA methylation disruption reshapes the

hematopoietic differentiation landscape. *Nature Genetics*, 52(4):378–387, April 2020. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-020-0595-4.

Dhawal Jain, Sandro Baldi, Angelika Zabel, Tobias Straub, and Peter B. Becker. Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments. *Nucleic Acids Research*, 43(14):6959–6968, August 2015. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkv637.

Anna M. Jankowska, Hadrian Szpurka, Ramon V. Tiu, Hideki Makishima, Manuel Afable, Jungwon Huh, Christine L. O'Keefe, Rebecca Ganetzky, Michael A. McDevitt, and Jaroslaw P. Maciejewski. Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood*, 113(25):6403–6410, June 2009. ISSN 0006-4971. doi: 10.1182/blood-2009-02-205690.

A. Janulaitis, S. Klimašauskas, M. Petrušyte, and V. Butkus. Cytosine modification in DNA by BcnI methylase yields N 4-methylcytosine. *FEBS Letters*, 161(1):131–134, 1983. ISSN 1873-3468. doi: 10.1016/0014-5793(83)80745-5.

Mira Jeong, Deqiang Sun, Min Luo, Yun Huang, Grant A Challen, Benjamin Rodriguez, Xiaotian Zhang, Lukas Chavez, Hui Wang, Rebecca Hannah, Sang-Bae Kim, Liubin Yang, Myunggon Ko, Rui Chen, Berthold Göttgens, Ju-Seog Lee, Preethi Gunaratne, Lucy A Godley, Gretchen J Darlington, Anjana Rao, Wei Li, and Margaret A Goodell. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nature Genetics*, 46(1):17–23, January 2014. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2836.

Lin Jia, Yichen Wang, Cong Wang, Zhonghua Du, Shilin Zhang, Xue Wen, Lei Zhou, Hui Li, Huiling Chen, Dan Li, Songling Zhang, Wei Li, Wei Xu, Andrew R. Hoffman, Jiuwei Cui, and Ji-Fan Hu. Oplr16 serves as a novel chromatin factor to control stem cell fate by modulating pluripotency-specific chromosomal looping and TET2-mediated DNA demethylation. *Nucleic Acids Research*, 48(7):3935–3948, April 2020. ISSN 1362-4962. doi: 10.1093/nar/gkaa097.

Shuai Jiang, Wei Yan, Shizhen Emily Wang, and David Baltimore. Dual mechanisms of posttranscriptional regulation of Tet2 by Let-7 microRNA in macrophages. *Proceedings of the National Academy of Sciences of the United States of America*, 116(25):12416–12421, 2019. ISSN 1091-6490. doi: 10.1073/pnas.1811040116.

Takuro Kameda, Kotaro Shide, Takumi Yamaji, Ayako Kamiunten, Masaaki Sekine, Yasuhiro Taniguchi, Tomonori Hidaka, Yoko Kubuki, Haruko Shimoda, Kousuke Marutsuka, Goro Sashida, Kazumasa Aoyama, Makoto Yoshimitsu, Taku Harada, Hiroo Abe, Tadashi Miike, Hisayoshi Iwakiri, Yoshihiro Tahara, Mitsue Sueta, Shojiro Yamamoto, Satoru Hasuike, Kenji Nagata, Atsushi Iwama, Akira Kitanaka, and Kazuya Shimoda. Loss of TET2 has dual roles in murine myeloproliferative neoplasms: disease sustainer and disease accelerator. *Blood*, 125(2):304–315, January 2015. ISSN 0006-4971. doi: 10.1182/blood-2014-04-555508.

Zhengyan Kan, Bijay S. Jaiswal, Jeremy Stinson, Vasantharajan Janakiraman, Deepali Bhatt, Howard M. Stern, Peng Yue, Peter M. Haverty, Richard Bourgon, Jianbiao Zheng, Martin Moorhead, Subhra Chaudhuri, Lynn P. Tomsho, Brock A. Peters, Kanan Pujara, Shaun Cordes, David P. Davis, Victoria E. H. Carlton, Wenlin Yuan, Li Li, Weiru Wang, Charles Eigenbrot, Joshua S. Kaminker, David A. Eberhard, Paul Waring, Stephan C. Schuster, Zora Modrusan, Zemin Zhang, David Stokoe, Frederic J. de Sauvage, Malek Faham, and Somasekar Seshagiri. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature*, 466(7308):869–873, August 2010. ISSN 1476-4687. doi: 10.1038/nature09208.

Masahiro Kaneda, Masaki Okano, Kenichiro Hata, Takashi Sado, Naomi Tsujimoto, En Li, and Hiroyuki Sasaki. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, 429(6994):900–903, June 2004. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature02633.

Rosa Karlić, Ho-Ryun Chung, Julia Lasserre, Kristian Vlahoviček, and Martin Vingron. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences*, 107(7):2926–2931, February 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0909344107.

Matthew W. Kellinger, Chun-Xiao Song, Jenny Chong, Xing-Yu Lu, Chuan He, and Dong Wang. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature Structural & Molecular Biology*, 19(8):831–833, August 2012. ISSN 1545-9985. doi: 10.1038/nsmb.2346.

Dae In Kim, Birendra Kc, Wenhong Zhu, Khatereh Motamedchaboki, Valérie Doye, and Kyle J. Roux. Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences*, 111(24):E2453–E2461, June 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1406459111.

Richard H. Kim, David Wang, Michael Tsang, Jennifer Martin, Carla Huff, Mark P. de Caestecker, W. Tony Parks, Xianwang Meng, Robert J. Lechleider, Tongwen Wang, and Anita B. Roberts. A novel Smad nuclear interacting protein, SNIP1, suppresses p300-dependent TGF- signal transduction. *Genes & Development*, 14(13):1605–1616, January 2000. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.14.13.1605.

Hiroshi Kimura. Histone modifications for human epigenome analysis. *Journal of Human Genetics*, 58(7):439–445, July 2013. ISSN 1434-5161, 1435-232X. doi: 10.1038/jhg.2013.66.

Ashwin Kishtagari, Babal K. Jha, and Jaroslaw P. Maciejewski. TET2 mutations and clonal dynamics. *Oncotarget*, 10(21):2010–2011, March 2019. ISSN 1949-2553. doi: 10.18632/oncotarget.26779.

M. Ko, H. S. Bandukwala, J. An, E. D. Lamperti, E. C. Thompson, R. Hastie, A. Tsangaratou, K. Rajewsky, S. B. Koralov, and A. Rao. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proceedings of the National Academy of Sciences*, 108(35):14566–14571, August 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1112317108.

Myunggon Ko, Yun Huang, Anna M. Jankowska, Utz J. Pape, Mamta Tahiliani, Hozefa S. Bandukwala, Jungeun An, Edward D. Lamperti, Kian Peng Koh, Rebecca Ganetzky, X. Shirley Liu, L. Aravind, Suneet Agarwal, Jaroslaw P. Maciejewski, and Anjana

Rao. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, 468(7325):839–843, December 2010. ISSN 1476-4687. doi: 10.1038/nature09586.

Myunggon Ko, Jungeun An, Hozefa S Bandukwala, Lukas Chavez, Tarmo Äijö, William A Pastor, Matthew F Segal, Huiming Li, Kian Koh, Harri Lähdesmäki, Patrick G Hogan, L Aravind, and Anjana Rao. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, 497 (7447):nature12052, 2013. ISSN 1476-4687. doi: 10.1038/nature12052.

Frederic Koch, Romain Fenouil, Marta Gut, Pierre Cauchy, Thomas K. Albert, Joaquin Zacarias-Cabeza, Salvatore Spicuglia, Albane Lamy de la Chapelle, Martin Heidemann, Corinna Hintermair, Dirk Eick, Ivo Gut, Pierre Ferrier, and Jean-Christophe Andrau. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural & Molecular Biology*, 18(8):956–963, August 2011. ISSN 1545-9985. doi: 10.1038/nsmb.2085.

Kian Peng Koh, Akiko Yabuuchi, Sridhar Rao, Yun Huang, Kerrianne Cunniff, Julie Nardone, Asta Laiho, Mamta Tahiliani, Cesar A. Sommer, Gustavo Mostoslavsky, Riitta Lahesmaa, Stuart H. Orkin, Scott J. Rodig, George Q. Daley, and Anjana Rao. Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells. *Cell Stem Cell*, 8(2):200–213, February 2011. ISSN 1934-5909, 1875-9777. doi: 10.1016/j.stem.2011.01.008.

Raivo Kolde. Pheatmap: pretty heatmaps. *R package version*, 1(2), 2012.

Lisa K. Kreppel, Melissa A. Blomberg, and Gerald W. Hart. Dynamic Glycosylation of Nuclear and Cytosolic Proteins CLONING AND CHARACTERIZATION OF A UNIQUE O-GlcNAc TRANSFERASE WITH MULTIPLE TETRATRICOPEPTIDE REPEATS. *Journal of Biological Chemistry*, 272(14):9308–9315, April 1997. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.272.14.9308.

S. Kriaucionis and N. Heintz. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*, 324(5929):929–930, May 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1169786.

Judith F. Kribelbauer, Oleg Laptenko, Siying Chen, Gabriella D. Martini, William A. Freed-Pastor, Carol Prives, Richard S. Mann, and Harmen J. Bussemaker. Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. *Cell Reports*, 19(11):2383–2395, June 2017. ISSN 2211-1247. doi: 10.1016/j.celrep.2017.05.069.

Kandasamy Krishnaraju, Barbara Hoffman, and Dan A. Liebermann. Early growth response gene 1 stimulates development of hematopoietic progenitor cells along the macrophage lineage at the expense of the granulocyte and erythroid lineages. *Blood*, 97(5):1298–1305, March 2001. ISSN 0006-4971. doi: 10.1182/blood.V97.5.1298.

Ivan Krivega and Ann Dean. Enhancer and promoter interactions-long distance calls. *Current Opinion in Genetics & Development*, 22(2):79–85, April 2012. ISSN 1879-0380. doi: 10.1016/j.gde.2011.11.001.

Hanna Krysinska, Maarten Hoogenkamp, Richard Ingram, Nicola Wilson, Hiromi Tagoh, Peter Laslo, Harinder Singh, and Constanze Bonifer. A Two-Step, PU.1-Dependent Mechanism for Developmentally Regulated Chromatin Remodeling and Transcription of the c-fms Gene. *Molecular and Cellular Biology*, 27(3):878–887, February 2007. ISSN 0270-7306. doi: 10.1128/MCB.01915-06.

Yotaro Kudo, Keisuke Tateishi, Keisuke Yamamoto, Shinzo Yamamoto, Yoshinari Asaoka, Hideaki Ijichi, Genta Nagae, Haruhiko Yoshida, Hiroyuki Aburatani, and Kazuhiko Koike. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Science*, 103(4):670–676, 2012. ISSN 1349-7006. doi: 10.1111/j.1349-7006.2012.02213.x.

Maxim V. Kuleshov, Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L. Jenkins, Kathleen M. Jagodnik,

Alexander Lachmann, Michael G. McDermott, Caroline D. Monteiro, Gregory W. Gundersen, and Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1):W90–97, 2016. ISSN 1362-4962. doi: 10.1093/nar/gkw377.

Anirban Kundu, Sandeep Shelar, Arindam P. Ghosh, Mary Ballestas, Richard Kirkman, Hyeyoung Nam, Garrett J. Brinkley, Suman Karki, James A. Mobley, Sejong Bae, Sooryanarayana Varambally, and Sunil Sudarshan. 14-3-3 proteins protect AMPK-phosphorylated ten-eleven translocation-2 (TET2) from PP2A-mediated dephosphorylation. *Journal of Biological Chemistry*, 295(6):1754–1766, February 2020. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.RA119.011089.

Alexander Lachmann, Huilei Xu, Jayanth Krishnan, Seth I. Berger, Amin R. Mazloom, and Avi Ma'ayan. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics (Oxford, England)*, 26(19):2438–2444, October 2010. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq466.

E Lainey, A Wolfromm, N Marie, D Enot, M Scoazec, C Bouteloup, C Leroy, J-B Micol, S De Botton, L Galluzzi, P Fenaux, and G Kroemer. Azacytidine and erlotinib exert synergistic effects against acute myeloid leukemia. *Oncogene*, 32(37):4331–4342, September 2013. ISSN 0950-9232, 1476-5594. doi: 10.1038/onc.2012.469.

Eric S. Lander, Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, and . et al Doyle. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, February 2001. ISSN 1476-4687. doi: 10.1038/35057062.

Jane M. Landolin, David S. Johnson, Nathan D. Trinklein, Shelly F. Aldred, Catherine Medina, Hennady Shulha, Zhiping Weng, and Richard M. Myers. Sequence features that drive human promoter function and tissue specificity. *Genome Research*, 20(7): 890–898, July 2010. ISSN 1549-5469. doi: 10.1101/gr.100370.109.

M. D. Lane, D. L. Young, and F. Lynen. THE ENZYMATIC SYNTHESIS OF HOLOTRANSCARBOXYLASE FROM APOTRANSCARBOXYLASE AND (+)-BIOTIN. I. PURIFICATION OF THE APOENZYME AND SYNTHETASE; CHARACTERISTICS OF THE REACTION. *The Journal of Biological Chemistry*, 239: 2858–2864, September 1964. ISSN 0021-9258.

Saskia M C Langemeijer, Roland P Kuiper, Marieke Berends, Ruth Knops, Mariam G Aslanyan, Marion Massop, Ellen Stevens-Linders, Patricia van Hoogen, Ad Geurts van Kessel, Reinier A P Raymakers, Eveline J Kamping, Gregor E Verhoef, Estelle Verburgh, Anne Hagemeijer, Peter Vandenberghe, Theo de Witte, Bert A van der Reijden, and Joop H Jansen. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature Genetics*, 41(7):838–842, July 2009. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.391.

Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, March 2012. ISSN 1548-7105. doi: 10.1038/nmeth.1923.

David Lara-Astiaso, Assaf Weiner, Erika Lorenzo-Vivas, Irina Zaretsky, Diego Adhemar Jaitin, Eyal David, Hadas Keren-Shaul, Alexander Mildner, Deborah Winter, Steffen Jung, Nir Friedman, and Ido Amit. Chromatin state dynamics during blood formation. *Science*, 345(6199):943–949, August 2014. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1256271.

Peter Laslo, Chauncey J. Spooner, Aryeh Warmflash, David W. Lancki, Hyun-Jun Lee, Roger Sciammas, Benjamin N. Gantner, Aaron R. Dinner, and Harinder Singh. Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates. *Cell*, 126(4):755–766, August 2006. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2006.06.052.

Louise Laurent, Eleanor Wong, Guoliang Li, Tien Huynh, Aristotelis Tsirigos, Chin Thing Ong, Hwee Meng Low, Ken Wing Kin Sung, Isidore Rigoutsos, Jeanne Loring, and Chia-Lin Wei. Dynamic changes in the human methylome during differentiation.

Genome Research, 20(3):320–331, March 2010. ISSN 1549-5469. doi: 10.1101/gr.101907.109.

Jens-Peter Holst Lauritsen, Sridevi Kurella, Sang-Yun Lee, Juliette M. Lefebvre, Michele Rhodes, José Alberola-Ila, and David L. Wiest. Egr2 is required for Bcl-2 induction during positive selection. *Journal of immunology (Baltimore, Md. : 1950)*, 181(11): 7778–7785, December 2008. ISSN 0022-1767.

Roland Lauster, Thomas A. Trautner, and Mario Noyer-Weidner. Cytosine-specific type II DNA methyltransferases: A conserved enzyme core with variable target-recognizing domains. *Journal of Molecular Biology*, 206(2):305–312, March 1989. ISSN 0022-2836. doi: 10.1016/0022-2836(89)90480-4.

Michael S. Lawrence, Petar Stojanov, Craig H. Mermel, James T. Robinson, Levi A. Garraway, Todd R. Golub, Matthew Meyerson, Stacey B. Gabriel, Eric S. Lander, and Gad Getz. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 505(7484):495–501, January 2014. ISSN 1476-4687. doi: 10.1038/nature12912.

Ryan M. Layer, Brent S. Pedersen, Tonya DiSera, Gabor T. Marth, Jason Gertz, and Aaron R. Quinlan. GIGGLE: a search engine for large-scale integrated genome analysis. *Nature Methods*, 15(2):123–126, 2018. ISSN 1548-7105. doi: 10.1038/nmeth.4556.

Katherine M. Lelli, Matthew Slattery, and Richard S. Mann. Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. *Annual Review of Genetics*, 46(1): 43–68, 2012. doi: 10.1146/annurev-genet-110711-155437.

François Lemonnier, Lucile Couronné, Marie Parrens, Jean-Philippe Jaïs, Marion Travert, Laurence Lamant, Olivier Tournillac, Therese Rousset, Bettina Fabiani, Rob A. Cairns, Tak Mak, Christian Bastard, Olivier A. Bernard, Laurence de Leval, and Philippe Gaulard. Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH-like features and adverse clinical parameters. *Blood*, 120(7):1466–1469, August 2012. ISSN 0006-4971. doi: 10.1182/blood-2012-02-408542.

Antonio Lentini, Cathrine Lagerwall, Svante Vikingsson, Heidi K. Mjoseng, Karolos Douvlataniotis, Hartmut Vogt, Henrik Green, Richard R. Meehan, Mikael Benson, and Colm E. Nestor. A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nature Methods*, 15(7):499–504, July 2018. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-018-0038-7.

Mads Lerdrup, Jens Vilstrup Johansen, Shuchi Agrawal-Singh, and Klaus Hansen. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nature Structural & Molecular Biology*, 23(4):349–357, April 2016. ISSN 1545-9985. doi: 10.1038/nsmb.3180.

En Li, Timothy H. Bestor, and Rudolf Jaenisch. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6):915–926, June 1992. ISSN 0092-8674. doi: 10.1016/0092-8674(92)90611-F.

Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–2079, August 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp352.

Meng Amy Li, Daniel J. Turner, Zemin Ning, Kosuke Yusa, Qi Liang, Sabine Eckert, Lena Rad, Tomas W. Fitzgerald, Nancy L. Craig, and Allan Bradley. Mobilization of giant piggyBac transposons in the mouse genome. *Nucleic Acids Research*, 39(22):e148, December 2011a. ISSN 1362-4962. doi: 10.1093/nar/gkr764.

Shaojuan Li, Changxin Wan, Rongbin Zheng, Jingyu Fan, Xin Dong, Clifford A Meyer, and X Shirley Liu. Cistrome-GO: a web server for functional enrichment analysis of transcription factor ChIP-seq peaks. *Nucleic Acids Research*, 47(W1):W206–W211, July 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz332.

Yuanyuan Li, Hui Zheng, Qiujun Wang, Chen Zhou, Lei Wei, Xuehui Liu, Wenhao Zhang, Yu Zhang, Zhenhai Du, Xiaowo Wang, and Wei Xie. Genome-wide analyses reveal a

role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biology*, 19(1):18, February 2018. ISSN 1474-760X. doi: 10.1186/s13059-018-1390-8.

Zhe Li, Xiaoqiang Cai, Chen-Leng Cai, Jiapeng Wang, Wenyong Zhang, Bruce E. Petersen, Feng-Chun Yang, and Mingjiang Xu. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*, 118(17):4509–4518, October 2011b. ISSN 0006-4971, 1528-0020. doi: 10.1182/blood-2010-12-325241.

Christine Guo Lian, Yufei Xu, Craig Ceol, Feizhen Wu, Allison Larson, Karen Dresser, Wenqi Xu, Li Tan, Yeguang Hu, Qian Zhan, Chung-wei Lee, Di Hu, Bill Q. Lian, Sonja Kleffel, Yijun Yang, James Neiswender, Abraham J. Khorasani, Rui Fang, Cecilia Lezcano, Lyn M. Duncan, Richard A. Scolyer, John F. Thompson, Hojabr Kakavand, Yariv Houvras, Leonard I. Zon, Martin C. Mihm, Ursula B. Kaiser, Tobias Schatton, Bruce A. Woda, George F. Murphy, and Yujiang G. Shi. Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell*, 150(6): 1135–1146, September 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.07.033.

Yang Liao, Gordon K. Smyth, and Wei Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, 30(7):923–930, April 2014. ISSN 1367-4811. doi: 10.1093/bioinformatics/btt656.

Charles Y. Lin, Jakob Lovén, Peter B. Rahl, Ronald M. Paranal, Christopher B. Burge, James E. Bradner, Tong Ihn Lee, and Richard A. Young. Transcriptional amplification in tumor cells with elevated c-myc. 151(1):56–67. ISSN 1097-4172. doi: 10.1016/j.cell.2012.08.026.

Anders M. Lindroth, Yoon Jung Park, Chelsea M. McLean, Gregoriy A. Dokshin, Jenna M. Persson, Herry Herman, Diego Pasini, Xavier Miró, Mary E. Donohoe, Jeannie T. Lee, Kristian Helin, and Paul D. Soloway. Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. *PLoS genetics*, 4(8):e1000145, August 2008. ISSN 1553-7404. doi: 10.1371/journal.pgen.1000145.

Chan-Wang Lio, Jiayuan Zhang, Edahí González-Avalos, Patrick G Hogan, Xing Chang, and Anjana Rao. Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. *eLife*, 5:e18290, November 2016. ISSN 2050-084X. doi: 10.7554/eLife.18290.

Chan-Wang J. Lio, Vipul Shukla, Daniela Samaniego-Castruita, Edahi González-Avalos, Abhijit Chakraborty, Xiaojing Yue, David G. Schatz, Ferhat Ay, and Anjana Rao. TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer. *Science Immunology*, 4(34), 2019a. ISSN 2470-9468. doi: 10.1126/sciimmunol.aau7523.

Chan-Wang J. Lio, Hiroshi Yuita, and Anjana Rao. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood*, 134(18): 1487–1497, October 2019b. ISSN 0006-4971. doi: 10.1182/blood.2019791475.

Ryan Lister, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, Leonard Lee, Zhen Ye, Que-Minh Ngo, Lee Edsall, Jessica Antosiewicz-Bourget, Ron Stewart, Victor Ruotti, A. Harvey Millar, James A. Thomson, Bing Ren, and Joseph R. Ecker. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–322, November 2009. ISSN 1476-4687. doi: 10.1038/nature08514.

Chungang Liu, Limei Liu, Xuejiao Chen, Junjie Shen, Juanjuan Shan, Yanmin Xu, Zhi Yang, Lin Wu, Feng Xia, Ping Bie, Youhong Cui, Xiu-wu Bian, and Cheng Qian. Decrease of 5-hydroxymethylcytosine is associated with progression of hepatocellular carcinoma through downregulation of TET1. *PloS One*, 8(5):e62828, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0062828.

Hang Liu, Lei Peng, Joan So, Ka Hing Tsang, Chi Ho Chong, Priscilla Hoi Shan Mak, Kui Ming Chan, and Siu Yuen Chan. TSPYL2 Regulates the Expression of EZH2 Target Genes in Neurons. *Molecular Neurobiology*, 56(4):2640–2652, April 2019a. ISSN 1559-1182. doi: 10.1007/s12035-018-1238-y.

Jianzhao Liu, Yuanxiang Zhu, Guan-Zheng Luo, Xinxia Wang, Yanan Yue, Xiaona Wang, Xin Zong, Kai Chen, Hang Yin, Ye Fu, Dali Han, Yizhen Wang, Dahua Chen, and Chuan He. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nature Communications*, 7(1):1–7, October 2016. ISSN 2041-1723. doi: 10.1038/ncomms13052.

Jun Liu, Junshik Hong, Heejoo Han, Jihyun Park, Dongchan Kim, Hyejoo Park, Myunggon Ko, Youngil Koh, Dong-Yeop Shin, and Sung-Soo Yoon. Decreased vitamin C uptake mediated by SLC2A3 promotes leukaemia progression and impedes TET2 restoration. *British Journal of Cancer*, pages 1–8, March 2020. ISSN 1532-1827. doi: 10.1038/s41416-020-0788-8.

Yibin Liu, Paulina Siejka-Zielińska, Gergana Velikova, Ying Bi, Fang Yuan, Marketa Tomkova, Chunsen Bai, Lei Chen, Benjamin Schuster-Böckler, and Chun-Xiao Song. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature Biotechnology*, 37(4):424–429, April 2019b. ISSN 1546-1696. doi: 10.1038/s41587-019-0041-2.

Leslie F. Lock, Nobuo Takagi, and Gail R. Martin. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell*, 48(1):39–46, January 1987. ISSN 0092-8674. doi: 10.1016/0092-8674(87)90353-9.

Christoph Loenarz and Christopher J. Schofield. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nature Chemical Biology*, 4(3):152–156, March 2008. ISSN 1552-4469. doi: 10.1038/nchembio0308-152.

Catherine Lofton-Day, Fabian Model, Theo DeVos, Reimo Tetzner, Juergen Distler, Matthias Schuster, Xiaoling Song, Ralf Lesche, Volker Liebenberg, Matthias Ebert, Bela Molnar, Robert Grützmann, Christian Pilarsky, and Andrew Sledziewski. DNA Methylation Biomarkers for Blood-Based Colorectal Cancer Screening. *Clinical Chemistry*, 54(2):414–423, February 2008. ISSN 0009-9147. doi: 10.1373/clinchem.2007.095992.

Hannah K. Long, Neil P. Blackledge, and Robert J. Klose. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society Transactions*, 41(Pt 3):727–740, June 2013. ISSN 0300-5127. doi: 10.1042/BST20130028.

R. B. Lorsbach, J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, and J. R. Downing. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, 17(3):637–641, March 2003. ISSN 0887-6924. doi: 10.1038/sj.leu.2402834.

Julie-Aurore Losman and William G. Kaelin. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes & Development*, 27(8):836–852, April 2013. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.217406.113.

Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, December 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8.

Jakob Lovén, David A. Orlando, Alla A. Sigova, Charles Y. Lin, Peter B. Rahl, Christopher B. Burge, David L. Levens, Tong Ihn Lee, and Richard A. Young. Revisiting global gene expression analysis. 151(3):476–482. ISSN 00928674. doi: 10.1016/j.cell.2012.10.012.

Falong Lu, Yuting Liu, Lan Jiang, Shinpei Yamaguchi, and Yi Zhang. Role of Tet proteins in enhancer activity and telomere elongation. *Genes & Development*, 28(19): 2103–2119, October 2014. ISSN 0890-9369. doi: 10.1101/gad.248005.114.

Chongyuan Luo, Petra Hajkova, and Joseph R. Ecker. Dynamic DNA methylation: In the right place at the right time. *Science*, 361(6409), September 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aat6806.

Magnus D. Lynch, Andrew J. H. Smith, Marco De Gobbi, Maria Flenley, Jim R. Hughes, Douglas Vernimmen, Helena Ayyub, Jacqueline A. Sharpe, Jacqueline A. Sloane-Stanley, Linda Sutherland, Stephen Meek, Tom Burdon, Richard J. Gibbons, David Garrick, and Douglas R. Higgs. An interspecies analysis reveals a key role for unmethylated

CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO journal*, 31(2):317–329, January 2012. ISSN 1460-2075. doi: 10.1038/emboj.2011.399.

Mary F. Lyon. Gene Action in the X -chromosome of the Mouse (Mus musculus L.). *Nature*, 190(4773), April 1961. ISSN 1476-4687. doi: 10.1038/190372a0.

Isaac F. López-Moyado, Ageliki Tsagaratou, Hiroshi Yuita, Hyungseok Seo, Benjamin Delatte, Sven Heinz, Christopher Benner, and Anjana Rao. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. *Proceedings of the National Academy of Sciences*, 116(34):16933–16942, August 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1903059116.

Philip Machanick and Timothy L. Bailey. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, 27(12):1696–1697, June 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr189.

Atanu Maiti and Alexander C. Drohat. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *The Journal of Biological Chemistry*, 286(41):35334–35338, October 2011. ISSN 1083-351X. doi: 10.1074/jbc.C111.284620.

Wolfgang Mayer, Alain Niveleau, Jörn Walter, Reinald Fundele, and Thomas Haaf. Demethylation of the zygotic paternal genome. *Nature*, 403(6769):501–502, February 2000. ISSN 1476-4687. doi: 10.1038/35000656.

Michael T. McCabe, Heidi M. Ott, Gopinath Ganji, Susan Korenchuk, Christine Thompson, Glenn S. Van Aller, Yan Liu, Alan P. Graves, Anthony Della Pietra Iii, Elsie Diaz, Louis V. LaFrance, Mark Mellinger, Celine Duquenne, Xinrong Tian, Ryan G. Kruger, Charles F. McHugh, Martin Brandt, William H. Miller, Dashyant Dhanak, Sharad K. Verma, Peter J. Tummino, and Caretha L. Creasy. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature*, 492(7427): 108–112, December 2012. ISSN 1476-4687. doi: 10.1038/nature11606.

Katy McLaughlin, Ilya M. Flyamer, John P. Thomson, Heidi K. Mjoseng, Ruchi Shukla, Iain Williamson, Graeme R. Grimes, Robert S. Illingworth, Ian R. Adams, Sari Pennings, Richard R. Meehan, and Wendy A. Bickmore. DNA Methylation Directs Polycomb-Dependent 3D Genome Re-organization in Naive Pluripotency. *Cell Reports*, 29(7):1974–1985.e6, November 2019. ISSN 22111247. doi: 10.1016/j.celrep.2019.10.031.

Alexander Meissner, Andreas Gnirke, George W. Bell, Bernard Ramsahoye, Eric S. Lander, and Rudolf Jaenisch. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, 33(18):5868–5877, 2005. ISSN 1362-4962. doi: 10.1093/nar/gki901.

Marian Mellén, Pinar Ayata, Scott Dewell, Skirmantas Kriaucionis, and Nathaniel Heintz. MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell*, 151(7):1417–1430, December 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.11.022.

Eric M. Mendenhall, Richard P. Koche, Thanh Truong, Vicky W. Zhou, Biju Issac, Andrew S. Chi, Manching Ku, and Bradley E. Bernstein. GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS genetics*, 6(12):e1001244, December 2010. ISSN 1553-7404. doi: 10.1371/journal.pgen.1001244.

S. Meyers, J. R. Downing, and S. W. Hiebert. Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the runt homology domain is required for DNA binding and protein-protein interactions. *Molecular and Cellular Biology*, 13(10):6336–6345, October 1993. ISSN 0270-7306. doi: 10.1128/mcb.13.10.6336.

Tarjei S. Mikkelsen, Manching Ku, David B. Jaffe, Biju Issac, Erez Lieberman, Georgia Giannoukos, Pablo Alvarez, William Brockman, Tae-Kyung Kim, Richard P. Koche, William Lee, Eric Mendenhall, Aisling O'Donovan, Aviva Presser, Carsten Russ, Xiaohui Xie, Alexander Meissner, Marius Wernig, Rudolf Jaenisch, Chad Nusbaum, Eric S. Lander, and Bradley E. Bernstein. Genome-wide maps of chromatin state

in pluripotent and lineage-committed cells. *Nature*, 448(7153), August 2007. ISSN 1476-4687. doi: 10.1038/nature06008.

Sara Montagner, Cristina Leoni, Stefan Emming, Giulia Della Chiara, Chiara Balestrieri, Iros Barozzi, Viviana Piccolo, Susan Togher, Myunggon Ko, Anjana Rao, Gioacchino Natoli, and Silvia Monticelli. TET2 Regulates Mast Cell Differentiation and Proliferation through Catalytic and Non-catalytic Activities. *Cell Reports*, 15(7):1566–1579, 2016. ISSN 2211-1247. doi: 10.1016/j.celrep.2016.04.044.

Kelly Moran-Crusio, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E. Figueroa, Aparna Vasanthakumar, Jay Patel, Xinyang Zhao, Fabiana Perna, Suveg Pandey, Jozef Madzo, Chunxiao Song, Qing Dai, Chuan He, Sherif Ibrahim, Miloslav Beran, Jiri Zavadil, Stephen D. Nimer, Ari Melnick, Lucy A. Godley, Iannis Aifantis, and Ross L. Levine. Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell*, 20 (1):11–24, July 2011. ISSN 15356108. doi: 10.1016/j.ccr.2011.06.001.

Leslie Morse, Dongquan Chen, David Franklin, Yue Xiong, and Selina Chen-Kiang. Induction of Cell Cycle Arrest and B Cell Terminal Differentiation by CDK Inhibitor p18 INK4c and IL-6. *Immunity*, 6(1):47–56, January 1997. ISSN 10747613. doi: 10.1016/S1074-7613(00)80241-1.

Noushine Mossadegh-Keller, Sandrine Sarrazin, Prashanth K. Kandalla, Leon Espinosa, E. Richard Stanley, Stephen L. Nutt, Jordan Moore, and Michael H. Sieweke. M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. *Nature*, 497(7448): 239–243, May 2013. ISSN 1476-4687. doi: 10.1038/nature12026.

Tomoya Muto, Goro Sashida, Motohiko Oshima, George R. Wendt, Makiko Mochizuki-Kashio, Yasunobu Nagata, Masashi Sanada, Satoru Miyagi, Atsunori Saraya, Asuka Kamio, Genta Nagae, Chiaki Nakaseko, Koutaro Yokote, Kazuya Shimoda, Haruhiko Koseki, Yutaka Suzuki, Sumio Sugano, Hiroyuki Aburatani, Seishi Ogawa, and Atsushi Iwama. Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis

of myelodysplastic disorders. *The Journal of Experimental Medicine*, 210(12):2627–2639, November 2013. ISSN 0022-1007. doi: 10.1084/jem.20131144.

Donna M. Muzny, Matthew N. Bainbridge, Kyle Chang, Huyen H. Dinh, Jennifer A. Drummond, Gerald Fowler, Christie L. Kovar, Lora R. Lewis, Margaret B. Morgan, Irene F. Newsham, Jeffrey G. Reid, Jireh Santibanez, Eve Shinbrot, Lisa R. Trevino, Yuan-Qing Wu, Min Wang, Preethi Gunaratne, Lawrence A. Donehower, Chad J. Creighton, David A. Wheeler, Richard A. Gibbs, Michael S. Lawrence, Douglas Voet, Rui Jing, Kristian Cibulskis, Andrey Sivachenko, Petar Stojanov, Aaron McKenna, Eric S. Lander, Stacey Gabriel, Gad Getz, Li Ding, Robert S. Fulton, Daniel C. Koboldt, Todd Wylie, Jason Walker, David J. Dooling, Lucinda Fulton, Kim D. Delehaunty, Catrina C. Fronick, Ryan Demeter, Elaine R. Mardis, Richard K. Wilson, Andy Chu, Hye-Jung E. Chun, Andrew J. Mungall, Erin Pleasance, A. Gordon Robertson, Dominik Stoll, Miruna Balasundaram, Inanc Birol, Yaron S. N. Butterfield, Eric Chuah, Robin J. N. Coope, Noreen Dhalla, Ranabir Guin, Carrie Hirst, Martin Hirst, Robert A. Holt, Darlene Lee, Haiyan I. Li, Michael Mayo, Richard A. Moore, Jacqueline E. Schein, Jared R. Slobodan, Angela Tam, Nina Thiessen, Richard Varhol, Thomas Zeng, Yongjun Zhao, Steven J. M. Jones, Marco A. Marra, Adam J. Bass, Alex H. Ramos, Gordon Saksena, Andrew D. Cherniack, Stephen E. Schumacher, Barbara Tabak, Scott L. Carter, Nam H. Pho, Huy Nguyen, Robert C. Onofrio, Andrew Crenshaw, Kristin Ardlie, Rameen Beroukhim, Wendy Winckler, Gad Getz, Matthew Meyerson, Alexei Protopopov, Juinhua Zhang, Angela Hadjipanayis, Eunjung Lee, Ruibin Xi, Lixing Yang, Xiaojia Ren, Hailei Zhang, Narayanan Sathiamoorthy, Sachet Shukla, Peng-Chieh Chen, Psalm Haseley, Yonghong Xiao, Semin Lee, Jonathan Seidman, Lynda Chin, Peter J. Park, Raju Kucherlapati, J. Todd Auman, Katherine A. Hoadley, Ying Du, Matthew D. Wilkerson, Yan Shi, Christina Liquori, Shaowu Meng, Ling Li, Yidi J. Turman, Michael D. Topal, Donghui Tan, Scot Waring, Elizabeth Buda, Jesse Walsh, Corbin D. Jones, Piotr A. Mieczkowski, Darshan Singh, Junyuan Wu, Anisha Gulabani, Peter Dolina, Tom Bodenheimer, Alan P. Hoyle, Janae V. Simons, Matthew Soloway, Lisle E. Mose, Stuart R. Jefferys, Saianand Balu, Brian D. O'Connor, Jan F. Prins, Derek Y. Chiang, D. Neil Hayes, Charles M. Perou, Toshinori Hinoue,

Daniel J. Weisenberger, Dennis T. Maglinte, Fei Pan, Benjamin P. Berman, David J. Van Den Berg, Hui Shen, Timothy Triche Jr, Stephen B. Baylin, Peter W. Laird, Gad Getz, Michael Noble, Doug Voet, Gordon Saksena, Nils Gehlenborg, Daniel DiCara, Juinhua Zhang, Hailei Zhang, Chang-Jiun Wu, Spring Yingchun Liu, Sachet Shukla, Michael S. Lawrence, Lihua Zhou, Andrey Sivachenko, Pei Lin, Petar Stojanov, Rui Jing, Richard W. Park, Marc-Danie Nazaire, Jim Robinson, Helga Thorvaldsdottir, Jill Mesirov, Peter J. Park, Lynda Chin, Vesteinn Thorsson, Sheila M. Reynolds, Brady Bernard, Richard Kreisberg, Jake Lin, Lisa Iype, Ryan Bressler, Timo Erkkilä, Madhumati Gundapuneni, Yuexin Liu, Adam Norberg, Tom Robinson, Da Yang, Wei Zhang, Ilya Shmulevich, Jorma J. de Ronde, Nikolaus Schultz, Ethan Cerami, Giovanni Ciriello, Arthur P. Goldberg, Benjamin Gross, Anders Jacobsen, Jianjiong Gao, Bogumil Kaczkowski, Rileen Sinha, B. Arman Aksoy, Yevgeniy Antipin, Boris Reva, Ronglai Shen, Barry S. Taylor, Timothy A. Chan, Marc Ladanyi, Chris Sander, Rehan Akbani, Nianxiang Zhang, Bradley M. Broom, Tod Casasent, Anna Unruh, Chris Wakefield, Stanley R. Hamilton, R. Craig Cason, Keith A. Baggerly, John N. Weinstein, David Haussler, Christopher C. Benz, Joshua M. Stuart, Stephen C. Benz, J. Zachary Sanborn, Charles J. Vaske, Jingchun Zhu, Christopher Szeto, Gary K. Scott, Christina Yau, Sam Ng, Ted Goldstein, Kyle Ellrott, Eric Collisson, Aaron E. Cozen, Daniel Zerbino, Christopher Wilks, Brian Craft, Paul Spellman, Robert Penny, Troy Shelton, Martha Hatfield, Scott Morris, Peggy Yena, Candace Shelton, Mark Sherman, Joseph Paulauskis, Julie M. Gastier-Foster, Jay Bowen, Nilsa C. Ramirez, Aaron Black, Robert Pyatt, Lisa Wise, Peter White, Monica Bertagnolli, Jen Brown, Timothy A. Chan, Gerald C. Chu, Christine Czerwinski, Fred Denstman, Rajiv Dhir, Arnulf Dörner, Charles S. Fuchs, Jose G. Guillem, Mary Iacocca, Hartmut Juhl, Andrew Kaufman, Bernard Kohl III, Xuan Van Le, Maria C. Mariano, Elizabeth N. Medina, Michael Meyers, Garrett M. Nash, Phillip B. Paty, Nicholas Petrelli, Brenda Rabeno, William G. Richards, David Solit, Pat Swanson, Larissa Temple, Joel E. Tepper, Richard Thorp, Efsevia Vakiani, Martin R. Weiser, Joseph E. Willis, Gary Witkin, Zhaoshi Zeng, Michael J. Zinner, The Cancer Genome Atlas Network, Genome Sequencing Center Baylor College of Medicine, Genome Sequencing Center Broad Institute,

Genome Sequencing Center Washington University in St Louis, Genome Characterization Center BC Cancer Agency, Genome-Characterization Center Broad Institute, Genome-Characterization Center Brigham and Women's Hospital and Harvard Medical School, Chapel Hill Genome-Characterization Center University of North Carolina, Genome-Characterization Centers University of Southern California and Johns Hopkins University, Genome Data Analysis Center Broad Institute, Genome Data Analysis Center Institute for Systems Biology, Genome Data Analysis Center Memorial Sloan-Kettering Cancer Center, Genome Data Analysis Center University of Texas MD Anderson Cancer Center, University of California Genome Data Analysis Centers, Santa Cruz and the Buck Institute, Biospecimen Core Resource International Genomics Consortium, Nationwide Children's Hospital Biospecimen Core Resource, and Tissue source sites and disease working group. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407):330–337, July 2012. ISSN 1476-4687. doi: 10.1038/nature11252.

Martin Münzel, Daniel Globisch, Tobias Brückl, Mirko Wagner, Veronika Welzmiller, Stylianos Michalakis, Markus Müller, Martin Biel, and Thomas Carell. Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angewandte Chemie International Edition*, 49(31):5375–5377, 2010. ISSN 1521-3773. doi: 10.1002/anie.201002033.

Tadashi Nakagawa, Lei Lv, Makiko Nakagawa, Yanbao Yu, Chao Yu, Ana C. D'Alessio, Keiko Nakayama, Heng-Yu Fan, Xian Chen, and Yue Xiong. CRL4VprBP E3 Ligase Promotes Monoubiquitylation and Chromatin Binding of TET Dioxygenases. *Molecular Cell*, 57(2):247–260, January 2015. ISSN 10972765. doi: 10.1016/j.molcel.2014.12.002.

Hideaki Nakajima and Hiroyoshi Kunimoto. TET2 as an epigenetic master regulator for normal and malignant hematopoiesis. *Cancer Science*, 105(9):1093–1099, 2014. ISSN 1349-7006. doi: 10.1111/cas.12484.

Toshinobu Nakamura, Yoshikazu Arai, Hiroki Umehara, Masaaki Masuhara, Tohru Kimura, Hisaaki Taniguchi, Toshihiro Sekimoto, Masahito Ikawa, Yoshihiro Yoneda, Masaru Okabe, Satoshi Tanaka, Kunio Shiota, and Toru Nakano. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biology*, 9(1):64–71, January 2007. ISSN 1465-7392. doi: 10.1038/ncb1519.

Francesco Neri, Danny Incarnato, Anna Krepelova, Stefania Rapelli, Andrea Pagnani, Riccardo Zecchina, Caterina Parlato, and Salvatore Oliviero. Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biology*, 14(8):R91, 2013. ISSN 1465-6906. doi: 10.1186/gb-2013-14-8-r91.

M. Okano, S. Xie, and E. Li. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, 19(3):219–220, July 1998. ISSN 1061-4036. doi: 10.1038/890.

Masaki Okano, Daphne W Bell, Daniel A Haber, and En Li. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, 99(3):247–257, October 1999. ISSN 0092-8674. doi: 10.1016/S0092-8674(00)81656-6.

Naoki Okashita, Shunsuke Kuroki, Ryo Maeda, and Makoto Tachibana. TET2 catalyzes active DNA demethylation of the Sry promoter and enhances its expression. *Scientific Reports*, 9(1):13462, September 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-50058-7.

T. Okuda, J. van Deursen, S. W. Hiebert, G. Grosveld, and J. R. Downing. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, 84(2):321–330, January 1996. ISSN 0092-8674. doi: 10.1016/s0092-8674(00)80986-1.

Ryoichi Ono, Tomohiko Taki, Takeshi Taketani, Masafumi Taniwaki, Hajime Kobayashi, and Yasuhide Hayashi. LCX, leukemia-associated protein with a CXXC domain,

is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Research*, 62(14):4075–4080, July 2002. ISSN 0008-5472.

Tõnis Org, Dan Duan, Roberto Ferrari, Amelie Montel-Hagen, Ben Van Handel, Marc A. Kerényi, Rajkumar Sasidharan, Liudmilla Rubbi, Yuko Fujiwara, Matteo Pellegrini, Stuart H. Orkin, Siavash K. Kurdistani, and Hanna Ka Mikkola. Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. *The EMBO journal*, 34(6):759–777, March 2015. ISSN 1460-2075. doi: 10.15252/embj.201490542.

Elizabeth L. Ostrander, Ashley C. Kramer, Cates Mallaney, Hamza Celik, Won Kyun Koh, Jake Fairchild, Emily Haussler, Christine R. C. Zhang, and Grant A. Challen. Divergent Effects of Dnmt3a and Tet2 Mutations on Hematopoietic Progenitor Cell Fitness. *Stem Cell Reports*, 14(4):551–560, April 2020. ISSN 2213-6711. doi: 10.1016/j.stemcr.2020.02.011.

J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, and J. Walter. Active demethylation of the paternal genome in the mouse zygote. *Current biology: CB*, 10(8):475–478, April 2000. ISSN 0960-9822. doi: 10.1016/s0960-9822(00)00448-6.

Junji Otani, Toshiyuki Nankumo, Kyohei Arita, Susumu Inamoto, Mariko Ariyoshi, and Masahiro Shirakawa. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO reports*, 10 (11):1235–1241, November 2009. ISSN 1469-3178. doi: 10.1038/embor.2009.218.

B. R. Palmer and M. G. Marinus. The dam and dcm strains of Escherichia coli — a review. *Gene*, 143(1):1–12, May 1994. ISSN 0378-1119. doi: 10.1016/0378-1119(94)90597-5.

Feng Pan, Ophelia Weeks, Feng-Chun Yang, and Mingjiang Xu. The TET2 interactors and their links to hematological malignancies. *IUBMB life*, 67(6):438–445, June 2015. ISSN 1521-6543. doi: 10.1002/iub.1389.

Feng Pan, Thomas S. Wingo, Zhigang Zhao, Rui Gao, Hideki Makishima, Guangbo Qu, Li Lin, Miao Yu, Janice R. Ortega, Jiapeng Wang, Aziz Nazha, Li Chen, Bing Yao, Can Liu, Shi Chen, Ophelia Weeks, Hongyu Ni, Brittany Lynn Phillips, Suming Huang, Jianlong Wang, Chuan He, Guo-Min Li, Tomas Radivoyevitch, Iannis Aifantis, Jaroslaw P. Maciejewski, Feng-Chun Yang, Peng Jin, and Mingjiang Xu. Tet2 loss leads to hypermutagenicity in haematopoietic stem/progenitor cells. *Nature Communications*, 8(1), April 2017. ISSN 2041-1723. doi: 10.1038/ncomms15102.

B. Panning and R. Jaenisch. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes & Development*, 10(16):1991–2002, August 1996. ISSN 0890-9369. doi: 10.1101/gad.10.16.1991.

William A. Pastor, Utz J. Pape, Yun Huang, Hope R. Henderson, Ryan Lister, Myunggon Ko, Erin M. McLoughlin, Yevgeny Brudno, Sahasransu Mahapatra, Philipp Kapranov, Mamta Tahiliani, George Q. Daley, X. Shirley Liu, Joseph R. Ecker, Patrice M. Milos, Suneet Agarwal, and Anjana Rao. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 473(7347):394–397, May 2011. ISSN 1476-4687. doi: 10.1038/nature10102.

William A. Pastor, L. Aravind, and Anjana Rao. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Reviews Molecular Cell Biology*, 14(6):341–356, June 2013. ISSN 1471-0072, 1471-0080. doi: 10.1038/nrm3589.

Lina Peng, Yan Li, Yanping Xi, Wei Li, Jin Li, Ruitu Lv, Lei Zhang, Qingping Zou, Shihua Dong, Huaibing Luo, Feizhen Wu, and Wenqiang Yu. MBD3L2 promotes Tet2 enzymatic activity for mediating 5-methylcytosine oxidation. *Journal of Cell Science*, 129(5):1059–1071, March 2016. ISSN 0021-9533, 1477-9137. doi: 10.1242/jcs.179044.

Len A. Pennacchio, Wendy Bickmore, Ann Dean, Marcelo A. Nobrega, and Gill Bejerano. Enhancers: five essential questions. *Nature Reviews Genetics*, 14(4):288–295, April 2013. ISSN 1471-0064. doi: 10.1038/nrg3458.

Toni Pfaffeneder, Benjamin Hackner, Matthias Truß, Martin Münzel, Markus Müller, Christian A. Deiml, Christian Hagemeier, and Thomas Carell. The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angewandte Chemie International Edition*, 50(31):7008–7012, 2011. ISSN 1521-3773. doi: 10.1002/anie.201103899.

P. Pinto do O. Expression of the LIM-homeobox gene LH2 generates immortalized Steel factor-dependent multipotent hematopoietic precursors. *The EMBO Journal*, 17(19): 5744–5756, October 1998. ISSN 14602075. doi: 10.1093/emboj/17.19.5744.

Elodie Pronier, Carole Almire, Hayat Mokrani, Aparna Vasanthakumar, Audrey Simon, Barbara da Costa Reis Monte Mor, Aline Massé, Jean-Pierre Le Couédic, Frédéric Pendino, Bruno Carbonne, Jérôme Larghero, Jean-Luc Ravanat, Nicole Casadevall, Olivier A. Bernard, Nathalie Droin, Eric Solary, Lucy A. Godley, William Vainchenker, Isabelle Plo, and François Delhommeau. Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. *Blood*, 118(9):2551–2555, September 2011. ISSN 1528-0020. doi: 10.1182/blood-2010-12-324707.

J Pósfai, A S Bhagwat, G Pósfai, and R J Roberts. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Research*, 17(7):2421–2435, April 1989. ISSN 0305-1048.

Jane Yuxia Qin, Li Zhang, Kayla L. Clift, Imge Hulur, Andy Peng Xiang, Bing-Zhong Ren, and Bruce T. Lahn. Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLOS ONE*, 5(5):e10611, May 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0010611.

Aaron R. Quinlan and Ira M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6):841–842, March 2010. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq033.

Cyril Quivoron, Lucile Couronné, Véronique Della Valle, Cécile K. Lopez, Isabelle Plo, Orianne Wagner-Ballon, Marcio Do Cruzeiro, Francois Delhommeau, Bertrand

Arnulf, Marc-Henri Stern, Lucy Godley, Paule Opolon, Hervé Tilly, Eric Solary, Yannis Duffourd, Philippe Dessen, Hélène Merle-Beral, Florence Nguyen-Khac, Michaëla Fontenay, William Vainchenker, Christian Bastard, Thomas Mercher, and Olivier A. Bernard. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell*, 20 (1):25–38, July 2011. ISSN 15356108. doi: 10.1016/j.ccr.2011.06.003.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

Günter Raddatz, Paloma M. Guzzardo, Nelly Olova, Marcelo Rosado Fantappié, Markus Rampp, Matthias Schaefer, Wolf Reik, Gregory J. Hannon, and Frank Lyko. Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences*, 110(21):8627–8631, May 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1306723110.

Eun-Ang Raiber, Pierre Murat, Dimitri Y. Chirgadze, Dario Beraldi, Ben F. Luisi, and Shankar Balasubramanian. 5-Formylcytosine alters the structure of the DNA double helix. *Nature Structural & Molecular Biology*, 22(1):44–49, January 2015. ISSN 1545-9985. doi: 10.1038/nsmb.2936.

Raajit Rampal, Altuna Alkalin, Jozef Madzo, Aparna Vasanthakumar, Elodie Pronier, Jay Patel, Yushan Li, Jihae Ahn, Omar Abdel-Wahab, Alan Shih, Chao Lu, Patrick S. Ward, Jennifer J. Tsai, Todd Hricik, Valeria Tosello, Jacob E. Tallman, Xinyang Zhao, Danette Daniels, Qing Dai, Luisa Ciminio, Iannis Aifantis, Chuan He, Francois Fuks, Martin S. Tallman, Adolfo Ferrando, Stephen Nimer, Elisabeth Paietta, Craig B. Thompson, Jonathan D. Licht, Christopher E. Mason, Lucy A. Godley, Ari Melnick, Maria E. Figueroa, and Ross L. Levine. DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia. *Cell Reports*, 9(5):1841–1855, December 2014. ISSN 22111247. doi: 10.1016/j.celrep.2014.11.004.

Fidel Ramírez, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S. Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1):W160–W165, July 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw257.

F. Ann Ran, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11): 2281–2308, November 2013. ISSN 1750-2799. doi: 10.1038/nprot.2013.143.

Kasper D. Rasmussen, Guangshuai Jia, Jens V. Johansen, Marianne T. Pedersen, Nicolas Rapin, Frederik O. Bagger, Bo T. Porse, Olivier A. Bernard, Jesper Christensen, and Kristian Helin. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes & Development*, 29(9): 910–922, January 2015. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.260174.115.

Kasper D. Rasmussen, Ivan Berest, Sandra Keler, Koutarou Nishimura, Lucía Simón-Carrasco, George S. Vassiliou, Marianne T. Pedersen, Jesper Christensen, Judith B. Zaugg, and Kristian Helin. TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. *Genome Research*, 29(4):564–575, April 2019. ISSN 1088-9051. doi: 10.1101/gr.239277.118.

Kasper Dindler Rasmussen and Kristian Helin. Role of TET enzymes in DNA methylation, development, and cancer. *Genes & Development*, 30(7):733–750, April 2016. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.276568.115.

David Ratel, Jean-Luc Ravanat, François Berger, and Didier Wion. N6-methyladenine: the other methylated base of DNA. *BioEssays*, 28(3):309–315, 2006. ISSN 1521-1878. doi: 10.1002/bies.20342.

James P Reddington, Sara M Perricone, Colm E Nestor, Judith Reichmann, Neil A Youngson, Masako Suzuki, Diana Reinhardt, Donncha S Dunican, James G Prendergast, Heidi Mjoseng, Bernard H Ramsahoye, Emma Whitelaw, John M Greally, Ian R Adams, Wendy A Bickmore, and Richard R Meehan. Redistribution of H3K27me3 upon DNA

hypomethylation results in de-repression of Polycomb target genes. *Genome Biology*, 14(3):R25, 2013. ISSN 1465-6906. doi: 10.1186/gb-2013-14-3-r25.

Daniel Renciuk, Olivier Blacque, Michaela Vorlickova, and Bernhard Spingler. Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Research*, 41(21):9891–9900, November 2013. ISSN 0305-1048. doi: 10.1093/nar/gkt738.

W. M. Rideout, G. A. Coetzee, A. F. Olumi, and P. A. Jones. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science (New York, N.Y.)*, 249(4974), September 1990. ISSN 0036-8075. doi: 10.1126/science.1697983.

Leonie Ringrose and Renato Paro. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annual Review of Genetics*, 38:413–443, 2004. ISSN 0066-4197. doi: 10.1146/annurev.genet.38.072902.091907.

Neil A. Robertson, Robert F. Hillary, Daniel L. McCartney, Maria Terradas-Terradas, Jonathan Higham, Duncan Sproul, Ian J. Deary, Kristina Kirschner, Riccardo E. Marioni, and Tamir Chandra. Age-related clonal haemopoiesis is associated with increased epigenetic age. *Current Biology*, 29(16):R786–R787, August 2019. ISSN 0960-9822. doi: 10.1016/j.cub.2019.07.011.

Kyle J. Roux, Dae In Kim, Manfred Raida, and Brian Burke. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology*, 196(6):801–810, March 2012. ISSN 1540-8140, 0021-9525. doi: 10.1083/jcb.201112098.

Julian M. Rozenberg, Andrey Shlyakhtenko, Kimberly Glass, Vikas Rishi, Maxim V. Myakishev, Peter C. FitzGerald, and Charles Vinson. All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC genomics*, 9:67, February 2008. ISSN 1471-2164. doi: 10.1186/1471-2164-9-67.

David A. Russler-Germain, David H. Spencer, Margaret A. Young, Tamara L. Lamprecht, Christopher A. Miller, Robert Fulton, Matthew R. Meyer, Petra Erdmann-Gilmore, R. Reid Townsend, Richard K. Wilson, and Timothy J. Ley. The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. *Cancer Cell*, 25(4):442–454, April 2014. ISSN 1535-6108, 1878-3686. doi: 10.1016/j.ccr.2014.02.010.

Takashi Sado, Masaki Okano, En Li, and Hiroyuki Sasaki. De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Development (Cambridge, England)*, 131(5):975–982, March 2004. ISSN 0950-1991. doi: 10.1242/dev.00995.

Kaoru Sakabe, Zihao Wang, and Gerald W. Hart. -N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences*, 107 (46):19915–19920, November 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1009023107.

Morito Sakaue, Hiroshi Ohta, Yuichi Kumaki, Masaaki Oda, Yuko Sakaide, Chisa Matsuoka, Akiko Yamagiwa, Hitoshi Niwa, Teruhiko Wakayama, and Masaki Okano. DNA Methylation Is Dispensable for the Growth and Survival of the Extraembryonic Lineages. *Current Biology*, 20(16):1452–1457, August 2010. ISSN 0960-9822. doi: 10.1016/j.cub.2010.06.050.

Jose Luis Sardina, Samuel Collombet, Tian V. Tian, Antonio Gómez, Bruno Di Stefano, Clara Berenguer, Justin Brumbaugh, Ralph Stadhouders, Carolina Segura-Morales, Marta Gut, Ivo G. Gut, Simon Heath, Sergi Aranda, Luciano Di Croce, Konrad Hochedlinger, Denis Thieffry, and Thomas Graf. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell*, 23(5): 727–741.e9, November 2018. ISSN 19345909. doi: 10.1016/j.stem.2018.08.016.

N. Max Schabla, Koushik Mondal, and Patrick C. Swanson. DCAF1 (VprBP): emerging physiological roles for a unique dual-service E3 ubiquitin ligase substrate receptor.

Journal of Molecular Cell Biology, 11(9):725–735, September 2019. doi: 10.1093/jmcb/mjy085.

Roland Schmitz, George W. Wright, Da Wei Huang, Calvin A. Johnson, James D. Phelan, James Q. Wang, Sandrine Roulland, Monica Kasbekar, Ryan M. Young, Arthur L. Shaffer, Daniel J. Hodson, Wenming Xiao, Xin Yu, Yandan Yang, Hong Zhao, Weihong Xu, Xuelu Liu, Bin Zhou, Wei Du, Wing C. Chan, Elaine S. Jaffe, Randy D. Gascoyne, Joseph M. Connors, Elias Campo, Armando Lopez-Guillermo, Andreas Rosenwald, German Ott, Jan Delabie, Lisa M. Rimsza, Kevin Tay Kuang Wei, Andrew D. Zelenetz, John P. Leonard, Nancy L. Bartlett, Bao Tran, Jyoti Shetty, Yongmei Zhao, Dan R. Soppet, Stefania Pittaluga, Wyndham H. Wilson, and Louis M. Staudt. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine*, 378(15):1396–1407, April 2018. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMoa1801445.

Stefan Schoenfelder and Peter Fraser. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics*, 20(8), August 2019. ISSN 1471-0064. doi: 10.1038/s41576-019-0128-0.

Laurianne Scourzic, Enguerran Mouly, and Olivier A Bernard. TET proteins and the control of cytosine demethylation in cancer. *Genome Medicine*, 7(1):9, 2015. ISSN 1756-994X. doi: 10.1186/s13073-015-0134-6.

Stefanie Seisenberger, Simon Andrews, Felix Krueger, Julia Arand, Jörn Walter, Fátima Santos, Christian Popp, Bernard Thienpont, Wendy Dean, and Wolf Reik. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular Cell*, 48(6):849–862, December 2012. ISSN 1097-4164. doi: 10.1016/j.molcel.2012.11.001.

Jeong-Sun Seo, Young Seok Ju, Won-Chul Lee, Jong-Yeon Shin, June Koo Lee, Thomas Bleazard, Junho Lee, Yoo Jin Jung, Jung-Oh Kim, Jung-Young Shin, Saet-Byeol Yu, Jihye Kim, Eung-Ryoung Lee, Chang-Hyun Kang, In-Kyu Park, Hwanseok Rhee, Se-Hoon Lee, Jong-Il Kim, Jin-Hyoung Kang, and Young Tae Kim. The transcriptional

landscape and mutational profile of lung adenocarcinoma. *Genome Research*, 22(11): 2109–2119, November 2012. ISSN 1088-9051. doi: 10.1101/gr.145144.112.

M. Serrano, A. W. Lin, M. E. McCurrach, D. Beach, and S. W. Lowe. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell*, 88(5):593–602, March 1997. ISSN 0092-8674. doi: 10.1016/s0092-8674(00)81902-9.

Manual Serrano, Gregory J. Hannon, and David Beach. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature*, 366(6456): 704–707, December 1993. ISSN 1476-4687. doi: 10.1038/366704a0.

Somasekar Seshagiri, Eric W. Stawiski, Steffen Durinck, Zora Modrusan, Elaine E. Storm, Caitlin B. Conboy, Subhra Chaudhuri, Yinghui Guan, Vasantharajan Janakiraman, Bijay S. Jaiswal, Joseph Guillory, Connie Ha, Gerrit J. P. Dijkgraaf, Jeremy Stinson, Florian Gnad, Melanie A. Huntley, Jeremiah D. Degenhardt, Peter M. Haverty, Richard Bourgon, Weiru Wang, Hartmut Koeppen, Robert Gentleman, Timothy K. Starr, Zemin Zhang, David A. Largaespada, Thomas D. Wu, and Frederic J. de Sauvage. Recurrent R-spondin fusions in colon cancer. *Nature*, 488(7413):660–664, August 2012. ISSN 1476-4687. doi: 10.1038/nature11282.

Li Shen, Azusa Inoue, Jin He, Yuting Liu, Falong Lu, and Yi Zhang. Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell*, 15(4):459–471, October 2014. ISSN 19345909. doi: 10.1016/j.stem.2014.09.002.

Qicong Shen, Qian Zhang, Yang Shi, Qingzhu Shi, Yanyan Jiang, Yan Gu, Zhiqing Li, Xia Li, Kai Zhao, Chunmei Wang, Nan Li, and Xuetao Cao. Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature*, 554(7690):123–127, February 2018. ISSN 1476-4687. doi: 10.1038/nature25434.

Mairi S. Shepherd, Juan Li, Nicola K. Wilson, Caroline A. Oedekoven, Jiangbing Li, Miriam Belmonte, Juergen Fink, Janine C. M. Prick, Dean C. Pask, Tina L. Hamilton,

Dirk Loeffler, Anjana Rao, Timm Schröder, Berthold Göttgens, Anthony R. Green, and David G. Kent. Single-cell approaches identify the molecular network driving malignant hematopoietic stem cell self-renewal. *Blood*, 132(8):791–803, August 2018. ISSN 0006-4971. doi: 10.1182/blood-2017-12-821066.

K. Shide, T. Kameda, H. Shimoda, T. Yamaji, H. Abe, A. Kamiunten, M. Sekine, T. Hidaka, K. Katayose, Y. Kubuki, S. Yamamoto, T. Miike, H. Iwakiri, S. Hasuike, K. Nagata, K. Marutsuka, A. Iwama, T. Matsuda, A. Kitanaka, and K. Shimoda. TET2 is essential for survival and hematopoietic stem cell homeostasis. *Leukemia*, 26(10): 2216–2223, October 2012. ISSN 1476-5551. doi: 10.1038/leu.2012.94.

Alan H. Shih, Yanwen Jiang, Cem Meydan, Kaitlyn Shank, Suveg Pandey, Laura Barreyro, Ileana Antony-Debre, Agnes Viale, Nicholas Socci, Yongming Sun, Alexander Robertson, Magali Cavatore, Elisa de Stanchina, Todd Hricik, Franck Rapaport, Brittany Woods, Chen Wei, Megan Hatlen, Muhamed Baljevic, Stephen D. Nimer, Martin Tallman, Elisabeth Paietta, Luisa Cimmino, Iannis Aifantis, Ulrich Steidl, Chris Mason, Ari Melnick, and Ross L. Levine. Mutational Cooperativity Linked to Combinatorial Epigenetic Gain of Function in Acute Myeloid Leukemia. *Cancer Cell*, 27(4):502–515, April 2015. ISSN 1535-6108, 1878-3686. doi: 10.1016/j.ccell.2015.03.009.

Pavithra Shyamsunder, Mahalakshmi Shanmugasundaram, Anand Mayakonda, Pushkar Dakle, Weoi Woon Teoh, Lin Han, Deepika Kanojia, Mei Chee Lim, Melissa Fullwood, Omer An, Henry Yang, Jizhong Shi, Mohammad Zakir Hossain, Vikas Madan, and H. Phillip Koeffler. Identification of a novel enhancer of CEBPE essential for granulocytic differentiation. *Blood*, 133(23):2507–2517, June 2019. ISSN 0006-4971. doi: 10.1182/blood.2018886077.

Alexander E. Smith, Azim M. Mohamedali, Austin Kulasekararaj, ZiYi Lim, Joop Gäken, Nicholas C. Lea, Bartlomiej Przychodzen, Syed A. Mian, Erick E. Nasser, Claire Shooter, Nigel B. Westwood, Corinna Strupp, Norbert Gattermann, Jaroslaw P. Maciejewski, Ulrich Germing, and Ghulam J. Mufti. Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with

early origins, but indicates no definite prognostic value. *Blood*, 116(19):3923–3932, November 2010. ISSN 0006-4971, 1528-0020. doi: 10.1182/blood-2010-03-274704.

Cynthia L. Smith and Janan T. Eppig. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, 1(3):390–399, December 2009. ISSN 1939-005X. doi: 10.1002/wsbm.44.

E Solary, O A Bernard, A Tefferi, F Fuks, and W Vainchenker. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia*, 28(3):485–496, March 2014. ISSN 0887-6924, 1476-5551. doi: 10.1038/leu.2013.337.

Tim D.D. Somerville, Fabrizio Simeoni, John A. Chadwick, Emma L. Williams, Gary J. Spencer, Katalin Boros, Christopher Wirth, Eleni Tholouli, Richard J. Byers, and Tim C.P. Somervaille. Derepression of the Iroquois Homeodomain Transcription Factor Gene IRX3 Confers Differentiation Block in Acute Leukemia. *Cell Reports*, 22 (3):638–652, January 2018. ISSN 22111247. doi: 10.1016/j.celrep.2017.12.063.

Chengli Song, Lina Wang, Xiaoyan Wu, Kai Wang, Dan Xie, Qi Xiao, Songyu Li, Kui Jiang, Lujian Liao, John R. Yates, Jiing-Dwan Lee, and Qingkai Yang. PML Recruits TET2 to Regulate DNA Modification and Cell Proliferation in Response to Chemotherapeutic Agent. *Cancer Research*, 78(10):2475–2489, 2018. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-17-3091.

Chun-Xiao Song, Keith E. Szulwach, Ye Fu, Qing Dai, Chengqi Yi, Xuekun Li, Yujing Li, Chih-Hsin Chen, Wen Zhang, Xing Jian, Jing Wang, Li Zhang, Timothy J. Looney, Baichen Zhang, Lucy A. Godley, Leslie M. Hicks, Bruce T. Lahn, Peng Jin, and Chuan He. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*, 29(1):68–72, January 2011a. ISSN 1546-1696. doi: 10.1038/nbt.1732.

Jikui Song, Olga Rechkoblit, Timothy H. Bestor, and Dinshaw J. Patel. Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA

Methylation. *Science (New York, N.Y.)*, 331(6020):1036–1040, February 2011b. ISSN 0036-8075. doi: 10.1126/science.1195380.

Su Jung Song, Keisuke Ito, Ugo Ala, Lev Kats, Kaitlyn Webster, Su Ming Sun, Mojca Jongen-Lavrencic, Katia Manova-Todorova, Julie Teruya-Feldstein, David E. Avigan, Ruud Delwel, and Pier Paolo Pandolfi. The Oncogenic MicroRNA miR-22 Targets the TET2 Tumor Suppressor to Promote Hematopoietic Stem Cell Self-Renewal and Transformation. *Cell Stem Cell*, 13(1):87–101, July 2013. ISSN 19345909. doi: 10.1016/j.stem.2013.06.003.

Cornelia G. Spruijt, Felix Gnerlich, Arne H. Smits, Toni Pfaffeneder, Pascal W. T. C. Jansen, Christina Bauer, Martin Münzel, Mirko Wagner, Markus Müller, Fariha Khan, H. Christian Eberl, Anneloes Mensinga, Arie B. Brinkman, Konstantin Lephikov, Udo Müller, Jörn Walter, Rolf Boelens, Hugo van Ingen, Heinrich Leonhardt, Thomas Carell, and Michiel Vermeulen. Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell*, 152(5):1146–1159, February 2013. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2013.02.004.

Michael B. Stadler, Rabih Murr, Lukas Burger, Robert Ivanek, Florian Lienert, Anne Schöler, Erik van Nimwegen, Christiane Wirbelauer, Edward J. Oakeley, Dimos Gaidatzis, Vijay K. Tiwari, and Dirk Schübeler. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378):490–495, December 2011. ISSN 1476-4687. doi: 10.1038/nature10716.

Rory Stark and Gord Brown. DiffBind: Differential binding analysis of ChIP-Seq peak data. page 33.

Yonatan Stelzer, Chikdu Shakti Shivalila, Frank Soldner, Styliani Markoulaki, and Rudolf Jaenisch. Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell*, 163(1):218–229, September 2015. ISSN 1097-4172. doi: 10.1016/j.cell.2015.08.046.

Melanie Stumpf, Claudia Waskow, Marit Krötschel, Dominic van Essen, Patrick Rodriguez, Xiaoting Zhang, Boris Guyot, Robert G. Roeder, and Tilman Borggrefe.

The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49):18504–18509, December 2006. ISSN 0027-8424. doi: 10.1073/pnas.0604494103.

Jie Sun, Xin He, Yinghui Zhu, Zonghui Ding, Haojie Dong, Yimei Feng, Juan Du, Hanying Wang, Xiwei Wu, Lei Zhang, Xiaochun Yu, Allen Lin, Tinisha McDonald, Dandan Zhao, Herman Wu, Wei-Kai Hua, Bin Zhang, Lifeng Feng, Kaoru Tohyama, Ravi Bhatia, Philipp Oberdoerffer, Yang Jo Chung, Peter D. Aplan, Jacqueline Boultwood, Andrea Pellagatti, Samer Khaled, Marcin Kortylewski, Flavia Pichiorri, Ya-Huei Kuo, Nadia Carlesso, Guido Marcucci, Hongchuan Jin, and Ling Li. SIRT1 Activation Disrupts Maintenance of Myelodysplastic Syndrome Stem and Progenitor Cells by Restoring TET2 Function. *Cell Stem Cell*, 23(3):355–369.e9, 2018. ISSN 1875-9777. doi: 10.1016/j.stem.2018.07.018.

Hiromu Suzuki, D Neil Watkins, Kam-Wing Jair, Kornel E Schuebel, Sanford D Markowitz, Wei Dong Chen, Theresa P Pretlow, Bin Yang, Yoshimitsu Akiyama, Manon van Engeland, Minoru Toyota, Takashi Tokino, Yuji Hinoda, Kohzoh Imai, James G Herman, and Stephen B Baylin. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nature Genetics*, 36(4):417–422, April 2004. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng1330.

Miho M. Suzuki and Adrian Bird. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9(6):465–476, June 2008. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg2341.

Takahiro Suzuki, Yuri Shimizu, Erina Furuhata, Shiori Maeda, Mami Kishima, Hajime Nishimura, Saaya Enomoto, Yoshihide Hayashizaki, and Harukazu Suzuki. RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. *Blood Advances*, 1(20):1699–1711, September 2017. ISSN 2473-9529. doi: 10.1182/bloodadvances.2017005710.

A. H. Swirnoff and J. Milbrandt. DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Molecular and Cellular Biology*, 15(4):2275–2287, April 1995. ISSN 0270-7306. doi: 10.1128/mcb.15.4.2275.

Keith E. Szulwach, Xuekun Li, Yujing Li, Chun-Xiao Song, Hao Wu, Qing Dai, Hasan Irier, Anup K. Upadhyay, Marla Gearing, Allan I. Levey, Aparna Vasanthakumar, Lucy A. Godley, Qiang Chang, Xiaodong Cheng, Chuan He, and Peng Jin. 5-hmC–mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature Neuroscience*, 14(12):1607–1616, December 2011. ISSN 1546-1726. doi: 10.1038/nn.2959.

Manuel Sánchez-Castillo, David Ruau, Adam C. Wilkinson, Felicia S.L. Ng, Rebecca Hannah, Evangelia Diamanti, Patrick Lombard, Nicola K. Wilson, and Berthold Gottgens. CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Research*, 43(D1): D1117–D1123, January 2015. ISSN 1362-4962, 0305-1048. doi: 10.1093/nar/gku895.

Phillippa C. Taberlay, Theresa K. Kelly, Chun-Chi Liu, Jueng Soo You, Daniel D. De Carvalho, Tina B. Miranda, Xianghong J. Zhou, Gangning Liang, and Peter A. Jones. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell*, 147(6):1283–1294, December 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.10.040.

Hiromi Tagoh, Roy Himes, Deborah Clarke, Pieter J. M. Leenen, Arthur D. Riggs, David Hume, and Constanze Bonifer. Transcription factor complex formation and chromatin fine structure alterations at the murine c-fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. *Genes & Development*, 16(13):1721–1737, January 2002. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.222002.

M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, 324 (5929):930–935, May 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1170116.

Franziska Taruttis, Maren Feist, Phillip Schwarzfischer, Wolfram Gronwald, Dieter Kube, Rainer Spang, and Julia C. Engelmann. External calibration with *Drosophila* whole-cell spike-ins delivers absolute mRNA fold changes from human RNA-seq and qPCR data. 62(2). ISSN 1940-9818, 0736-6205. doi: 10.2144/000114514.

Katrina Tatton-Brown, Sheila Seal, Elise Ruark, Jenny Harmer, Emma Ramsay, Silvana Del Vecchio Duarte, Anna Zachariou, Sandra Hanks, Eleanor O'Brien, Lise Aksglaede, Diana Baralle, Tabib Dabir, Blanca Gener, David Goudie, Tessa Homfray, Ajith Kumar, Daniela T. Pilz, Angelo Selicorni, I. Karen Temple, Lionel Van Maldergem, Naomi Yachelevich, Childhood Overgrowth Consortium, Robert van Montfort, and Nazneen Rahman. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nature Genetics*, 46(4):385–388, April 2014. ISSN 1546-1718. doi: 10.1038/ng.2917.

A Tefferi, A Pardanani, K-H Lim, O Abdel-Wahab, T L Lasho, J Patel, N Gangat, C M Finke, S Schwager, A Mullally, C-Y Li, C A Hanson, R Mesa, O Bernard, F Delhommeau, W Vainchenker, D G Gilliland, and R L Levine. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. *Leukemia*, 23(5):905–911, May 2009. ISSN 0887-6924, 1476-5551. doi: 10.1038/leu.2009.47.

Bing Tian, Jun Yang, and Allan R. Brasier. Two-step Crosslinking for Analysis of Protein-Chromatin Interactions. *Methods in molecular biology (Clifton, N.J.)*, 809:105–120, 2012. ISSN 1064-3745. doi: 10.1007/978-1-61779-376-9_7.

Magda Tomala, Antonina Lavrentieva, Pierre Moretti, Ursula Rinas, Cornelia Kasper, Frank Stahl, Axel Schambach, Eva Warlich, Ulrich Martin, Tobias Cantz, and Thomas Scheper. Preparation of bioactive soluble human leukemia inhibitory factor from recombinant Escherichia coli using thioredoxin as fusion partner. *Protein Expression and Purification*, 73(1):51–57, September 2010. ISSN 1046-5928. doi: 10.1016/j.pep.2010.04.002.

Serena Tommasini-Ghelfi, Kevin Murnan, Fotini M. Kouri, Akanksha S. Mahajan, Jasmine L. May, and Alexander H. Stegh. Cancer-associated mutation and beyond: The emerging biology of isocitrate dehydrogenases in human disease. *Science Advances*, 5 (5):eaaw4543, May 2019. ISSN 2375-2548. doi: 10.1126/sciadv.aaw4543.

Laura Trinkle-Mulcahy. Recent advances in proximity-based labeling methods for interactome mapping. *F1000Research*, 8, January 2019. ISSN 2046-1402. doi: 10.12688/f1000research.16903.1.

Akiko Tsumura, Tomohiro Hayakawa, Yuichi Kumaki, Shin-ichiro Takebayashi, Morito Sakaue, Chisa Matsuoka, Kunitada Shimotohno, Fuyuki Ishikawa, En Li, Hiroki R. Ueda, Jun-ichi Nakayama, and Masaki Okano. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells*, 11(7):805–814, 2006. ISSN 1365-2443. doi: 10.1111/j.1365-2443.2006.00984.x.

Sevin Turcan, Daniel Rohle, Anuj Goenka, Logan A. Walsh, Fang Fang, Emrullah Yilmaz, Carl Campos, Armida W. M. Fabius, Chao Lu, Patrick S. Ward, Craig B. Thompson, Andrew Kaufman, Olga Guryanova, Ross Levine, Adriana Heguy, Agnes Viale, Luc G. T. Morris, Jason T. Huse, Ingo K. Mellinghoff, and Timothy A. Chan. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*, 483 (7390):479–483, March 2012. ISSN 1476-4687. doi: 10.1038/nature10866.

Stefka Tyanova, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y. Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13(9), September 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3901.

Jeffrey W. Tyner, Cristina E. Tognon, Daniel Bottomly, Beth Wilmot, Stephen E. Kurtz, Samantha L. Savage, Nicola Long, Anna Reister Schultz, Elie Traer, Melissa Abel, Anupriya Agarwal, Aurora Blucher, Uma Borate, Jade Bryant, Russell Burke, Amy Carlos, Richie Carpenter, Joseph Carroll, Bill H. Chang, Cody Coblentz, Amanda d'Almeida, Rachel Cook, Alexey Danilov, Kim-Hien T. Dao, Michie Degnin, Deirdre

Devine, James Dibb, David K. Edwards, Christopher A. Eide, Isabel English, Jason Glover, Rachel Henson, Hibery Ho, Abdusebur Jemal, Kara Johnson, Ryan Johnson, Brian Junio, Andy Kaempf, Jessica Leonard, Chenwei Lin, Selina Qiuying Liu, Pierrette Lo, Marc M. Loriaux, Samuel Luty, Tara Macey, Jason MacManiman, Jacqueline Martinez, Motomi Mori, Dylan Nelson, Ceilidh Nichols, Jill Peters, Justin Ramsdill, Angela Rofelty, Robert Schuff, Robert Searles, Erik Segerdell, Rebecca L. Smith, Stephen E. Spurgeon, Tyler Sweeney, Aashis Thapa, Corinne Visser, Jake Wagner, Kevin Watanabe-Smith, Kristen Werth, Joelle Wolf, Libbey White, Amy Yates, Haijiao Zhang, Christopher R. Cogle, Robert H. Collins, Denise C. Connolly, Michael W. Deininger, Leylah Drusbosky, Christopher S. Hourigan, Craig T. Jordan, Patricia Kropf, Tara L. Lin, Micaela E. Martinez, Bruno C. Medeiros, Rachel R. Pallapati, Daniel A. Pollyea, Ronan T. Swords, Justin M. Watts, Scott J. Weir, David L. Wiest, Ryan M. Winters, Shannon K. McWeeney, and Brian J. Druker. Functional genomic landscape of acute myeloid leukaemia. *Nature*, 562(7728):526–531, October 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0623-z.

Santiago Uribe-Lewis, Rory Stark, Thomas Carroll, Mark J Dunning, Martin Bachman, Yoko Ito, Lovorka Stojic, Silvia Halim, Sarah L Vowler, Andy G Lynch, Benjamin Delatte, Eric J de Bony, Laurence Colin, Matthieu Defrance, Felix Krueger, Ana-Luisa Silva, Rogier ten Hoopen, Ashraf EK Ibrahim, François Fuks, and Adele Murrell. 5-hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer. *Genome Biology*, 16(1), December 2015. ISSN 1474-760X. doi: 10.1186/s13059-015-0605-5.

Christopher R. Vakoc, Danielle L. Letting, Nele Gheldof, Tomoyuki Sawado, M. A. Bender, Mark Groudine, Mitchell J. Weiss, Job Dekker, and Gerd A. Blobel. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular Cell*, 17(3):453–462, February 2005. ISSN 1097-2765. doi: 10.1016/j.molcel.2004.12.028.

Christopher R. Vakoc, Mira M. Sachdeva, Hongxin Wang, and Gerd A. Blobel. Profile of histone lysine methylation across transcribed mammalian chromatin. *Molecular*

and Cellular Biology, 26(24):9185–9195, December 2006. ISSN 0270-7306. doi: 10.1128/MCB.01529-06.

Victoria Valinluck and Lawrence C. Sowers. Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1. *Cancer Research*, 67(3):946–950, February 2007. ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-06-3123.

Pietro Vella, Andrea Scelfo, SriGanesh Jammula, Fulvio Chiacchiera, Kristine Williams, Alessandro Cuomo, Alessandra Roberto, Jesper Christensen, Tiziana Bonaldi, Kristian Helin, and Diego Pasini. Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Molecular Cell*, 49(4):645–656, February 2013. ISSN 1097-2765. doi: 10.1016/j.molcel.2012.12.019.

John J. Vincent, Yun Huang, Pao-Yang Chen, Suhua Feng, Joseph H. Calvopiña, Kevin Nee, Serena A. Lee, Thuc Le, Alexander J. Yoon, Kym Faull, Guoping Fan, Anjana Rao, Steven E. Jacobsen, Matteo Pellegrini, and Amander T. Clark. Stage-Specific Roles for Tet1 and Tet2 in DNA Demethylation in Primordial Germ Cells. *Cell Stem Cell*, 12(4):470–478, April 2013. ISSN 1934-5909. doi: 10.1016/j.stem.2013.01.016.

C. P. Walsh, J. R. Chaillet, and T. H. Bestor. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, 20(2):116–117, October 1998. ISSN 1061-4036. doi: 10.1038/2413.

Dongxue Wang, Hideharu Hashimoto, Xing Zhang, Benjamin G. Barwick, Sagar Lonial, Lawrence H. Boise, Paula M. Vertino, and Xiaodong Cheng. MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Research*, 45(5):2396–2407, March 2017. ISSN 0305-1048. doi: 10.1093/nar/gkw1184.

Haoyi Wang, Hui Yang, Chikdu S. Shivalila, Meelad M. Dawlaty, Albert W. Cheng, Feng Zhang, and Rudolf Jaenisch. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*, 153(4): 910–918, May 2013. ISSN 0092-8674. doi: 10.1016/j.cell.2013.04.025.

Lanfeng Wang, Yu Zhou, Liang Xu, Rui Xiao, Xingyu Lu, Liang Chen, Jenny Chong, Hairi Li, Chuan He, Xiang-Dong Fu, and Dong Wang. Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*, 523(7562):621–625, July 2015a. ISSN 1476-4687. doi: 10.1038/nature14482.

Lu Wang, Jun Zhang, Jialei Duan, Xinxing Gao, Wei Zhu, Xingyu Lu, Lu Yang, Jing Zhang, Guoqiang Li, Weimin Ci, Wei Li, Qi Zhou, Neel Aluru, Fuchou Tang, Chuan He, Xingxu Huang, and Jiang Liu. Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell*, 157(4):979–991, May 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.04.017.

Lu Wang, Patrick A. Ozark, Edwin R. Smith, Zibo Zhao, Stacy A. Marshall, Emily J. Rendleman, Andrea Piunti, Caila Ryan, Anna L. Whelan, Kathryn A. Helmin, Marc Alard Morgan, Lihua Zou, Benjamin D. Singer, and Ali Shilatifard. TET2 coactivates gene expression through demethylation of enhancers. *Science Advances*, 4 (11):eaau6986, 2018. ISSN 2375-2548. doi: 10.1126/sciadv.aau6986.

Yiping Wang, Mengtao Xiao, Xiufei Chen, Leilei Chen, Yanping Xu, Lei Lv, Pu Wang, Hui Yang, Shenghong Ma, Huaipeng Lin, Bo Jiao, Ruibao Ren, Dan Ye, Kun-Liang Guan, and Yue Xiong. WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation. *Molecular Cell*, 57(4):662–673, February 2015b. ISSN 10972765. doi: 10.1016/j.molcel.2014.12.023.

Yiping Wang, Mengtao Xiao, Xiufei Chen, Leilei Chen, Yanping Xu, Lei Lv, Pu Wang, Hui Yang, Shenghong Ma, Huaipeng Lin, Bo Jiao, Ruibao Ren, Dan Ye, Kun-Liang Guan, and Yue Xiong. WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation. *Molecular Cell*, 57(4):662–673, February 2015c. ISSN 10972765. doi: 10.1016/j.molcel.2014.12.023.

Yu Wang and Yi Zhang. Regulation of TET Protein Stability by Calpains. *Cell Reports*, 6(2):278–284, January 2014. ISSN 2211-1247. doi: 10.1016/j.celrep.2013.12.031.

Zhibin Wang, Chongzhi Zang, Jeffrey A. Rosenfeld, Dustin E. Schones, Artem Barski, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Weiqun Peng, Michael Q. Zhang, and Keji Zhao. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, 40(7):897–903, July 2008. ISSN 1546-1718. doi: 10.1038/ng.154.

Patrick S. Ward, Jay Patel, David R. Wise, Omar Abdel-Wahab, Bryson D. Bennett, Hilary A. Coller, Justin R. Cross, Valeria R. Fantin, Cyrus V. Hedvat, Alexander E. Perl, Joshua D. Rabinowitz, Martin Carroll, Shinsan M. Su, Kim A. Sharp, Ross L. Levine, and Craig B. Thompson. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell*, 17(3):225–234, March 2010. ISSN 1878-3686. doi: 10.1016/j.ccr.2010.01.020.

Alain R. Weber, Claudia Krawczyk, Adam B. Robertson, Anna Kuśnierczyk, Cathrine B. Vågbø, David Schuermann, Arne Klungland, and Primo Schär. Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications*, 7(1):1–13, March 2016. ISSN 2041-1723. doi: 10.1038/ncomms10806.

Erin L. Weber and Paula M. Cannon. Promoter Choice for Retroviral Vectors: Transcriptional Strength Versus Trans-Activation Potential, October 2007.

S Weissmann, T Alpermann, V Grossmann, A Kowarsch, N Nadarajah, C Eder, F Dicker, A Fasan, C Haferlach, T Haferlach, W Kern, S Schnittger, and A Kohlmann. Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia*, 26(5):934–942, May 2012. ISSN 0887-6924, 1476-5551. doi: 10.1038/leu.2011.326.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Use R! Springer-Verlag, New York, 2009. ISBN 978-0-387-98141-3. doi: 10.1007/978-0-387-98141-3.

Laura Wiehle, Günter Raddatz, Tanja Musch, Meelad M. Dawlaty, Rudolf Jaenisch, Frank Lyko, and Achim Breiling. Tet1 and Tet2 Protect DNA Methylation Canyons against

Hypermethylation. *Molecular and Cellular Biology*, 36(3):452–461, February 2016. ISSN 0270-7306, 1098-5549. doi: 10.1128/MCB.00587-15.

Malgorzata Wiench, Sam John, Songjoon Baek, Thomas A. Johnson, Myong-Hee Sung, Thelma Escobar, Catherine A. Simmons, Kenneth H. Pearce, Simon C. Biddie, Pete J. Sabo, Robert E. Thurman, John A. Stamatoyannopoulos, and Gordon L. Hager. DNA methylation status predicts cell type-specific enhancer activity. *The EMBO journal*, 30 (15):3028–3039, June 2011. ISSN 1460-2075. doi: 10.1038/emboj.2011.210.

Kristine Williams, Jesper Christensen, Marianne Terndrup Pedersen, Jens V. Johansen, Paul A. C. Cloos, Juri Rappsilber, and Kristian Helin. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, 473(7347):343–348, May 2011. ISSN 1476-4687. doi: 10.1038/nature10066.

Geoffrey G. Wilson and Noreen E. Murray. Restriction and Modification Systems. *Annual Review of Genetics*, 25(1), 1991. doi: 10.1146/annurev.ge.25.120191.003101.

Nicola K. Wilson, Stefan Schoenfelder, Rebecca Hannah, Manuel Sánchez Castillo, Judith Schütte, Vasileios Ladopoulos, Joanna Mitchelmore, Debbie K. Goode, Fernando J. Calero-Nieto, Victoria Moignard, Adam C. Wilkinson, Isabel Jimenez-Madrid, Sarah Kinston, Mikhail Spivakov, Peter Fraser, and Berthold Göttgens. Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood*, 127(13):e12–e23, March 2016. ISSN 0006-4971. doi: 10.1182/blood-2015-10-677393.

Juliane Winkelmann, Ling Lin, Barbara Schormair, Birgitte R. Kornum, Juliette Faraco, Giuseppe Plazzi, Atle Melberg, Ferdinando Cornelio, Alexander E. Urban, Fabio Pizza, Francesca Poli, Fabian Grubert, Thomas Wieland, Elisabeth Graf, Joachim Hallmayer, Tim M. Strom, and Emmanuel Mignot. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Human Molecular Genetics*, 21(10):2205–2210, May 2012. ISSN 0964-6906. doi: 10.1093/hmg/dds035.

S. F. Wolf, D. J. Jolly, K. D. Lunnen, T. Friedmann, and B. R. Migeon. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proceedings of the National Academy of Sciences*, 81(9):2806–2810, May 1984. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.81.9.2806. URL https://www.pnas.org/content/81/9/2806.

Di Wu, Di Hu, Hao Chen, Guoming Shi, Irfete S. Fetahu, Feizhen Wu, Kimberlie Rabidou, Rui Fang, Li Tan, Shuyun Xu, Hang Liu, Christian Argueta, Lei Zhang, Fei Mao, Guoquan Yan, Jiajia Chen, Zhaoru Dong, Ruitu Lv, Yufei Xu, Mei Wang, Yong Ye, Shike Zhang, Danielle Duquette, Songmei Geng, Clark Yin, Christine Guo Lian, George F. Murphy, Gail K. Adler, Rajesh Garg, Lydia Lynch, Pengyuan Yang, Yiming Li, Fei Lan, Jia Fan, Yang Shi, and Yujiang Geno Shi. Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer. *Nature*, 559(7715):637–641, July 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0350-5.

H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, and Y. Zhang. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development*, 25(7):679–684, April 2011a. ISSN 0890-9369. doi: 10.1101/gad.2036011.

Hao Wu, Ana C. D'Alessio, Shinsuke Ito, Kai Xia, Zhibin Wang, Kairong Cui, Keji Zhao, Yi Eve Sun, and Yi Zhang. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, 473(7347):389–393, May 2011b. ISSN 1476-4687. doi: 10.1038/nature09934.

Xiaoji Wu and Yi Zhang. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics*, 18(9):517–534, September 2017. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg.2017.33.

G. R. Wyatt and S. S. Cohen. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochemical Journal*, 55(5):774–782, December 1953. ISSN 0264-6021.

Chuan-Le Xiao, Song Zhu, Minghui He, De Chen, Qian Zhang, Ying Chen, Guoliang Yu, Jinbao Liu, Shang-Qian Xie, Feng Luo, Zhe Liang, De-Peng Wang, Xiao-Chen Bo, Xiao-Feng Gu, Kai Wang, and Guang-Rong Yan. N6-Methyladenine DNA Modification in the Human Genome. *Molecular Cell*, 71(2):306–318.e7, July 2018. ISSN 10972765. doi: 10.1016/j.molcel.2018.06.015.

Wei Xie, Matthew D. Schultz, Ryan Lister, Zhonggang Hou, Nisha Rajagopal, Pradipta Ray, John W. Whitaker, Shulan Tian, R. David Hawkins, Danny Leung, Hongbo Yang, Tao Wang, Ah Young Lee, Scott A. Swanson, Jiuchun Zhang, Yun Zhu, Audrey Kim, Joseph R. Nery, Mark A. Urich, Samantha Kuan, Chia-an Yen, Sarit Klugman, Pengzhi Yu, Kran Suknuntha, Nicholas E. Propson, Huaming Chen, Lee E. Edsall, Ulrich Wagner, Yan Li, Zhen Ye, Ashwinikumar Kulkarni, Zhenyu Xuan, Wen-Yu Chung, Neil C. Chi, Jessica E. Antosiewicz-Bourget, Igor Slukvin, Ron Stewart, Michael Q. Zhang, Wei Wang, James A. Thomson, Joseph R. Ecker, and Bing Ren. Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells. *Cell*, 153 (5):1134–1148, May 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.04.022.

Jun Xiong, Zhuqiang Zhang, Jiayu Chen, Hua Huang, Yali Xu, Xiaojun Ding, Yong Zheng, Ryuichi Nishinakamura, Guo-Liang Xu, Hailin Wang, She Chen, Shaorong Gao, and Bing Zhu. Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. *Molecular Cell*, 64(5):913–925, December 2016. ISSN 10972765. doi: 10.1016/j.molcel.2016.10.013.

Bowen Xu, Ling Cai, Jason M. Butler, Dongliang Chen, Xiongdong Lu, David F. Allison, Rui Lu, Shahin Rafii, Joel S. Parker, Deyou Zheng, and Gang Greg Wang. The Chromatin Remodeler BPTF Activates a Stemness Gene-Expression Program Essential for the Maintenance of Adult Hematopoietic Stem Cells. *Stem Cell Reports*, 10(3): 675–683, 2018. ISSN 2213-6711. doi: 10.1016/j.stemcr.2018.01.020.

Wei Xu, Hui Yang, Ying Liu, Ying Yang, Ping Wang, Se-Hee Kim, Shinsuke Ito, Chen Yang, Pu Wang, Meng-Tao Xiao, Li-xia Liu, Wen-qing Jiang, Jing Liu, Jin-ye Zhang,

Bin Wang, Stephen Frye, Yi Zhang, Yan-hui Xu, Qun-ying Lei, Kun-Liang Guan, Shimin Zhao, and Yue Xiong. Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of -Ketoglutarate-Dependent Dioxygenases. *Cancer Cell*, 19(1):17–30, January 2011a. ISSN 1535-6108, 1878-3686. doi: 10.1016/j.ccr.2010.12.014.

Yufei Xu, Feizhen Wu, Li Tan, Lingchun Kong, Lijun Xiong, Jie Deng, Andrew J. Barbera, Lijuan Zheng, Haikuo Zhang, Stephen Huang, Jinrong Min, Thomas Nicholson, Taiping Chen, Guoliang Xu, Yang Shi, Kun Zhang, and Yujiang Geno Shi. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Molecular Cell*, 42(4):451–464, May 2011b. ISSN 1097-2765. doi: 10.1016/j.molcel.2011.04.005.

Yufei Xu, Chao Xu, Akiko Kato, Wolfram Tempel, Jose Garcia Abreu, Chuanbing Bian, Yeguang Hu, Di Hu, Bin Zhao, Tanja Cerovina, Jianbo Diao, Feizhen Wu, Housheng Hansen He, Qingyan Cui, Erin Clark, Chun Ma, Andrew Barbara, Gert Jan C. Veenstra, Guoliang Xu, Ursula B. Kaiser, X. Shirley Liu, Stephen P. Sugrue, Xi He, Jinrong Min, Yoichi Kato, and Yujiang Geno Shi. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell*, 151(6):1200–1213, December 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2012.11.014.

Shinpei Yamaguchi, Kwonho Hong, Rui Liu, Li Shen, Azusa Inoue, Dinh Diep, Kun Zhang, and Yi Zhang. Tet1 controls meiosis by regulating meiotic gene expression. *Nature*, 492(7429):443–447, December 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11709.

Ryuya Yamanaka, Carrolee Barlow, Julie Lekstrom-Himes, Lucio H. Castilla, Pu P. Liu, Michael Eckhaus, Thomas Decker, Anthony Wynshaw-Boris, and Kleanthis G. Xanthopoulos. Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein -deficient mice. *Proceedings of the National Academy of Sciences*, 94(24):13187–13192, November 1997. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.94.24.13187.

Jumpei Yamazaki, Rodolphe Taby, Aparna Vasanthakumar, Trisha Macrae, Kelly R Ostler, Lanlan Shen, Hagop M Kantarjian, Marcos R Estecio, Jaroslav Jelinek, Lucy A Godley, and Jean-Pierre J Issa. Effects of TET2 mutations on DNA methylation in chronic myelomonocytic leukemia. *Epigenetics*, 7(2):201–207, February 2012. ISSN 1559-2294. doi: 10.4161/epi.7.2.19015.

H Yang, Y Liu, F Bai, J-Y Zhang, S-H Ma, J Liu, Z-D Xu, H-G Zhu, Z-Q Ling, D Ye, K-L Guan, and Y Xiong. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene*, 32(5):663–669, January 2013. ISSN 0950-9232, 1476-5594. doi: 10.1038/onc.2012.67.

Xin Yang, Ying Yang, Bao-Fa Sun, Yu-Sheng Chen, Jia-Wei Xu, Wei-Yi Lai, Ang Li, Xing Wang, Devi Prasad Bhattarai, Wen Xiao, Hui-Ying Sun, Qin Zhu, Hai-Li Ma, Samir Adhikari, Min Sun, Ya-Juan Hao, Bing Zhang, Chun-Min Huang, Niu Huang, Gui-Bin Jiang, Yong-Liang Zhao, Hai-Lin Wang, Ying-Pu Sun, and Yun-Gui Yang. 5-methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m 5 C reader. *Cell Research*, 27(5), May 2017. ISSN 1748-7838. doi: 10.1038/cr.2017.55. Number: 5 pages = 606–625.

S. R. Yant, L. Meuse, W. Chiu, Z. Ivics, Z. Izsvak, and M. A. Kay. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nature Genetics*, 25(1):35–41, May 2000. ISSN 1061-4036. doi: 10.1038/75568.

Yimeng Yin, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, Teemu Kivioja, Kashyap Dave, Fan Zhong, Kazuhiro R. Nitta, Minna Taipale, Alexander Popov, Paul A. Ginno, Silvia Domcke, Jian Yan, Dirk Schübeler, Charles Vinson, and Jussi Taipale. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, 356 (6337), May 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaj2239.

J. A. Yoder, N. S. Soman, G. L. Verdine, and T. H. Bestor. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe.

Journal of Molecular Biology, 270(3):385–395, July 1997a. ISSN 0022-2836. doi: 10.1006/jmbi.1997.1125.

J. A. Yoder, C. P. Walsh, and T. H. Bestor. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics: TIG*, 13(8):335–340, August 1997b. ISSN 0168-9525. doi: 10.1016/s0168-9525(97)01181-5.

Bin Yu, Liu Bi, Binglian Zheng, Lijuan Ji, David Chevalier, Manu Agarwal, Vanitharani Ramachandran, Wanxiang Li, Thierry Lagrange, John C. Walker, and Xuemei Chen. The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proceedings of the National Academy of Sciences*, 105(29):10073–10078, July 2008a. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0804218105.

Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, 16(5):284–287, March 2012a. doi: 10.1089/omi.2011.0118.

Haiyuan Yu, Pascal Braun, Muhammed A. Yildirim, Irma Lemmens, Kavitha Venkatesan, Julie Sahalie, Tomoko Hirozane-Kishikawa, Fana Gebreab, Na Li, Nicolas Simonis, Tong Hao, Jean-François Rual, Amélie Dricot, Alexei Vazquez, Ryan R. Murray, Christophe Simon, Leah Tardivo, Stanley Tam, Nenad Svrzikapa, Changyu Fan, Anne-Sophie de Smet, Adriana Motyl, Michael E. Hudson, Juyong Park, Xiaofeng Xin, Michael E. Cusick, Troy Moore, Charlie Boone, Michael Snyder, Frederick P. Roth, Albert-László Barabási, Jan Tavernier, David E. Hill, and Marc Vidal. High-quality binary protein interaction map of the yeast interactome network. *Science (New York, N.Y.)*, 322(5898):104–110, October 2008b. ISSN 1095-9203. doi: 10.1126/science.1158684.

Miao Yu, Gary C. Hon, Keith E. Szulwach, Chun-Xiao Song, Liang Zhang, Audrey Kim, Xuekun Li, Qing Dai, Yin Shen, Beomseok Park, Jung-Hyun Min, Peng Jin, Bing Ren, and Chuan He. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell*, 149(6):1368–1380, June 2012b. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2012.04.027.

A. Zemach, I. E. McDaniel, P. Silva, and D. Zilberman. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science*, 328(5980):916–919, May 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1186366.

Assaf Zemach and Daniel Zilberman. Evolution of Eukaryotic DNA Methylation and the Pursuit of Safer Sex. *Current Biology*, 20(17):R780–R785, September 2010. ISSN 0960-9822. doi: 10.1016/j.cub.2010.07.007.

Haikuo Zhang, Xin Zhang, Erin Clark, Michelle Mulcahey, Stephen Huang, and Yujiang Geno Shi. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Research*, 20(12): 1390–1393, December 2010. ISSN 1748-7838. doi: 10.1038/cr.2010.156.

Pu Zhang, Junko Iwasaki-Arai, Hiromi Iwasaki, Maris L. Fenyus, Tajhal Dayaram, Bronwyn M. Owens, Hirokazu Shigematsu, Elena Levantini, Claudia S. Huettner, Julie A. Lekstrom-Himes, Koichi Akashi, and Daniel G. Tenen. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity*, 21(6):853–863, December 2004. ISSN 1074-7613. doi: 10.1016/j.immuni.2004.11.006.

Qiao Zhang, Xiaoguang Liu, Wenqi Gao, Pishun Li, Jingli Hou, Jiwen Li, and Jiemin Wong. Differential Regulation of the Ten-Eleven Translocation (TET) Family of Dioxygenases by O-Linked -N-Acetylglucosamine Transferase (OGT). *Journal of Biological Chemistry*, 289(9):5986–5996, February 2014. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M113.524140.

S. Zhang, Y. Feng, O. Narayan, and L. J. Zhao. Cytoplasmic retention of HIV-1 regulatory protein Vpr by protein-protein interaction with a novel human cytoplasmic protein VprBP. *Gene*, 263(1-2):131–140, January 2001. ISSN 0378-1119. doi: 10.1016/s0378-1119(00)00583-7.

Ting Zhang, Xiaowen Guan, Un Lam Choi, Qiang Dong, Melody M. T. Lam, Jianming Zeng, Jun Xiong, Xianju Wang, Terence C. W. Poon, Hongjie Zhang, Xuanjun Zhang,

Hailin Wang, Ruiyu Xie, Bing Zhu, and Gang Li. Phosphorylation of TET2 by AMPK is indispensable in myogenic differentiation. *Epigenetics & Chromatin*, 12(1):32, 2019. ISSN 1756-8935. doi: 10.1186/s13072-019-0281-x.

Xiaotian Zhang, Jianzhong Su, Mira Jeong, Myunggon Ko, Yun Huang, Hyun Jung Park, Anna Guzman, Yong Lei, Yung-Hsin Huang, Anjana Rao, Wei Li, and Margaret A. Goodell. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nature Genetics*, 48(9):1014–1023, September 2016. ISSN 1546-1718. doi: 10.1038/ng.3610.

Yang W. Zhang, Zhihong Wang, Wenbing Xie, Yi Cai, Limin Xia, Hariharan Easwaran, Jianjun Luo, Ray-Whay Chiu Yen, Yana Li, and Stephen B. Baylin. Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress. *Molecular Cell*, 65(2):323–335, January 2017. ISSN 1097-4164. doi: 10.1016/j.molcel.2016.12.013.

Yong Zhang, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, Richard M. Myers, Myles Brown, Wei Li, and X. Shirley Liu. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9):R137, September 2008. ISSN 1474-760X. doi: 10.1186/gb-2008-9-9-r137.

Zhigang Zhao, Li Chen, Meelad M. Dawlaty, Feng Pan, Ophelia Weeks, Yuan Zhou, Zeng Cao, Hui Shi, Jiapeng Wang, Li Lin, Shi Chen, Weiping Yuan, Zhaohui Qin, Hongyu Ni, Stephen D. Nimer, Feng-Chun Yang, Rudolf Jaenisch, Peng Jin, and Mingjiang Xu. Combined Loss of Tet1 and Tet2 Promotes B Cell, but Not Myeloid Malignancies, in Mice. *Cell Reports*, 13(8):1692–1704, November 2015. ISSN 22111247. doi: 10.1016/j.celrep.2015.10.037.

Rongbin Zheng, Changxin Wan, Shenglin Mei, Qian Qin, Qiu Wu, Hanfei Sun, Chen-Hao Chen, Myles Brown, Xiaoyan Zhang, Clifford A Meyer, and X Shirley Liu. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Research*, 47(D1):D729–D735, January 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gky1094.