Genome-wide epigenetic profiling of B cell leukemia and lymphoma

Mohammad Hamdy Abdelrazak Morsy

Department of Clinical Chemistry and Transfusion Medicine
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2020

Genome-wide epigenetic profiling of B cell leukemia and lymphoma

Mohammad.hamdy@gu.se

ISBN 978-91-7833-734-7 (PRINT)
ISBN 978-91-7833-735-4 (PDF)

Printed in Gothenburg, Sweden 2019
Printed by BrandFactory

Dedicated to my beloved mother, my supportive brother and my father's soul

& to the souls of uncle Magdi and uncle Atef who will always be remembered

Genome-wide epigenetic profiling of B cell leukemia and lymphoma

Mohammad Hamdy Abdelrazak Morsy
Department of Clinical Chemistry and Transfusion Medicine, Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

ABSTRACT

Epigenetic modifications, at the level of DNA methylation and post-translational modifications of histone tails cooperatively function in the organization of the genome, and thereby establish the gene expression profiles, phenotypes, and cellular fates. In this work, we investigated the aberrant epigenome in chronic lymphocytic leukemia (CLL) which is one of the most frequent lymphoid malignancies in the west including the Nordic countries. The overall aim of this work is to address the impact of altered epigenetic patterns in CLL on the disease progression with respect to gene expression profile and gain mechanistic insights on the interplay between the different epigenetic mechanisms, such as DNA methylation and histone modifications, in regulating the expression of CLL signature genes. The first study in this thesis aims to investigate the impact of gene body hypermethylation on transcriptional activation which was not completely understood then. Based on our previous MBD seq data (Methyl-CpG-Binding Domain based next generation Sequencing) datasets on CLL samples, of the top differentially methylated genes in CLL compared to normal B cells, we nominated Ten-eleven translocation (*TET1*) which was shown to harbor hypermethylation at CpG islands within gene body. We found that gene body of *TET1* harbors an overlapping cryptic promoter, the transcript of which attenuates the corresponding gene transcription when unmethylated and its hypermethylation in CLL was found to be associated with the overexpression of *TET1*. The second study aimed at globally mapping the genomic targets of enhancer of zeste homolog 2 (EZH2) the catalytic subunit of Polycomb repressive complex 2 (PRC2) in CLL by chromatin immunoprecipitation followed by sequencing (ChIP-seq) along with its prototypical repressive chromatin feature (H3K27me3). The findings of this study unraveled a non-canonical implication of EZH2 in transcriptional activation apart from PRC2. We show a mechanism by which EZH2 transactivates *IGF1R* gene in the more adverse CLL subgroups with IGHV mutations (mutated CLL) and how it contributes to activating PI3K/AKT pathway through IGF1R signaling. The third project is somehow pertinent to the aforementioned first study and aims at drawing a more detailed mechanistic link between CpG methylation and transcriptional regulation in terms of the residence of PRC2, as it preferentially locates GC-rich elements. Integration of our previous global methylome datasets in CLL patients and transcriptome analysis by RNA-seq after induction of global demethylation in CLL cell lines has revealed a set of genes that are supposedly prone to hypermethylation within their intragenic regions in CLL, and such hypermethyation is found to be positively correlated with their overexpression in CLL. Out of the top significant genes, *MNX1* was selected to probe the mutual exclusivity of PRC2 and intragenic CpG islands and the possible implication of gene body hypermethylation in upregulating *MNX1* in CLL through impeding the PRC2-mediated

repression. Altogether, the findings of our work underscore that aberrant epigenome is more likely to be the niche within which the cancer type-relevant aggressive traits are acquired and might pave the way for further detailed investigations that look forward to improve the therapy options and accordingly the clinical outcomes in CLL.

Keywords: CLL, PRC2, EZH2, Epigenetics, CpG islands, DNA methylation, ChIP-seq, RNA-seq
ISBN 978-91-7833-734-7 (PRINT)
ISBN 978-91-7833-735-4 (PDF)

SAMMANFATTNING PÅ SVENSKA

Förändringar i arvsmassan som leder till ändrat uttryck av gener utan att DNA-sekvensen ändras kallas epigenetik och inkluderar metylering av DNA och post-translationella modifieringar av histonernas svansar. Dessa samverkar med varandra och kan leda till ändrat genuttryck och därigenom också en ändrad fenotyp hos celler, utan att DNA-sekvensen påverkats. I denna studie undersökte vi det avvikande epigenetiska mönster som ses vid kronisk lymfocytisk leukemi (KLL), som är den vanligaste lymfoida neoplasi i västvärlden inklusive de nordiska länderna. Det övergripande syftet med denna studie är att avgöra betydelsen av dessa epigenetiska förändringar och deras relation till förändrat genuttryck och sjukdomsprogress som ses vid KLL samt förstå samspelet mellan olika epigenetiska mekanismer, så som DNA-metylering och modifieringar av histoner, och genuttryck.

Den första studien syftade till att undersöka hypermetylerings betydelse vid aktivering av genen Ten eleven translocation 1 (*TET1*), där våra tidigare resultat från kartläggning av den globala DNA-metylering vid KLL visat att genen var differentiellt metylerad vid jämförelse mellan KLL-celler och normala B-celler. Genen för *TET1* visade sig ha en överlappande kryptisk promotor, vars transkript dämpar transkriptionen av *TET1* när den inte är metylerad.

Den andra studien syftade till att globalt kartlägga vilka gener som Enhancer of zeste homolg 2 (EZH2), den katalytisk enheten i polycomb repressive complex 2 (PRC2), reglerar genom att analysera väl karakteriserade KLL prover genom användning av kromatin-immunprecipitering riktade mot H3K27me3, följt av sekvensering (s.k. ChIP-seq). På detta sätt identifierade vi en PCR2 oberoende icke-kanonisk funktion hos EZH2 vid transkriptionell aktivering. Sammantaget visar våra resultat att EZH2 transaktiverar *IGF1R*-genen vilket bidrar till aktivering av PI3K / AKT-signalvägen i den prognostiskt mer ogynnsamma subgruppen av KLL.

Vårt tredje projekt syftade till att ge en mer detaljerad mekanistisk förklaring mellan CpG-metylering och transkriptionell reglering av *MNX1*-genen och dess roll för bindning av PRC2 till GC-rika element och aktivering av promotorer.

KLL är en obotlig sjukdom men med mycket varierande kliniskt förlopp. Ökad förståelse för de epigenetiska förändringarna som ligger bakom en mer ogynnsam sjukdomsprogression kan bana väg för utveckling av mer riktade behandlingsstrategier med avsikt att reversera dessa förändringar för att förbättra de kliniska resultaten vid KLL.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Pradeep Kumar Kopparapu*, **Mohammad Hamdy Abdelrazak Morsy***,Chandrasekhar Kanduri and Meena Kanduri.Gene-body hypermethylation controlled cryptic promoter and miR26A1-dependent EZH2 regulation of TET1 gene activity in chronic lymphocytic leukemia.Oncotarget ,2017, Sep 29; 8(44): 77595-77608.

* Equally contributing first author.

II. Kosalai ST, **Morsy MHA**, Papakonstantinou N, Mansouri L, Stavroyianni N, Kanduri C, Stamatopoulos K, Rosenquist R, Kanduri M: EZH2 upregulates the PI3K/AKT pathway through IGF1R and MYC in clinically aggressive chronic lymphocytic leukaemia. Epigenetics 2019:1-16.

III. **Mohammad Hamdy Abdelrazak Morsy**, Mohamad Moustafa Ali, Chandrasehkar Kanduri and Meena Kanduri. DNA methylation at intragenic CpG islands controls PRC2-mediated transcriptional regulation of *MNX1* in Chronic lymphocytic leukemia. (Manuscript)

CONTENT

ABBREVIATIONS

5-cC	5-carboxy Cytosine
5-fC	5-formyl Cytosine
5-mC	5-methyl Cytosine
AID	Activation induced deaminase
BER	Base excision repair
CATCH-IT	Covalent attachment of tags to capture histones and identify turnover
CGI	CpG island
CIMP	CpG methylator phenotype
CLL	Chronic lymphocytic leukemia
cPRC1	Canonical polycomb repressive complex1
DNMT	DNA methyl transferase
dsDNA	Double stranded DNA
EED	Embryonic ectoderm development
EZH1	Enhancer of zeste homolog 1
EZH2	Enhancer of zeste homolog 2
GEP	Gene expression profile
GRO-seq	Global Run on-sequencing
HCP	High CGI-promoter
HDAC	Histone deacetylase
HMT	Histone methyl transferase
HP1	Heterochromatin protein 1
HSC	Hematopoietic stem cell
IGHV	Immunoglobulin heavy chain variable gene
JARID2	Jumonji and ARID domain containing protein 2
JmjC	Jumonji C domain
JmjN	Jumonji N domain

LncRNA	Long non-coding RNA
MBD	Methyl Binding Domain
M-CLL	IGHV-mutated CLL
mESC	Mouse embryonic stem cell
MLL	Mixed lineage leukemia
ncPRC1	Non-canonical polycomb repressive complex1
PcG	Polycomb group protein
PCGF	Polycomb group ring-finger domain proteins
PCL	Polycomb like protein
PHD	Plant Homeodomain
PRC1	Polycomb repressive complex
PRC2	Polycomb repressive complex 2
PRE	Polycomb response elements
PWWP	Pro-Trp-Trp-Pro domain
RNAPII	RNA polymerase II
RYBP	Zinc finger and YY1 binding protein
SAM	S-adenosyl methionine
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SNV	single nucleotide variation
SUZ12	Suppressor of zeste 12
SWI/SNF	Switch/ sucrose non-fermentable
TDG	Thymine DNA glycosylase
TET	Ten eleven translocation
TrxG	Trithorax group protein
TSS	Transcriptional start site
U-CLL	IGHV-unmutated CLL
WD40	Tryptophan-aspartate repeat domain

1 INTRODUCTION

The rise of high throughput and massively parallel technologies has revolutionized the field of genomics and epigenomics. This revolution has in turn enhanced understanding of genome organization and the mechanisms orchestrating the interplay between the genome and epigenome towards driving both normal and diseased phenotypes.
The emergence of regulatory genomic elements and the expansion of the haploid genome size in higher organisms entail highly sophisticated mechanisms, so that they govern and regulate the genome function in establishing gene expression profiles in response to either normal cellular conditions or stochastic environmental circumstances. Along with the sequence-based intrinsic features of the genomic domains, epigenetic mechanisms serve an integral part in this regard without altering the genomic sequences. Epigenetic mechanisms involve chemical modification of DNA bases such as cytosine methylation and histone post-translational modifications, which in harmony with long non-coding RNAs (LncRNAs) organize the eukaryotic, particularly the mammalian genome into functionally distinct compartments, thereby control gene expression in a spatial and temporal manner.
Alterations that encounter any of these mechanisms have extensively been reported in many diseases including cancer. Aberrant epigenetic patterns have been shown to adopt the mutational load and encompass many notorious cytogenetic lesions that arise with the progression of malignant diseases, suggesting that cancer is more than a disease of mutations. Altered DNA methylation has been well addressed in cancers and viewed to cooperate with abnormal histone modification pattern to foster the acquisition of the malignant hallmarks.
By means of high through-put sequencing techniques and with integration of current along with previously published global datasets by our team, this work mainly aims at identification of the altered epigenome in terms of DNA methylation and polycomb repressive complex 2 (PRC2) deregulation in chronic lymphocytic leukemia (CLL). In addition, we aimed at dissecting the crosstalk between these epigenetic mechanisms in reprogramming and altering gene expression. More specifically, the main concern of this work is drawing more detailed mechanistic links between DNA methylation and PRC2 in regulating sets of biologically relevant genes that are related to the proliferative pathways such as IGF1R signaling, PI3K/AKT pathway as well as homeobox-related genes and how that contributes to CLL pathogenesis.
Comprehensive understanding of the altered aspects of epigenetic mechanisms in CLL and other cancers in general, along with translational medicine may pave the way for improving and elaborating promising and more efficient lines of therapy and improve the patients' prognosis.

1.1 Genome: The pivotal dimensions of complexity in normal development and diseases

1.1.1 Genome and complexity of living organisms

The last two decades following the completion of Human Genome and ENCODE projects, have witnessed a stunning paradigm shift. It was previously believed that the living organisms are hard-wired by their genes, as implied by the central dogma of molecular biology.[1] The central dogma defines life or a biological phenomenon as a flow of genetic information embedded in the DNA into a messenger RNA (mRNA), which is in turn translated into a biologically functional protein that contributes to manifestation of phenotypes.
Throughout the evolution of the living organisms, a wide range of biological diversity and scalable broad spectrum of organismal complexities are observed from symbiotic bacteria to humans.[2] The intriguing question is that, what stands behind and what confers such diverse biological complexities?
It was believed until the early 1970s of the last century that the organismal complexity is scaled by the cellular contents of DNA (the C-value).[3] The notion that the C-value failed to show a

significantly consistent relationship to biological complexity raised the C-value paradox.[2, 4-7] The C-value paradox was rationally resolved by the fact that the C-value is not more than a crude measure of cellular DNA content and is not normalized to varying degrees of ploidy and the emergence of non-protein coding elements into the genomes of higher organisms.[8, 9]
Another intriguing paradox in molecular biology is the inconsistent relationship between developmental complexity and the number of protein-coding genes, referred to as the G-value paradox.[10] According to the central dogma, one would expect that the organismal complexity is supposedly scaled up by the increase of the protein-coding sequences. However, one striking example amidst myriads, has casted a remarkable skepticism on this tenet; the mean number of protein-coding genes in human and in the microscopic nematode *C.elegans* is 20,000 genes.[2] Moreover, the human genome project revealed that the sequences of DNA that are synonymous with functional proteins represent less than 2% of the whole haploid genome, while the rest was termed as "Junk DNA" or "Selfish DNA".[8, 9] After the revolution in the field of genomics, the Junk DNA turned out to be biologically relevant to the development of the higher organisms. This 98% of the human genome is broadly partitioned into *cis*-acting regulatory elements and non-protein-coding sequence that are not inert and code for *trans*-acting non-coding RNAs (ncRNAs), which in turn fulfil a broad spectrum of cellular and developmental functions. Together, these elements cooperatively contribute to the establishment of the highly sophisticated architecture and dynamic organization of the genome, and thus control gene expression in a spatial and temporal manner and account for the gene product diversity throughout the multicellular ontogeny.[2]
The non-random overrepresentation of non-genic elements including introns and intergenic elements would evoke our thinking of the mechanisms beyond the intrinsic sequence properties and the information stored in our genomes, and how these mechanisms shape our genomes and accordingly control the flow of genetic information towards a stable phenotype.[2, 11] These mechanisms are known as epigenetic mechanisms which control, influence and get influenced by the genome without altering the DNA sequence, and thus account for cellular memory upon reaching the terminal stage of differentiation.[12]
By virtue of next generation sequencing and massively-parallel technologies, our comprehension of biological and physiological processes and the development of diseases such as cancer, has experienced a transition from a gene-centered to a genome-wide approach. Also, as best expressed by Adrian Bird[12], the revolution in the field of epigenetics has offered an antidote to the tenet that a living organism is equal to the summation of its genes; an organism is rather a function of its genome.

1.1.2 The problem of cancer

During the multistep process of cancer development from normal cellular incipient, new biological capabilities are acquired, following the cumulative disruptions to normal growth controls. These new traits comprise the hallmarks of cancer and confer the survival and aggressiveness of cancer through providing the greatest clonal advantages, that in turn qualify the progression of the neoplastic cells that are ultimately becoming malignant tumor.[13] These hallmarks were first introduced in the year 2000 by Douglas Hanahan and Robert A.Weinberg.[14] This proposition constituted an organizing principle and a conceptual framework for understanding the diversity and complexity of human neoplasms.[15]
These hallmarks include six biological traits:

1. Sustainable proliferative signaling
2. Evading growth suppressors
3. Resistance to cellular death
4. Enabling replicative immortality
5. Induction of angiogenesis
6. Activation of invasion and metastasis

A decade later and with the conceptual progress and accumulation of a wealth of knowledge, Douglas and Weinberg introduced an ancillary proposition which implied additional two hallmark traits. Reprogramming of energy metabolism and evasion of immune-mediated destruction are two emerging capabilities have been added to the list of cancer hallmarks.
The fairly intriguing question is: **what underlies the acquisition, preservation and fostering of these traits? And which physiological niche would embrace the clonal selection of the competent neoplastic cells?**
Genomic instabilities and inflammation underlie these hallmarks through generating genetic diversity, thereby expedite their acquisition and foster their functions. Both are considered as enabling characteristics that account for the Darwinian selection of these new traits, thus provide the greatest clonal advantages for survival and generation of macroscopic tumors. The dimensions of complexity of cancers is superimposed by their ability to recruit apparently normal cells that help in establishing the so called tumor microenvironment which constitutes as a physiological niche encompassing the acquisition and contribution of the hallmarks traits in tumor progression.[16, 17]
Throughout the course of tumorigenesis, the cells destined to become malignant experience successions of drastic conditions elicited by the natural barriers which are hard wired into the cells to impede outgrowth of either pre-neoplasms or frank neoplastic lesions and functions in various combinations of tumor suppressive modes of action.[15] Neoplastic cells relentlessly invest the malignant capabilities they have been acquiring to circumvent such anticancer defense activities. Therefore, the developing neoplastic cells can overcome such bottle neck by compromising the gene expression profile (GEP) for the tumor advantage.[18, 19]
By means of epigenetic mechanisms, the expression of the components of surveillance system that ensure genetic and genomic integrity is altered [20-22]. Most importantly gate-keepers and care-takers that evoke genomic maintenance and DNA repair or otherwise trigger senescence and/or apoptosis of genetically damaged cells are compromised; thereby cancer increases the mutational load and guarantees the accumulation of genetic aberration and cytogenetic lesions that are well suited as vehicles of persistent phenotypical oncogenic changes and candidates for clonal selection.
The dynamic reprogramming of gene expression profile and the shaping of malignant phenotype is a culminate result of cooperative integration between altered genetic and epigenetic mechanisms like methylation of DNA or post-translational modification of histones that foster acquisition of hallmark traits.[13, 23, 24] Here comes the significance of epigenetic mechanism in providing alternative and flexible, yet persistent ways to regulate and acquire stable oncogenic traits throughout multiple cellular division cycles that would potentially boost the neoplastic phenotypes.[25]

1.1.3 Cancer epigenome

During the last century, cancer has generally been viewed as a genetic disease; the general tenet that was adopted then implied that mutations are the driving forces that initiate the neoplastic transformation.[13] However, it has become clear that rate of mutations, copy number alterations and insertion/deletions take place at relatively low frequencies; thus rendering them inefficient means of initially driving neoplastic transformation.
Considering the local influence of base composition on single nucleotide variation (SNV), chromatin structure, replication timing and regional effects of sequence composition and mutation rates across the genome vary markedly.[26] Notably, the mutational load in cancer can by no means be inferred directly from the number or frequency of observed mutations, without taking the number of cell divisions that have occurred or the influence of epigenetic mechanism on the rate of generation and repair of such altered genetic lesions into account.
Epigenetic mechanisms have an influential role in mutational rate in several ways. Most importantly, certain subsets of colorectal cancers and glioblastomas show epigenetically-addictive phenotypes, wherein the incipient pre-neoplastic cells exhibit a predisposition towards

exceptionally high frequency of cytosine methylation at CpG-rich promoter of certain genes the silencing of which is of a molecular and biological relevance to the respective malignancies. This phenomenon is referred to as CpG island methylator phenotype (CIMP) [27, 28], where the exceeding frequency of promoter methylation is neither stochastic nor spontaneous, but instead an epigenetically attributed and highly coordinated event. Such predisposition towards epigenetic addicted phenotypes may be a consequence of germline variation that increases the likelihood of cancer development.[29, 30] Another facet of epigenetic addition is exhibited in aberrant chromatin organization that is attributed to gain-of function of chromatin modifiers as evident in Polycomb repressive complex addiction in diffuse large B cell lymphoma (DLBCL) [31-33] and DOT1L-addiction in subsets of mixed lineage leukemia (MLL) which is associated to altered regulation of Homoebox-related genes.[34-36]

Also, cytosine methylation in the context of CIMP has been shown to play a crucial role in the boosting the rate of C-to-T mutation within CpG islands up to tens of folds of magnitudes.[26] The C-to-T mutations during DNA replication in highly proliferative cells result in T:A substitution lesion which are not recognized by the DNA repair machinery. These kind epigenetically-attributed mutations constitute about 25% of all *TP53* mutations in human cancers.[37]

Altogether, the aforementioned facts support the contribution of epigenetics to cancer mutational load and that epigenetic alterations might act upstream to foster the acquisition of the cytogenetic lesions throughout the course of tumorigenesis and maintain them to boost the malignant phenotypes and achieve the best clonal traits.[13, 38]

It has become evident that both genome and epigenome influence each other as a mean of high-fidelity and tight control of gene expression profile and fate specification. The next chapter explains the different epigenetic mechanisms at the levels both DNA and histone modification, the regulation of epigenetic marks at the levels of writing and erasure, and the interaction between the different epigenetic patterns and show that they are not operating in isolation, but rather within a an integrative network of mutually exclusive, reinforcing and counteracting signals.

1.2 Epigenetics: the flexible mediators of genome dynamicity

1.2.1 Epigenetics at a glance

The term "Epigenetics" was first coined in 1939 by Conrad Hal Waddington, nevertheless, so many definitions have been suggested for epigenetics thereafter.[12] According to Waddington, and apart from classical genetics, epigenetics concerns about how genotypes manifest phenotypes. In his own words, epigenetics is defined as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being".[39, 40]
The definition of epigenetics remained subject of debate, until the mid-1970s, when the field of epigenetics was revived once again by Arthur Riggs and coworkers.[41] They could reformulate the definition and reintroduce the biological importance of epigenetics, so that the term refers to the study of mitotically and/or meiotically heritable changes in genetic functions beyond sequence alteration of DNA itself. More specifically, epigenetics describes the interplay between the reversible chemical modifications of histone tails and/or DNA without changing the genetic information, and chromatin-associated proteins, in regulating chromatin structure and transcriptional programs of the cells.[42]
In brief, epigenetics is inheritance, but not as we know. It is suggested to be described as "the soul" of the genome that organizes its function in response to the environmental conditions or developmental demands, thereby coins the cellular identity and guarantees the phenotypic distinctiveness of each type of the genetically identical cells that comprise a multicellular organism.

1.2.2 Epigenetics: what stands behind genome's function?

The succession of the technological advances in the post-sequencing era and the rise of the high throughput technologies have revolutionized the field of epigenetics and offered several genome-wide-based lines of investigations. By virtue of such technological revolution, the functions and distribution of several epigenetic marks over the genome have been accurately mapped and comprehensively understood. The different epigenetic modifications at the levels of DNA and histones modification are considered as surrogates of the functional genomic elements, structural organization of chromatin domains and transcriptional potentials across the chromatin domains.[43]
As explained in section1.1, the evolution of the higher organisms had experienced genomic size expansion, as a consequence of the emergence of non-protein-coding elements including repetitive elements which are mainly transposons-derived, intergenic regulatory elements and intragenic introns. This latter set of elements largely accounts for the expansion of eukaryotic proteomes by alternative splicing.[44] Such complexity of the eukaryotic genomes, in particular the mammalian genomes, entails highly sophisticated patterns of organization, that aim at not only structural packaging of the genome within the nuclear vicinity, but also functional arraying of the genomic elements and delineating the distinct chromatin domains. The mammalian genome is viewed as a series of superimposed organizational layers rooted in the double stranded DNA (dsDNA).[13, 45, 46] Approximately 147 bp of DNA is wrapped twice around histone octamer that is composed of dimers of H2A, H2B, H3 and H4 core histones, comprising a nucleosome, the building unit of chromatin.[47-49]
The chromatin is compartmentalized into structurally and functionally distinct domains that are established and demarcated by epigenetic modifications. These domains are principally classified into euchromatin regions, the conformation of which is accessible to the transcriptional machinery, and heterochromatin, which is on the contrary, closed and is transcriptionally inactive.[50] This latter is further classified based on several factors, most importantly the intrinsic sequence-based criteria and their replicative timing of the genomic sequences that

comprise these domains, into facultative heterochromatin and constitutive heterochromatin.[51-53]
This organization is highly context-dependent, in the sense that chromatin is not static, but rather dynamically varies across the cellular conditions, either during development and lineage specification, or upon cellular reprogramming and neoplastic transformation.[13, 43] Also, the higher order of chromatin organization is cell-type specific and is established by interplay between the genomic features and epigenetic machinery. Thus, epigenetic patterns elicit a "memory" for the cell, that they mediate and implement the decisions that has been taken by the cell during development from totipotent towards its terminally differentiated state, through shaping the gene expression profile of the cell and manifesting the unique phenotype. Perhaps, the key fundamental facet of epigenetic patterns is that they themselves are faithfully maintained throughout the successive cell division cycles, and in turn reciprocate by stabilizing the phenotype of the terminally differentiated cell and maintaining its identity.[54] Accordingly, the cellular epigenome seems more likely to be the "mind" of the cell.

1.2.3 Epigenetic patterns establishment

Establishing the epigenome demands reversible chemical modification either at the DNA level; most famously, DNA methylation which is achieved by adding methyl group on the fifth carbon atom of cytosine bases forming 5-methyl cytosine (5-mC), or by post translational modification of histone proteins by acetylation, phosphorylation, sumoylation, ubiquitinylation or most prominently methylation. Such chemical modifications are referred to as "epigenetic marks". These marks are either surrogates of open chromatin structures and active transcription, or prototypes of closed conformation and diminished transcriptional potential. Each of these marks is regulated by counterbalancing activities of "writers" and "erasers". The writers catalyze the establishment of the chemical modification at DNA or histones levels, while the erasers catalyze the removal of the epigenetic mark.[55] For example, a histone lysine methylation mark is catalyzed by a dedicated histone methyl transferase and erased by a demethylase. The balance between writers and erasers is tightly controlled based on the cellular conditions and developmental needs. As pointed out above, epigenetic patterns are not insulated from each other and they are operating within an integrative network governed by a collaboration of multiple regulatory mediators including:

1. **Readers:** Readers are proteins with featured domains by which they recognize specific epigenetic marks. These domains include PHD fingers, WD40, Ankyrin, PWWP, MBT, Chromodomains, Tudor domains and others.[55] Upon recognizing a certain epigenetic feature, readers can either convey signals to downstream effectors that implement a dedicated epigenetic function in establishing local chromatin structure (e.g. ATP-dependent chromatin remodelers), or recruit other epigenetic regulators and moderate the crosstalk between the mutually reinforcing epigenetic machineries. Also, some readers of certain epigenetic marks function as erasers of another antagonizing epigenetic feature. For example, PHF8 reads the active mark H3K4me3 via its PHD domain in the euchromatic contexts, and under certain circumstances it serves in erasure of the repressive H3K27me3 mark from the same context by its Jumonji C (JmjC) domain.[56, 57]
2. **Long non-coding RNAs (LncRNAs):** LncRNAs are also known as "initiators" that serve in recognizing the chromatin status in terms of conformation and transcriptional activity and recruiting the epigenetic machineries to their targets in a spatial and temporal manner. Many lncRNAs have been shown to recruit polycomb repressive complex 2 (PRC2) to their targets over the genome.[58-61]
3. **Genomic sequence:** DNA sequences in part, play a role for directing and shaping epigenetic patterns. Among several examples, DNA methylation is highly dependent on and influenced by the genomic context. For instance, CpG islands (CGIs) are by default unmethylated and are rendered protected from methylation, in spite of the

methylation attrition, by several factors, most prominently, the strand asymmetry at the GC-rich regions which is in turn associated with R-loop formation.[62]
On the other hand, CpG dinucleotides in the context of CpG oceans are more prone to methylation, as explained in details below (section1.2.4).[63-65] It is also noteworthy that the functionally distinct chromatin domains that are studded by counteracting epigenetic marks are delineated, in a way that prevents the spreading of the repressive regions or the firing of unintended promoters or enhancers activities. This is mainly accomplished by the binding of CTCF proteins to insulator elements, so that the different chromatin domains are stably demarcated.[66-68]

4. **Nucleosome positioning and turnover:** Nucleosome turnover has been suggested to play a role in regulating the pattern of histone methylation, in terms of influencing the processive kinetics of the corresponding histone methyl transferases and the valence of methylation mark on histone lysine residues. For instance H3K79 methylation is known to be written by DOT1L, while no dedicated eraser has been characterized so far. Recently, it was found that the rate of nucleosomal turnover preferentially associate with the lower states of H3K79 methylation valence and in turn influence the dynamicity of the local chromatin and its transcriptional potential as well.[69] Also, it was shown that the rate of nucleosomal turnover affect the inheritance of repressive chromatin regions that are marked by H3K9me3, in a way that entails suppression of the turnover rates by SMARCAD1 of the SNF2 family proteins to faithfully maintain the transmission of such epigenetic feature to the progeny cell.[70] In brief, the rate of nucleosome turnover is likely to be negatively correlated with suppressive epigenetic features, while directly proportional to active chromatin marks.
5. **DNA replication timing:** The maintenance and faithful copying of epigenetic patterns is influenced by the replication timing, in the sense that the epigenetic features on the early replicated regions are rapidly restored on the daughter DNA strand compared to the late replicating regions that are mainly decorated with repressive chromatin marks like H3K9/27 methylation which might take longer time that exceeds the time of the cell division, so that the restoration completes after the entry into G1 phase of the next cycle.[54, 71-73]

1.2.4 DNA methylation

DNA methylation has been viewed as an epigenetic prototype of transcriptional silencing that characterizes long-term stable repression.[74] In contrast to histone modification, DNA methylation is less dynamic and not easily reversed, consistent to its implication in long-range silencing that is most evident in X-chromosome inactivation and genomic imprinting.[75, 76]
Methylation of cytosine bases within the genomic DNA is the most prominent methylation mark that has been a matter of deep interrogation, and the succession of the findings that have come out since the mid-1970s raised remarkable debate regarding the relationship between DNA methylation and transcriptional silencing.[41, 77]
To accurately describe the functional link between DNA methylation and transcriptional regulation, the distribution of methylation across the different genomic elements and its location within the transcriptional unit should be considered.[64, 74, 78] DNA methylation takes place to modify either Cytosine or Guanine residues yielding 5-methyl Cytosine (5-mC or O6-methyl Guanine, respectively.[79, 80]
I hereby in this thesis focus on the methylation at cytosine and its implication in establishing chromatin structure and regulation of gene expression.

DNA methylation machinery: Writing, reading and erasing

The deposition of a methyl group at the fifth carbon atom of cytosine bases is catalyzed by a set of evolutionarily conserved DNA methyl transferases (DNMTs).[81, 82] More tellingly, the DNA methylation patterns are established *de novo* during the early embryonic stages by DNMT3A, DNMT3B, and a non-catalytic DNMT3L, the function of which is to recognize chromatin status in terms of histone modification (discussed below) and accordingly direct the *de novo* methyl transferases DNMT3A/B to the target genomic sequence.[83, 84] On the other hand, the maintenance of DNA methylation is carried out principally by DNMT1 along with the E3 ubiquitin-protein ligase (UHRF1).[85] Together, DNMT1 and UHRF1 associate with the replication machinery at the hemi-methylated parental DNA strand and faithfully transmit the DNA methylation pattern to the newly synthesized DNA.[86, 87] It was previously believed that DNMT1 alone can copy DNA methylation pattern and pass it across the successive cellular generations. However, it turned out that the maintenance of DNA methylation upon cellular division is cooperatively achieved by both the de novo DNMT3A/B and DNMT1.[78, 88, 89]
The implication of DNA methylation in establishing chromatin conformation and regulating transcription cannot be explained by the conviction that the presence of 5-mC *per se* controls the access of either transcriptional machinery or chromatin modifiers. One would rather consider the crosstalk between such an epigenetic mark with the rest of epigenetic mechanisms including histone modifiers and chromatin remodelers. Here comes the importance of the reader proteins that recognize methylated CpG sites and accordingly mediate the function of DNA methylation. Two families of proteins read methylated CpGs, namely, the MBD family (MeCP2 and MBD1-4) and the BTB/POZ zinc finger containing family (ZBTB 4, 33 and 38 and Kaiso).[90, 91] These reader proteins implement the crosstalk between DNA methylation and histone modifications that are known to establish local or long-range heterochromatin structure. For instance, upon reading 5-mC in the CpG contexts, MeCP2 recruits Suv39h1/2 which in turn catalyzes the methylation on lysine 9 on histone H3 tail, thus influences heterochromatinization and reinforces silencing.[92-94]
Despite the fact that DNA methylation is stable and not easily reversed, DNA methylation can be erased either passively during DNA replication or actively by dedicated demethylases including Ten Elven translocation family (TET1-3), activation induced deaminase (AID) and thymine DNA glycosylase (TDG).[95-97] The TETs belong to the Fe^{2+} and α-ketoglutarate (α-KG)-dependent dioxygenases, they catalyzes oxidative deamination of 5-mC, yielding intermediate products starting from 5-hydroxymethyl cytosine (5-hmC), which is further oxidized into 5-formyl cytosine (5-fC) and eventually into 5-carboxy cytosine (5-cC). Both 5-fC and 5-cC are subject to excision by TDG giving rise to (abasic) positions at the sites that were marked by methylation. Base excision repair (BER) is then elicited to restore cytosine residues in their original positions.[97, 98]

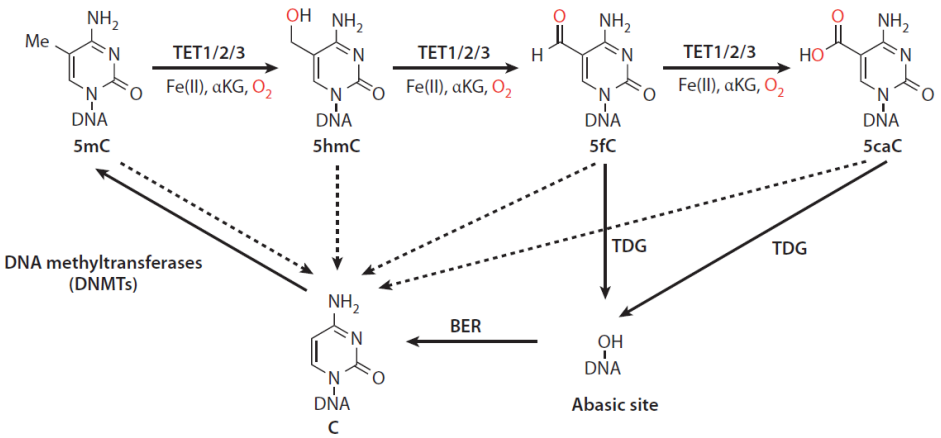

Figure 1. Integrative mode of action of TETs and TDG in the erasure of 5-mC by oxidative deamination. Shen L et al, Annu Rev Biochem, p588. The figure is modified and taken from[98]

Establishment of DNA methylation pattern is not merely a culminate consequence of the balance between the DNMTs and the TETs. It is rather governed by other factors. Perhaps, the bimodal distribution of DNA methylation at CpG sites across the different genomic context suggests an impact of intrinsic sequence-based features on setting the global pattern of DNA methylation.[64, 86, 99] Moreover, with the attrition of indiscriminate methylation during the early stages of embryonic development, CpG dinucleotides in the context of CpG islands (CGIs) show diminished degrees of methylation, if any; which would support the aforementioned suggestion that DNA methylation, in the sense of its establishment and function, is first and foremost dependent on the genomic context. (Explained in the next section)

Establishing bimodal pattern of DNA methylation:

During the pre-implantation stage of the mammalian embryo's life, the genome is subjected to a massive DNA methylation at CpG dinucleotides. Upon implantation, the differential DNA methylation pattern is established over the genome, by two counteracting waves: [63, 64, 74, 99]

- An indiscriminate surge of *de novo* methylation of the majority, if not all CpG dinucleotides, that are sparsely distributed in a context called CpG oceans.
- A mechanism comprised of a group of reinforcing loops that keep CpG dinucleotides within the context of CpG islands (CGIs) unmethylated. The protection of CGIs from methylation could be a consequence of either spontaneous or enzymatic demethylation. However, the shared features of CGIs better explain this predominant deficiency of methylated CpG dinucleotides within CGIs context.

As pointed out above, it is very unlikely to accurately explain DNA methylation in terms of writers/erasers balance, without considering the genomic targets for methylation and how their intrinsic characteristics direct the bimodal DNA methylation profiles. Since DNA methylation takes place on CpG dinucleotides, it is of a prime importance to understand the distribution of these CpG dinucleotides over the mammalian genome, and to hyperlink this understanding to DNA methylation.

The majority of the mammalian genomic elements are deprived of CpG dinucleotides; most probably because they are subject to intensive methylation during the early primordial stages of embryonic development. As a consequence of that, the majority of the methylated cytosines in CpG context undergo spontaneous deamination into thymine, thus explains the lack of CpG dinucleotides over the mammalian genomes.[64] The persistent CpG dinucleotides are scattered within long genomic elements (CpG oceans). On the other hand, CpG islands are relatively short (up to 1 KB), interspersed and both CpG and GC-rich genomic elements, that are punctuated by the CpG oceans. Such genomic elements harbor some intrinsic criteria that supposedly stand behind the predominant lack of methylation within their vicinities and place them apart from the rest of the bulky genomic elements. CGIs share general sequence-based features, in spite of their sequence heterogeneity; they are adapted for a transcriptionally permissive and nucleosome-depleted chromatin structure.[100, 101] It is suggested that the open structure of CGI regions is by default and is transcription-independent, evidenced by the finding that CGIs showed reluctance to assemble into nucleosomes in vitro.[102, 103] Consistently, CGIs have been shown to be well suited for promoter activity, even if they are remotely located with respect to any of the currently annotated promoters.[76]

In embryonic stem cells and during pluripotency, CGIs have been shown to recruit transcription factors and RNA polymerase II (RNAPII), consistent to the fact that the genome of pluripotent cells is more permissive.[104] RNA polymerase then recruits the writers of the active histone mark H3K4me3, for example SETD1A/B, which in turn catalyzes the deposition of such histone mark that features active transcription. The CGIs are almost nucleosome depleted, yet flanked by nucleosomes that are embellished by H3K4me3. Methylation at any level of valence on H3K4 impedes the binding of DNMT3L to the target CGI, thus prohibiting the assembly of the *de novo* methylation machinery that is comprised of DNTM3A/B and the non-catalytic DNMT3L.[105-107] This in part explains the diminished level of CpG methylation within the CGI contexts and the bimodal patterns of DNA methylation. (Figure.2)

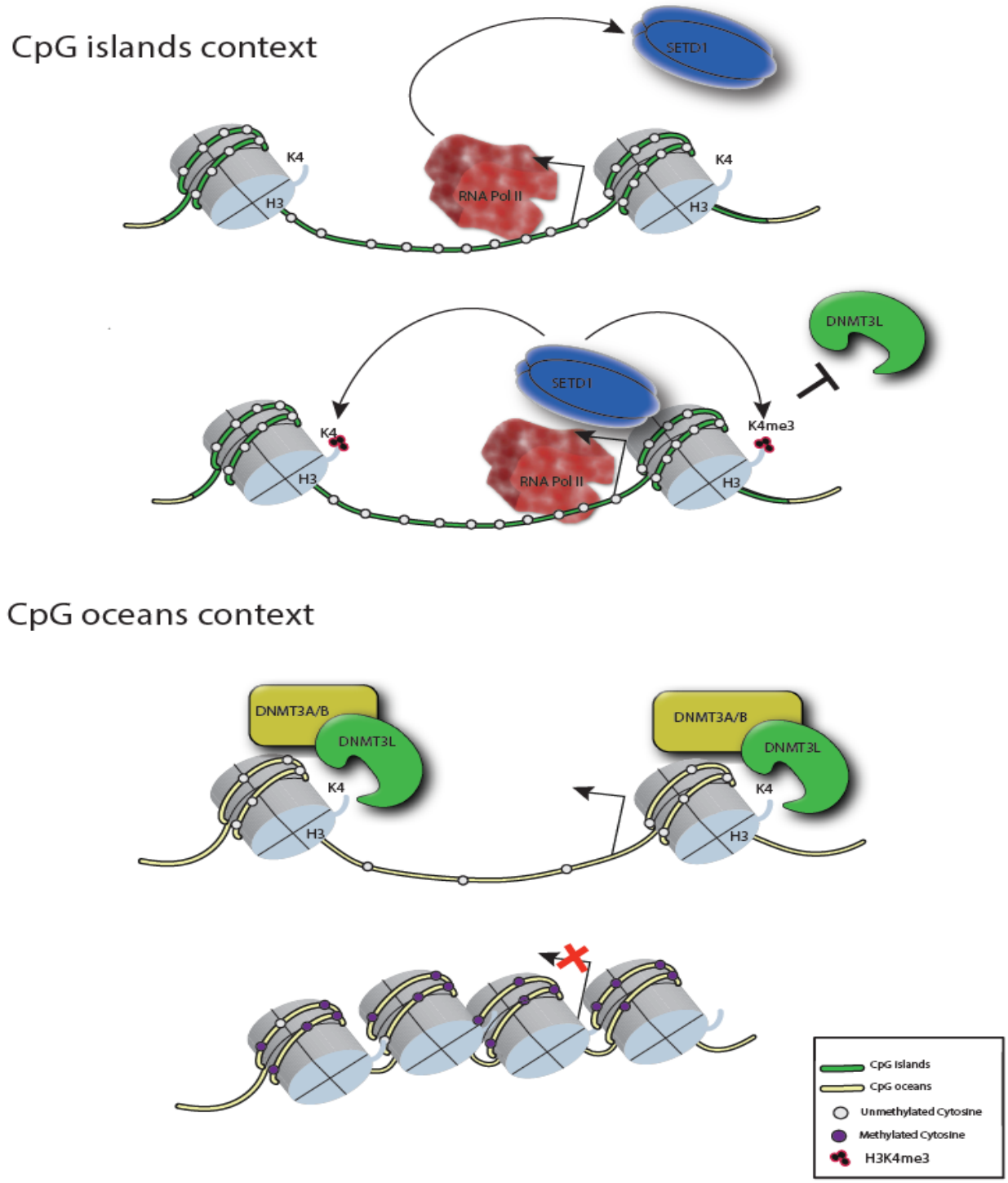

Figure 2. *The establishment of bimodal DNA methylation pattern during pluripotency in CpG islands (top panel) and in CpG ocean context (bottom panel).*

DNA methylation and transcriptional regulation:

DNA methylation as an epigenetic mark was first introduced in 1948 by R.Hotchiss; however, the functions of DNA methylation was kind of obscure until the mid-1970s. The year 1975 witnessed two featured and independent publications by Arthur Riggs and Ryan Holliday; these key papers

suggested for the first time the direct implication of DNA methylation at CpG context in transcriptional silencing.[41, 77]
What has been proposed by the 1970s epigenesists, has actually offered a principal conceptual frame for our understanding of epigenetic mechanisms. Nevertheless some of the concepts that were adopted that time became debatable. For instance, the relationship between DNA methylation and transcriptional regulation has raised remarkable controversies.
Since the distribution of CpG dinucleotides and CGIs is bimodal, it is plausible enough to think of the influence of such genomic elements on DNA methylation and how the distribution of DNA methylation itself in the transcriptional unit affects gene expression.
About 50% of the GCIs in human genome coincide with transcriptional start sites (TSS), while the rest is distributed between intergenic and intragenic regions.[64] The annotated promoters are classified generally into high-CGIs promoters (HCP) which accounts for 70% of human promoters and the rest is comprised of low to intermediate CGI contents. The former set of promoters is annotated to crucial genes that code for proteins that are involved in development and lineage specification, and require tightly controlled modes of regulation.[78]
It is presumable that methylation at CGIs within TSS is associated to silencing of transcriptional initiation. However, the timing of DNA methylation with respect silencing is disputable and has not been clearly affirmed. Here is raised the intriguing question below:

Does DNA methylation precede or pursue silencing of transcriptional initiation?

Notably, based on the perceptions of how *de novo* methylation is established, it has been realized that nucleosomal DNA is a perfect substrate for the *de novo* methylation machinery comprised of DNMTs. Moreover, the *de novo* DNMTs are influenced by the epigenetic marks decorating the target nucleosomes; for instance DNMT3L binds to DNA wrapped on nucleosome lacking H3K4me3(Figure.2), emphasizing the adversarial relationship between DNA methylation and gene expression.
This would raise the speculation that DNA methylation directly induces silencing of transcription. DNA methylation is rather supposed to be an additional regulatory layer that reinforces the long-term silencing, evidenced by the finding that the promoters of the genes on the inactive X chromosomes become methylated after they are already silenced.[108]
It was found that during differentiation, the pluripotency-related genes that are destined for silencing are silenced by certain repressors which in turn recruit a complex comprised of G9a, LSD1 and HDAC1 (Figure.3). Each component in this complex contribute to local heterochromatinization that is eventually mediated by heterochromatin protein (HP1), then *de novo* DNMTs are recruited to establish DNA methylation that serves as a “Lock” to reinforce heterochromatinization, and long-term silencing.[109]
Nevertheless, the findings that knocking *de novo* DNMTS out is associated with impaired differentiation of hematopoietic stem cells suggest the indispensability of DNA methylation for silencing pluripotency-associated genes and proper development and that in turn favors the postulation that DNA methylation acts upstream to silencing of transcriptional initiation.
Bottom line, the relationship between DNA methylation and silencing seems to be bi-directional, it can proceed in either direction; however, no universal conceptual conclusion has come to affirm the issue of the timing of methylation with respect to silencing so far.

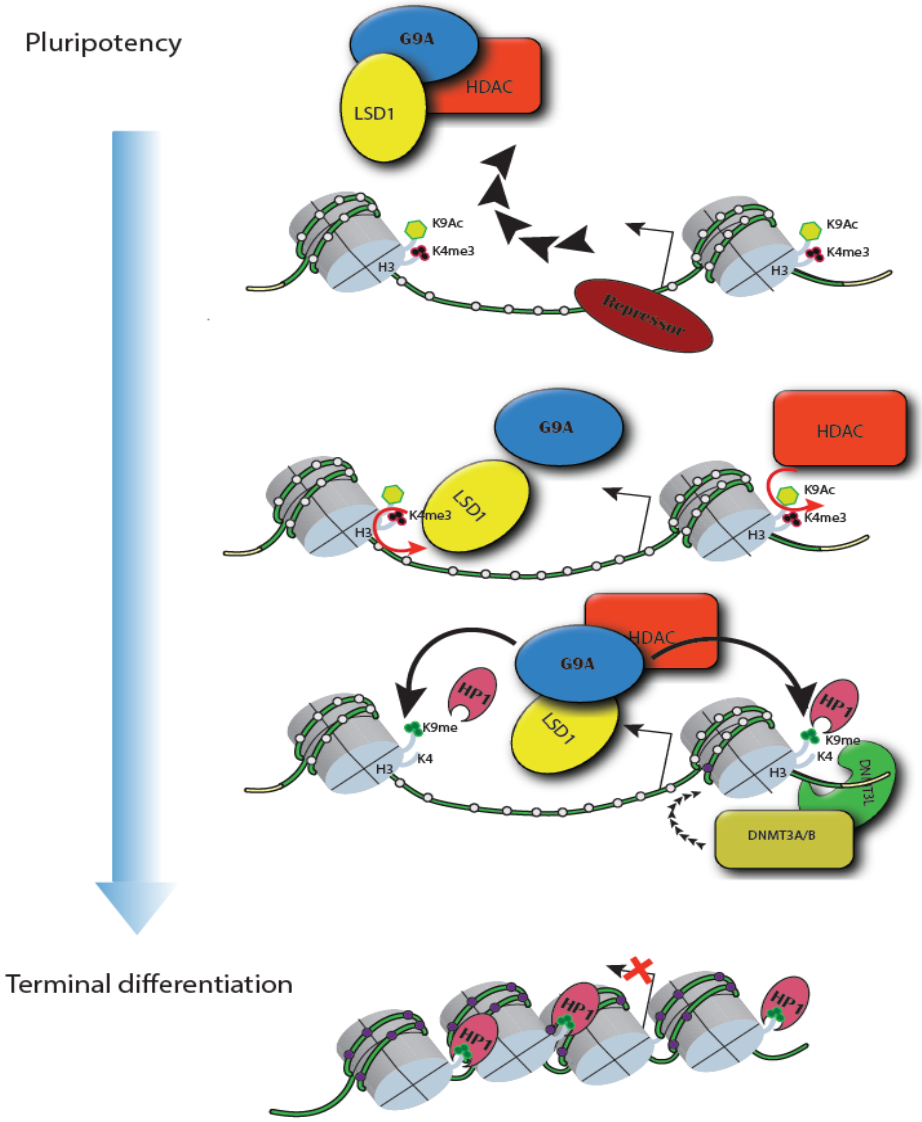

Figure 3. DNA methylation in silencing of pluripotency-genes

Does DNA methylation involve in active gene expression?

Interestingly, DNA methylation in the intragenic CGIs (also called Orphan GCIs) has been shown to be associated with active gene expression. As mentioned above, the location of methylated CpG dinucleotides in the transcriptional unit is a determining factor that directs the functionality of DNA methylation. Consistent to the bimodal distribution of CGIs, methylation of CpG in their vicinities show a differential behavior regarding the regulation of gene expression.
Several functional lines have explained how CpG methylation marks the gene body regions of actively transcribed genes.[110-113] Perhaps, the most prominent mechanism is that intragenic CpG methylation enhances transcriptional elongation through inhibiting the cryptic intragenic transcription that overlap with the corresponding transcription.[114, 115] In line with the notion that CGIs are adapted for promoter activity, integrative analyses of Global Run on (GRO-seq) and transcriptome datasets have characterized many intragenic CGIs that code for non-coding or anti-sense transcripts. These transcripts have been shown to attenuate the transcriptional elongation of the corresponding protein coding transcript.[116, 117] It is most likely that methylation of these cryptic promoters, functions in the suppression of their transcriptional potentials, thereby boosting the elongation of the corresponding gene transcription. In line with this notion, it was found that

the recruitment of DNMT3 at gene bodies is dependent on SETD2-mediated H3K36me3 deposition, the intragenic mark of active transcription.[118, 119]
Thus, despite the universality of the *de novo* methylation enzymes, the mechanism of establishing DNA methylation, in the sense of the factors that recruit DNMTs, is differential between promoter/TSS and gene bodies. This suggests that DNA methylation is not functioning separately from the genomic contexts and other epigenetic mechanisms, and casts doubts on the hypothesis that DNA methylation is the direct cause of transcriptional silencing, or transactivation.

DNA methylation in Cancer:

Altered methylome is considered as one of the most notorious features of human malignant diseases. Aberrant DNA methylation pattern has been found to contribute to the neoplastic transformation, as it is followed by altered gene expression profile which in turn expedites the acquisition of cancer hallmarks, boost mutational rate, and implement the phenotypic transformation of incipient cell into neoplastic cell with the greatest clonal advantages and eventually form a macroscopic tumor.
In most solid as well as hematopoietic malignancies, a global hypomethylation, in particular in the intergenic CGIs has been observed.[120-124] In addition to global loss of methylation, regional hypermethylation is also a feature of human neoplasms. The exceptionally high frequency of hypermethylation at certain CpG-rich promoters of tumor suppressor genes in cancer is not taking place spontaneously or stochastically as was thought previously. This has turned out to be a rather highly coordinated and epigenetically attributed event referred to as CpG islands methylator phenotype (CIMP).[28] CIMP is a phenomenon of an epigenetically addictive phenotype, where regional hypermethylation at certain CpG contexts has been found to be of a significant biological relevance to certain subsets of colorectal cancer and glioblastoma; meaning that the genes that are prone to hypermethylation are not the same in all cancer types or subsets.[27, 119, 125-127]
Recently, Pan-cancer analysis including whole-genome bisulfite sequencing data from normal tissues and their respective solid tumors showed a strong correlation between gene body hypermethylation and increased expression of 43% of homeobox genes in several cancers, supporting the idea that aberrant DNA methylation in cancer is not random.[128] The integrative Pan-cancer analysis revealed that a subset of under-methylated regions (> 3.5 kb) in normal tissues and are mostly enriched in proto-oncogenes involved in transcriptional regulation and control of differentiation. These regions are called (Methylation Canyons) and are prone to hypermethylation in cancers; in particular at the intragenic locations of signature oncogenes, thereby increase their expression in cancer.[128-130]
It was found that certain sets of genes in embryonic stem cells that are silenced by PRC2-mediated H3K27 trimethylation during pluripotency are prone to methylation in a cancer-type dependent manner.[131-133] Upon normal differentiation and lineage specification, PRC2 is released from the promoters CGIs, so that these genes are actively transcribed. Interestingly, with the development of neoplasms, the regions that were pre-marked with PRC2 during pluripotency become methylated and silenced[133, 134]; indicating that DNA methylation mirrors the pattern of PRC2-mediated silencing in stem cells, and thus establishes a state of pseudo-pluripotency that favors the unrestrained proliferation and cellular plasticity.[132, 134-136]
In conclusion, DNA methylation, in terms of its functions, establishment and crosstalk to other epigenetic mechanisms, is highly influenced by genomic context, as well as both normal cellular conditions and disease types, and thus reciprocates to cellular programming in either situation.

1.2.5 Histone modifications

Another facet of epigenetic mechanisms is the post-translational modification of certain amino acid residues on histone tails, most prominent candidate is lysine modification. Several chemical modifications of histones take place and involve in establishing structurally and functionally distinct chromatin domains.[43] Notably, post-translational modification is accomplished by adding a group or moiety to the side chain of the amino acid residue. These moieties could be phosphate, sumoyl, ubiquitinyl, acetyl, or methyl groups.[137-140] Similar to DNA methylation, a histone modification is established by a "writer" and removed by an "eraser" and is also recognized by a reader via a special domain and function in recruiting downstream effectors, thus local chromatin features are established.[55, 141] Hereby, I focus mainly on histone lysine methylation marks, in particular, methylation of lysine 27 on histone 3 (H3K27me) the repressive prototype of PRC2.

What makes histone lysine methylation distinct amid the rest of post-translational modification features, for instance lysine acetylation?

It is actually an intriguing question that comes to the minds of those who are into the field of epigenetics and chromatin. Several approaches can offer explanation to the distinctiveness of histone lysine methylation; the chemistry of lysine side chain, the functionalities of histone methylation marks and their crosstalk with the other epigenetic modifications, and the influence of the level of methylation in terms of valence on the local chromatin conformation and transcriptional potentials.
Based on the chemical structure, lysine is one of the basic amino acid, its side chain contains ε-amino group ($-NH_3^+$), rendering the side chain positively charged. When the side chain amino group is acetylated, the net positive charge is then neutralized, thus attenuating the electrostatic interaction between the net positively charged core histones within the nucleosome with the negatively charged backbone of the DNA wrapping around. Stressing that histone lysine acetylation is a mark of actively transcribed promoters and active enhancer elements; histone lysine acetylation, for example H3K27Ac, is associated with open chromatin structure and transcriptionally permissive conformation.[142-144]
On the other hand, methylation of histone lysine can take place at three levels of valence, mono-, di- and up to trimethylation, without altering the electrostatic interaction between DNA and core histones (Figure.4).[55, 145] This suggests that the epigenetic functioning of histone methylation marks does not hinge on disrupting the packaging of DNA around histones, and in agreement with the fact that histone lysine methylation marks can be associated with either active or repressed chromatin structures, depending on the position of lysine residue on histone tails as well as the valence of methylation. For instance, methylation at lysine residues 4, 36 and 79 are marks of active euchromatic loci, while methylation on lysine residues 9 and 27 are prototypical marks of long-range silencing at constitutive heterochromatin and of facultative heterochromatin, respectively.[40, 55, 145-149]

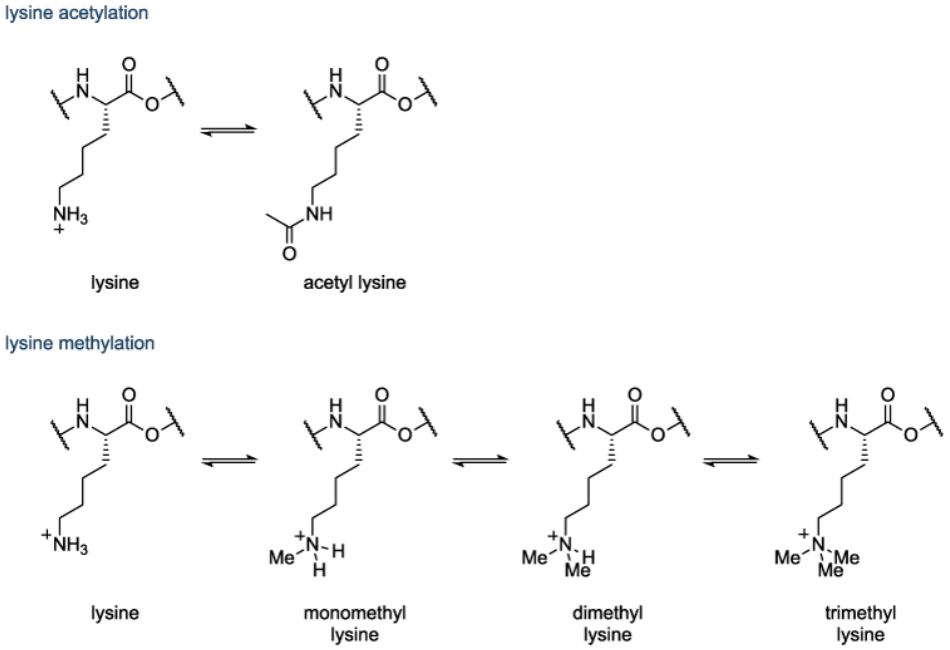

Figure 4. Lysine acetylation and methylation, edited and taken from ATDBio Nucleic Acids Book (www.atdbio.com)

Thus, lysine methylation exerts differential functions in different genomic context and not tethered to a single mode of action, on the contrary to histone lysine acetylation.

The establishment of histone methylation at lysine residues is accomplished by writers that harbor histone methyl transferase activity (HMT); such catalytic activity is mainly mediated by Su(var)3-9, Enhancer of Zeste, and Trithorax (SET) domain[150, 151]. While the reversal of histone lysine methylation marks is mainly catalyzed by Jumonji C (JmjC) domain-containing lysine demethylases (KDMs).[152-154] However, Some histone lysine methylation features such as methylation at lysine 79 on H3 tails has been shown to be catalyzed by DOT1L, while no dedicated eraser has been characterized so far.[155] With the lack of an eraser to such mark, the levels of methylation in terms of methylation valence have been shown to be in part dependent on the rate of nucleosomal turnover. Recently, it has been found that the processive kinetics of writing H3K79 methylation is adversely influenced by the high rate of nucleosome exchange.[69] By devising a chemical induced proximity model, a DOT1L fusion protein was targeted to certain loci with different transcriptional potential and different nucleosomal exchange rates, yet, deprived of H3K79me marks in mouse embryonic stem cells (mESC). Interestingly, the findings suggested that the dynamicity of the genomic locus favors the lower methylation status, independently of an eraser activity. Integration of datasets from ChIP seq, RNA seq and CATCH-IT seq in mESCs, provided parameters to Monte Carlo simulations that were extended to other histone methylation marks including H3K4 and H3K27 methylation marks. These simulations depicted the influence of nucleosome turnover on the methylation valences and revealed that lower methylation states are associated with the highly dynamic regions with higher transcriptional potentials. Consistently, it was surprisingly found that monomethylation of H3K27 (H3K27me1) is a mark of actively transcribed genes when deposited at their respective intragenic regions (gene bodies); whilst dimethylation (H3K27me2) marks were shown to be enriched at the intergenic regions, preventing the firing of non-cell type specific enhancers and are associated with a poised transcriptional activity.[156] Later, it was confirmed that the gene body-deposited (H3K27me1) marks are not intermediate products of PRC2 towards the ultimate trimethylation[157], in an agreement with the finding that H3K27me1 is dependent in large part on SETD2-mediated H3K36me3 deposition at active gene bodies.[156]

Altogether these findings reveals that histone methylations in terms of their establishment, epigenetic functioning is highly governed by the lysine residue, the methylation valence and

genomic context, and thus emphasize the distinctive placement of histone lysine methylation among the epigenetic mechanisms ever known.

Establishing histone methylation patterns:

As pointed out above, histone lysine methylation marks are not exclusive signs of a solitary chromatin status, they rather show a diversity of epigenetic facets of chromatin domains organization. In this section, I briefly explain the different classes of histone methylation machinery, specifically the antagonizing complexes that belong to two broad categories of chromatin modifying factors, Polycomb and Trithorax families.

Throughout the past seven decades, two key evolutionarily conserved chromatin modifiers, namely, Polycomb (PcG) and Trithorax (TrxG) group proteins has been viewed as integral parts of epigenetic cellular memory system, wherein they function in opposition to maintain repressed and active gene expression states, respectively. They were first discovered in *Drosophila melanogaster*, and were introduced as epigenetic gatekeepers that orchestrate the expression of Homeotic (*HOX*) genes, thereby, function in controlling body segmentation plans during embryonic development.[158-162]

During the last decade, extensive research has highlighted the functional diversity of both PcG and TrxG group proteins, which has expanded beyond merely regulating *HOX* genes expression. Such diversity is lucid in their implication in controlling a plethora of cellular and developmental processes including cell cycle control, proliferation, X chromosome inactivation, genomic imprinting, and regulation of stemness and development of cancer upon their aberrations.[163] These multiple functionalities hinge mainly on their ability to regulate chromatin on a wide-scale, ranging from local chromatin structural conformation to the three-dimensional organization of the genome. The wide-range chromatin organization is highly context dependent, in the sense that the assembly of the components of PcG and TrxG complexes takes place in a cell type- and developmental stage-specific manner; thus granting their abilities to spatio-temporally regulate genome in response to the varying cellular conditions.[163-166]

I. Trithorax (TrxG) complexes:

Trithorax members are highly heterogeneous, and thus play a widespread role in transcriptional activation and oppose the repressive activities exerted by polycomb in multiple ways. Trithorax category is subdivided in three different subcategories:

- **Switch/ sucrose non-fermentable (SWI/SNF):** a family that includes two groups, namely BAF and PBAF. Each is comprised of up to 15 accessory subunits that are uniquely assembled around core subunits (BRM/BRG1) that harbor ATPase activity, by which they mediate chromatin remodeling.[162, 165-168]
- **ASH1:** this subcategory includes proteins, the HMT activity of which is directed to write H3K36 methylation marks [169-171]; in addition, CBP which catalyzes H3K27 acetylation is also included.[172, 173]
- **COMPASS family:** the COMPASS subcategory of TrxG is associated with the histone lysine4 methylation which is attributed to active transcription. Perhaps the most prominent group in terms of its evolutionary conservation and the unique functionality in establishing the bulk H3K4 trimethylation is the SET1/COMPASS group.[174] The different complexes of COMPASS share a consensus group of core subunits, around which other accessory proteins assemble. The core complex is comprised of WDR5, ASH2, RBBP5 and DPY30; collectively the core complex is abbreviated as "WARD".[163, 174, 175]

  With the evolution of metazoans and the emergence of more complicated *cis*-elements, the SET1/COMPASS complex as experience divergence that gave rise to other combinations, namely the MLL1/2 COMPASS-like and MLL3/4 COMPASS-like. The former is functioning in catalyzing H3K4 trimethylation at certain subset of *HOX* genes and at bivalent promoter[176, 177]; while the latter mediates monomethylation (H3K4me1) at active enhancers.[174] Thus, the emergence of

these new combinations of COMPASS complex matches the increasing evolutionary complexity of the metazoan genomes.

II. Polycomb (PcG) complexes:

Despite their adversarial relationship, the complexities and evolutionary conservation of PcG and TrxG are matching to each other. Biochemical purification-based studies have viewed polycomb group proteins as of two cooperating, even if not necessarily redundant entities: Polycomb repressive complex1 (PRC1) which catalyzes mono-ubiquitination of lysine119 on histone H2A (H2AK119ub) and Polycomb repressive complex2 (PRC2) that functions in depositing the repressive prototype of facultative heterochromatin (H3K27me3).[163] Evolutionarily, PRC2 seems to have evolved later than PRC1, evidenced by the absence of PRC2 in the common model organisms, budding and fission yeast.[178, 179] The evolution of metazoans was accompanied by the emergence of a partnership between both PRC1 and 2, suggesting their Lamarckian evolution to match the complexity of metazoan genome organization in terms of establishment of facultative heterochromatin.

In mammals, both PRC1 and PRC2 exist in different combinations, wherein mutually exclusive sub-stoichiometric components that assemble to their respective core complexes; thus, exhibiting different modes of interactions (Figure.5). More tellingly, PRC1 for example is subdivided into: Canonical (cPRC1) and non-canonical (ncPRC1).[180] Both cPRC1 and ncPRC1 share a highly interspecies-conserved core complex, which comprised of:

- RING1A/B: the catalytic subunit that harbors E3 ubiquitin ligase activity that establishes H2AK119ub.
- Polycomb group ring-finger domain proteins (PCGF1-PCGF6), one of these six proteins should exist and considered as determining factor of the other accessory subunits that assemble to it. For example for the specification of cPRC1, a chromobox (CBX2, 4, 6-8), in addition to one of the polyhomeotic homologous proteins (PHC1-3) assemble around PCGF2/4.[163, 181] Whilst, the presence of zinc-finger and YY1 binding protein (RYBP) or its paralog YAF2 specifies the non-canonical PRC1 and define the enzymatic activity and the mode of interaction with PRC2 as well.[182, 183].

It was previously though that the relationship between PRC1 and PRC2 is unidirectional, in the sense that H3K27me3 the product of PRC2 is recognized by the chromodomain of CBX subunit of cPRC1, and then triggers the RING 1A/B to catalyze H2AK119 mono-ubiquitination which leads to stoppage of the RNA polymerase II to reinforce silencing. Nevertheless, the discovery of the diverse combinations of PRC1 revealed that the crosstalk between PRC1 and 2 is bi-directional, so that PRC1 can function independently of PRC2 or may even function in its recruitment as well. It has become well understood that the relationship between them is rather cooperative not hierarchical. The next section focuses on PRC2 functions and modes of recruitment to its genomic targets; therein, the crosstalk between PRC2 and PRC1 is explained in more details.

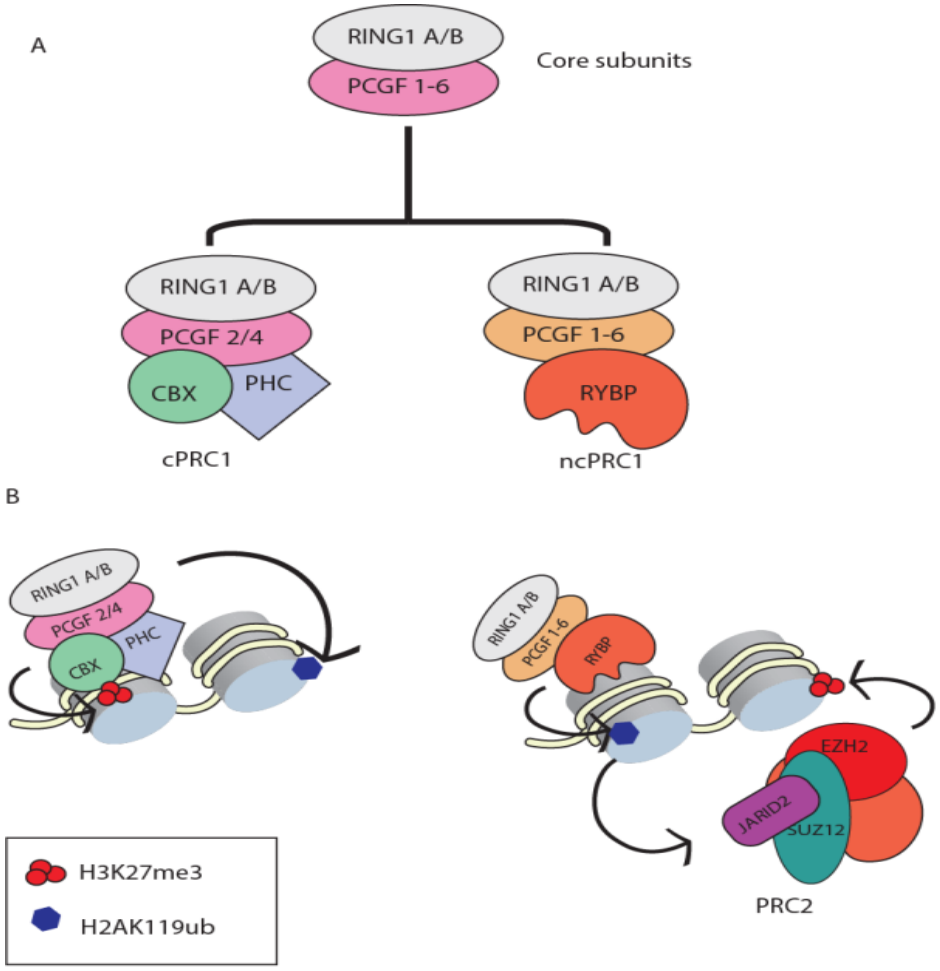

Figure 5. A) PRC1 assembly and specifications B) PRC1/PRC2 interaction

1.2.6 Polycomb repressive complex 2 (PRC2)

The specification of cellular lineage and preservation of cellular identities through the successive generations require tight control over transcriptional states. This in turn entails an intricate network of transcription factors that work in harmony with chromatin associated factors that represent additional regulatory layers contributing to transcriptional control.[184] Of these chromatin associated factors, are the PcG group proteins that are known to function in reinforcing transcriptional repression. Perhaps PRC2 is well studied and well known to be altered in many solid as well as hematopoietic malignancies; thus, has been rendered a promising therapeutic target for many cancer types.[185, 186]

PRC2 has been both intensively and extensively studied, and known to exclusively establish all levels of methylation at lysine 27 on H3 (H3K27me1/me2/me3), each shows distinct distribution over the different genomic loci.[156, 157, 184] The establishment of either methylation valence on H3K27 is directly implemented by the SET domain of the core subunit *Ehnancer of Zeste homolog 2* (EZH2) or its closely related homolog EZH1.[187] The SET domain harbors histone lysine methyl transferase activity (HMT), that is it has a binding pocket that accommodate the substrate H3K27, which is right adjacent the site where the cofactor *S-adenosyl Methionine* (SAM) resides. The SAM represents the methyl group donor; the methyl group is then transferred to the amino group of H3K27 side chain in an ordered manner so that the ultimate product is trimethylated H3K27 (H3K27me3).[188, 189]

The HMT activity exerted by the SET domain of EZH2 has been shown to be exclusively responsible for establishing the three valences of H2K27me which are characterized by distinct functional disruptions across the genomic domains (as discussed in 1.2.5). This actually points out to the diversity of the aspects of PRC2, in terms of the factors influencing the HMT activity, the residence time of PRC2 subunits, the mechanisms of PRC2 recruitment that constitutes the regional preferences of PRC2 and differential deposition of the various H3K27 methylation marks. One general and plausible explanation of such diversity is the existence of PRC2 in different subsets like PRC1.[183, 186] Biochemical and structural studies have revealed that the catalytic SET domain of EZH2 is rendered "auto-inhibited" if solitarily occurs, indicating the importance of the other regulatory subunits for the functionality of the PRC2 holoenzyme.[189, 190] Analogous to PRC1, It has been shown that PRC2 occurs in two main subsets, namely PRC2.1 and PRC2.2 (Figure.6). These two subsets share three core subunits including the aforementioned catalytic EZH2, in addition to *Suppressor of Zeste12* (SUZ12) and *Embryonic ectoderm development* (EED).[184, 191] Together, in an equimolar stoichiometry, the three subunits comprise the core PRC2 complex around which other accessory subunits assemble in sub-stoichiometric combinations, giving rise to either PRC2.1 or PRC2.2. It is suggested that these sub-stoichiometric interactions are transient and are more likely to be lineage- and developmental stage-specific.

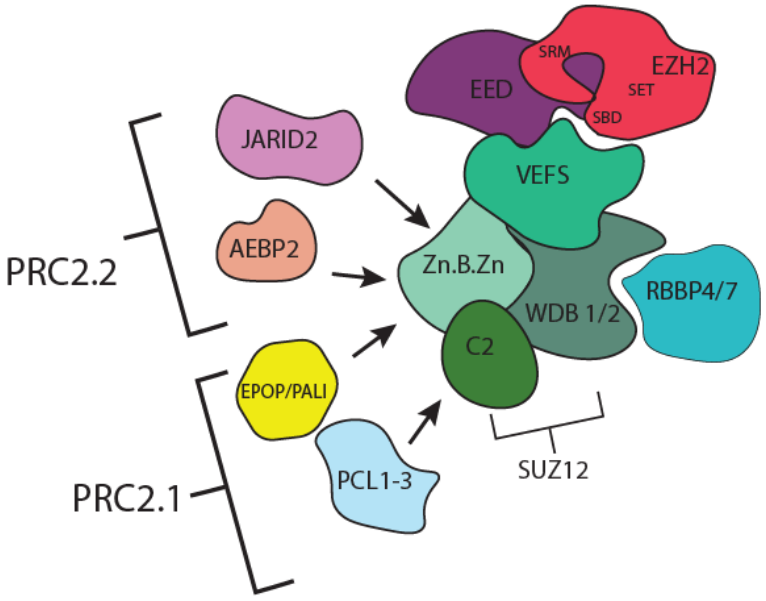

Figure 6. The assembly of PRC2 core subunits with accessory components into PRC2.1 and PRC2.2 subsets

..

The core subunit SUZ12 doesn't possess any enzymatic activities; however, it represents a docking station for the structural assembly of the other core as well as the accessory subunits (Figure.6). The C-terminus of SUZ12 contains VEFS domain around which assemble the core subunits EZH2 and EED; while the accessory subunits assemble to either of the N-terminal domains. For instance, the specification of PRC2.1 is achieved upon chromo-like domain mediated the interaction of one of three Polycomb-like proteins (PCL1-3) to the C2 domain of SUZ12; in addition, in mutually exclusive manner, either EPOP or PALI interacts to the Zn.B.Zn domain of SUZ12. On the other hand, for the assembly of PRC2.2, AEBP2 and JARID2 bind to Zn.B.Zn domain of SUZ12 via C2BH3K4 and trans-repression (TR) domains respectively.[190, 192, 193]
The occurrence of these subsets of PRC2 matches to its diversity and may in part explain the mechanisms of influencing the catalytic activity of PRC2 holoenzyme and the cellular condition-dependent recruitment of PRC2. In other words, this might explain the plasticity of PRC2 in controlling mammalian genomes. (See below)

The "Write and read" activity of PRC2:

Given that the absence of at least the core regulatory subunits EED and SUZ12 renders the SET domain of EZH2 auto-inhibited and apart from the modes of PRC2 recruitment to their genomic targets, the catalytic activity mediated by SET domain is allosterically regulated. Moreover, a key feature of PRC2 is the self-contained or self-perpetuation mechanisms, which explains the positive feedback loop that functions in the further enhancement of the HMT activity of SET domain upon recognizing its product (H3K27me3) by the WD40 domains within the aromatic cage of EED. It is true that this feature is not exclusive to PRC2; SUV39H1/2, the H3K9 methyl transferase exhibits a double activity of both writing and reading H3K9me.[191] On the other hand, the writing and reading activities are mediated by two core subunits, EZH2 and EED, respectively. Thus, the mode and pattern of interaction between these two subunits is crucial for understanding the "write and read" mechanism and how such interactions functions in the allosteric regulation of the SET domain of EZH2. Structural studies have unraveled the domains of EZH2 and the modes of interaction between EZH2 and EED (figure).
The EZH2 protein is subdivided into three main lobes, the N-terminal lobe (N-lobe), the middle lobe, and the C-terminal catalytic lobe that harbors the SET domain. As illustrated in figure, the N-lobe is comprised of six domains, the first of them (SBD) packs against the last (SANT1), forming a closed lasso around the aromatic cage of EED.[189] The WD40 domains of EED´s aromatic cage bind and recognize H3K27me3, thus undergoing conformational restructure which in turn induce the reorganization of SRM domain of EZH2 from its basal disordered state. The SRM is comprised of helical structures which interact to the helices of the post-SET portion (SET-I) by hydrophobic interactions, leading to release of SET-I portion that basally keeps the structure of SET domain suboptimal.[189, 190]
The self-perpetuating mode of action of PRC2 and the allotment of the writer and reading activities between EZH2 and EED respectively highlights the flexibility and dynamicity of PRC2 that matches its implication in establishing local facultative heterochromatin regions.[191, 194, 195] Furthermore, the existence of PRC2 in sub-stoichiometric subsets depending on the accessory subunits supports the idea that PRC2 is dynamically influenced in context dependent fashion. The modes of PRC2 recruitment over the genome during the different cellular conditions and how this in turn influence the catalytic activity would foster the understanding of the diversity and dynamicity of PRC2 in chromatin regulation.

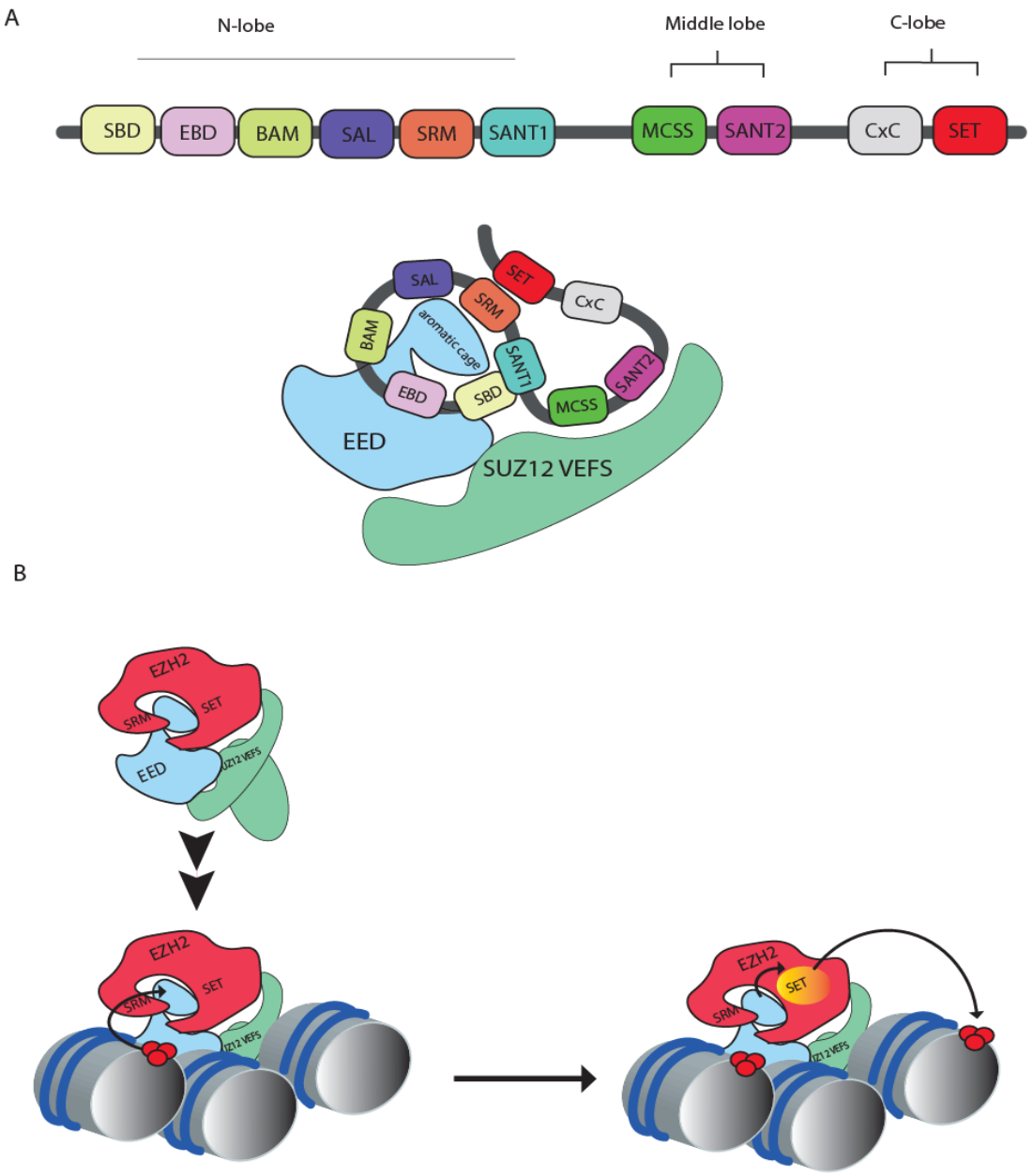

Figure 7. A) Map of EZH2 domains and the pattern of interaction with and looping around the aromatic cage-containing portion of EED. B) An illustration of self "write and read" mechanism of PRC2.

Targeting and recruitment of PRC2 over chromatin:

With the evolution of mammalian genomes, some elements have emerged, while other disappeared. Of note, in *Drosophila melanogaster*, polycomb complexes have been shown to be recruited at *cis*-elements termed Polycomb-response elements (PREs), thus providing a direct mechanism for locating PcG over the genome.[196, 197] On the contrary, the targeting of polycomb proteins in mammals is not that straightforward, evidenced by ChIP-seq studies that revealed that the regions enriched with H3K27 methylation, except few, are lacking PRC2 subunits, which highlight the "hit and run" nature of PRC2 and its transient residence over the majority of the genomic regions. PRC2 has no sequence-specific DNA binding motifs and its recruitment to its target loci relies on a plethora of factors that are after all governed by the cell type and the stage of differentiation.[184, 191, 198]

As mentioned in the previous sections, PRC2 exist in different combinations that share the core subunits EZH2, EED and SUZ12 and depending on the accessory proteins that assemble to the core complex, the subsets are differentially targeted to the genomic regions. In addition to the structural importance of SUZ12 for assembling PRC2 subsets, it plays an integral role in locating PRC2 preferentially at unmethylated, transcriptionally repressed CGIs. In this regard, SUZ12 relies on its N-terminal portion for recruiting proteins that possess DNA binding properties. For instance AEBP2 which binds at either C2 or Zn.B.Zn domains of SUZ12, harbors at its C-terminus a zinc finger domain that binds to DNA and thus was suggested as a recruiter of PRC2; however it was proven not be that efficient.[184, 191]
During pluripotency, PRC2 was found to preferentially occur in the PRC2.2 subset that contains both AEBP2 and JARID2.[199, 200] This latter contains JmjN and ARID domains that serve in loading PRC2 on nucleosomes and CGIs respectively; also JARID2 functions in allosteric regulation of SET domain activity in a way that mimics the write and read mechanism, thus establishing H3K27me3 profiles during the early stages of differentiation. On the other side, terminally differentiated cells prefers the PRC2.1 subset, wherein either of PCL proteins assemble to SUZ12, and by virtue of EH domain, it mediates the recruitment of PRC2 to CGI-rich regions. Accordingly, unmethylated CGIs are suggested to be the mammalian alternatives for PREs in flies and underscore the importance to investigate the mechanistic link between DNA methylation and PRC2.[201, 202]
Long non-coding RNAs have been also proven as efficient means for targeting PRC2 at the parentally imprinted alleles and inactive X-chromosome in females.[59, 203] Surprisingly, the actively transcribed genes during pluripotency have been involved in capturing PRC2[58], in the sense that nascent mRNAs compete with chromatin to bind at PRC2, in particular the middle lobe of EZH2. Notably the binding of the nascent transcripts of active genes and the presence of other active marks including H3K4me3 and H3K36me3 inhibit the catalytic activity of PRC2, yet tethering it within the vicinities of the actively expressed gene that might be destined for silencing upon lineage specification, when PRC2 is then liberated to catalyze heterochromatinization silencing.[184, 191] Thus, PRC2 is suggested to serve a transcriptional sensory function. The recruitment of PRC2 and how it influences its catalytic activity is becoming increasingly understood, yet, needs to be more deeply portrayed.

PRC2 accurately establishes H3K27 methylation patterns *de novo*:

Given that PRC2 has no consensus binding sites over the mammalian genome, the self-perpetuating, or the “write and read” mechanism is suggested to function in faithful restoration of H3K27 methylation patterns across the successive cellular division cycles. However, this postulation has become a matter of controversy. It has been found that as the genome is copied, H3K27 methylation shows two-fold dilution. In line with this, the knock-out of PRC2 core components in mESCs was found associated with global loss of H3K27 methylation pattern, while ectopic re-expression of PRC2 components resulted in restoration of H3K27me patterns at the same original loci in wild-type counterparts.[157] This suggests that PRC2 can establish H3K27me profiles *de novo*, and that the self-maintenance mechanisms partially serve to this endeavor, in particular during DNA replication.
This postulation is supported by several aspects:

- First, the deposition of lower levels of H3K27me within the genomic regions of differential transcriptional properties is less likely to be consequences of demethylase (UTX) gain of function, and neither are these lower valences intermediates towards the ultimate trimethylation of H3K27. This notion indicate the deliberately diverse functionality of PRC2 in catalyzing the *ad hoc* H3K27 methylation patterns and support that these patterns are not established nor maintained haphazardly.[156, 157]
- Second, the core subunit SUZ12 has been proven to play a substantial role in regulating PRC2-mediated H3K27 methylation.[184] Along with the assembly of the PRC2 core complex, it has been shown to locate at the PRC2 genomic targets independently of the other core subunits. Depending on either the EH domain of PCL

proteins or the ARID domain of JARID2, PRC2 binds to CGIs, hence implying an oriented functioning of PRC2.

- Third, as explained above, the differential occurrence of PRC2 in sub-stoichiometries is specified by the stage of differentiation. More tellingly, JARID2 plays a dual role for directing and influencing PRC2 functions, in the sense that it not only helps direct placement of PRC2.2 to target regions during pluripotency, but also serves to allosterically activate SET domain of EZH2 to enhance its HMT activity to *de novo* establish H3K27 methylation at the transcriptionally permissive loci that are deprived of H3K27me3 and destined for silencing upon differentiation. Notably, JARID2 and EZH2 decline upon reaching terminal differentiation and the HMT activity is then dispensed by EZH1[200], emphasizing that the H3K27 patterns are established *de novo* by PRC2 during lineage commitment and point out to the active implication of PRC2 in maintaining H3K27me3 even in terminally differentiated cells.

Altogether, these findings underscore the inadequacy of the self-maintenance mechanism to set the H3K27me profiles and confirm that they are deliberately established by PRC2 not passively by the self-perpetuating mechanism.

Non-canonical implication of EZH2 in transcriptional activation:

Along with the intensively understood facet of PRC2 in constituting additional regulatory layer that reinforce transcriptional silencing and establishing heterochromatin, other non-canonical aspects of PRC2 has been highlighted.[204-206] Interestingly, the catalytic subunit EZH2 has been viewed to serve oncogenic functions independently of its HMT activity and involves in methylating non-histone substrates; for instance it was found that upon its phosophorylation by Akt, EZH2 methylates and activates STAT3 in glioblastoma.[207] Also, EZH2 functions in inhibiting the interaction between the transcription factor GATA4 and p300 by methylating GATA4.[208]

Surprisingly, anomalous behaviors of EZH in transcriptional activation have been reported.[209-212] most prominently, EZH2 was found to exhibit a context-dependent double faced behavior in transcriptional regulation; more tellingly, EZH2 behave differently in two prognostic subgroups of breast cancer, namely the basal and luminal subsets.[213]

Interestingly, in the aggressive basal breast carcinoma, EZH2 interacts with RelA and RelB and functions in activating the expression of NF-κB-related genes. On the other hand, these genes were found to be repressed by PRC2-dependent activity of EZH2 in the luminal subgroup with more favorable clinical prognosis. This further underscores the dynamicity and plasticity of PRC2 in response to the cellular context. Given that the sole occurrence of EZH2 is associated with a loss of HMT activity and the interaction between the nascent transcripts of the actively expressed genes, the implication of PRC2 in transcriptional activation is not that strange and accordingly stresses on the importance of dissecting these non-canonical facets of PRC2 for comprehensive understanding of PRC2 in normal and diseased conditions.

PRC2 a promising therapeutic target for cancer treatment:

Several lines of evidence have reported the deregulation of PRC2-mediated H3K27 methylation patterns in solid as well as hematopoietic cancers.[163, 186, 189, 204] The aspects of PRC2 aberration are differential among the various cancer types; either gain- or loss-of-functions encounter the different subunits of PRC2 in human malignancies. Accordingly, it has become of a prime importance to develop lines of treatment to circumvent the cancer-associated PRC2 alterations.[214-216]

Structural and cryo-EM-based studies have portrayed the mode of HMT activity of EZH2; this approach initiated developing chemical compound that selectively target the HMT activity of EZH2. Some of the potent inhibitors have been designed to competitively inhibit the binding of the cofactor SAM at its binding pockets, such as EPZ005687 and other similar compounds.[216] although several of these SAM-competitive inhibitors are currently in clinical trials, their therapeutic indices are restricted by the imperfect pharmacological properties in terms of the bio-

availabilities and short half-time. Moreover, many tumors have been shown to exhibit resistance to these inhibitors; most citable is the gain-of function mutations (Y641X) in diffuse large B cell lymphoma (DLBCL). These substitution mutations along with other mutations in the vicinities of SET and SRM domains have been found to impede the binding of the SAM-competitive inhibitors, thus eliciting resistance to such therapies. Thus, the need to elaborate other lines of therapies that target PRC2 is tremendous.

Recently, other lines of treatment have been developed to target the self-maintenance of PRC2, in particular the aromatic cage of EED and block its binding to H3K27me3 and abrogate the initiation of H3K27me3-potentiated positive feedback loop.[190] Also, the other oncogenic activities exerted by EZH2 in a PRC2-independent manner have also to be considered. In this regard, other chemical compounds, like gambogenic acid derivatives have been developed to target the stability of EZH2 at protein level and induce its ubiquitin-mediated proteolysis; however, this line has to be adequately studied.[217]

Throughout the past decade, many ground-breaking studies have enhanced our understanding of PRC2 in terms of its recruitment, catalysis and diverse functionalities. The progressive understanding of PRC2 and its crosstalk with PRC1 and other epigenetic mechanism has proved such chromatin associated factor as highly dynamic and context specific that needs comprehensive and detailed understanding, which along with the personalized medicine might pave platforms for developing more efficient lines of treatment and improve the life quality of cancer patients.

1.3 Chronic lymphocytic leukemia (CLL)

1.3.1 CLL overview

Chronic Lymphocytic Leukemia (CLL) is a lymphoproliferative malignancy of CD5+ B lymphocytes that is characterized by increased number of CD5+, CD19+, CD23+ small and mature-looking leukemic B cells in peripheral blood and secondary lymphoid tissues, resulting in lymphocytosis, splenomegaly, bone marrow infiltration and lymphadenopathy.[218, 219] It is the most frequent hematopoietic malignancy in the west[220] and is common among the elders in particular and the median age of CLL diagnosis is ~70 years.[221-223]

The risk of CLL incidence is two-folds in males compared to females.[224, 225] CLL shows a heterogeneous clinical progression among patients; so that it ranges from patients with indolent disease and do not necessarily require therapy for so many years post-diagnosis, if any at all, to a more deadly aggressive disease that requires immediate treatment soon after diagnosis.

Such clinical heterogeneity renders CLL a hard-to-cure disease. Multiple factors like patients' age, genomic alterations and mutations, presence of comorbidities, in addition to the mutational status of immunoglobulin´s heavy chain variable region gene segment (IGHV) should be taken into the physicians´ considerations to decide the treatment lines and define the optimal therapeutic strategies.[218, 219, 226, 227] The latter factor is among the most important to consider, in this case the inhibitors of B cell receptor signaling or receptor tyrosine kinase inhibitors are suitable treatment options.[228-231]

New insights have progressively been gained and shed lights on both genomic and epigenomic patterns in CLL, which turned out to be at least in part associated to the clinical heterogeneity and disease aggressiveness.[232, 233] Here, CLL is our cancer model where we investigate the aberrant epigenetic profile in terms of DNA methylation and chromatin architectural anomalies as a consequence of polycomb repressive complex deregulation and show how this influences the gene expression profile of CLL which is in turn implicated in the disease aggression.

1.3.2 CLL pathobiology

Normal B cell development:

To understand the pathobiology of CLL, the development and maturation of B lymphocytes has to be understood. In bone marrow, the process of B cell development starts at hematopoietic stem cells (HSC), wherein interactions with stromal cells and nurse-like cells that comprise the specialized bone marrow microenvironment influence the maturation and activation of B cells in three ways:

- First, they specifically form adhesive contact with the developing lymphocytes via cell-adhesion molecules and their respective ligand on the stromal cells
- Second, the stromal cells provide either soluble or membrane-bound cytokines and chemokines that control differentiation and proliferation
- Third, these interactions leads to activation of downstream signaling that eventually leads to activation of lineage-specific transcription factors and then changes in the gene expression core regulatory circuits that serve in the commitment of specific cell lineage.

More tellingly, the stromal cells secretes interleukin-7 (IL-7) that binds to its cognate receptor on the surface of the common lymphoid progenitor (CLP), the IL-7 signaling then induces the expression of PU.1 and E2A transcription factors, which in turn function in activation of early B cell factor (EBF).[234] EBF specifies B-lineage via activating other B-lineage specific transcription factor PAX-5 and inducing the expression of the components of V[D]J recombination machinery (RAG-1 and RAG-2).[235] PAX-5 is known as a B-cell activator which

induces the expression of cell surface signatures of B lymphocytes CD19, BLNK and Igα; thus promulgating the development of Progenitor B cells (Pro-B cells).[236]
After the B-lineage is specified in the bone marrow, the process of immunoglobulin gene rearrangement start in early progenitor B cells (early Pro-B cells), wherein, the D[J] segment rearrangement of the heavy chain portion starts and upon its completion, early Pro-B gives rise to late Pro-B cell. In late Pro-B cell, the already rearranged D[J] segment recombines with one variable (v) segment and upon the completion of $V_H[D]J_H$ recombination process, late Pro-B cell develop into a large precursor B cell (Pre-B cell). The Pre-B cells carry a μ-heavy chain and with the development into small Pre-B cell, the process of light chain of immunoglobulin gene V_LJ_L rearrangement starts. After the V_LJ_L rearrangement is successfully done, the small Pre-B cell gives rise to an immature B cell that carries on its surface a pre-B cell receptor (pre-BCR), which is comprised of a μ-heavy chain and a surrogate light chain (Vpre B, λ5).[237]
The immature B lymphocytes are then subject for antigen experiencing; as centroblasts, immature B cells migrate to lymph nodes where the process of maturation and antigen stimulation takes place. The maturation of B cells is bifurcate, in the sense that it can either take place inside the germinal centers of the lymphoid follicles or at the marginal zone region. In the germinal centers, the centroblast immature B cell, experience antigen stimulation in T helper cell-dependent manner and the rearranged immunoglobulin heavy chain variable gene (IGHV) is subject to somatic hypermutations; thus coding for variable binding pockets to a variety of antigens. Whereas, the marginal zone maturation takes place in a T cell independent way and the somatic hypermutations are not obligatory, if any, the difference in homology compared to germline would be less than 2%, and the antigen stimulation takes place in response to bacterial lipopolysaccharides.[238, 239]
During antigen stimulation, those centroblasts that show no binding affinities towards antigens or bind to self-antigens are subject for negative selection and elimination by apoptosis; while those the properly bind to antigens would mature into centrocytes, that can mature to either antigen-experience memory B cells or plasma cells. Bottom line, the mature B cells are not monoclonal and can either harbor somatic hypermutations to their IGHV genes or do not. The process of B cell development and maturation from Pro-B cells till mature antigen-experienced B lymphocytes is illustrated in Figure.8.

Where does B-CLL originate from?

In the past, chronic lymphocytic leukemia was viewed as a disease of homogeneous origin and heterogeneous clinical course.[240] It was thought that CLL originates from long lived, immature and immuno-incompetent (Naïve) B lymphocytes. Until the end of the nineties of the last century, CLL was considered as a disease of accumulation, in the sense that leukemic B cells of diminished self-renewal potential relentlessly accumulate as a consequence of an inherent apoptotic aberrations.[239, 240]

This view has been dramatically transformed thanks to the wealth of progressive insights have been gained. CLL is by now viewed as a both clinically and clonally heterogeneous disease of two related entities originating from antigen-experienced B lymphocytes that differ in their activation, interaction to the microenvironment and immunoglobulin heavy chain variable region gene (IGHV) mutational status.[239, 241] Such heterogeneity ranges from relatively indolent disease course to a more aggressive, incurable and poor prognostic disease.[242] In addition, there is no inherent faulty apoptotic mechanisms encounter the entire mass of leukemic B cells. [243, 244] In this sense, CLL is not merely considered as a disease of accumulation anymore.

Based on the mutational status of IGHV genes, CLL is sub-grouped into two entities of distinct clinical outcomes, the more favorable IGHV-mutated CLL that originate from follicular mature B cell that has matured inside the germinal centers, and the clinically adverse IGHV-unmutated CLL that develop form marginal zone mature B lymphocytes, where very few or no mutations has been introduced to the rearranged IGHV genes. The latter is attributed with poor prognosis and

inferior clinical outcomes that can be rationalized by BCR restriction and sustainable proliferative signaling (explained below).[239, 242]

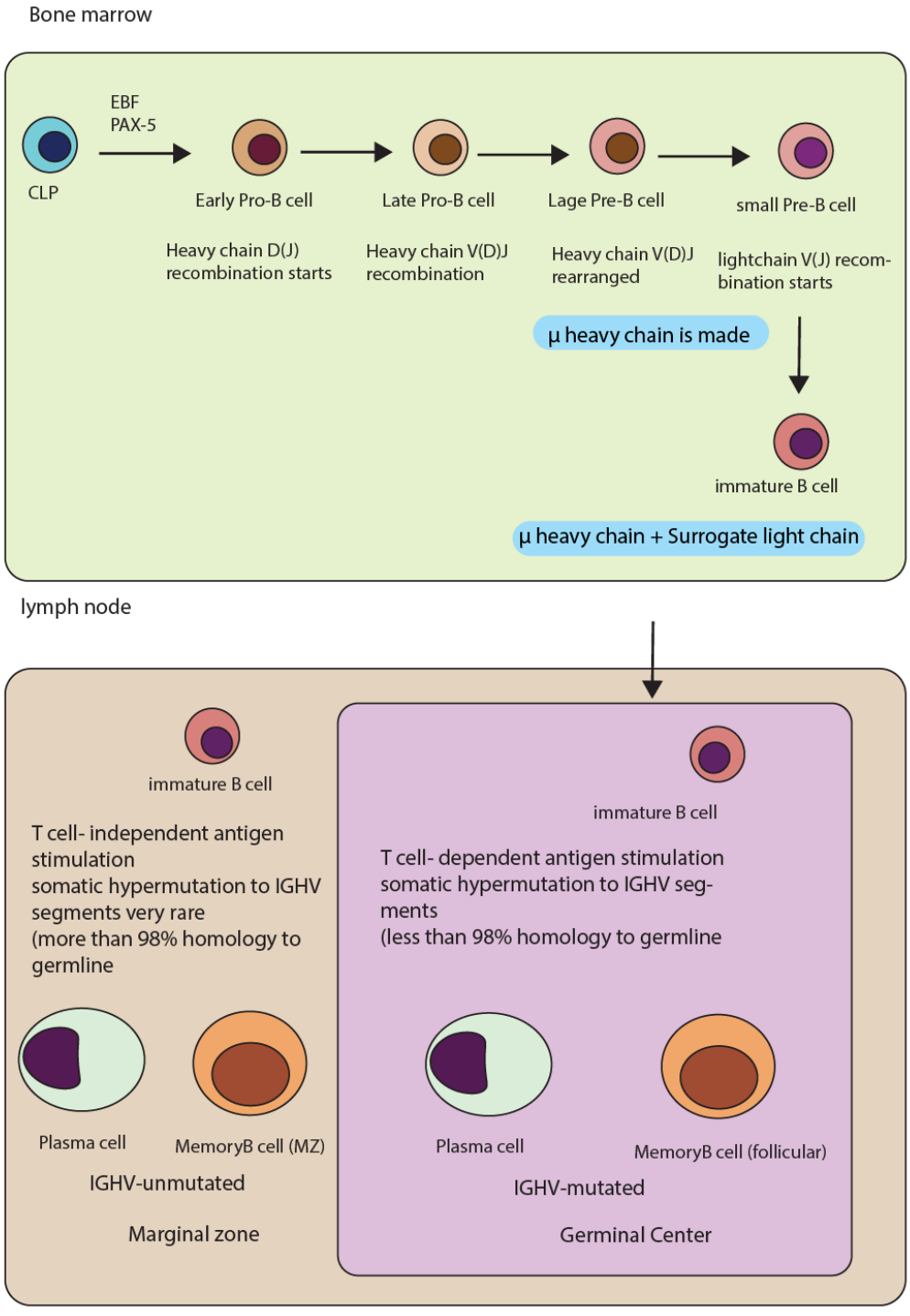

Figure 8. B lymphocytes development and maturation

CLL pathogenesis:

In most cases, CLL is preceded by an indolent disease called monoclonal B cell lymphocytosis (MBL)[245], which arises from either follicular mature B cells or marginal zone mature B cells, giving rise to either IGHV-mutated or unmutated CLL, respectively.
Over the time of CLL development, a small fraction can give rise to an atrociously aggressive stage of the disease called Richter's syndrome that is associated with poor survival.[246]
The progression of CLL hinges on some promoting factors including the reactivity of BCR, the interaction of the microenvironment and the altered signaling pathways. I hereby focus on BCR signaling and its link to the microenvironment and the affected pathways that contribute to the disease progression and how it impacts on the clinical outcomes of the disease. Also, I explain some of the recurrent mutations and cytogenetic aberration in CLL and whether they are cut out for being inducing factors of CLL development or promoting agents that emerge during the disease progression. Thus, I pave the way of the discussion of the significance of epigenetic patterns in CLL development.

B Cell Receptors (BCRs) signaling:

The immunoglobulin molecules repertoire made by CLL cells is more limited than that made by normal B lymphocytes. Given the number of possible V-gene segment recombinations that code for antigen-binding site on the immunoglobulin molecules, it would never be anticipated to find two among one million of CLL cases express structurally similar BCRs.[247] However, it has surprisingly found that various CLL patients show high degree of structural similarity, indicating that BCRs in CLL preferentially bind to certain antigens that are relevant to CLL pathogenesis. This further suggests that in CLL, there is a bias toward a special pattern of V_HDJ_H recombination, rendering antigen-binding pocket more restricted to certain antigens to which BCR are hyper-reactive compared to the much broader diversity in normal B cell counterparts. These certain promoting antigens are unknown; however it is believed that CLL may result directly or indirectly from specific latent viral perpetuated infections in a way that resembles the case of gastric carcinoma development as a consequence of *Helicobacter pylori* infection.[248]

The structural similarity of BCRs is considered as a brand by which the various CLL cases can be grouped into distinct categories. Such BCR restriction along with poly-reactivity highlights the substantial role of BCR signaling in CLL pathogenesis and promotion. The restricted BCR repertoire in CLL may also explain why CLL patients with unmutated IGHV genes exhibit much more adverse clinical outcomes compared to those with IGHV mutations.[247, 249]

Engagement of BCR with microenvironment:

As aforementioned that CLL has been shown to arise from mature, antigen-stimulated B lymphocytes that show abnormal B cell receptors (BCR) reactivity, in the sense that these receptors may show hyper-reactivity towards the microenvironment and autoantigens which can provoke clonal expansion. Upon interaction with microenvironment, BCRs interact to some accessory cells e.g. stromal cells and nurse-like cells and others that respond to cytokines and chemokines produced by those accessory cells. Thus, stimuli are delivered through receptor tyrosine kinases like Lyn, Syk and BTK which leads to stimulation of proliferative pathways including NOTCH, PI3K/AKT, mTOR, NF-κB and wnt/ β-Catenin pathways leading to uncontrolled proliferation and neoplastic transformation.[250, 251]

Cytogenetic lesions and mutations in CLL:

In line with the heterogeneity of CLL clinical course, cytogenetic alterations that have been associated with the prognosis of CLL patients are also heterogeneous.[252] Early cytogenetic analysis revealed a marked heterogeneity, at both inter-patient and intra-patient levels.
A landmark study was conducted on a large CLL cohort to comprehensively identify the mutational landscape in CLL using fluorescence *in situ* hybridization (FISH).[221] the findings of

this study underscored the most prominent cytogenetic lesions and copy number alteration in CLL and showed their differential occurrence among the different CLL patients.
Perhaps the most frequent cytogenetic lesion is del (13q14) which takes place in approximately 50% of CLL patients and is associated with kind of a good clinical outcome. The genes that are affected upon this deletion are mainly tumor suppressors and they include *MIR15A/MIR16-1* and *DLEU2*. Despite the fact that this cytogenetic aberration is associated with good prognosis, some sets of patients carrying this alteration at the locus of the tumor suppressor *RB1*, are showing more aggressive disease course.[253] The second most common cytogenetic alteration is trisomy 12, and it is associated with some driver mutations that affect *NOTCH1* and *BIRC3* genes, which are in turn associated with altered NOTCH signaling and NF-κB pathway, respectively.[254] Mutations that afflict *NOTCH1* gene occurs at a frequency of 10% of CLL patients and this frequency increases to up to 40% in the most adverse stage of CLL, the Richter syndrome.[255]
Coming after these two hallmark lesions, are del (11q22-23), del (17p13) and del (6q21) and others that take place at much lower frequencies in CLL patients. These lesions are associated with notorious mutations that are correlated with poor prognosis and much more aggressive clinical course of CLL.[256] The cytogenetic lesion del (11q22-23) occurs at up to 20% frequency and is associated with altered DNA repair system, as a consequence of *ATM* loss[257]; whereas del (17p13) is associated with the loss of the gate-keeper *TP53* leading to a serious adverse clinical outcome. Less frequently, del (6q21) occurs in CLL patients and is mainly associated with aberration in transcriptional regulation as a result of *ZNF292* loss.[221, 239, 252]
Several ominous point mutations to crucial genes for CLL biology have been identified. Perhaps most prominently are mutations of *SF3B1* which is of 20% frequency and is associated to altered RNA splicing and processing and eventually altered gene expression in CLL. Mutations and cytogenetic lesions are heterogeneous among CLL patients and are *bona fide* contributors to CLL prognosis.

1.3.3 CLL epigenome

Genetic lesions: inducing or promoting factors?

Considering the earlier adopted approach that cancer is a disease of mutations and the monoclonal nature of leukemic B cells (B-CLL), one might think that an inducing lesion must have existed in the progenitor clone and accordingly such lesion is supposed to be an inducing factor of the disease. However, even before the rise of the massively parallel technologies, the cytogenetic lesions were found to be very rare during the early course of CLL; thus arguing that mutations and cytogenetic lesions are the actual inducer of transformation of normal B lymphocytes to B-CLL. Moreover, among the other hematological malignancies and compared to other solid tumors, CLL is distinct by its relatively stable genome and the scarcity of chromosomal translocations.
One more aspect of CLL distinctiveness is that the mutational load is much less compared to other cancers; the average of mutational burden in CLL is approximately 1 mutation / megabase. Therefore, it is very unlikely that mutations and cytogenetic aberrations are inducing factors of CLL development; they rather emerge with the progression of the disease to a range of clinically heterogeneous courses. As pointed out above, the heterogeneity is not only observed between the different patients. This heterogeneity of cytogenetic lesion is also intra-patient, in the sense that some mutations and lesions are sub-clonal, meaning that they are acquired or present in only a fraction of the B-CLL pool in the same patient.
Bottom line, there is no unifying cytogenetic lesion or mutation identified so far as a universal attribute of CLL; thus supporting the counter argument to the idea that mutations/lesions are the factors that stand behind the development of CLL from normal B lymphocytes. This is also implies that there should be highly context specific mechanisms that encompass the emergence of mutations and the impaired proliferative pathways that are most likely to be of the hallmark

characteristics that are acquired as CLL progresses and underscore the importance of epigenetic mechanisms as candidate "enabling" factors of CLL development.

CLL aggressiveness is embedded in an aberrant epigenome:

As discussed previously that that mutational load of cancer might not qualify "mutations" as inducing factors of neoplastic transformation. Also, aberrant gene expression profile is surrogate of the cellular reprogramming and implements the manifestation of neoplastic or malignant phenotype. In CLL, the complex molecular pathogenesis substantially hinges on aberrant BCR signaling, making it a promising therapeutic target. The significance of BCR signaling in CLL progression is evidenced by the success of BCR inhibitors like ibrutinib in improving patients' prognosis. However, there are no unifying or recurring mutations found to encounter BCR signaling-related genes in CLL. Thus, it is of utmost importance to dissect the epigenome of CLL and its link to inter- and intra-patients heterogeneity.

Altered DNA methylation is of the most prominent hallmarks of CLL and has been extensively studied. A global loss of DNA methylation in CLL and regional preference to hypermethylation and the relevance of such altered methylome to CLL progression are well documented by several studies.[258-262] the findings of these studies suggested DNA methylation as a prognostic factor that can serve in stratification of CLL subtypes and shed the lights on the relevance of altered DNA methylation in reprogramming of transcription factors dependencies and altering gene expression regulatory circuits that establish a malignancy-specific epigenome.

Also, it was found that hypomethylation at LINE and Alu elements is associated with genomic instabilities in CLL, which supports the idea that altered epigenetic mechanisms encompass the cytogenetic lesions, the impact of which in CLL progression is secondary to aberrant epigenome.[263]

Moreover, it has been shown that all cell types display asymmetrical patterns of enhancer potentials, pointing out to a set of super-enhancers that have been found to be cancer-type-specific.[264] In CLL, integration of global histone acetylation and chromatin accessibility data revealed CLL-specific super-enhancers mediating the CLL-transcription factor core circuitry.[265] This study identified PAX-5 as a core regulator of CLL super-enhancers essential for B-CLL survival and pointed that targeting of CLL-super-enhancer circuits is a promising line of therapy.

A recent study has been conducted on a cohort of 23 CLL patients and 17 pools of CD19+ sorted out from healthy age-matched participants to comprehensively portray the altered epigenome in CLL.[266] Histone modification of 7 marks (H3K4me1, H3K4me3, H3K27me3, H3K9me3, H3K36me3, H3K27Ac and H3K9Ac) were assessed by ChIP-seq; the peak regions of each of these marks, along with transcriptome analysis by RNA-seq, were used as input to the 12 state ChromHMM Hidden Markov model for genome segmentation and accordingly addressing the epigenetic transitions of chromatin states between normal B cells and CLL. Also, this study mapped the differentially accessible regions by ATAC-seq and MNase-seq, in addition to global methylome assessment by whole genome bisulfite sequencing (WGBS). Integration of ChIP-seq, RNA-seq and MNase-seq data revealed a set of genes regulated by alternative promoters in CLL; in the sense that some promoters were found to extend their H3K4me3-regions by 2 nucleosomes (≈400bp). Consistently, these extended promoter loci lost ATAC-signals and found to be mainly enriched to BCR signaling-related genes. Moreover, 1700 promoters that are bivalent in normal B cells were found to lose bivalency, in the sense that they lose H3K4me3 and retain H3K27me3 and were found to be binding motifs for MEF2 family of transcription factors.

ATAC-seq revealed that about 30% of the differentially accessible loci gained ATAC-signals in CLL compared to normal counterparts, while 70% of these loci were accessible in normal B cells and lost ATAC-signals in CLL. The loci that gained ATAC-seq were supposed to be *bona fide* transcription factor binding sites specific to CLL, and thus subjected to motif analysis of these loci by HOMER software. Gene regulatory network analysis using expression data of 264 B cell samples and RNA-seq of this study were intersected to motif analysis and revealed 1378 regulators displayed differential activity between CLL and normal B cells (FDR > 0.05), the top four of which were FOXA1, LEF1, POU3F1 and REPIN1.

This integrative analysis could also predict potential CLL-specific enhancers; that is super-enhancers were annotated by clusters of open chromatin regions (ATAC-signal gain), H3K27Ac (ChIP-seq) and bidirectional RNA expression (RNA-seq). Out of 1289 union super-enhancer, 310 showed differential super-enhancer activities; 219 of these 310 showed upregulation of enhancer activity and defined CLL-specific super-enhancers. It was found that *TCF4*, *CREB3L2*, *FMOD*, *LEF1*, *ETV6* and *NFATC1* were supposedly upregulated genes by CLL-specific super-enhancer.
Integration of all the data included in that study along with other ENCODE data underscored the deregulated CLL-signature genes and intersection of these genes with transcription factors of differential activities highlighted the aberrant regulators and target genes in CLL.
In line with others, the data of WGBS of that study revealed a global hypomethylation in CLL compared to normal B cells that was rationalized by partially methylated domains (PMDs) in 50% of the CLL genome. Interestingly, the repressive domains were overrepresented in PDMs and 75% of the recurrent somatic mutations in CLL were found located in these PMD regions, consistent to the increased mutational burden at repressed domains and further confirming the idea that aberrant epigenome is upstream to mutations and cytogenetic lesions.
Furthermore, 91% of the differentially methylated regions which are mainly hypomethylated were overlapped with predicted enhancers, suggesting the impact of differentia DNA methylation pattern in the asymmetric enhancer activity between normal B cells and B-CLL.
What has been discovered so far indicate that inter- and intra- patient heterogeneity is epigenetically attributed; thus superimposing the complication of CLL as an incurable disease. This necessitates collaboration between epigenesists and physicians to stratify CLL patients' prognosis based on their altered epigenome and dig for a unifying epigenetic aberration that might be a promising therapeutic target for generating more efficient treatment options for better prognoses.

2 AIM

This work mainly focuses on probing the aberrant epigenetic patterns in CLL, in terms of global alteration of DNA methylation within the transcriptional units and the abnormal patterns of H3K27 methylation, and how these altered epigenetic patterns are involved in shaping the gene expression profile in CLL. The thesis includes three studies the unifying goal of which is to scrutinize the mechanisms by which the aberrant epigenome is implicated in the deregulation of signature genes that are of biological relevance to CLL progression.

2.1 Specific aims

I. Investigate the mechanism elicited by gene body methylation in transcriptional activation of *TET1* as a candidate model gene to address the implication of methylation within the intragenic regions of the gene that are prone to hypermethyltion in CLL compared to normal B cell counterparts. (Paper 1)

II. Global mapping of H3K27 methylation patterns and genome-wide distribution of its writer EZH2 in two prognostic subgroups of CLL patients. In addition to uncover a possible PRC2-independent implication of EZH2 in activation of PI3K/AKT pathway through transcriptional activation of *IGF1R* in IGHV-unmutated CLL the worse prognostic subtype. (Paper 2)

III. Draw a functional link between intragenic CpG islands methylation and the functionality of PRC2, in terms of its localization and catalytic activity and probe the possible mechanism behind the adversarial relationship between PRC2 and CpG islands hypermethylation in controlling the expression of *MNX1* in CLL. (manuscript)

3 PATIENTS AND METHODS

3.1 Patients cohort

Total 40 CLL patient samples (twenty U-CLL and twenty M-CLL) were used in Paper 1 and Paper 3. The peripheral blood mononuclear cells (PBMCs) CLL samples were collected from different hematology departments in the western part of Sweden after written consent had been obtained and is approved by Regionala etikprovningsniimnden in Gothenburg (decision number: 197-18). All samples were diagnosed according to contemporary iWCLL criteria for the time of sampling, showing typical CLL immunophenotype and 70% or more tumor percentage. The median age at diagnosis was 66 years, with a male: female ratio of 3: 2. For comparison to normal counterparts, five DNA samples of CD+19 sorted normal B cells from healthy age matched controls were purchased from 3H biomedical (Uppsala, Sweden).
Sorting of B lymphocytes was performed using autoMacs protocol from the respective buffy coats.
Paper 2 was conducted using a cohort of 113 CLL patients with evaluated molecular and clinical data. Out of these samples, 12 cases (six U-CLL and six M-CLL) were subjected to ChIP-seq to map EZH2 and H3K27me3 patterns, 16 samples were included in validation experiments to ChIP-seq data; these 16 samples along with the rest (n=82) were used for assessing expression levels of *EZH2* and *IGF1R* using real-time PCR analysis, and the data were used for Pearson correlation analysis. The sample provision has been approved by the G. Papanicolaou Hospital ethics committee (decision number: 149/27.3.2015). CLL patients were diagnosed according to the iwCLL criteria.[267] The patients were either untreated by the sampling time or off-therapy for at least 6 months before the time of sampling.
All patients provided informed consents in accordance with the Helsinki Declaration and the study was approved by the local ethics review boards.

3.2 Methods

1. DNA methylation assessment by Pyrosequencing:
The principle on which pyrosequencing is based, is the sequence-by-synthesis approach[268]; more tellingly, a single stranded DNA molecule serves as a template for progressively incorporated labelled deoxyribonucleotide triphosphate (dNTPs) by DNA polymerase. This method is also substantially based on bisulfite conversion of DNA, the subject of methylation assessment. Unmethylated cytosine bases in either CGI or non-CGI contexts are subject for deamination into uracil upon bisulfite conversion reaction, while methylation at fifth position of cytosine (5-mC) remains intact. Thus, methylated cytosine can be detected by other downstream applications including whole genome bisulfite sequencing (WGBS) or reduced representation bisulfite sequencing (RRBS) and other methods. In this work, pyrosequencing was the mean of assessing DNA methylation patterns in CLL patients. The bisulfite converted DNA of probe is subjected to PCR amplification using two primers designed by Pyromark software, so that one of them is biotinylated. The PCR amplicon is then immobilized to high-performance streptavidin sepharose beads by CLICK-IT chemistry and is then denatured into single stranded template that anneals to a designed sequencing primer; from this step, starts the sequencing-by-synthesis process. The sequencing-by-synthesis process takes place inside Pyromark machine, wherein the sequencing immobilized biotinylated single stranded on sepharose beads which is also annealed to the respective sequencing primer is in contact with dNTPs, enzyme mixture which is comprised of DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrate mixture that contains luciferin and adenosine 5’ phosphosulfate (APS). After incorporation of each dNTP onto the single stranded DNA template by the polymerase, a pyrophosphate is released and reacts to APS yielding an ATP by ATP sulfurylase. The ATP is in turn drives the action of luciferase in converting luciferin into oxyluciferin that generates visible light proportional to the consumed

ATP that corresponds to each incorporated dNTPs. The light signal is then detected and recorded as a peak in the pyrogram. The peaks are surrogates of nucleotides sequence and the CGI sites as well as methylation % are inferred on the pyrogram.

2. Measuring promoter activity by luciferase reporter assay:
The transcriptional potential of either enhancer or promoter genomic elements can be assessed by inserting the genomic sequence of interest into luciferase reporter vector (e.g. PGL3 vectors) in comparison of coexisting β-galactosidase reporter gene.[269] The high sensitivity of the assay allows the discrimination of transcriptional activities of elements that show minor difference in their respective potentials. The transcriptional potential of the inserted sequence accordingly influence the expression of the reporter gene (luciferase). The sequence of investigation is cloned into reporter vector and transfected into transfection-tolerant cell lines; perhaps MCF-7 and HEK are of the best candidates. After a considerable time of growing the cells post-transfection, preferably at a time that exceeds the doubling time of the cells, the transfected cells are then lysed and incubated with luciferin and other cofactors like ATP. As described in the previous section, the expressed luciferase catalyzed the oxidation of luciferin into oxyluciferin in an ATPase-coupled reaction. The light emitted is detected and quantified by luminometer as light units which are in turn normalized over the β-galactosidase activity; thus the relative light units are measures of the relative luciferase activity which represents the transcriptional potential in turn.

3. Chromatin immunoprecipitation (ChIP)
The ChIP method aims at assessing DNA-protein interaction and identification of transcription factor binding sites at certain genomic regions, in addition it can assess chromatin status by determination of certain histone post-translational modification marks.[270, 271] The principle of this method hinges on fixing the cells of study by 1% formaldehyde for 10-15 minutes, so that chromatin is crosslinked to its associated proteins or histones. The fixation process is supposed to be not longer than 15 minutes and to quench the formaldehyde mediated crosslinking, 125 uM Glycine should be added to the fixed cells. Next, the fixed cells are subject to lysis and the crosslinked chromatin should be exposed to shearing by either enzymatic means using restriction enzymes, or by random sonication (30 second ON/ 30 seconds OFF) so that the sheared smear is supposed to range from 100- 500 bp long. The sheared chromatin is then subject for immunoprecipitation using high quality ChIP-grade antibody for the transcription factor, chromatin modifier or histone mark of probe. The antibody can be crosslinked to protein A/G-coated magnetic beads. The immunoprecipitation reaction takes up to 16 hours at $4^{o}C$, then the beads crosslinked with the antibody-DNA is then subject for washing by buffers containing gradient concentrations of detergent. The washed beads-antibody-DNA is then subjected for de-crosslinking in the presence of proteinase K, so that all DNA-binding proteins, histones and beads are released and the DNA sequences are then eluted and used for downstream applications, either high throughput sequencing platforms (explained below) or amplified by ChIP-real time PCR (explained in section 3.3).

4. Hight throughput sequencing and bioinformatics analysis:

4.A. ChIP-sequencing (ChIP-seq):
ChIP was performed on 12 CLL samples using the Ideal ChIP seq kit (Diagenode, C01010051) according to the manufacturer's protocol using an EZH2 polyclonal antibody (Diagenode, C15410039) and an H3K27me3 polyclonal antibody (Diagenode, C15410069).
The general workflow of ChIP includes:

- Preparation of cDNA libraries, with considerable number of reads (40-50 million reads/ sample) and the read length at 2 * 100 bp.
- A quality check is performed using FastQ software, and then the reads are filtered based on their quality using TrimeGalore software.
- The filtered reads are then subject for alignment to human reference genome assembly GRCh37/hg19 using Bowtie2.

- Quality check by cross correlation, cumulative enrichment and clustering were performed using phantompeakqualtools and deepTools.
- Removal of the duplicates and blacklisted-mapped-reads is carried out by Picard
- Peak calling is performed with MACS2, with a minimum FDR (q-value) cut-off for peak detection of 0.05.
- Normalization of peaks with their corresponding input using deepTools2.
- Annotation of peaks in different genomic regions using W-score with cutoff 0.1 and 0.01and finding unique peaks by BedTools.
- The average profile and density heatmaps obtained for window sizes +/- 5KB of transcriptional start sites (TSS) in M-CLL and U-CLL were plotted using deepTools2.
- The pathways enriched in EZH2 and H3K27me3 overlapping regions of each CLL subgroups were identified using ChIP-Enrich.

The general workflow is described in detail in paper 2[272] and illustrated in figure.8.

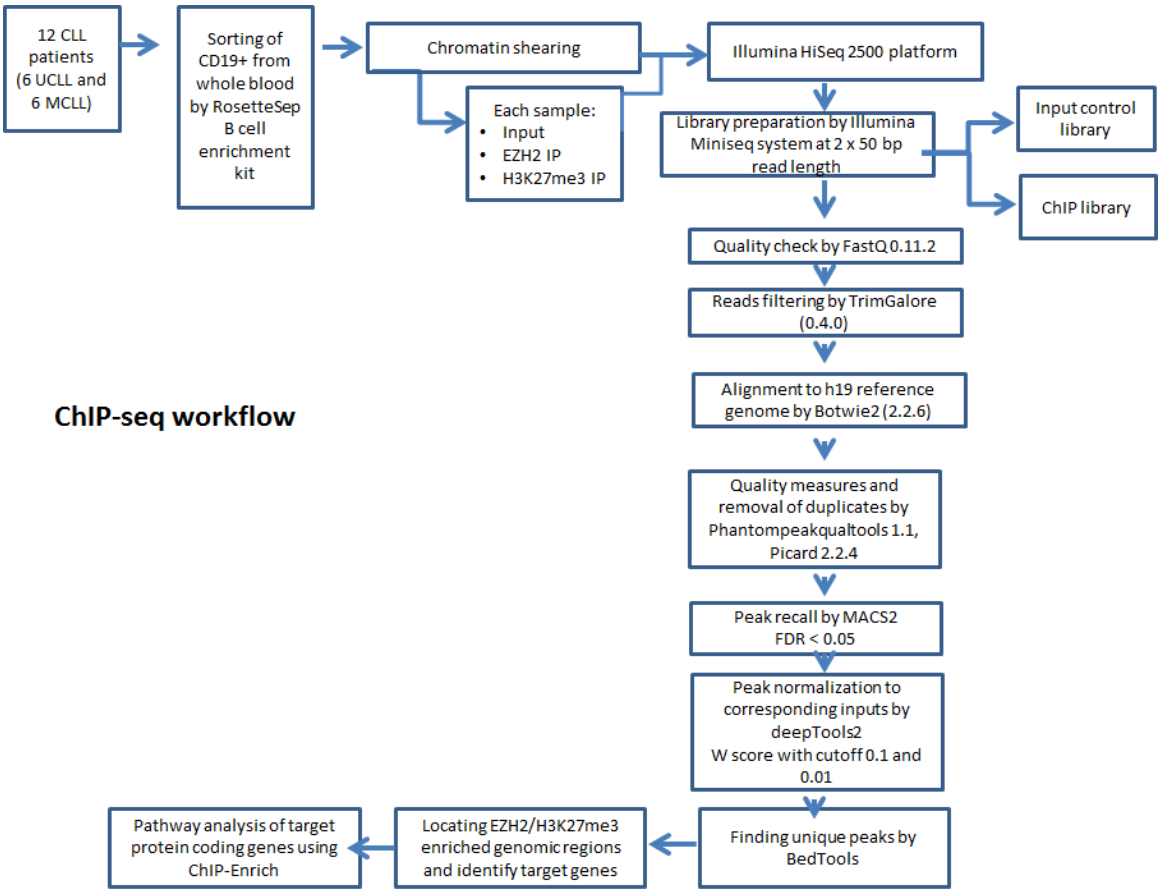

Figure 9. ChIP-seq workflow

4.B. RNA-sequencing (RNA-seq):

RNA samples extracted form CLL cell lines were subject to cDNA library preparation. The libraries construction and sequencing were performed by Novogene (Hong Kong) using Illumina HiSeq 2500 platform. The single-end cleaned reads were aligned to human genome assembly GRCh37/hg19 using STAR aligner version 2.5.2b [273], and the alignment quality was further assessed using SAMStat tool[274]. The high quality of reads mapping score was considered (MAPQ≥30) for downstream quantification based on gencode version 19 annotation using with

the aid of featureCounts [275], excluding the multi-mapped reads. To calculate the differential expression statistics, edgeR Bioconductor package was used[276]. The statistical false discovery rate (FDR) was set at a cut-off <0.05 for multiple comparisons and expression log2 fold change ≥ ±1. Pathways enrichment analysis of differentially expressed genes was done using GeneSCF tool (5) and considered a statistical cut-off value of 0.05 for the enriched pathways.[277]

3.3 Laboratory assays

Cell culture and maintenance:
Two CLL cell lines, namely HG3[278] and MEC1[279] along with other mantle cell lymphoma cell lines (Z138, GRANTA 519, Jeko and Mino) were used to conduct the experiments of the three studies included in this thesis. Cell lines were cultured in RPMI 1640 containing L-glutamine and supplemented with 10% fetal bovine serum and 1X penicillin/streptomycin and maintained at 37oC and 5% CO2 as earlier described [280]. To investigate the effect of CpG methylation on gene expression in paper 1 and the manuscript, cells were treated with the 5'-Aza-2'-deoxycytidine (DAC) methyl inhibitor using CLL cell lines cultured in RPMI media at an optimal concentration 30uM according to a previously described protocol.[123] also, 3-Deazaneplanocin (DZNep) and JQ treatment of cell lines were performed to inhibit EZH2 and c-MYC respectively for three days using different concentrations. Similarly, the usage of chemical inhibitors that target the catalytic activity (GSK 343 and UNC1999), in addition to Idelalisib the inhibitor of PI3K/AKT pathway were performed using cell lines as described in details in paper 2.[272]

Transfection:
Transient transfection by electroporation was carried out by Amaxa 4D-Nucleofector™ System (Lonza group AG, Basel, Switzerland) using the SF cell line Amaxa kit (V4XC-2032) according to the manufacturer's instruction. 50nM of Predesigned Stealth RNAi siRNAs were utilized against *EZH2*, *c-MYC*, *IGF1R* (ThermoFischer Scientific, Waltham, USA) in equal concentrations. The silencer negative control siRNA (ThermoFischer Scientific, Waltham, USA) was used as control siRNA. The transfected cells were kept and maintained for one doubling time (48-72 h) depending on the type on the used cell lines. This protocol was also applied for overexpressing miR26A1 and PGL3 vector containing-*TET1* HMR as described in paper 1.[281]

RNA extraction and cDNA synthesis:
RNA was extracted from CLL PBMC samples using RNA Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNAse1 treatment was performed to remove any residual genomic DNA during RNA extractions. The total cDNA synthesis was performed using Superscript III FS synthesis supermix kit (Life technologies, Carlsbad, USA) according to the manufacturer's protocol. For cell lines, RNA extraction was performed using Relia Prep RNA miniprep system kit, Promega according to the protocol provided by the manufacturer. cDNA samples was prepared from their respective RNA using GoScript Reverse transcription system kit (A5000, Promega) in compliance to the producer´s protocol.
The extracted RNA samples that were destined for RNA sequencing were subjected for concentration and quality analysis by Bioanalyzer 2100 before cDNA library preparation.

Gene expression analysis by RT-PCR:
Real-time PCR was performed using Power SYBR Green PCR master mix and the 7900HT fast real-time PCR instrument (Applied Biosystems,Warrington, UK). Relative expression was calculated and normalized to *HPRT* and/or *GAPDH* genes. The relative expression levels were calculated by the ΔΔCt method.

DNA extraction, Bisulfite conversion:
DNA extraction from CLL PBMC samples and CLL cell lines was performed using DNA Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Bisulfite conversion of DNA was performed to assess DNA methylation by using EZ DNA Methylation-Gold kit (Cat # D5005). The eluted bisulfite converted DNA was subject for methylation analysis by pyrosequencing

Pyrosequencing:
Pyrosequencing was conducted on bisulfite converted DNA using PyroMark Q24, to assess differential methylation across some selected regions across *MNX1* gene obtained from our previous global MBD-seq analysis and the primer sets were designed using (software) and listed in Supplementary table. The analysis of data obtained from PyroMark machine was analyzed by PyroMark Q24 Advance software.

Protein expression estimation by western blotting:
Western blot analysis was performed using total cell lysates lysed from transfected or drug-treated CLL cell line samples in RIPA buffer (Sigma-aldrich, St.Louis, USA) with PI inhibitors (Roche, Basel, Switzerland). Equal amounts of lysates were loaded on NuPAGE 4-12% Bis-Tris gels (Invitrogen, Carlsbad, USA) and transferred to membranes (Amersham Hybond ECL; GE Health Care Life Sciences, Sweden). Blocking of the membranes is performed in 5% BSA with TBS with addition of 0.1% Triton X-100, then the membranes were incubated with the appropriate primary and secondary antibodies, followed by washes with TBS containing 0.05% Triton X-100. The primary antibodies used for western blotting were: TET1 (ab191698, Abcam, Cambridge, UK), EZH2 (3147, Cell Signaling Technology, Danvers, USA) and GAPDH (SC-25778; Santa Cruz Biotechnology, Dallas, USA), MYC (ab17355, Abcam), H3K27me3 (C15410069), H3K4me3 (C15410003), H3K27Ac (C15410174) polyclonal antibodies (Diagenode, Liege, Belgium), H3 (ab1791, Abcam), IGF1R (ab39675, Abcam), Phospho-Akt (Ser473; 9271) and AKT (9272) antibodies (Cell signaling technology, Netherlands). The secondary antibodies were anti-mouse IgG (7076S, Cell Signaling Technology, Danvers, USA) and anti-rabbit IgG (7074S, Cell Signaling Technology, Danvers, USA). Blots were developed using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific, Rockford, USA) and the visualization is performed using the ChemiDoc XRSC (Bio-Rad) instrument.

ChIP assay:
ChIP assays were performed using the iDeal ChIP-seq kit (Diagenode, Liege, Belgium), according to the manufacturer's instructions. The immunoprecipitation was conducted using antibodies against EZH2 (39876, Active motif), H3K27me3 (C15410069, Diagenode), H3K27me1 (C15410045-10, Diagenode), H3K36me3 (C1541092-10, Diagenode) and polyclonal rabbit anti-IgG (c15410206, Diagenode) as negative control.
The final eluted DNA was analyzed by real-time quantitative PCR with the Power SYBR Green PCR master mix (Applied Biosystems, Warrington, UK). The PCR conditions were: 50 °C for 2 min, 95 °C for 10 min, and 40 cycles of 95 °C for 15 s and 60 °C for 1 min. The PCR analysis was performed with the 7900HT fast real-time PCR system instrument and software (Applied Biosystems, Warrington, UK). All primers were designed by using PRIMER3 online tools.
The calculation of enrichment was done according to the following formulas:

- The % of IP to input formula:
  % recovery = 2^[(Ct input-log2(5%)) – Ct sample] * 100%
- The enrichment was calculated as: %recovery IP / % recovery IgG

The eluted ChIP DNA was quantified using Qubit 2 Florometer (ThermoFischer Scientific, Q32866) and around 10ng of ChIP DNA was used for preparing sequencing libraries by ThruPLEX-FD Prep Kit (Rubicon Genomics, Ann Arbor, USA) according to the manufacturer's protocol.

Cloning:
Two regions within the HMR region of TET1 that was identified by MBD-sequencing data analysis from a published study [120] (519 and 632 bp long) were amplified by PCR. The PCR primers were designed to contain *Kpn1* and *XhoI* restriction sites at 5'- and 3'- ends respectively for positive orientation insertion of the sequence of interest and the other way around for negative orientation insertion. Restriction digestion of both PGL3 vector and the amplified sequences was performed by *Kpn1* and *XhoI*; following that, mounting of both oriented sequences was performed by T4 ligase. The ligated inserts-vectors constructs were then transformed into TOPO10 competent bacterial cells by heat shock followed by outgrowth in SOC medium and plating on LB agar containing ampicillin. Miniprep experiments were performed to select the optimal colonies for further outgrowth in LB medium containing ampicillin. Midiprep was then performed to eventually extract and purify the constructs in TE buffer.

Dual Luciferase reporter assay:
The dual-luciferase reporter assay system (Promega, Madison, USA) was performed in MCF-7 cells transfected with PGL3 constructs by Lipofectamine LTX plus reagent kit (Invitrogen, Carlsbad, USA) according to the protocol supplied by the manufacturers. The relative light units were measured with a luminometer (Glomax 20/20 luminometer, Promega, Madison, USA). Each experiment included at least three independent biological replicates, and the final luciferase values (relative light units) were calculated by dividing the luciferase activity by the β-galactosidase activity.

Strand specific PCR:
Specifically designed 5'GSP (Gene specific Primer) and 3'GSP were used for synthesizing gene specific cDNA from DNase1 treated RNA samples. The 5'GSP specifically binds to antisense strand and the 3'GSP binds to the sense strand. After cDNA synthesis, to quantify the strand specific cDNA synthesis we used two sets of primers which are located downstream to the 5'GSP and 3'GSP. The PCR conditions for amplification were 95°C for 10mins, 95°C for 30seconds, 55°C for 30seconds, 72°C for 30seconds and 30 cycles of 95°C, 55°C and 72°C for 30seconds and finally 72°C for 3minutes. The amplified product was run in 1% agarose gel and visualized using the ChemiDoc XRS+ (Bio-Rad) instrument.

3.4 Statistical analysis

Statistical analysis was performed with STATISTICA 12.0 (Stat Soft, Tulsa, USA) and GraphPad Prism version 6 (GraphPad Software, La Jolla California, USA). Comparisons between groups were performed so that the cut-off of the level of statistical significance was set as ($P < 0.05$)
For comparing any parameter (relative expression or methylation %) in CLL samples, unpaired t-test was performed with Welch´s correction. The experiments that were conducted using cell lines included at least three independent biological replicates and the data were represented as (Mean ± SD). For comparisons, two-tailed t-test was performed, and the replicates were considered of equal variance for calculating p-values.
For correlation analysis, Pearson correlation was performed on CLL patient samples to correlate two variables: (expression of EZH2 vs IGF1R; n=96). For the manuscript, Pearson correlation test was performed between -log10 relative expression values of *MNX1* and –log10 of intragenic methylation percentage of *MNX1* and the correlation was further supported by performing linear regression test, wherein methylation was set as independent variable (X) and expression as dependent (Y). Two-tailed t-test was performed to calculate p-value of Pearson correlation.

4 RESULTS AND DISCUSSION

4.1 Paper 1

Gene-body hypermethylation controlled cryptic promoter and miR26A1-dependent EZH2 regulation of TET1 gene activity in chronic lymphocytic leukemia:

A previous study of our group, by means of MBD-sequencing technique, has mapped the differential DNA methylation patterns in two subsets of chronic lymphocytic leukemia (CLL) patients (IGHV-mutated and unmutated CLL) in comparison to normal B cell counterparts that are sorted from peripheral blood samples of healthy donors.[120] This methylome analysis revealed sets of biologically relevant genes that are differentially hypermethylated in CLL versus healthy B cells. Out of these genes is Ten Eleven translocation 1 (*TET1*), the prominent eraser of 5-methyl cytosine that elicits active demethylation of DNA and under certain circumstances might potentiate passive demethylation as well. According to this data, the gene body region of TET1 was found to harbor higher degree of methylation in CLL; whilst the promoter does not show any differential pattern between CLL and normal B cells. The hypermethylated intragenic region of *TET1* in CLL that is identified by MBD-seq is referred to as HMR. Interestingly, analyzing a published RNA-sequencing dataset on a large cohort of CLL patients (n=96) and 9 healthy controls revealed that TET1 is overexpressed in CLL compared to normal counterparts (logFC ≈ 2; FDR < 0.001).
The differential methylation and expression patterns of *TET1* is validated in sorted B-CLL cells from an independent CLL sample cohort (n=40) versus sorted normal B cells from healthy donors (n=5) by Pyrosequencing and RT-PCR, respectively. Altogether, these findings suggest that the expression of *TET1* is likely to be positively regulated by intragenic DNA methylation, in agreement to what has been found by other studies regarding the relationship between gene body methylation and transcriptional activation.
To gain further insights on the mechanism underlying the transcriptional activation by gene body methylation, four leukemic cell lines were used as model system, wherein global demethylation was induced by 5-aza-2'-deoxycytidine (DAC) which competitively inhibit S-adenosyl methionine (SAM) binding at DNMTs. Upon DAC treatment in the cell lines, a significant reduction of the relative expression of *TET1* mRNA, as well as reduced protein level were observed; thus increasing the likelihood that gene body hypermethylation functions in transcriptional activation of *TET1*. Moreover, transcriptome analysis in CLL cell lines treated with DAC and negative control counterparts, showed a significant reduction of *TET1* upon DAC treatment (unpublished data). Intrigued by such finding, it was postulated that the HMR region of *TET1* might harbor a cryptic promoter activity, which might be associated with antisense transcription that overlap with or attenuate the corresponding *TET1* gene transcription. Cloning of this part of TET1 (519 bp) into a promoter-less PGL3 expression vector in both sense and antisense orientation revealed an exceedingly higher luciferase activity of the antisense orientation compared to a basic control vector, sense oriented HMR and SV40 as a positive control. This notion was then supported by the detection of intronic transcripts in B-CLL and normal B cells by RT-PCR using 7 sets of primers designed across HMR region.
Of note, DAC treatment of cell lines was found associated with a significant increase of this intronic transcription along with reduction of *TET1* mRNA, suggesting that methylation at this region contributes to activation of *TET1* gene expression by suppressing this supposed intronic transcription. The postulation was then that intronic transcription might overlap with the corresponding gene expression; however fine mapping of the intronic transcripts by RT-PCR and strand-specific PCR revealed that the intronic transcription is not detectable at the regions downstream to HMR, indicating that the intronic transcription is not overlapping with the

corresponding promoter of *TET1* and argue against the assumption that cryptic transcription converge and attenuate the corresponding transcription. In line with what has been viewed by other studies, this might be explained that the intronic transcription may be adversarial to transcriptional elongation; thereby it interferes with the corresponding transcript with no effect on the *TET1* promoter activity itself. Global DNA hypomethylation with regional hypermethylation has been reported in CLL.[120, 122, 123, 282] This might consistently match to our observation that is *TET1* overexpression in CLL, and might also explain one possible regulatory loop of active demethylation across the genome in CLL.

Conclusion:

Integration between our previous methylome dataset and a published transcriptome dataset of independent CLL cohorts, this study drew a functional link between intragenic DNA methylation and transcriptional activation of *TET1* gene in terms of controlling a cryptic intronic transcription. Loss of methylation within the gene body regions is associated with firing of cryptic promoter and in turn intronic transcription that interfere with the corresponding gene activity. Thus, explaining the implication of preferential gene body hypermethylation in increased *TET1* gene expression in CLL compared to normal B cellular counterparts.

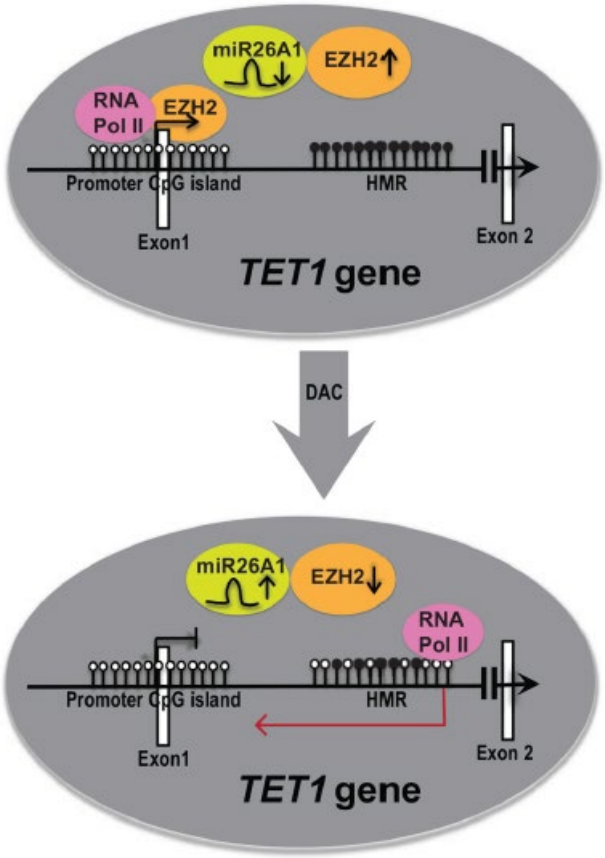

Figure 10. Graphical illustration explains the conclusion of paper 1.[281]

4.2 Paper 2

EZH2 upregulates the PI3K/AKT pathway through IGF1R and MYC in clinically aggressive chronic lymphocytic leukaemia:

Employing ChIP-seq technology, this study identified for the first time EZH2 target genes in two prognostic subsets of CLL, namely IGHV-mutated and unmutated CLL (M-CLL & U-CLL) patients. In addition, the genome-wide patterns of H3K27me3, the prototypical repressive mark of PRC2 has been identified in both subsets as well. By mapping and aligning EZH2-peak regions to the reference human genome (hg19), it was found that the majority of EZH2-target loci (63-67%) are annotated to promoter/TSS regions in both M-CLL and U-CLL patients, indicating a functional significance of EZH2 in transcriptional regulation in CLL. The similar pattern of EZH2 distribution might somehow indicate a unifying role of EZH2 in both prognostic subgroups; supported by the notion that 90% of the total EZH2-target genes are found overlapped between MCLL and UCLL. Intriguingly, upon overlapping total EZH2-target genes with H3K27me3 peaks with minimum 1 bp overlap, 2676 genes were obtained and are called EZH2-overlapped (EZH2-OP). The remaining 10239 genes (EZH2-NOP) were not overlapped with H3K27me3 according to the strategy we have set for overlapping with at least 1 bp overlap. This could be explained by the transient nature of PRC2 and its "hit-and-run" mode of action, along with the lack of consensus binding sites over human genome. [184, 191] Moreover, the finding that 80% of EZH2 peaks in CLL (EZH2-NOP) are not enriched with H3K27me3, suggests that EZH2 might function at certain loci independently of its being an integral part of PRC2.
The focus of the study was then narrowed to the genes that are enriched with both EZH2 and its associated mark H3K27me3 (EZH2-OP), so that the likelihood that these genes are *bona fide* targets of PRC2 is enclosed. The overlap of EZH2-taget genes between mutated and unmutated CLL has reduced from 90% that was observed in the total EZH2-targets in CLL to 45% in the filtered EZH2-OP list of genes; thus offered a counterargument to the tentative assumption that EZH2 exhibit unifying functionality in both subsets of CLL. It turned out that EZH2 shows differential pattern of distribution in both subsets, evidenced by the result that 44.5% of EZH2-OP are specifically enriched with EZH2 in U-CLL. These findings so far suggest a functional relevance of EZH2 enrichment at the ad hoc set of genes in either subgroup.
Analyzing a published RNA-seq datasets [283] revealed that approximately 30% of these EZH2-OP genes are differentially expressed between MCLL and UCLL, hence further supports the subdivision of EZH2 mode of action in the two prognostic subgroups. Interestingly, the majority of these differentially expressed genes is upregulated in U-CLL; thus, indicating an implication of EZH2 in transcriptional activation in U-CLL. Next, KEGG pathway analysis of EZH2-OP genes revealed several oncogenic pathways; among the top significant of which is PI3K/Akt pathway. Interestingly, PI3K/Akt pathway-clustered genes were shown to be differentially enriched with EZH2 between both subgroups. Furthermore, integrative analysis of our ChIP-seq data and the published RNA-seq data revealed some of the 86 PI3K-genes including *IGF1R* are found to be significantly enriched with EZH2 and overexpressed in the poor prognostic UCLL compared to MCLL. According to our ChIP seq, EZH2 occurs at *IGF1R* promoter in unmutated CLL but not in mutated and RNA-seq analysis showed a significant overexpression of *IGF1R* in UCLL compared to MCLL (LogFC 1.41; p value 0.0001). Intrigued by such notion, *IGF1R* was nominated as a model gene to probe a possible mechanism underlying a non-canonical implication of EZH2 in activating PI3K/AKT pathway-related genes in the CLL subgroup with worse prognosis. Using a large cohort of CLL (n=96), a positive correlation between relative expression of *EZH2* and *IGF1R* was observed ($r = 0.58$; p value < 0.0001); thus, the expression of *IGF1R* is more likely to be positively regulated by EZH2 in a PRC2-independent manner. This was evidenced by the significant reduced expression of *IGF1R* upon siRNA-mediated silencing of *EZH2* in CLL cell line (HG3). Moreover, the reduced enrichment of EZH2 at IGF1R was not associated by any significant change of H3K27me3, highlighting an HMT-independent mode of action of EZH2; this was evidenced by the insignificant effect of using SAM-inhibitors (GSK343 and UNC1999) on the level of phosphorylated AKT (p-AKT) in HG3 cell line.

Our findings also revealed that EZH2 recruits c-MYC at *IGF1R* promoter, which in turn supposedly functions through its trans-activating domain in active expression of *IGF1R*, evidenced by the significant *c-MYC*-siRNA-associated *IGF1R* reduction in a way comparable to the effect of EZH2. The recruitment of c-MYC by EZH2 was further confirmed by ChIP-PCR which showed nonsignificant effect of *c-MYC* silencing on EZH2 enrichment; and neither does EZH2 silencing result in affecting *c-MYC* at mRNA level, as revealed by expression analysis by RT-PCR.

Conclusion:

Our ChIP-seq data identified for the first time the target genes of EZH2 and the global pattern of H3K27me3 in two subsets of CLL and such pattern led to characterize a hitherto unknown mode of action of EZH2 in active gene expression independently on PRC2 in a clinically adverse subgroup of CLL. This role of EZH2 seems to be of biological relevance to cancer type or subtype, illustrated here by activating PI3K-related signature genes including *IGF1R* specifically unmutated CLL (UCLL), thereby it partially contribute to the disease aggressiveness. Understanding the non-canonical facets of EZH2 might pave the way for more effective therapeutic lines that target PRC2 and circumvent the resistance exhibited by many cancers to SAM-inhibitors that target the HMT activity of PRC2.

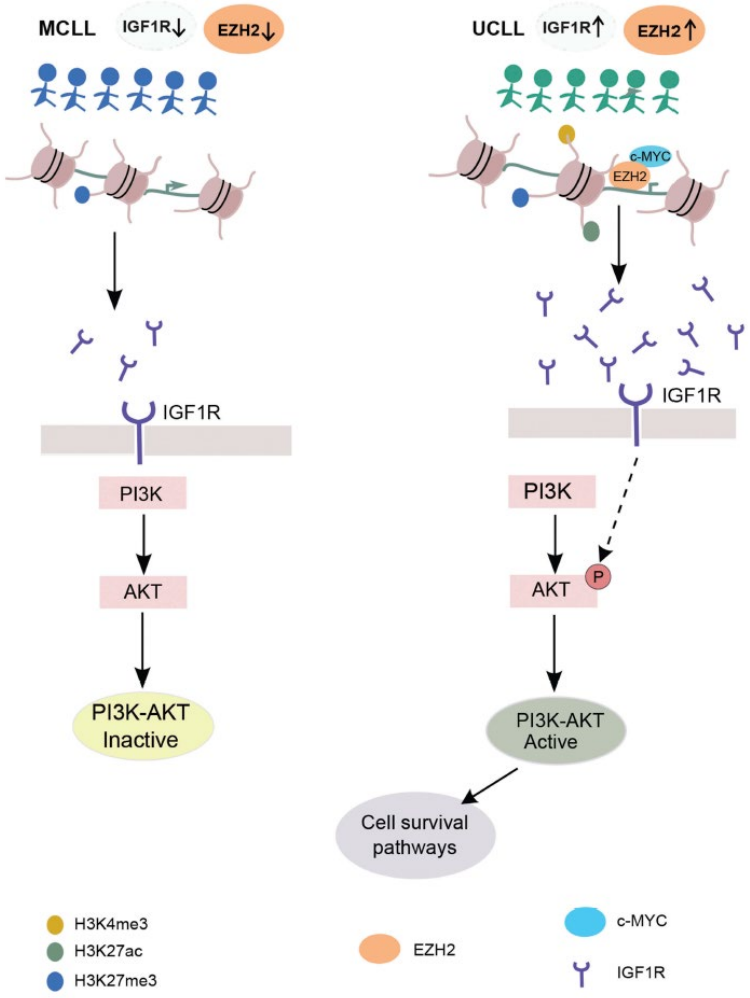

Figure 11. Graphical illustration explains the conclusion of paper 2.[272]

4.3 Paper 3 (manuscript)

DNA methylation at intragenic CpG islands controls PRC2-mediated transcriptional regulation of MNX1 in Chronic lymphocytic leukemia

This study is pertinent to paper 1 of this thesis, in the sense it is based on the global methylome data obtained from our previous study. The finding of paper 1 that there is a set of genes that are supposed to be upregulated by hypermethylation intrigued us to investigate other facet of the implication of methylation at intragenic CGIs in regulating gene expression in terms of the crosstalk with PRC2. In this regard, CLL cell lines were used as model system to dissect the interplay between DNA methylation and PRC2 within the intragenic context of the transcriptional unit upon global demethylation using DAC. Transcriptome analysis (RNA-seq) revealed that 35% of the differentially expressed genes are significantly downregulated in DAC-treated MEC1 cell lines compared to control counterparts. Molecular function analysis of this set of genes using GO terms for complete molecular function showed significant enrichment of DNA binding, transcriptional regulation and RNA polymerase II regulation and other relevant molecular processes.

Of these genes, Motor Neuron and Pancreas Homeobox (*MNX1*), which harbors high degree of methylation within gene body region according to MBD-seq data, was selected to follow the placement and function of PRC2 after removal of DNA methylation by DAC treatment (LogFC = -1.58; FDR= 4.10691690348078e-154). According to our recently published study, transcriptome analysis of CLL patients revealed a significant overexpression of *MNX1* in CLL patients compared to normal cells from healthy donors.[282]Validation of MBD-seq[120] and RNA-seq[282] was performed using an independent cohort of CLL patients (n = 20) and sorted normal B cells from healthy donors (n = 8) by Pyrosequencing and RT-PCR, respectively. Interestingly, intragenic methylation % at *MNX1* gene body showed a significant positive correlation to relative expression of *MNX1* ($r = 0.472$; $p < 0.05$) using the aforementioned independent CLL patient samples used for validation.

The relationship between intragenic DNA methylation at *MNX1* gene and PRC2 was then tested in CLL cell lines. Silencing of EZH2 in HG3 cell lines showed no significant effect on *MNX1* expression; nevertheless, this observation did not rule out the implication of PRC2 in regulating *MNX1* expression. Notably, silencing EZH2 in HG3 cells that were maintained in DAC-containing medium resulted in 45% in *MNX1* expression compared to DAC-free maintained HG3 cells ($p < 0.005$). In line with the downregulation of *MNX1* expression upon DAC treatment, ChIP experiments revealed significant increase of the enrichment of both EZH2 and its prototypical mark H3K27me3 at promoter region and the intragenic region of *MNX1* that was identified by MBD-seq analysis. This suggests that PRC2 in terms of occupancy and functionality of its catalytic subunit EZH2 is governed by the degree of intragenic CGIs methylation. Given that PRC2 has a preference towards locating unmethylated CGI, it is more likely that the relationship between intragenic CGIs methylation and PRC2 is adversarial. More tellingly, higher degrees of intragenic CGIs methylation might elicit a barrier that impede PRC2 binding to its target genomic loci, this was evidenced by the increase of EZH2 and H3K27me3 enrichment. Moreover the significant increase of *MNX1* expression upon EZH2 silencing in HG3 growing in DAC-supplemented medium, even though the effect of EZH2 silencing in control counterparts is insignificant supports the postulation that PRC2 is impeded by hypermethylation. Also, the significant increased enrichment of H3K27me1 after DAC treatment is suggested to be an intermediate product of PRC2 towards the ultimate trimethylation of H3K27 and point out to the processive mode of action of PRC2 that requires longer time for H3K27 trimethylation. Since monomethylation of H3K27 mark the bodies of actively transcribed genes [156, 157], H3K27me1 is not an ad hoc product of PRC2 in this context, as the expression of *MNX1* reduces upon DAC treatment. Furthermore, H3K36me3 enrichment across the gene body region that harbors hypermethylation showed no significant change upon DAC treatment; thus further support that H3K27me1 is an intermediate product of PRC2 matching the relatively short time if DAC treatment that covers only one doubling time of HG3 cells. Together, these results highlight the contextual mutual exclusivity of CGI methylation and PRC2, in the sense that PRC2 access to its

target genomic regions might be blocked by DNA methylation. This in part could be supported by the finding that most of the pluripotency-related genes that are pre-marked and silenced by PRC2 in embryonic stem cells are prone to hypermethylation in cancer, a process that mirrors the unrestricted proliferation during pluripotency.[135]

Conclusion:

The results of this work address an adversarial relationship between intragenic CGIs methylation, and PRC2 and give an insight to non-canonical contribution of DNA methylation in transcriptional activation of *MNX1* a putative oncogene. Herein, intragenic CGIs hypermethylation is found to render the expression of *MNX1* high in CLL, by eliciting a barrier that impedes the access of EZH2 the catalytic subunit of PRC2 and thus protects its expression form the spread of the repressive mode of action of PRC2.

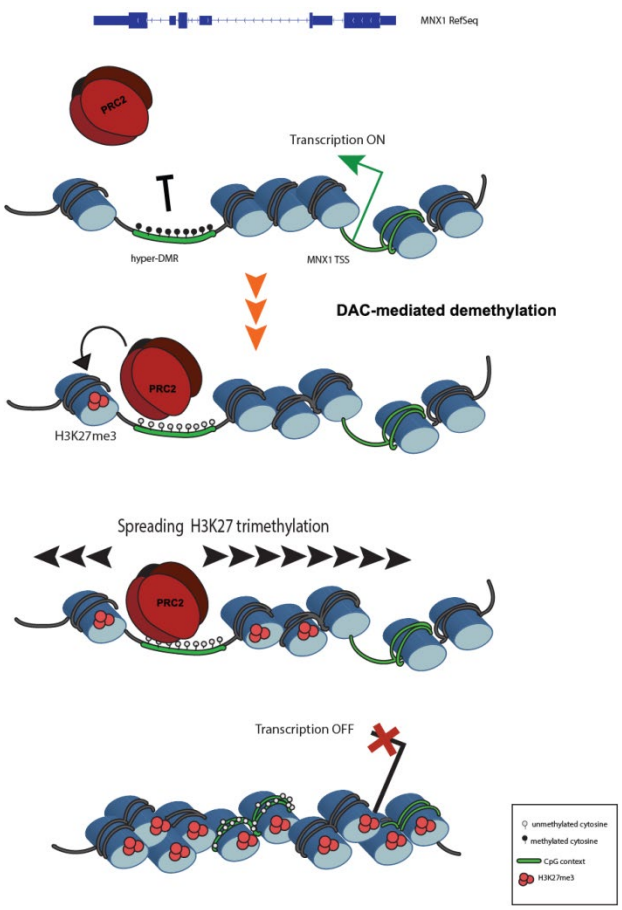

Figure 12. Graphical illustration explains the conclusion of the manuscript

5 FUTURE PERSPECTIVES

The post-sequencing era has witnessed a remarkable conceptual progress, which has in turn offered new insights on biological and physiological processes and has enriched understanding of the development and progression of diseases that jeopardize mankind. However, with the developments of efficient anticancer therapeutic strategies, still cancer represents a major threat to humankind all over the world.

Aberrant epigenome has been proven to embrace the generation of the notorious cytogenetic lesions that expedite tumor progression and account for poor prognoses of several human cancers. The modalities of epigenetic machineries that control DNA methylation and histone modifications are highly influenced by contexts, in particular cancer-type contexts; thus entailing complete understanding of the aspects of epigenetic mechanisms including the non-canonical modalities. For instance, the deregulation of Polycomb proteins is differential among different cancers; loss-of-function aberrations are documented in some malignancies, while gain-of-function are hallmarks of others. In addition, apart from mutations, EZH2 the catalytic core subunit of PRC2 throughout the past decade has been shown to function non-canonically and independently of being a substantial part of PRC2. Studies on breast cancer subtypes, T-cell lymphoma and poor prognostic prostate cancer subset revealed that EZH2 functions in transcriptional activation of some genes of relevance to the corresponding cancer type/subtype.

The development of therapeutic agents that target the PRC2-attributed HMT activity of EZH2 has proven high efficiency of treatment of several cancers and some of these agents have become in use for clinical trials. Nevertheless, these agents might not properly circumvent the HMT-independent oncogenic modalities of EZH2 in some cancers. Accordingly, it is of an utmost importance to investigate the non-canonical behavior of EZH2 in Pan cancer-based studies and identify the target genes and their relevance to the respective malignancies. Moreover, the domains or subdomains of EZH2 protein have to be subject of intense interrogation in order to dissect the involvement of EZH2 in transcriptional activation, in terms of:

- The domain(s) that contribute to enhancing gene expression of cancer-signature genes
- The interface between such domain(s) of EZH2 with the transcriptional machinery, other chromatin-associated factors, and chromatin remodelers and how this influences the shift of mode of EZH2 from trans-repression towards trans-activation
- How likely EZH2 assembles to or disassembles from the other core and non-core subunits of PRC2 genome-wide
- The target genes in each cancer type/subtype, and how likely that EZH2 non-canonically trans-activates universal oncogenes that generally contributes to cancer progression

Comprehensive understanding of all the possible aspects of PRC2 deregulation and pinpointing its hitherto uncharacterized oncogenic facets would offer on the long-term a compendium of the different PRC2 modalities in different cancers and that would participate in paving the way for devising alternative therapeutic lines that specifically target the cancer type/subtype-attributed modality of PRC2 deregulation and thereby improve the clinical outcomes of cancer patients. I would adopt such approach and line of research for my future career and I would consider investing the progressive advancement of CRISPR/Cas9 technologies, life cell imaging, proteomics, and bioinformatics along with high through-put based sequencing techniques for this short-term endeavor.

ACKNOWLEDGEMENT

It has been such a good life experience to live in Sweden and conduct my PhD study at Gothenburg University, which allowed me to circle Europe for participating in conferences and meet people from every walk of life. During the whole four years since I moved here, I have perceived Sweden as a nice and calm place to live in; and the people are cultured enough to easily deal with and interact to. I did not find any barrier to live normally and manage my life smoothly in Gothenburg, where some good moments and extraordinary dreams found their way to me.

First of all, I would like to express my gratitude to my supervisor **Dr. Meena Kanduri** for giving me the opportunity to join her group, for supervising my thesis, for the discussions, and for her sustainable optimism that made the hard times much easier than it would have actually been. It was actually a good experience that has opened my eyes to the contemporary revolution in the field epigenetics and chromatin.

I was really lucky to work at the department of clinical chemistry and transfusion medicine, a highly organized place, nice and supportive people around; and it was of a great pleasure to get to know to a refined gentleman like **Professor. Göran Larson**, the head of the department. I still remember that day when I first joined and my supervisor **Meena Kanduri** introduced me to him, he was such a hospitable person and such hospitality was an important factor that empowered me to overcome the overwhelming feeling of alienation anyone might experience during the first days in a new place.

It is of my deep pleasure to thank my co-supervisor **Prof. Lars Palmqvist**, I still remember the seminars that were held for the journal clubs and the Fridays´ lunch seminars and it was a good opportunity to present my project and hear the feedback from **Lars** and his group members, it was a good chance to get in contact to the clinical perspective upon participating in these seminars.

I would also say that it was so influential to get acquainted to **Prof. Linda Fogelstrand**, she is such a clever, enthusiastic and sociable scientist that is capable of explaining a highly complicated clinical issue in a nutshell and so many thanks to her and **Prof. Lars Palmqvist** for organizing the scientific day that was held at ChalmerskaHuset and was sort of a kick-off event where everybody participated and though loudly so that we were all connected together in the sense of what we are doing and what we were thinking of.

I would also like to thank my colleague **Pradeep Kumar Kopparapu** who was part of this team, and who took care of my practice in the cell laboratory and how to handle and manage the work in many units in our department. I will also remember the good times i have spent having coffee and chatting with my colleague **Mukesh Yadav Pratap** who has joined our group in 2018.

As I mentioned, it was a milestone in my life to join the department of clinical chemistry and transfusion medicine and meet different people, of different backgrounds and different expertise, yet they all agreed on sincere work for keeping everything in order. Thanks to **Camilla**, **Marianne**, **Maria O**, **Maria H**, **Marion**, **Josefine**, **Aida**, **Pegah**, **Alexandra**, **Petra**, **Lotta**, **Petronella**, **Elisa**, **Susann**, **Carina**, **Narmin** and **Ondina Tomar** who used to make our cell culture lab perfect.

A well organized and willful person, **Laleh Arabanian**, it was nice to get to know to you.

Almost all the time, I am sitting with him, in the same office, having lunch together, having coffee and chat, sharing some points of view in science and other things. **Ahmed Waraky**, it was nice to have you an officemate and a colleague and thanks for those times you helped me with the FACS and it is nice to be involved in part of your work as well.

I also thank **Christina Nilson** (**Tina**) for guiding everyone and her sincerity to provide a highly organized cell culture lab and perfect environment for proper culture-based work, for always finding solutions to any crisis that encounters either cell lab or Stora lab, and also for guidance to ensure the effectiveness of the techniques that are used by almost everybody in the department.

No one would even forget the highly skilled manager who is always able to organize, lead, find solutions and tackle problems smoothly, **Cecilia Boreström** the chef of the unit; So many thanks for helping everyone.

So many thanks to meet my colleagues **Erik Malmberg**, **Faisal**, **Rakesh**, **Mahnaz**, **Jasmine**, **Laia, Anders** and special thanks to the supportive and cheerful **Tugce Satir** who is always capable of

making the good times and organize the gathering with the rest of the colleagues at wonderful places and throwing the parties at her place, she used to gather us always.
It is also of my great pleasure to acknowledge the **Egyptian ministry of higher education** and the **Mission sector** for supporting me through granting a scholarship in July 2015 for studying at Gothenburg University and getting my PhD degree; and my sincere gratitude to **Professor. Ahmed Farouk Ghoneim**, the Egyptian cultural counselor in Berlin and all the staff members of the **Egyptian cultural Bureau in Berlin**, especially **Mrs.Nabila Hassan** for taking care of the logistic work and making contact and practicing as focal point between me and the Ministry in Egypt and for their commitment support to deliver my salaries in time, receiving my queries and give answers and for assuring me and make sure that I am OK either by email or phone conversation, in particular when the terroristic episodes sadly took place in Stockholm in 2017.
I also acknowledge **Assar Gabrielsson's foundation for the research grants** (**AG fonden)** for the grant that I have got in December 2018 which helped me to do some substantial parts of my work and covered my participation in EMBL conference in Germany in November 2019. And again thanks to my supervisor **Meena Kanduri** for encouraging me to apply to this grant, revising the proposal and cheering me up when I finally got it.
It is of pleasure also to acknowledge the contribution of **Professor. Chandrasehkar Kanduri**, for his cooperation and participation to our studies, and above that, it was a good opportunity for me to join the weekly journal club and also participate in giving seminars. And also my gratitude to his lab members, **Dr.Subazini** who helped us a lot with the bioinformatics analysis especially the second publication, **Santhilal** who also participated in the bioinformatics analysis of a former study in collaboration with **Meena** and based on the findings of this study I came with the idea of my last project along with **Meena**. I also thank **Sara**, **Luisa**, **Lily**, **Caroline**, **Prasanna**, **Sanhita**, **Silke**, **Sagar**, **Daniel, Tanushree** and **Mirko**. No one would even dare to forget the influential **Kankadeb Mishra** a former member of this lab but a not-to-be forgotten person, even when he left us and flew to New York to get started with his post-PhD career. The big brother, as everyone used to call him, **Dr.Tanmoy Mondal**, he is such a kind person, diplomatic and sober for the most part and his feedback, discussion in the journal clubs is really fruitful and motivating one to know more and more and develop the critical thinking, It was of my big pleasure to get to know to you big bother.
Still we are in the confines of Chandra´s lab, but this time I am talking about closest among them to me **Mohamad Gendy** whom I knew about 15 years back, since the time we were studying at the Faculty of Science-Alexandria University in Egypt. He was the one, who made a contact between me and my current supervisor **Meena**. Thanks for this, thanks for the good times at your home, with your family, **Walaa** and my little friend **Noor**, without you guys, life would have been much more boring and meaningless. Thanks for the chats, travel together, having lunch together, thanks for helping in the bioinformatics analysis work for the third manuscript. I would also love to express my pleasure to get to know to **Wassim** and **Laila** and the little **Karl** and thank them for the good times I have had when we all gathered along with the **Gendys**.
Last but not the least, my whole gratitude to my Family, by which I have always been backed up. My Mother **Sanaa Badr-Eldeen**, who I used to look at my life through her eyes, the one that I have been tethered to since the moment I was born of her, until now, and forever long. I have never found such a strong woman like you and seriously it would take me an unlimited number of pages to talk about you. **Marwan Hamdy**, my only brother, you have been such a support, a real man. It is true that you are my younger brother; nevertheless, I have never felt that you are the younger; you are rather much more durable than I am. In spite of what the three of us have seen throughout our life, what we faced, what we experienced, what we dreamed of, the good moments, the harsh times and bad memories, we still support each other, we are still there. And as long as we are together, clenching to our hope and perseverance, recalling our conviction that something extraordinary is possible, we have fairly adequate reasons to continue our lives. Love you both; I am a no body without you.
I think I have got an opportunity to get into a new environment, see things, learn things, and know what I haven´t even imagined I will. Again and again thanks to everyone who has been to my life.

REFERENCES

1. Crick F: **Central dogma of molecular biology**. *Nature* 1970, **227**(5258):561-563.
2. Taft RJ, Pheasant M, Mattick JS: **The relationship between non-protein-coding DNA and eukaryotic complexity**. *BioEssays : news and reviews in molecular, cellular and developmental biology* 2007, **29**(3):288-299.
3. Swift H: **The constancy of desoxyribose nucleic acid in plant nuclei**. *Proceedings of the National Academy of Sciences of the United States of America* 1950, **36**(11):643-654.
4. Bird AP: **Gene number, noise reduction and biological complexity**. *Trends in genetics : TIG* 1995, **11**(3):94-100.
5. Thomas CA, Jr.: **The genetic organization of chromosomes**. *Annual review of genetics* 1971, **5**:237-256.
6. Vinogradov AE: **Genome size and chromatin condensation in vertebrates**. *Chromosoma* 2005, **113**(7):362-369.
7. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD: **Eukaryotic genome size databases**. *Nucleic acids research* 2007, **35**(Database issue):D332-338.
8. Ohno S: **So much "junk" DNA in our genome**. *Brookhaven symposia in biology* 1972, **23**:366-370.
9. Doolittle WF, Sapienza C: **Selfish genes, the phenotype paradigm and genome evolution**. *Nature* 1980, **284**(5757):601-603.
10. Hahn MW, Wray GA: **The g-value paradox**. *Evolution & development* 2002, **4**(2):73-75.
11. Mattick JS: **Introns: evolution and function**. *Current opinion in genetics & development* 1994, **4**(6):823-831.
12. Bird A: **Perceptions of epigenetics**. *Nature* 2007, **447**(7143):396-398.
13. Shen H, Laird PW: **Interplay between the cancer genome and epigenome**. *Cell* 2013, **153**(1):38-55.
14. Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**(1):57-70.
15. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.
16. Egeblad M, Nakasone ES, Werb Z: **Tumors as Organs: Complex Tissues that Interface with the Entire Organism**. *Developmental Cell* 2010, **18**(6):884-901.
17. Pietras K, Östman A: **Hallmarks of cancer: Interactions with the tumor stroma**. *Experimental Cell Research* 2010, **316**(8):1324-1331.
18. Sherr CJ, DePinho RA: **Cellular Senescence: Minireview Mitotic Clock or Culture Shock?** *Cell* 2000, **102**(4):407-410.
19. Artandi SE, DePinho RA: **Telomeres and telomerase in cancer**. *Carcinogenesis* 2009, **31**(1):9-18.
20. Negrini S, Gorgoulis VG, Halazonetis TD: **Genomic instability — an evolving hallmark of cancer**. *Nature Reviews Molecular Cell Biology* 2010, **11**:220.
21. Luo J, Solimini NL, Elledge SJ: **Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction**. *Cell* 2009, **136**(5):823-837.
22. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A: **Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability**. *Carcinogenesis* 2009, **30**(7):1073-1081.
23. Choi JD, Lee J-S: **Interplay between Epigenetics and Genetics in Cancer**. *Genomics Inform* 2013, **11**(4):164-173.
24. Jones PA: **Functions of DNA methylation: islands, start sites, gene bodies and beyond**. *Nature Reviews Genetics* 2012, **13**:484.
25. You Jueng S, Jones Peter A: **Cancer Genetics and Epigenetics: Two Sides of the Same Coin?** *Cancer cell* 2012, **22**(1):9-20.
26. Hodgkinson A, Eyre-Walker A: **Variation in the mutation rate across mammalian genomes**. *Nature Reviews Genetics* 2011, **12**:756.
27. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D: **CpG island methylator phenotype underlies**

sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature genetics* 2006, **38**(7):787.
28. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa J-PJ: **CpG island methylator phenotype in colorectal cancer**. *Proceedings of the National Academy of Sciences* 1999, **96**(15):8681-8686.
29. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schübeler D: **Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors**. *Molecular cell* 2008, **30**(6):755-766.
30. Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G: **Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation**. *Cell stem cell* 2008, **2**(2):160-169.
31. McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra A, 3rd, Diaz E *et al*: **EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations**. *Nature* 2012, **492**(7427):108-112.
32. Beguelin W, Popovic R, Teater M, Jiang Y, Bunting KL, Rosen M, Shen H, Yang SN, Wang L, Ezponda T *et al*: **EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation**. *Cancer cell* 2013, **23**(5):677-692.
33. Lee C-H, Yu J-R, Kumar S, Jin Y, LeRoy G, Bhanu N, Kaneko S, Garcia BA, Hamilton AD, Reinberg D: **Allosteric Activation Dictates PRC2 Activity Independent of Its Recruitment to Chromatin**. *Molecular cell* 2018, **70**(3):422-434.e426.
34. Bernt KM, Zhu N, Sinha AU, Vempati S, Faber J, Krivtsov AV, Feng Z, Punt N, Daigle A, Bullinger L *et al*: **MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L**. *Cancer cell* 2011, **20**(1):66-78.
35. Chen S, Yang Z, Wilkinson AW, Deshpande AJ, Sidoli S, Krajewski K, Strahl BD, Garcia BA, Armstrong SA, Patel DJ *et al*: **The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79**. *Molecular cell* 2015, **60**(2):319-327.
36. Erb MA, Scott TG, Li BE, Xie H, Paulk J, Seo HS, Souza A, Roberts JM, Dastjerdi S, Buckley DL *et al*: **Transcription control by the ENL YEATS domain in acute leukaemia**. *Nature* 2017, **543**(7644):270-274.
37. Olivier M, Hollstein M, Hainaut P: **TP53 mutations in human cancers: origins, consequences, and clinical use**. *Cold Spring Harbor perspectives in biology* 2010, **2**(1):a001008.
38. Choi JD, Lee JS: **Interplay between Epigenetics and Genetics in Cancer**. *Genomics Inform* 2013, **11**(4):164-173.
39. Waddington CH: **The epigenotype. 1942**. *International journal of epidemiology* 2012, **41**(1):10-13.
40. Goldberg AD, Allis CD, Bernstein E: **Epigenetics: A Landscape Takes Shape**. *Cell* 2007, **128**(4):635-638.
41. Riggs AD: **X inactivation, differentiation, and DNA methylation**. *Cytogenetics and cell genetics* 1975, **14**(1):9-25.
42. Brien GL, Valerio DG, Armstrong SA: **Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs**. *Cancer cell* 2016, **29**(4):464-476.
43. Zhou VW, Goren A, Bernstein BE: **Charting histone modifications and the functional organization of mammalian genomes**. *Nature Reviews Genetics* 2010, **12**:7.
44. Nilsen TW, Graveley BR: **Expansion of the eukaryotic proteome by alternative splicing**. *Nature* 2010, **463**(7280):457.
45. Schones DE, Zhao K: **Genome-wide approaches to studying chromatin modifications**. *Nature Reviews Genetics* 2008, **9**(3):179.
46. Felsenfeld G, Groudine M: **Controlling the double helix**. *Nature* 2003, **421**(6921):448.
47. Zhou BR, Bai Y: **Chromatin structures condensed by linker histones**. *Essays in biochemistry* 2019, **63**(1):75-87.
48. Grigoryev SA, Schubert M: **Unraveling the multiplex folding of nucleosome chains in higher order chromatin**. *Essays in biochemistry* 2019, **63**(1):109-121.
49. Kornberg RD: **Chromatin Structure: A Repeating Unit of Histones and DNA**. *Science (New York, NY)* 1974, **184**(4139):868.
50. Wolffe AP: **Gene Regulation: Insulating chromatin**. *Current Biology* 1994, **4**(1):85-87.

51. Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B: **Xist RNA and the mechanism of X chromosome inactivation**. *Annual review of genetics* 2002, **36**(1):233-278.
52. Dimitri P, Caizzi R, Giordano E, Carmela Accardo M, Lattanzi G, Biamonti G: **Constitutive heterochromatin: a surprising variety of expressed sequences**. *Chromosoma* 2009, **118**(4):419-435.
53. Brown SW: **Heterochromatin**. *Science (New York, NY)* 1966, **151**(3709):417-425.
54. Reveron-Gomez N, Gonzalez-Aguilera C, Stewart-Morgan KR, Petryk N, Flury V, Graziano S, Johansen JV, Jakobsen JS, Alabert C, Groth A: **Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication**. *Molecular cell* 2018, **72**(2):239-249.e235.
55. Hyun K, Jeon J, Park K, Kim J: **Writing, erasing and reading histone lysine methylations**. *Experimental & molecular medicine* 2017, **49**(4):e324.
56. Fortschegger K, de Graaf P, Outchkourov NS, van Schaik FM, Timmers HT, Shiekhattar R: **PHF8 targets histone methylation and RNA polymerase II to activate transcription**. *Molecular and cellular biology* 2010, **30**(13):3286-3298.
57. Feng W, Yonezawa M, Ye J, Jenuwein T, Grummt I: **PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation**. *Nature structural & molecular biology* 2010, **17**(4):445-450.
58. Kaneko S, Son J, Bonasio R, Shen SS, Reinberg D: **Nascent RNA interaction keeps PRC2 activity poised and in check**. *Genes & development* 2014, **28**(18):1983-1988.
59. Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E *et al*: **MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures**. *Nature communications* 2015, **6**:7743.
60. Beltran M, Yates CM, Skalska L, Dawson M, Reis FP, Viiri K, Fisher CL, Sibley CR, Foster BM, Bartke T *et al*: **The interaction of PRC2 with RNA or chromatin s mutually antagonistic**. *Genome Research* 2016, **26**(7):896-907.
61. Subhash S, Mishra K, Akhade VS, Kanduri M, Mondal T, Kanduri C: **H3K4me2 and WDR5 enriched chromatin interacting long non-coding RNAs maintain transcriptionally competent chromatin at divergent transcriptional units**. *Nucleic acids research* 2018, **46**(18):9384-9400.
62. Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F: **R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters**. *Molecular cell* 2012, **45**(6):814-825.
63. Bird A: **DNA methylation patterns and epigenetic memory**. *Genes & development* 2002, **16**(1):6-21.
64. Deaton AM, Bird A: **CpG islands and the regulation of transcription**. *Genes & development* 2011, **25**(10):1010-1022.
65. Zabet NR, Catoni M, Prischi F, Paszkowski J: **Cytosine methylation at CpCpG sites triggers accumulation of non-CpG methylation in gene bodies**. *Nucleic acids research* 2017, **45**(7):3777-3784.
66. Ausio J, Georgel PT: **MeCP2 and CTCF: enhancing the cross-talk of silencers**. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 2017, **95**(6):593-608.
67. Shin HY: **The structural and functional roles of CTCF in the regulation of cell type-specific and human disease-associated super-enhancers**. *Genes & genomics* 2019, **41**(3):257-265.
68. Ma Z, Li M, Roy S, Liu KJ, Romine ML, Lane DC, Patel SK, Cai HN: **Chromatin boundary elements organize genomic architecture and developmental gene regulation in Drosophila Hox clusters**. *World journal of biological chemistry* 2016, **7**(3):223-230.
69. Chory EJ, Calarco JP, Hathaway NA, Bell O, Neel DS, Crabtree GR: **Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome**. *Molecular cell* 2019, **73**(1):61-72.e63.
70. Taneja N, Zofall M, Balachandran V, Thillainadesan G, Sugiyama T, Wheeler D, Zhou M, Grewal SI: **SNF2 Family Protein Fft3 Suppresses Nucleosome Turnover to Promote Epigenetic Inheritance and Proper Replication**. *Molecular cell* 2017, **66**(1):50-62.e56.
71. Stewart-Morgan KR, Reverón-Gómez N, Groth A: **Transcription Restart Establishes Chromatin Accessibility after DNA Replication**. *Molecular cell* 2019, **75**(2):408-414.

72. Petryk N, Dalby M, Wenger A, Stromme CB, Strandsby A, Andersson R, Groth A: **MCM2 promotes symmetric inheritance of modified histones during DNA replication**. *Science (New York, NY)* 2018, **361**(6409):1389-1392.
73. Alabert C, Jasencakova Z, Groth A: **Chromatin Replication and Histone Dynamics**. In: *DNA Replication: From Old Principles to New Discoveries*. edn. Edited by Masai H, Foiani M. Singapore: Springer Singapore; 2017: 311-333.
74. Cedar H, Bergman Y: **Linking DNA methylation and histone modification: patterns and paradigms**. *Nature reviews Genetics* 2009, **10**(5):295-304.
75. Payer B, Lee JT: **X chromosome dosage compensation: how mammals keep the balance**. *Annual review of genetics* 2008, **42**:733-772.
76. Klose RJ, Bird AP: **Genomic DNA methylation: the mark and its mediators**. *Trends in biochemical sciences* 2006, **31**(2):89-97.
77. Holliday R, Pugh JE: **DNA modification mechanisms and gene activity during development**. *Science (New York, NY)* 1975, **187**(4173):226-232.
78. Jones PA: **Functions of DNA methylation: islands, start sites, gene bodies and beyond**. *Nature reviews Genetics* 2012, **13**(7):484-492.
79. Meng W, Jiang Y, Ma J: **Is the prognostic significance of O6-methylguanine- DNA methyltransferase promoter methylation equally important in glioblastomas of patients from different continents? A systematic review with meta-analysis**. *Cancer management and research* 2017, **9**:411-425.
80. Chen Y, Hu F, Zhou Y, Chen W, Shao H, Zhang Y: **MGMT promoter methylation and glioblastoma prognosis: a systematic review and meta-analysis**. *Archives of medical research* 2013, **44**(4):281-290.
81. Edwards JR, Yarychkivska O, Boulard M, Bestor TH: **DNA methylation and DNA methyltransferases**. *Epigenetics & chromatin* 2017, **10**(1):23.
82. Aoki A, Suetake I, Miyagawa J, Fujio T, Chijiwa T, Sasaki H, Tajima S: **Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases**. *Nucleic acids research* 2001, **29**(17):3506-3512.
83. Razin A, Szyf M: **DNA methylation patterns. Formation and function**. *Biochimica et biophysica acta* 1984, **782**(4):331-342.
84. Razin A, Riggs AD: **DNA methylation and gene function**. *Science (New York, NY)* 1980, **210**(4470):604-610.
85. Bronner C, Alhosin M, Hamiche A, Mousli M: **Coordinated Dialogue between UHRF1 and DNMT1 to Ensure Faithful Inheritance of Methylated DNA Patterns**. *Genes (Basel)* 2019, **10**(1).
86. Nishiyama A, Yamaguchi L, Nakanishi M: **Regulation of maintenance DNA methylation via histone ubiquitylation**. *Journal of biochemistry* 2016, **159**(1):9-15.
87. Bronner C, Krifa M, Mousli M: **Increasing role of UHRF1 in the reading and inheritance of the epigenetic code as well as in tumorogenesis**. *Biochemical pharmacology* 2013, **86**(12):1643-1649.
88. Jeltsch A, Jurkowska RZ: **New concepts in DNA methylation**. *Trends Biochem Sci* 2014, **39**(7):310-318.
89. Gowher H, Jeltsch A: **Mammalian DNA methyltransferases: new discoveries and open questions**. *Biochemical Society transactions* 2018, **46**(5):1191-1202.
90. Franklin D: **P152R Mutation Within MeCP2 Can Cause Loss of DNA-Binding Selectivity**. *Interdisciplinary sciences, computational life sciences* 2019, **11**(1):10-20.
91. Lee CH, Shih YP, Ho MR, Wang AH: **The C-terminal D/E-rich domain of MBD3 is a putative Z-DNA mimic that competes for Zalpha DNA-binding activity**. *Nucleic acids research* 2018, **46**(22):11806-11821.
92. Nan X, Cross S, Bird A: **Gene silencing by methyl-CpG-binding proteins**. *Novartis Foundation symposium* 1998, **214**:6-16; discussion 16-21, 46-50.
93. Jorgensen HF, Bird A: **MeCP2 and other methyl-CpG binding proteins**. *Mental retardation and developmental disabilities research reviews* 2002, **8**(2):87-93.
94. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T: **The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation**. *The Journal of biological chemistry* 2003, **278**(6):4035-4040.

95. Zhenwei J, Shuxin G, Yongchun Z, Xianhua Z: **Mechanisms of TET protein-mediated DNA demethylation and its role in the regulation of mouse development**. *Yi chuan = Hereditas* 2015, **37**(1):34-40.
96. Ko M, An J, Pastor WA, Koralov SB, Rajewsky K, Rao A: **TET proteins and 5-methylcytosine oxidation in hematological cancers**. *Immunological reviews* 2015, **263**(1):6-21.
97. Zhao H, Chen T: **Tet family of 5-methylcytosine dioxygenases in mammalian development**. *Journal of human genetics* 2013, **58**(7):421-427.
98. Shen L, Song CX, He C, Zhang Y: **Mechanism and function of oxidative reversal of DNA and RNA methylation**. *Annual review of biochemistry* 2014, **83**:585-614.
99. Jones PA, Liang G: **Rethinking how DNA methylation patterns are maintained**. *Nature Reviews Genetics* 2009, **10**:805.
100. Tazi J, Bird A: **Alternative chromatin structure at CpG islands**. *Cell* 1990, **60**(6):909-920.
101. Schones DE, Cui K, Cuddapah S, Roh T-Y, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome**. *Cell* 2008, **132**(5):887-898.
102. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST: **A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling**. *Cell* 2009, **138**(1):114-128.
103. Choi JK: **Contrasting chromatin organization of CpG islands and exons in the human genome**. *Genome biology* 2010, **11**(7):R70.
104. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA: **A chromatin landmark and transcription initiation at most promoters in human cells**. *Cell* 2007, **130**(1):77-88.
105. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD *et al*: **DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA**. *Nature* 2007, **448**(7154):714-717.
106. Okitsu CY, Hsieh CL: **DNA methylation dictates histone H3K4 methylation**. *Molecular and cellular biology* 2007, **27**(7):2746-2757.
107. Singh P, Li AX, Tran DA, Oates N, Kang ER, Wu X, Szabo PE: **De novo DNA methylation in the male germ line occurs by default but is excluded at sites of H3K4 methylation**. *Cell reports* 2013, **4**(1):205-219.
108. Lock LF, Takagi N, Martin GR: **Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation**. *Cell* 1987, **48**(1):39-46.
109. Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS, Bock C, Vasanthakumar A, Gu H, Xi Y: **Dnmt3a is essential for hematopoietic stem cell differentiation**. *Nature genetics* 2012, **44**(1):23.
110. Hellman A, Chess A: **Gene body-specific methylation on the active X chromosome**. *Science (New York, NY)* 2007, **315**(5815):1141-1143.
111. Larsen F, Solheim J, Prydz H: **A methylated CpG island 3'in the apolipoprotein-E gene does not repress its transcription**. *Human molecular genetics* 1993, **2**(6):775-780.
112. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP: **Orphan CpG islands identify numerous conserved promoters in the mammalian genome**. *PLoS genetics* 2010, **6**(9):e1001134.
113. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME *et al*: **Conservation and divergence of methylation patterning in plants and animals**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(19):8689-8694.
114. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y: **Conserved role of intragenic DNA methylation in regulating alternative promoters**. *Nature* 2010, **466**(7303):253.
115. Nguyen CT, Gonzales FA, Jones PA: **Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation**. *Nucleic acids research* 2001, **29**(22):4598-4606.
116. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs**. *Cell* 2007, **129**(7):1311-1323.

117. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S *et al*: **Widespread transcription at neuronal activity-regulated enhancers**. *Nature* 2010, **465**:182.
118. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S: **CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing**. *Nature* 2011, **479**(7371):74-79.
119. Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP: **Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks**. *PloS one* 2011, **6**(4):e18844.
120. Subhash S, Andersson PO, Kosalai ST, Kanduri C, Kanduri M: **Global DNA methylation profiling reveals new insights into epigenetically deregulated protein coding and long noncoding RNAs in CLL**. *Clinical epigenetics* 2016, **8**:106.
121. Martinelli S, Kanduri M, Maffei R, Fiorcari S, Bulgarelli J, Marasca R, Rosenquist R: **ANGPT2 promoter methylation is strongly associated with gene expression and prognosis in chronic lymphocytic leukemia**. *Epigenetics* 2013, **8**(7):720-729.
122. Kanduri M, Marincevic M, Halldorsdottir AM, Mansouri L, Junevik K, Ntoufa S, Kultima HG, Isaksson A, Juliusson G, Andersson PO *et al*: **Distinct transcriptional control in major immunogenetic subsets of chronic lymphocytic leukemia exhibiting subset-biased global DNA methylation profiles**. *Epigenetics* 2012, **7**(12):1435-1442.
123. Kanduri M, Cahill N, Goransson H, Enstrom C, Ryan F, Isaksson A, Rosenquist R: **Differential genome-wide array-based methylation profiles in prognostic subsets of chronic lymphocytic leukemia**. *Blood* 2010, **115**(2):296-305.
124. Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, Martinez-Trillos A, Castellano G, Brun-Heath I, Pinyol M *et al*: **Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia**. *Nat Genet* 2012, **44**(11):1236-1242.
125. Hinoue T, Weisenberger DJ, Lange CP, Shen H, Byun H-M, Van Den Berg D, Malik S, Pan F, Noushmehr H, van Dijk CM: **Genome-scale analysis of aberrant DNA methylation in colorectal cancer**. *Genome research* 2012, **22**(2):271-282.
126. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP: **Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma**. *Cancer cell* 2010, **17**(5):510-522.
127. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AW, Lu C, Ward PS: **IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype**. *Nature* 2012, **483**(7390):479.
128. Su J, Huang YH, Cui X, Wang X, Zhang X, Lei Y, Xu J, Lin X, Chen K, Lv J *et al*: **Homeobox oncogene activation by pan-cancer DNA hypermethylation**. *Genome biology* 2018, **19**(1):108.
129. Jeong M, Sun D, Luo M, Huang Y, Challen GA, Rodriguez B, Zhang X, Chavez L, Wang H, Hannah R *et al*: **Large conserved domains of low DNA methylation maintained by Dnmt3a**. *Nat Genet* 2014, **46**(1):17-23.
130. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D *et al*: **Epigenomic analysis of multilineage differentiation of human embryonic stem cells**. *Cell* 2013, **153**(5):1134-1148.
131. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB *et al*: **Genome-scale DNA methylation maps of pluripotent and differentiated cells**. *Nature* 2008, **454**:766.
132. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I: **Epigenetic stem cell signature in cancer**. *Nature genetics* 2007, **39**(2):157.
133. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W: **A stem cell–like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing**. *Nature genetics* 2007, **39**(2):237.
134. Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, Pikarski E, Young RA, Niveleau A, Cedar H: **Evidence for an instructive mechanism of de novo methylation in cancer cells**. *Nature genetics* 2006, **38**(2):149.

135. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE: **Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer**. *Nature genetics* 2007, **39**(2):232.
136. Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM: **Reprogramming towards pluripotency requires AID-dependent DNA demethylation**. *Nature* 2010, **463**(7284):1042-1047.
137. Bannister AJ, Kouzarides T: **Regulation of chromatin by histone modifications**. *Cell research* 2011, **21**(3):381.
138. Kouzarides T: **Chromatin modifications and their function**. *Cell* 2007, **128**(4):693-705.
139. Izzo A, Schneider R: **Chatting histone modifications in mammals**. *Briefings in functional genomics* 2010, **9**(5-6):429-443.
140. Banerjee T, Chakravarti D: **A peek into the complex realm of histone phosphorylation**. *Molecular and cellular biology* 2011, **31**(24):4858-4873.
141. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh T-Y, Peng W, Zhang MQ: **Combinatorial patterns of histone acetylations and methylations in the human genome**. *Nature genetics* 2008, **40**(7):897.
142. Barnes CE, English DM, Cowley SM: **Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription**. *Essays in biochemistry* 2019, **63**(1):97-107.
143. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri PC, Harte PJ: **CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing**. *Development* 2009, **136**(18):3131-3141.
144. Hebbes TR, Thorne AW, Crane-Robinson C: **A direct link between core histone acetylation and transcriptionally active chromatin**. *The EMBO journal* 1988, **7**(5):1395-1402.
145. Martin C, Zhang Y: **The diverse functions of histone lysine methylation**. *Nature reviews Molecular cell biology* 2005, **6**(11):838-849.
146. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome**. *Nature genetics* 2007, **39**(3):311.
147. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T: **Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins**. *Nature* 2001, **410**(6824):116.
148. Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, Perez-Burgos L, Kohlmaier A, Opravil S, Tachibana M, Shinkai Y: **Partitioning and plasticity of repressive histone methylation states in mammalian chromatin**. *Molecular cell* 2003, **12**(6):1577-1589.
149. Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, Shinkai Y, Allis CD: **Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains**. *Molecular cell* 2003, **12**(6):1591-1598.
150. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD *et al*: **Regulation of chromatin structure by site-specific histone H3 methyltransferases**. *Nature* 2000, **406**(6796):593-599.
151. Dillon SC, Zhang X, Trievel RC, Cheng X: **The SET-domain protein superfamily: protein lysine methyltransferases**. *Genome biology* 2005, **6**(8):227.
152. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y: **Histone demethylation by a family of JmjC domain-containing proteins**. *Nature* 2006, **439**(7078):811-816.
153. Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M *et al*: **Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases**. *Cell* 2006, **125**(3):467-481.
154. Chen Z, Zang J, Whetstine J, Hong X, Davrazou F, Kutateladze TG, Simpson M, Mao Q, Pan CH, Dai S *et al*: **Structural insights into histone demethylation by JMJD2 family members**. *Cell* 2006, **125**(4):691-702.
155. Krivtsov AV, Armstrong SA: **MLL translocations, histone modifications and leukaemia stem-cell development**. *Nature Reviews Cancer* 2007, **7**:823.
156. Ferrari KJ, Scelfo A, Jammula S, Cuomo A, Barozzi I, Stutzer A, Fischle W, Bonaldi T, Pasini D: **Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity**. *Molecular cell* 2014, **53**(1):49-62.

157. Hojfeldt JW, Laugesen A, Willumsen BM, Damhofer H, Hedehus L, Tvardovskiy A, Mohammad F, Jensen ON, Helin K: **Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2**. *Nature structural & molecular biology* 2018, **25**(3):225-232.
158. Lewis E, Mislove R: **New mutants report**. *Drosophila Information Service* 1947, **21**:69.
159. Lewis EB: **A gene complex controlling segmentation in Drosophila**. In: *Genes, Development and Cancer*. edn.: Springer; 1978: 205-217.
160. Ingham P: **Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes**. *Nature* 1983, **306**(5943):591.
161. Ingham P: **A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila**. *Journal of embryology and experimental morphology* 1985, **89**:349-365.
162. Kennison JA, Tamkun JW: **Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila**. *Proceedings of the National Academy of Sciences* 1988, **85**(21):8136-8140.
163. Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G: **Genome Regulation by Polycomb and Trithorax: 70 Years and Counting**. *Cell* 2017, **171**(1):34-57.
164. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G: **Genome regulation by polycomb and trithorax proteins**. *Cell* 2007, **128**(4):735-745.
165. Kadoch C, Crabtree GR: **Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics**. *Science advances* 2015, **1**(5):e1500447.
166. Hodges C, Kirkland JG, Crabtree GR: **The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer**. *Cold Spring Harbor perspectives in medicine* 2016, **6**(8):a026930.
167. Dingwall AK, Beek SJ, McCallum C, Tamkun J, Kalpana GV, Goff SP, Scott M: **The Drosophila snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex**. *Molecular Biology of the Cell* 1995, **6**(7):777-791.
168. Khavari PA, Peterson CL, Tamkun JW, Mendel DB, Crabtree GR: **BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription**. *Nature* 1993, **366**(6451):170.
169. Schmahling S, Meiler A, Lee Y, Mohammed A, Finkl K, Tauscher K, Israel L, Wirth M, Philippou-Massier J, Blum H *et al*: **Regulation and function of H3K36 di-methylation by the trithorax-group protein complex AMC**. *Development* 2018, **145**(7).
170. Huang C, Yang F, Zhang Z, Zhang J, Cai G, Li L, Zheng Y, Chen S, Xi R, Zhu B: **Mrg15 stimulates Ash1 H3K36 methyltransferase activity and facilitates Ash1 Trithorax group protein function in Drosophila**. *Nature communications* 2017, **8**(1):1649.
171. Gregory GD, Vakoc CR, Rozovskaia T, Zheng X, Patel S, Nakamura T, Canaani E, Blobel GA: **Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes**. *Molecular and cellular biology* 2007, **27**(24):8466-8479.
172. Tie F, Banerjee R, Saiakhova AR, Howard B, Monteith KE, Scacheri PC, Cosgrove MS, Harte PJ: **Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing**. *Development* 2014, **141**(5):1129-1139.
173. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri PC, Harte PJ: **CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing**. *Development* 2009, **136**(18):3131-3141.
174. Piunti A, Shilatifard A: **Epigenetic balance of gene expression by Polycomb and COMPASS families**. *Science (New York, NY)* 2016, **352**(6290):aad9780.
175. Steffen PA, Ringrose L: **What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory**. *Nature reviews Molecular cell biology* 2014, **15**(5):340-356.
176. Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A: **The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells**. *Nature structural & molecular biology* 2013, **20**(9):1093.
177. Denissov S, Hofemeister H, Marks H, Kranz A, Ciotta G, Singh S, Anastassiadis K, Stunnenberg HG, Stewart AF: **Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant**. *Development* 2014, **141**(3):526-537.

178. Lhuillier-Akakpo M, Frapporti A, Wilkes CD, Matelot M, Vervoort M, Sperling L, Duharcourt S: **Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements**. *PLoS genetics* 2014, **10**(9):e1004665.
179. Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, Strahl BD, Fujimori DG, Yates III JR, Madhani HD: **Product binding enforces the genomic specificity of a yeast polycomb repressive complex**. *Cell* 2015, **160**(1-2):204-218.
180. Blackledge NP, Rose NR, Klose RJ: **Targeting Polycomb systems to regulate gene expression: modifications to a complex story**. *Nature reviews Molecular cell biology* 2015, **16**(11):643.
181. Robinson AK, Leal BZ, Chadwell LV, Wang R, Ilangovan U, Kaur Y, Junco SE, Schirf V, Osmulski PA, Gaczynska M: **The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain**. *Journal of Biological Chemistry* 2012, **287**(12):8702-8713.
182. Wilkinson F, Pratt H, Atchison ML: **PcG recruitment by the YY1 REPO domain can be mediated by Yaf2**. *Journal of cellular biochemistry* 2010, **109**(3):478-486.
183. Rose NR, King HW, Blackledge NP, Fursova NA, Ember KJ, Fischer R, Kessler BM, Klose RJ: **RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes**. *eLife* 2016, **5**:e18591.
184. Laugesen A, Hojfeldt JW, Helin K: **Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation**. *Molecular cell* 2019, **74**(1):8-18.
185. Laugesen A, Helin K: **Chromatin repressive complexes in stem cells, development, and cancer**. *Cell Stem Cell* 2014, **14**(6):735-751.
186. Laugesen A, Hojfeldt JW, Helin K: **Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer**. *Cold Spring Harb Perspect Med* 2016, **6**(9).
187. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y: **Role of histone H3 lysine 27 methylation in Polycomb-group silencing**. *Science (New York, NY)* 2002, **298**(5595):1039-1043.
188. Cao R, Zhang Y: **The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3**. *Current opinion in genetics & development* 2004, **14**(2):155-164.
189. Justin N, Zhang Y, Tarricone C, Martin SR, Chen S, Underwood E, De Marco V, Haire LF, Walker PA, Reinberg D *et al*: **Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2**. *Nature communications* 2016, **7**:11316.
190. Lee CH, Yu JR, Kumar S, Jin Y, LeRoy G, Bhanu N, Kaneko S, Garcia BA, Hamilton AD, Reinberg D: **Allosteric Activation Dictates PRC2 Activity Independent of Its Recruitment to Chromatin**. *Molecular cell* 2018, **70**(3):422-434.e426.
191. Yu JR, Lee CH, Oksuz O, Stafford JM, Reinberg D: **PRC2 is high maintenance**. *Genes & development* 2019, **33**(15-16):903-935.
192. Lee C-H, Holder M, Grau D, Saldana-Meyer R, Yu J-R, Ganai RA, Zhang J, Wang M, LeRoy G, Dobenecker M-W: **Distinct stimulatory mechanisms regulate the catalytic activity of polycomb repressive complex 2**. *Molecular cell* 2018, **70**(3):435-448. e435.
193. Kasinath V, Poepsel S, Nogales E: **Recent structural insights into Polycomb repressive complex 2 regulation and substrate binding**. *Biochemistry* 2018, **58**(5):346-354.
194. Zhang K, Mosch K, Fischle W, Grewal SI: **Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin**. *Nature structural & molecular biology* 2008, **15**(4):381.
195. Müller MM, Fierz B, Bittova L, Liszczak G, Muir TW: **A two-state activation mechanism controls the histone methyltransferase Suv39h1**. *Nature chemical biology* 2016, **12**(3):188.
196. Lynch MD, Smith AJH, De Gobbi M, Flenley M, Hughes JR, Vernimmen D, Ayyub H, Sharpe JA, Sloane-Stanley JA, Sutherland L *et al*: **An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment**. *The EMBO Journal* 2012, **31**(2):317-329.
197. Coleman RT, Struhl G: **Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene**. *Science (New York, NY)* 2017, **356**(6333):eaai8236.
198. Laprell F, Finkl K, Müller J: **Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA**. *Science (New York, NY)* 2017, **356**(6333):85-88.

199. Sanulli S, Justin N, Teissandier A, Ancelin K, Portoso M, Caron M, Michaud A, Lombard B, Da Rocha ST, Offer J: **Jarid2 methylation via the PRC2 complex regulates H3K27me3 deposition during cell differentiation**. *Molecular cell* 2015, **57**(5):769-783.
200. Son J, Shen SS, Margueron R, Reinberg D: **Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin**. *Genes & development* 2013, **27**(24):2663-2677.
201. Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE: **GC-rich sequence elements recruit PRC2 in mammalian ES cells**. *PLoS genetics* 2010, **6**(12):e1001244.
202. Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K: **Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide**. *Molecular cell* 2014, **55**(3):347-360.
203. Bousard A, Raposo AC, Zylicz JJ, Picard C, Pires VB, Qi Y, Gil C, Syx L, Chang HY, Heard E *et al*: **The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation**. *EMBO reports* 2019:e48019.
204. Yamaguchi H, Hung MC: **Regulation and Role of EZH2 in Cancer**. *Cancer research and treatment : official journal of Korean Cancer Association* 2014, **46**(3):209-222.
205. Deb G, Singh AK, Gupta S: **EZH2: not EZHY (easy) to deal**. *Molecular cancer research : MCR* 2014, **12**(5):639-653.
206. Lu H, Li G, Zhou C, Jin W, Qian X, Wang Z, Pan H, Jin H, Wang X: **Regulation and role of post-translational modifications of enhancer of zeste homologue 2 in cancer development**. *American journal of cancer research* 2016, **6**(12):2737-2754.
207. Kim E, Kim M, Woo DH, Shin Y, Shin J, Chang N, Oh YT, Kim H, Rheey J, Nakano I *et al*: **Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells**. *Cancer cell* 2013, **23**(6):839-852.
208. He A, Shen X, Ma Q, Cao J, von Gise A, Zhou P, Wang G, Marquez VE, Orkin SH, Pu WT: **PRC2 directly methylates GATA4 and represses its transcriptional activity**. *Genes & development* 2012, **26**(1):37-42.
209. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT, Wu X, Stack EC, Loda M, Liu T *et al*: **EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent**. *Science (New York, NY)* 2012, **338**(6113):1465-1469.
210. Luo J, Wang K, Yeh S, Sun Y, Liang L, Xiao Y, Xu W, Niu Y, Cheng L, Maity SN *et al*: **LncRNA-p21 alters the antiandrogen enzalutamide-induced prostate cancer neuroendocrine differentiation via modulating the EZH2/STAT3 signaling**. *Nature communications* 2019, **10**(1):2571.
211. Yan J, Ng SB, Tay JL, Lin B, Koh TL, Tan J, Selvarajan V, Liu SC, Bi C, Wang S *et al*: **EZH2 overexpression in natural killer/T-cell lymphoma confers growth advantage independently of histone methyltransferase activity**. *Blood* 2013, **121**(22):4512-4520.
212. Li J, Xi Y, Li W, McCarthy RL, Stratton SA, Zou W, Li W, Dent SY, Jain AK, Barton MC: **TRIM28 interacts with EZH2 and SWI/SNF to activate genes that promote mammosphere formation**. *Oncogene* 2017, **36**(21):2991-3001.
213. Lee ST, Li Z, Wu Z, Aau M, Guan P, Karuturi RK, Liou YC, Yu Q: **Context-specific regulation of NF-kappaB target gene expression by EZH2 in breast cancers**. *Molecular cell* 2011, **43**(5):798-810.
214. Kim KH, Roberts CW: **Targeting EZH2 in cancer**. *Nature medicine* 2016, **22**(2):128-134.
215. Han Li C, Chen Y: **Targeting EZH2 for cancer therapy: progress and perspective**. *Current protein & peptide science* 2015, **16**(6):559-570.
216. Xu B, Konze KD, Jin J, Wang GG: **Targeting EZH2 and PRC2 dependence as novel anticancer therapy**. *Experimental hematology* 2015, **43**(8):698-712.
217. Wang X, Cao W, Zhang J, Yan M, Xu Q, Wu X, Wan L, Zhang Z, Zhang C, Qin X *et al*: **A covalently bound inhibitor triggers EZH2 degradation through CHIP-mediated ubiquitination**. *Embo j* 2017, **36**(9):1243-1260.
218. Kipps TJ, Stevenson FK, Wu CJ, Croce CM, Packham G, Wierda WG, O'Brien S, Gribben J, Rai K: **Chronic lymphocytic leukaemia**. *Nature reviews Disease primers* 2017, **3**:16096.
219. Kokhaei P, Palma M, Mellstedt H, Choudhury A: **Advances in molecular biology diagnostic and treatment of B-cell malignancies: Biology and treatment of chronic lymphocytic leukemia**. *Annals of Oncology* 2005, **16**(suppl_2):ii113-ii123.

220. Dighiero G, Hamblin TJ: **Chronic lymphocytic leukaemia**. *Lancet (London, England)* 2008, **371**(9617):1017-1029.
221. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, Dohner K, Bentz M, Lichter P: **Genomic aberrations and survival in chronic lymphocytic leukemia**. *The New England journal of medicine* 2000, **343**(26):1910-1916.
222. Swerdlow SH: **WHO classification of tumours of haematopoietic and lymphoid tissues**: International Agency for Research on Cancer; 2017.
223. Siegel R, DeSantis C, Virgo K, Stein K, Mariotto A, Smith T, Cooper D, Gansler T, Lerro C, Fedewa S *et al*: **Cancer treatment and survivorship statistics, 2012**. *CA: a cancer journal for clinicians* 2012, **62**(4):220-241.
224. Pulte D, Redaniel MT, Bird J, Jeffreys M: **Survival for patients with chronic leukemias in the US and Britain: Age-related disparities and changes in the early 21st century**. *European journal of haematology* 2015, **94**(6):540-545.
225. Li Y, Wang Y, Wang Z, Yi D, Ma S: **Racial differences in three major NHL subtypes: descriptive epidemiology**. *Cancer epidemiology* 2015, **39**(1):8-13.
226. Guieze R, Ysebaert L, de Guibert S, Dhedin N, Bay JO, Troussard X, Tournilhac O: **[Conventional therapeutic strategies for relapsed chronic lymphocytic leukemia]**. *Bulletin du cancer* 2012, **99**(12):1123-1132.
227. Hallek M: **Therapy of chronic lymphocytic leukaemia**. *Best practice & research Clinical haematology* 2010, **23**(1):85-96.
228. de Rooij MF, Kuil A, Geest CR, Eldering E, Chang BY, Buggy JJ, Pals ST, Spaargaren M: **The clinically active BTK inhibitor PCI-32765 targets B-cell receptor- and chemokine-controlled adhesion and migration in chronic lymphocytic leukemia**. *Blood* 2012, **119**(11):2590-2594.
229. Ponader S, Chen SS, Buggy JJ, Balakrishnan K, Gandhi V, Wierda WG, Keating MJ, O'Brien S, Chiorazzi N, Burger JA: **The Bruton tyrosine kinase inhibitor PCI-32765 thwarts chronic lymphocytic leukemia cell survival and tissue homing in vitro and in vivo**. *Blood* 2012, **119**(5):1182-1189.
230. Guo A, Lu P, Galanina N, Nabhan C, Smith SM, Coleman M, Wang YL: **Heightened BTK-dependent cell proliferation in unmutated chronic lymphocytic leukemia confers increased sensitivity to ibrutinib**. *Oncotarget* 2016, **7**(4):4598-4610.
231. Byrd JC, Brown JR, O'Brien S, Barrientos JC, Kay NE, Reddy NM, Coutre S, Tam CS, Mulligan SP, Jaeger U *et al*: **Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia**. *The New England journal of medicine* 2014, **371**(3):213-223.
232. Mallm J-P, Iskar M, Ishaque N, Klett LC, Kugler SJ, Muino JM, Teif VB, Poos AM, Großmann S, Erdel F *et al*: **Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks**. *Mol Syst Biol* 2019, **15**(5):e8339-e8339.
233. Georgiadis P, Gavriil M, Rantakokko P, Ladoukakis E, Botsivali M, Kelly RS, Bergdahl IA, Kiviranta H, Vermeulen RCH, Spaeth F *et al*: **DNA methylation profiling implicates exposure to PCBs in the pathogenesis of B-cell chronic lymphocytic leukemia**. *Environment international* 2019, **126**:24-36.
234. Lin H, Grosschedl R: **Failure of B-cell differentiation in mice lacking the transcription factor EBF**. *Nature* 1995, **376**(6537):263.
235. Borghesi L, Aites J, Nelson S, Lefterov P, James P, Gerstein R: **E47 is required for V (D) J recombinase activity in common lymphoid progenitors**. *Journal of Experimental Medicine* 2005, **202**(12):1669-1677.
236. Souabni A, Cobaleda C, Schebesta M, Busslinger M: **Pax5 promotes B lymphopoiesis and blocks T cell development by repressing Notch1**. *Immunity* 2002, **17**(6):781-793.
237. Eibel H, Kraus H, Sic H, Kienzler AK, Rizzi M: **B cell biology: an overview**. *Current allergy and asthma reports* 2014, **14**(5):434.
238. Hamel KM, Liarski VM, Clark MR: **Germinal center B-cells**. *Autoimmunity* 2012, **45**(5):333-347.
239. Chiorazzi N, Rai KR, Ferrarini M: **Chronic lymphocytic leukemia**. *The New England journal of medicine* 2005, **352**(8):804-815.
240. Rozman C, Montserrat E: **Chronic lymphocytic leukemia**. *The New England journal of medicine* 1995, **333**(16):1052-1057.

241. Raheem RM: **Immunohistochemical Expression of Vascular Endothelial Growth Factor in Chronic Lymphocytic Leukemia and its Relation to Laboratory and Clinical Findings**. *Medical Journal of Babylon* 2012, **9**(1):27-35.
242. Gaidano G, Foà R, Dalla-Favera R: **Molecular pathogenesis of chronic lymphocytic leukemia**. *J Clin Invest* 2012, **122**(10):3432-3438.
243. Mangiola M, Cutrona G, Colombo M, Rossi E, Spriano M, Morabito F: **Mechanisms that control cell apoptosis or survival through surface Ig in B-CLL**. *Leukemia & lymphoma* 2003, **44**:S4.
244. Chu P, Deforce D, Pedersen IM, Kim Y, Kitada S, Reed JC, Kipps TJ: **Latent sensitivity to Fas-mediated apoptosis after CD40 ligation may explain activity of CD154 gene therapy in chronic lymphocytic leukemia**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(6):3854-3859.
245. Rawstron AC, Bennett FL, O'Connor SJ, Kwok M, Fenton JA, Plummer M, de Tute R, Owen RG, Richards SJ, Jack AS: **Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia**. *New England Journal of Medicine* 2008, **359**(6):575-583.
246. Jain N, Keating MJ: **Richter transformation of CLL**. *Expert review of hematology* 2016, **9**(8):793-801.
247. Messmer BT, Albesiano E, Messmer D, Chiorazzi N: **The pattern and distribution of immunoglobulin VH gene mutations in chronic lymphocytic leukemia B cells are consistent with the canonical somatic hypermutation process**. *Blood* 2004, **103**(9):3490-3495.
248. Cavalli F, Isaacson PG, Gascoyne RD, Zucca E: **MALT lymphomas**. *ASH Education Program Book* 2001, **2001**(1):241-258.
249. Damle RN, Batliwalla FM, Ghiotto F, Valetto A, Albesiano E, Sison C, Allen SL, Kolitz J, Vinciguerra VP, Kudalkar P: **Telomere length and telomerase activity delineate distinctive replicative features of the B-CLL subgroups defined by immunoglobulin V gene mutations**. *Blood* 2004, **103**(2):375-382.
250. Bernal A, Pastore RD, Asgary Z, Keller SA, Cesarman E, Liou H-C, Schattner EJ: **Survival of leukemic B cells promoted by engagement of the antigen receptor**. *Blood* 2001, **98**(10):3050-3057.
251. Panayiotidis P, Jones D, Ganeshaguru K, Foroni L, Hoffbrand A: **Human bone marrow stromal cells prevent apoptosis and support the survival of chronic lymphocytic leukaemia cells in vitro**. *British journal of haematology* 1996, **92**(1):97-103.
252. Ghamlouch H, Nguyen-Khac F, Bernard OA: **Chronic lymphocytic leukaemia genomics and the precision medicine era**. *British journal of haematology* 2017, **178**(6):852-870.
253. Ouillette P, Collins R, Shakhan S, Li J, Li C, Shedden K, Malek SN: **The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011, **17**(21):6778-6790.
254. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Bottcher S *et al*: **Mutations driving CLL and their evolution in progression and relapse**. *Nature* 2015, **526**(7574):525-530.
255. Oscier DG, Rose-Zerilli MJ, Winkelmann N, Gonzalez de Castro D, Gomez B, Forster J, Parker H, Parker A, Gardiner A, Collins A *et al*: **The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial**. *Blood* 2013, **121**(3):468-475.
256. Dohner H, Stilgenbauer S, Dohner K, Bentz M, Lichter P: **Chromosome aberrations in B-cell chronic lymphocytic leukemia: reassessment based on molecular cytogenetic analysis**. *Journal of molecular medicine (Berlin, Germany)* 1999, **77**(2):266-281.
257. Dohner H, Stilgenbauer S, James MR, Benner A, Weilguni T, Bentz M, Fischer K, Hunstein W, Lichter P: **11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis**. *Blood* 1997, **89**(7):2516-2522.
258. Kanduri M, Cahill N, Göransson H, Enström C, Ryan F, Isaksson A, Rosenquist R: **Differential genome-wide array–based methylation profiles in prognostic subsets of chronic lymphocytic leukemia**. *Blood* 2010, **115**(2):296-305.
259. Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, Martínez-Trillos A, Castellano G, Brun-Heath I, Pinyol M: **Epigenomic analysis detects widespread gene-body**

DNA hypomethylation in chronic lymphocytic leukemia. *Nature genetics* 2012, **44**(11):1236.
260. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S: **Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia**. *Cancer cell* 2014, **26**(6):813-825.
261. Oakes CC, Claus R, Gu L, Assenov Y, Hüllein J, Zucknick M, Bieg M, Brocks D, Bogatyrova O, Schmidt CR: **Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia**. *Cancer discovery* 2014, **4**(3):348-361.
262. Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, Wang Q, Imbusch CD, Serva A, Koser SD: **DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia**. *Nature genetics* 2016, **48**(3):253.
263. Kulis M, Merkel A, Heath S, Queiros AC, Schuyler RP, Castellano G, Beekman R, Raineri E, Esteve A, Clot G *et al*: **Whole-genome fingerprint of the DNA methylome during human B cell differentiation**. *Nat Genet* 2015, **47**(7):746-756.
264. Chapuy B, McKeown MR, Lin CY, Monti S, Roemer MG, Qi J, Rahl PB, Sun HH, Yeda KT, Doench JG: **Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma**. *Cancer cell* 2013, **24**(6):777-790.
265. Ott CJ, Federation AJ, Schwartz LS, Kasar S, Klitgaard JL, Lenci R, Li Q, Lawlor M, Fernandes SM, Souza A *et al*: **Enhancer Architecture and Essential Core Regulatory Circuitry of Chronic Lymphocytic Leukemia**. *Cancer cell* 2018, **34**(6):982-995.e987.
266. Mallm JP, Iskar M, Ishaque N, Klett LC, Kugler SJ, Muino JM, Teif VB, Poos AM, Grossmann S, Erdel F *et al*: **Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks**. *Mol Syst Biol* 2019, **15**(5):e8339.
267. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H, Hillmen P, Keating M, Montserrat E, Chiorazzi N *et al*: **iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL**. *Blood* 2018, **131**(25):2745-2760.
268. Harrington CT, Lin EI, Olson MT, Eshleman JR: **Fundamentals of pyrosequencing**. *Archives of pathology & laboratory medicine* 2013, **137**(9):1296-1303.
269. Smale ST: **Luciferase assay**. *Cold Spring Harbor protocols* 2010, **2010**(5):pdb.prot5421.
270. Gade P, Kalvakolanu DV: **Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity**. *Methods in molecular biology (Clifton, NJ)* 2012, **809**:85-104.
271. Jiang S, Mortazavi A: **Integrating ChIP-seq with other functional genomics data**. *Brief Funct Genomics* 2018, **17**(2):104-115.
272. Kosalai ST, Morsy MHA, Papakonstantinou N, Mansouri L, Stavroyianni N, Kanduri C, Stamatopoulos K, Rosenquist R, Kanduri M: **EZH2 upregulates the PI3K/AKT pathway through IGF1R and MYC in clinically aggressive chronic lymphocytic leukaemia**. *Epigenetics* 2019:1-16.
273. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics (Oxford, England)* 2013, **29**(1):15-21.
274. Lassmann T, Hayashizaki Y, Daub CO: **SAMStat: monitoring biases in next generation sequencing data**. *Bioinformatics (Oxford, England)* 2011, **27**(1):130-131.
275. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics (Oxford, England)* 2014, **30**(7):923-930.
276. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics (Oxford, England)* 2010, **26**(1):139-140.
277. Subhash S, Kanduri C: **GeneSCF: a real-time based functional enrichment tool with support for multiple organisms**. *BMC bioinformatics* 2016, **17**(1):365.
278. Rosen A, Bergh AC, Gogok P, Evaldsson C, Myhrinder AL, Hellqvist E, Rasul A, Bjorkholm M, Jansson M, Mansouri L *et al*: **Lymphoblastoid cell line with B1 cell characteristics established from a chronic lymphocytic leukemia clone by in vitro EBV infection**. *Oncoimmunology* 2012, **1**(1):18-27.
279. Stacchini A, Aragno M, Vallario A, Alfarano A, Circosta P, Gottardi D, Faldella A, Rege-Cambrin G, Thunberg U, Nilsson K *et al*: **MEC1 and MEC2: two new cell lines derived**

from B-chronic lymphocytic leukaemia in prolymphocytoid transformation. *Leukemia research* 1999, **23**(2):127-136.
280. Kopparapu PK, Miranda C, Fogelstrand L, Mishra K, Andersson PO, Kanduri C, Kanduri M: **MCPH1 maintains long-term epigenetic silencing of ANGPT2 in chronic lymphocytic leukemia**. *The FEBS journal* 2015, **282**(10):1939-1952.
281. Kopparapu PK, Abdelrazak Morsy MH, Kanduri C, Kanduri M: **Gene-body hypermethylation controlled cryptic promoter and miR26A1-dependent EZH2 regulation of TET1 gene activity in chronic lymphocytic leukemia**. *Oncotarget* 2017, **8**(44):77595-77608.
282. Wernig-Zorc S, Yadav MP, Kopparapu PK, Bemark M, Kristjansdottir HL, Andersson PO, Kanduri C, Kanduri M: **Global distribution of DNA hydroxymethylation and DNA methylation in chronic lymphocytic leukemia**. *Epigenetics & chromatin* 2019, **12**(1):4.
283. Ferreira PG, Jares P, Rico D, Gomez-Lopez G, Martinez-Trillos A, Villamor N, Ecker S, Gonzalez-Perez A, Knowles DG, Monlong J *et al*: **Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia**. *Genome Res* 2014, **24**(2):212-226.