

Inaugural dissertation
for
obtaining the doctoral degree
of the
Combined Faculty Of Mathematics, Engineering and Natural Sciences
of the
Ruprecht – Karls – University
Heidelberg

Presented by:
M.Sc. Matteo Trovato

born in: Pisa, Italy

Oral examination: 23rd January 2023

The role of histone H3.3 K9 and K27 residues in transcriptional regulation
in mouse embryonic stem cells

Referees:

Dr. Eileen Furlong

Prof. Dr. Sylvia Erhardt

Provando e riprovando

Galileo Galilei
Accademia del Cimento (Firenze, 1657)

Summary

Histone post-translational modifications (PTMs) are generally associated with defined transcriptional states. For instance, histone H3 lysine acetylation (e.g., H3K27ac, H3K9ac) conventionally marks cell-type specific active regulatory elements, whereas H3 lysine 9 tri-methylation (H3K9me3) and lysine 27 tri-methylation (H3K27me3) are enriched at repressed repeat elements and promoters, respectively. Despite strong correlations, it is still unclear whether histone modifications play an instructive role in transcriptional regulation. Histone residue mutagenesis represents an elegant approach to address this question. Nevertheless, applying histone mutagenesis is challenging in the vast majority of metazoans because canonical histone genes are found in large clusters on different chromosomes. The histone variant H3.3 is instead encoded by two genes outside the clusters, which makes mutagenesis more feasible in mammalian systems. Here, I applied H3.3 mutagenesis to study the relevance of H3.3 K9 and K27 residues in gene expression regulation, with the primary goal of characterizing the contribution of these histone marks in repressive chromatin environments. Using CRISPR-Cas9, I generated mouse embryonic stem cells (mESCs) bearing lysine-to-alanine substitutions and applied multi-omics to profile changes in transcriptome and epigenome. H3.3 K9A and K27A mutants resulted in severe gene expression changes whereby a large proportion of deregulated genes overlapped in both mutants. Developmentally-regulated genes harbouring bivalent (i.e., H3K4me3-H3K27me3) promoters were de-repressed in both mutants, and the in-vitro differentiation potential of the mESCs was disrupted. The H3.3K27A mutation affected local deposition and spreading of the H3K27me3 mark, which particularly impacts a subset of bivalent genes with a higher basal level of H3.3 on their promoters. On the other hand, the H3.3K9A mutation resulted in the global loss of H3K9me3 at endogenous retroviruses (ERVs) and distal intergenic regions. These regions gained histone H3 acetylation and exhibited increased levels of nascent transcription. Constructing a gene regulatory network (GRN) revealed that the de-repression of these territories leads to a profound rewiring of transcriptional regulatory networks in mESCs. This rewiring resulted in the upregulation of bivalent genes and genes related to immune system processes in H3.3K9A and also culminated in the expression of proteins generally restricted to highly specialized immune cell types.

At canonical enhancers, H3.3K27A displayed decreased H3K27ac and induced a collateral loss of acetylation at other H3 lysines. Nascent transcription levels were reduced, and the genes connected to these regions through the GRN were downregulated, indicating that

comprehensive—rather than residue-specific—H3 acetylation might play a crucial role in maintaining enhancers in an active state.

In H3.3K9A mESCs, H3K27ac was primarily affected at enhancers, and the nascent transcription was reduced at regulatory elements particularly active in mESCs. I propose that the reduced activity of canonical enhancers in H3.3K9A mESCs is a consequence of the alien activation of ERV-derived regulatory elements, which potentially compete with canonical enhancers resulting in rerouting of transcription factors to binding sites that are not available in the basal state.

Collectively, this study demonstrates that H3.3 K9 and K27 residues are critical for maintaining repressive chromatin states at promoters and distal genomic regions and are, therefore, essential for transcriptional homeostasis in mESCs.

Zusammenfassung

Im Allgemeinen werden posttranslationale Histonmodifikationen (PTMs) mit bestimmten Transkriptionszuständen in Verbindung gebracht. So markiert beispielsweise die Acetylierung von Histon-H3 (z. B. H3K27ac, H3K9ac) zelltypspezifische aktive regulatorische Elemente, während die H3-Lysin-9-Tri-Methylierung (H3K9me3) und die Lysin-27-Tri-Methylierung (H3K27me3) an reprimierten repetitiven DNA-Elementen bzw. an reprimierten Promotoren angereichert sind. Trotz starker Korrelationen ist immer noch unklar, ob Histonmodifikationen eine instruktive Rolle bei der Regulation der Gentranskription spielen. Die Mutagenese von Aminosäureresten der Histonproteine ist ein eleganter Ansatz, um diese Frage zu klären. Dennoch ist die Anwendung der Histon-Mutagenese bei der überwiegenden Mehrheit der vielzelligen Tiere eine Herausforderung, da die kanonischen Histongene in großen Clustern auf verschiedenen Chromosomen verteilt sind. Die Histonvariante H3.3 wird stattdessen lediglich von zwei Genen außerhalb dieser Cluster kodiert, welches wiederum ihre Mutagenese in Säugerzellen vereinfacht. Im Rahmen dieses Forschungsprojekts, habe ich die H3.3-Mutagenese angewandt, um die Bedeutung der H3.3-K9- und -K27-Aminosäurereste bei der Regulierung der Genexpression zu untersuchen, mit dem primären Ziel, den Beitrag dieser Histonmarkierungen in repressiver Chromatinlandschaft zu charakterisieren. Mithilfe von CRISPR-Cas9 erzeugte ich murine embryonale Stammzellen (mESCs), welche Lysin-zu-Alanin-Substitutionen trugen, und wendete multi-omics Techniken an, um Veränderungen im Transkriptom und im Epigenom zu erfassen. Hierbei führten H3.3 K9A- und K27A-Mutationen zu schwerwiegenden Veränderungen der Genexpression, wobei sich ein großer Anteil deregulierter Gene in beiden Mutationen überschneidet. Gene mit bivalenten (d. h. H3K4me3-H3K27me3) Promotoren, die während der Entwicklung des Organismus eine entscheidende Rolle spielen, wurden in beiden Mutanten dereprimiert, und das in-vitro Differenzierungspotenzial der mESCs war gestört. Die H3.3K27A-Mutation beeinträchtigte die lokale Ablagerung und Ausbreitung der H3K27me3-Markierung, welches sich insbesondere auf eine Untergruppe von bivalenten Genen mit einem höheren Basalniveau von H3.3 an ihren Promotoren auswirkt. Andererseits führte die H3.3K9A-Mutation zu einem globalen Verlust von H3K9me3 bei endogenen Retroviren (ERVs) und nicht-kodierenden distalen Zwischengen-Abschnitten des Genoms. Diese Regionen gewannen an Histon-H3-Acetylierung zu und wiesen ein erhöhtes Maß an naszierender Transkription auf. Die bioinformatische Konstruktion eines genregulatorischen Netzwerks (GRN) zeigte, dass die Derepression dieser Gebiete zu einer tiefgreifenden

Neuverdrahtung der transkriptionellen regulatorischen Netzwerke in mESCs führt. Diese Neuverdrahtung führte zu einer Hochregulierung von bivalenten und immunregulatorischen Genen in H3.3K9A Zellen, und kulminierte in der Expression von Proteinen, die unter normalen Umständen nur in hochspezialisierten Immunzelltypen exprimiert würden.

An kanonischen Enhancerregionen wies die H3.3K27A Mutation eine verringerte Acetylierung des H3K27 auf und führte ebenso zu einem Verlust der Acetylierung an anderen H3-Lysinresten. Die naszierende Transkription war reduziert, und die Gene, die über das GRN mit diesen Regionen verbunden waren, wurden herunterreguliert. Dies deutet darauf hin, dass die umfassende - und nicht die histonrestspezifische - H3-Acetylierung eine entscheidende Rolle bei der Aufrechterhaltung von Enhancern in einem aktiven Zustand spielen könnte.

In H3.3K9A mESCs war die Acetylierung des H3K27 hauptsächlich an Enhancern betroffen, und die naszierende Transkription war an regulatorischen DNA-Elementen, die in mESCs besonders aktiv sind, reduziert. Daher stelle ich die Hypothese auf, dass die verringerte Aktivität kanonischer Enhancer in H3.3K9A mESCs eine Folge der Derepression von ERV-abgeleiteten regulatorischen DNA-Elementen ist, die möglicherweise mit kanonischen Enhancern konkurrieren. Dies wiederum könnte zu einer veränderten Interaktion von Transkriptionsfaktoren und DNA-Bindestellen führen, die im Grundzustand nicht zugänglich sind.

Insgesamt zeigt diese Studie, dass die H3.3 K9- und K27-Aminosäurereste für die Aufrechterhaltung repressiver Chromatinzustände an Promotoren sowie distalen genomischen Regionen entscheidend sind und daher unverzichtbar für die Transkriptionshomöostase in mESCs sind.

Acknowledgments

It is a strange feeling to write the acknowledgements section of this thesis because, despite the challenges faced along the “PhD journey”, this time has passed really quickly. I think this probably happens only when we do something we like and, most importantly, when we are lucky enough to share the path with truly good people.

First and foremost, I thank my supervisor Kyung-Min Noh for giving me the opportunity to join her research group. Thanks for your guidance, for believing in me and for being able to strike the right chords to always spur me on to do my best.

I would also like to express my gratitude to my TAC members Eileen Furlong, Judith Zaugg and Sylvia Erhardt for the time dedicated to help in developing the project and for having always shown interest in it. I would also like to acknowledge Prof. Haikun Liu and Prof. Simon Anders for agreeing on participating to the thesis defence. The realisation of this project would not have been possible without the help of several people. Thanks to Andrea Callegari and Moritz Kueblbeck for sharing their expertise in CRISPR-Cas9 editing. EMBL wouldn't be EMBL without the amazing core facilities! So thanks to Pepcore, FACS facility (especially Diana and Michael) and Genecore (especially Vladimir, Laura and Mireia). Also many thanks to Charles and the GB computational support. I would also like to thank Elena Vizcaya Molina for providing the *Drosophila* cells.

I feel really lucky to have had, in a way, the privilege to bridge between the “*first gen-*” and “*second gen-Noh Group*”. Thanks to Niccolò, Maja and Nichole for welcoming me in the lab. Special thanks to Dasha for all the guidance and help with the analyses. Thanks to Lotta for being always there to dispense advice, every time I needed it. Thanks to Lucia, for making this PhD not all about pain and struggle, but mostly fun and laughter. You have been the best PhD-buddy and I will never stop being grateful to you for this. Thanks to Umut, since we first met I realised *in the blink of an eye* the great lab mate you would be, I can say now I wasn't wrong. Thanks to Sara, for being the nicest *surprise lab mate* I could have ever wished for. Thanks to Nadine, for being always there to help and for spoiling us with your *sweetness*, in every possible sense. Thanks to Victor, for being always around the corner ready for a joke, a meme or just a good chat. Thanks to the amazing postdocs of the group, Thomas, Jenny, Vibha, Mikael and Inyoung for always humbly sharing your experience and for finding the right encouraging word whenever needed. I would also like to thank Ana and Mohamad, for working with me on part of the project and for teaching me how to teach. Thanks also to the *guest stars* María and Victor for bringing a breath of joy, crazy stories and motivational phrases in these weeks of thesis writing.

I am also incredibly grateful to all the predocs of the 2018 batch—with a special mention to Anna, Agata, Karolina, Alberto, Gilberto and Jesus—for all the great moments spent together in these years. I could also not fail to mention *the Italians* (Matteo, Michael, Mattia, Roby, Alice, Silvia and all the others) for making me feel at home, abroad.

I am deeply thankful to my family and lifelong friends for always being there.

A heartfelt thank you to Mari, because despite the distance I could always count on you at all times, all this time.

Finally, the warmest thank you to my parents for teaching me to believe in myself and for always supporting me in my choices. Grazie Babbo e Mamma, vi voglio bene!

Contents

Summary	VII
Zusammenfassung.....	IX
Acknowledgments	XI
1. Introduction	1
1.1 Genomic determinants of transcriptional regulation: promoters and enhancers .	1
1.2 Nucleosome and histone modifications	3
1.2.1 Histone acetylation	3
1.2.2 Histone acetylation and transcription	5
1.2.3 Histone methylation	6
1.2.4 Histone methylation and transcription	8
1.3 The residues studied in my PhD.....	9
1.3.1 H3K9 acetylation and methylation.....	9
1.3.2 H3K27 acetylation and methylation	11
1.3.3 H3K79: DOT1L-mediated methylation and proposed functions	13
1.4 Histone modifications define chromatin states	15
1.4.1 Bivalent promoters: a closer look.....	16
1.4.2 Histone modifications: cause or (dispensable) cog? Insights from latest studies, evolutionary perspectives and synthetic biology	18
1.4.3 Studying the contribution of histone modifications: limitations of knockouts, inactivating mutations and chemical inhibition of histone-modifying enzymes.....	20
1.4.4 Histone mutagenesis to investigate the function of histone residues and histone PTMs.....	22
1.5 A brief summary of histone H3 variants.....	24
1.6 Histone H3.3	25
1.6.1 Histone H3.3 modifications	26
1.6.2 Histone H3 mutations in cancers	27
1.6.3 H3.3 mutagenesis as a platform to study histone residues and PTMs in mammals	28
1.7 Endogenous retroviruses (ERVs): origins and classification	30
1.7.1 Epigenetic mechanisms of ERVs repression.....	31
1.7.2 H3.3 and ERVs repression.....	33
1.7.3 Roles of ERVs in transcriptional regulation.....	33
1.8 Aim of the project.....	35
2. Results	37

2.1	Generation of H3.3 mutant mESC lines using CRISPR-Cas9	37
2.1.1	H3.3 K79A mESCs lines generation and validation	37
2.1.2	H3.3 K9A and K27A mESC lines generation and validation.....	38
2.2	Characterization of H3.3 K9A and K27A mESCs.....	42
2.2.1	H3.3 K9A and K27A mESCs display wide transcriptional changes.....	42
2.2.2	Global histone modifications changes in H3.3 K9A and K27A mESCs.....	45
2.2.3	H3.3 K9A (but not K27A) mESCs display growth delay and increased cell death	46
2.2.4	Derepression of developmentally-regulated genes in K9A and K27A mESCs impacts the differentiation potential.....	49
2.2.5	Developmental genes upregulated in K9A and K27A mESCs have bivalent promoters.....	52
2.2.6	H3K27me3 signal is decreased at bivalent promoters and H3.3K27A results in a reduction in SUZ12 binding and H3K27me3 spreading.....	55
2.2.7	Bivalent promoters do not display enrichment in H3K9me2/3 and local changes in histone PTMs fail to explain the H3K27me3 decrease observed in K9A mESCs	57
2.2.8	H3K9me3 signal is reduced at distal intergenic regions in K9A mESCs	60
2.2.9	Genomic regions with reduced H3K9me3 signal in H3.3K9A mESCs gain H3K27ac and exhibit increased levels of nascent transcription	61
2.2.10	H3.3K9 maintains ERVs in a repressed state in mESCs and regions with reduced H3K9me3 signal contain ERV-derived regulatory elements	65
2.2.11	Activation of ERV-derived regulatory elements in H3.3K9A mESCs drives upregulation of both bivalent and immune-related genes	67
2.2.12	Regulatory elements activated in H3.3K9A mESCs overlap with CREs active in immune cell types.....	70
2.2.13	Differential nascent transcription analysis verifies increased activity of ERV-derived CREs enriched for motifs of immune-related and developmental TFs.....	71
2.2.14	Analysis of differential TF activity reveals NF-KB and other immune-related TFs among the most significantly active in H3.3K9A mESCs.....	73
2.2.15	H3.3K9A mESCs express immune-related surface proteins	74
2.2.16	H3.3K9A mESCs exhibit reduced levels of lipid peroxidation.....	75
2.2.17	Effects of H3.3 K9A and K27A on histone acetylation and nascent transcription at distal CREs	77
2.2.18	Analysis of expression of genes connected to distal CREs through the GRN	80
3.	Discussion and outlook.....	83
3.1	Single H3.3 amino-acid substitutions at K27 and K9 residues have a large impact on transcription in mESCs.....	83

3.2	H3.3 K27A mutant discloses a prominent role of H3.3 in maintaining promoter H3K27me3 levels and repression of certain Polycomb-target genes in mESCs.....	84
3.3	H3.3K9 safeguards repression of ERVs and distal <i>cis</i> -regulatory elements in mESCs	86
3.4	H3.3 K27A and K9A affect (likely in different ways) activity of canonical distal <i>cis</i> -regulatory elements and expression of putative target genes	93
4.	Materials and methods	97
	Appendix	111
	I. Characterization of H3.3 K79A mESCs	111
	Bibliography.....	114

1. Introduction

1.1 Genomic determinants of transcriptional regulation: promoters and enhancers

Regulation of gene expression is a complex and multi-layered process that involves several factors. Chromatin, the molecular entity constituted of DNA and histone proteins, represents the canvas on which diverse molecular cues are integrated to ultimately direct towards either expression or repression of genes. At the genomic level, only a small portion of the DNA carries the information to produce amino-acids and about 98% of the mammalian genome is non-coding. Throughout evolution, many of these non-coding DNA sequences have acquired crucial roles in gene expression regulation. Promoters, for instance, are DNA sequences located upstream to genes and act as platforms for the recruitment of transcription factors (TFs) and RNA polymerase II (RNA-Pol II)¹, representing the regions where transcription of genes starts [1, 2]. Only a small portion of promoters are constitutively active in all cell types (i.e., housekeeping genes' promoters) and most of the promoters require the presence of specific TFs in the nucleus in order to be activated. Transcription factors rely therefore on DNA-binding domains which confer the ability to bind specific DNA sequences or *motifs*, and eventually interpret the genomic context.

Promoters are not the only *cis*-regulatory elements that TFs can interact with and other classes of distal elements exist, such as the so-called enhancers [3]. Historically, enhancers are defined as regulatory elements capable of boosting the expression of a gene, regardless of distance, location and orientation [4]. Enhancers can indeed be located far from gene promoters, with an average distance in vertebrates ranging between 20 to 50 kb from the transcription start site (TSS) (Figure 1.1) [5]. More rare cases are also reported, in which enhancers are situated 500 kb or further away from target genes [6]. As a result, the identification of enhancers for a certain gene is not a trivial task, because besides being located far from the target genes, they can also be found either upstream, downstream or within intronic sequences², and they don't necessarily regulate only the most proximal genes [7]. The current model suggests that an enhancer regulates a target gene by establishing a direct physical contact with the gene's promoter, or at least by

¹ Three classes of polymerases exist; RNA Pol II produces messenger RNA (mRNA) and several kinds of non-coding RNAs. Mechanisms of regulation and accessory factors differ among polymerases.

² Introns are non-coding DNA sequences present in a gene, which are not retained in the final mRNA product. For their discovery, Sharp and Roberts were awarded the Nobel Prize in Physiology or Medicine in 1993.

coming in close proximity with it (Figure 1.1) [8]. The formation of such enhancer-promoter contacts is favoured and/or restricted by the organization of the genome in so-called topologically associating domains (TADs). TADs, which in mammals span 800 kb on average and up to 2 Mb, are usually well conserved among different cell types. On the other hand, the long-range enhancer-promoter contacts taking place within TADs tend to change significantly throughout differentiation, in line with the notion that specific genes need to be switched on or off to confer cell type-specific transcriptional signatures [9, 10]. Overall, except for the genomic distance and location, the logic of enhancer function is not exceedingly different from the one of promoter elements [11]. When active, cis-regulatory elements are transcribed bi-directionally and produce short non-coding RNAs, which are called enhancer RNAs (eRNAs) [12]. The production of eRNAs is therefore considered an important hallmark of these regulatory elements and is used for the identification of active enhancers in distinct cell types or tissues [13]. Enhancers are generally bound by specific combinations of TFs and such combinatorial activity is fundamental for the activation of tightly controlled regulatory programs, for instance in the context of cell differentiation and developmental processes [3, 14]. In some cases, multiple enhancers can work in a synergistic manner to regulate the expression of genes in a specific cell type, forming the so-called enhancer clusters (also known under the name of “*super-enhancers*”) [15, 16].

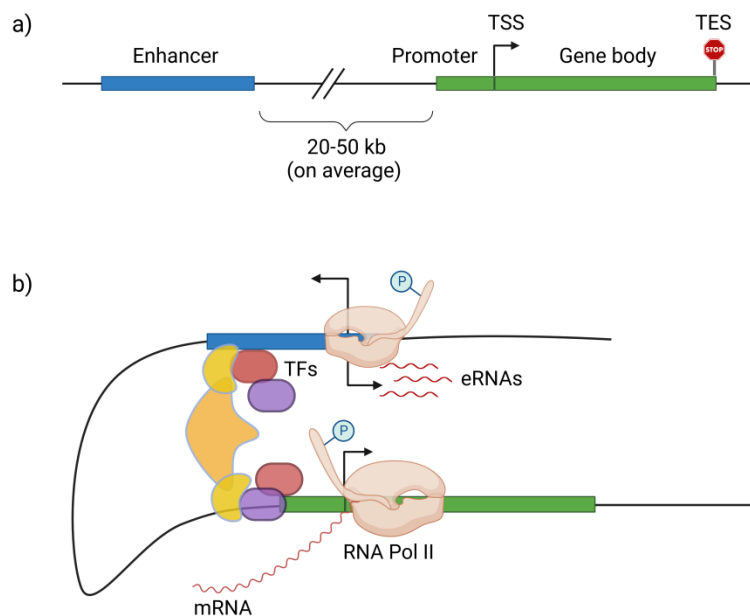


Figure 1.1 Enhancer-promoter schematic. (a) Enhancers and promoters are cis-regulatory elements involved in transcriptional regulation. (b) These genomic regions recruit TFs, transcriptional co-activators (e.g., Mediator complex) and RNA Pol II. Enhancers regulate target genes by coming in close proximity to their promoters. Figure created with BioRender.com.

Promoters and enhancers can hence interact with trans-acting transcription factors to regulate gene expression, yet these regions of DNA are protected by histone proteins. In

the next section, I will introduce the principles and players involved in chromatin structure regulation and how they can influence gene expression regulation.

1.2 Nucleosome and histone modifications

Histones are the building blocks of chromatin, responsible for highly compacting the DNA in the nucleus of eukaryotic cells. Four main canonical histones (i.e., H2A; H2B; H3 and H4) constitute an octamer that is wrapped 1.6 times with approximately 146 bp of DNA [17]. This protein-DNA complex, known as nucleosome is the basic structural unit of chromatin, [17, 18]. A common feature of the four canonical histone proteins is the presence of a structured core domain in close contact with the DNA, and an unstructured amino-terminal tail domain. The flexible N-terminal tails are the substrate of histone-modifying enzymes which deposit (i.e., *writers*) or remove (i.e., *erasers*) covalent chemical groups to specific amino-acidic histone residues. Another class of proteins is capable of interpreting (i.e., *readers*) the histone modifications and help translating them in a functional output. Although histone post-translational modifications (PTMs) were identified on the N-terminal tails, PTMs on residues of the core histone domain are also reported [19]. Several histone PTMs have been described so far, and additional newly discovered chemical modifications are continuously added to the list of known PTMs [20–22]. Nevertheless, much still needs to be learned about the four more common modifications, namely: acetylation, methylation, ubiquitination and phosphorylation.

1.2.1 Histone acetylation

In the early-1960s, based on the outcome of *in-vitro* experiments, Allfrey and colleagues formulated for the first time the hypothesis that the acetylation status of histones in the context of chromatin could be important for gene expression regulation³ [23]. Histone acetylation occurs on lysine residues and histone acetyltransferases (HATs)⁴ are the writers catalysing this reaction. HATs are a diverse group of enzymes classified into several families, based on the structural similarities of the catalytic domains [27–29]. Briefly, the main families are:

- **GNAT family**, annotated as “*GCN5-related N-acetyltransferases*”, includes GCN5 (KAT2A), PCAF (KAT2B) and HAT1, among others. Members generally present a

³ Using preparations from calf nuclei, Allfrey *et al.* monitored RNA synthesis *in-vitro* and noticed that, while addition of arginine-rich histones resulted in inhibition of the process, the supplement of acetylated histones reduced the degree of this inhibition.

⁴ The first HAT was identified by Brownell and Allis in 1996 [24, 25]; in their studies they used macronuclei obtained from the unicellular eukaryote *Tetrahymena thermophile*, which have remarkably high levels of hyperacetylated histones [26].

HAT domain (about 160 amino-acids) and a C-terminal bromodomain—a reader protein module binding acetylated histone lysine residues.

- **MYST family**, owes its name to the four main members of the group MOZ (KAT6A), YBF2 (SAS3), SAS2 and TIP60 [30]. Besides the HAT domain (about 250 amino-acids), common features are the presence of a zinc-finger binding domain and a N-terminal chromodomain for the interaction with methylated lysine.
- **p300/CBP family**, characterized by the presence of a larger HAT domain (about 500 amino-acids [31]) and several other domains, such as a bromodomain and cysteine-histidine rich domains (mainly TAZ, PHD and ZZ motifs). These latter domains can be involved in protein-protein interactions and are important for substrate specificity [32].

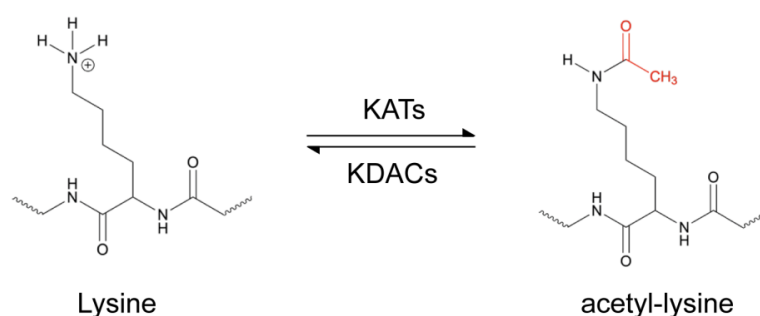


Figure 1.2 Lysine acetylation reaction. Lysine acetyl-transferases (KATs) catalyse the addition of an acetyl group on the ϵ -amino group of the residue, neutralizing the positive charge. Lysine deacetylases (KDACs) catalyse the opposite reaction. Figure created by Mariachiara Pucci (Dept. of Molecular and Translational Medicine – University of Brescia, Italy).

Histone	Acetyl mark(s)	HATs	HDACs
H3	K9ac	GCN5; PCAF; p300; CBP; MOZ; MYST2	SIRT1/6; HDAC1/2/3/8
	K14ac	GCN5; PCAF; p300; CBP; MOZ; MORF; TIP60; MYST2; TAF1	SIRT1; HDAC1/3/6/8
	K18ac	p300; CBP	SIRT7; HDAC1
	K23ac	GCN5; PCAF; p300; CBP; MOZ; MORF	HDAC1
	K27ac	p300; CBP	HDAC1
	K36ac	GCN5; PCAF; p300; CBP	/
	K56ac	GCN5; p300; CBP	SIRT1/6; HDAC8
	K122ac	p300; CBP	/
H4	K5ac; K8ac; K12ac; K16ac	HAT1; p300; CBP; MOF; MYST2; TIP60	HDAC3/6
H2A	K5ac; K7ac; K11ac; K13ac	p300; CBP; TIP60	/
H2B	K5ac; K11ac; K12ac; K15ac; K16ac; K20ac; K21ac; K23ac	p300; CBP	/

Table 1. Histone residues with respective targeting HATs and HDACs. HATs are color-coded by family (red: GNAT; blue: p300/CBP; purple: MYST)

Different HATs display a preference towards specific histones and lysine residues. For instance, with regard to histone H3, GCN5/PCAF preferentially deposits lysine 9 acetylation (H3K9ac), while p300/CBP favours acetylation of K18 and K27 residues [33]. In Table 1, I summarized a non-exhaustive list of HAT enzymes for each histone residue. HATs are generally quite promiscuous and individual histone residues can often be acetylated by several different enzymes. Most HATs can also acetylate non-histone proteins, including transcription factors (e.g., p53; YY1; GATA1; NF- κ B and others) [34–38].

1.2.2 Histone acetylation and transcription

HATs are considered important transcriptional co-activators present at active promoters and enhancers. How do HATs contribute to gene activation? The first step in the process of transcriptional activation consists in the binding of *pioneer* TFs, that have the special ability to bind nucleosomal DNA [39, 40]. The next step consists in the recruitment of both HATs and chromatin remodelling complexes that exert their activity to open chromatin and hence favour binding of Pol II and additional TFs [41, 42]. The impact of histone acetylation in this process is two-fold: i) direct effect on chromatin compaction and ii) indirect effect through recruitment of effector proteins. In terms of direct effect, acetylation of the ϵ -amino group of a lysine residue results in neutralization of a positive charge (Figure 1.2), which eases the electrostatic interaction between the positively charged histones and the negatively charged DNA molecule [43]. However, studies have shown that the interaction between histone tails and DNA is only weakened, when histone tails are hyperacetylated [44]⁵. Histone acetylation can also impact chromatin structure by providing docking-sites for the recruitment of protein factors equipped with acetyl-reader domains. Bromodomains (BRDs)⁶ are acetyl-readers often present in HATs enzymes and contribute to substrate specificity [50]. BRDs are also present in chromatin remodelling complexes, a vast group of factors capable of changing the space in between nucleosomes, by sliding or evicting them, and hence regulate chromatin accessibility [51, 52]. SMARCA2 (brahma/BRM or SNF2 α) and SMARCA4 (BRG1 or SNF2 β), for instance, are the two mutually exclusive catalytic subunits of the mammalian ATP-dependent SWI/SNF chromatin remodeller, and contain a bromodomain [28]. BET (*bromodomain and extra-terminal motif*) proteins are another representative family of BRD-containing factors, for

⁵ Histone H4K16ac was suggested to have a prominent role in affecting directly the interaction between neighbouring nucleosomes [45–48].

⁶ Bromodomains owe their name to the *Drosophila melanogaster* brahma (brm) protein, in which they were first discovered in 1992 [49].

which crucial roles in transcriptional regulation are reported [53]. BET proteins typically contain two N-terminal bromodomains for binding to acetylated histones and one extra-terminal (ET) domain for recruitment of chromatin factors, transcription factors or histone-modifying enzymes [54, 55]. The BET family comprises four main proteins: BRD2 and BRD3 (bind H4K5ac; H4K12ac) [56], BRD4 (binds H3K18ac, H3K27ac, H3K36ac, H3K56ac, H4K5ac, H4K20ac, H2AK5ac, H2AK36ac and others) [57] and the testis-specific variant BRDT. BET proteins were shown to be involved in diverse transcriptional regulatory processes: for instance, BRD4 plays an essential role in positively regulating transcriptional elongation, through recruitment of the P-TEFb complex which activates paused Pol II [58, 59]. Also, direct interaction between BRD4 and Mediator subunits has been reported, which points towards a central role of this protein in enhancer function and cell type-specific gene expression regulation [60–63]. Small-molecule BET proteins inhibitors were developed in the last couple of decades and some of them have ongoing clinical trials, mainly for treatment of cancers [64]. Besides bromodomains, other acetyl-binding modules have been identified such as the YEATS (from Yaf9, ENL, AF9, Taf14 and Sas5) domains [65] and the double plant homeodomains (double PHD fingers, aka DPFs) [66].

Histone deacetylases (HDACs) are the enzymes in charge of erasing histone acetylation and they are categorized in four main classes based on structural similarities [67]. Just like BET proteins, HDACs also represent attractive candidates for the development of small-molecule inhibitors for treatment of diseases characterized by anomalous transcriptional profiles. However, given the broad substrate specificity of HDACs and histone-modifying enzymes in general, also the potential side effects of such compounds are under constant examination [68–70].

1.2.3 Histone methylation

Unlike histone acetylation, methylation of histone residues correlates with transcriptional activation and repression, depending on the specific residue that is modified. Lysine and arginine are the two main histone residues subjected to the addition of methyl groups, and up to three methyl groups can be added to the ϵ -amino group of a lysine (Figure 3) in reactions catalysed by histone methyltransferases (HMTs)⁷.

⁷ The first HMT was described by Rea *et al.* in a study published in 2000 [71], i.e. 36 years after histone methylation was originally reported by Allfrey *et al.*, and also by Murray in an independent study published in the same year [72].

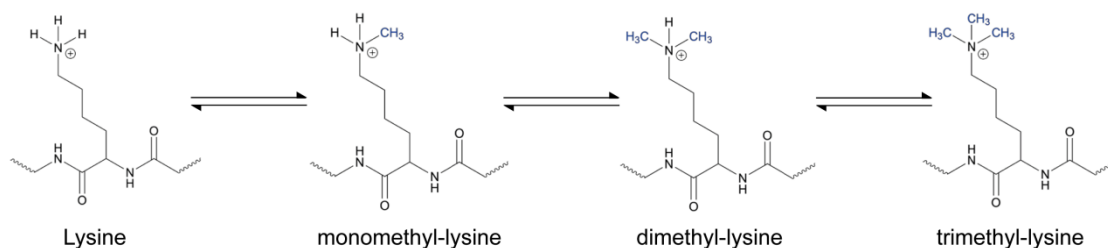


Figure 1.3 Lysine methylation reaction scheme. Lysine methyl-transferases (KMTs) catalyse the addition of up to three methyl groups on the ϵ -amino group of the lysine residue. The opposite reaction of removal of the modification is catalysed by lysine demethylases (KDMs). Figure created by Mariachiara Pucci (Dept. of Molecular and Translational Medicine – University of Brescia, Italy).

Based on the type of the catalytic domain, lysine methyltransferase are classified in two main families:

- **SET-domain family**, as of the SUV39H1 enzyme originally reported, which contained the SET (from Su(var)3-9, Enhancer of Zeste and Trithorax) domain of about 140 amino-acids [71]. This family comprises most of the HMTs known so far.
- **non-SET-domain family**, essentially composed of DOT1 and DOT1-like proteins, capable of methylating the lysine 79 residue in the globular domain of histone H3 [73, 74].

Residue	HMTs			HDMs		
	me1	me2	me3	me1	me2	me3
H3K4	SETD1A/B; MLL1/2/3/4			KDM1A/B; KDM5B	KDM1A/B; KDM5A/B/C/D	KDM2B; KDM5A/B/C/D
H3K9	SETDB1; G9a; GLP	SUV39H1/2; SETDB1; G9a; GLP	SUV39H1/2 SETDB1	KDM3A; JHDM1D	KDM3A; KDM4A/B/C/D; KDM1A; JHDM1D	KDM4A/B/C/D
H3K27	EZH2; EZH1			JHDM1D	KDM6A/B; JHDM1D	KDM6A/B
H3K36	SETD2; NSD1/2/3	SETD2; NSD1/2/3	SETD2	KDM2A; KDM2B	KDM2A/B; KDM4A/B/C	KDM4A/B/C
H3K79	DOT1L			/		

Table 2. Histone H3 lysine residues with respective targeting HMTs and HDMs.

When compared to histone acetyltransferases, HMTs generally display a higher selectivity towards one or few histone lysines. For instance, DOT1L is responsible for all three forms of H3K79me, and EZH2, the catalytic subunit of the Polycomb Repressive Complex 2 (PRC2), methylates H3K27. A peculiar example is the MLL (*Mixed Lineage Leukaemia*) family of proteins that in mammals comprises six members, all methylating H3K4 [75]. In

Table 2, I recapitulate some of the main known HMTs and their respective histone H3 lysines substrates.

Not surprisingly, in recent years some studies have shown that several HMTs can also methylate non-histone proteins [76]. One such example is the methyltransferase G9a (EHMT2)—responsible for histone H3 K9 and K56 mono-/di-methylation—that can also methylate different transcription factors [77, 78] and other proteins [79].

1.2.4 Histone methylation and transcription

How does histone methylation participate in transcriptional regulation? Unlike lysine acetylation, the deposition of one or more methyl groups does not affect the charge of the side chain of lysine residues (despite resulting in increased hydrophobicity; Figure 3). Therefore, any possible effect of the methylation is mainly ascribable to changes in protein-protein interactions. Many methyllysine-reader domains such as chromodomains, PHD fingers, WD40 repeats, PWWP and Tudor domains allow chromatin factors to interact with the histone methylation marks [80–83]. The outcome of such interactions depends on both the specific lysine residue and on the properties of the effector protein recruited. For example, histone H3K9me_{2/3} is a mark associated with transcriptional repression and represents the docking site for HP1 (*Heterochromatin Protein 1*) proteins [84–86]. The HP1 family of proteins includes three members in mice and humans, namely HP1 α (CBX5), HP1 β (CBX1) and HP1 γ (CBX3) that share a common structure consisting of one N-terminal chromodomain and one C-terminal chromoshadow domain [87]. This latter domain is important for both dimerization and interaction with other chromatin factors involved in the establishment of a repressive chromatin environment (e.g., ATRX and SUV39H1), as well as in other processes [87, 88].

Context matters when it comes to histone methylation recognition and highly similar reader modules can acquire histone residue-selectivity. As I will introduce more extensively in the next sections, histone H3K9me₃ is associated with constitutive heterochromatin, while H3K27me₃ is a hallmark of facultative heterochromatin. These two marks serve different functions and they usually don't co-localize. Nevertheless, it was shown in *Drosophila* that the HP1 chromodomain (H3K9me₃-reader) is highly similar to the Polycomb chromodomain (H3K27me₃-reader) and the segment of histone H3 tail in which K9 and K27 are located contains the same ARKS motif of residues. A comparative structure study of HP1 and Polycomb chromodomains bound to H3 peptides showed that HP1 chromodomain binding relies more on the H3 residues situated in close proximity to the methyllysine, whereas Polycomb chromodomain establishes intramolecular

attractions with residues of the H3 peptide placed more distant from the methyllysine that provide selectivity towards methylated lysine 27 [89]. This example illustrates how the same covalent modification deposited on two different histone residues can acquire a distinct functional relevance in genomic regulation, thanks to reader domains that have attained specialized abilities in interpreting the chromatin marks.

Histone lysine methylation is removed by histone demethylase enzymes⁸, a vast group of proteins which is also responsible for non-histone proteins demethylation. Histone demethylases are classified in two main groups: i) amine oxidases and ii) jumonji C-domain containing dioxygenases [82]. These enzymes typically display a certain degree of residue-specificity, with most of the demethylases that can act on one or only a few lysines (some examples are also reported in Table 2).

During my PhD I focused on studying three histone residues, namely H3 lysine 9, lysine 27 and lysine 79. I will summarize in the next section specific features of these residues.

1.3 The residues studied in my PhD

1.3.1 H3K9 acetylation and methylation

Like virtually every lysine, the H3 K9 residue can be either acetylated or methylated. H3K9ac is linked to active transcription and for instance it highly associates with H3K14ac and H3K27ac at active distal regulatory elements in mESCs [91]. H3K9ac also decorates active promoters and a subset of poised promoters (referred to as *bivalent*, see section 1.4.1 for details) [91]. H3K9ac is deposited by GCN5 and PCAF HATs; GCN5 is part of the SAGA (*Spt-Ada-Gcn5 acetyltransferase*) complex which was shown to directly bind H3K4me3 in yeast and thus increase deposition of H3K9ac at promoter regions [92]. From a mechanistical point of view, H3K9ac might play a central role in promoting Pol II release and hence favour transcriptional elongation. H3 peptides containing the K9ac mark bind the main subunits of the super-elongation complex (SEC), such as AF9 and ENL (equipped with the K9ac-reader YEATS domain [93]) as well as CDK9 and Cyclin T1 (i.e., P-TEFb complex) [94]. Pol II pausing index⁹ was increased at a subset of genes upon GCN5 inhibition [94], however, a study in *Drosophila* showed that upon introduction of a loss-of-function mutation in one of the dSAGA complex subunits and consequent reduction of the K9ac mark, the expression of only a small number of genes was affected [96]. This and

⁸ Giving temporal perspective to what we study highlights how fast the chromatin biology field developed in the last decades: until the discovery in 2004 of the first histone demethylase called LSD1 (KDM1A) [90], histone methylation was considered to be a stable chromatin mark (unlike acetylation).

⁹ The pausing index is defined as the ratio between promoter proximal and gene body Pol II signal [95].

other studies indicate that the redundant presence of acetylation on many histone lysines can buffer the decreased acetylation levels on one or a few residues, which will likely translate in limited consequences on transcriptional programmes.

As opposed to acetylation, H3K9 methylation is typically associated with transcriptional repression. The previously introduced G9a together with GLP (*G9a-Like Protein*, aka EHMT1¹⁰) are the HMTs responsible for H3K9me1/2 deposition in mammals [98, 99]. These enzymes constitute a heterodimer *in-vivo* and the formation of this complex is crucial for the methyltransferase activity [100]. They are primarily involved in depositing H3K9me1/2 in euchromatic regions of the genome, where they contribute to gene silencing. H3K9me2 usually covers long stretches of the genome, with these K9 methylation *blocks* that tend to increase in the transition from embryonic to differentiated cells [101]. The establishment of such large H3K9me2 domains is thought to be favoured by two mechanisms: i) direct H3K9me2-binding of G9a/GLP through reader domains [102] and ii) HP1-mediated H3K9me2 recognition and recruitment of G9a/GLP [99]. G9a/GLP-dependent H3K9me2 is also considered an important player in the process of X chromosome inactivation and in regulation of imprinted genes [98, 103].

Deposition of H3K9me3 is catalysed by a different set of enzymes, namely SUV39H1, SUV39H2 and SETDB1 (ESET). As revealed by gene knockouts studies in mice, the SUV39H proteins are crucial contributors for H3K9me3 build-up and consequent heterochromatin formation at pericentromeric and telomeric regions [104–106]. SETDB1, on the other hand, is mainly in charge of depositing the trimethyl K9 mark to ensure euchromatic genes silencing [107] as well as repression of proviral sequences [108]. To exert these functions, SETDB1 relies on the presence of the co-repressor known as KAP1 (*KRAB-Associated Protein 1*, aka TRIM28). This protein interacts with the KRAB (*Kruppel-Associated Box Repression*) domain found in a large family of zinc-finger TFs that act as transcriptional repressors [109], and hence targets SETDB1 to the genomic loci destined to be repressed. For example, a RNA interference screening conducted in mESCs, identified SETDB1 (but also GLP and SUV39H2) as an important factor for silencing of developmental genes [110]. Notably, in line with a crucial role of this histone PTM in development, a recent study conducted in mouse embryos showed that H3K9me3 is dynamically redistributed to allow activation of the correct set of lineage-specific gene programs [111, 112].

The KAP1-SETDB1 pathway is also involved in the repression of a specific class of transposable elements known as endogenous retroviruses (ERVs) and I will return to this

¹⁰ The acronym EHMT is an abbreviation for *Eu-HMTase* or *euchromatin histone methyltransferase* [97].

in more detail in the section “1.7.1 Epigenetic mechanisms of ERVs repression”. In general, where there is H3K9me2/3 then HP1 proteins can bind. According to the most accredited model, HP1 ensures spreading of the H3K9 methylation on adjacent nucleosomes and helps to maintain the heterochromatic state [113]. However, it is noteworthy that the presence of HP1 proteins (and of other H3K9me-readers) is not required to guarantee silencing of ERVs in mESCs [114].

1.3.2 H3K27 acetylation and methylation

Histone H3 K27 acetylation has acquired particular relevance in genomics and chromatin biology because it typically decorates active enhancers [115–117]. At enhancers, H3K27ac occurs together with H3K4me1 and the presence of both modifications at distal genomic regions is widely used to predict the location of enhancers in a cell type-specific manner. In terms of mechanism, H3K4me1 is deposited at enhancers by the MLL3/4 enzymes for commissioning of distal regulatory elements, where it precedes and favours p300/CBP-mediated histone acetylation [118, 119]. How p300/CBP is directly recruited by H3K4me1 is still not clear, but a direct cross-talk between H3K4me1 and the acetyltransferase TIP60 (i.e., H2A and H4 acetylation) was described [120]. This reinforces the concept that, besides the good predictive power for activity, it is limiting to grant H3K27ac an excessive functional role in enhancer regulation. A recent study in mESCs also reported critical roles of p300/CBP activity in PIC assembly and Pol II recruitment at enhancers, preceding acetylation-mediated BRD4 recruitment which induces Pol II release [121].

Similarly to other histone acetylations, H3K27ac is found at the promoter of active genes. Thus, the histone H3 K27 residue is disputed between two enzymatic activities: p300/CBP on one side that deposit acetylation to promote activation, and EZH2/1 on the other side that deposit methylation to favour repression. This antagonistic activity has been extensively documented in several model systems such as *Drosophila* and mESCs [122, 123]. A recent study in mESCs, using a combination of knockout and catalytically-dead EZH2 and EZH1 lines, showed that H3K27ac takes over at Polycomb target promoters when H3K27me2/3 is completely wiped out [124]. To preserve this balance, it is really the presence of the K27 methylation—rather than the recruitment of the enzyme—to be essential [124].

Although the tri-methylated form of H3K27 has always drawn great attention due to the well documented role in transcriptional repression, also mono- and di-methylated forms of this residue exist. H3K27me1 is generally present at the gene bodies of actively transcribed genes, while H3K27me2 is a broadly distributed mark present at 70% of total

histone H3 in mammals [125]. A suggested role for this latter mark is to prevent improper activation of enhancers [125]. All forms of K27 methylation are catalysed by the Polycomb repressive complex 2 (PRC2)¹¹, which can contain two mutually exclusive catalytic subunits: i) EZH2, more abundant in stem cells and better performing HMT, hence critical for H3K27me3 deposition and ii) EZH1, more abundant in post-mitotic cells and weaker HMT [127]. Besides the catalytic subunit, the PRC2 complex includes three additional core subunits known as SUZ12, EED and RBBP4/7 [128]. The core subunits are essential for the catalytic activity of the complex, likely because they stabilize the interaction between PRC2 and the nucleosome substrate [129]. Based on the presence of ancillary subunits, PRC2 complexes can then be distinguished in two main variant complexes called PRC2.1 (with MTF2, PHF1, or PHF19 + EPOP or PAL1/2) and PRC2.2 (with JARID2 and AEBP2)[128]. Although not directly responsible for H3K27 methylation, another complex known as Polycomb repressive complex 1 (PRC1) join forces with PRC2 to ensure silencing of target genes. Many different variant complexes have been reported for PRC1 (which I will not describe here), but the critical subunits are the RING1A/B (*Really Interesting New Genes 1A/B*) E3 ligases that catalyse mono-ubiquitination of histone H2A on lysine 119 (H2AK119ub1) [130]. This modification can be read by the JARID2 and AEBP2 subunits of PRC2.2. At the same time, the H3K27me3 mark is read by the CBX proteins (e.g., CBX7) of PRC1 and this establishes a positive feedback loop between the activities of the two complexes (Figure 1.4) [128]. Transcriptional repression of target genes through PRC2-mediated H3K27me3 is crucial for both maintenance of pluripotency and cell differentiation, however how PRC2 and PRC1 are targeted to specific loci in the genome remains an open question. The so-called Polycomb response elements (PREs) present in *Drosophila*—described as DNA motifs capable of recruiting specific TFs and in turn Polycomb complexes—are not conserved in mammals. Nevertheless, a common feature of the PRC2/1 target promoters is the high CpG content (also referred to as CpG-islands or CGIs¹²) [132] and some PRC2 accessory subunits, like MTF2 and PHF1, have affinity for these DNA sequences [133]. In addition, PRC2 can interact with its own histone PTM via the core subunit EED (Figure 1.4) [134]. This H3K27me3-binding is critical to ensure spreading of the modification on adjacent nucleosomes in a mechanism known as *allosteric activation* [134, 135].

¹¹ The name Polycomb comes from the phenotype observed in *Drosophila* in 1947: mutation of this gene leads to appearance of ectopic bristles called sex combs. Three decades later, it was reported that this is caused by ectopic Hox genes expression, causing a posterior transformation [126].

¹² CGIs are on average 1kb long, show an elevated G+C base composition and are in general devoid of DNA methylation. In vertebrates, about 70% of annotated promoters are associated with a CGI [131].

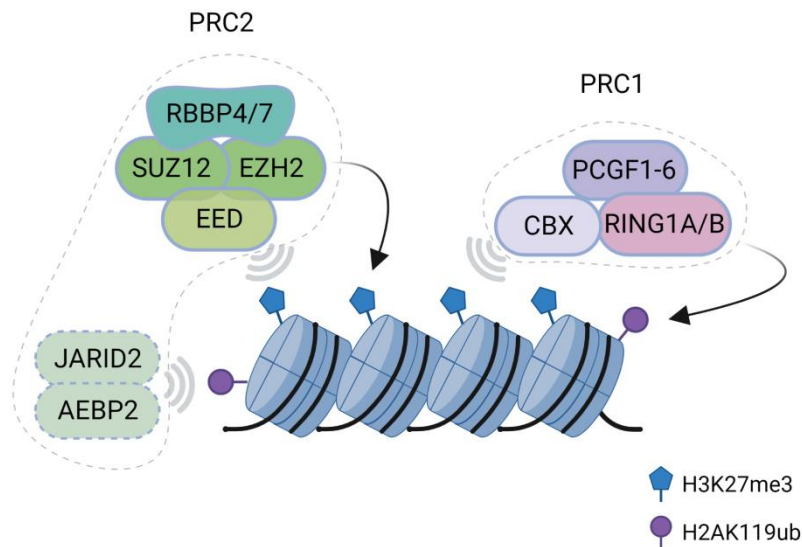


Figure 1.4 Schematic of PRC1 and PRC2 activity and cross-talk. The PRC2 complex catalyses the deposition of histone H3K27 methylation. The H3K27me3 mark can be “read” by the CBX proteins of the PRC1 complex, which deposits the H2AK119ub mark. PRC2 can interact with the H3K27me3 mark (through the core subunit EED) and also with the H2AK119ub mark by means of the JARID2 and AEBP2 subunits. Figure created with BioRender.com.

Presence of histone H3K36me_{2/3} can instead inhibit PRC2 activity both *in-vitro* and *in-vivo* [136, 137]. The Tudor domain contained in the PHF1 subunit binds to H3K36me₃, and this interaction results in decreased PRC2 activity [137]. The significance of this cross-talk is that the active mark H3K36me₃ decorates gene bodies of transcribed genes and the interaction of PRC2 with this PTM can prevent the spreading of H3K27me₃ on genes that need to be expressed, but also prepare the ground for repression of pluripotency-associated genes during differentiation [138].

1.3.3 H3K79: DOT1L-mediated methylation and proposed functions

The distinctive feature of the lysine 79 residue is that it lies within the histone H3 globular domain, and not on its N-terminal tail. H3K79 is reached by the methyltransferase DOT1L¹³ (KMT4) which is the only HMT targeting this residue known so far. As introduced in the previous section, DOT1L lacks a SET domain [73]. It is unable to methylate its target on free histone H3 and the catalytic process is favoured by a direct cross-talk with other histone PTMs in the chromatin context. The main interaction occurs with the mono-ubiquitinated lysine 120 of histone H2B (H2BK120ub1). Recent structural studies showed that DOT1L binding to H2BK120ub1 restricts the enzyme in a position that promotes the catalysis and, aided by the histone H4 tail, induces a conformational change of the H3

¹³ The yeast ortholog known as DOT1p (*Disruptor Of Telomeric silencing 1p*), was originally identified in *S. cerevisiae* in a screening aimed at finding proteins involved in telomeric silencing [139].

globular domain which reorients K79 in a position suitable for the reaction [140, 141]. H3K9ac also plays a role in promoting K79 methylation, specifically through the binding of the AF9 protein (via YEATS domain) that can recruit DOT1L [93]. Another example is the DOT1L cofactor AF10 that binds unmodified H3K27 and favours H3K79me2 deposition [142]. This specific interaction is prevented in the presence of H3K27me3 [142]. Very recently, one more cross-talk was described, between DOT1L and H4K16ac, which also results in activation of the enzyme [143]. Therefore, H3K79 methylation is generally considered an active chromatin mark. H3K79me2/3 is found on the gene bodies of actively transcribed genes, with a signal more enriched in proximity of the TSS [144]. H3K79me1 is instead a more broadly distributed mark, generally spanning intergenic regions [144].

Some studies reported the presence of H3K79 methylation at enhancers. This observation was first made in *Drosophila*, where H3K79me3 is found at specific developmental enhancers [145]. Most of the investigations in this direction were conducted in cell lines carrying MLL rearrangements, in which H3K79me2 was enriched at super-enhancers [146]. By means of DOT1L inhibition, a potential functional role of H3K79me2/3 in maintaining these enhancers in an active state in MLL-AF4 leukemic cells was reported [147, 148].

DOT1L knockout is embryonic lethal in mice and studies suggested that DOT1L and H3K79 methylation levels could be crucial for proper epigenetic reprogramming during early developmental stages, with a particular relevance in cardiac differentiation [149, 150]. Also, while mESCs can survive DOT1L deficiency and low levels of H3K79 methylation, the HMT is required for proper differentiation [151]. Despite DOT1L being associated with active transcription, mESCs lacking DOT1L showed defects in heterochromatin formation and in distribution of heterochromatin-associated marks (i.e., H4K20me3 and H3K9me2), as well as in telomere structure [149]. These results point towards multiple roles of DOT1L, which might go beyond deposition of the H3K79 methylation. In this regard, DOT1L catalytic activity is dispensable for transcription elongation under stress condition as well as for lineage specification, while DOT1L KO shows more severe effects [152].

So far, no demethylase enzymes targeting H3K79 have been discovered (with the exception of one study that reported KDM2B as a potential candidate [153]). Furthermore, to the best of my knowledge, there are no studies yet reporting to what extent H3K79 acetylation occurs *in-vivo* and what could be the potential molecular effects of this modification.

1.4 Histone modifications define chromatin states

In 2007, two landmark papers profiled histone modifications genome-wide [154, 155] paving way for the concept of *chromatin states*, recurrent patterns of histone modifications decorating genomic regions that share common features [156]. Nowadays, the presence of specific modifications, or a defined combination of those, is interpreted as a hallmark of specific chromatin environments and histone marks are commonly used to infer the transcriptional state of genes or the activity of genomic regions.

Most chromatin states are extremely conserved; here I will briefly summarize the most common ones (Figure 1.5). Active promoters are generally marked by H3K4me3 together with several acetylation marks like H3K9ac and H3K27ac, among others. The gene bodies of active genes are often decorated with H3K79me3, together with H3K36me3 more spread towards the 3'-end. Histone H3K4me1 is found at enhancer elements, which are considered active when gaining also H3K27ac.

Histone H3K9me3 and H4K20me3 mark genomic regions stably maintained in a repressed state—referred to as constitutive heterochromatin—which comprehends pericentromeric regions and telomeres [104]. These marks are also found at transposable elements (TEs)—repetitive or low-complexity DNA sequences of various nature—long considered inert, but for which possible regulatory activities were reported. Genomic regions containing genes that are switched off, but that might be activated depending on developmental stage or cell type, are instead referred to as facultative heterochromatin [157]. H3K27me3 usually decorates the promoters of such repressed genes. Other modifications such as H2AK119ub1 and H3K9me2 also serve for local gene silencing.

In some cases, histone modifications that are typical of active or repressed states can co-exist. It is the case of bivalent promoters which present both the active H3K4me3 and the repressive H3K27me3 marks. As I will come back to these genes in the Results section, I will focus here on describing some features of bivalent genes in more detail.

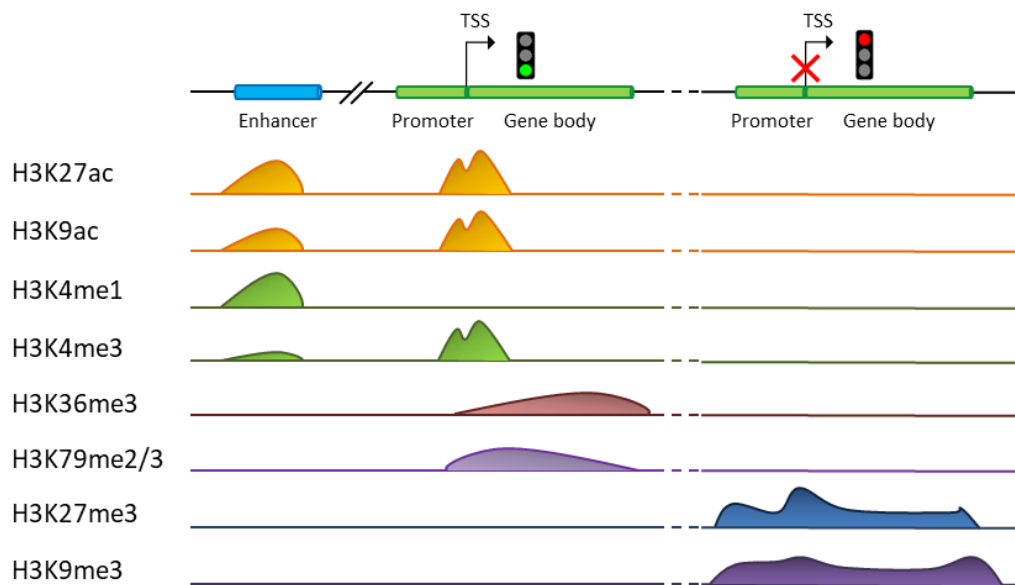


Figure 1.5 Histone modifications and chromatin states. Schematic of genomic location and distribution of selected histone modifications.

1.4.1 Bivalent promoters: a closer look

The Polycomb group (PcG) and Trithorax group (TrxG) proteins were originally characterized in *Drosophila melanogaster* as two opposing molecular machineries capable of influencing the repression or activation of developmental genes—by depositing the H3K27me3 and the H3K4me3 marks—respectively [158]. Later studies conducted in embryonic stem cells showed the concomitant presence of these two activating and silencing chromatin marks on promoters of lineage-specific genes, henceforth referred to as *bivalent* chromatin state [159]. Bivalency was described in several metazoans since then, but mESCs are the system in which it remains best characterized.

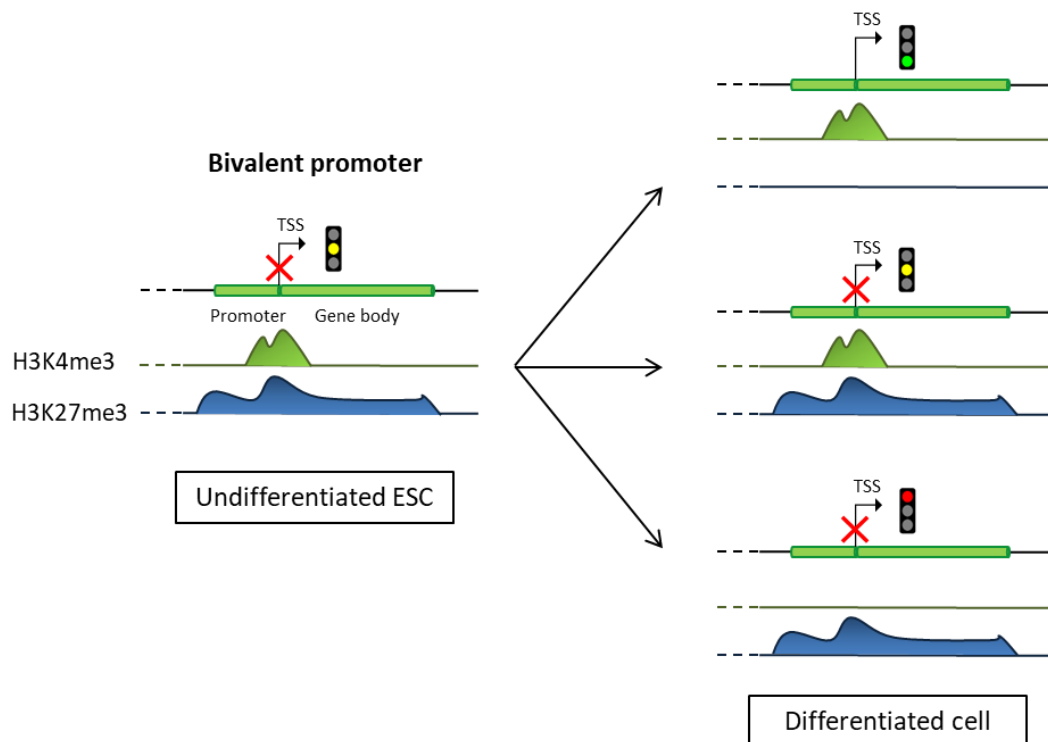


Figure 1.6 Bivalent promoters. Bivalent promoters are defined by the co-occurrence of the H3K4me3 and H3K27me3 marks. This bivalency can eventually be resolved during differentiation, whereby only one of the two modifications is retained depending on the transcriptional state of the gene.

The general concept is that the expression of developmental regulators—such as lineage-specific TFs—needs to be tightly regulated when differentiation takes place. In this regard, a bivalent chromatin state might poise genes and prepare them to be either activated or permanently repressed in the presence of the right molecular circumstances (Figure 1.6). This model derives from the observation that the bivalent state of several genes is resolved in differentiated cells [159], and bivalent promoters generally show enrichment for TF motifs of developmental regulators [160]. During differentiation, histone H3K4 and H3K27 demethylases are crucial in resolving the bivalent state [161, 162], but interestingly also chromatin remodellers and activation of distal “*poised enhancers*” play a role [117, 161]. From a phenotypic perspective, loss of bivalent promoter state doesn’t affect mESCs self-renewal but rather results in improper lineage specification [161].

MLL2 (KMT2B) is the methyltransferase primarily involved in H3K4me3 deposition at bivalent loci [163]. PRC2 and PRC1 are found together at ~40% of bivalent promoters, the ones often coding for developmental TFs [160]. Removal of H3K27me3 from bivalent promoters (e.g., as consequence of PRC2 subunits KO) doesn’t always translate in transcriptional activation. This can be likely due to the lack of the proper activating molecular cues, or conversely to the presence of additional repressive mechanisms. For instance, SETDB1 was reported to occupy a subset of bivalent genes in mESCs and

contribute to their repression [110]. H3K9me2 and G9a/GLP were also reported to be present at specific bivalent loci, which were ineffectively repressed upon G9a depletion [164]. Of note, the histone variant H3.3 was proposed to favour a dynamic chromatin environment at bivalent promoters to facilitate PRC2 recruitment [165].

1.4.2 Histone modifications: cause or (dispensable) cog? Insights from latest studies, evolutionary perspectives and synthetic biology

Specific histone modifications display high correlation with defined transcriptional features, however in most of the cases it is still largely unclear whether these marks have a direct impact on the regulation of transcription. Two key questions arise: does the histone PTM or the activation/repression come first? Are histone PTMs a dispensable gear in the transcriptional regulatory machinery?

The idea that histone PTMs are instructive in dictating either activation or repression of genes comes from the *histone code* hypothesis formulated in the year 2000 by Strahl and Allis [166]. According to this hypothesis, histone marks or combinations of those are interpreted by reader proteins and ultimately dictate whether genes will be activated or repressed. This idea has been challenged by several studies, in which drastic reductions in the levels of histone PTMs considered to be crucial had negligible effects on gene expression [167]. The comparison of downstream effects of the complete knockout for a certain histone-modifying enzyme with the effects of catalytically inactive mutants, has shown that histone writers have functions independent from the deposition of the chemical modifications. For instance, mESCs with KO of MLL3/4 enzymes responsible for H3K4me1 deposition at enhancers showed loss of the mark and also drastic decrease in H3K27ac at enhancers as well as changes in gene expression patterns. However, MLL3/4 catalytically-dead mutants displayed much less severe effects on both H3K27ac and gene expression [168, 169]. Despite being dispensable at the ESC-stage, MLL3/4 catalytic activity appeared to be crucial for proper gene regulation upon induction of differentiation [169]. Similar outcomes were reported for the SETD1A enzyme depositing another active histone PTM, such as promoter H3K4me3: substantial defects in *null* mESCs and reduced effects in catalytically-dead mutants, but with defects that appear upon differentiation [170, 171]. Another study from the Shilatifard laboratory aimed at characterizing the relevance of H3K4me3 deposition in the context of bivalent promoters. The conclusion was that the MLL2 enzyme in this context might act to prevent silencing activities (i.e., DNA methyltransferases and PRC2) rather than directly promoting gene activation [172].

Notably, the reacquisition of H3K4me3 in MLL2 KO cells was not necessary to rescue gene expression.

On the cause-and-consequence matter regarding histone PTMs, I would like to mention here two recent intriguing studies. First, Martin and colleagues observed in yeast global and local loss of H3 and H4 acetylation upon treatment with a compound (i.e., 1,10-phenanthroline) that inhibits Pol II recruitment [173]. Based also on other experiments, the authors concluded that HATs depend on transcription for their activity¹⁴ [173]. Second, Wang & Chivu et al. predicted the location and distribution of histone PTMs using nascent transcriptomics data [176]. The authors then used a small molecule (i.e., Triptolide) to block Pol II and study the relation between transcription and histone marks. Upon treatment, global levels as well as promoter H3K4me3 and H3K27ac signals were reduced. The authors concluded that deposition of these two active histone PTMs is dependent on Pol II transcription [176]. These studies showed once again that histone marks and transcription correlate very well with one another. Nevertheless, there is need for caution in claiming that histone modifications do not serve an essential regulatory role. Even on the assumption that the inhibitors used in the two studies to block Pol II have negligible side effects, this experiment is in my opinion not different from other experiments in which transcriptional co-activators (e.g., HATs) are inhibited/deleted and this translates in defective transcription [177]. One important aspect to take into consideration is that histone proteins and their modifications are extremely conserved [178]. If “*nothing in biology makes sense except in the light of evolution*”¹⁵, why would histones be so conserved across ~1.5 billion years of eukaryotes evolution? In a study that looked at the conservation of histone PTMs across species, the authors suggested that defined genomic features (e.g., GC content, CTCF binding and clustering of TFBS) might help stabilize histone modifications throughout evolution [179]. A positive feedback loop between histone modifications and TF binding may then have evolved in eukaryotes to ensure stability of regulatory networks [179]. Some important lessons can also be learned from recent advances in synthetic biology. For instance, CRISPR-Cas9-based transcriptional activators are nowadays a widely used tool to induce expression of specific genes of interest [180]. A study comparing CRISPR-based activators showed that targeting of p300 or CBP is sufficient to induce transcription, but at the same time gene activation can be detected in some cases without a concomitant H3K27ac increase [181]. Another recent preprint took advantage of a synthetic system to assess the effect of *de-novo* and

¹⁴ For the sake of completeness, a comment and a reply to it were published after the paper; the attention was brought on the fact that the specific compound used to inhibit Pol II might affect also TFs and HATs/HDACs recruitment [174, 175].

¹⁵ Famous quote from the geneticist and evolutionary biologist T. Dobzhansky (1900–1975).

controlled deposition of histone PTMs [182]. A crucial conclusion presented is that the effects of depositing histone PTMs are variable and often depend on several factors, such as the presence of specific TFs motif [182]. So how do we put all this in a coherent model? It is likely that individual histone modifications might not have an instructive role in directly causing transcription. Nevertheless, even not being causative doesn't mean they are useless or mere by-products. Based on both past and more recent literature, I think that a "*cog model*" (to which also the paper from the Danko lab referred to [176]) nicely summarizes the great complexity of transcriptional regulation: a "high-precision mechanical watch" in which histone modifications represent one of the important—and sometimes redundant—gears. Finally, repressive histone marks might be indeed instructive in driving repression, while activating marks would act to prevent such non-permissive state, rather than being strictly necessary for transcriptional activation [172, 183].

1.4.3 Studying the contribution of histone modifications: limitations of knockouts, inactivating mutations and chemical inhibition of histone-modifying enzymes

Probing the contribution of histone modifications is not a trivial task. Genetic intervention and chemical perturbations of histone-modifying enzymes can be in theory extremely effective when a specific histone mark is known to be deposited by only one enzyme. Nonetheless, a growing body of literature confirms that this is likely to be a simplistic scenario and hardly happening in living cells. In mESCs with EZH2 KO, levels of H3K27 methylation are maintained by the non-canonical PRC2 complex that contains the methyltransferase EZH1, able to partially compensate for the absence of its homolog [124, 184]. So even if the levels of a specific PTM are severely reduced genome-wide upon the deletion of dedicated enzymes, compensatory mechanisms might help to retain the modification on specific genomic locations via non-canonical or less characterized pathways. Cases in which a one-to-one relation between writers and histone marks take place are also extremely rare. Most of the PTMs like K9 methylation can be deposited by several distinct enzymes and hence is necessary to generate multiple knock-outs to deplete the levels of the PTM, with the risk to introduce severe collateral effects. On the other hand, most of the histone writers can modify many different residues and the genetic deletion in this case is either lethal or makes very hard the interpretation of phenotypic effects. Another relevant aspect is that histone writers are typically one component of larger protein complexes. Their deletion can lead to the destabilization of

such complexes and therefore to the loss of functions that might be independent from the catalytic activity of the deleted subunits. As already discussed, this drawback can be overcome by introducing mutations that abolish only the catalytic activity of the histone writers, without affecting their expression. However, the pleiotropic activity characteristic of many histone-modifying enzymes—often capable of exerting their function on multiple histone residues and on non-histone proteins—poses a limitation to this approach.

More recently, the growing interest in the epigenetic drugs and in the possibility to modulate the activity of epigenetic factors for therapeutic purposes, paved the way to the use of small-molecules as a complementary approach to study the activity of histone modifying enzymes [185, 186]. When chemical perturbations are employed as an alternative approach, an additional layer of complexity is introduced, and must be considered in the interpretation of the results. Compounds designed to inhibit histone modifiers could also act on different targets, leading to interference with other cellular pathways. Substrate specificity of such compounds has to be carefully validated as a recent study reported how several widely used HAT inhibitors show poor target selectivity both *in-vitro* and *in-vivo* [187]. Inferring the role of histone modifications through the inhibition of histone writers can be therefore challenging (Figure 1.7).

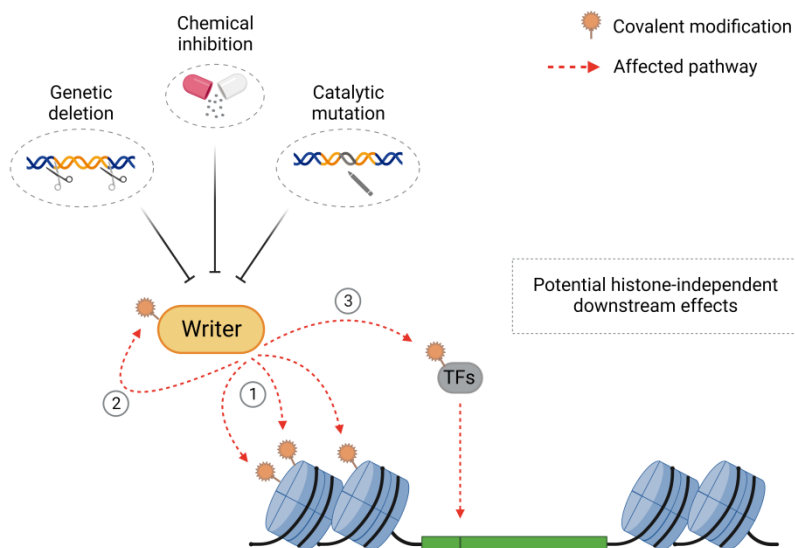


Figure 1.7 Chemical and genetic perturbation of chromatin factors. Histone writers can often deposit modifications on multiple histone residues (1). Additionally, they can also modify themselves (2) as well as other non-histone substrates such as TFs (3). Therefore, when chemical or genetic perturbations of such enzymes is applied, multiple cellular pathways can be affected (red dashed arrows). Figure created with BioRender.com.

1.4.4 Histone mutagenesis to investigate the function of histone residues and histone PTMs

A more direct approach to expand and refine our understanding of histone PTMs consists in the generation of histone mutants. By mutating modifiable histone residues (e.g., lysine) to non-modifiable ones (e.g., alanine or arginine) the observed phenotypic effects are a direct consequence of the absence of specific residues and of their chemical modifications, hence removing layers of complexity that are inevitable when histone writers are genetically ablated or chemically inhibited (Figure 1.8).

This approach has been systemically applied in *Saccharomyces cerevisiae* in which nearly all the residues of the four histones have been mutated and their phenotypes described [188–191]. Overall, these studies confirmed that mutations on specific residues can affect transcription, nucleosome positioning, heterochromatin organization and DNA damage response. Moreover, 18 residues in the globular domains of the four histones have proven to be essential for viability in yeast [191]. In yeast, the genetic dissection of histones function is possible due to the fairly simple organization of histone genes, with each one of the four histones encoded by only two non-allelic copies [192]. However, the matter gets more complicated in multicellular and higher order eukaryotes. In metazoans, each canonical histone is encoded by ten or more genes organized in clusters that are located on different chromosomes [193, 194]. These genes lack introns, their expression is coordinately regulated during S-phase, and the mRNA produced lacks a poly(A) tail which is replaced by a characteristic 3'-stem-loop structure [194, 195]. The presence of multiple gene copies coding for the same histone protein, poses a severe obstacle to mutational studies of histones in metazoans. Up until recently, among all multicellular organisms this goal was accomplished only in *Drosophila melanogaster*. The fruit fly represents an exception among metazoans because all canonical histone genes are organized in tandem repeats located at a single chromosomal locus, and different gene editing tools (e.g., BAC-based transgenes and CRISPR-Cas9) were used to replace the wild-type genes with mutated versions [196–198]. Collectively, these studies showed that histone function gained complexity during the evolution of multicellular organisms, to the effect that many modifiable residues located on the histone tails (e.g., histone H3 K4; K9; K14; K27 and K36) are essential for viability in *Drosophila*, unlike in yeast. Among the mutants that survive the early developmental stages, many show reduction in fertility (e.g., H4K16A; H3T3A; H3K56A), susceptibility to DNA damage (e.g., H3K56A) as well as defects in gene silencing (e.g., H3K9A and H3K23A) [198]. Therefore, with the increasing complexity of multicellular organisms, histone residues may have acquired new roles in DNA-related

processes, especially considering the importance of cell type- and developmental stage-specific gene expression programmes that need to be tightly regulated during the life cycle of high order multicellular eukaryotes [199].

It was only this year that the first study was published in which Sankar, Mohammad, Sundaramurthy *et al.* managed to mutate the lysine 27 codon (i.e., AAG) to an arginine codon (i.e., AGG) in all the endogenous canonical histone H3 gene copies of mESCs [200] (due to its relevance for my project, I will come back to some of the salient findings of this paper later on). The authors employed an adenine base editor which converts adenine to guanine. In particular, they took advantage of the fact that the DNA sequence in proximity of the K27 codon is conserved in all the 12 canonical H3 genes in mice, meaning that all genes could be targeted using one single Cas9 guide RNA (gRNA). While this approach could be applied in the near future to mutate other residues of canonical histones in mammals, with the gene editing tools currently available this remains a challenge and the sequence similarity that favoured editing of the K27 is not a recurrent theme.

How can we then overcome this problem to apply histone mutagenesis in mammalian systems? The approach that I used in my PhD project is to take advantage of a specific histone variant. Histone variants are isoforms of the canonical histones which have acquired specialized functions in chromatin regulation. Mainly histone H3 and H2A variants have been best characterized so far, and incorporation of such isoforms is thought to contribute to the establishment of specific chromatin states. Importantly, unlike canonical histones, one or a couple of genes expressed throughout the cell cycle encode histone variants. The presence of few genes—not included in histone genes clusters—makes histone mutagenesis feasible in mammalian systems, paving the way for a systematic dissection of histone residues function in mammals.

In the next sections I will introduce the histone variant H3.3, main focus of my PhD, and also illustrate the current state of mutagenesis studies of H3.3 (also recently reviewed in Trovato *et al.* [201]).

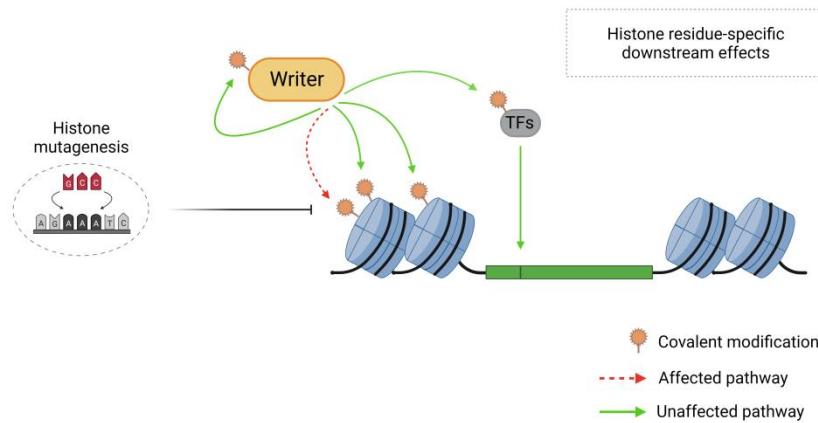


Figure 1.8 Histone mutagenesis approach. Mutation of specific histone residues (e.g., H3.3 residues in mammalian systems) is expected to elicit downstream effects which are residue-specific. The deposition of the modifications on other histone residues, and modification of other non-histone substrates is not affected (green arrows). Figure created with BioRender.com.

1.5 A brief summary of histone H3 variants

The number of histone H3 isoforms differs among species, but there are two main subgroups known as *replication-dependent* and *-independent* variants. In humans and mice, H3.1 and H3.2 are the canonical isoforms usually referred to as *replication-dependent* histones as they are deposited into chromatin in a process coupled with replication and their expression peak occurs in S-phase. H3.1 and H3.2 genes are present in multiple copies that are organized into clusters. Moreover, H3.1 and H3.2 genes do not have introns and their gene transcripts lack poly(A) tails but instead they contain a 3'-stem loop structure important for the tight regulation of replication-dependent histone gene expression. With regard to the histone H3 replacement variants, several specialized isoforms exist in different species. For instance, the two testis-specific H3 variants H3.4 (or H3.t) and H3.5 have been reported in primates [202]. Similarly, H3.Y (or H3.Y1) and H3.X (or H3.Y2) have been found in primate neuronal subpopulations [203]. The variant which is most conserved in function is the centromeric H3 variant CenH3 (aka CENP-A). This variant is specifically localized at centromeres and plays an essential role in regulating chromosome segregation [204]. CENP-A null mice—besides premature death—show severe defects in the mitotic process, supporting the hypothesis that this variant is crucial for kinetochore attachment and acts as an epigenetic mark for the formation of the centromere [205].

1.6 Histone H3.3

The histone variant H3.3 represents the most ancestral histone H3 isoform. Phylogenetic analyses showed that the histone H3 found in fungi is indeed more closely related to H3.3, compared to the canonical H3.1 and H3.2 variants (and for this reason often referred to as H3.3-like). In many metazoans species (e.g., human, mouse, *Xenopus* and *Drosophila*) the H3.3 protein sequence is the same [178]. The primary structure of H3.3 does not present many differences when compared to the canonical H3.1 and H3.2 variants, with only five and four different residues respectively. Three of the distinctive residues (i.e., A87, I89, G90) are located in the so-called histone chaperone recognition domain and they are unique to H3.3; a fourth residue of the histone fold domain (i.e., S96) is present in both H3.3 and H3.2, while H3.1 has an alanine in this position. Serine 31 is the only distinctive H3.3 residue located on the N-terminal tail [206]. The different residues in the core domain are critical for the recognition of H3.3 by dedicated protein chaperones: the HIRA and the DAXX/ATR complex. The HIRA complex is composed of three main evolutionary conserved factors, known as HIRA, CABIN1 and UBN1 [207], and strictly cooperates with the ASF1 chaperone for loading H3.3 onto chromatin. UBN1 (*Ubinuclein-1*) specifically interacts with the G90 residue of H3.3, thus conferring binding specificity for H3.3 to the whole HIRA complex [208]. HIRA deposits H3.3 in genomic regions with an open chromatin state, such as promoters, enhancers and gene bodies of actively transcribed genes [209, 210]. With regard to the DAXX/ATR (*death domain-associated protein 6/α-thalassemia/mental retardation syndrome X-linked*) chaperone complex, structural and biochemical studies showed that this chaperone specifically binds H3.3 on the basis of the selective recognition of the G90 (and to a lesser extent A87) residue mediated by DAXX [211, 212]. Despite H3.3 being originally associated to euchromatic regions only, some studies have highlighted the relevance of this variant in a broader plethora of chromatin environments. DAXX/ATR are required for the deposition of H3.3 at telomeres and pericentromeric heterochromatin. This chaperone deposits H3.3 also at endogenous retroviral elements (ERVs), thus depicting an overall complex role for the histone H3.3 variant in chromatin function [213–215]. Several studies are now shedding light on other networks of chaperones, protein factors and chromatin remodelling complexes that fine-tune H3.3 dynamics. For instance, the histone chaperone DEK regulates the balance between free and nucleosomal H3.3 and cooperates with ATRX for establishing the proper telomeric architecture [216]. The chromatin remodelling factors CHD1 and EP400, that are also specifically involved in ATP-dependent processes during H3.3 deposition [217, 218].

The fact that H3.3 is enriched both at active genes and regulatory elements by processes linked to active transcription, as well as in heterochromatic regions, leads to a complex role for this variant in both the regulation of gene expression and chromatin conformation. At the organismal level, studies in mice showed that H3.3 plays a critical role in the very early stages of embryonic development, and knockout of both H3.3-coding genes is embryonic lethal [219]. *H3f3a* and *H3f3b* genes are partially redundant and the presence of one of the two genes can sustain mouse development, although phenotypes such as shortened viability and adult infertility can arise [220–223]. In general, *H3f3b* KO gives rise to more severe phenotypes [204]. Partially in line with this, recent analyses focusing on *H3f3a* and *H3f3b* genes evolution in metazoans suggested that the *H3f3b* gene could represent a more ancestral form of H3.3 gene, likely adapted for expression in a wide array of cell types. *H3f3a*, on the other hand, might have appeared later in metazoans evolution—probably as a consequence of duplication events—and it might have evolved with expression more restricted to specific cell types [224].

Histone H3.3 is deposited in a *replication-independent* fashion, and for this reason acquires crucial importance in post-replicative cells. In the nervous system of adult mice and humans, H3.3 becomes the predominant component of the histone H3 pool as a consequence of the progressive accumulation in brain tissues throughout life [225]. Notably, H3.3 maintains a dynamic nature also in adult neurons, with high rates of turnover at the active genes. The H3.3 turnover is directly induced by neuronal activity and, at the same time, is essential for the acquisition and maintenance of proper brain memory [225].

1.6.1 Histone H3.3 modifications

A question of great interest is whether H3.3 can show enrichment in specific PTMs compared to the canonical variants, and if this contributes to define distinct chromatin states. Mass-spectrometry studies conducted in multiple laboratory models, ranging from *Drosophila* and mammals to plants, reported that H3.3 is enriched in PTMs that correlate with active transcription [226–228]. For instance, K4 methylation, K9 acetylation, K36me1/2 and K79me1/2 display relatively higher levels on H3.3, when compared to H3.1 and H3.2 isoforms. Most of these covalent modifications were found to be enriched on H3.3 after its deposition into chromatin. However, with regard to K9 methylation in particular, specific patterns of enrichment were found on non-nucleosomal H3.1 and H3.3 variants. The degree of methylation on H3.1 K9 and H3.3 K9 is indeed different and the modifications present prior deposition might play a role in modulating the activity of

histone-modifying enzymes (e.g., H3.1 is enriched in K9me1, preferential substrate of SUV39H enzymes, while H3.3 is enriched in K9ac and K9me2) [229]. In line with H3.3 residency at heterochromatic regions, H3.3 also contributes to build-up of K9me3 mark at telomeres [230]. Phosphorylation of serine 31 is a H3.3-specific PTM which was originally found to be enriched in chromosomal regions adjacent to centromeres [226]. More recently, H3.3S31ph was also shown to increase p300 activity, therefore crucially contributing to H3.3's supportive role in enhancers activation [231, 232]. Another recent study reported how H3.3S31ph can also recruit the HMT enzyme SETD2 to favour rapid stimulus-induced gene expression in macrophages [233].

1.6.2 Histone H3 mutations in cancers

Specific mutations of H3.3 and canonical H3 variants can naturally occur in multiple types of cancers [234–242]. This implies that only a small amount of mutated protein within the total histone H3 pool is sufficient to affect chromatin function and promote tumorigenesis. Several studies focused on characterizing the oncogenic mechanisms of the lysine-to-methionine mutations occurring at lysines 27 (K27M) and 36 (K36M). The two substitutions share a similar molecular mechanism and they ultimately induce a decrease in the methylation levels of the respective residue on wild-type H3 isoforms. The K27M mutation has high affinity for EZH2, and it was originally proposed to compete with wild-type K27 by sequestering PRC2 and hence precluding proper H3K27me3 deposition [243–245]. More recent studies have however illustrated that PRC2 and H3K27M often do not co-localize [246–248], and that H3K27M inhibits PRC2 at genomic regions where the H3K27me3 mark is present, such as at CpG islands: here the allosterically-activated PRC2 is stalled, and the H3K27me3 levels are therefore maintained locally but globally reduced [249]. Reports have shown that the K27M mutation can affect H3K27ac levels too, resulting in aberrant gene expression and enhancer dysfunction [246, 250]. In particular, the loss of H3K27me3 in the presence of K27M clears the way for H3K27ac accumulation at intergenic regions, leading to de-repression of repetitive elements such as ERVs [250]. The K36M mutation found in chondroblastomas has a similar effect: inhibition of K36-specific methyltransferases (e.g., NSD1/2 and SETD2) with consequent reduction of the global H3K36 methylation levels [251, 252]. This H3K36me2/me3 reduction is also generally coupled with increased H3K27me3 in intergenic regions [251], thus supporting the previous studies reporting that H3K36 methylation prevents H3K27me3 from spreading [136, 253]. Glycine 34 mutations occurring on the *H3f3a* gene were also reported in paediatric gliomas and bone tumors [235, 237]. The G34 residue is not

modifiable and studies have shown that substitution with other bulkier amino-acids results in increased steric hindrance and consequent effect on H3K36 methyltransferases, resulting in imbalanced H3K36me2/3 and H3K27me3 modifications [254–256].

1.6.3 H3.3 mutagenesis as a platform to study histone residues and PTMs in mammals

The fact that only two genes encode H3.3, makes histone mutational studies feasible in mammalian systems. In this section I will briefly summarize some of the main studies that used this approach to expand our understanding of histone residues and H3.3 contribution to chromatin function [201]. In a study conducted in our laboratory, we examined the function of the H3.3 K4 and K36 residues in mESCs, by introducing a lysine-to-alanine substitution using CRISPR-Cas9 editing [257]. To assess the relevance of these residues in stem cells biology, the lines were differentiated to glutamatergic neurons and the H3.3K4A mutant showed severe defects in the formation of both embryoid bodies (EBs) and neurons. The H3.3K36A lines were instead capable of differentiating to neurons, although displaying a higher cell density. Transcriptomic analysis reflected the differentiation phenotypes, with K36A showing limited changes in mESCs but intensified gene expression changes in neurons. This observation is in line with the notion that H3.3 contributes more to the total H3 pool in differentiated cells. Nevertheless, K4A lines displayed wide transcriptional changes already at the stem cell stage, indicating that H3.3 crucially contributes to chromatin regulation despite accounting for a relatively small portion of the total H3 pool in stem cells. Notably, the study revealed that the K4A mutation translates in reduced stability of H3.3 at regulatory elements. From a mechanistic point of view, the K4A mutation affected binding of subunits of the NuRD and SWI/SNF remodelling complexes at active promoters and enhancers, with consequent increased Pol II occupancy and increased nascent RNA signal. This finding was consistent with another study in which the authors compared by mass-spectrometry protein interactors of wild-type or lysine-to-arginine H3.3 mutants at residues K4, K9, K27 and K36 and reported reduced interaction with NuRD components [258]. While K4 modifications were overall well compensated by the canonical H3 in case of the K4A mutation, H3.3K36A mESCs showed decreased H3K36me2 levels in intragenic regions and a concurrent spread of H3K27me3 [257]. Notably, *in-vitro* differentiation to the mesoderm-cardiac lineage disclosed a higher activation of early mesoderm markers (i.e., *Tbxt* aka *Brachyury*) and the K36A mutant was unable to properly differentiate to contracting cardiomyocytes. Given the importance of H3K27me3 in safeguarding the

expression of developmental genes, this observation suggests that a redistribution of the H3K27me3 mark as a consequence of the K36 residues removal might cause lineage-specific defects.

In another recent paper, histone H3.3 mutagenesis was employed to study the contribution of H3.3 and H3K27ac at enhancers. The histone variant H3.3 is indeed enriched at promoters and enhancers of active genes, where is generally subjected to a high histone turnover [259–261]. Consistently, mESCs bearing H3.3K27R mutation displayed loss of H3K27ac specifically at regulatory elements and more prominently at putative enhancers [262]. Only modest transcriptional changes were however observed in this experimental setup and they did not show a strong correlation with the H3K27ac decrease. Given that chromatin accessibility and Pol II occupancy at TSSs were not altered, the authors concluded that H3K27ac is not strictly necessary for enhancer function and transcription [262]. These results were partially mirroring the findings of another study in which the authors introduced a H3.3K27R mutation in *Drosophila* [263]. A wide-spread decrease in H3K27ac was reported but with little consequences on gene expression. An important finding of the study was that H3.3K27R also affected H3K27me3 deposition, indicating a relevant participation of H3.3 to Polycomb-mediated silencing in *Drosophila* [263].

So far, not many H3.3 mutagenesis studies were carried out in *in-vivo* mammalian model systems. One of such studies described the relevance of the H3.3 K27 residue in establishing heterochromatin in the mouse embryo [264]. Another study investigated the effects of H3.3 K9- and K36-to-M mutations in mouse development, by using an inducible system that allowed to uncover residue-specific developmental defects [265].

All in all, H3.3 mutagenesis represents a powerful approach for studying histone residues and PTMs. One crucial aspect to consider is that the chemical properties of the amino acid introduced to substitute the target residue can have critical effects on the outcome. For instance, the lysine-to-methionine mutations have a dominant gain-of-function effect by inhibiting residue-specific HMTs [243, 266]. Substitutions with amino acids mimicking persistent PTMs (e.g., glutamic acid as phosphomimetic or glutamate as acetylmimetic) represent useful gain-of-function mutations when studying specific modifications [231], and even synthetic biology approaches exist to produce unnatural amino acids such as genetically encoded acetyllysines [267]. On the other hand, the introduction of loss-of-function mutations (e.g., lysine-to-arginine or -alanine) can serve two purposes: i) prevent the deposition of PTMs on specific residues of interest, and ii) completely abolish potential function of the histone residue itself. In this regard, it is indeed important to note that also

the unmodified residues can have critical properties, for example for interaction with specific chromatin factors (e.g., H3.3K4 for chromatin remodelling complexes). Loss-of-function residues mutations are helpful to investigate how loss of histone marks can affect biological processes, without the need to delete/inhibit histone modifiers. One possible drawback of the approach is that by mutating a residue that can be for instance both acetylated and methylated, the downstream effects need to be carefully examined to disentangle the contribution of one versus the other modification.

At last, one should consider that histone H3.3 deposition is tightly regulated and therefore when aiming at characterizing histone residues and PTMs contribution by means of H3.3 mutagenesis, the localization pattern of H3.3 will likely affect specific chromatin pathways. In this regard, in the next and last section of the Introduction I will move the focus on describing in more details a specific group of transposable elements previously introduced and known as endogenous retroviruses. This will be propaedeutic for the rest of my study, because potential roles of H3.3 in the repression of such elements were reported, yet they remain to some extent debated.

1.7 Endogenous retroviruses (ERVs): origins and classification

Transposable elements (TEs)¹⁶ are repetitive DNA sequences that constitute up to 40-60% of the mammalian genome and they can be classified in two large groups: DNA transposons and retrotransposons. Among the latter, long-terminal-repeat (LTR) retrotransposons also known as endogenous retroviruses (ERVs), represent ~10% of the mouse genome. ERVs derive from past retroviruses infections occurred in the germline and—as a result of millions of years of neutral evolution—they display various degrees of sequence degeneration. Based on their genomic structure, four main types of ERVs are defined [269]:

- **Complete**, resemble integrated proviruses and contain ~6–9 kb composing the viral genes *gag*, *pro*, *pol* and *env* flanked by the two LTRs (~ 0.3–1 kb).
- **“Slimmed down”**, ERVs in which some of the viral coding sequences are lost.
- **“Substituted”**, the entire viral coding sequences of these ERVs are lost and only the two flanking LTRs are retained.
- **Solitary LTRs**, individual LTRs, likely originated from homologous recombination between 5' and 3' LTRs of an ERV.

¹⁶ McClintock was awarded the Nobel Prize in Physiology or Medicine in 1983 "for her discovery of mobile genetic elements", as a recognition for her ground-breaking studies conducted in the '50s [268].

In general, only few full-length ERVs capable of retrotransposition exist, not many maintain full viral coding sequences, and also when *gag*, *pro*, *pol* and *env* genes are retained, they often contain inactivating missense mutations. ERVs are further classified in three classes based on the similarity of their *pol* gene sequence to exogenous retroviruses (main groups reported in Table 3) [270]. The Class I, II and III ERVs are found in all mammalian species but the relative abundance of the classes differs between species. For instance, Class II ERVs (e.g., including IAP and MusD) are particularly abundant in the mouse genome compared to other mammals [269].

ERV Class	ERV group	Full name
Class I	MLV	Murine leukemia virus
	MuERV-C	/
	MuRVY	/
	MmERV/VL30	<i>Mus musculus</i> endogenous retrovirus
	GLN	Murine retrovirus using tRNA ^{Gln}
Class II	IAP/IAPE	Intracisternal A particle
	MusD/Etn	MusD retrotransposon/early transposon
Class III	MuERV-L/MaLR	Murine endogenous retrovirus-like/Mammalian apparent LTR retrotransposon

Table 3 Main ERV groups in mice (adapted from Johnson 2015 [270])

1.7.1 Epigenetic mechanisms of ERVs repression

ERVs are generally kept in a silent state in most cell types and they usually reside in heterochromatin domains. Two major repressive mechanisms are primarily involved in ERVs repression: DNA methylation and repressive histone modifications. Although both machineries collaborate in silencing, several *in-vitro* and *in-vivo* studies suggest that the contribution of these two mechanisms differs among cell types. In the traditional view, DNA methyltransferases have a critical role in differentiated cells, while histone methyltransferases prevail in stem cells [271, 272]. As DNA methylation levels are normally reduced in pre-implantation embryos [273], alternative pathways have evolved to ensure TEs repression. In mESCs, deposition of H3K9me3 plays a critical role in ERVs repression. Various knockout studies have shown that the KAP1-SETDB1 pathway is crucial for silencing of several classes of ERVs [108, 274]. Similarly to what happens for

genes, the repression is initiated by KRAB-ZnF (KRAB-zinc-finger) proteins¹⁷ that have the ability to recognize specific regions of the LTR sequence (such as the primer binding site) and in turn recruit KAP1-SETDB1 (Figure 1.9) [276–278]. KAP1 and SETDB1 could also be involved in repression of distinct repetitive elements in mESCs; for instance, KAP1 KO displayed derepression of class II and III ERVs (e.g., IAP and MERV-L) [279–281], while SETDB1 KO typically shows derepression of class I and II elements [108, 274]. Studies conducted in differentiated cell types showed that KAP1-SETDB1 are also important for repression of specific ERVs which are usually different from the ones controlled by this pathway in stem cells (as expression also depends on TFs availability [280]).

With regard to the other H3K9 methyltransferases—G9a and SUV39H enzymes—they overall seem to be less involved in ERVs repression in mESCs [282, 283] (with the exception of a specific effect of G9a on silencing of the class III MERV element [284]).

Many other protein factors have been reported as potential players in ERVs repression. Sumo2-dependent sumoylation of KAP1 can for instance favour the KAP1-SETDB1 interaction and also proper KAP1 recruitment at LTRs [285]. The histone demethylase LSD1 controls MERV expression, likely by keeping low H3K4me and H3K27ac levels [286]. The chromatin remodeler SMARCA1 was also recently identified as a central player in class I and II ERVs repression, by establishing a direct functional interaction with KAP1 [287].

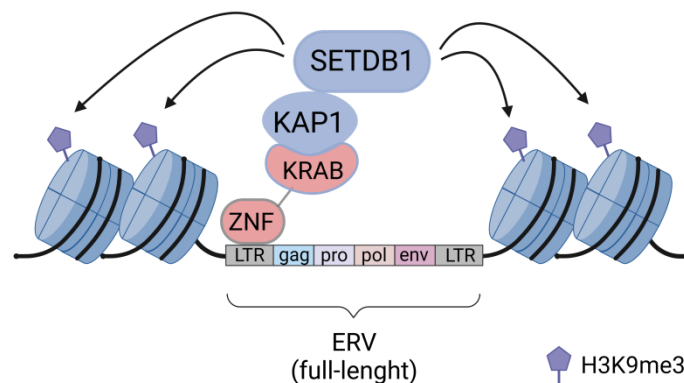


Figure 1.9 Schematic of KAP1-SETDB1 mediated repression of ERVs. KRAB-ZnF proteins interact with LTRs in a sequence-specific manner and recruit KAP1 and SETDB1, which ultimately deposits H3K9me3. Figure created with BioRender.com.

¹⁷ KRAB-ZnF are one of the largest family of TFs in higher vertebrates and evidences suggest that the rapid evolution and expansion of this protein family might be driven by the acquisition of novel ERVs [275].

1.7.2 H3.3 and ERVs repression

H3.3 is deposited specifically at ERVs, predominantly at class I and II elements [215], and the deposition of H3.3 at these loci is mediated by the DAXX/ATRX chaperone complex [213, 288]. H3.3 KO results in reduced levels of H3K9me3 and in increased transcription of IAP elements, with a concomitant decrease in KAP1 and DAXX recruitment at these ERVs [215]. Despite these initial findings being reported, the role of H3.3 deposition in ERV silencing is still unclear [280, 281, 289]. For instance, using a reporter assay coupled with shRNA screening, ATRX was identified as a critical player for heterochromatin formation at IAP elements [288]. DAXX is also necessary for proper establishment of heterochromatin, however DAXX mutant proteins unable to bind H3.3 could still rescue repression of the reporter [288]. Moreover, the silencing of the reporter was maintained in H3.3 KO cells [288].

Evidences collected in a recent study led to a model for IAP silencing in mESCs that foresees a continuous SMARCAD1-mediated eviction of nucleosomes, which is counteracted by the DAXX-mediated incorporation of H3.3-containing nucleosomes at ERVs [290]. This high H3.3 turnover would then contribute to maintain a closed chromatin state, aided by SETDB1 which is called upon to continuously re-establish high H3K9me3 levels [290]. The MORC3 ATPase protein was also recently identified as a regulator of the DAXX-mediated deposition of H3.3 at ERVs [291]. Nevertheless, the number of ERVs and the degree of derepression in H3.3 KO mESCs or in mESCs depleted of DAXX, ATRX or MORC3 is generally considered to be marginal [280, 289, 291], leading to the hypothesis that H3.3 deposition is not crucial for ERVs silencing and that DAXX and ATRX may even have additional functions besides H3.3 incorporation [280, 291].

1.7.3 Roles of ERVs in transcriptional regulation

As summarized above, cells employ many mechanisms to ensure repression of ERVs. Nonetheless, according to several lines of evidence, ERVs can be co-opted by the host cells to exert regulatory functions in gene expression. Multiple TF binding sites (TFBS) are located within the sequence of LTRs, enabling these elements to function as potential landing sites for cell type-specific TFs. In stem cells, over 15% of pluripotency factors binding sites are placed within LTRs sequences [292], and the presence of species-specific TEs was suggested as a critical driver of regulatory networks divergence between human

and mouse pluripotent cells¹⁸ [294]. ERVs can operate as either alternative promoters or enhancer elements to regulate the expression of host genes [295, 296]. When ERVs are co-opted as cis-regulatory elements, they tend to acquire the features characteristic of active chromatin states expected for canonical enhancers/promoters (e.g., H3K27ac; H3K4me1/3; nascent transcription etc.). This co-option can occur as a consequence of defective silencing mechanisms, or in a more physiological manner (context-dependent and cell type-specific). One example of unwanted co-option was described for KAP1 KO mESCs, in which derepressed ERVs can function as *de-novo* enhancers of bivalent genes, ultimately resulting in their upregulation [297]. A fascinating example of physiological co-option is the one that is reported for ERVs in the regulation of immune-related gene regulatory networks [298–300], whereby ERVs were co-opted to exert mainly an enhancer function for several immune-related genes. Recent studies have shown that a relevant fraction—ranging between 15 and 30%—of active enhancers in immune cell types are represented by ERVs [298, 299]. Several ERVs-derived enhancers and promoters can be readily activated by infectious stimuli, thus providing an inducible and responsive immune regulatory platform [298, 301, 302]. On the other hand, a phenomenon known as “*viral mimicry*” has also been described, in which de-repression of ERVs leads to production of double-stranded RNA (dsRNA) which is sensed by pattern recognition receptors such as RIG-1 [303], that ultimately triggers an interferon response [304, 305]. This mechanism might have evolved as a way to aid and boost the antiviral response in case of infection [306].

¹⁸ In 1950, McClintock originally noticed that transposons could alter expression of proximal genes in plants [268]. This concept was then expanded in the model formulated in 1971 by Britten and Davidson, who postulated that TEs could promote the evolution of gene regulatory networks [293].

1.8 Aim of the project

Histone H3.3 is enriched at distal gene regulatory elements and active genes in mESCs. Thus, mutagenesis of the specific residues within histone variant H3.3 offers an opportunity to study the role of histone post-translational modifications in a mammalian system at the genomic regions enriched with H3.3. Histone H3.3 has been associated with dynamic and active chromatin environments, and in line with this, mass-spectrometry studies showed that histone marks found in H3.3 are related to active transcription. Nevertheless, it is now known that in mESCs, H3.3 is also located at the promoters of repressed genes, such as bivalent promoters, and at specific classes of transposable elements known as ERVs. Whether and how H3.3 and its associated histone marks contribute to maintaining repressive chromatin states at these genomic regions is largely unclear.

The main aim of the project is to investigate to what extent histone H3.3 and its associated histone marks contribute to repressive mechanisms. Over my PhD, I took advantage of the H3.3 mutagenesis strategy to generate several mutants¹⁹. To study the repressive roles of H3.3, I focused on the lysine 9 and lysine 27 residues. By employing CRISPR-Cas9 genome editing tools, I introduced lysine-to-alanine substitutions at K9 and K27 residues in mESCs. The mutation to a non-modifiable amino acid prevents the deposition of modifications at these sites. In the first part of the Results, I describe the optimization process of the gene editing strategy. I then present the experiments conducted to characterize the role of H3.3 K9 and K27 in mESCs (i.e., control versus mutant cells). First, I profiled gene expression changes by means of mRNA-seq and conducted phenotypic assays to verify the effect of the mutations on stem cell biology. *In-vitro* differentiation to ectoderm and mesoderm lineages was also carried out in order to uncover possible defects in the differentiation potential of the mutant cells. Second, I assessed the disrupted chromatin landscape in mESCs by means of ChIP-seq and PRO-seq to investigate how genome-wide changes in histone modifications and nascent transcription relate to the changes in gene expression.

¹⁹ Among the mutants I generated there are H3.3 K79A and triple K14A-K18A-K27A mutants. In the Results section I will briefly summarize the editing strategy adopted to generate the H3.3K79A mESCs, because the data obtained with these cell lines were used to generate the gene regulatory network.

2. Results

2.1 Generation of H3.3 mutant mESC lines using CRISPR-Cas9

Histone H3.3 is encoded in mammals by two independent genes residing outside of the canonical histone clusters, which makes this protein an attractive candidate for histone mutagenesis studies in mammalian systems. To ensure that the entire H3.3 protein pool would carry the mutations of interest, I used for this project *H3f3a* knockout mESC lines (i.e., *H3f3a*^{KO}/*H3f3b*^{WT} —previously generated by former PhD student Maja Gehre) and introduced lysine-to-alanine (K-to-A) substitutions in the *H3f3b* coding gene.

I employed two different CRISPR-Cas9 gene editing approaches to introduce the mutations of interest, namely:

- **Plasmid-based approach**, for H3.3 K79A lines (see Appendix for details)
- **Ribonucleoparticle (RNP)-based approach**, for H3.3 K9A and K27A lines (and triple mutant K14-K18-K27A lines, not shown)

In the next sections I will present both technologies and describe the outcomes of the gene editing rounds.

2.1.1 H3.3 K79A mESCs lines generation and validation

To generate mESCs bearing lysine-to-alanine substitution at lysine 79, I used the plasmid-based CRISPR-Cas9 method previously established by Maja Gehre [257, 307]. This strategy is based on the protocol published by Ran *et al.* [308] and consists in the co-transfection of a plasmid—encoding both a GFP-tagged Cas9 and the single-guide RNA (sgRNA)—together with a single-stranded oligonucleotide (ssODN) repair template for the homology-directed repair (HDR). Briefly, the expression of Cas9-GFP allows fluorescence activated cell sorting (FACS) of efficiently transfected cells, which I used to isolate single cells that were then expanded to isolate clonal lines. After clonal expansion, I employed a mismatch-qPCR pre-screening to identify clones potentially carrying the K79A substitution. The editing was then confirmed by means of PCR amplification followed by Sanger sequencing of the targeted locus. A total of 5 independent H3.3 K79A mutant clonal lines were isolated.

Due to both CRISPR-Cas9 gene editing and clonal selection, chromosomal abnormalities can arise, and it is crucial to validate the clones before conducting further experiments. I

performed low-coverage whole genome sequencing to spot possible abnormalities on a genomic scale. Two of the five H3.3K79A clones generated displayed chromosomal deletions or duplications on chromosome 11. The affected sites are close to the *H3f3b* locus, indicating that these chromosomal abnormalities are likely due to the aberrant repair of double-strand breaks (DSBs). These clones were discarded and only the three K79A clones with intact chromosomes were used for experiments. See Appendix for details.

2.1.2 H3.3 K9A and K27A mESC lines generation and validation

The plasmid-based gene editing method resulted in overall a low efficiency for editing the *H3f3b* locus (i.e., < 0.5% of transfected GFP+ clones). This can be due to: i) delayed expression of Cas9-GFP protein from plasmid and ii) reduced stability of ssODN repair template. Given that the two components are co-transfected, by the time the Cas9-GFP protein reaches the peak of expression it is likely that the single-stranded repair template will be partially degraded. Therefore, with the aim to improve HDR efficiency, I participated in optimizing a novel CRISPR-Cas9 system based on Cas9 ribonucleoparticles (RNPs). Several papers reported enhanced HDR efficiency using Cas9 proteins covalently linked to the repair template [309–311]. With the same rationale, Andrea Callegari and Moritz Kueblbeck (at the time technicians in the Ellenberg Group in the Cell Biology and Biophysics Unit at EMBL Heidelberg) teamed up with the Protein Expression and Purification Facility at EMBL Heidelberg and developed a new CRISPR-Cas9 system. Here, the Cas9 protein is tagged with a monomeric streptavidin (mSA) that can be theoretically coupled *in-vitro* with a 5'-biotinylated repair template. To select transfected cells upon electroporation, I took advantage of a commercially available tracrRNA labelled with the ATTO-550 fluorophore. Specifically, I optimized a system in which the sgRNA is generated by annealing *in-vitro* two RNA oligonucleotides: i) the ATTO-550-tracrRNA and ii) the target-specific crRNA and the RNP is formed via binding of the Cas9-mSA to the sgRNA. I then included a biotinylated-ssODN to conjugate to the Cas9-mSA and increase stability inside the cells (Figure 2.1a). Based on the ATTO-550 fluorescent signal, single cells are sorted by FACS 24 hours after electroporation (Figure 2.1b). The main advantage of this approach is that the sorted single cells are more likely to contain both the Cas9 protein as well as the repair template. Overall, a substantial enhancement of the HDR rate was confirmed by Sanger sequencing and the correct homozygous mutation was detected in 10% or more of the sorted clones (Figure 2.1c). With the nearly 100-fold increase in gene editing efficiency compared to the plasmid-based method, the Cas9 RNP-based approach

allows screening a significantly lower number of clones, resulting in a much faster and streamlined generation of mutant clones.

Using this approach, I could generate 22 K27A and 9 K9A mESC clonal lines in total²⁰. Of those, I performed low-coverage whole genome sequencing on 10 K27A and 4 K9A clones and I could confirm the integrity of 7 K27A clones and 4 K9A clones (Figure 2.2). For each genotype, I selected three independent clones to be used for the project.

²⁰ The K9A mESCs were generated with the help of the intern student Mohamad Wahidi.

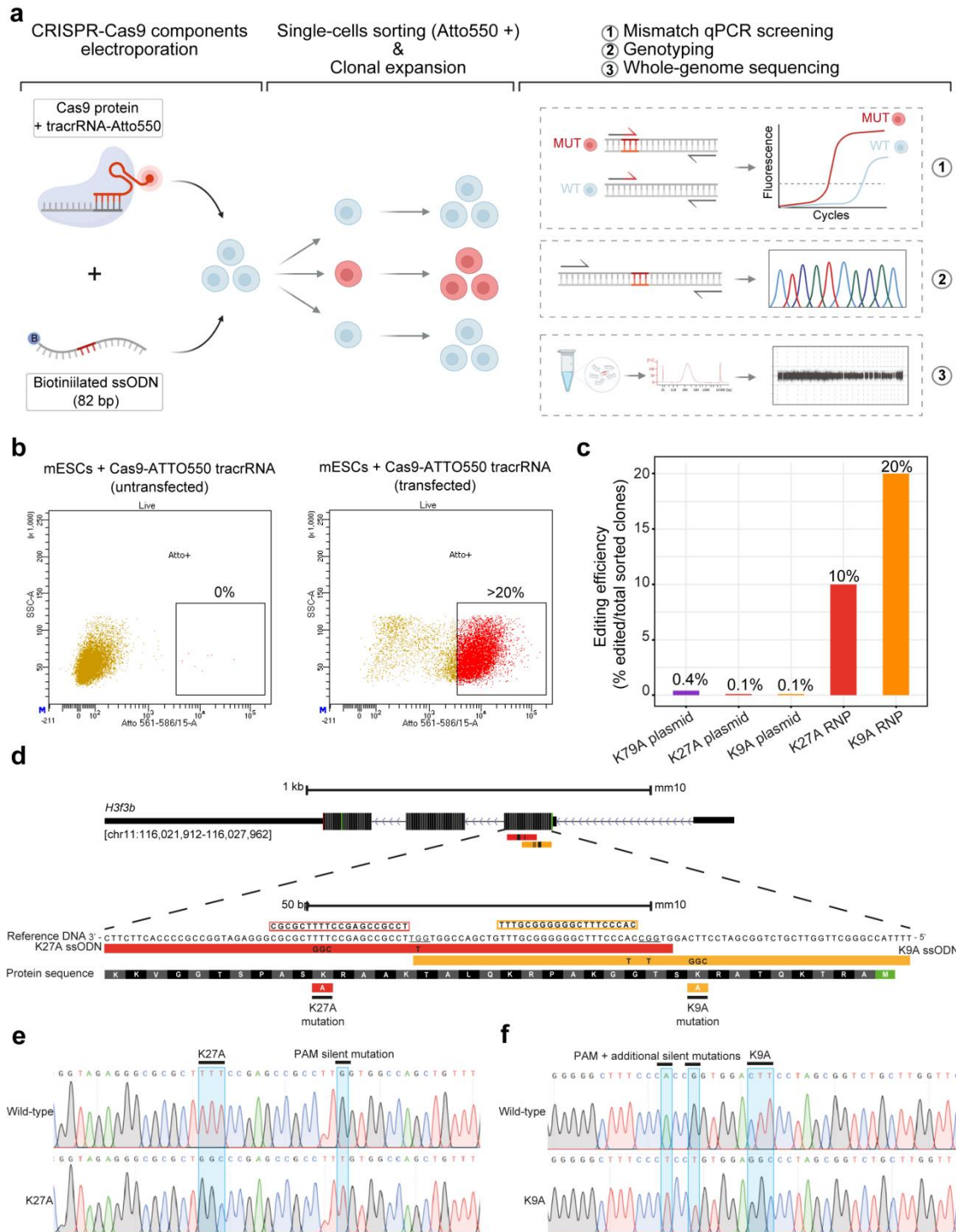


Figure 2.1 CRISPR-Cas9 editing strategy. **(a)** Schematic describing the Cas9 RNP editing workflow; mESCs are co-transfected with Cas9 RNP + 80mer repair template and single-cell sorted. After clonal expansion, screening and validation is carried out in three steps. **(b)** Transfection efficiency with Cas9-tracrRNA-ATTO550 of about 20% or higher is reached. Control cells subjected to Cas9 RNP complex but not transfected are used to set the gates for flow cytometry analysis. **(c)** Editing efficiency with Cas9 plasmid versus Cas9 RNP approaches for the editing of the *H3f3b* locus (cumulative data from 5 K79A-plasmid rounds, 5 K27A-plasmid rounds, 3 K9A-plasmid rounds, 1 K27A-RNP round and 1 K9A-RNP round). **(d)** Schematic with guide RNAs and ssODN designed to introduce K27A (red) and K9A (orange) mutations in the *H3f3b* locus. **(e-f)** Representative Sanger sequencing chromatograms showing the comparison between wild-type and K27A (e) or K9A (f) mESCs at the *H3f3b* endogenous locus.

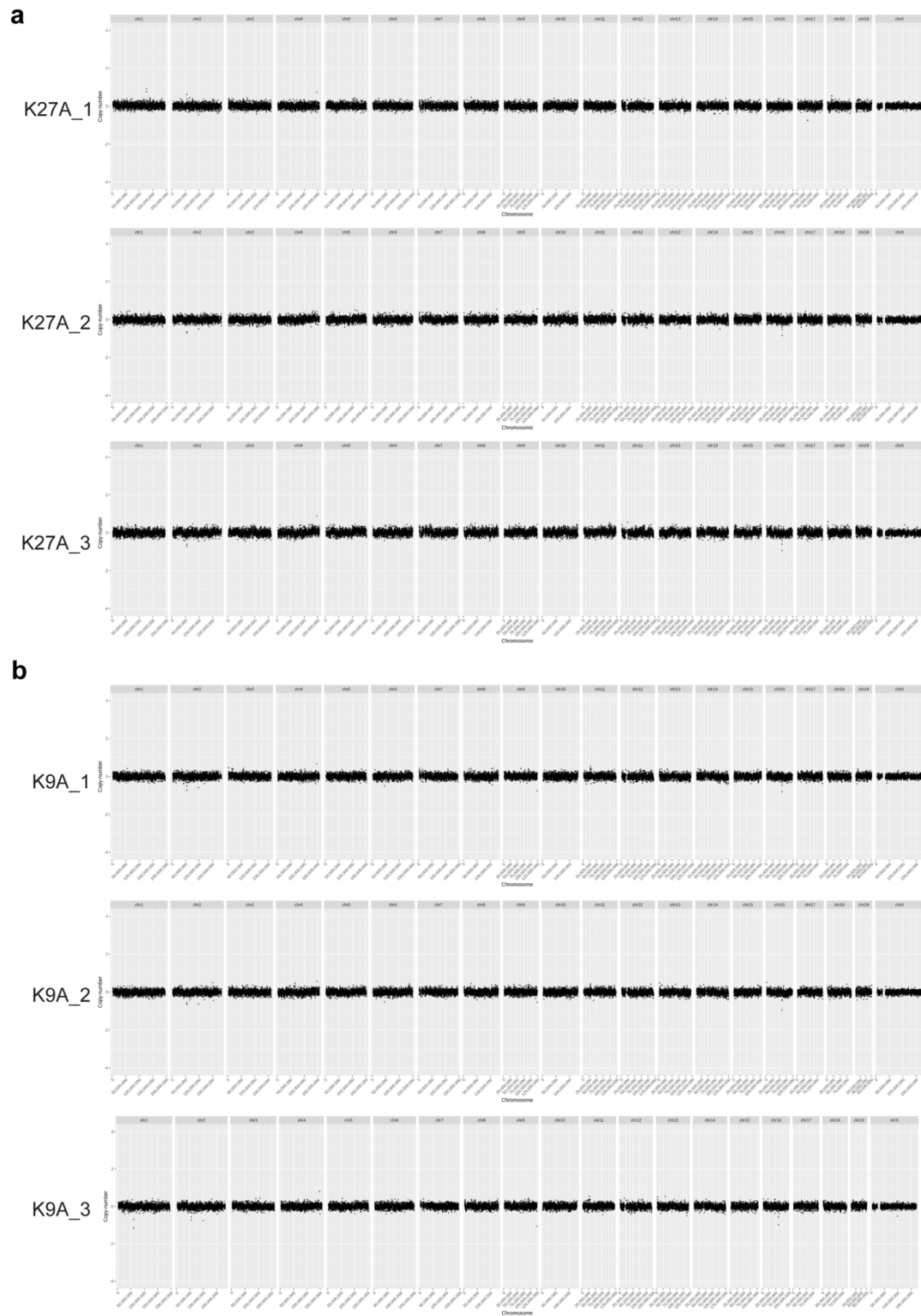


Figure 2.2 Chromosomal integrity assessment for the edited clones. Whole-genome plots show the copy-number difference of 100 kb genomic bins between edited clones and the parental control line. Plots of the three independent replicates generated for K27A **(a)** and K9A **(b)** mESCs are shown.

2.2 Characterization of H3.3 K9A and K27A mESCs

In the course of my PhD, Kyung-Min Noh and I decided to focus our efforts on characterizing the effects of the H3.3 lysine 9 and 27 substitutions. Therefore, most of the data I generated and analysed were obtained from these two mutants. The goal of the project is to employ a multiomic approach to profile changes in the transcriptome and in the epigenome, with the aim to probe the contribution of H3.3 K9 and K27 residues to regulation of gene expression in mESCs.

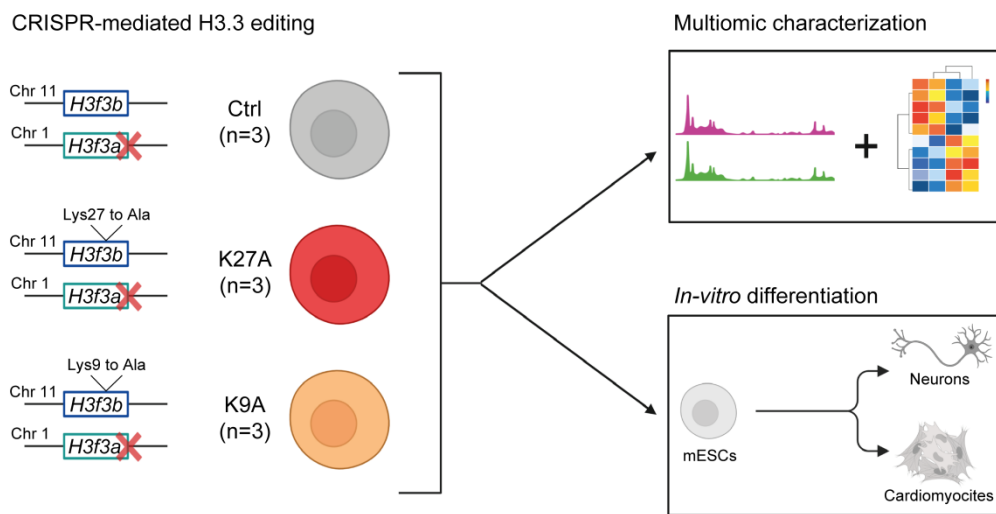


Figure 2.3 Graphical project outline.

2.2.1 H3.3 K9A and K27A mESCs display wide transcriptional changes

To evaluate the impact on transcription in the absence of H3.3 K9 or K27 residue, I performed mRNA-seq with three independent biological replicates each for control ($H3f3a^{K0}/H3f3b^{WT}$), K9A ($H3f3a^{K0}/H3f3b^{K9A}$) and K27A ($H3f3a^{K0}/H3f3b^{K27A}$) genotypes (Figure 2.3). Hierarchical clustering showed that K9A and K27A lines display distinct transcriptional profiles compared to controls (Figure 2.4a). Analysis of differentially expressed genes (DEGs) showed substantial gene expression changes for K9A (2585 DEGs, $\text{padj} < 0.05$ & FC cut-off=2) and K27A mESCs (1463 DEGs, $\text{padj} < 0.05$ & FC cut-off=2). Notably, the number of DEGs and the magnitude of gene expression changes were greater in K9A than K27A (Figure 2.4b).

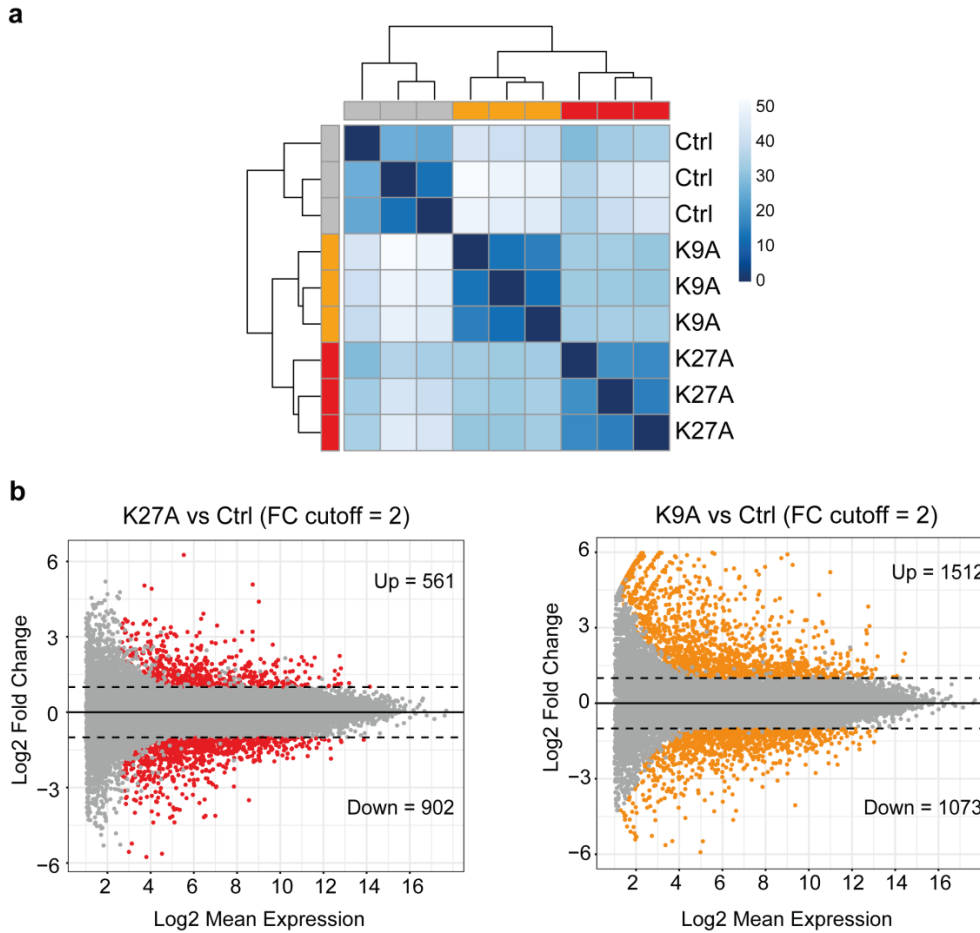


Figure 2.4 Transcriptional changes in H3.3 K27A and K9A mESCs. (a) Hierarchical clustering with sample-to-sample distances computed from rlog-transformed mRNA-seq counts. **(b)** MA-plots of DEGs in K27A (left) and K9A (right) mESCs. Coloured genes represent significantly DEGs defined with $\text{padj} < 0.05$ and $\log_2(\text{fold-change})$ cut-off of 1.

Both residues targeted in the project can be subjected to acetylation and methylation, thus being potentially involved in activating (e.g., H3K9ac and H3K27ac) and silencing (e.g., H3K9me3 and H3K27me3) mechanisms. Therefore, to better characterise the transcriptional changes in mESCs lacking H3.3 K9 or K27, I decided to classify the DEGs based on their de-regulation in only one or both H3.3 mutants (Figure 2.5). The largest group of genes ($n=1084$) was upregulated explicitly in K9A and showed enrichment for gene ontology terms related to immune response, extracellular matrix organization and differentiation. A smaller fraction of genes ($n=183$) upregulated in K27A was specifically enriched for development and differentiation processes (e.g., regionalization, pattern specification processes, embryonic organ development). Downregulated genes showed a similar trend, with a considerable group of genes ($n=525$) related to protein secretion processes with reduced expression in K9A and a smaller group ($n=304$) only affected in K27A. Notably, many genes showed concomitant upregulation ($n=377$) or downregulation

(n=547) in mutant cells. Gene ontology analysis of the commonly upregulated genes showed enrichment for differentiation and development-related terms, while downregulated genes in both mutants exhibited enrichment for terms related to meiosis. Only a small group of genes (n=52) was differentially expressed in opposite directions in K9A and K27A. Despite representing only about 20-30% of the total histone H3 pool in mESCs, single amino-acid substitutions of histone H3.3 can lead to extensive changes in the transcriptional programmes of mESCs. This result suggests that specific residues of this histone variant are critically required for maintaining steady-state transcriptional regulation in mESCs.

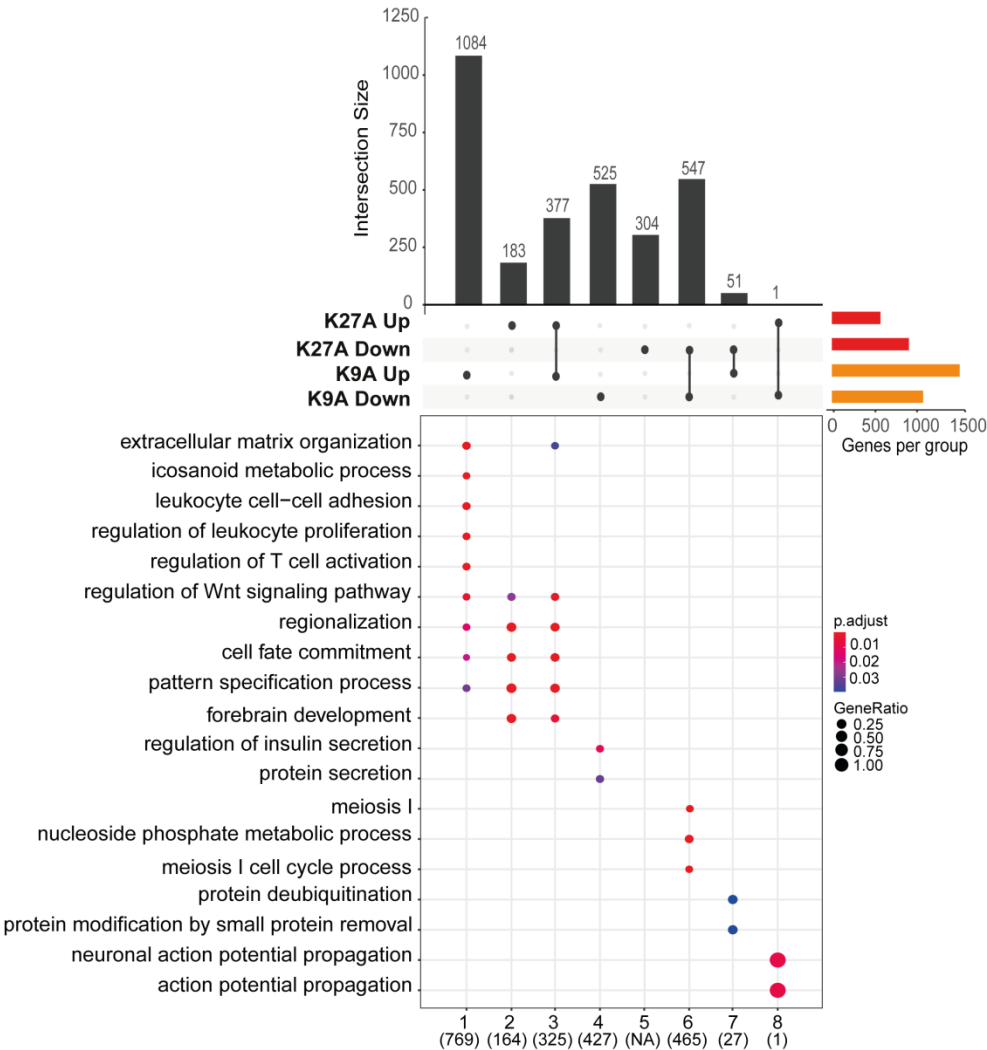


Figure 2.5 Gene ontology enrichment analysis shows terms shared for DEGs in K27A and K9A mESCs. On the top, upset plot with DEGs up-/down-regulated exclusively in one mutant or shared between the two mutants. The last two groups on the right (n=51 and n=1) are genes de-regulated in opposite directions. On the bottom, the most significant GO terms for each group of genes are reported.

2.2.2 Global histone modifications changes in H3.3 K9A and K27A mESCs

Despite H3.3 representing a relatively small fraction of the total H3 pool in mESCs, individual histone residue substitutions may affect the overall abundance of histone modifications. To investigate the global levels of histone modifications, the intern student Ana Janeva performed western-blots for selected histone H3 modifications. However, by means of western-blot we did not detect significant changes in PTMs in mESCs (Figure 2.6a). With the aim to carry out a more thorough analysis, Yekaterina Kori (PhD student from the Garcia Lab, University of Pennsylvania) performed middle-down mass spectrometry to detect global changes in histone H3, H3.3 and H4 post-translational modifications. With regard to histone H3 modifications (Figure 2.6b), both K27A and K9A mESCs showed increased levels of H3K79me_{1/2}. Levels of H3K27me_{2/3} were mildly reduced in K27A mESCs. Especially in K9A mESCs, H3K36me_{1/2} was reduced and the same trend was also observed for H3K27me_{1/2/3}. On the other hand, levels of histone H3 acetylation (e.g., H3K9ac; H3K14ac; H3K18ac; H3K23ac) were in general increased. Interestingly, global levels of K27 and K36 methylation were reduced on H3.3 peptides (Figure 2.6c). Histone H4 PTMs also showed global changes in two out of three K9A replicates, whereby several acetylation marks were overall increased.

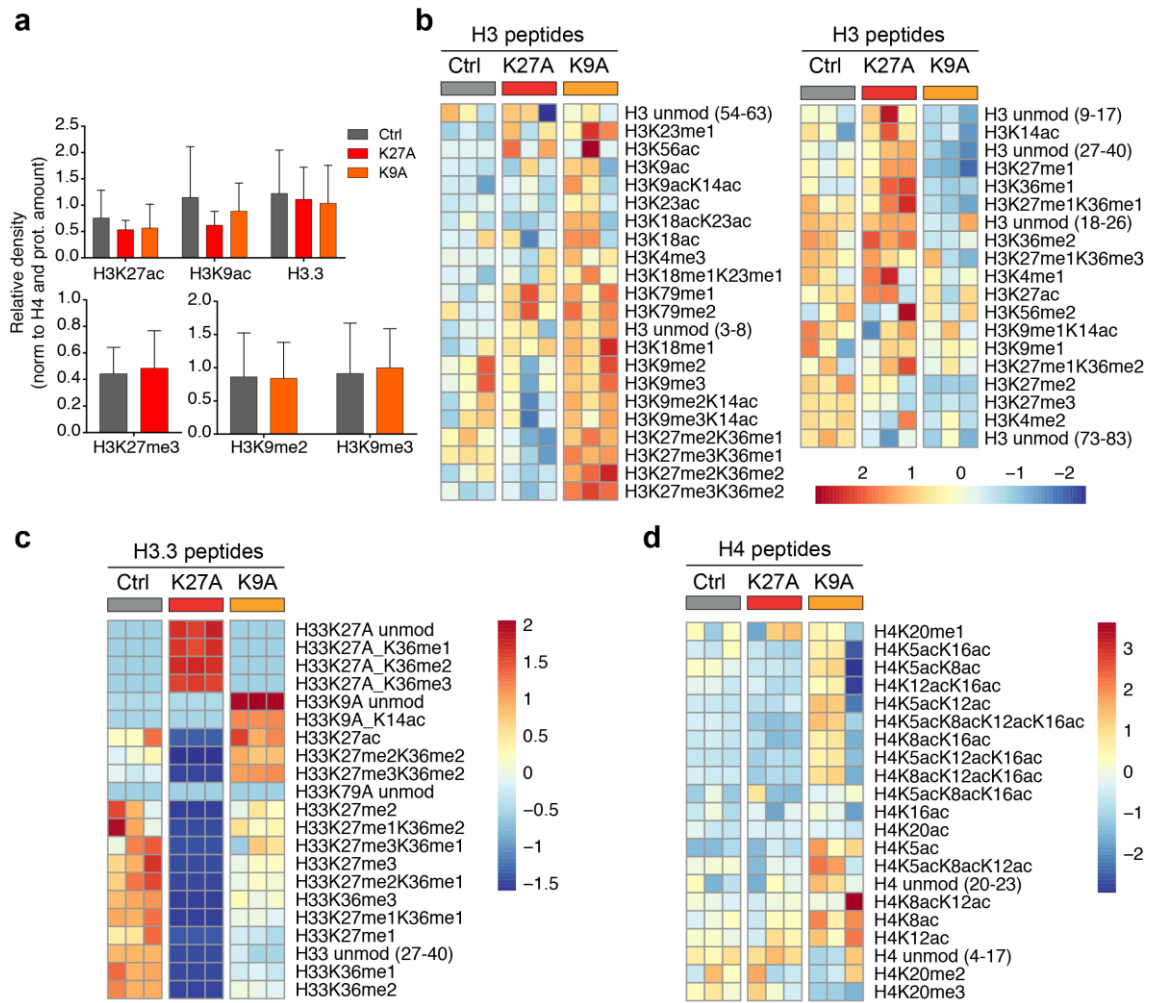


Figure 2.6 Global differences in histone PTMs measured by immunoblot and mass-spectrometry. (a) Quantification of western-blot for selected histone modifications in K27A and K9A mESCs. Samples were loaded three times in increasing amounts (0.5, 1 and 2 μ g) and quantified considering the “peak area”. Density was then normalized over histone H4 density and over the loaded protein amount. Data reported are mean $\pm$ standard deviation (n=2 replicates for control; n=3 replicates for mutants). Differences were not significant, according to unpaired t-test. **(b-d)** Heatmaps displaying the relative abundance (row z-score) of individual or combinatorial post-translational modifications of histone H3 (b), H3.3 (c) or H4 (d). Lysine-to-alanine mutated H3.3 peptides were used as controls in panel (c). The peptides length was omitted for clarity from the figure and is reported here: H3K4 (3-8); H3K9/K14 (9-17); H3K18/K23 (18-26); H3K27/K36 (27-40); H3K56 (54-63); H3K79 (73-83); H4K20 (20-23); H4K5/K8/K12/K16 (4-17); H4K20 (20-23).

2.2.3 H3.3 K9A (but not K27A) mESCs display growth delay and increased cell death

To test whether the absence of either one or the other residue could affect the growth of mESCs, I monitored the cell proliferation over six days. H3.3K9A cells showed decreased cell numbers compared to the control, while the number of H3.3K27A cells did not deviate significantly (Figure 2.7a). The decreased cell numbers in K9A mESCs could be due to i) reduced proliferation or ii) increased cell death. The expression of the pluripotency marker genes *Nanog* and *Sox2* was modestly decreased in K9A and K27A mESCs, while

Pou5f1 (aka *Oct4*) was not significantly differentially expressed in K9A (and mildly upregulated in K27A). Additional pluripotency markers such as *Klf4*, *Lifr* and *Esrrb* did not show a significant difference (Figure 2.7b). The observation that these pluripotency markers were equally affected in both mutants, indicates that these changes cannot account for the growth delay observed for K9A mESCs.

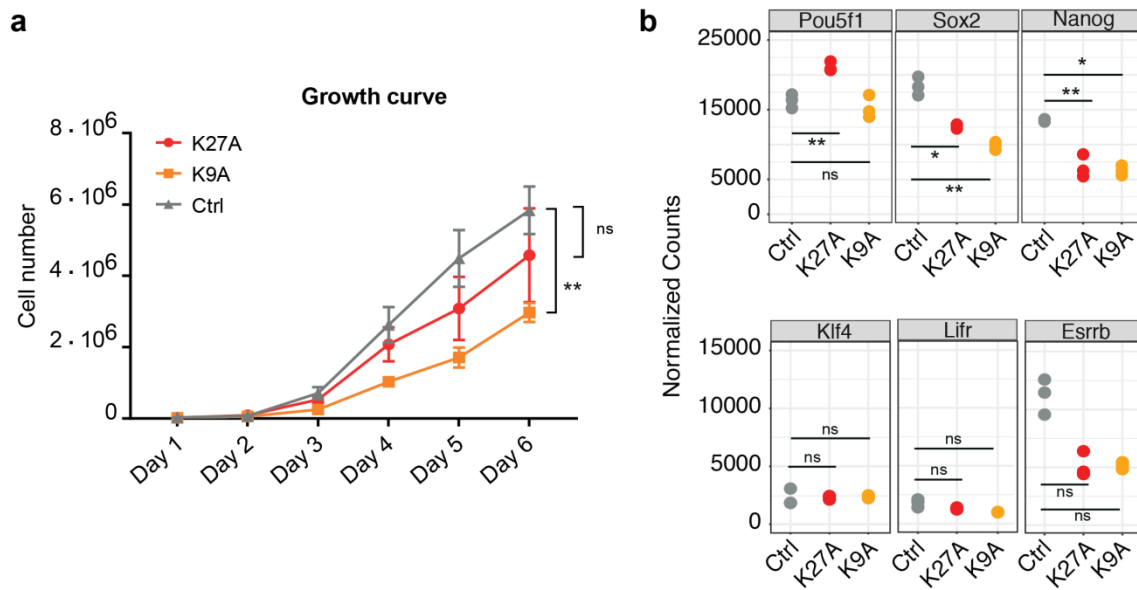


Figure 2.7 Mutant mESCs growth curve and pluripotency markers expression. **(a)** Growth curve for control and mutant mESCs. Error bars indicate the standard deviation of $n=3$ biological replicates. **(b)** Normalized mRNA-seq counts for six selected pluripotency markers (*Pou5f1/Oct4*; *Sox2*; *Nanog*; *Klf4*; *Lifr* and *Esrrb*) in control, K27A and K9A mESCs. Significant differences were calculated with DESeq2 and adjusted with Benjamini Hochberg's correction. (*: $\text{padj} < 0.05$; **: $\text{padj} < 0.01$; ns: $\text{padj} > 0.05$).

To further assess proliferation in mESCs, Umut Yildiz and Nadine Fernandez-Novel Marx performed a cell cycle analysis, by means of EdU labelling and flow-cytometry analysis. The percentage of cells in the G1, S or G2/M phase was not significantly different in K9A nor K27A mESCs, when compared to control lines (Figure 2.8).

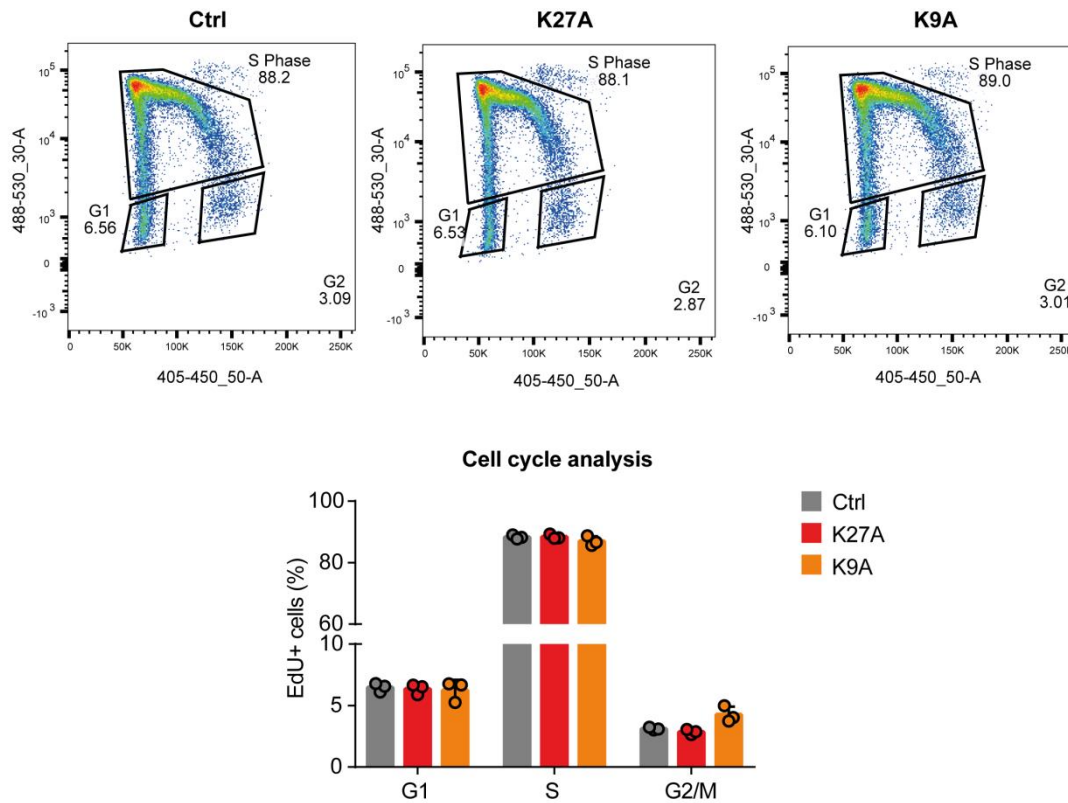


Figure 2.8 Cell cycle assay in mESCs. Representative flow cytometry plots (top) displaying gating strategy for measuring 5-ethynyl-2'-deoxyuridine (EdU) incorporation in control, K27A and K9A mESCs. Barplot (bottom) with percentages of cells in G1, S or G2/M phases are reported. Unpaired t-test was used to calculate significance and no significant differences were detected.

We asked whether an increase in cell death could explain the reduced cell numbers observed for K9A mESCs. Annexin V staining and flow cytometry analysis (performed by Umut Yildiz and Nadine Fernandez-Novel Marx) revealed a significant increase in cell death for K9A mESCs (Figure 2.9). In line with the growth curve result, K27A mESCs did not show significant changes in cell death rates.

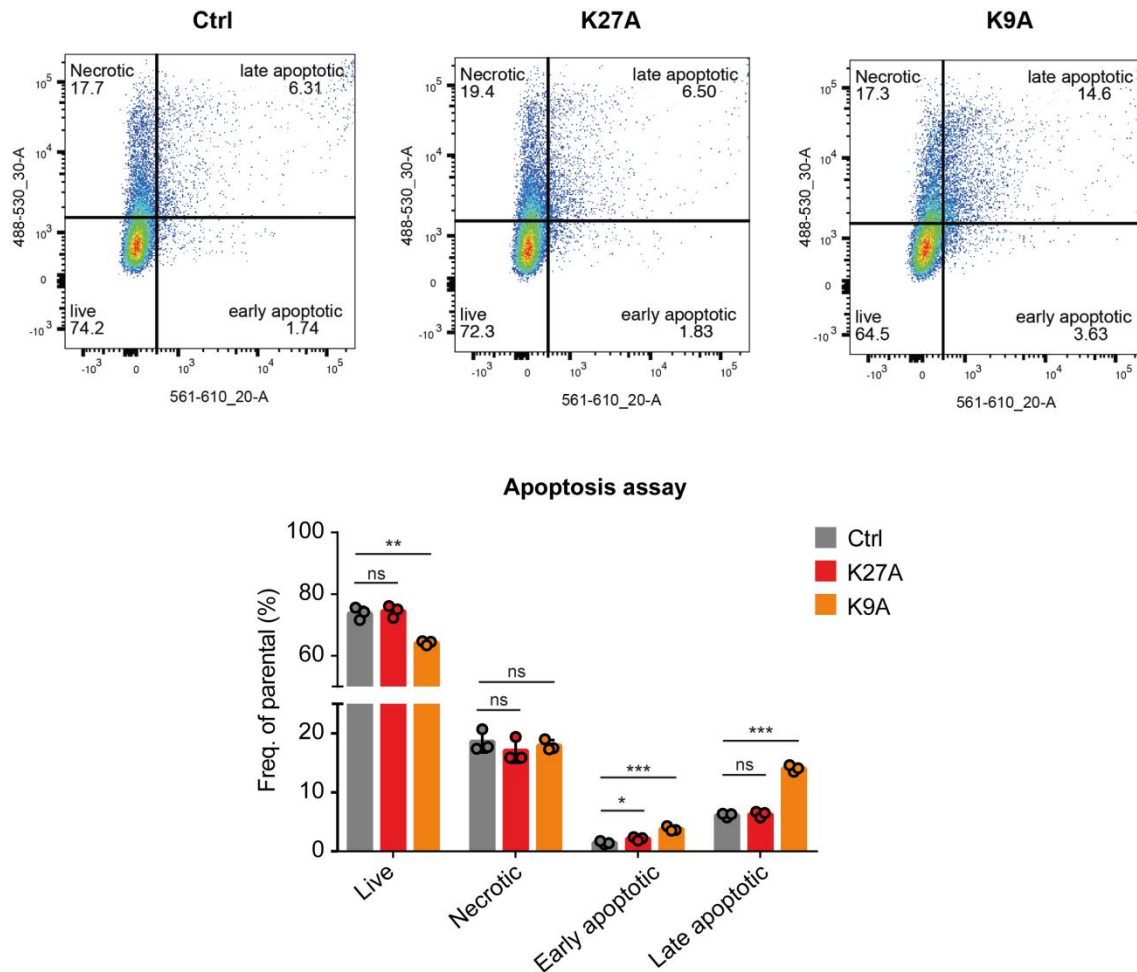


Figure 2.9 Apoptosis assay in mESCs. Cell death assay was conducted by means of Annexin V staining of mESCs; annexin V has affinity for the phospholipid phosphatidylserine translocated on the outer side of the cell membrane in dying cells. Three representative flow cytometry plots (top) for control, K27A and K9A mESCs are reported. Barplot (bottom) of percentages of cells identified as live, necrotic, early apoptotic or late apoptotic are reported. Unpaired t-test was used to calculate significance (*:p-value<0.05; **: p-value<0.01; ***: p-value<0.001; ns: p-value>0.05).

2.2.4 Derepression of developmentally-regulated genes in K9A and K27A mESCs impacts the differentiation potential

Given that mESCs lacking either H3.3K9 or H3.3K27 upregulated expression of genes related to developmental processes, I investigated whether this premature activation of development-related genes in mESCs could affect *in-vitro* differentiation. To assess the ability of the H3.3 mutant mESCs to commit to the neuro-ectodermal lineage, I performed *in-vitro* neuronal differentiation. Upon induction of the differentiation triggered by the removal of leukaemia inhibitory factor (LIF) and subsequent addition of retinoic acid, H3.3K9-deficient mESCs formed significantly smaller embryoid bodies (EBs) and fewer neuronal progenitor cells (NPCs) (Figure 2.10). H3.3K9A clones also failed to properly

differentiate into neurons and formed a heterogeneous population of cells characterized by a non-neuronal like morphology (Figure 2.10a). In contrast, the H3.3K27A cells were able to form a homogeneous population of neurons with regular morphology and dense axonal networks.

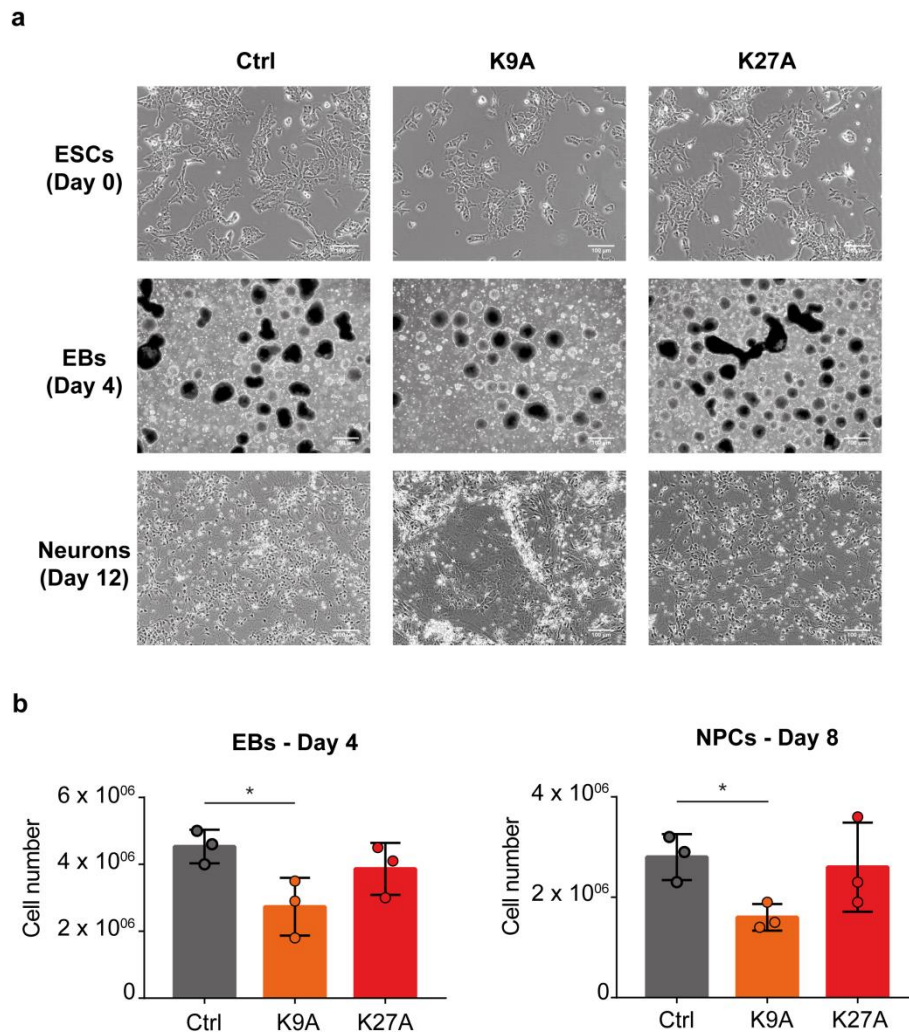


Figure 2.10 In-vitro neuronal differentiation. (a) Bright-field images of control, K27A and K9A feeder-depleted mESCs (top), embryoid bodies at day 4 of the in-vitro differentiation (middle) and terminally differentiated glutamatergic neurons at day 12 of the differentiation (bottom). Scale bar 100 μ m. **(b)** Barplot of cell numbers counted after dissociation of the embryoid bodies at day 4 (left) or day 8 (right) of the in-vitro neuronal differentiation. Error bars represent standard deviation of n=3 biological replicates and significance was calculated with unpaired t-test (*: p-value<0.05).

Consistently with the results from cell counts and cell morphologies, the expression of NPCs markers (i.e., *Hash1*, *Neurod1*, *Pax6* and *Sox2*) at day 8 of the differentiation was severely affected only in the H3.3K9A mutant (Figure 2.11a). Similarly, immunofluorescence staining for neuronal markers (i.e., *Map2* and *beta-III tubulin*) at day 12 of the differentiation showed a decrease in signal in H3.3K9A cells but not in K27A (Figure 2.11b).

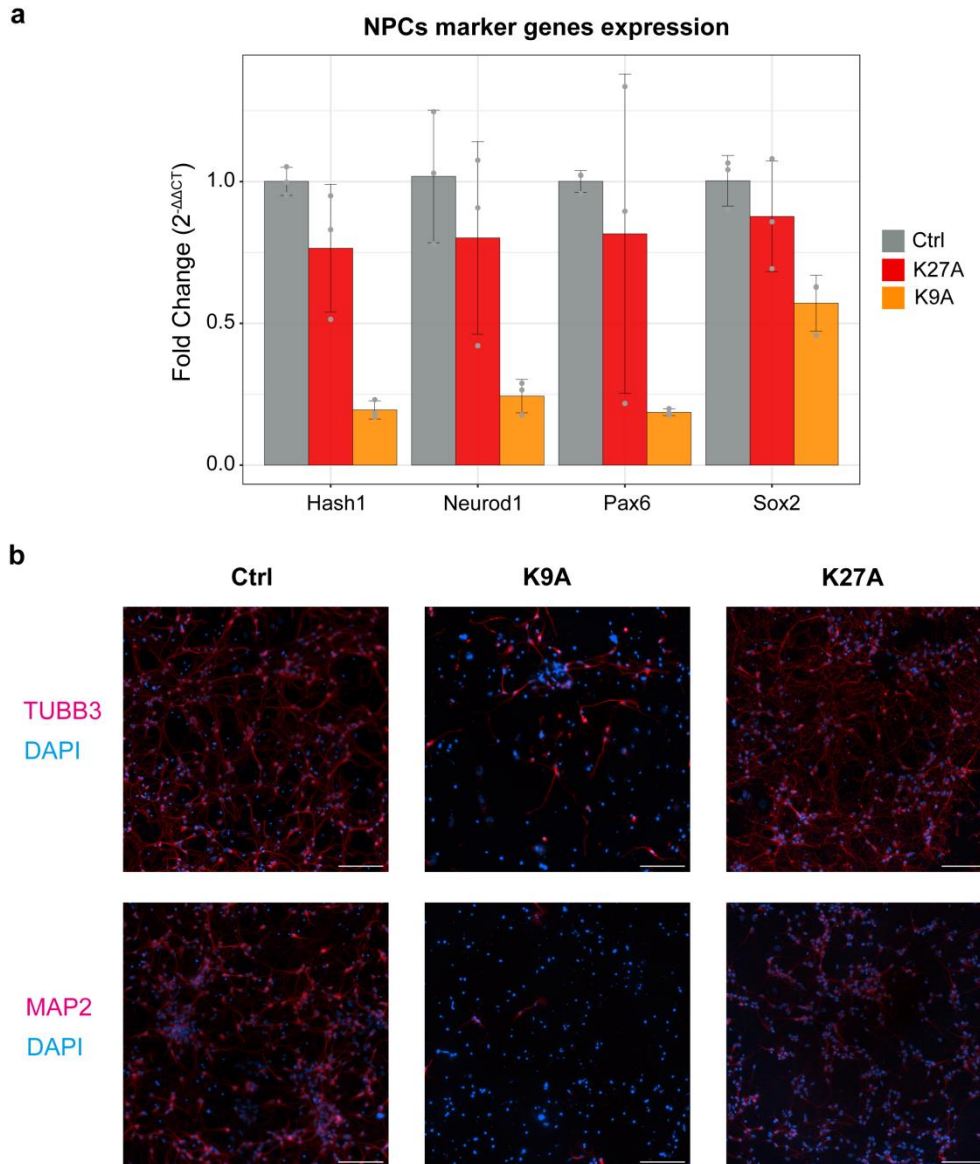


Figure 2.11 Neuronal progenitors and neuronal markers expression in differentiated cells. (a) qRT-PCR results for neuronal progenitor cell markers (day 8) in control, K9A and K27A (n=3 independent clonal lines/condition). Data are the mean $\pm$ standard deviation of gene expression values ($2^{-\Delta\Delta C_t}$) normalized to Rpl13 housekeeping gene and Ctrl. **(b)** Merged immunofluorescence images of neurons on day 12 of the *in-vitro* differentiation, stained with antibodies against TUBB3 (top, magenta) or MAP2 (bottom, magenta) and DAPI to detect nuclei. Scale bar 100 μ m.

Previous studies suggested that PRC2 may suppress meso-endoderm marker genes through H3K27me3 deposition, to counteract a neuro-ectodermal lineage bias intrinsic to pluripotent stem cells [312, 313]. A closer look at the transcriptomic data revealed upregulation of master regulators of trophoblast (e.g., *Cdx1*; *Cdx2*), mesoderm (e.g., *Tbxt*; *Snai1*; *Twist*) and cardiac-mesoderm lineage markers (e.g., *Mef2c*; *Tbx1*; *Fgf10*) (Figure 2.12a). Therefore, I performed *in-vitro* cardiac-mesoderm differentiation to examine whether and how other lineages are affected upon introducing the H3.3 mutations. The number of H3.3 K9A cells was significantly reduced at the intermediate

related to developmental processes (Figure 2.5), I examined whether the de-repression of these genes could be due to reduced H3K27me3 signal in the context of bivalent promoters.

I performed H3K27me3 chromatin-immunoprecipitation followed by sequencing (ChIP-seq) with the addition of reference exogenous genome from *Drosophila melanogaster* (spike-in, ChIP-Rx) and assessed the overall histone modifications landscape at the promoter regions of the genes upregulated in K27A, K9A or both. To visually identify bivalent promoters (i.e., H3K4me3 and H3K27me3 marks), I integrated the H3K4me3 ChIP-seq dataset from the control lines, which was previously generated by Maja Gehre [257]. In line with the gene ontology (GO) enrichment analysis, most of the promoters of the genes upregulated in H3.3K27A (K27A Up) and in both H3.3K27A and K9A (K27A_K9A Up) showed the characteristic bivalent signature (Figure 2.13a).

Among the genes upregulated only in H3.3K9A (K9A Up), many promoters showed a bivalent signature. To clarify the exact proportion of bivalent genes in each group, I generated a bivalent genes annotations (n=4117) from our ChIP-seq data set. To refine the annotation, I used publicly available data from the ChIP-Atlas database [314], and retained only the promoters which are bound by core PRC2 subunits (i.e., SUZ12, EZH2, EZH1, EED) in mESCs. The robustness of this custom annotation was confirmed using ChromHMM state models for mESCs (Figure 2.13b). About 60% of the “K27A Up” and “K27A_K9A Up” genes were confirmed to be bivalent, whereas about 32% of the “K9A Up” genes were bivalent. A minor proportion of bivalent genes was downregulated either in individuals or in both mutants (Figure 2.13c). However, the H3K27me3 signal on promoters of downregulated bivalent genes was significantly lower, likely indicating that other molecular mechanisms predominate in the regulation of these genes. All in all, 1041 bivalent genes were significantly differentially expressed with a fold-change cut-off of 2.

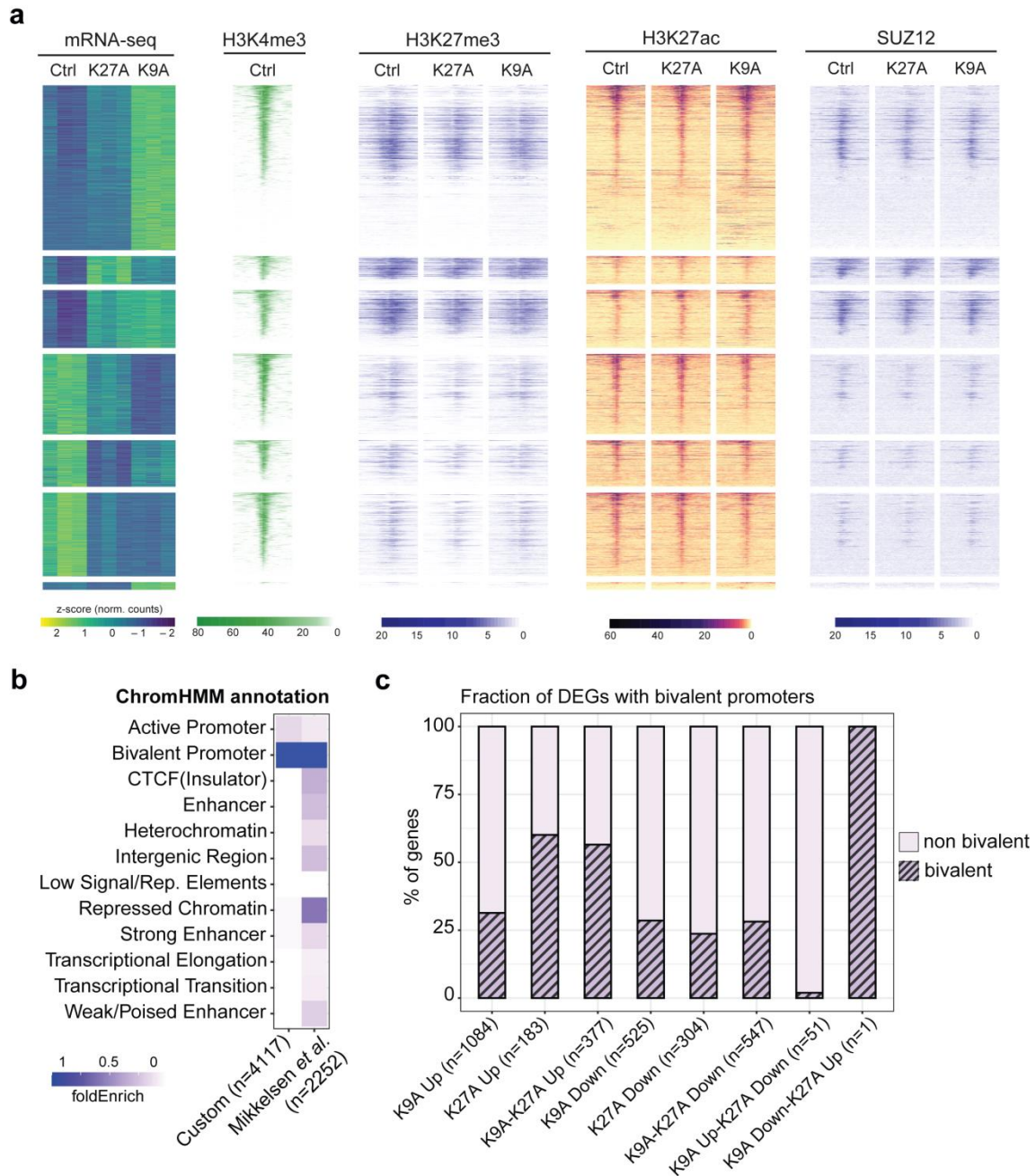


Figure 2.13 Profiling of histone modifications at promoters of DEGs in K27A and K9A mESCs. (a) Heatmap of z-score mRNA-seq normalized counts for the three biological replicates of control, K27A and K9A mESCs (left). Genes are clustered in eight groups according to the upset-plot from Fig. 2.4, based on significant de-regulation in one or in both mutants. Next, are reported heatmaps of H3K4me3, H3K27me3, H3K27ac and SUZ12 ChIP-seq signal at TSS +/- 5 kb of the genes in the same eight clusters. ChIP-seq signal was calculated as the average of three replicates for control or mutant mESCs, in 250 bp sliding genomic windows. **(b)** ChromHMM validation of the custom bivalent annotation generated using in-house H3K4me3 + H3K27me3 ChIP-seq data and compared with the annotation from Mikkelsen *et al.*, 2007 [154]. Reported is the enrichment score normalized over the maximum value among the twelve states. **(c)** Stacked barplot depicting the percentage of genes with a bivalent promoter signature in each group of differentially expressed genes.

2.2.6 H3K27me3 signal is decreased at bivalent promoters and H3.3K27A results in a reduction in SUZ12 binding and H3K27me3 spreading

H3K27me3 is deposited by PRC2. To investigate how H3K27me3 deposition is affected in the H3.3 mutant mESCs, I profiled the genome-wide distribution of other chromatin marks as well as H3.3 occupancy and PRC2 binding. Metaprofile plots of the promoter regions of the differentially expressed bivalent genes showed decreased H3K27me3 signal in both H3.3 K27A and K9A (Figure 2.14a). Conversely, H3K27ac levels were increased in K27A and—to a lesser extent—in K9A mutants at these promoters. Despite the promoter H3K27me3 signal being comparably reduced in both K27A and K9A, SUZ12 binding was more affected in K27A mESCs. I could confirm this also by performing differential SUZ12 peak-calling analysis: 1062 peaks were significantly decreased in K27A, as opposed to only 41 in K9A (Figure 2.14b).

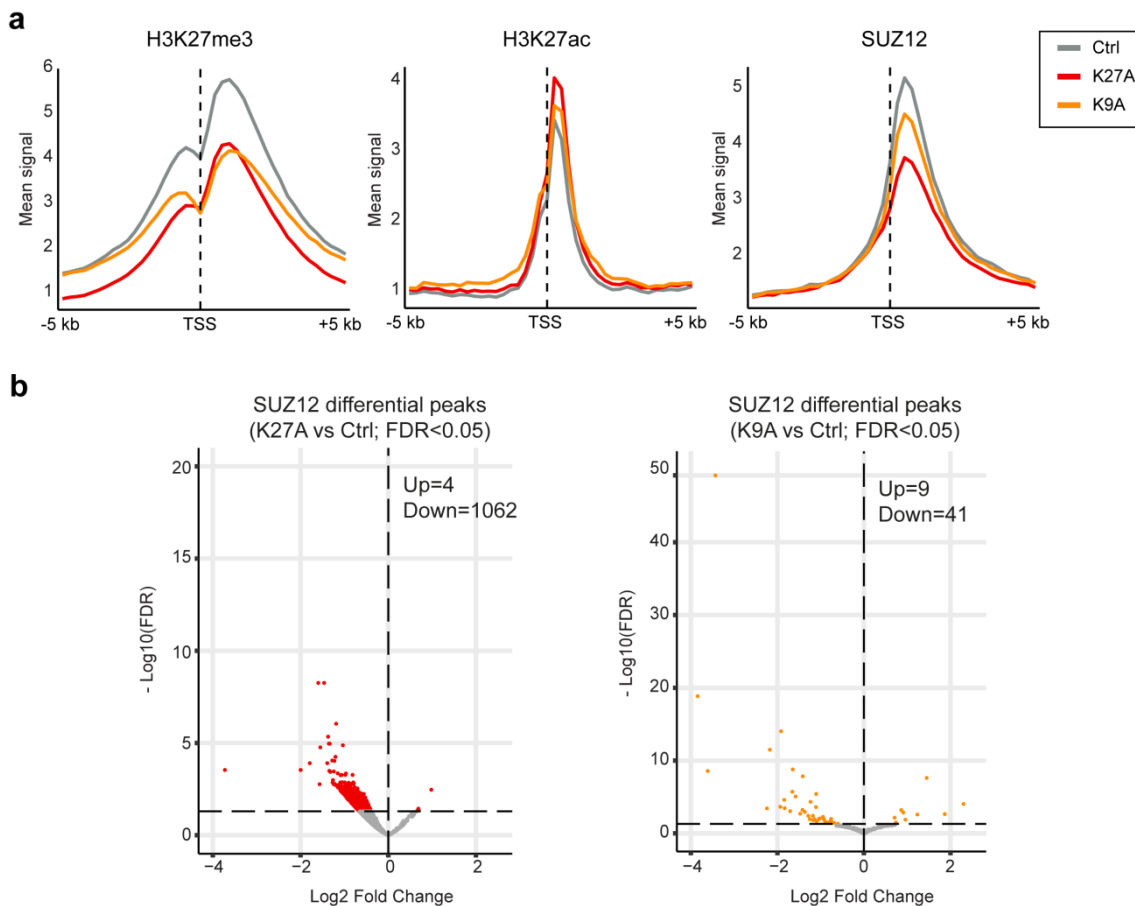


Figure 2.14 Analysis of H3K27 modifications and SUZ12 enrichment at promoters of differentially expressed bivalent genes. (a) Metaprofile plots of H3K27me3, H3K27ac and SUZ12 at TSS +/- 5kb of differentially expressed bivalent genes (n=1041). **(b)** Volcano plots showing differential SUZ12 ChIP-seq peaks in K27A (left) or K9A (right) versus control (consensus peak-set with n=3356 peaks).

The mechanism of PRC2 allosteric activation foresees that the presence of the H3K27me3 mark can establish a positive feedback loop whereby PRC2 binds and spreads the

modification. Therefore, I inspected whether the spreading of the H3K27me3 mark was more affected in the absence of the H3.3K27 residue. I measured the H3K27me3 signal within SUZ12 peaks and in the 1.5 kb regions immediately upstream or downstream of the peaks (similarly to the analysis reported in Lavarone *et al.*, 2019 [124]). Notably, while the H3K27me3 signal within SUZ12 peaks was similarly reduced in K27A and K9A, H3K27me3 levels were further decreased only in K27A mESCs at the immediate proximity of the SUZ12-bound regions (Figure 2.15a). This result indicates that the H3.3K27 residue contributes to establishing the proper H3K27me3 mark at PRC2-bound promoters and to the spreading of the modification. In addition, the bivalent genes upregulated in K27A mESCs (n=323, including also the ones shared with K9A) exhibited significantly higher levels of H3.3 at promoter regions, when compared to promoters of the bivalent genes upregulated only in K9A mESCs or to a comparable number of randomly selected non-differential bivalent genes (Figure 2.15b).

All in all, in the absence of the H3.3K27 residue, H3K27me3 deposition and its spreading are affected at PRC2-target promoters. A specific subset of bivalent promoters with a higher basal level of H3.3 is more susceptible to this decrease and these genes are de-repressed in K27A mESCs.

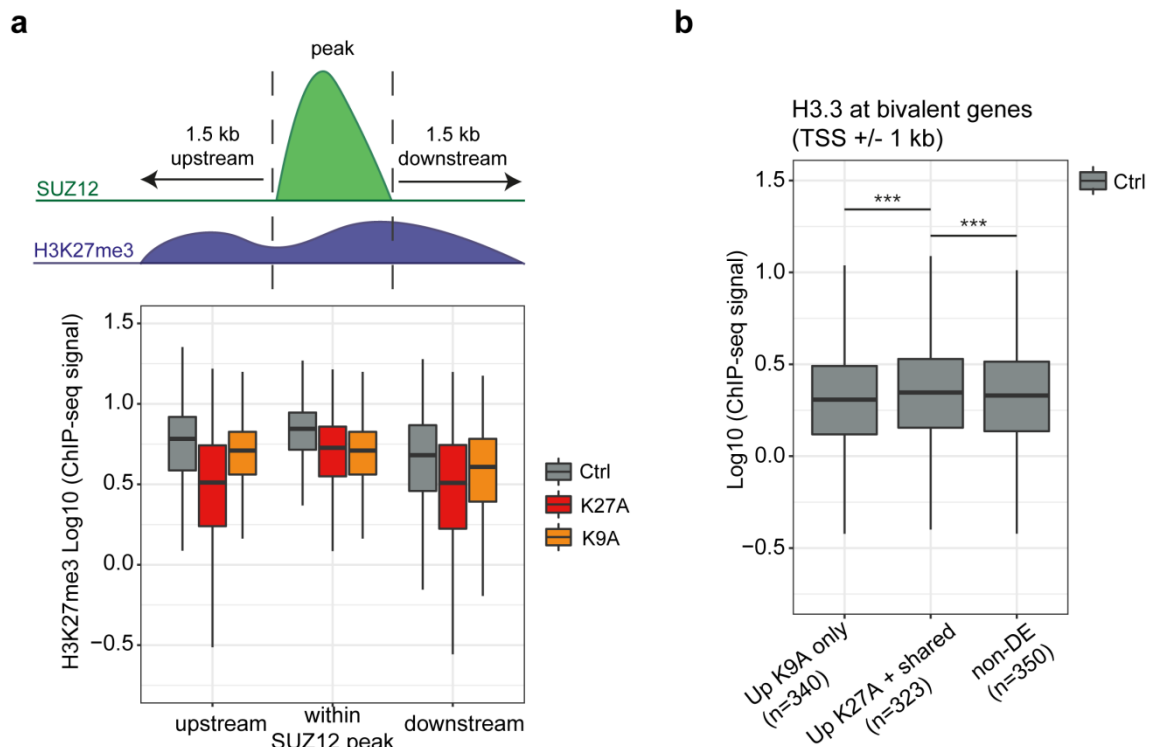


Figure 2.15 H3K27me3 signal spreading at SUZ12-bound regions and H3.3 signal at bivalent promoters. **(a)** Analysis of H3K27me3 signal within SUZ12 peaks (n=3356) and in the 1.5 kb genomic region immediately upstream or downstream. ChIP-seq signal is calculated as the average of three replicates for control or mutant mESCs, in 250 bp sliding genomic windows in the three intervals defined. SUZ12 peaks were re-sized to 2 kb (peak summit +/- 1 kb). **(b)** H3.3 ChIP-seq signal in control mESCs, at the promoter regions of

bivalent genes up-regulated only in K9A or K27A. A group of comparable size with non-significantly differentially expressed bivalent genes was randomly selected for comparison. ChIP-seq signal calculated as described above. Significance was calculated with unpaired t-test (***: p-value<0.001).

2.2.7 Bivalent promoters do not display enrichment in H3K9me2/3 and local changes in histone PTMs fail to explain the H3K27me3 decrease observed in K9A mESCs

Next, I sought to determine why H3K27me3 was also affected in the absence of the H3.3K9 residue. As summarized in the Introduction, a few reports have shown that H3K9me2 and/or H3K9me3 can contribute to maintaining Polycomb target genes in a repressed state [110, 164]. First, to better characterize the promoter regions of the genes selectively upregulated in H3.3K9A, I checked for relative enrichment of chromatin and transcription factors at these genomic loci using publicly available data from the ChIP-Atlas database (n=292 total factors for mESCs). When limiting the analysis to the differentially expressed bivalent genes, all the upregulated genes—regardless of the mutant line—displayed enrichment only for PRC2 and PRC1 subunits (e.g., SUZ12, MTF2, JARID2, RING2, PCGF2, CBX7) (Figure 2.16).

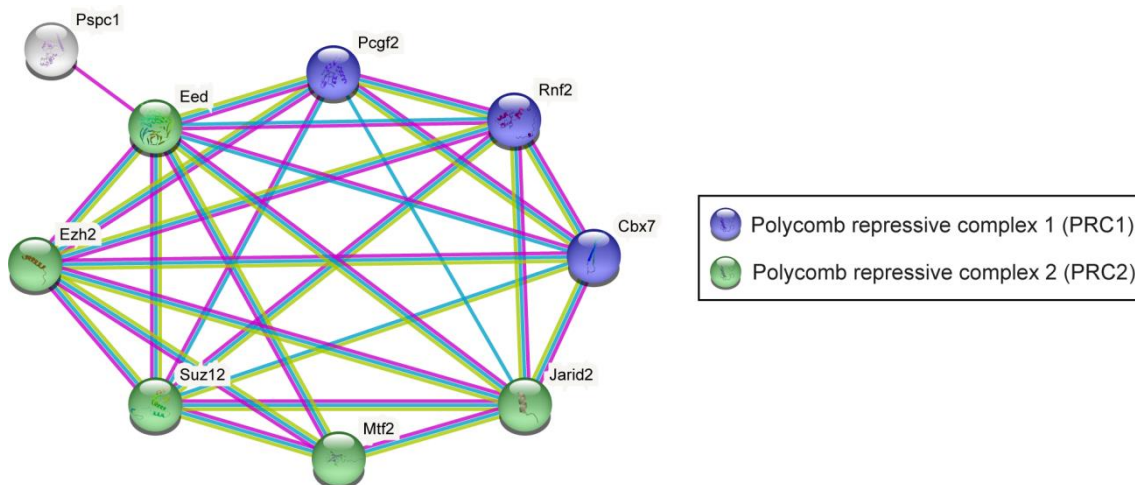


Figure 2.16 STRING network analysis of chromatin factors enriched at promoters of upregulated bivalent genes. PRC1 (blue) and PRC2 (green) subunits show enrichment at promoters (TSS +/- 0.5 kb) of upregulated bivalent genes. Significant enrichment was determined by means of Fisher's exact test, comparing significantly differentially expressed bivalent genes with non-significantly expressed genes. STRING settings: physical subnetwork (the edges indicate that the proteins are part of a physical complex) – medium confidence (interaction score 0.4).

Expanding the analysis to all the upregulated genes, the ones specific for K27A and shared between K27A and K9A preserved enrichment for Polycomb subunits. However, 63 additional chromatin factors were enriched at the promoters of genes exclusively upregulated in K9A (Figure 2.17). STRING analysis revealed that 11 are factors involved in

H3K9me3 deposition, DNA methylation and heterochromatin maintenance in general (e.g., G9a/EHMT2; SETDB1; KAP1/TRIM28; CBX3; CBX5; SMARCA5; etc.). Factors of the SWI/SNF remodeler complex (e.g., SMARCD1; SMARCA5; etc.) and components of the Paf1 and transcription elongation complexes (e.g., PAF1; LEO1; CDC73; ELL2; etc.) were also found as enriched at these promoters. Other hits were factors of the NuA4 acetyltransferase complex (YEATS4; RUVBL1 and RUVBL2), factors of the MLL3/4 complex (KDM6B and WDR5) and the three main factors involved in miRNA biogenesis (DROSHA; DGCR8 and DICER1). Notably, several of these factors have reported roles in heterochromatin formation and maintenance [315, 316].

All in all, the enrichment of these factors is in line with the chromatin modifications state observed at promoters of the genes selectively upregulated in H3.3K9A mESCs, whereby over 70% show no H3K4me3 TSS signal hence indicating inactive transcription in the basal state.

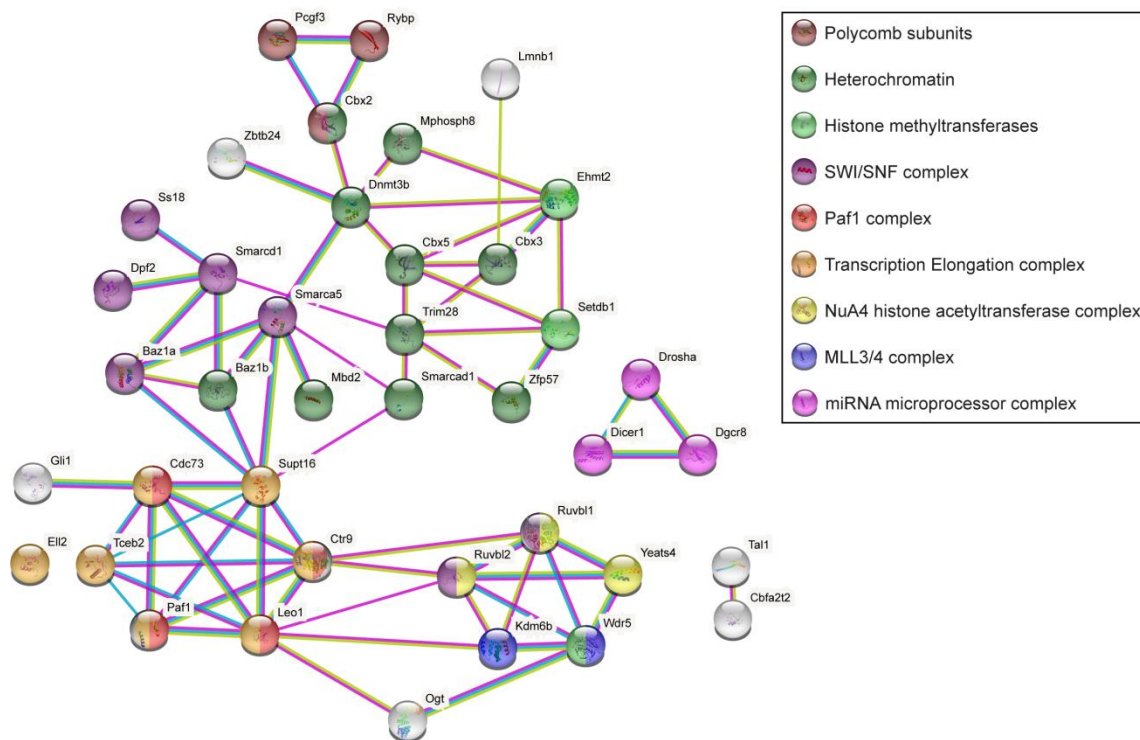


Figure 2.17 STRING network analysis of chromatin factors enriched at promoters of genes upregulated only in K9A. Reported are 38/63 factors enriched at TSS +/- 0.5 kb of genes upregulated solely in K9A mESCs. Colours indicate the chromatin complex/category to which the individual proteins belong. STRING settings: physical subnetwork (the edges indicate that the proteins are part of a physical complex) – medium confidence (interaction score 0.4).

A previous study reported the role of G9a (H3K9 methyltransferase) in reinforcing PRC2 activity at bivalent promoters [164], implying a function of H3K9 methylation in PRC2 action. I investigated this possibility in collaboration with the Davidovich Lab at EMBL Australia and Monash University. They performed a PRC2 activity assay using *in-vitro*

reconstituted chromatin with DNA sequence from the ATOH1 bivalent locus. We did not observe enhanced PRC2 activity (e.g., H3K27me3 increase) in the chromatin substrate presenting H3K9me2 or H3K9me3 modification. Instead, chromatin reconstituted with non-methylated H3.3K9 was a better substrate for PRC2 than H3.1K9 and di-/tri-methylated H3.3/H3.1 (Figure 2.18b).

I also performed H3K9me2 and H3K9me3 ChIP-seq and profiled the signals at TSSs of the bivalent genes whose expression was disrupted in mutant mESCs. H3K9me2 and H3K9me3 signals were very scarce at bivalent promoters (Figure 2.18a) and upregulated genes in general (despite K9A Up genes exhibiting a higher basal level, in accordance with the ChIP-Atlas integration). These results indicate that reduced H3K27me3 at a subset of bivalent promoters and de-repression of the bivalent genes in H3.3K9A mESCs are not due to the disrupted PRC2 activity, nor coupled with the abundance of promoter H3K9 modifications.

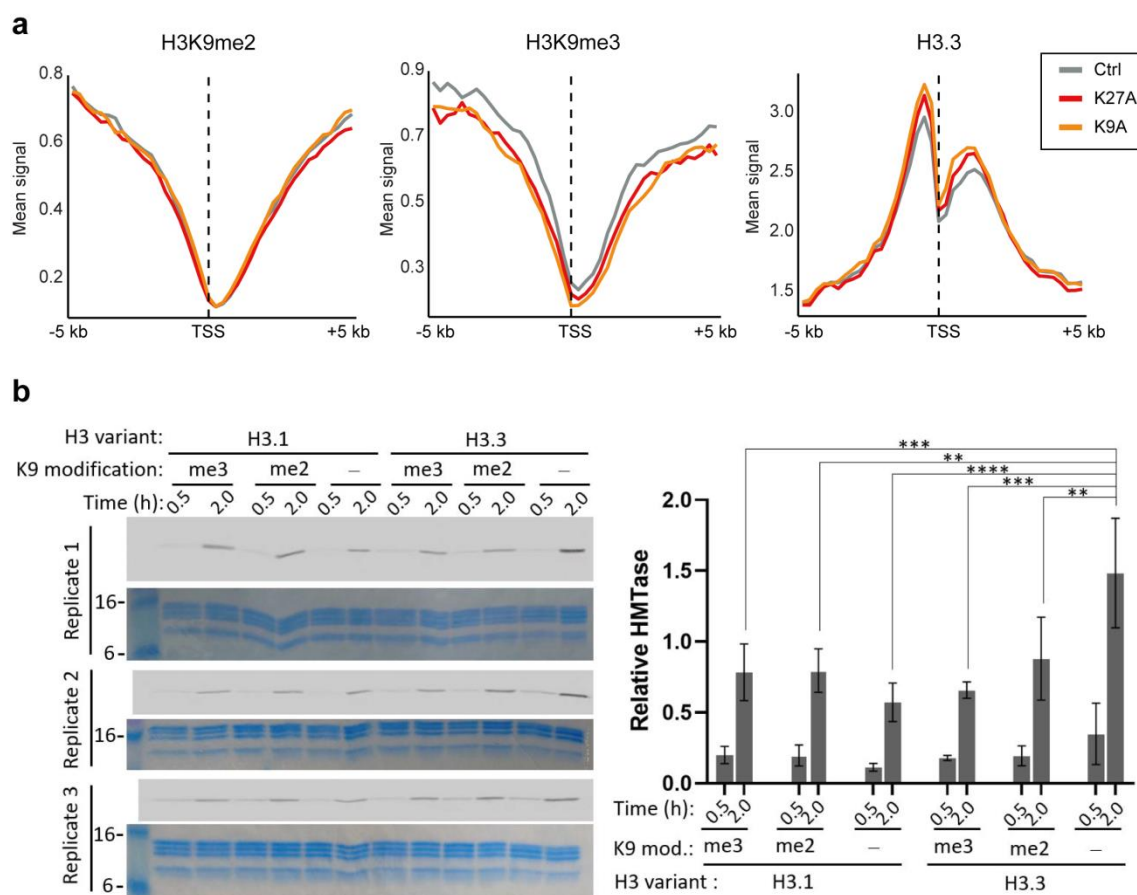


Figure 2.18 Analysis of H3K9 modifications at promoters of differentially expressed bivalent genes and in-vitro Polycomb HMT assay. (a) Metaprofile plots of H3K9me2, H3K9me3 and H3.3 at TSS +/- 5 kb of differentially expressed bivalent genes (n=1041). **(b)** Histone methyltransferase assays of PRC2 in the presence of chromatin reconstituted using different H3 variants and H3K9 modifications. In all cases, “me2” and “me3” represent the respective methyl lysine analogue (MLA) and “-” represents an unmodified lysine residue. The barplot (right) represents the mean densitometry values as recorded from the radiograms (left). The error bars represent the standard deviation over three replicates carried out on different days. Significance was calculated with Tukey’s multiple comparisons test (****: $P < 0.0001$, ***: $P < 0.001$, **: $P < 0.005$, *: $P < 0.05$).

2.2.8 H3K9me3 signal is reduced at distal intergenic regions in K9A mESCs

Given that I could not detect local changes in H3K9 methylation at promoters of genes de-repressed in K9A mESCs, H3.3 K9 might contribute to gene silencing via repression of distal regulatory elements. These repressed distal elements might be disrupted in H3.3K9A mESCs and activate putative target genes with bivalent promoters. To investigate this possibility, I defined a consensus peak-set of H3K9me3 broad regions and performed differential peak analysis to narrow down the analysis to regions with changes in H3K9me3 abundance. Out of ~48300 H3K9me3 broad peaks identified, 8685 were differentially abundant (DA) in K9A mESCs, against only 2 in K27A (Figure 2.19a). Almost 95% of the DA peaks in K9A (i.e., 8203) were less enriched in the mutant compared to the control. Unsurprisingly, annotation of the DA H3K9me3 peaks with ChromHMM showed that most of the peaks are located in heterochromatic regions (Figure 2.19b). Further genomic annotation of the H3K9me3 regions affected by the K9A mutation revealed that over 68% of the peaks fall in distal intergenic regions, followed by ~18% in introns, ~3% within 2 kb window from TSSs and the rest located in exons and 5'-/3'-UTRs (Figure 2.19c).

To verify that the affected 3% H3K9me3 signal in promoter regions could not account for the majority of the transcriptional changes detected, I retrieved these genes and checked their expression: out of 238 genes detected, only 50 were included among the ones differentially expressed and 58% of those (n=29) were among the genes upregulated exclusively in K9A.

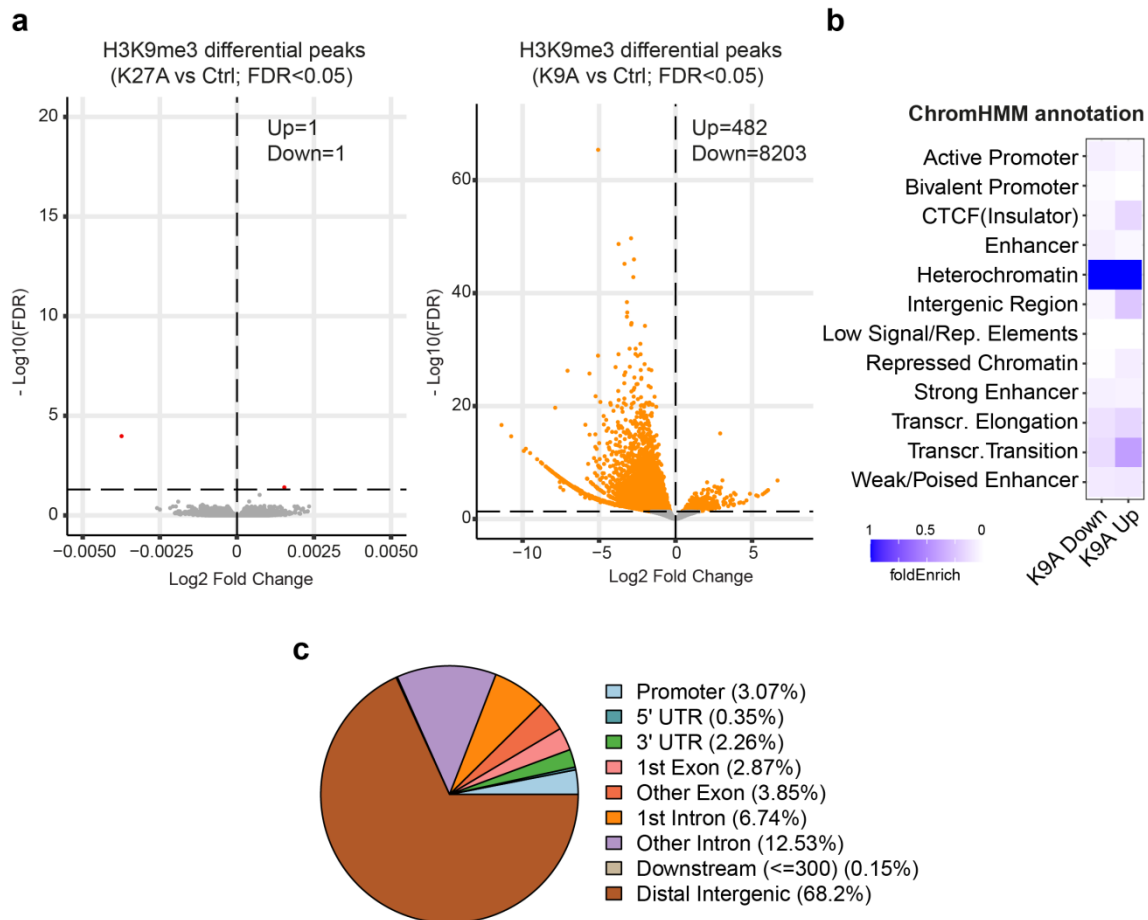


Figure 2.19 H3K9me3 ChIP-seq differential analysis. (a) Volcano plots showing differential H3K9me3 ChIP-seq peaks in K27A (left) or K9A (right) versus control (consensus peak-set with $n=48346$ broad peaks). (b) ChromHMM annotation of differentially abundant H3K9me3 regions in K9A mESCs. Reported is the enrichment score normalized over the maximum value among the twelve states. (c) Genomic annotation of the H3K9me3 regions with significantly reduced signal in K9A mESCs. Promoter defined as TSS $\pm$ 1 kb.

2.2.9 Genomic regions with reduced H3K9me3 signal in H3.3K9A mESCs gain H3K27ac and exhibit increased levels of nascent transcription

On the assumption that H3.3 K9 might be involved in safeguarding the repression of distal regulatory elements, the anticipated consequence of a decreased H3K9me3 signal would be the acquisition of an active chromatin state. H3K27ac differential peak-calling analysis showed a significant increase at 12723 regions (Figure 2.20a). ChromHMM annotation showed that most of these gained H3K27ac peaks in K9A are located in heterochromatic regions (Figure 2.20b). The genomic distribution of the gained H3K27ac peaks also fit with the distribution of the H3K9me3 lost regions (Figure 2.20c and Figure 2.19c). Notably, the H3K27ac signal was reduced at over 14000 peaks and I will return to this in section 2.2.17.

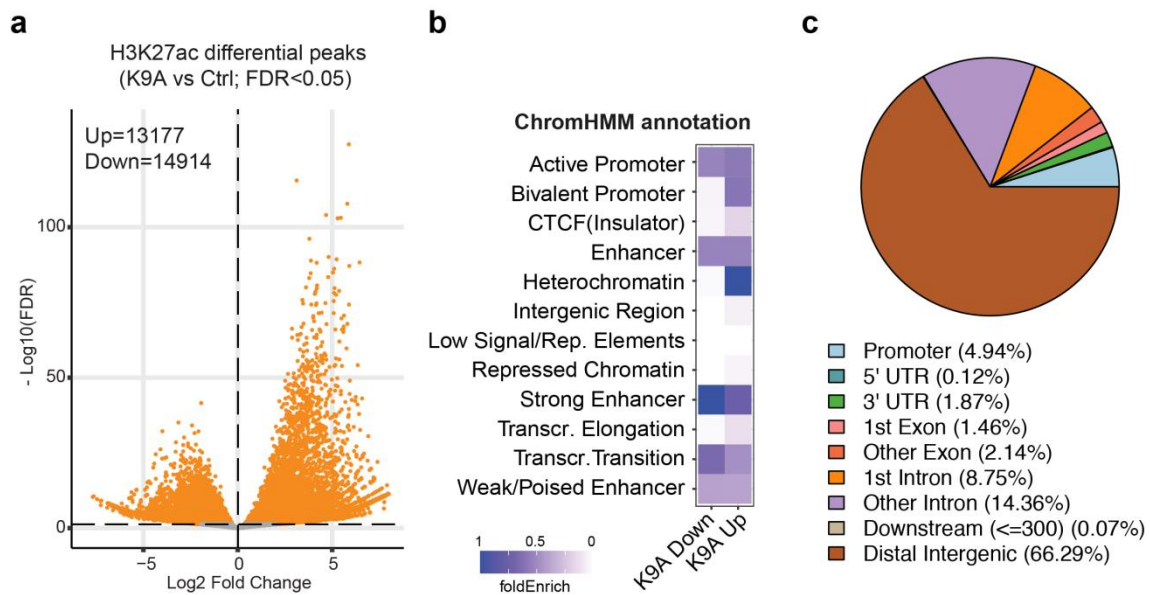


Figure 2.20 H3K27ac ChIP-seq differential analysis in K9A mESCs. (a) Volcano plots showing differential H3K27ac ChIP-seq peaks in K9A mESCs (consensus peak-set with $n=94000$ peaks). **(b)** ChromHMM annotation of differentially abundant H3K27ac regions in K9A mESCs. Reported is the enrichment score normalized over the maximum value among the twelve states. **(c)** Genomic annotation of the H3K27ac regions with significantly increased signal in K9A mESCs. Promoter defined as TSS $\pm$ 1 kb.

To find out whether the observed increase in H3K27ac happened in the same genomic territories with decreased H3K9me3, I clustered the H3K9me3-differentially abundant peaks in two groups—based on the degree of signal reduction in K9A—and sorted the peaks by decreasing signal intensity in control. For comparison, I included the analysis of 8500 randomly selected H3K9me3 peaks with non-differentially abundant (non-DA) (Figure 2.21). The H3K9me3 signal was completely gone in the first cluster (henceforth defined as “Cluster 1”). The deposition of the H3.3 was higher in Cluster 1, and the K9A variant was maintained at these regions, indicating that H3.3 is the major contributor to H3K9me3 build-up at these genomic regions. The basal level of H3K27ac was also higher in Cluster 1 compared to the other regions included in the second cluster (henceforth called “Cluster 2”).

Cluster 1 regions exhibited a prominent increase in H3K27ac levels in K9A mESCs. This gain was not limited to one acetylation mark, as also H3K9ac signal was increased. While Cluster 1 regions exhibited a certain degree of basal histone H3 acetylation, Cluster 2 regions were instead immaculate. Yet, H3K9me3 was also reduced here and I could detect a modest increase in both H3K27ac and H3K9ac signals (Figure 2.21).

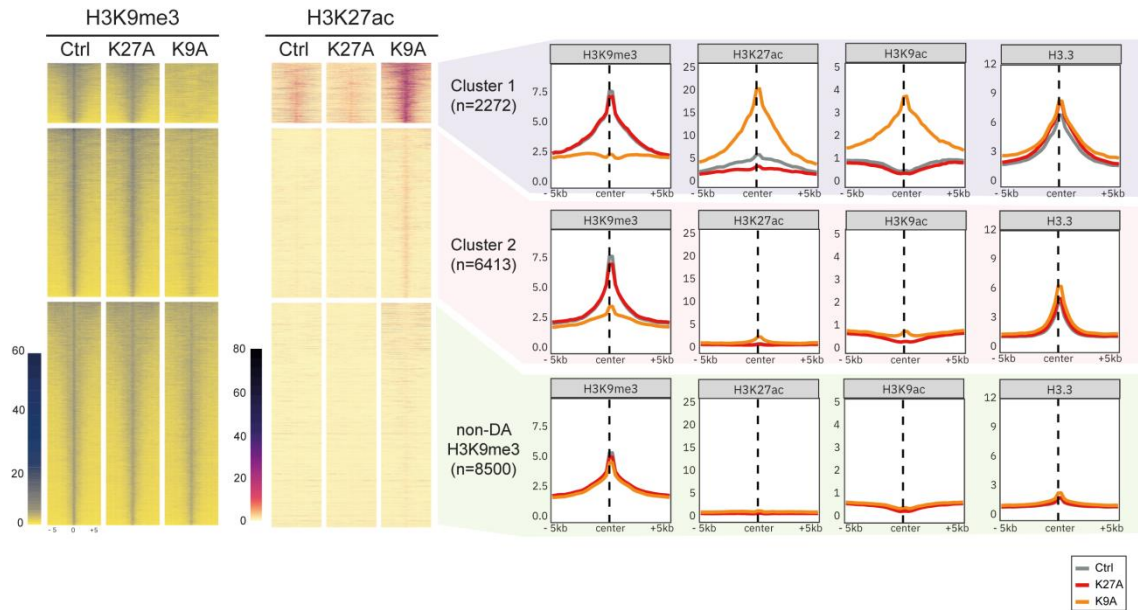


Figure 2.21 Histone H3 acetylation is increased at regions with reduced H3K9me3 in K9A mESCs. Heatmaps of H3K9me3 and H3K27ac ChIP-seq signal at H3K9me3 regions (summit +/- 5 kb). Differentially abundant regions are divided into two clusters based on H3K9me3 signal reduction and sorted by H3K27ac signal. ChIP-seq signal calculated as the average of three replicates for control or mutant mESCs, in 250 bp sliding genomic windows. On the right, metaprofile plots with H3K9me3, H3K27ac, H3K9ac and H3.3 average ChIP-seq signals are reported.

Because Cluster 1 and Cluster 2 regions displayed a different degree of changes in histone modifications in the absence of the H3.3K9 residue, I sought to determine if the two clusters would show any intrinsic differences. In line with the higher basal level of histone H3 acetylation detected in Cluster 1, integration of data from the ChIP-Atlas collection revealed relative enrichment of several transcriptional co-activators (e.g., P300; BRD4; MED24; etc.) as well as subunits of chromatin remodellers and of the elongation complex (i.e., SMARCA4; CHD7). Interestingly, also TFs linked to pluripotency (i.e., NANOG; SOX2; OCT4/POU5F1) and development (i.e., SOX17; SMAD2/4; YAP1) showed relative enrichment at Cluster 1 regions, supporting the hypothesis that these regions could regulate developmental genes (Figure 2.22a). ChIP-Atlas data integration at Cluster 2 regions showed a relatively higher enrichment for classic heterochromatin factors such as HP1 proteins (i.e., CBX1; CBX3; CBX5), as well as KAP1/TRIM28, SETDB1 and other transcriptional repressors (Figure 2.22a).

Active distal regulatory elements typically show increased transcriptional activity. I therefore performed PRO-seq to assess whether Cluster 1 and 2 regions also acquired this characteristic feature. Cluster 1 regions—which displayed a more prominent H3K9me3 loss and H3 acetylation increase—showed a marked gain in nascent transcription levels (Figure 2.22b). In line with the ChIP-seq data, Cluster 2 regions showed modestly increased nascent transcription. Altogether, I described how the H3.3K9 residue critically

contributes to maintaining certain distal intergenic regions in a repressed state. In the absence of this residue (but not H3.3K27), H3K9me3 is severely reduced at these genomic regions, with a concomitant increase in H3 acetylation and transcriptional activity.

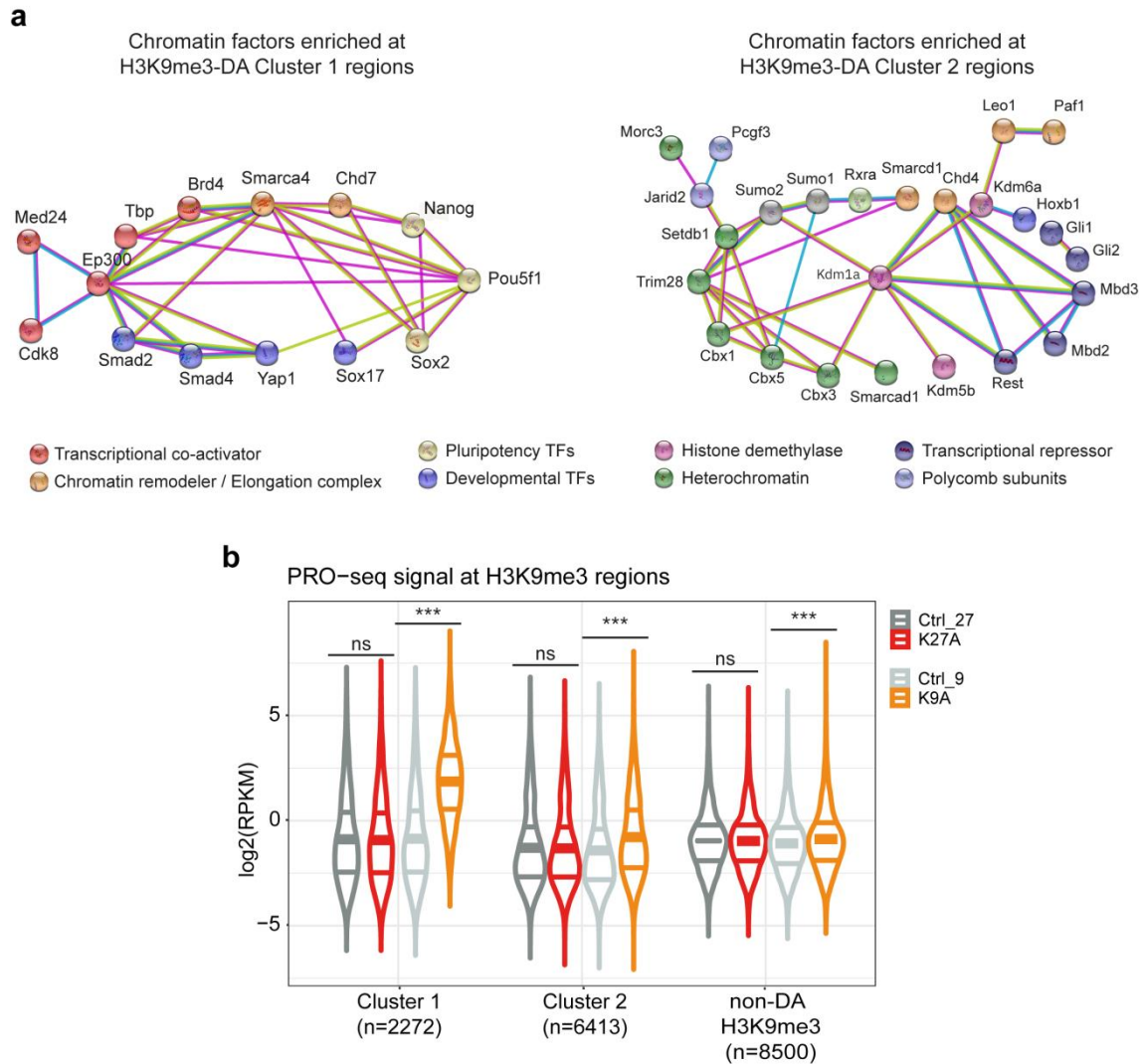


Figure 2.22 Cluster 1 H3K9me3-DA regions are more active in the basal state and display greater increase in nascent transcription activity upon H3K9me3 loss. (a) STRING network analysis of chromatin and transcription factors enriched at Cluster 1 (left) and Cluster 2 (right) H3K9me3-DA regions. Colours indicate the chromatin complex/category to which the individual proteins belong to. STRING settings: physical subnetwork (the edges indicate that the proteins are part of a physical complex) – medium confidence (interaction score 0.4). **(b)** Nascent transcription signal at H3K9me3-DA regions reported as $\log_2(\text{RPKM})$ values. Significance was calculated with unpaired t-test (***: $p\text{-value} < 0.001$; ns: $p\text{-value} > 0.05$). Due to technical reasons, I performed the PRO-seq experiment independently for K27A and K9A. Control samples were repeated in each batch and reported in two different shades of grey.

2.2.10 H3.3K9 maintains ERVs in a repressed state in mESCs and regions with reduced H3K9me3 signal contain ERV-derived regulatory elements

In mESCs, H3K9me3 deposition is considered the predominant mechanism for ERVs repression. Nevertheless, as summarized in the Introduction section “1.7.2 H3.3 and ERVs repression”, the exact contribution of the H3.3 histone variant to this regulatory mechanism is still debated. The H3.3 K9A mutant generated in this project represents a valid means to test H3.3 involvement in ERVs silencing in mESCs. I hence analysed the expression of mouse LTR elements using annotations retrieved from the UCSC RepeatMasker track. Principal component analysis (PCA) revealed that most of the variance could be explained by the K9A samples (Figure 2.23a). Differential expression analysis showed the de-repression of ERVs selectively in H3.3K9A mESCs, with IAP and IAPE elements being among the most highly transcribed repeats (Figure 2.23b).

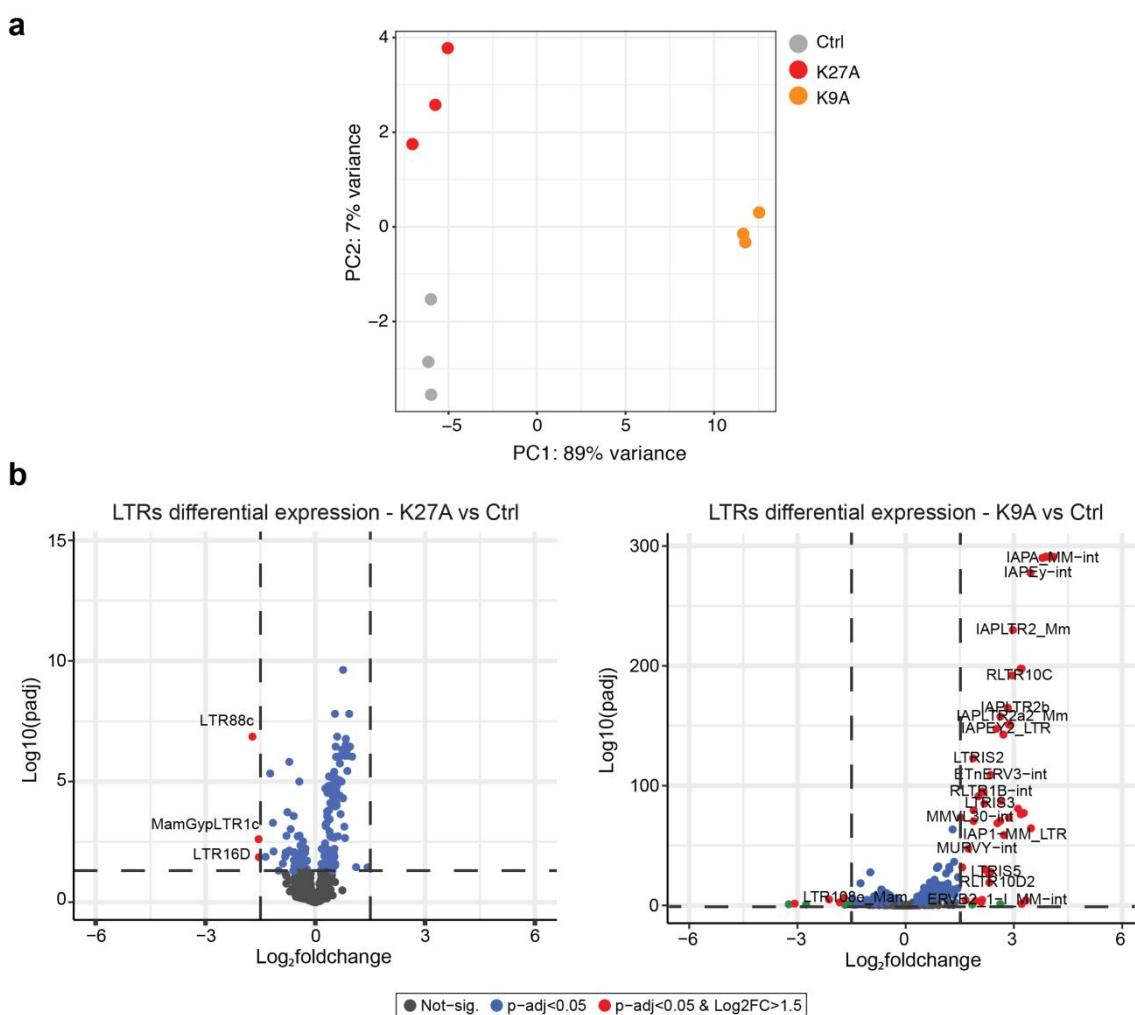


Figure 2.23 LTRs differential expression analysis. (a) Principal component analysis. **(b)** Volcano plots showing differential ERVs expression in K27A (left) and K9A (right versus control).

To validate this analysis, I checked the expression of selected TEs using RT-qPCR and I could confirm selective de-repression of ERVs, but not of LINE and SINE elements (Figure 2.24a). At these same loci, I also assessed the H3K9me3 level using ChIP-qPCR and the H3K9me3 enrichment was only affected at ERVs (Figure 2.24b), in line with the RT-qPCR. Given that H3.3K9A mESCs displayed activation of potential distal *cis*-regulatory elements (CREs), I investigated the degree of overlap between these regions and mouse ERVs. In accordance with the higher levels of histone acetylation and nascent transcription, Cluster 1 regions showed a greater percentage of ERVs presence in the reference mouse genome, with more than 60% of the H3K9me3 peaks included in this cluster displaying one or more ERVs (Figure 2.24c). The fraction of regions with ERVs was reduced to 40% for Cluster 2 regions and down to 25% in the control set of non-DA H3K9me3 peaks. Therefore, many genomic regions gaining an active chromatin state in K9A mESCs are likely to be ERV-derived enhancers.

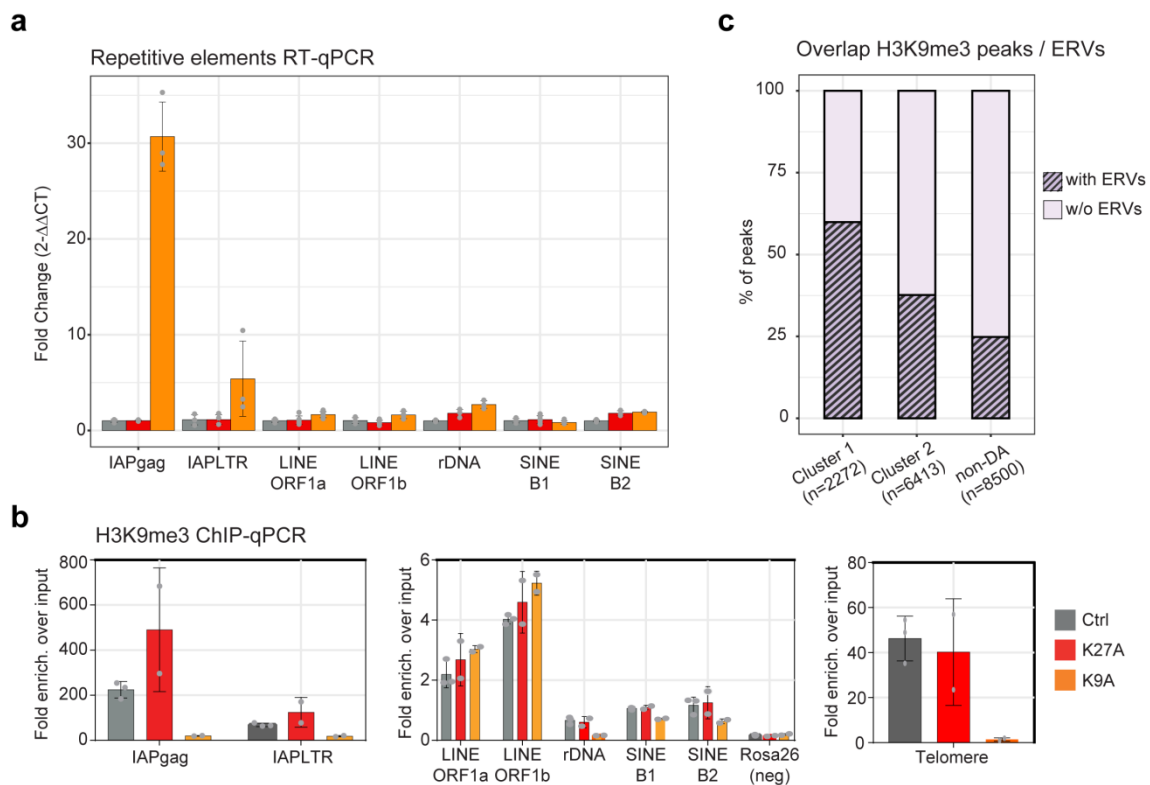


Figure 2.24 Validation of ERVs expression and overlap between ERVs and H3K9me3-DA regions. (a) Real-time qPCR using primers for selected transposable elements shows selective upregulation of ERVs in K9A mESCs. **(b)** H3K9me3 ChIP-qPCR shows signal reduction selectively at ERVs (left) and not at other transposable elements (right). Rosa26 locus was used as a negative control for H3K9me3 enrichment. Telomeric H3K9me3 signal (right) was also reduced (in agreement with Udugama *et al.*, 2015 [230]) **(c)** Stacked barplot display percentage of H3K9me3-DA with at least one ERV (size cut-off of 500bp was applied for ERVs).

2.2.11 Activation of ERV-derived regulatory elements in H3.3K9A mESCs drives upregulation of both bivalent and immune-related genes

Having identified putative distal CREs that are activated without the H3.3K9 residue, the next step was to find out whether these elements could act as regulatory elements driving the expression of genes selectively upregulated in K9A mESCs.

To accomplish this, Daria Bunina and Brian Mang Ching Lai from the Noh and Zaugg Groups generated a gene regulatory network (GRN) based on co-variation between mRNA-seq and H3K27ac ChIP-seq data across multiple samples [317]. A total of 12 samples were used—specifically, the three biological replicates for K9A, K27A, K79A and control mESCs. The algorithm employed allows us to infer interactions between enhancer/promoter and genes as well as between TFs and enhancers/promoters.

A total of 657 H3K9me3-DA regions (n=592 from Cluster 1; n=65 from Cluster 2) were retained in the GRN and I confirmed that the original overall genomic distribution was still maintained, with ~65% of H3K9me3-GRN regions annotated as distal intergenic (reasonably, I could detect a mild increase in the fraction of regions overlapping with promoters – from ~3% of the total DA H3K9me3 to ~5% of the DA H3K9me3-GRN regions – Figure 2.25a). Through the GRN, we could link these regions to a total of 789 genes (average peak-gene distance $\approx$ 108 kb). Analysis of the expression of these genes revealed upregulation of the genes linked to H3K9me3 DA regions, exclusively in H3.3K9A mESCs (Figure 2.25b). To answer whether the ERV-derived regulatory elements could be responsible for the activation of bivalent genes, I checked the proportion of H3K9me3 DA-linked genes with a bivalent promoter signature. About 20% (n=153) exhibited bivalent promoters and, among those, 130 were affected explicitly in K9A or at most shared with K27A, while only 8 were exclusive for K27A mESCs. A similar trend was also observed for the remaining 636 genes, with ~80% of the genes being affected in K9A (Figure 2.25c).

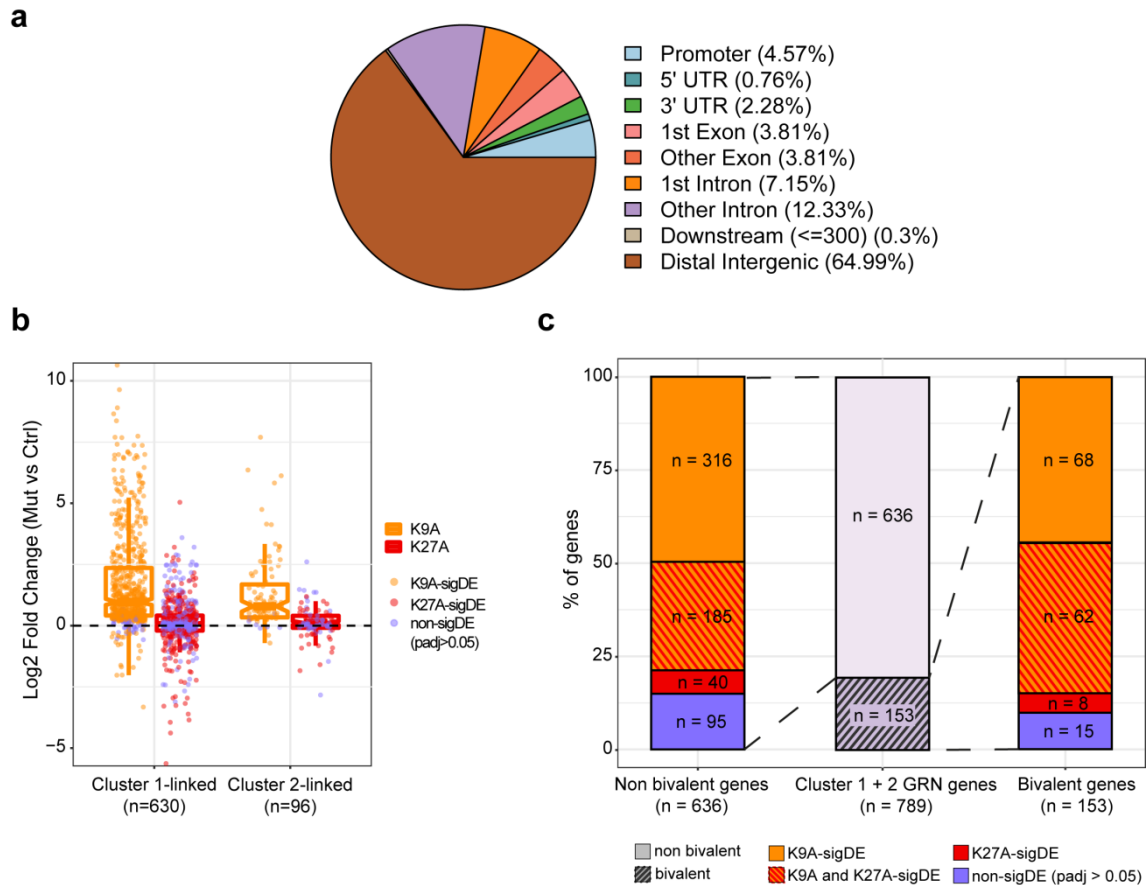


Figure 2.25 Genes connected to H3K9me3-DA regions are upregulated in K9A mESCs. (a) Genomic annotation of H3K9me3-DA regions retained in the GRN (n=657). **(b)** Expression of DEGs connected to H3K9me3-DA regions through the GRN; $\log_2(\text{fold-change})$ values calculated with DESeq2 are plotted, and individual genes (data points) are coloured to indicate if the differential expression is significant (orange/red) or not (purple). **(c)** Stacked barplot indicating the percentage of genes with or without bivalent promoters (middle), and details of significant differential expression for genes without (left) or with (right) bivalent promoters.

From the mRNA-seq, GO analysis of the genes specifically upregulated in K9A mESCs had previously shown enrichment for terms related to immune processes (e.g., leukocyte cell-cell adhesion; regulation of T cell activation; etc. – Figure 2.5). As outlined in the Introduction section “1.7.3 Roles of ERVs in transcriptional regulation”, the de-repression of ERVs and thus activation of ERV-derived regulatory elements in H3.3K9A mESCs could provide a plausible explanation for the ectopic expression of such genes in stem cells. To verify that the upregulation of immune-related genes could be directly driven by alien activation of CREs, I performed GO analysis of the genes connected to these elements through the GRN (Figure 2.26). The analysis revealed enrichment for terms such as leukocyte-mediated cytotoxicity and regulation of response to a biotic stimulus. Inspection of individual genes underlying these terms showed presence of Toll-like receptor genes (Tlr2), genes encoding for proteins of class II major histocompatibility complex (MHC, e.g., H2-M2 and H2-T23) and also interferon inducible proteins. Two other major groups of GO

terms were enriched: the first was related to the inhibition of protease activities, with several genes encoding for members of the serine protease inhibitors (serpins) being upregulated. The other terms were instead related to fatty acid and eicosanoids metabolism. Examples of genes connected to these terms were genes belonging to the cytochrome P450 monooxygenase (CYP) superfamily, aldo-keto reductases and acyl-coenzyme A thioesterase coding genes. Interestingly, serpin family members—in particular the ones belonging to the ovalbumin (ov)-serpin subfamily—were reported to play a role in immune cell response [318], and among the serpin genes de-repressed in H3.3K9A mESCs I could identify some that are known to be expressed specifically in immune cell types (i.e., *Serpinb1*; *Serpinb6*; *Serpina3* genes). Similarly, also fatty acids and the eicosanoids derived from them through oxygenation reactions are reported as critical pro-inflammatory compounds playing roles in innate immunity [319].

Taken together, the genes connected to the ERV-derived CREs through the GRN are de-repressed specifically in H3.3K9A mESCs, and display a strong enrichment for terms related to immune-related processes. Some of these genes also show a bivalent promoter signature, indicating that the H3K27me3 decrease observed at the promoters might be forced by the activation of distal putative regulatory elements.

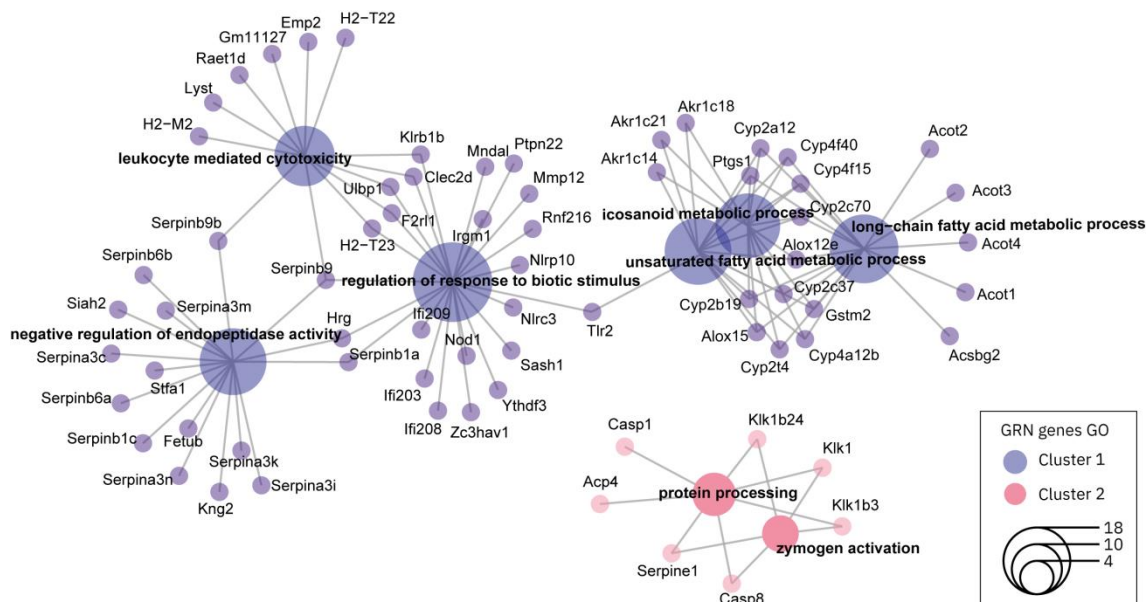


Figure 2.26 Gene ontology analysis of genes connected to H3K9me3-DA regions. Cnet plot with GO terms enriched for genes connected to Cluster 1 (purple) or Cluster 2 (pink) H3K9me3-DA regions. Individual genes associated with the GO terms are reported.

2.2.12 Regulatory elements activated in H3.3K9A mESCs overlap with CREs active in immune cell types

The GRN helped to disclose a connection between the activation of the CREs activated in H3.3K9A mESCs and the expression of immune-related genes, pointing towards the involvement of the putative ERV-derived CREs in the direct transcriptional regulation of these genes. I therefore asked whether a fraction of these elements is active in immune cell types. To test this hypothesis, I retrieved the data from the mouse immune-system *cis*-regulatory atlas which comprises over 500,000 putative CREs identified using ATAC-seq in 86 primary immune cell types [320]. Overlap between the H3K9me3-DA regions and the collection of immune open-chromatin regions (Imm-OCRs) revealed that ~30% of Cluster 1 (n=707) and ~25% of Cluster 2 (n=1542) regions overlapped with the Imm-OCRs (the fraction of overlapping regions was reduced to ~14% when checking non-DA H3K9me3 regions – Figure 2.27a). Next, I sought to determine if the overlapping CREs (n=2249) are active in specific groups of immune cells. I first checked that the mean activity of the Imm-OCRs (intended as chromatin accessibility) correlated well between similar cell types and then merged closely-related cell types to reduce the number of populations. I then sorted the cell types by decreasing mean activity of the Imm-OCRs overlapping with H3K9me3-DA regions (Figure 2.27b). Stromal cells were the cell type in which the overlapping Imm-OCRs are most active, followed by B cells, T cells, macrophages, granulocytes etc. Stromal cells—also known as mesenchymal stem cells (MSCs)—play crucial roles in modulating the immune response. They are multipotent cells deriving directly from embryonic precursors, rather than from the hematopoietic lineage [321]. Altogether, this analysis confirms that the putative CREs activated in H3.3K9A mESCs likely directly regulate the expression of the de-repressed immune-related genes.

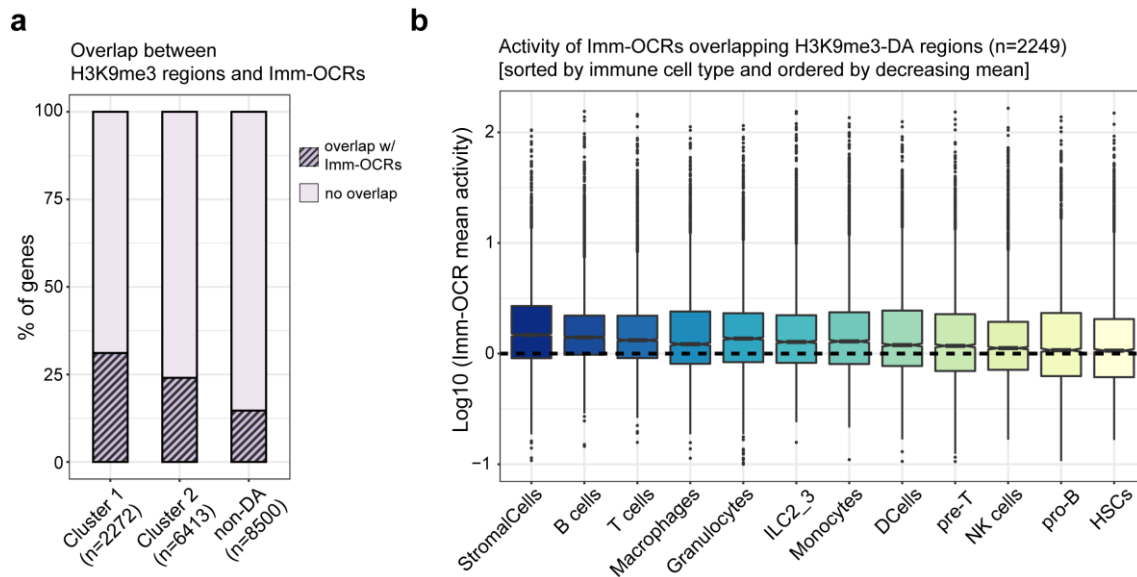


Figure 2.27 Overlap between H3K9me3-DA regions and putative CREs active in immune cells. **(a)** Stacked barplot displaying fraction of H3K9me3-DA regions overlapping with Imm-OCRs (Imm-OCRs were extended to a total size of 2 kb). **(b)** Mean activity of Imm-OCRs overlapping with H3K9me3-DA regions in immune cell populations; cell types are sorted by decreasing mean activity of the open chromatin regions.

2.2.13 Differential nascent transcription analysis verifies increased activity of ERV-derived CREs enriched for motifs of immune-related and developmental TFs

Nascent transcription analysis implied increased levels of active Pol II at the distal H3K9me3-DA regions in H3.3K9A mESCs. Therefore, I further analysed the PRO-seq data—not restricted to the H3K9me3-DA regions previously identified—to verify the findings derived from ChIP-seq and mRNA-seq integrations. I used two packages developed from the Danko Lab that take nascent transcription data as input: i) dREG, for the identification of regulatory DNA sequences [322, 323] and ii) tfTarget, for the identification of *TF-CRE-gene* networks differentially regulated between two conditions [324]. This analysis revealed the presence of 2345 and 596 putative CREs with increased and decreased activity respectively, in H3.3K9A mESCs (Figure 2.28a). Overall, the tfTarget analysis captured mostly changes at promoter regions and at putative enhancers, but the regions more active in H3.3K9A mESCs included heterochromatin territories according to ChromHMM (Figure 2.28b–c). Over 32% of the regions with higher activity in H3.3K9A mESCs had at least one ERV (Figure 2.28d), in support of the findings reported in the previous sections.

A total of 98 TF motifs were returned as significantly enriched at the CREs with increased activity and I could notice a high representation of TFs related to immune cells specification and function (e.g., Bhlhe40; Tcf3; etc.), including interferon response factors

(i.e., Irf) and NF-KB subunits (e.g., Nfkb1/p50; Nfkb2/p52; Rela/p65; Rel). Also, several TFs related to developmental processes (e.g., Tbx2; Tbx3; Sox17; Sox9; Hoxa3; Hoxd3; etc.) and maintenance of pluripotency (e.g., Pou5f1/Oct4; Sox2; Myc) were included (Figure 2.28e).

Therefore, the orthogonal analysis solely based on PRO-seq data and the overlap of H3K9me3-DA regions with Imm-OCRs substantiate that the putative ERV-derived CREs are involved in the regulation of immune-related gene transcription.

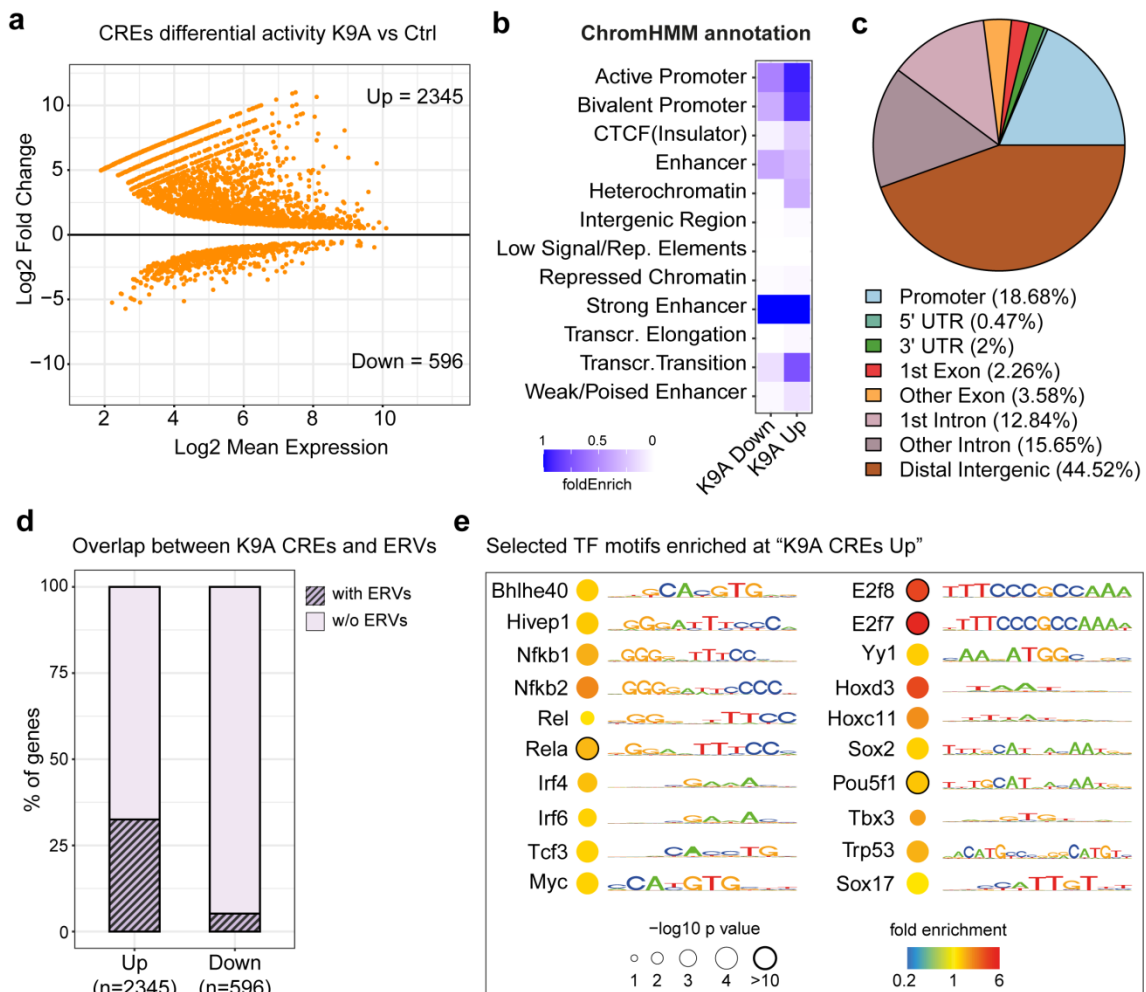


Figure 2.28 PRO-seq tfTarget analysis. (a) MA-plot showing differentially active CREs, according to tfTarget analysis (retained CREs as significant with $\text{padj} < 0.01$). **(b)** ChromHMM annotation of differentially active CREs. **(c)** Genomic annotation of differentially active CREs. **(d)** Stacked barplot indicating the fraction of differentially active CREs with at least one ERV. **(e)** Selected TF motifs significantly enriched at CREs with increased activity in K9A mESCs.

2.2.14 Analysis of differential TF activity reveals NF-KB and other immune-related TFs among the most significantly active in H3.3K9A mESCs

Having identified an increased activation of putative CREs and the TF motifs that are enriched at these genomic regions according to the differential PRO-seq analysis, I wanted to verify whether we could pinpoint specific TFs that have increased or decreased activity in H3.3K9A mESCs. To accomplish this, Daria Bunina performed differential transcription factors activity analysis using the diffTF package [325]. H3K27ac ChIP-seq and mRNA-seq data were used as input for the analysis which allows to compute the differential TF activity between two experimental conditions and classify them as either activators or repressors. A total of 84 TFs were identified as differentially active when comparing K9A and control mESCs ($p_{adj} < 0.001$ – Figure 2.29). Among the most significant TFs with higher activity in H3.3K9A mESCs, I noticed two zing-finger TFs known as HIVEP1 (aka ZEP1, ZNF40A) and HIVEP2 (aka ZEP2, ZNF40B). This two closely related TFs can bind regulatory elements of several viral genes (e.g., HIV-1), as well as enhancers/promoters of host cell immune-related genes (e.g., MHC, interleukin receptors and interferon genes) and hence considered as potential players in T cell activation. Other top hits were three NF-KB subunits (REL/C-REL; RELA/p65; NFKB1/p50), having roles in inflammation and immune response. TAL1, a TF with reported roles in haematopoiesis, was also among the hits with higher activity in H3.3K9A (of note, TAL1 was also identified as relatively enriched at the promoters of the genes selectively upregulated in K9A according to the ChIP-Atlas data integration – Figure 2.17). In addition, among the TFs less active in K9A, I could identify NF2L2/NRF2 that has reported roles in the repression of pro-inflammatory genes. TFs related to NRF2 and important for erythroid differentiation such as NFE2, BACH1 and the MAFK protein partners were also less active in K9A mESCs. This finding is potentially interesting because in hematopoietic differentiation, one of the first ramifications separates the megakaryocyte-erythroid lineage from the myeloid lineage—which gives rise to leukocytes [326]. Therefore, it is tempting to speculate that upon CREs activation in H3.3K9A mESCs, transcriptional programmes are more closely related to immune activity and leukocyte lineages, and TFs involved in erythroid differentiation are found to be less active.

Altogether, the diffTF analysis is concordant with the tfTarget PRO-seq analysis and corroborates the previous findings indicating a global rewiring of regulatory networks in H3.3K9A mESCs resulting in the expression of immune-related genes.

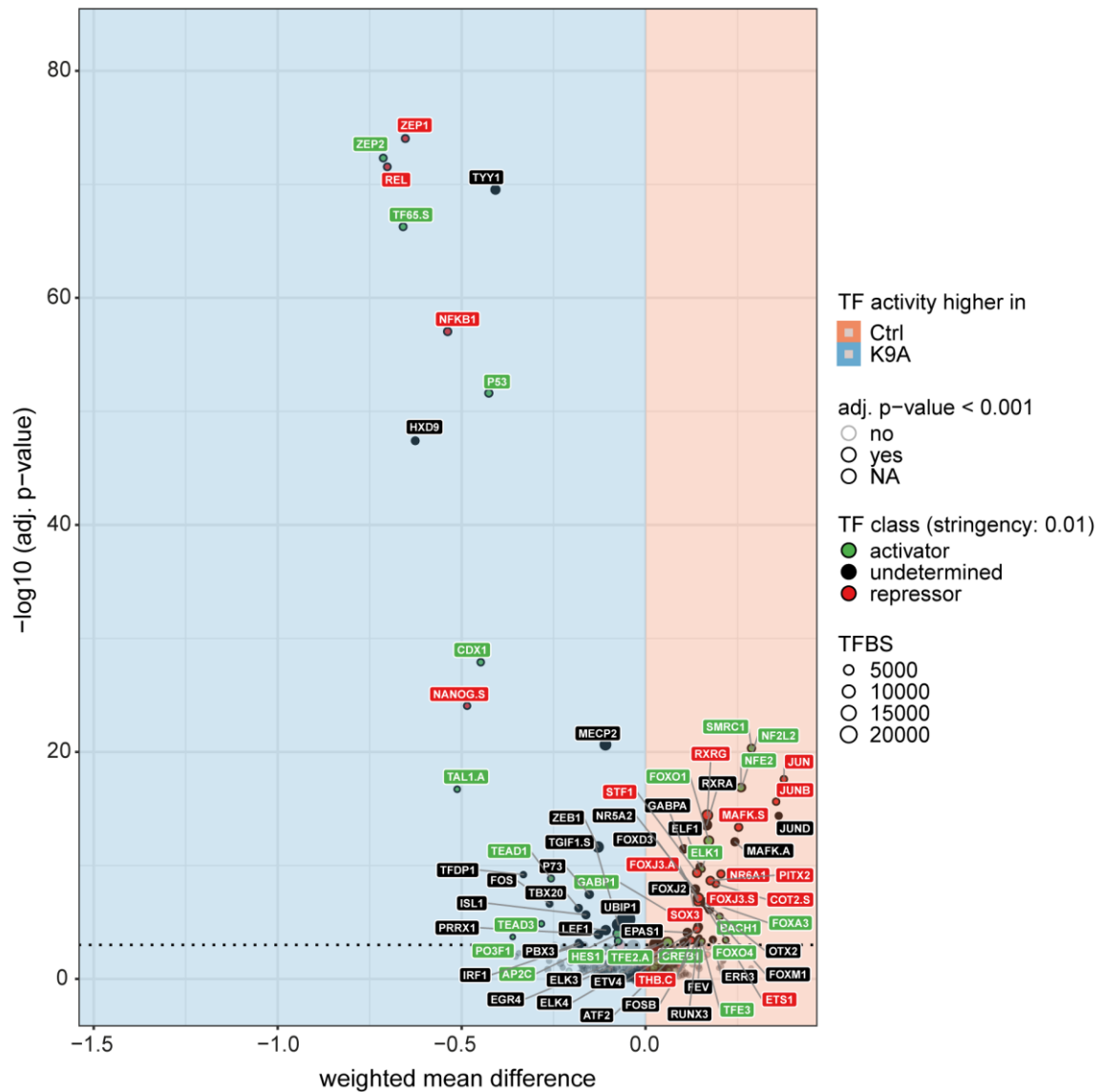


Figure 2.29 diffTF analysis for K9A mESCs. Volcano plot showing TFs with higher activity in K9A (left - blue) or in control (right - pink). TFs are classified as activators (green) or repressors (red). Only expressed TFs are displayed.

2.2.15 H3.3K9A mESCs express immune-related surface proteins

Immune-related genes are generally not expressed in mESCs. Given that H3.3K9A mESCs showed upregulation of many genes involved in immune function, I sought to determine if I could detect such changes at the protein level. I selected three targets among the genes upregulated in H3.3K9A: CD59a (surface glycoprotein involved in inhibition of the complement membrane attack complex and in T-cell activation), TLR2 (surface protein implicated in activation of innate immunity) and MHC Class II (surface proteins typically expressed only by immune cells, dedicated to the presentation of antigens derived from exogenous sources). Nadine Fernandez-Novel Marx and Michael Bonadonna (from the

EMBL Flow Cytometry Core Facility) performed immunolabeling for these markers, followed by flow cytometry analysis. All three protein markers were significantly more expressed selectively in H3.3K9A mESCs (Figure 2.30). As expected, H3.3K27A mESCs—which were also included in the analysis—did not show significantly higher expression of the selected proteins compared to control mESCs. This result highlights that the gene regulatory network rewiring taking place in K9A mESCs not only activates ectopic expression of immune-related genes but also culminates in the production of proteins generally restricted to highly specialized cell types.

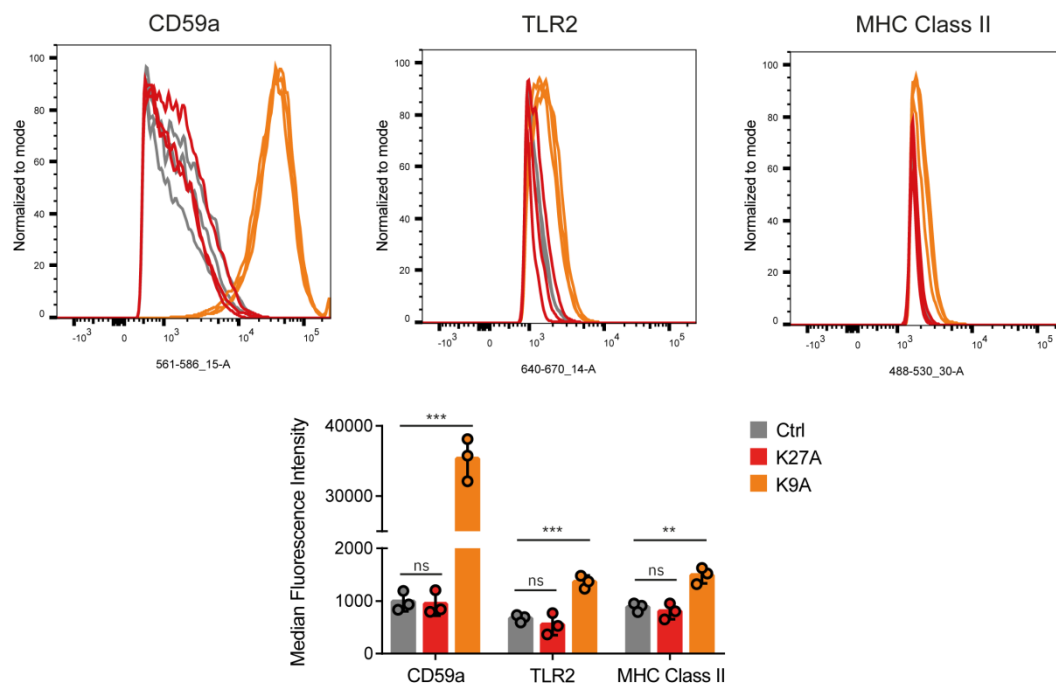


Figure 2.30 Flow cytometry stainings of immune-related marker proteins. Density plots (top) with the signal detected from three biological replicates of control (grey), K27A (red) and K9A (orange) mESCs stained against selected surface markers expressed in immune cell types. Barplots (bottom) of median fluorescence intensity. Significance was calculated with unpaired t-test (**: p-value<0.01; ***: p-value>0.001; ns: p-value>0.05).

2.2.16 H3.3K9A mESCs exhibit reduced levels of lipid peroxidation

To further characterize the phenotypic consequences of the rewired transcriptional programmes in H3.3K9A mESCs, we assessed the levels of reactive oxygen species (ROS). Cellular ROS are produced in physiological conditions, but due to their unstable nature and to the ability to interact with other molecules, several enzymatic and non-enzymatic mechanisms exist to prevent oxidative stress. As presented in section 2.2.3, H3.3K9A mESCs displayed increased cell death (Figure 2.9), and excessive production of ROS can contribute to different types of cell death. For instance, lipid peroxidation occurs when

membrane lipids are damaged by ROS and the accumulation of lipid ROS leads to regulated cell death known as ferroptosis.

On the other hand, increased ROS levels can also trigger NF-KB activation—which I observed in H3.3K9A mESCs. Therefore, to check the presence of oxidative stress in mESCs, Umut Yildiz measured the amount of lipid peroxides using the BODIPY-C11 probe coupled with flow cytometry analysis. To our surprise, H3.3K9A mESCs displayed lower lipid peroxidation levels, effectively halved compared to H3.3K27A or control mESCs (Figure 2.31).

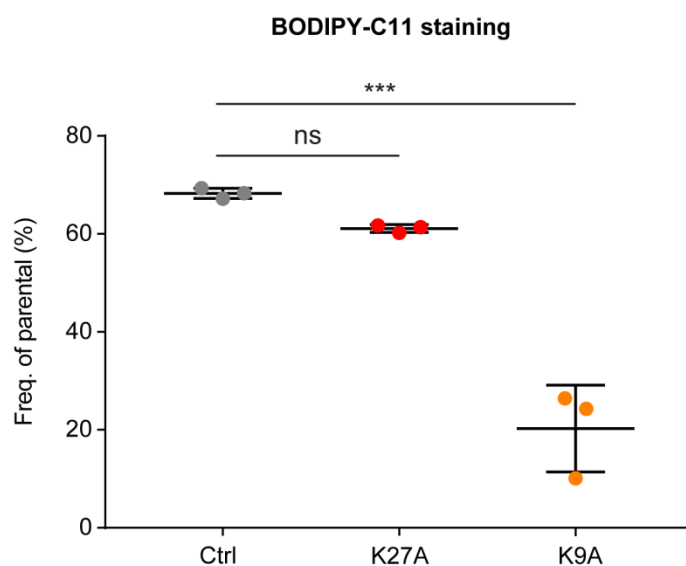


Figure 2.31 Lipid peroxidation flow cytometry staining. Levels of lipid peroxides were measured using BODIPY-C11 staining of mESCs. Significance was calculated with unpaired t-test (***: p-value<0.001; ns: p-value>0.05).

A complete understanding of this finding is still lacking and would require additional experiments, which might be beyond the scope of this study. Nevertheless, it is tempting to speculate the following: firstly, an aspect to consider is that low levels of lipid peroxides might imply lower ROS levels and therefore reinforce the finding that NF-kB activation is likely driven in H3.3K9A mESCs by ERVs de-repression (and not by increased ROS), ultimately resulting in the immune-like shift. The lower level of lipid ROS also fits with the reduced activity of the NRF2 TF detected by diffTF in H3.3K9A mESCs (Figure 2.29), because in the presence of high ROS levels NRF2 is activated and translocates to the nucleus to bind specific antioxidant-response elements (AREs) found in enhancers/promoters of NRF2 target genes.

A second consideration is that the BODIPY-C11 probe used for the staining is a fatty acid analogue. As presented in section 2.2.11, among the genes upregulated in H3.3K9A mESCs—and likely directly regulated by the H3K9me3-DA regions—we find many genes involved in fatty acid metabolism (Figure 2.26). Hence, the decreased lipid peroxidation

detected in K9A might be related to a different metabolic activity resulting in the unusual processing of fatty acids. Aside from these considerations, more experiments would be necessary to fully comprehend this phenotype and why K9A mESCs display decreased lipid peroxidation remains an open question.

2.2.17 Effects of H3.3 K9A and K27A on histone acetylation and nascent transcription at distal CREs

As I observed for specific bivalent promoter regions, the de-regulation of shared groups of genes among K9A and K27A mutants may indicate that similar transcriptional regulatory pathways are affected upon the removal of the K9 and K27 residues from H3.3. Up to now, I focused on the impact of H3.3 residues substitution on repressive chromatin states and highlighted the contribution of H3.3 K9 and K27 residues to silencing mechanisms. However, H3.3 is expected to be enriched at active enhancers, so I investigated whether the absence of these residues impacts enhancers activity. Histone H3K27ac marks active CREs and I looked at the distribution of this modification genome-wide in control and mutant mESC lines. Differential peak analysis revealed that H3K27ac levels were significantly reduced at several genomic regions in both K9A and K27A mESCs (n=18592 and n=17146 down-represented peaks respectively – Figure 2.20a and Figure 2.32a). ChromHMM annotation of the regions with reduced H3K27ac in H3.3 K9 or K27 mutant revealed that loss of this mark occurred primarily at strong enhancers and enhancers regions overall (Figure 2.20b and Figure 2.32b). A fraction of the peaks with reduced signal was then annotated as active promoters and regions close to transcription start sites (“Transcriptional transition”). Of note, in line with the results previously reported, gained H3K27ac peaks in the absence of H3.3K27 were primarily assigned to bivalent promoters and to a minor extent to active promoters.

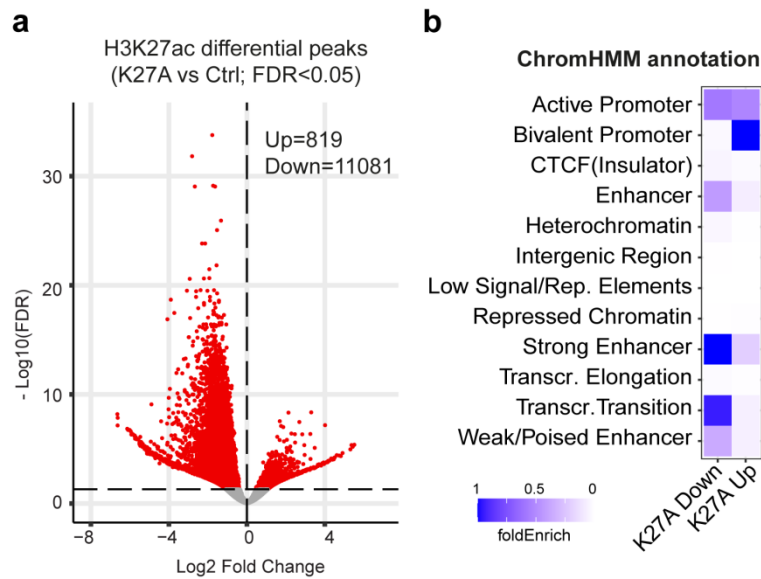


Figure 2.32 H3K27ac ChIP-seq differential analysis in K27A mESCs. (a) Volcano plots showing differential H3K27ac ChIP-seq peaks in K27A mESCs (consensus peak-set with $n=94000$ peaks). **(b)** ChromHMM annotation of differentially abundant H3K27ac regions in K27A mESCs. Reported is the enrichment score normalized over the maximum value among the twelve states.

To better explore the degree of H3K27ac decrease at enhancers and to assess whether reduced levels of histone H3 acetylation correspond to decreased enhancer activity, I used the PRO-seq data and the dREG package to define a set of regulatory regions active in our control mESCs. I excluded promoter-proximal regions from the initial set of regulatory elements and genomic annotations of the regions predicted to be enhancers showed that over 50% were located in distal intergenic territories, while about 30% were in introns (Figure 2.33a). I classified these predicted CREs as “Strong”, “Medium” or “Weak” enhancers based on their PRO-seq signal (upper, middle and lower quartiles, respectively).

H3K27ac signal was primarily reduced at Strong enhancers in both H3.3 K27A and K9A mESCs and also at Medium enhancers, albeit to a lesser extent (Figure 2.33b). This indicates that the absence of the H3.3K9 residue can also impact H3K27ac levels at enhancers. Weak enhancers showed a minor H3K27ac decrease in H3.3K27A (Figure 2.33b and Figure 2.34a). Deposition of the H3.3 histone variant at these elements was not affected by the absence of the K9 or K27 residues, as measured by H3.3 ChIP-seq. Aside from H3K27ac, other histone H3 acetylation marks are found at active enhancers. I monitored possible changes in other acetylation marks by means of H3K9ac ChIP-seq and H3K18ac ChIP-qPCR. In H3.3K27A mESCs, H3K9ac was prominently reduced at Strong enhancers, while Medium and Weak enhancers displayed a less severe yet significant decrease (Figure 2.33b and Figure 2.34a). In H3.3K9A mESCs I observed a slight reduction of H3K9ac at Medium and Weak enhancers.

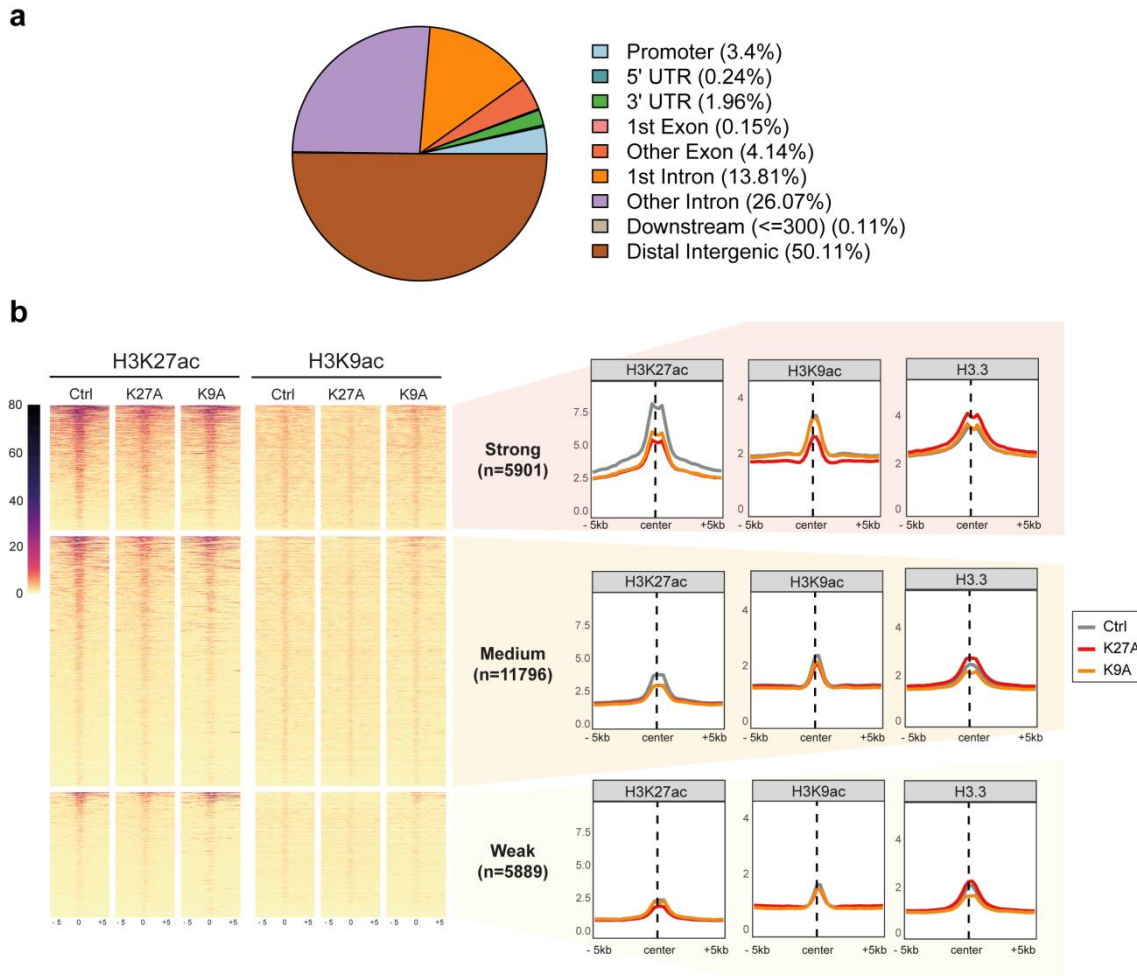


Figure 2.33 Analysis of histone H3 acetylation at putative distal enhancers. (a) Genomic annotation of regions identified as putative CREs in control mESCs, according to dREG analysis of nascent transcription (n=23586). TSS +/- 5kb and CTCF-bound regions were removed from the set of CREs to include only distal and intragenic enhancers. **(b)** Heatmaps of H3K27ac and H3K9ac ChIP-seq signal at dREG enhancers (centre +/- 5 kb). ChIP-seq signal calculated as the average of three replicates for control and mutant mESCs, in 250 bp sliding genomic windows. dREG enhancers are clustered in three groups based on PRO-seq signal in control mESCs ("Weak": first quartile; "Medium": middle; "Strong": third quartile). On the right, metaprofile plots with H3K27ac, H3K9ac and H3.3 average ChIP-seq signals are reported.

H3K18ac showed reduced enrichment at selected Strong enhancers, more consistently in H3.3K27A mESCs (Figure 2.34b). To evaluate whether the reduction in H3 acetylation impacts the activity of the CREs, I compared nascent transcription levels of control and mutant cell lines. At Strong and Medium enhancers, nascent transcription levels were significantly reduced in K27A and K9A (Figure 2.34c). At Weak enhancers, I observed an increased nascent transcription in K27A and reduced in K9A.

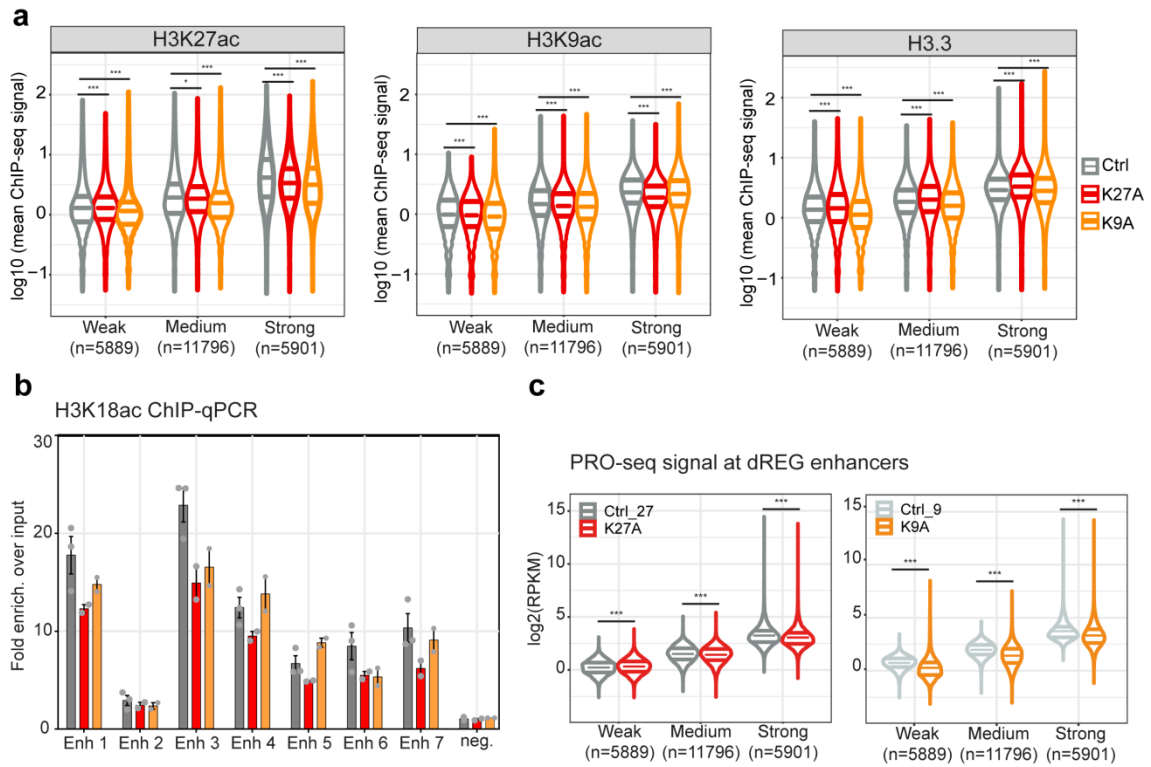


Figure 2.34 H3K18ac ChIP-qPCR and nascent transcription levels at dREG enhancers. (a) Violin plots showing H3K27ac, H3K9ac and H3.3 mean ChIP-seq signal at peak centre +/- 1 kb. Significance was calculated with unpaired t-test (*: p-value<0.05; **: p-value<0.01; ***: p-value<0.001). **(b)** H3K18ac ChIP-qPCR at selected Strong enhancers regions. The relative enrichment over the input is reported and data are mean $\pm$ standard deviation (n=3 replicates Ctrl; n=2 replicates mutants). **(c)** PRO-seq signal at Weak, Medium and Strong dREG enhancers for K27A (left) and K9A (right), with respective controls. Log2(RPKM) values are reported and significance was calculated with unpaired t-test (***: p-value<0.001).

2.2.18 Analysis of expression of genes connected to distal CREs through the GRN

Next, to assess whether changes in enhancers activity reflect changes in gene expression, I used the enhancer-gene connections established with the GRN generated by Daria Bunina. From the entire set of CREs, 934 were included in the GRN (n=568 Strong; n=309 Medium and n=57 Weak). I focused the analysis on this subset to look for possible co-dependency between changes in histone acetylation, enhancer activity, and gene expression. The overall genomic distribution of the CREs retained in the GRN resembled the entire set (Figure 2.35a and Figure 2.33a). H3K27ac showed a significantly decreased signal for Strong, Medium and Weak GRN-enhancers in both K27A and K9A lines (Figure 2.35b). H3K9ac on the other hand displayed opposite trends, with overall reduced signal in K27A and a moderately increased signal in K9A (Figure 2.35b).

Nascent transcription levels were mainly reduced at GRN-Strong enhancers in both K27A and K9A, while GRN-Medium enhancers showed a significantly decreased PRO-seq signal only in K27A (Figure 2.35c). GRN-Weak enhancers showed increased activity in K9A and

no significant change in K27A. Regarding the potential target genes, I could identify from the GRN 639, 391, and 83 DEGs connected to Strong, Medium and Weak enhancers, respectively (average peak-gene distance $\approx$ 111 kb). In line with the reduction in H3 acetylation and enhancer activity, genes connected to Strong enhancers were on average downregulated (Figure 2.35d). Genes connected to Medium enhancers also showed a trend towards downregulation which was more robust for K27A than for K9A, hence again correlating with the differences in enhancer activity. Although the number of genes connected to Weak enhancers through the GRN was rather small, K9A displayed upregulation, and K27A showed no changes.

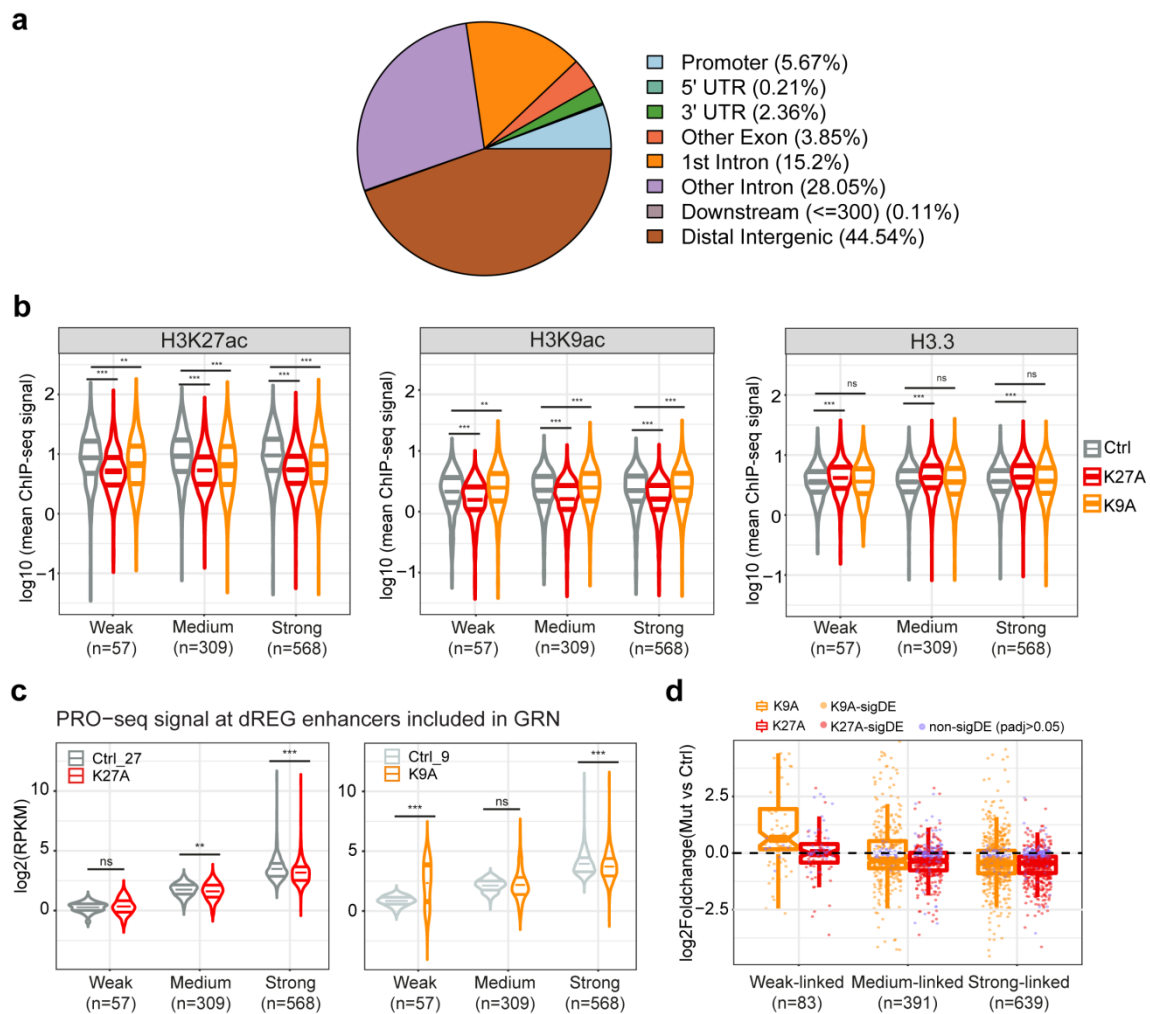


Figure 2.35 Histone modifications and nascent transcription at dREG enhancers retained in the GRN and expression of connected genes. (a) Genomic annotation of dREG enhancers retained in the GRN (n=934). **(b)** Violin plots showing H3K27ac, H3K9ac and H3.3 mean ChIP-seq signal at peak centre $\pm$ 1 kb of enhancers retained in the GRN. Significance was calculated with unpaired t-test (*: p-value<0.05; **: p-value<0.01; ***: p-value<0.001, ns: p-value>0.05). **(c)** PRO-seq signal at Weak, Medium and Strong dREG enhancers included in the GRN for K27A (left) and K9A (right), with respective controls. Log₂(RPKM) values are reported, and significance was calculated with unpaired t-test (**: p-value<0.01; ***: p-value<0.001; ns: p-value>0.05). **(d)** Expression of DEGs connected to dREG enhancers through the GRN; log₂(fold-change) values calculated with DESeq2 are plotted, and individual genes (data points) are coloured to indicate if the differential expression is significant (orange/red) or not (purple).

		Log2FC expression GRN dREG enhancers		
		Weak	Medium	Strong
K9A	Median	0.63	−0.33	−0.44
	Mean	1.37	0.08	−0.35
	95% CI	0.82, 1.82	−0.09, 0.25	−0.49, −0.44
K27A	Median	0	−0.35	−0.44
	Mean	0	−0.39	−0.55
	95% CI	−0.24, 0.24	−0.48, −0.31	−0.61, −0.49

Table 1 Median, mean and confidence interval of log2(fold-change) values from Figure 2.35d

Taken together, these analyses revealed that in the absence of H3.3K27, the levels of histone H3 acetylation are reduced at other lysine residues. I could indeed detect decreased H3K27ac and H3K9ac, as well as H3K18ac, at selected loci. When analysing all the CREs, this reduction was more prominent at Strong CREs. These elements also displayed a more pronounced decrease in the PRO-seq signal. The analysis restricted to the set of CREs included in the GRN confirmed the robust decrease in H3K27 and H3K9 acetylation in H3.3K27A mESCs, accompanied by a reduction in the activity of Medium and Strong putative regulatory elements. The genes connected to these distal elements were downregulated in H3.3K27A mESCs, suggesting that the reduced activity of the CREs might directly impact the expression of potential target genes.

In H3.3K9A mESCs, histone H3 acetylation was also affected, with H3K27ac primarily reduced at Strong and Medium enhancers. H3K9ac was less impacted in the absence of H3.3K9, and I observed the same tendency for H3K18ac enrichment at the selected loci. This general trend was also conserved when focusing the analysis on the CREs included in the GRN—for which H3K9ac was more affected in K27A than K9A. Nevertheless, nascent transcription was reduced at CREs in H3.3K9A mESCs. Strong CREs-linked genes were on average downregulated. Medium CREs-linked genes, on the other hand, did not exhibit the same degree of downregulation, in line with a non-significant difference in the activity of the CREs.

3. Discussion and outlook

Histone modifications correlate with defined transcriptional states, yet the exact contribution of these covalent modifications in the regulation of gene expression remains elusive. Histone mutagenesis has proven to be a valuable approach to identifying the role of histone modifications in yeast and *Drosophila* [189, 198]. As histone-modifying enzymes can target several residues and non-histone proteins, substituting a histone residue represents a more elegant approach to tackle this question. Nevertheless, applying histone mutagenesis is challenging in most metazoans because canonical histone genes have a peculiar genomic organization whereby multiple genes are found in large clusters, often located on different chromosomes. In my PhD project, I exploited the fact that histone variants—such as H3.3—are encoded by fewer genes and can therefore be more easily subjected to mutagenesis. By targeting H3.3, I primarily aimed to characterize the contribution of this histone variant in maintaining repressive chromatin states. In particular, I mutated H3.3 K9 and K27 residues to study how the removal of such residues affects histone modifications globally and locally in mESCs, and how this relates to variations in transcription.

3.1 Single H3.3 amino-acid substitutions at K27 and K9 residues have a large impact on transcription in mESCs

Histone H3.3 represents only about 20-30% of the total H3 pool in proliferating cells, yet the introduction of either H3.3 K27A or K9A mutations resulted in wide gene expression changes. The impact of H3.3 residue substitutions is more significant than the effect of complete H3.3 knockouts [165, 231]. This might be because in the absence of H3.3, proliferating cells can compensate for this absence with canonical wild-type H3. On the other hand, when H3.3 is mutated, it can still participate in chromatin-templated processes. In the case of both K27A and K9A mutations, I did not detect defects in H3.3 deposition. This likely indicates that the two targeted residues are not primarily involved in the binding of chaperone proteins and/or chromatin remodelling complexes (as it was for example reported for lysine 4, in the case of the H3.3K4A mutation [257, 258]). Having assessed that the mutant H3.3 is deposited correctly into chromatin at specific genomic locations, allows to evaluate the contribution of H3.3 K9 and K27 residues and the respective modifications. For both mutants, the changes detected in gene expression are likely a result of altered chromatin states.

3.2 H3.3 K27A mutant discloses a prominent role of H3.3 in maintaining promoter H3K27me3 levels and repression of certain Polycomb-target genes in mESCs

Histone H3K27me3 is a histone modification found at the promoters of repressed genes. A study conducted in *Drosophila* reported that H3.3K27 critically participates in Polycomb-target gene silencing [263]. Whether H3.3 fulfils a similar function in mammals is, however, still not clear. A previous paper indicated that H3.3 could be deposited by the chaperone HIRA at Polycomb-target promoters in mESCs [165]. The authors observed reduced H3K27me3 signal at bivalent promoters upon depletion of H3.3, and a small subset of those genes was upregulated. Interestingly, the repression of the affected genes was restored with overexpression of H3.3K27A/R mutants in H3.3 depleted mESCs. Given that H3.3K27me3 was not necessary to restore the repression, the interpretation was that H3.3 is not the primary substrate of PRC2 but is instead the HIRA-mediated H3.3 deposition that contributes to creating a dynamic chromatin environment facilitating PRC2 recruitment and/or activity [165]. H3.3 is not enriched in H3K27me3, and also the recent paper from the Helin Laboratory in which all canonical copies of histone H3 were edited to carry a K27R mutation, displayed an almost complete loss in H3K27me3 [200]. Nevertheless, the differentiation phenotype observed in the study was exacerbated when—on top of the canonical histone H3—also H3.3K27 was mutated, indicating a crucial contribution of the H3.3 residue. Another recent paper, adopted a strategy similar to ours and generated H3.3K27R mutant mESCs [262]. The authors reported a decrease in H3K27me3 at bivalent promoters, and ~30% of the genes significantly differentially expressed were upregulated (a fraction similar to the one observed in our H3.3K27A mESCs). However, this result was not further commented on or investigated.

The mESCs with endogenous H3.3K27A mutation generated in my PhD project exhibited upregulation of a substantial number of developmental genes with bivalent promoters. H3K27me3 was also generally decreased at Polycomb-target genes. This indicates that even though H3.3K27 is not the primary PRC2 substrate, the presence of the H3.3K27 residue is necessary to guarantee proper deposition of the tri-methylation mark at promoters. A plausible explanation for the H3K27me3 decrease is the mechanism of PRC2 allosteric activation. The H3.3 variant is deposited at bivalent promoters, and in the absence of the H3.3K27 residue the levels of H3K27me3 would be consequently reduced. This local reduction in H3K27me3—or the continuous incorporation of a histone devoid of the PRC2 substrate residue at promoters—would result in a less effective allosteric activation of PRC2 (Figure 3.1). In line with this model, K27A mESCs showed decreased

SUZ12 binding at bivalent promoters indicating that Polycomb recruitment is affected. Accordingly, the spreading of the H3K27me3 mark was affected in K27A mESCs. As reported in other studies, the reduction of H3K27me3 at promoters doesn't always alter the expression of the genes. This can be due to several possible explanations, such as the absence of required transcription factors or the presence of alternative repressive mechanisms—for instance, DNA methylation or histone methylation of other residues. Although H3K27me3 was generally reduced at bivalent promoters, I could notice that the promoters of the genes upregulated in K27A mESCs have, on average, a higher basal level of both H3K27me3 and H3.3. Thus, the specific subset of bivalent genes upregulated in K27A mESCs relies more on H3K27me3 to maintain low levels of transcription, and—having more H3.3 incorporated at their promoters—they are also more susceptible to the absence of the H3.3K27 residue.

Notably, these promoters having more H3K27me3 and H3.3 in the basal state is also in line with the *in-vitro* HMT assay performed by the collaborators in the Davidovich Lab, which essentially showed that H3.3 (with unmodified lysine 9) is a better substrate for PRC2.

To reinforce this finding, it would be compelling to perform a rescue experiment in which wild-type H3.1 or H3.3 are ectopically expressed in H3.3K27A mESCs. This experiment could tell whether H3.3 is effectively preferentially incorporated at the developmental genes upregulated in the mutant line.

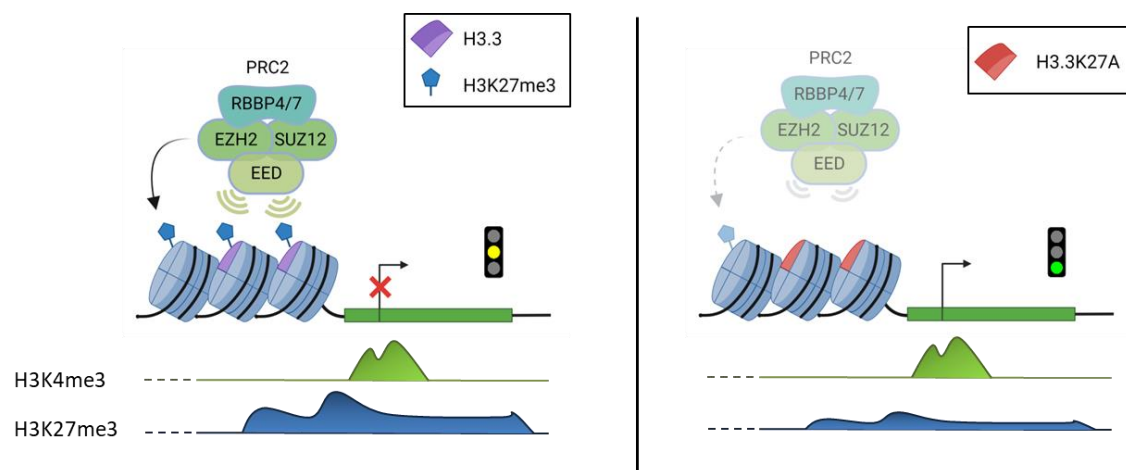


Figure 3.1 H3.3K27A effect at bivalent promoters. In the basal state (left) H3.3 contributes with the lysine 27 residue to the maintenance of the proper repressive environment. In the presence of the K27A mutation (right), levels of H3K27me3 are decreased locally on promoters, impacting PRC2 allosteric activity and spreading of the modification. Figure created with BioRender.com.

Among the developmental genes more prominently affected in K27A mESCs, I could detect several involved in ectoderm and mesoderm lineage differentiation. The premature upregulation of neuro-ectoderm marker genes in mESCs did not strongly impact *in-vitro* neuronal differentiation. On the contrary, I could identify a phenotype upon induction of *in-vitro* mesoderm-cardiac differentiation, whereby the K27A mutants formed smaller EBs and were unable to maintain the functional cardiomyocyte differentiated state. This observation is partially in line with a recent preprint in which the authors suggested that proper H3K27me3 deposition could be important for preserving the differentiated state, rather than for initiating the differentiation process [327]. To verify this, it would be important to perform additional experiments. Differentiation is generally irreversible as cells cannot spontaneously revert to a pluripotent state. In this regard, an interesting experiment would be to check whether mesoderm-cardiac cells derived from K27A cells can revert to a more pluripotent-resembling state (e.g., high alkaline phosphatase activity and expression of pluripotency markers) without receiving any external stimulus.

Taken together, H3.3K27 is essential for establishing the correct H3K27me3 landscape in mESCs. The reduction in H3K27me3 levels results in the de-repression of specific developmentally-regulated genes that have more H3.3 at their promoters and premature activation of mesoderm lineage markers impacts the differentiation potential of the mESCs.

3.3 H3.3K9 safeguards repression of ERVs and distal *cis*-regulatory elements in mESCs

Histone H3K9me3 is a repressive mark found at several genomic locations, such as distal intergenic regions and repeat elements. I sought to determine the contribution of H3.3K9 in establishing this repressive modification in mESCs. Mass-spectrometry studies have shown that in eukaryotes K9 di-methylation is relatively more enriched on the canonical H3 variants [226, 227]. One study in particular, reported that the K9 residue is the only lysine found to be methylated on non-nucleosomal histones and that non-nucleosomal H3.1 is relatively more enriched in mono-methylation when compared to H3.3 [229]; this modification state can influence the differences in the H3 variants-specific levels of K9 modifications following deposition into chromatin [229]. Histone mutagenesis studies conducted in *Drosophila* showed that K9R mutation on canonical H3 leads to a significant loss of HP1 binding at pericentromeric heterochromatin, while the euchromatic H3K9me3 signal is less impacted [328]. Consistently with this, not many protein-coding genes were affected in this experimental setup, but several transposons on the other hand were de-repressed. In this regard, comparing H3K9R and H3.3K9R mutant flies helped to disclose a

certain degree of redundancy between canonical H3 and H3.3 K9 residues. Flies carrying both mutations were not viable, and the changes in the expression of genes and transposons were much more pronounced when all H3 variants were mutated [329].

In mammals, H3.3 is also located at telomeres [214], and H3.3 knockout studies in mESCs reported that at telomeric regions H3.3 crucially contributes to establishing the proper levels of H3K9me3 [215, 230]. Consistently with this, I could confirm that in H3.3K9A mESCs the H3K9me3 signal was reduced at telomeric regions (Figure 2.24b). Such decrease is not expected to have a consequence per se in mESCs [230] and indeed H3.3K9A cells did not show defects in the cell cycle (Figure 2.8). Histone H3.3 also contributes to the properly establishing H3K9me3 at specific repeat elements, especially Class I and II ERVs [215]. The study that investigated this possible role also employed a H3.3 knockout approach in mESCs, which resulted in de-repression of such ERVs [215]. However, the contribution of H3.3 to ERVs silencing has been debated, as previously summarized in the section “1.7.2 H3.3 and ERVs repression”. Knockout studies are beneficial, yet histone mutagenesis represents a more refined approach that can help to clarify the contribution of the H3.3K9 residue in regulatory mechanisms.

In the present study, I generated mESCs bearing K9A mutation on the endogenous H3.3 histone variant. Upon loss of the H3.3K9 residue, mESCs displayed broad transcriptional changes that imply an extensive rewiring of gene regulatory networks. In particular, I focused on the upregulated genes in H3.3K9A mESCs to determine whether increased expression of these genes could be directly linked to the remodelling of repressive chromatin states. Among the upregulated genes, I noted two major ontology categories, such as differentiation and immune system processes, whereby several genes exhibited a bivalent promoter and—consistently with the increased expression level—H3K27me3 was reduced at these loci. The decrease in H3K27me3 at bivalent promoters in the absence of the H3.3K9 residue can be explained by i) direct involvement of H3.3K9 in maintaining local levels of H3K27me3 or ii) indirect effect due to a coerced upregulation of genes. Regarding the first hypothesis, two studies reported that H3K9 di-/tri-methylation could be located at mESCs promoters to assist and reinforce PRC2-mediated silencing [110, 164]. However, the levels of these modifications at the promoters of upregulated genes targeted by Polycomb were very low in this case. In general, I could not detect changes in H3K9me2/3 that could explain the H3K27me3 reduction. An *in-vitro* methyltransferase assay performed by collaborators from the Chen Davidovich Laboratory at EMBL Australia and Monash University, further confirmed that di-/tri-methylation of the lysine 9 residue of H3.1/H3.3 does not increase PRC2 activity *in-vitro*.

On the other hand, the *in-vitro* experiment showed for the first time that histone H3.3 with unmodified lysine 9 is the optimal PRC2 substrate. This observation is in line with the rest of the genomic data because, among the bivalent genes, the ones with more H3K27me3 signal on their promoters showed a higher level of promoter H3.3 signal (while the H3K9me2/3 signal was rather low). Also, according to the histone modifications mass-spectrometry data generated by collaborators from the Benjamin Garcia Laboratory at Pennsylvania University, global K27 methylation levels within H3 and H3.3 variant were reduced in H3.3K9A mESCs.

If H3.3K9 is a better substrate for PRC2, removal of the K9 residue might have a direct local effect on Polycomb activity. Histone H3.3 deposition is not affected at Polycomb target genes, so the presence of H3.3K9A at these loci might partially explain the reduction in the H3K27me3 signal, but it is unlikely to be the sole cause of the observed decrease. The decrease in promoter H3K27me3 could result from the aberrant activation of distal regulatory elements. To investigate this possibility, I analysed whether H3K9me3 was affected at distal intergenic territories in H3.3K9A mESCs. Several H3K9me3 broad regions were impacted by the H3.3K9A mutation. About 25% of these regions were particularly affected, and the H3K9me3 mark was practically erased. In the remaining regions, the signal was reduced less dramatically but still to a considerable extent. The reduction in the H3K9me3 mark was accompanied by an increase in histone H3 acetylation and nascent transcription signal, indicating that these regions acquired in H3.3K9A mESCs a chromatin state typical of active *cis*-regulatory elements.

The construction of a gene regulatory network from the transcriptomic and epigenomic data highlighted that the genes upregulated in H3.3K9A mESCs are putative targets of the CREs activated upon loss of H3K9me3. In particular, several genes were related to immune system processes. Also, ~20% of the genes upregulated in H3.3K9A mESCs and connected to the H3K9me3-DA regions presented a bivalent promoter. Of note, the bivalent genes upregulated only in H3.3K9A were not enriched for ontology terms related to general processes of embryonic development, but were rather associated with more specialized pathways such as myeloid leukocyte differentiation. This is to some extent in line with previous reports which showed the involvement of Polycomb in the regulation of immune-related genes in specialized cell types [330, 331]. It would therefore make sense that in mESCs, the repression of such genes would be secured by two converging regulatory mechanisms: one local, represented by the PRC2-mediated deposition of H3K27me3 on promoters, and one distal, via repression of CREs through H3K9me3. The fact that the level of H3K27me3 was lower on the promoter of the H3.3K9A-specific upregulated genes

also implies that the repression through silencing of distal putative CREs represents the prevailing regulatory mechanism (Figure 3.2).

The upregulation of genes associated with immune function in mESCs is unusual, yet a reasonable explanation for the expression of such specific genes can be linked to the regulatory potential of ERVs. I observed that ERVs were de-repressed only in H3.3K9A mESCs. H3K9me3 ChIP-qPCR confirmed that this modification was reduced at IAP elements. Given that decreasing H3K9me3 by means of histone mutagenesis is sufficient to unleash ERVs expression, this result highlights how repressive histone modifications represent the major silencing mechanism for retroelements in mESCs, and that DNA methylation is not of primary importance in this context [274]. As the de-repression is achieved in this study solely by removing the lysine 9 residue of the H3.3 histone variant—and not by acting on more broad players such as SETDB1 or KAP1—I could argue that H3.3 is indeed a key element required for silencing of ERVs. Global levels of H3K9me2/3 were not impacted in H3.3K9A mESCs (Figure 2.6b). H3.3 itself likely contributes to the H3K9me3 build-up at distinct regions, and the decrease is not just initiated by a local decrease and secondly aggravated by an impaired spreading of the modification on adjacent canonical H3 (as opposed to what I envisioned for the H3.3K27A effect at promoters). In support of this hypothesis, H3K9me3-reader proteins (e.g., HP1, CDYL, CBX proteins etc.) were previously shown to be dispensable for silencing of retroelements [114].

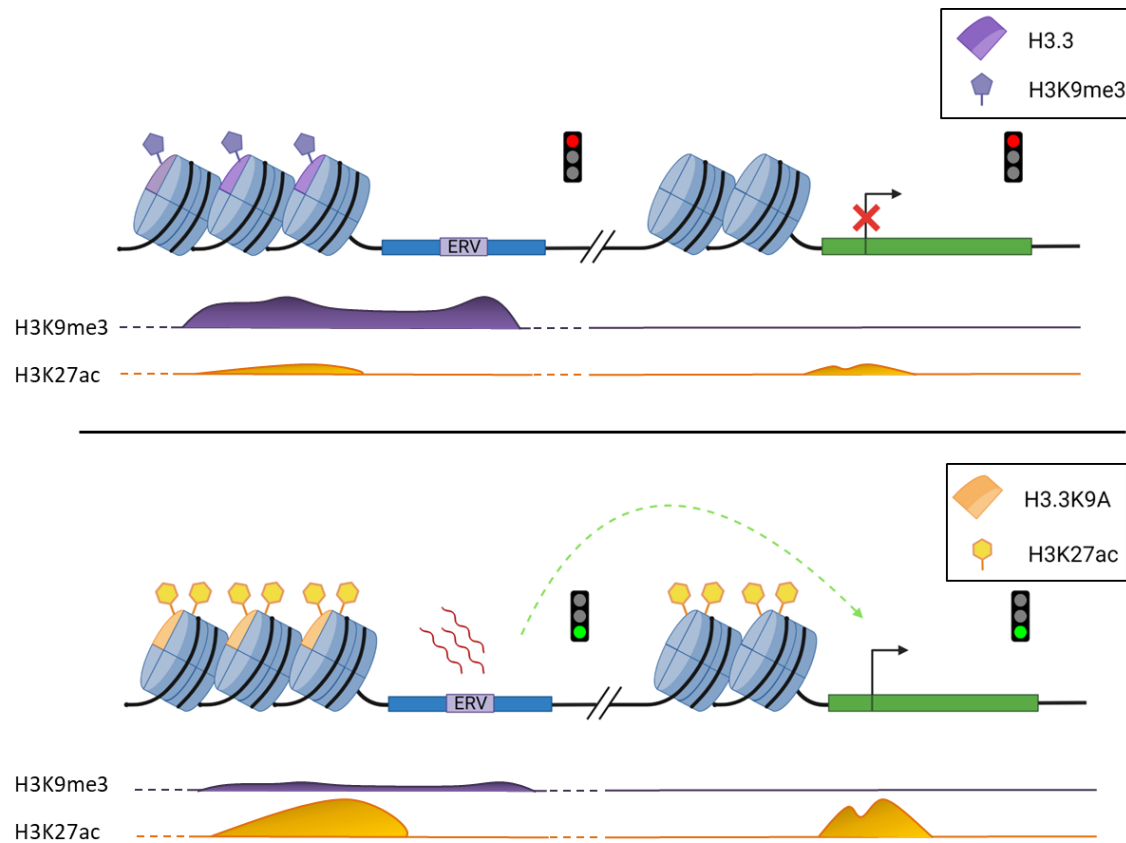


Figure 3.2 H3.3K9A impact on distal ERV-derived regulatory elements. In the basal state (top), distal genomic regions with regulatory potential are repressed by H3.3K9me3. When the K9A mutation is introduced (bottom), these regions acquire an active chromatin state (histone acetylation + nascent transcription) and can drive the expression of putative target genes. Figure created with BioRender.com.

The GRN demonstrated that the ERV-derived distal CREs activated in H3.3K9A could potentially drive the upregulation of the genes exclusively expressed in this mutant. Nevertheless, there is another possibility that I must consider. When ERVs are de-repressed, they may produce double-stranded RNA sensed by the host cell [303, 304]. This triggers an interferon-mediated antiviral response, similar to the innate immune response taking place in case of infections from exogenous viruses. This process is known as viral mimicry and in recent years it has been considered as a potential way to treat cancer [332, 333]. For instance, DNA methyltransferase inhibitors (DNMTi) can lead to ERVs activation and to the production of dsRNA in cancer cells. The ensuing interferon response leads to the production of cytokines and tumor-related neoantigens that can be presented on the surface of tumor cells, ultimately eliciting a cytotoxic T cell response [333].

From a molecular point of view, activation of the interferon pathway results in the translocation of interferon regulatory factors (e.g., IRF-3/7) and NF- κ B to the nucleus, leading to the expression of interferon-stimulated genes [334]. Because of the higher activity of NF- κ B TFs observed in H3.3K9A mESCs (according to diffTF and tfTarget analyses), it is reasonable to think that an interferon response might take place in this

mutant, which leads to two main considerations. Firstly, activation of the interferon response as an effect of ERVs de-repression in the H3.3K9A mutant may dispel several doubts on the relevance of H3.3 in ERVs silencing. One of the main objections so far was that H3.3 KO or depletion of factors involved in H3.3 deposition generally resulted in a mild upregulation of ERVs [280, 288]. Nevertheless, if the ERVs de-repression in H3.3K9A mESCs is sufficient to trigger an interferon response, this objection could be overcome. The second aspect concerns the regulatory effect of the ERVs, which I described before. Activation of the interferon response from transcribed repeats could provide another means for the upregulation of specific genes in H3.3K9A mESCs, potentially independent from the activity of distal CREs. This could, in part, explain the activation of some immune-related genes. Nevertheless, the fact that also other genes—such as the one with bivalent promoters associated with different lineages—are also upregulated, supports the hypothesis that a direct transcriptional regulation mediated by ectopically activated CREs is likely to take place in H3.3K9A mESCs. For instance, a previous study reported upregulation of specific bivalent genes in KAP1/TRIM28 knockout mESCs as a consequence of ERVs activation [297]. These findings are to some extent concordant with the results I presented here. As one possibility doesn't rule out the other, it is plausible to think that both mechanisms of viral mimicry and direct transcriptional regulation through ectopically activated distal CREs take place in H3.3K9A mESCs (Figure 3.3).

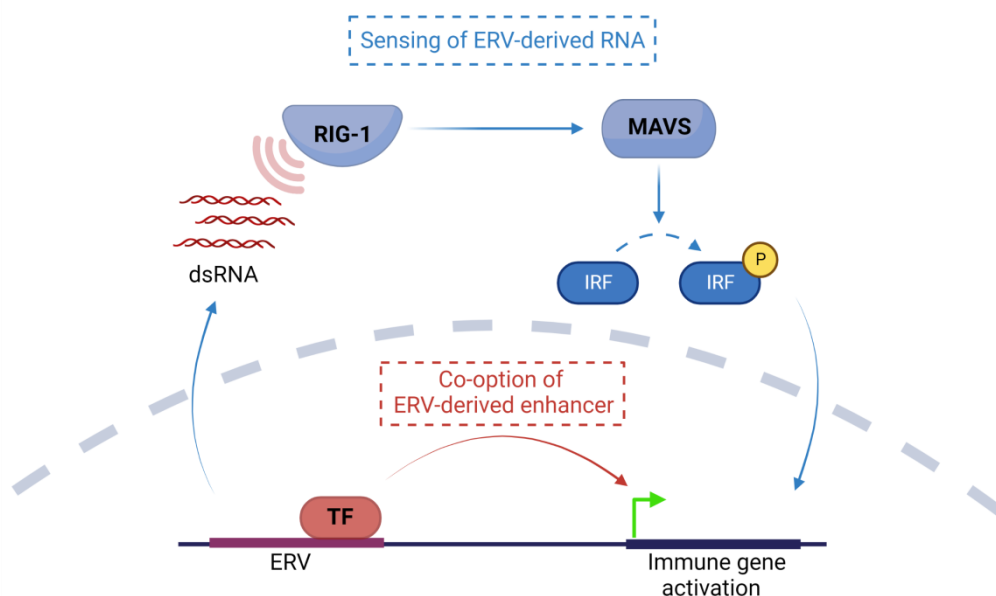


Figure 3.3 Impact of ERVs on transcription. Schematic representation of the viral mimicry mechanism which can lead to the expression of innate immune genes, upon translocation of IRF or NF- κ B TFs. Alternatively, ERVs can directly recruit TFs and act as cis-regulatory elements. Figure created with BioRender.com.

A compelling experiment to perform would be to treat the H3.3K9A mESCs with a recently reported antagonist of the pattern recognition receptor RIG-1 [335]. This antagonist is expected to prevent dsRNA sensing and inhibit the interferon response. Upon treatment, analysis of expression of genes identified from the GRN as putative direct targets of the distal CREs, coupled with analysis of histone modifications (e.g., H3K27ac ChIP-qPCR at distal CREs and genes promoters) could tell whether upregulation of these genes is a consequence of direct transcriptional regulation. There are also additional experiments that in my opinion could further corroborate this hypothesis. For instance, despite the solid evidence already provided by the GRN of the potential enhancers-promoters pairs, targeted (e.g., 3c-qPCR) or genome-wide (e.g., H3K27ac HiChIP) conformation capture experiments would demonstrate direct transcriptional regulation exerted by the H3.3-safeguarded distal CREs. Lastly, CRISPR-inhibition (CRISPRi) and -activation (CRISPRa) approaches would allow the selective switching-on (in control mESCs) or switching-off (in K9A mESCs) of specific H3.3-safeguarded CREs and analysis of the putative target genes expression could verify if the transcriptional state would change accordingly with the on/off state of the distal CREs.

Concerning viral mimicry, one debated aspect of exploiting this mechanism in cancer treatment is attributed to the fact that by using so-called “broad-brush approaches” such as DNMTis, several other protein-coding and tumor suppressor genes could also be reactivated [333]. For this reason, it would be very useful to pinpoint specific ERVs with immunogenic potential that could be selectively targeted for activation (e.g., by means of CRISPRa) to trigger an interferon response in tumors. Given that H3.3 is expected to affect a limited number of ERVs (at least compared to KAP1, STEDB1 or DNA methyltransferase depletion), the H3.3K9 mutagenesis approach could be a good starting point to screen for immunogenic ERVs. In the first phase, H3.3K9A/R mutant human stem cells or cancer cells could be generated to narrow down to potential candidate ERVs. In a second phase, a CRISPRa screening could allow for verifying the immunogenic potential of individual retroelements in a high-throughput fashion, for instance, by means of FACS analysis for interferon-related markers.

3.4 H3.3 K27A and K9A affect (likely in different ways) activity of canonical distal *cis*-regulatory elements and expression of putative target genes

Although I focused mainly on investigating the contribution of H3.3 to repressive mechanisms, mutation of the H3.3 K9 and K27 residues also affects deposition of acetylation at these sites. In the last few years, the role of histone H3 acetylation in gene expression regulation has been, to some extent, reconsidered [167]. The studies that looked at the contribution of H3.3K9 and K27 residues in *Drosophila* concluded that acetylation of these residues is not strictly necessary for maintaining gene expression [263, 329]. The same conclusion was also derived from the H3.3K27R in mESCs, in which the authors focused on the relevance of this mark at enhancers and concluded that H3K27ac is not needed to sustain enhancer activity [262]. These results partially contrast with other studies that employed inhibition of HATs enzymes—like p300/CBP—and showed that H3K27ac is required for enhancer activity [177]. One possibility is that the overall histone acetylation state, rather than the acetylation of a specific residue, might be relevant for maintaining transcriptional activity.

In H3.3K27A mESCs, I could observe that in the absence of the lysine 27 residue, H3K27ac as well as acetylation at other sites, such as K9 and K18, were affected. The distal CREs at which the acetylation decrease was more prominent exhibited reduced activity (as measured by means of nascent transcription), and the genes connected through the GRN were downregulated. This result contrasts with the findings reported in the H3.3K27R mESCs paper, despite the experimental approach adopted being almost the same. In this study, the authors could not detect changes in the acetylation of lysine residues other than lysine 27. This could explain why gene expression changes did not correlate with changes in enhancer acetylation. One critical difference between the cell lines generated in my project and the ones from the previous study is the type of amino acid substitution introduced. The lysine-to-arginine mutation allows to retain of a positive charge that is also present on unmodified lysine. An often-underrated aspect is that histone residues themselves—in their unmodified state—can be involved in critical interactions with chromatin factors [142, 257]. Although more experiments would be necessary to corroborate this hypothesis, it is tempting to speculate that the lysine-to-alanine mutation might disrupt protein-protein interactions with chromatin factors (e.g., p300/CBP) that are essential to maintain an active chromatin environment at enhancers (Figure 3.4). A chromatin immunoprecipitation experiment coupled with mass-spectrometry (ChIP-MS) could help to determine if specific interactions are disrupted in the absence of the H3.3K27 residue. For this purpose, I generated mESCs lines in which the H3.3 protein is

fused to a SNAP-tag and the aim of this experiment would be to compare how the H3.3 protein interactome is impacted in the presence of wild-type, K27A and K27R H3.3 histone.

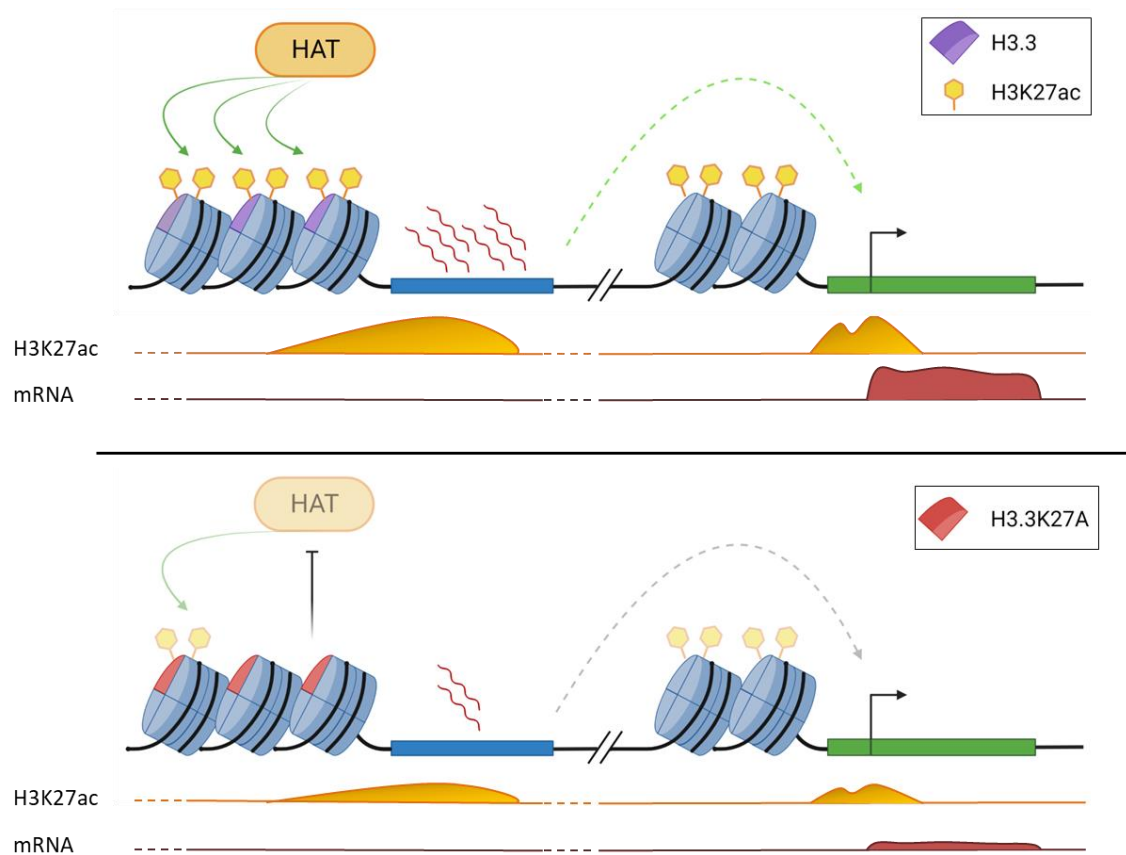


Figure 3.4 H3.3K27A impact on canonical enhancers. In the basal state (top) H3.3 is deposited at active distal regulatory elements, but the introduction of the K27A mutation (bottom) results in reduced levels of acetylation at several histone H3 residues and decreased nascent transcription levels. One hypothesis is that the H3.3K27 residue is important to maintain an active chromatin environment, perhaps by being directly involved in protein-protein interactions. Figure created with BioRender.com.

Regarding H3.3K9A mESCs, at enhancers, the H3K27ac signal was primarily affected while the H3K9ac was less impacted overall. This was also the case for the CREs included in the GRN, for which potential target genes were identified. Mostly the activity of the Strong-CREs in the GRN was reduced. The expression of the genes potentially regulated by these elements was concordant, and the ones connected to the more active enhancers were, in general, more prominently downregulated. Of note, despite the changes in gene expression being—in terms of magnitude—overall more dramatic in H3.3K9A mESCs (Figure 2.4b), for the set of genes linked to distal CREs through the GRN, the degree of downregulation was on average more prominent in H3.3K27A mESCs (Table 1). The decrease in H3K27ac and activity of the CREs might be an indirect consequence of the deep rewiring of gene regulatory networks taking place in H3.3K9A. Several ERV-derived

regulatory elements can recruit pluripotency transcription factors [292] and the ectopic activation of ERV-derived CREs in H3.3K9A mESCs could provide a plethora of new binding sites which are generally not available in the basal state. Assuming that the pool of available TFs does not change, this would ultimately result in a partial diversion of TFs to alien CREs due to the competition between these newly available binding sites with canonical enhancers (Figure 3.5). This hypothesis was recently tested experimentally and reported in a preprint, in which the authors observed a reduction in histone acetylation, chromatin accessibility and TF binding at active chromatin regions upon acute KAP1 depletion in mESCs [336]. This was attributed to the de-repression of ERVs, and the effect was modelled by inserting thousands of ectopic binding sites of pluripotency transcription factors, which had a similar effect on canonical regulatory elements [336]. More experiments would be necessary to verify whether a similar mechanism occurs in H3.3K9A mESCs, but the data support this hypothesis to some extent. Firstly, the ChIP-Atlas data integration showed that the Cluster 1 H3K9me3-DA regions (i.e., the ones gaining more activity) are relatively enriched in NANOG, SOX2 and POU5F1 binding compared to other heterochromatic regions (Figure 2.22a). Secondly, the orthogonal analysis based on differential PRO-seq signal revealed that SOX2 and particularly POU5F1 binding sites are among the most significantly enriched at the regions with higher activity in H3.3K9A mESCs (Figure 2.28e). The diffTF analysis showed higher activity only for NANOG (Figure 2.29), and in this case it might not be surprising that pluripotency TFs do not display higher/lower activity because—if the model is correct—then the TFs would be essentially redistributed on different sites without necessarily resulting in an overall change in their activity. Finally, the “TFs rerouting model” would also fit with the observation that—according to the GRN analysis—the Strong-CREs are the most affected in H3.3K9A mESCs, in terms of decreased activity. This might make sense because, being the most active CREs in the basal state, they would be more touched upon by the diversion of pluripotency TFs. Although proving this hypothesis in our system might be challenging, ChIP-seq experiments for pluripotency factors could help to uncover whether the TFs are partially relocated on the H3K9me3-DA regions and hence concomitantly hijacked from canonical CREs.

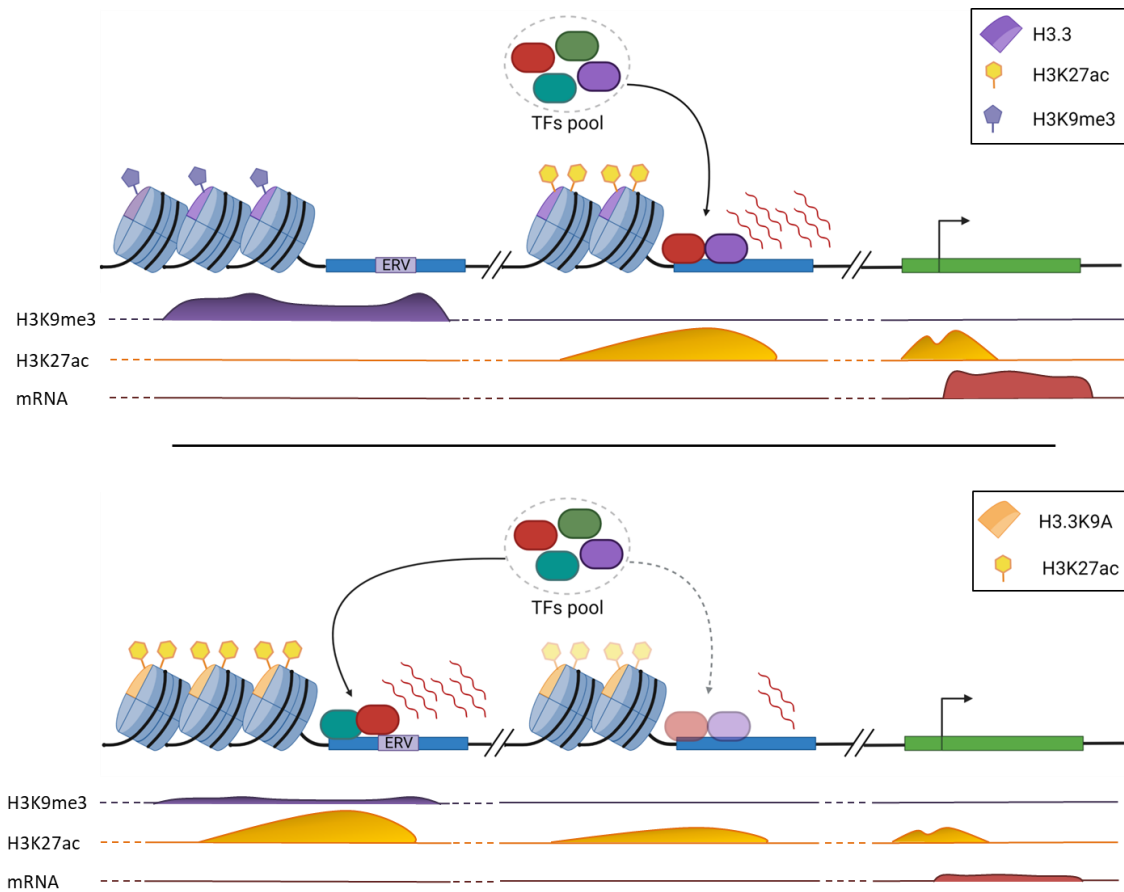


Figure 3.5 Schematic of the TFs rerouting model. In the basal state (top), pluripotency TFs are recruited at canonical enhancers and sustain transcription of target genes. When alien distal regulatory elements are activated in H3.3K9A mESCs (bottom), new binding sites for these TFs are made available and compete with canonical enhancers. As a result of decreased TF binding, the activity of enhancers and expression of target genes are reduced. Figure created with BioRender.com.

4. Materials and methods

Cell culture. Mouse embryonic stem cells (129XC57BL/6J) were cultured in ESC medium containing Knockout-DMEM (Thermo Fisher Scientific) supplemented with 15% EmbryoMax fetal bovine serum (ES-FBS Merck-Millipore) and 20 ng/ml leukemia inhibitory factor (LIF, produced by Protein Expression and Purification Facility at EMBL Heidelberg), 1% non-essential amino acids (Thermo Fisher Scientific), 1% Glutamax (Thermo Fisher Scientific), 1% Pen/Strep (Thermo Fisher Scientific), 1% of 55 mM beta-Mercaptoethanol solution (Merck Millipore). Upon thawing, cells were cultured on a layer of mouse embryonic fibroblasts (MEFs) and afterwards passaged on 0.1% gelatine-coated culture dishes, without feeder cells. Cells were maintained at 37°C with 5% CO₂ and routinely tested for mycoplasma. The medium was changed daily, and cells were passaged every two days.

Plasmid-based CRISPR-Cas9 editing strategy. This approach was used to generate K79A clones and first generation of K27A clones. The editing strategy was previously described in Gehre *et al.* [257, 307]. In brief, the guide RNA (sgRNA) was designed using the Benchling CRISPR guide RNA design tool and cloned into the pSpCas9(BB)-2A-GFP vector (PX458 – Addgene), according to the method described by Ran *et al.* [308]. A total of 2×10⁶ mESCs (*H3f3a*^{-/-}; *H3f3b*^{wt}) were co-transfected with 2 µg of pSpCas9(BB)-2A-GFP plasmid with sgRNA and 100 µM of single-stranded oligonucleotide (ssODN) repair template (180 bp – Integrated DNA Technologies Ultramers). A Lonza Nucleofector (program CG-104) was used for electroporation. After 48 hours, single-cell sorting was performed to select GFP positive cells and single clones were expanded for genotyping and freezing. Clonal cell lines were lysed in a 96-well plate (DirectPCR Viagen – cat#302-C) and screened for homozygous mutations by means of mismatch qPCR and Sanger sequencing of the target locus. Guide RNAs and repair templates used are reported in the following tables.

sgRNA	Target	Sequence (5'-3')	PAM
sgRNA_H3f3b_g1M_K79	H3f3b K79 codon	ATTGGTGAGGGAGATCGCCC	AGG
sgRNA_H3f3b_g5_K27	H3f3b K27 codon	CGCGCTTTTCCGAGCCGCCT	TGG

Repair template	Sequence (5'-3')
ssODN_g1M_K79A_mut	CGTGGCTCTGAGAGAGATCCGTCGTTACCAGAAATCGACTGAGC TGCTCATCCGGAAGCTGCCATTCCAGAGACTGGTGAGAGAGATC GCCCCAAGATTTCGCCACCGACTTGAGGTTTCAAAGTGCAGCCAT CGGTGCCCTTCAGGTAAAGGGACAAAAGGCTTGGTGGCACTCGG GGCGGCG
ssODN_g5_K27A_mut	CCGGCGGCCAAGTCATTGTTCCCAGCCCTCTGCCTACCTGTAGCG GTGAGGCTTCTTCACCCGCGGTAGAGGGCGCGCTGGCCCCGAGC CGCCTTTGTGGCCAGCTGTTTTCGGGGGGCTTTCCACCGGTGGA CTTCCTAGCGGTCTGCTTGGTTCGGGCCATTTTTCACCTGCG AGA

RNP-based CRISPR-Cas9 editing strategy. This approach was used to generate K9A and K27A clones. First, tracrRNA-ATTO550 and the target specific crRNA-XT (Integrated DNA Technologies) were resuspended in IDT Nuclease-Free Duplex Buffer to a final concentration of 200 μ M. The single-guide RNA (sgRNA) was assembled *in-vitro* by mixing tracrRNA and crRNA in equimolar ratio and the mixture was heated at 95°C for 5 min and cooled at RT on bench-top for ~ 20 minutes. Second, the Cas9 RNP was assembled by mixing the Cas9-mSA protein (produced by the EMBL Protein Expression and Purification Facility) with the sgRNA in a molar ratio of 1:1.5 Cas9:sgRNA, in a final volume of 10 μ l of Neon Buffer R per electroporation reaction and incubated at RT for 10-20 minutes. The biotinylated repair template (Integrated DNA Technologies Ultramers) was then added to the mix prior to electroporation, considering the same molar ratio of 1:1.5. Cells were washed once with PBS and 10⁵ cells/reaction were resuspended in 10 μ l of mix previously prepared (Buffer R + Cas9 RNP complex). The Neon Transfection System (Thermo Fisher Scientific) was used for electroporation using the following program: 1600 V; 10 ms; 3 pulses (from Liang *et al.* [337]). After electroporation, cells were transferred to a tube containing 500 μ l of pre-warmed RPMI medium supplemented with 8% ES FBS and left to recover for 15 minutes at 37°C in the incubator. Cells were then washed centrifuged and washed once with RPMI medium supplemented with 8% ES FBS and plated in 24-well plate containing MEFs. After 24 hours, single-cell sorting was performed to select ATTO550 positive cells and single clones were expanded and genotyped as previously described for the plasmid-based method.

sgRNA	Target	Sequence (5'-3')	PAM
crRNA_H3f3b_g3n_K9A	H3f3b_K9_codon	TTTGCGGGGGGCTTTCCAC	CGG
crRNA_H3f3b_g5_K27	H3f3b K27 codon	CGCGCTTTTCCGAGCCGCCT	TGG

Repair template	Sequence (5'-3')
ssODN_g3_new_K9A_82mer	GAAAAAAAAATGGCCCGAACCAAGCAGACCGCTAGGGCCT CCACAGGAGGGAAAGCCCCCGCAAACAGCTGGCCACCA AGGC
ssODN_g5_K27A_82mer	CACCGGTGGGAAAGCCCCCGCAAACAGCTGGCCACAAA GGCGGCTCGGGCCAGCGCGCCCTCTACCGGCGGGGTGAA GAAG
ssODN_g5M_K27-18- 14A_112mer	GAACCAAGCAGACCGCTAGGAAGTCCACCGGTGGGGCTG CCCCCGCGCTCAGCTGGCCACAAAGGCGGCTCGGGCCA GCGCGCCCTCTACCGGCGGGGTGAAGAAGCCTCA

Chromosomal integrity check of CRISPR-edited clones. The chromosomal integrity of the homozygous clones generated was assessed via low-coverage whole-genome sequencing. Briefly, genomic DNA was extracted from $\sim 10^6$ cells using the Gentra Puregene Cell Kit (Qiagen) and sonicated at 4°C using a Bioruptor Pico sonication device (Diagenode) for 10 cycles (30s ON/30s OFF) to obtain fragments of ~ 200 bp. The fragmentation pattern was assessed via agarose gel electrophoresis and 0.5-1 μ g of fragmented gDNA was used to prepare sequencing libraries using the NEBNext Ultra II DNA Library Preparation Kit (New England Biolabs), according to manufacturer's instructions. Libraries were then quantified using the Qubit fluorometer (Thermo Fisher Scientific) and the quality was assessed on a Bioanalyzer using the Agilent High Sensitivity DNA Kit. Samples were then pooled and sequenced on a Illumina NextSeq-500 platform (75 bp single-end reads).

Low-coverage whole-genome sequencing analysis. Quality of the sequencing run was assessed using FastQC (v 0.11.5 [338]) and adapter sequences were trimmed using TrimGalore (v 0.4.3 – <https://github.com/FelixKrueger/TrimGalore>). Sequencing reads were aligned to the mouse reference genome (GRCm38/mm10 assembly) using Bowtie2 (v 2.3.4 [339]) with default settings. Uniquely mapped reads (MAPQ $\geq$ 30) were retained for the subsequent steps. The bam files thus generated were then indexed and used for copy number variation analysis (bin size 100 kb) using the Coral script developed by Tobias Rausch (<https://github.com/tobiasrausch/coral>). Pre-processing steps were conducted with the Galaxy platform (v 21.09 [340]).

Protein extraction and western-blots. *Protein extractions and immunoblots were performed by Ana Janeva.* For crude nuclear extraction, cell pellets from 10^6 cells were resuspended in 600 μ l ice-cold lysis buffer (D-PBS, 1X cOmplete EDTA-free protease inhibitor, 1X Trichostatin A and 0.05% Triton-X) and rotated at 4°C for 30 min to lyse cells.

Nuclei were pelleted by centrifugation at 5000 g at 4°C for 5 min, resuspended in 150 µl RIPA buffer (25 mM TRIS pH 7.6, 150 mM NaCl, 1% NP-40, 1% Sodium-deoxycholate, 0.5% SDS) and sonicated with EpiShear Sonicator (Active Motif). Protein concentration was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For immunoblotting, protein samples were separated in 1X MES running buffer (50 mM MES, 50 mM Tris base Trizma, 3.4 mM SDS, 1 mM EDTA) on precast 4-12% Bis-Tris SDS-PAGE gels (Invitrogen) at 150 V for 1 hour and transferred onto Immobilon PVDF membranes (Merck Millipore) in 1X transfer buffer (25 mM Bicine, 25 mM Bis-Tris, 1 mM EDTA) at 30 V for 1 hour. The blots were incubated with primary antibodies overnight at 4°C under constant agitation, using the following antibodies diluted 1:1000 in 5% BSA/TBS-T: H3K27ac (39685, Active Motif), H3K27me3 (9733, Cell Signaling Technologies), H3.3 (09-838, Millipore) and H4 (10158, Abcam). Secondary antibody incubation was performed at RT for 1 hour with the following antibody: goat anti-rabbit IgG-HRP (1706515, BioRad), diluted 1:4000 in 5% BSA/TBS-T. Membranes were developed using the Immobilon Western Chemiluminescence HRP substrate (Millipore), diluted 1:5 in TBS-T, on a ChemiDoc Touch Imaging System by BioRad. Blots were quantified using Fiji Image J to calculate the band intensity of targets vs. loading control (H4) and normalized to protein input amount in controls and mutants.

Histones extraction and mass-spectrometry analysis. *These experiments and analyses were performed by Yekaterina Kori (from Benjamin Garcia Laboratory at University of Pennsylvania).* Histones were extracted, derivatized with propionic anhydride, and digested as previously described [341]. All samples were desalted prior to nanoLC-MS/MS analysis using in-house prepared C18 stage-tips. Histone samples were analyzed by nanoLC-MS/MS with a Dionex-nanoLC coupled to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). The column was packed in-house using reverse-phase 75 µm ID × 17 cm Reprosil-Pur C18-AQ (3 µm; Dr. Maisch GmbH). The HPLC gradient was: 4% solvent B (A = 0.1% formic acid; B = 80% acetonitrile, 0.1% formic acid) for 2 minutes, then from 4% to 34% solvent B over 48 min, then from 34% to 90% solvent B in 2 minutes, hold at 90% solvent B for 2 minutes, and from 90% solvent B down to 4% in 1 minute followed by a hold at 4% solvent B for 5 minutes. The flow rate was at 300 nL/min. Data were acquired using a data-independent acquisition method, consisting of a full scan MS spectrum (m/z 300–1200) performed in the Orbitrap at 120,000 resolution with an AGC target value of 5e5, followed by 16 MS/MS windows of 50 m/z using HCD fragmentation and detection in the orbitrap. HCD collision energy was set to 27, and AGC target at 1e4. Histone samples were resuspended in buffer A and 1 µg of total histones was

injected. Histone data was analyzed using a combination of EpiProfile 2.0 [342], Skyline [343], and manual analysis with Xcalibur (Thermo Fisher Scientific). The peptide relative ratio was calculated by using the area under the curve (AUC) for that particular peptide over the total AUC of all possible modified forms of that peptide. Data analysis was performed using Microsoft Excel to calculate averages and standard deviations.

PRC2 *in-vitro* HMT assay. *The HMTase activity assays were performed by Michael Uckelmann and Vita Levina (from Chen Davidovich Laboratory at EMBL Australia and Monash University).* The assay was performed as previously described [344], with the addition of normalisation samples that were loaded on each of the gels to allow for a comparison across multiple gels. Each 10 μ L HMTase reaction contained 500 nM PRC2 complex, chromatinized DNA with sequence from the ATOH1 locus (see Zhang *et al.*, 2021 [345]) at a total concentration equivalent to 0.8 μ M nucleosomes and 5.0 μ M 14C-SAM (S-[methyl-14C]-adenosyl-l-methionine (PerkinElmer, no. NEC363050UC)). All chromatinized DNA constructs were identical except for the H3 histone constructs, which included either H3.1 or H3.3, as indicated, and lysine 9 was either unmodified or modified to include a methyl lysine analog (MLA), as indicated. Each reaction was incubated for 30 minutes and 2 h at 30 °C in HMTase buffer (50 mM Tris-HCl pH 8.0 (at 30°C), 100 mM KCl, 2.5 mM MgCl₂, 0.1 mM ZnCl₂, 2.0 mM 2-mercaptoethanol and 0.1 mg/mL-1 bovine serum albumin) and then stopped by adding 4 $\times$ LDS sample buffer ($\sim$ 3.3 μ L) (Thermo Fisher Scientific, no. NP0007) to a final concentration of 1 $\times$ (final volume $\sim$ 13.3 μ L). 10 μ L of each sample was then incubated at 95 °C for 5 min and subjected to SDS-PAGE (16.5% acrylamide). Gels were first stained with InstantBlue Coomassie protein stain (Expedeon, no. ISB1L) and then vacuum-dried for 60 min at 80 °C with the aid of a VE-11 electric aspirator pump (Jeio Tech). Dried gels were exposed to a storage phosphor screen (Cytiva) for 5 days and the signal was acquired using a Typhoon Trio imager (Cytiva). All experiments were performed in triplicate.

Growth assay. mESCs were seeded at a density of 10⁴ cells per well and I seeded 12 wells/line on day 0. Two wells per cell line were trypsinized and counted each day for a total of 6 days, while the remaining wells received fresh media. I generated growth curves using the averaged duplicate cell counts.

Cell cycle assay. *Cell cycle analysis was performed by Umut Yildiz, with the help of Nadine Fernandez-Novel Marx.* The percentage of proliferating cells in S-Phase was measured using the EdU Click FC-488 Kit (Carl Roth – BA-7779) following the manufacturer's

instructions. Briefly, EdU was added at a final concentration of 10 μ M and cells were incubated for 2.5 hours at 37°C. Cells were washed once with PBS and then dissociated using accutase. Cells were fixed by resuspension in 100 μ l (every 1×10^6) of 4% PFA in PBS and incubation 15 minutes at RT, in the dark. After quenching with PBS + 1% BSA, cells were centrifuged and the pellet resuspended in 1X saponin-based permeabilization buffer. A total of 4.5×10^5 cells were used for each condition and 500 μ l of Click reaction master mix (PBS, catalyst solution, dye azide and 10X buffer additive) were added to each tube. After 30 minutes incubation at RT in the dark, cells were washed once with 3 ml of saponin-based permeabilization buffer and DNA content was stained with propidium iodide (7 minutes incubation at RT). Cells were washed once with saponin-based permeabilization buffer and resuspended in 300-500 μ l of buffer and analysed by flow cytometry. Three controls were included for flow cytometry analysis: i) w/o EdU, w/o click and w/o DAPI; ii) w/o Edu, with click and DAPI; iii) with EdU, w/o click and with DAPI.

Cell death assay. *Cell death analysis was performed by Umut Yildiz, with the help of Nadine Fernandez-Novel Marx.* The FITC Annexin V Kit (BD Pharmingen – 556419) was used, which allows to detect earlier stages of cell death, preceding loss of membrane integrity. Briefly, cells were washed twice with PBS and resuspended in 1X binding buffer to a concentration of 1×10^6 cells/ml. To 100 μ l of the mix ($\sim 10^5$ cells), 5 μ l of FITC Annexin V and 10 μ l of propidium iodide (50 μ g/ml stock) were added and cells were incubated for 15 minutes at RT, in the dark. After incubation, 400 μ l of 1X binding buffer were added to each tube and cells were analysed by flow cytometry. Three controls were included for flow cytometry analysis: i) unstained cells; ii) cells with FITC Annexin V, w/o PI; iii) cells w/o FITC-Annexin V, with PI.

Immune proteins staining. *This staining was performed by Nadine Fernandez-Novel Marx and Michael Bonadonna performed the flow cytometry analysis.* Cells were harvested and washed with FACS buffer (PBS + 2% FBS). Cells were then resuspended in 100 μ l of master mix composed of FACS buffer and desired antibodies, each one diluted 1:100. Alternatively, cells were resuspended in FACS buffer only for the unstained controls. After 30 minutes incubation on ice and in the dark, cells were centrifuged and washed with FACS buffer. Antibodies used for the staining were the following: PE-CD59a (143103 – Biolegend); APC-TLR2/CD282 (153005 – Biolegend); FITC-MHC Class II (107605 – Biolegend).

Bodipy-C11 staining. *Lipid peroxidation analysis was performed by Umut Yildiz.* The staining was performed following the guidelines described by Martinez *et al.* [346]. In brief, the BODIPY-C11 probe was added at a final concentration of 2.5 μ M and cells were incubated for 30 minutes at 37°C. Cells were then washed with HBSS buffer, collected and analysed by flow cytometry.

Neuronal differentiation. *Neuronal differentiations were performed by myself, Nadine Fernandez-Novel Marx and Ana Janeva.* mESCs were differentiated into glutamatergic neurons, as described by Bibel *et al.* [347], with modifications. mESCs were cultured in differentiation medium containing high-glucose DMEM, 10% FBS (Gibco), 1% non-essential amino acids (Thermo Fisher Scientific), 1% penicillin/streptomycin (Thermo Fisher Scientific), 1% GlutaMAX, 1% sodium pyruvate (Thermo Fisher Scientific), 0,1% of 14,5 M β -mercaptoethanol (Sigma Aldrich) solution. To promote the formation of embryoid bodies, cells were grown in suspension on non-adherent dishes (Greiner Petri dishes – 633102). The differentiation medium was changed every two days. From day 4 until day 8, the EBs were treated with 5 μ M retinoic acid every two days to induce neuronal lineage commitment. On day 8, EBs were dissociated with trypsin and neural progenitor cells (NPCs) were plated on poly-D-lysine hydrobromide (Sigma Aldrich) and laminin (Roche) coated dishes in N2 medium (high-glucose DMEM, 1% N-2, 2% B-27 and 1% penicillin/streptomycin, from Thermo Fisher Scientific) at a density of 2×10^5 cells/cm². The medium was changed after 2 h and after 24 h. On day 10, the medium was changed to complete medium (neurobasal, 2% B-27 and 1% penicillin/streptomycin). Glutamatergic neurons were harvested on day 12.

Mesoderm-cardiomyocyte differentiation. *Mesoderm-cardiomyocyte differentiations were performed by myself, Nadine Fernandez-Novel Marx and Ana Janeva.* mESCs were differentiated into cardiomyocytes adapting a protocol from previous publications [348, 349]. Briefly, 7.5×10^5 mESCs were resuspended in differentiation medium (DMEM, 20% FBS, 1% NEAA, 1% P/S, 1% GlutaMAX, 100 μ M ascorbic acid) and grown in suspension on non-adherent dishes to promote EB formation. After 4 days, EB were plated onto 0,1% gelatine-coated plates. The medium was changed every two days. Formation of contracting colonies was observed after 8 days of differentiation and time-course quantification of contracting colonies was performed at day 8, 10 and 12. Cardiomyocytes were harvested via trypsinization on day 14.

Quantitative real-time PCR. RNA was extracted from approximately 10^6 cells using RNeasy Kit (Qiagen), followed by DNase digestion using TURBO DNase (Thermo Fisher Scientific). A total of 1 μ g of RNA was reverse transcribed with random primers to cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). For qRT-PCR reaction, 9 ng of cDNA were used as template and reactions were performed using Power SYBR Green master mix (Applied Biosystems) on a StepOne Plus Real-Time PCR machine. The comparative Ct method ($\Delta\Delta$ Ct method) was used to calculate normalized gene expression values. Ct values of target genes were normalized to Ct values of the housekeeping gene Rpl13 to obtain Δ Ct values and to control samples to obtain $\Delta\Delta$ Ct values. Primers used for the NPCs qRT-qPCR are listed in the following table.

Primer	Forward	Reverse
Hash1	TTCTCCGGTCTCGTCCTACTC	CCAGTTGGTAAAGTCCAGCAG
Neurod1	ATGACCAAATCATACAGCGAGAG	TCTGCCTCGTGTTCTCTCGT
Pax6	GCGCAGACGGCATGTATGATA	GGGTTGCCCTGGTACTGAAG
Sox2	CGCGGCGGAAAACCAAGACG	GCCGGCGCCCAACCAACC
Rpl13	AACTTAAGCTGGCCACCCA	ACTCTGGCCTTTTCCTTTTGT

Immunofluorescence. *The stainings were performed by Nadine Fernandez-Novel Marx and Ana Janeva.* NPCs were seeded on glass coverslips at day 8 of neuronal differentiation and fixed at day 12 for neuronal staining. Cells were washed with PBS briefly and fixed with 3% paraformaldehyde (PFA) (Electron Microscopy Sciences) in PBS for 20 minutes at RT, then quenched with 30 mM glycine in PBS for 5 minutes at RT. Cells were washed three times with PBS and stored in PBS at 4°C until needed. Permeabilization was conducted with 0.1% Triton-X in PBS for 5 minutes at RT, followed by blocking with 0.5% BSA in PBS for 30 minutes at RT. Primary antibody incubation was performed for 1 hour at RT under constant shaking. The following antibodies were diluted 1:200 in 0.5% BSA and used for stainings: Map2 (Sigma Aldrich, 9942) and β -III tubulin (abcam, ab78078). Cells were washed three times with PBS and incubated for 30 minutes at RT with secondary antibody (goat anti-mouse IgG Alexa 594, Thermo Fisher Scientific – A11005), diluted 1:1000 in 0.5% BSA. After washing twice with PBS, cells were counterstained with 5 μ g/ml DAPI for 5 minutes at RT and washed three times with PBS. Coverslips were mounted with Mowiol mounting medium (Calbiochem) and imaged on a Nikon Eclipse Ti fluorescence microscope. Images were processed with Fiji Image J.

mRNA-seq. RNA was extracted from $\sim 10^6$ cells using the RNeasy Mini kit (Qiagen) and DNase digestion was performed using TURBO DNase (Thermo Fisher Scientific), according to the manufacturer's instructions and an additional clean-up of the RNA was then

performed using the RNeasy Mini kit (Qiagen). Quality of the RNA was assessed on Bioanalyzer using the Agilent RNA 6000 Nano Kit and samples with a RNA integrity number (RIN) of 10 were used for library generation. For each sample, 100 ng of total RNA was used as input for mRNA selection and conversion to cDNA using the NEBNext Poly(A) Magnetic Isolation Module (New England BioLabs). Sequencing libraries were prepared using the NEBNext Ultra II RNA Library Preparation Kit for Illumina (New England BioLabs), according to the manufacturer's instructions. After Qubit quantification, individual the quality of individual libraries was assessed on a Bioanalyzer using the Agilent High Sensitivity DNA Kit. Samples were then pooled and sequenced on an Illumina NextSeq500 sequencer (read length of 75 bp in single-end mode).

mRNA-seq analysis. Quality of the sequencing reads was assessed using FastQC (v 0.11.5 [338]) and adapter sequences were trimmed using TrimGalore (v 0.4.3). Sequencing reads were aligned to the mouse reference genome (GRCm38/mm10 assembly) using STAR (v 2.5.2b [350]) with default settings. Uniquely mapped reads ($\text{MAPQ} \geq 20$) were retained for the subsequent steps. Gene count tables were generated with featureCounts (subread v 1.6.2 [351]), using gencode gene annotations (release M10). Coverage files were generated using bamCoverage (deeptools v 2.4.1 [352]). Differential expression analysis was performed in R using the DESeq2 package [353]. Genes were considered differentially expressed using a false discovery rate (FDR) cut-off of 0.05 and the fold-change cut-off applied is indicated in the main text and in figure captions. MA-plots were generated using the ggmaplot function (from the ggpubr R package – v 0.4.0; <https://github.com/kassambara/ggpubr>). Upset plot in Figure 2.5 was generated using the UpSetR package [354]. Gene ontology enrichment analysis was conducted using ClusterProfiler [355, 356]. Other plots were generated using the ggplot2 R package (<https://github.com/tidyverse/ggplot2>). Pre-processing steps were conducted with the Galaxy platform (v 21.09 [340]).

Repeat elements analysis. For endogenous retroviruses (ERVs) expression analysis, the RepEnrich tool [357] was used in a dedicated conda environment, in its updated version (<https://github.com/nerettilab/RepEnrich2>). LTRs annotation for the mm10 mouse genome assembly was retrieved using the RepeatMasker track from the UCSC genome table browser and prepared as described in the github repository. After running RepEnrich2 independently for all samples, the individual output tables with estimated counts were merged, and the differential expression analysis was performed in R using DESeq2. Repeats were considered differentially expressed using a false discovery rate

(FDR) cut-off = 0.05 and abs(log2 fold-change) cut-off = 0.585. The PCA plot was generated in R with ggplot2 and volcano plots were generated with the EnhancedVolcano package (<https://github.com/kevinblighe/EnhancedVolcano>).

Native ChIP-seq. DNA for native ChIP was digested by MNase treatment to obtain mainly mono-nucleosomes using a modified protocol from Barski *et al.*, 2007 [155]. For each IP, 20x10⁶ cells were resuspended in digestion buffer (50 mM Tris-HCl pH 7.6; 1 mM CaCl₂; 0.2% Triton-X), treated with 100 U MNase (Worthington) and incubated at 37°C for 5 minutes while shaking at 500 rpm. Samples were quickly moved to ice and MNase was quenched by addition of 5 mM EDTA (final). Lysates were sonicated for 3 cycles using Bioruptor Pico (Diagenode) and dialyzed against RIPA buffer for 3 hours at 4°C. Insoluble materials was pelleted at 10,000 rpm for 10 minutes at 4°C and supernatant was used as input for ChIP. To check the fragmentation pattern, input DNA was analysed by agarose gel electrophoresis. A 5% fraction of input was set aside for sequencing. For native ChIP of histone modifications, Protein-G-Dynabeads (Invitrogen) were pre-coated with antibodies for 4 hours at 4°C and the following antibodies were used for this protocol: H3K27ac (39685 – Active Motif); H3K9ac (C5B11-9549 – Cell Signalling Technology). After overnight incubation of the pre-coated beads with chromatin lysates, beads were washed (3X with RIPA; 2X with RIPA + 300 mM NaCl; 2X with LiCl buffer–250 mM LiCl, 0.5% NP-40, 0.5% NaDoc) and finally rinsed with TE + 50 mM NaCl Buffer. Samples were eluted from ProteinG Dynabeads using SDS elution buffer (50 mM Tris-HCl pH 8.0; 10 mM EDTA; 1% SDS) at 65°C for 30 minutes with shaking at 1500 rpm. Proteinase K was added (0.2 µg/ml final) to the eluted samples and digestion was carried out for 2 hours at 55°C followed by PCR purification (Qiagen). Sequencing libraries were prepared using the NEBNext Ultra II DNA Library Preparation Kit (New England Biolabs) and sequenced on Illumina NextSeq500 platform (75 bp in single-end mode).

Cross-link ChIP-seq. The protocol was adapted from Groh *et al.*, 2021 [291] and Navarro *et al.*, 2020 [290]. Cells were harvested and crosslinked in 3 ml of pre-tempered (25°C) ES medium supplemented with 1% formaldehyde (10⁶ cells/3 ml) for 10 minutes at RT, with rotation. Formaldehyde was then quenched with glycine (125 mM final) and incubated 5 minutes at RT, with rotation. Cells were washed twice with ice-cold PBS containing 10%FBS (centrifugation at 1100 rpm for 4 minutes, at 4°C). Pellets can be stored at –80°C for several months.

To prepare lysates, fixed cells were resuspended in 300 µl of Sonication Buffer 1 (50 mM Tris-HCl pH 8.0; 0.5% SDS) and incubated for 10 minutes on ice. Sonication was performed

with Bioruptor Pico (Diagenode) for 20 cycles (30"ON/30"OFF). Sonicated lysates were then diluted 1:6 with lysis buffer (10 mM Tris-HCl pH 8.0; 100 mM NaCl; 1% Triton-X; 1 mM EDTA; 0.5 mM EGTA; 0.1% NaDoc; 0.5% N-lauroylsarcosine) and the soluble fraction was collected after centrifugation at full speed for 10 minutes at 4°C. A 2.5% fraction of the supernatant was set aside for input and the fragmentation pattern (~150–500 bp) was checked by agarose gel electrophoresis. Lysates can be aliquoted and stored at –80°C with 10% glycerol (final). For each IP, 30 µl Protein G Dynabeads (Invitrogen) were washed twice with 1 ml PBS-T (PBS + 0.01% Tween-20) and incubated with the desired antibody for at least 1 hour at RT (or > 1 hour at 4°C).

Antibodies used with this protocol were the following: H3.3 (09-838 – Merck-Millipore); H3K9me3 (D4W1U – Cell Signalling Technology); H3K9me2 (D85B4 – Cell Signalling Technology); H3K27me3 (C36B11 – Cell Signalling Technology); H3K18ac (9675 – Cell Signalling Technology); SUZ12 (D39F6-3737 – Cell Signalling Technology).

Coated beads were then washed (1X with PBS-T; 2X with lysis buffer), resuspended in 30 µl of lysis buffer/IP and added to the desired amount of chromatin (generally 15–25 µg of chromatin were used, corresponding to 2–4x10⁶ cells). After over-night incubation at 4°C with rotation, beads-immunocomplexes were washed twice—each time for 5 minutes—with the following buffers: RIPA; RIPA + 360 mM NaCl; LiCl buffer (10 mM Tris-HCl pH 8.0; 250 mM LiCl, 0.5% NP-40, 0.5% NaDoc; 1 mM EDTA) and finally quickly rinsed with TE buffer and eluted in ChIP SDS elution buffer (10 mM Tris-HCl pH 8.0; 300 mM NaCl; 5 mM EDTA; 0.5% SDS). RNA/protein digestion was performed on beads by adding 2 µl RNaseA (10 mg/ml stock) incubated 30 minutes at 37°C, followed by addition of 1.5 µl Proteinase K (20 mg/ml stock) incubated 1h @ 55°C. Reverse cross-link was performed with over-night incubation at 65°C. DNA was purified with 1.4X SPRI-select beads. For H3K27me3 ChIP-Rx, spike-in chromatin was prepared from *Drosophila* Schneider 2 (S2) cells following the same protocol and aliquots were stored at –80°C with 10% glycerol. Before immunoprecipitation, 3% of exogenous chromatin was added to each reaction. Sequencing libraries were prepared using the NEBNext Ultra II DNA Library Preparation Kit (New England Biolabs) and sequenced on Illumina NextSeq500 (75 bp in single-end mode) or NextSeq2000 (P3 kit – 88 bp in single-end mode).

ChIP-seq analysis. Quality of the sequencing run was assessed using FastQC (v 0.11.5 [338]) and adapter sequences were trimmed using TrimGalore (v 0.4.3 – <https://github.com/FelixKrueger/TrimGalore>). Sequencing reads were aligned to the mouse reference genome (GRCm38/mm10 assembly) using Bowtie2 (v 2.3.4 [339]). Uniquely mapped reads (MAPQ ≥ 20) aligning to major chromosomes were retained for

the subsequent steps. ChIP signal strength and sequencing depth were assessed using the plotFingerprint and plotCoverage (deeptools v 2.4.1 [352]). Coverage files (bigwig format) were generated using deeptools bamCoverage: 10 bp was used as bin size, the “reads per genomic content (RPGC)” method was used for normalization and reads were extended to an average fragment size of 150 bp. For the H3K27me3 ChIP-seq with exogenous spike-in, reads were aligned to a mouse (mm10) + Drosophila (dm6) combined reference genome. The number of uniquely mapped reads aligning to major mouse and fly chromosomes was retrieved and used to calculate normalization factors that were used with deeptools bamCoverage to generate scaled bigwig files. Peak calling was performed with MACS2 [358, 359], by providing ChIP and respective input bam files and considering a minimum FDR cut-off for peak detection of 0.05; narrow or broad peak calling was performed for histone modifications following ENCODE guidelines [360]. Differential peak analysis was performed in R using the DiffBind package [361], only peaks detected in at least two replicates were retained for the analysis and peaks were called as significantly differential considering an FDR cut-off = 0.05. Annotation of ChIP-seq peaks was performed with the ChIPseeker package [362] and with publicly available ChromHMM state maps for mESCs (https://github.com/guifengwei/ChromHMM_mESC_mm10). Heatmaps and metagene plots were generated in R using a script developed by Daria Bunina: genomic regions of interest were binned in 40 windows of 250 bp each, the ChIP-seq signal from individual bigwig files was measured in each window and all bigwig regions overlapping the same window were averaged. The signal for replicates in each condition was then averaged. Heatmaps were generated using the pheatmap package (<https://cran.r-project.org/web/packages/pheatmap/index.html>). Boxplots and violin plots were generated considering the ChIP-seq signal measured at TSS or peak summit +/- 1 kb. Overlaps between genomic regions of interest were quantified using the GenomicRanges objects and findOverlaps-methods [363].

ChIP-Atlas data integration. Publicly available ChIP-seq data for chromatin factors and transcription factors in mESCs (mm10 assembly) were retrieved from the ChIP-Atlas database [314] and the analysis was performed with a script developed by Daria Bunina, with an approach similar to the one previously described [364]. Briefly, genomic coordinates of peaks for all mESCs factors (n=292) were retrieved from the ChIP-Atlas Peak Browser and the findOverlaps function was used to overlap these peaks with genomic regions of interest. The significance of the relative enrichment of chromatin/transcription factors was assessed using the Fisher’s exact test.

ChIP-qPCR. Prepared libraries from ChIP experiments were diluted and 5 ng of DNA was used as input for each qPCR reaction with SYBR Green PCR Master Mix (Applied Biosystems). The qPCR reactions were performed on a StepOnePlus Real-Time PCR machine. For each condition biological replicates (ChIP material from two independent mutant cell lines) were used and measured in technical duplicates. Primers used for H3K9me3 and H3K18ac ChIP-qPCR are reported in the following table.

Primer	Forward	Reverse
IAPgag	AGCAGGTGAAGCCACTG	CTTGCCACACTTAGAGC
IAPLTR	TTGATAGTTGTGTAGTGGTAAATAAA	AAAACACCACAAACAAAATCTTCTAC
LINE_ORF1a	ACTCAAAGCGAGGCAACACTAGA	GTTCCAGATTTCTTTCCTAGGGTTTC
LINE_ORF1b	AGGCTACTATACCCAGCCAACTCT	TACTTTGGTTTCTTCTATGATAATTG
Rdna	CCTGTGAATTCTCTGAACTC	CCTAAACTGCTGACAGGGTG
SINE_B1	GTGGCGCACGCCTTTAATC	GACAGGGTTTCTCTGTGTAG
SINE_B2	GAGATGGCTCAGTGGTTAAG	CTGTCTTCAGACACTCCAG
Rosa26_US	GGAATTCAGCTGCTTCCTG	GTGGGGACGTAAGTGAAGGA
Telomere	CGGTTTGTGGGTTTGGGTTTGGGTT TGGGTTTGGGTT	GGCTTGCTTACCCTTACCCTTACCCTT ACCCTTACCCT
Dppa5a_Enh	TGTTTCAGTTTGGCAAGTCCA	GAAGATGCGAAGAGCCGAGA
Ssr2_Enh	GTTTCTGTTGGTCTCTGGTAT	CTTGGTGGGCAAAGTTACATCA
Epha2_Enh	CCACGACCCAGTCTTCTGAG	CCTGCAGTAGAGAGCCATGC
Slc5a12a_Enh	GGACACTAAAGAGACCGCCA	GTATGCATGAAACAGTGGGACT
Lncenc1_Enh	TCTTTGTGTGAGAAATGGGCAC	TGATGCACTGTTACAATGGGGA
Ercc6l2_Enh	GCTAAGGCCTGGGTCATCAG	AGCCTAAGCAGTCAAGGCAG
Klf9_Enh	CTGTTGTCTGCTGGAACCCA	GGGGAGAGAGCTGCAATAGG

PRO-seq. Precision nuclear run-on transcription sequencing (PRO-seq) experiments were performed following the protocol previously described [365]. PRO-seq was performed in two batches, one with two replicates each of control and H3.3K27A mESCs and one with three replicates each of control and H3.3K9A mESCs. Briefly, 10⁷ mESCs were permeabilized and used for each reaction and 5% (i.e., 500k) of permeabilized *Drosophila* S2 cells were added prior to the *in-vitro* run-on reaction. Run-on reactions were carried out by adding 100 µl of 2× NRO reaction mixture (10 mM of Tris-HCl, pH 8.0; 5 mM MgCl₂; 300 mM KCl; 1 mM dithiothreitol (DTT); 1% sarkosyl; 375 µM biotin-11-CTP/-UTP; 375 µM ATP/GTP; 2 µl RNase inhibitor) to 100 µl of the permeabilized cells, and incubated at 30 °C for 3 min. Addition of TRIzol LS (Thermo Fisher Scientific) terminated the reaction and RNA was extracted and precipitated in 75% ethanol. Nuclear RNA was fragmented by base hydrolysis in 0.2 N of NaOH on ice for 10 min and subsequently neutralized by adding the same volume of 1 M of Tris-HCl, pH 6.8. Fragmented biotinylated RNA was bound to the streptavidin magnetic beads (Thermo Fisher Scientific), washed, eluted using TRIzol and precipitated in ethanol. Following 3'-RNA and 5'-RNA adapter ligation, the

precipitated RNA was reverse-transcribed and PCR-amplified. DNA libraries were purified using SPRI-select beads and quality was assessed on a Bioanalyzer using the Agilent High Sensitivity DNA Kit, pooled and sequenced on an Illumina NextSeq 500 platform (75 bp in single-end mode).

PRO-seq analysis. The analysis of the PRO-seq data was performed according to the guidelines provided by the Danko Laboratory. In particular, pre-processing and alignment was performed using the PROseq2.0 pipeline (<https://github.com/Danko-Lab/proseq2.0>). The consensus set of distal *cis*-regulatory elements active at the basal state in mESCs was obtained by running the dREG package (<https://github.com/Danko-Lab/dREG>) on the PRO-seq data from control cell lines. The differential PRO-seq analysis was performed using the output of the dREG package from control and H3.3K9A lines as input for the tfTarget package (<https://github.com/Danko-Lab/tfTarget>). The violin plots in Figure 2.22, Figure 2.34 and Figure 2.35 were generated by calculating RPKM values for mutant and control samples within the genomic regions of interest.

Gene regulatory network construction and differential transcription factor activity analysis. *These analyses were performed by Daria Bunina, with the help of Brian Mang Ching Lai (from Judith Zaugg Laboratory at EMBL Heidelberg).* The GRN was generated using the GRaNIe and GRaNPAs packages developed by Kamal & Arnold *et al.*, 2022 [317]. As introduced in the results section, mRNA-seq and H3K27ac ChIP-seq datasets from 12 independent samples generated in the project (i.e., three biological replicates each for control, H3.3 K9A, K27A and K79A mESCs) were used to construct the GRN. The differential transcription factor activity analysis was performed using the diffTF method described in Berest & Arnold *et al.*, 2019 [325], using the same 12 mRNA-seq and ChIP-seq datasets as input.

Appendix

I. Characterization of H3.3 K79A mESCs

The H3.3K79A mESCs were generated as described in the Results section 2.1.1. Given that the mRNA-seq and the H3K27ac ChIP-seq data generated with these lines were also used for the construction of the GRN that I used to connect CREs to putative target genes, I will report these data in the appendix. In Figure A.1, editing strategy and validation are reported.

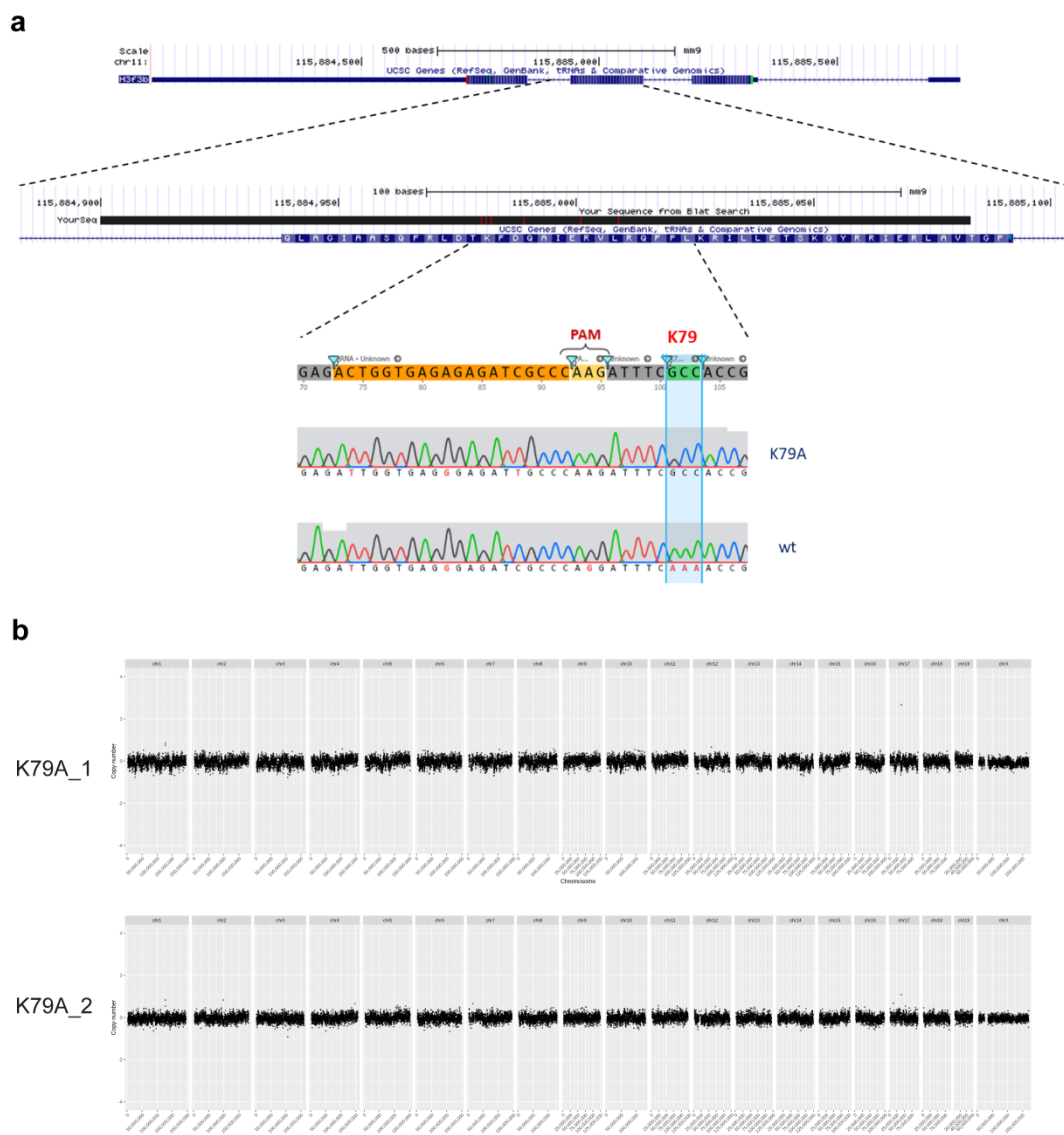


Figure A.1 *H3f3b* K79 editing strategy and validation of chromosomal integrity. (a) The second exon of the *H3f3b* locus was targeted for editing and the repair template used is shown in black (red vertical bars represent inserted mutations). At the bottom, a representative Sanger sequencing chromatogram of a successfully edited K79A clone is shown, in comparison with the wild-type sequence. **(b)** Representative

whole-genome plots show the copy-number difference of 100 kb genomic bins between edited clones and the parental control line.

The H3.3K79A clones did not show any evident phenotypic difference, compared to control. Cell cycle analysis by means of EdU labelling and flow cytometry analysis (performed with the help of Eva Lotta Kasper) showed mild yet significant decrease in the percentage of cells in G2/M phase and a mild significant increase of cells in G1 phase for K79A mESCs (Figure A.2c). The K79A mESCs could be successfully differentiated to both neuronal and meso-cardiomyocyte lineages *in-vitro* (data not shown).

With three biological replicates of H3.3K79A (*H3f3a*^{KO}/*H3f3b*^{K79A}) which did not show chromosomal abnormalities, I performed mRNA-seq to evaluate the effect of the inserted mutation. Differential gene expression analysis showed only mild transcriptional changes, with 235 genes identified as differentially expressed (Figure A.2a) when comparing mutant with control (padj < 0.05 & FC cut-off=1.5). GO analysis showed enrichment for terms related to mesenchyme development and regulation to chemotaxis for the upregulated genes (Figure A.2b). Downregulated genes did not show significant enrichment for specific ontology terms. H3K27ac ChIP-seq differential peak analysis revealed the presence of 8388 differential peaks (n=6453 down and n=1928 up; data not shown). Further experiments would be necessary for a better characterization of the downstream effects of the H3.3K79A mutation in mESCs.

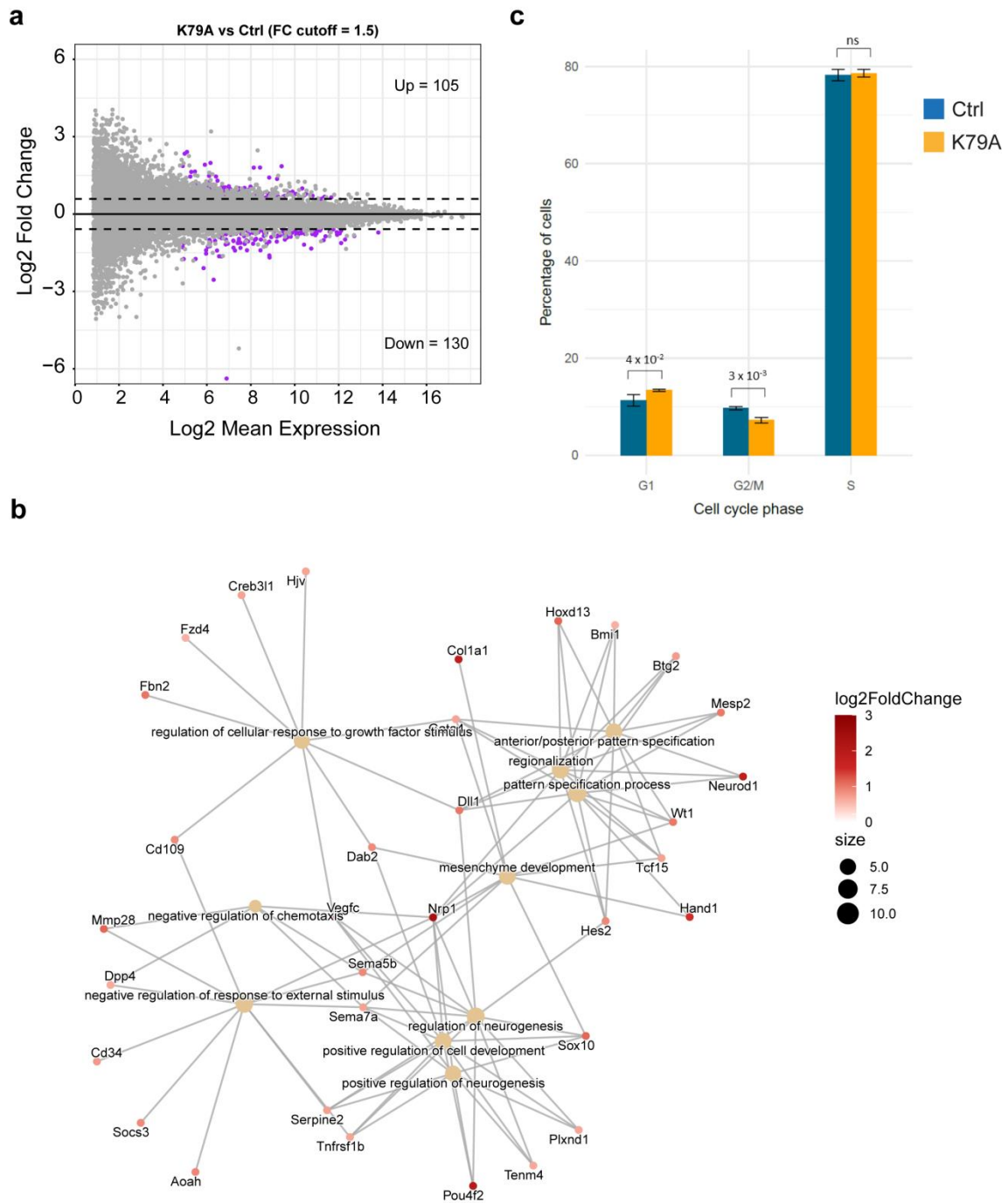


Figure A.2 Differential gene expression and cell cycle analysis of H3.3K79A mESCs. (a) MA-plot of DEGs in K79A mESCs. Coloured genes represent significantly DEGs defined with $\text{padj} < 0.05$ and fold-change cut-off=1.5. (b) Cnet-plot with most significant GO terms for genes upregulated in K79A mESCs (colors of individual genes represent $\log_2$ Fold-change values). (c) Barplot with percentages of cells in G1, S or G2/M phases are reported and unpaired t-test was used to calculate significance.

Bibliography

1. Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* 13:233–245
2. Cramer P (2019) Organization and regulation of gene transcription. *Nature* 573:45–54
3. Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 13:613–626
4. Banerji J, Rusconi S, Schaffner W (1981) Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27:299–308
5. Furlong EEM, Levine M (2018) Developmental enhancers and chromosome topology. *Science* 361:1341–1345
6. Bulger M, Groudine M (2011) Functional and Mechanistic Diversity of Distal Transcription Enhancers. *Cell* 144:327–339
7. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G (2013) Enhancers: five essential questions. *Nat Rev Genet* 14:288–295
8. Schoenfelder S, Fraser P (2019) Long-range enhancer–promoter contacts in gene expression control. *Nat Rev Genet* 20:437–455
9. Dekker J, Heard E (2015) Structural and functional diversity of Topologically Associating Domains. *FEBS Lett* 589:2877–2884
10. Di Giammartino DC, Kloetgen A, Polyzos A, et al (2019) KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat Cell Biol* 21:1179–1190
11. Andersson R (2015) Promoter or enhancer, what’s the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays* 37:314–323
12. Sartorelli V, Lauberth SM (2020) Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* 27:521–528
13. Andersson R, Gebhard C, Miguel-Escalada I, et al (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507:455–461
14. Reiter F, Wienerroither S, Stark A (2017) Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev* 43:73–81
15. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153:307–319
16. Blobel GA, Higgs DR, Mitchell JA, Notani D, Young RA (2021) Testing the super-enhancer concept. *Nat Rev Genet* 22:749–755

17. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260
18. Kornberg R (1974) Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. *Science* 184:868–871
19. Tessarz P, Kouzarides T (2014) Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* 15:703–708
20. Tan M, Luo H, Lee S, et al (2011) Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* 146:1016–1028
21. Farrelly LA, Thompson RE, Zhao S, et al (2019) Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* 1
22. Lepack AE, Werner CT, Stewart AF, et al (2020) Dopaminylation of histone H3 in ventral tegmental area regulates cocaine seeking. *Science* 368:197–201
23. Allfrey VG, Faulkner R, Mirski AE (1964) Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* 51:786–94
24. Brownell JE, Allis CD (1995) An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proc Natl Acad Sci* 92:6364–6368
25. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD (1996) *Tetrahymena* Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. *Cell* 84:843–851
26. Vavra KJ, Allis CD, Gorovsky MA (1982) Regulation of histone acetylation in *Tetrahymena* macro- and micronuclei. *J Biol Chem* 257:2591–2598
27. Marmorstein R (2001) Structure of Histone Acetyltransferases. *J Mol Biol* 433–444
28. Marmorstein R, Zhou MM (2014) Writers and readers of histone acetylation: Structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol* 6:1–26
29. Lee KK, Workman JL (2007) Histone acetyltransferase complexes : one size doesn ' t fit all. *Nat Rev Mol Cell Biol* 8:284–295
30. Avvakumov N, Côté J (2007) The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene* 26:5395–5407
31. Liu X, Wang L, Zhao K, Thompson PR, Hwang Y, Marmorstein R, Cole PA (2008) The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nat* 2008 4517180 451:846–850
32. Zhang Y, Xue Y, Shi J, et al (2018) The ZZ domain of p300 mediates specificity of the adjacent HAT domain for histone H3. *Nat Struct Mol Biol* 2018 259 25:841–849
33. Jin Q, Yu L-R, Wang L, Zhang Z, Kasper LH, Lee J-E, Wang C, Brindle PK, Dent SYR, Ge K (2011) Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J* 30:249–262

34. Glozak MA, Sengupta N, Zhang X, Seto E (2005) Acetylation and deacetylation of non-histone proteins. *Gene* 363:15–23
35. Bannister* AJ, Miska EA (2000) Regulation of gene expression by transcription factor acetylation. *Cell Mol Life Sci CMLS* 57:1184–1192
36. Nagy Z, Tora L (2007) Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* 26:5341–5357
37. Narita T, Weinert BT, Choudhary C (2019) Functions and mechanisms of non-histone protein acetylation. *Nat Rev Mol Cell Biol* 20:156–174
38. Shvedunova M, Akhtar A (2022) Modulation of cellular processes by histone and non-histone protein acetylation. *Nat Rev Mol Cell Biol* 23:329–349
39. Iwafuchi-Doi M, Zaret KS (2016) Cell fate control by pioneer transcription factors. *Development* 143:1833–1837
40. Zhu F, Farnung L, Kaasinen E, et al (2018) The interaction landscape between transcription factors and the nucleosome. *Nature* 562:76–81
41. Utley RT, Ikeda K, Grant PA, Côté J, Steger DJ, Eberharter A, John S, Workman JL (1998) Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* 394:498–502
42. Fuda NJ, Ardehali MB, Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461:186–192
43. Shahbazian MD, Grunstein M (2007) Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu Rev Biochem* 76:75–100
44. Mutskov V, Gerber D, Angelov D, Ausio J, Workman J, Dimitrov S (1998) Persistent Interactions of Core Histone Tails with Nucleosomal DNA following Acetylation and Transcription Factor Binding. *Mol Cell Biol* 18:6293–6304
45. Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* 311:844–847
46. Kalashnikova AA, Porter-Goff ME, Muthurajan UM, Luger K, Hansen JC (2013) The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J R Soc Interface* 10:20121022
47. Zhang R, Erler J, Langowski J (2017) Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction. *Biophys J* 112:450–459
48. Nitsch S, Zorro Shahidian L, Schneider R (2021) Histone acylations and chromatin dynamics: concepts, challenges, and links to metabolism. *EMBO Rep* 22:e52774
49. Tamkun JW, Deuring R, Scott MP, Kissinger M, Pattatucci AM, Kaufman TC, Kennison JA (1992) *brahma*: A regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. *Cell* 68:561–572

50. Cieniewicz AM, Moreland L, Ringel AE, Mackintosh SG, Raman A, Gilbert TM, Wolberger C, Tackett AJ, Taverna SD (2014) The Bromodomain of Gcn5 Regulates Site Specificity of Lysine Acetylation on Histone H3. *Mol Cell Proteomics* 13:2896–2910
51. Längst G, Manelyte L (2015) Chromatin Remodelers: From Function to Dysfunction. *Genes* 6:299–324
52. Clapier CR, Iwasa J, Cairns BR, Peterson CL (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* 18:407–422
53. Wu T, Kamikawa YF, Donohoe ME (2018) Brd4's Bromodomains Mediate Histone H3 Acetylation and Chromatin Remodeling in Pluripotent Cells through P300 and Brg1. *Cell Rep* 25:1756–1771
54. Belkina AC, Denis GV (2012) BET domain co-regulators in obesity, inflammation and cancer. *Nat Rev Cancer* 12:465–477
55. Fujisawa T, Filippakopoulos P (2017) Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat Rev Mol Cell Biol* 18:246–262
56. LeRoy G, Rickards B, Flint SJ (2008) The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription. *Mol Cell* 30:51–60
57. Josling GA, Selvarajah SA, Petter M, Duffy MF (2012) The Role of Bromodomain Proteins in Regulating Gene Expression. *Genes* 3:320–343
58. Jang MK, Mochizuki K, Zhou M, Jeong H-S, Brady JN, Ozato K (2005) The Bromodomain Protein Brd4 Is a Positive Regulatory Component of P-TEFb and Stimulates RNA Polymerase II-Dependent Transcription. *Mol Cell* 19:523–534
59. Yang Z, Yik JHN, Chen R, He N, Jang MK, Ozato K, Zhou Q (2005) Recruitment of P-TEFb for Stimulation of Transcriptional Elongation by the Bromodomain Protein Brd4. *Mol Cell* 19:535–545
60. Bhagwat AS, Roe J-S, Mok BYL, Hohmann AF, Shi J, Vakoc CR (2016) BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. *Cell Rep* 15:519–530
61. Lee J-E, Park Y-K, Park S, Jang Y, Waring N, Dey A, Ozato K, Lai B, Peng W, Ge K (2017) Brd4 binds to active enhancers to control cell identity gene induction in adipogenesis and myogenesis. *Nat Commun* 8:2217
62. Crump NT, Ballabio E, Godfrey L, et al (2021) BET inhibition disrupts transcription but retains enhancer-promoter contact. *Nat Commun* 12:1–15
63. Sooraj D, Sun C, Doan A, et al (2022) MED12 and BRD4 cooperate to sustain cancer growth upon loss of mediator kinase. *Mol Cell* 82:123–139.e7
64. Shorstova T, Foulkes WD, Witcher M (2021) Achieving clinical success with BET inhibitors as anti-cancer agents. *Br J Cancer* 124:1478–1490

65. Zhao D, Li Y, Xiong X, Chen Z, Li H (2017) YEATS Domain—A Histone Acylation Reader in Health and Disease. *J Mol Biol* 429:1994–2002
66. Khan A, Bridgers JB, Strahl BD (2017) Expanding the Reader Landscape of Histone Acylation. *Structure* 25:571–573
67. Haberland M, Montgomery RL, Olson EN (2009) The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* 10:32–42
68. Falkenberg KJ, Johnstone RW (2014) Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat Rev Drug Discov* 13:673–691
69. Jones PA, Issa JPI, Baylin S (2016) Targeting the cancer epigenome for therapy. *Nat Rev Genet* 17:630–641
70. Freitas MF, Cuendet M, Bertrand P, Faria M, Cuendet M, Hdac PB, Bertrand P (2018) HDAC inhibitors : a 2013 – 2017 patent survey. *Expert Opin Ther Pat* 28:365–381
71. Rea S, Eisenhaber F, O’Carroll D, et al (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406:593–599
72. Murray K (1964) The Occurrence of ϵ -N-Methyl Lysine in Histones. *Biochemistry* 3:10–15
73. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y (2002) Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain. *Curr Biol* 12:1052–1058
74. van Leeuwen F, Gafken PR, Gottschling DE (2002) Dot1p Modulates Silencing in Yeast by Methylation of the Nucleosome Core. *Cell* 109:745–756
75. Crump NT, Milne TA (2019) Why are so many MLL lysine methyltransferases required for normal mammalian development? *Cell Mol Life Sci* 76:2885–2898
76. Cornett EM, Ferry L, Defossez P-A, Rothbart SB (2019) Lysine Methylation Regulators Moonlighting outside the Epigenome. *Mol Cell* 75:1092–1101
77. Choi J, Jang H, Kim H, Lee J-H, Kim S-T, Cho E-J, Youn H-D (2014) Modulation of lysine methylation in myocyte enhancer factor 2 during skeletal muscle cell differentiation. *Nucleic Acids Res* 42:224–234
78. Tsusaka T, Kikuchi M, Shimazu T, Suzuki T, Sohtome Y, Akakabe M, Sodeoka M, Dohmae N, Umehara T, Shinkai Y (2018) Tri-methylation of ATF7IP by G9a/GLP recruits the chromodomain protein MPP8. *Epigenetics Chromatin* 11:56
79. Rathert P, Dhayalan A, Murakami M, Zhang X, Tamas R, Jurkowska R, Komatsu Y, Shinkai Y, Cheng X, Jeltsch A (2008) Protein lysine methyltransferase G9a acts on non-histone targets. *Nat Chem Biol* 4:344–346
80. Vermeulen M, Eberl HC, Matarese F, et al (2010) Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* 142:967–980

81. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* 14:1025–1040
82. Greer EL, Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* 13:343–357
83. Mozzetta C, Boyarchuk E, Pontis J, Ait-Si-Ali S (2015) Sound of silence: the properties and functions of repressive Lys methyltransferases. *Nat Rev Mol Cell Biol* 16:499–513
84. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410:116–120
85. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410:120–124
86. Stewart MD, Li J, Wong J (2005) Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. *Mol Cell Biol* 25:2525–38
87. Lomberk G, Wallrath L, Urrutia R (2006) The Heterochromatin Protein 1 family. *Genome Biol* 7:228
88. Li Y, Kirschmann DA, Wallrath LL (2002) Does heterochromatin protein 1 always follow code? *Proc Natl Acad Sci* 99:16462–16469
89. Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* 17:1870–1881
90. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y (2004) Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* 119:941–953
91. Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L (2012) H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* 13:424
92. Bian C, Xu C, Ruan J, et al (2011) Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J* 30:2829–2842
93. Li Y, Wen H, Xi Y, et al (2014) AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* 159:558–571
94. Gates LA, Shi J, Rohira AD, et al (2017) Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J Biol Chem* 292:14456–14472
95. Chen FX, Woodfin AR, Gardini A, Rickels RA, Marshall SA, Smith ER, Shiekhata R, Shilatifard A (2015) PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* 162:1003–1015

96. Zsindely N, Pankotai T, Újfaludi Z, Lakatos D, Komonyi O, Bodai L, Tora L, Boros IM (2009) The loss of histone H3 lysine 9 acetylation due to dSAGA-specific dAda2b mutation influences the expression of only a small subset of genes. *Nucleic Acids Res* 37:6665–6680
97. Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y (2002) A Complex with Chromatin Modifiers That Occupies E2F- and Myc-Responsive Genes in G0 Cells. *Science* 296:1132–1136
98. Shinkai Y, Tachibana M (2011) H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev* 25:781–788
99. Collins R, Cheng X (2010) A case study in cross-talk: the histone lysine methyltransferases G9a and GLP. *Nucleic Acids Res* 38:3503–3511
100. Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, Sakihama T, Kodama T, Hamakubo T, Shinkai Y (2005) Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev* 19:815–826
101. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* 41:246–250
102. Collins RE, Northrop JP, Horton JR, Lee DY, Zhang X, Stallcup MR, Cheng X (2008) The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat Struct Mol Biol* 15:245–250
103. Escamilla-Del-Arenal M, Rocha ST da, Spruijt CG, Masui O, Renaud O, Smits AH, Margueron R, Vermeulen M, Heard E (2013) Cdy1, a New Partner of the Inactive X Chromosome and Potential Reader of H3K27me3 and H3K9me2. *Mol Cell Biol* 33:5005–5020
104. Saksouk N, Simboeck E, Déjardin J (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* 8:3
105. Peters AHFM, O'Carroll D, Scherthan H, et al (2001) Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell* 107:323–337
106. Schoeftner S, Blasco MA (2009) A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *EMBO J* 28:2323–2336
107. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* 16:919–932
108. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y (2010) Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464:927–931
109. Urrutia R (2003) KRAB-containing zinc-finger repressor proteins. *Genome Biol* 4:231

110. Bilodeau S, Kagey MH, Frampton GM, Rahl PB, Young RA (2009) SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 23:2484–2489
111. Nicetto D, Donahue G, Jain T, et al (2019) H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science* 363:294–297
112. Ninova M, Fejes Tóth K, Aravin AA (2019) The control of gene expression and cell identity by H3K9 trimethylation. *Development* 146:dev181180
113. Maison C, Almouzni G (2004) HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* 5:296–305
114. Maksakova IA, Goyal P, Bullwinkel J, Brown JP, Bilenky M, Mager DL, Singh PB, Lorincz MC (2011) H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. *Epigenetics Chromatin* 4:1–18
115. Creighton MP, Cheng AW, Welstead GG, et al (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 107:21931–21936
116. Heintzman ND, Hon GC, Hawkins RD, et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459:108–112
117. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470:279–285
118. Lee J-E, Wang C, Xu S, et al (2013) H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* 2:e01503
119. Lai B, Lee J-E, Jang Y, Wang L, Peng W, Ge K (2017) MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. *Nucleic Acids Res* 45:6388–6403
120. Jeong KW, Kim K, Situ AJ, Ulmer TS, An W, Stallcup MR (2011) Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. *Nat Struct Mol Biol* 18:1358–1365
121. Narita T, Ito S, Higashijima Y, et al (2021) Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. *Mol Cell* 81:2166–2182.e6
122. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri PC, Harte PJ (2009) CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* 136:3131–3141
123. Pasini D, Malatesta M, Jung HR, et al (2010) Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* 38:4958–4969

124. Lavarone E, Barbieri CM, Pasini D (2019) Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nat Commun* 10:1–16
125. Ferrari KJ, Scelfo A, Jammula SG, Cuomo A, Barozzi I, Stützer A, Fischle W, Bonaldi T, Pasini D (2014) Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Mol Cell* 53:49–62
126. Schuettengruber B, Bourbon H-M, Croce LD, Cavalli G (2017) Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 171:34–57
127. Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D (2008) Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Mol Cell* 32:503–518
128. Loh CH, Veenstra GJC (2022) The Role of Polycomb Proteins in Cell Lineage Commitment and Embryonic Development. *Epigenomes* 6:23
129. Chammas P, Mocavini I, Di Croce L (2020) Engaging chromatin: PRC2 structure meets function. *Br J Cancer* 122:315–328
130. Tamburri S, Conway E, Pasini D (2022) Polycomb-dependent histone H2A ubiquitination links developmental disorders with cancer. *Trends Genet* 38:333–352
131. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes Dev* 25:1010–1022
132. Laugesen A, Højfeldt JW, Helin K (2019) Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol Cell* 74:8–18
133. Li H, Liefke R, Jiang J, et al (2017) Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549:287–291
134. Margueron R, Justin N, Ohno K, et al (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461:762–767
135. Lee CH, Yu JR, Kumar S, Jin Y, LeRoy G, Bhanu N, Kaneko S, Garcia BA, Hamilton AD, Reinberg D (2018) Allosteric Activation Dictates PRC2 Activity Independent of Its Recruitment to Chromatin. *Mol Cell* 70:422–434.e6
136. Yuan W, Xu M, Huang C, Liu N, Chen S, Zhu B (2011) H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem* 286:7983–7989
137. Musselman CA, Avvakumov N, Watanabe R, et al (2012) Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat Struct Mol Biol* 19:1266–1272
138. Brien GL, Gambero G, O’Connell DJ, et al (2012) Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol* 19:1273–1281
139. Singer MS, Kahana A, Wolf AJ, Meisinger LL, Peterson SE, Goggin C, Mahowald M, Gottschling DE (1998) Identification of High-Copy Disruptors of Telomeric Silencing in *Saccharomyces cerevisiae*. *Genetics* 150:613–632

140. Worden EJ, Hoffmann NA, Hicks CW, Wolberger C (2019) Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell*. <https://doi.org/10.1016/j.cell.2019.02.002>
141. Valencia-Sánchez MI, De Ioannes P, Wang M, Vasilyev N, Chen R, Nudler E, Armache J-P, Armache K-J (2019) Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. *Mol Cell* 74:1010-1019.e6
142. Patel DJ, Krajewski K, Sidoli S, et al (2015) The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79. *Mol Cell* 60:319-327
143. Valencia-Sánchez MI, De Ioannes P, Wang M, Truong DM, Lee R, Armache J-P, Boeke JD, Armache K-J (2021) Regulation of the Dot1 histone H3K79 methyltransferase by histone H4K16 acetylation. *Science* 371:eabc6663
144. Steger DJ, Lefterova MI, Ying L, et al (2008) DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells. *Mol Cell Biol* 28:2825-2839
145. Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EEM (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* 44:148-156
146. Gilan O, Lam EYN, Becher I, et al (2016) Functional interdependence of BRD4 and DOT1L in MLL leukemia. *Nat Struct Mol Biol* 23:673-681
147. Godfrey L, Crump NT, Thorne R, et al (2019) DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat Commun*. <https://doi.org/10.1038/s41467-019-10844-3>
148. Godfrey L, Crump NT, O'Byrne S, et al (2020) H3K79me2/3 controls enhancer-promoter interactions and activation of the pan-cancer stem cell marker PROM1/CD133 in MLL-AF4 leukemia cells. *Leukemia* 1-17
149. Jones B, Su H, Bhat A, et al (2008) The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. *PLoS Genet* 4:e1000190
150. Cattaneo P, Kunderfranco P, Greco C, et al (2016) DOT1L-mediated H3K79me2 modification critically regulates gene expression during cardiomyocyte differentiation. *Cell Death Differ* 23:555-564
151. Barry ER, Krueger W, Jakuba CM, Veilleux E, Ambrosi DJ, Nelson CE, Rasmussen TP (2009) ES Cell Cycle Progression and Differentiation Require the Action of the Histone Methyltransferase Dot1L. *Stem Cells* 27:1538-1547
152. Cao K, Ugarenko M, Ozark PA, et al (2020) DOT1L-controlled cell-fate determination and transcription elongation are independent of H3K79 methylation. *Proc Natl Acad Sci U S A* 117:27365-27373
153. Kang J, Kim J, Kim K, et al (2018) KDM2B is a histone H3K79 demethylase and induces transcriptional repression *via* sirtuin-1-mediated chromatin silencing. *FASEB J* 32:5737-5750

154. Mikkelsen TS, Ku M, Jaffe DB, et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560
155. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129:823–837
156. Mendenhall EM, Bernstein BE (2008) Chromatin state maps: new technologies, new insights. *Curr Opin Genet Dev* 18:109–115
157. Trojer P, Reinberg D (2007) Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Mol Cell* 28:1–13
158. Papp B, Müller J (2006) Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev* 20:2041–2054
159. Bernstein BE, Mikkelsen TS, Xie X, et al (2006) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125:315–326
160. Ku M, Koche RP, Rheinbay E, et al (2008) Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLOS Genet* 4:e1000242
161. Macrae TA, Fothergill-Robinson J, Ramalho-Santos M (2022) Regulation, functions and transmission of bivalent chromatin during mammalian development. *Nat Rev Mol Cell Biol* 1–21
162. Dhar SS, Lee S-H, Chen K, et al (2016) An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic Acids Res* 44:3659–3674
163. Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A (2013) The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol* 20:1093–1097
164. Mozzetta C, Pontis J, Fritsch L, Robin P, Portoso M, Proux C, Margueron R, Ait-Si-Ali S (2014) The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. *Mol Cell* 53:277–289
165. Banaszynski LA, Wen D, Dewell S, et al (2013) Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. *Cell* 155:107–20
166. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403:41–45
167. Morgan MAJ, Shilatifard A (2020) Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nat Genet* 52:1271–1281
168. Dorigi KM, Swigut T, Henriques T, Garcia BA, Adelman K, Correspondence JW (2017) Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation Loss of Mll3/4 from enhancers reduces Pol II density in bodies of adjacent genes. *Mol Cell* 66:568–576

169. Cao K, Collings CK, Morgan MA, Marshall SA, Rendleman EJ, Ozark PA, Smith ER, Shilatifard A (2018) An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. *Sci Adv* 4:eaap8747
170. Bledau AS, Schmidt K, Neumann K, et al (2014) The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development* 141:1022–1035
171. Sze CC, Cao K, Collings CK, Marshall SA, Rendleman EJ, Ozark PA, Chen FX, Morgan MA, Wang L, Shilatifard A (2017) Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation. *Genes Dev.* <https://doi.org/10.1101/gad.303768.117>
172. Douillet D, Sze CC, Ryan C, et al (2020) Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. *Nat Genet* 52:615–625
173. Martin BJE, Brind'Amour J, Kuzmin A, Jensen KN, Liu ZC, Lorincz M, Howe LJ (2021) Transcription shapes genome-wide histone acetylation patterns. *Nat Commun* 12:210
174. Zencir S, Dilg D, Shore D, Albert B (2022) Pitfalls in using phenanthroline to study the causal relationship between promoter nucleosome acetylation and transcription. *Nat Commun* 13:3726
175. Martin BJE, Howe LJ (2022) Reply to: Pitfalls in using phenanthroline to study the causal relationship between promoter nucleosome acetylation and transcription. *Nat Commun* 13:3725
176. Wang Z, Chivu AG, Choate LA, et al (2022) Prediction of histone post-translational modification patterns based on nascent transcription data. *Nat Genet* 54:295–305
177. Raisner R, Kharbanda S, Jin L, et al (2018) Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. *Cell Rep* 24:1722–1729
178. Postberg J, Forcob S, Chang W-J, Lipps HJ (2010) The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol* 10:259
179. Woo YH, Li W-H (2012) Evolutionary Conservation of Histone Modifications in Mammals. *Mol Biol Evol* 29:1757–1767
180. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33:510–517
181. Wang K, Escobar M, Li J, Mahata B, Goell J, Shah S, Cluck M, Hilton IB (2022) Systematic comparison of CRISPR-based transcriptional activators uncovers gene-regulatory features of enhancer–promoter interactions. *Nucleic Acids Res* 50:7842–7855
182. Policarpi C, Munafò M, Tsagkris S, Carlini V, Hackett JA (2022) Systematic Epigenome Editing Captures the Context-dependent Instructive Function of Chromatin Modifications. 2022.09.04.506519

183. Snowden AW, Gregory PD, Case CC, Pabo CO (2002) Gene-Specific Targeting of H3K9 Methylation Is Sufficient for Initiating Repression In Vivo. *Curr Biol* 12:2159–2166
184. Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, Mao X, Yuan GC, Orkin SH (2008) EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Mol Cell* 32:491–502
185. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M (2012) Epigenetic protein families: A new frontier for drug discovery. *Nat Rev Drug Discov* 11:384–400
186. Fierz B, Muir TW (2012) Chromatin as an expansive canvas for chemical biology. *Nat Chem Biol* 8:417–427
187. Dahlin JL, Nelson KM, Strasser JM, et al (2017) Assay interference and off-target liabilities of reported histone acetyltransferase inhibitors. *Nat Commun* 8:1–14
188. Hyland EM, Cosgrove MS, Molina H, Wang D, Pandey A, Cottee RJ, Boeke JD (2005) Insights into the Role of Histone H3 and Histone H4 Core Modifiable Residues in *Saccharomyces cerevisiae*. *Mol Cell Biol* 25:10060–10070
189. Dai J, Hyland EM, Yuan DS, Huang H, Bader JS, Boeke JD (2008) Probing Nucleosome Function: A Highly Versatile Library of Synthetic Histone H3 and H4 Mutants. *Cell* 134:1066–1078
190. Huang H, Maertens AM, Hyland EM, Dai J, Norris A, Boeke JD, Bader JS (2009) HistoneHits: A database for histone mutations and their phenotypes. *Genome Res* 19:674–681
191. Nakanishi S, Sanderson BW, Delventhal KM, Bradford WD, Staehling-Hampton K, Shilatifard A (2008) A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat Struct Mol Biol* 15:881–888
192. Smith MM, Santisteban MS (1998) Genetic dissection of histone function. *Methods Companion Methods Enzymol* 15:269–281
193. Marzluff W, Gongidi P, Woods K, Jin J, Maltais L (2002) The Human and Mouse Replication-Dependent Histone Genes. *Genomics* 80:487–498
194. Marzluff WF, Wagner EJ, Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nat Rev Genet* 9:843–854
195. Dominski Z, Marzluff WF (2007) Formation of the 3' end of histone mRNA: Getting closer to the end. *Gene* 396:373–390
196. Günesdogan U, Jäckle H, Herzig A (2010) A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes. *EMBO Rep* 11:772–776
197. McKay DJ, Klusza S, Penke TJR, et al (2015) Interrogating the function of metazoan histones using engineered gene clusters. *Dev Cell* 32:373–386
198. Zhang W, Zhang X, Xue Z, et al (2019) Probing the Function of Metazoan Histones with a Systematic Library of H3 and H4 Mutants. *Dev Cell* 48:406–419.e5

199. Atlasi Y, Stunnenberg HG (2017) The interplay of epigenetic marks during stem cell differentiation and development. *Nat Rev Genet* 18:643–658
200. Sankar A, Mohammad F, Sundaramurthy AK, Wang H, Lerdrup M, Tatar T, Helin K (2022) Histone editing elucidates the functional roles of H3K27 methylation and acetylation in mammals. *Nat Genet* 2022 1–7
201. Trovato M, Patil V, Gehre M, Noh KM (2020) Histone Variant H3.3 Mutations in Defining the Chromatin Function in Mammals. *Cells* 9:2716
202. Filipescu D, Müller S, Almouzni G (2014) Histone H3 Variants and Their Chaperones During Development and Disease: Contributing to Epigenetic Control. *Annu Rev Cell Dev Biol* 30:615–646
203. Wiedemann SM, Mildner SN, Bönisch C, et al (2010) Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *J Cell Biol* 190:777–791
204. Buschbeck M, Hake SB (2017) Variants of core histones and their roles in cell fate decisions, development and cancer. *Nat Rev Mol Cell Biol* 18:299–314
205. Howman EV, Fowler KJ, Newson AJ, Redward S, MacDonald AC, Kalitsis P, Choo KH (2000) Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc Natl Acad Sci U S A* 97:1148–53
206. Maze I, Noh K-M, Soshnev AA, Allis CD (2014) Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nat Rev Genet* 15:259–271
207. Rai TS, Puri A, McBryan T, Hoffman J, Tang Y, Pchelintsev NA, Tuyn J van, Marmorstein R, Schultz DC, Adams PD (2011) Human CABIN1 Is a Functional Member of the Human HIRA/UBN1/ASF1a Histone H3.3 Chaperone Complex. *Mol Cell Biol* 31:4107–4118
208. Daniel Ricketts M, Frederick B, Hoff H, Tang Y, Schultz DC, Singh Rai T, Grazia Vizioli M, Adams PD, Marmorstein R (2015) Ubinuclein-1 confers histone H3.3-specific-binding by the HIRA histone chaperone complex. *Nat Commun* 6:7711
209. Ray-Gallet D, Quivy JP, Scamps C, Martini EMD, Lipinski M, Almouzni G (2002) HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Mol Cell* 9:1091–1100
210. Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116:51–61
211. DeNizio JE, Elsässer SJ, Black BE (2014) DAXX co-folds with H3.3/H4 using high local stability conferred by the H3.3 variant recognition residues. *Nucleic Acids Res* 42:4318–31
212. Liu CP, Xiong C, Wang M, Yu Z, Yang N, Chen P, Zhang Z, Li G, Xu RM (2012) Structure of the variant histone H3.3-H4 heterodimer in complex with its chaperone DAXX. *Nat Struct Mol Biol* 19:1287–1292

213. Lewis PW, Elsaesser SJ, Noh K-M, Stadler SC, Allis CD (2010) Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A* 107:14075–80
214. Goldberg AD, Banaszynski LA, Noh K-M, et al (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* 140:678–91
215. Elsässer SJ, Noh K-M, Diaz N, Allis CD, Banaszynski LA (2015) Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* 522:240–244
216. Ivanauskiene K, Delbarre E, McGhie JD, Küntziger T, Wong LH, Collas P (2014) The PML-associated protein DEK regulates the balance of H3.3 loading on chromatin and is important for telomere integrity. *Genome Res* 24:1584–94
217. Konev AY, Tribus M, Park SY, et al (2007) CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. *Science* 317:1087–90
218. Pradhan SK, Su T, Yen L, Jacquet K, Huang C, Côté J, Kurdistani SK, Carey MF (2016) EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. *Mol Cell* 61:27–38
219. Jang C-W, Shibata Y, Starmer J, Yee D, Magnuson T (2015) Histone H3.3 maintains genome integrity during mammalian development. *Genes Dev* 29:1377–92
220. Couldrey C, Carlton MBL, Nolan PM, Colledge WH, Evans MJ (1999) A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. *Hum Mol Genet* 8:2489–2495
221. Bush KM, Yuen BT, Barrilleaux BL, Riggs JW, O’Geen H, Cotterman RF, Knoepfler PS (2013) Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. *Epigenetics Chromatin* 6:7
222. Tang MCW, Jacobs SA, Wong LH, Mann JR (2013) Conditional allelic replacement applied to genes encoding the histone variant H3.3 in the mouse. *genesis* 51:142–146
223. Tang MCW, Jacobs SA, Mattiske DM, et al (2015) Contribution of the Two Genes Encoding Histone Variant H3.3 to Viability and Fertility in Mice. *PLOS Genet* 11:e1004964
224. Muhire BM, Booker MA, Tolstorukov MY (2019) Non-neutral evolution of H3.3-encoding genes occurs without alterations in protein sequence. *Sci Rep* 9:1–11
225. Maze I, Wenderski W, Noh K-M, et al (2015) Critical Role of Histone Turnover in Neuronal Transcription and Plasticity. *Neuron* 87:77–94
226. Hake SB, Garcia BA, Duncan EM, Kauer M, Dellaire G, Shabanowitz J, Bazett-Jones DP, Allis CD, Hunt DF (2006) Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem* 281:559–68

227. McKittrick E, Gafken PR, Ahmad K, Henikoff S (2004) Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A* 101:1525–30
228. Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE (2004) Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res* 32:6511–6518
229. Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G (2006) PTMs on H3 Variants before Chromatin Assembly Potentiate Their Final Epigenetic State. *Mol Cell* 24:309–316
230. Udugama M, Chang FTM, Chan FL, Tang MC, Pickett HA, McGhie JDR, Mayne L, Collas P, Mann JR, Wong LH (2015) Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic Acids Res* 43:gkv847
231. Martire S, Gogate AA, Whitmill A, Tafessu A, Nguyen J, Teng Y-C, Tastemel M, Banaszynski LA (2019) Phosphorylation of histone H3.3 at serine 31 promotes p300 activity and enhancer acetylation. *Nat Genet* 51:941–946
232. Sitbon D, Boyarchuk E, Dingli F, Loew D, Almouzni G (2020) Histone variant H3.3 residue S31 is essential for *Xenopus* gastrulation regardless of the deposition pathway. *Nat Commun* 11:1–15
233. Armache A, Yang S, Martínez de Paz A, et al (2020) Histone H3.3 phosphorylation amplifies stimulation-induced transcription. *Nature* 583:852–857
234. Khuong-Quang D-A, Buczkowicz P, Rakopoulos P, et al (2012) K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathol (Berl)* 124:439–447
235. Schwartzentruber J, Korshunov A, Liu XY, et al (2012) Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* 482:226–231
236. Wu G, Broniscer A, McEachron TA, et al (2012) Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet* 44:251–253
237. Behjati S, Tarpey PS, Presneau N, et al (2013) Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet* 45:1479–1482
238. Gessi M, Gielen GH, Hammes J, Dörner E, Mühlen AZ, Waha A, Pietsch T (2013) H3.3 G34R mutations in pediatric primitive neuroectodermal tumors of central nervous system (CNS-PNET) and pediatric glioblastomas: Possible diagnostic and therapeutic implications? *J Neurooncol* 112:67–72
239. Fontebasso AM, Papillon-Cavanagh S, Schwartzentruber J, et al (2014) Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma. *Nat Genet* 46:462–466
240. Kernohan KD, Grynspan D, Ramphal R, et al (2017) H3.1 K36M mutation in a congenital-onset soft tissue neoplasm. *Pediatr Blood Cancer* 64:e26633

241. Koelsche C, Schrimpf D, Tharun L, et al (2017) Histone 3.3 hotspot mutations in conventional osteosarcomas: a comprehensive clinical and molecular characterization of six H3F3A mutated cases. *Clin Sarcoma Res* 7:9
242. Shah MA, Denton EL, Arrowsmith CH, Lupien M, Schapira M (2014) A global assessment of cancer genomic alterations in epigenetic mechanisms. *Epigenetics Chromatin* 7:29
243. Lewis PW, Müller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, Garcia BA, Muir TW, Becher OJ, Allis CD (2013) Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* 340:857–861
244. Chan KM, Fang D, Gan H, et al (2013) The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev* 27:985–990
245. Justin N, Zhang Y, Tarricone C, et al (2016) Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun* 7:11316
246. Piunti A, Hashizume R, Morgan MA, et al (2017) Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. *Nat Med* 23:493–500
247. Mohammad F, Weissmann S, Leblanc B, et al (2017) EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat Med* 23:483–492
248. Stafford JM, Lee CH, Voigt P, et al (2018) Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma. *Sci Adv* 4:5935–5966
249. Jain SU, Rashoff AQ, Krabbenhoft SD, et al (2020) H3 K27M and EZHIP Impede H3K27-Methylation Spreading by Inhibiting Allosterically Stimulated PRC2. *Mol Cell*. <https://doi.org/10.1016/j.molcel.2020.09.028>
250. Krug B, De Jay N, Harutyunyan AS, et al (2019) Pervasive H3K27 Acetylation Leads to ERV Expression and a Therapeutic Vulnerability in H3K27M Gliomas. *Cancer Cell* 35:782-797.e8
251. Lu C, Jain SU, Hoelper D, et al (2016) Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science* 352:844–9
252. Fang D, Gan H, Lee JH, et al (2016) The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science* 352:1344–1348
253. Schmitges FW, Prusty AB, Faty M, et al (2011) Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol Cell* 42:330–341
254. Fang J, Huang Y, Mao G, Yang S, Rennert G, Gu L, Li H, Li GM (2018) Cancer-driving H3G34V/R/D mutations block H3K36 methylation and H3K36me3-MutSα interaction. *Proc Natl Acad Sci U S A* 115:9598–9603
255. Voon HPJ, Udugama M, Lin W, Hii L, Law RHP, Steer DL, Das PP, Mann JR, Wong LH (2018) Inhibition of a K9/K36 demethylase by an H3.3 point mutation found in paediatric glioblastoma. *Nat Commun* 9:1–10

256. Jain SU, Khazaei S, Marchione DM, et al (2020) Histone H3.3 G34 mutations promote aberrant PRC2 activity and drive tumor progression. *Proc Natl Acad Sci* 202006076
257. Gehre M, Bunina D, Sidoli S, Lübke MJ, Diaz N, Trovato M, Garcia BA, Zaugg JB, Noh KM (2020) Lysine 4 of histone H3.3 is required for embryonic stem cell differentiation, histone enrichment at regulatory regions and transcription accuracy. *Nat Genet* 52:273–282
258. Kraushaar DC, Chen Z, Tang Q, Cui K, Zhang J, Zhao K (2018) The gene repressor complex NuRD interacts with the histone variant H3.3 at promoters of active genes. *Genome Res* 28:1646–1655
259. Henikoff S (2008) Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet* 9:15–26
260. Chen P, Zhao J, Wang Y, et al (2013) H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes Dev* 27:2109–2124
261. Kraushaar DC, Jin W, Maunakea A, Abraham B, Ha M, Zhao K (2013) Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol* 14:R121
262. Zhang T, Zhang Z, Dong Q, Xiong J, Zhu B (2020) Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol* 21:45
263. Leatham-Jensen M, Uyehara CM, Strahl BD, Matera AG, Duronio RJ, McKay DJ (2019) Lysine 27 of replication-independent histone H3.3 is required for Polycomb target gene silencing but not for gene activation. *PLOS Genet* 15:e1007932
264. Santenard A, Ziegler-Birling C, Koch M, Tora L, Bannister AJ, Torres-Padilla ME (2010) Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat Cell Biol* 12:853–862
265. Brumbaugh J, Kim IS, Ji F, et al (2019) Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nat Cell Biol* 21:1449–1461
266. Herz H-M, Morgan M, Gao X, Jackson J, Rickels R, Swanson SK, Florens L, Washburn MP, Eissenberg JC, Shilatifard A (2014) Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. *Science* 345:1065–70
267. Elsässer SJ, Ernst RJ, Walker OS, Chin JW (2016) Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat Methods* 13:158–164
268. McClintock B (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci* 36:344–355
269. Mager DL, Stoye JP (2015) Mammalian Endogenous Retroviruses. *Microbiol Spectr.* <https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0009-2014/FORMAT/EPUB>
270. Johnson WE (2015) Endogenous Retroviruses in the Genomics Era. *Annu Rev Virol* 2:135–159

271. Hutnick LK, Huang X, Loo T-C, Ma Z, Fan G (2010) Repression of Retrotransposal Elements in Mouse Embryonic Stem Cells Is Primarily Mediated by a DNA Methylation-independent Mechanism*. *J Biol Chem* 285:21082–21091
272. Leung DC, Lorincz MC (2012) Silencing of endogenous retroviruses: When and why do histone marks predominate? *Trends Biochem Sci* 37:127–133
273. Hemberger M, Dean W, Reik W (2009) Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat Rev Mol Cell Biol* 10:526–537
274. Karimi MM, Goyal P, Maksakova IA, et al (2011) DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* 8:676–687
275. Bruno M, Mahgoub M, Macfarlan TS (2019) The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu Rev Genet* 53:393–416
276. Wolf D, Goff SP (2009) Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* 458:1201–1204
277. Wolf G, Yang P, Füchtbauer AC, Füchtbauer E-M, Silva AM, Park C, Wu W, Nielsen AL, Pedersen FS, Macfarlan TS (2015) The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev* 29:538–554
278. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D (2016) Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Dev Cell* 36:611–623
279. Rowe HM, Jakobsson J, Mesnard D, et al (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463:237–240
280. Groh S, Schotta G (2017) Silencing of endogenous retroviruses by heterochromatin. *Cell Mol Life Sci* 74:2055–2065
281. Geis FK, Goff SP (2020) Silencing and Transcriptional Regulation of Endogenous Retroviruses: An Overview. *Viruses* 12:884
282. Dong KB, Maksakova IA, Mohn F, et al (2008) DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J* 27:2691–2701
283. Bulut-Karslioglu A, De La Rosa-Velázquez IA, Ramirez F, et al (2014) Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Mol Cell* 55:277–290
284. Maksakova IA, Thompson PJ, Goyal P, Jones SJ, Singh PB, Karimi MM, Lorincz MC (2013) Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* 6:15
285. Yang BX, EL Farran CA, Guo HC, et al (2015) Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* 163:230–245

286. Macfarlan TS, Gifford WD, Agarwal S, et al (2011) Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* 25:594–607
287. Sachs P, Ding D, Bergmaier P, Lamp B, Schlagheck C, Finkernagel F, Nist A, Stiewe T, Mermoud JE (2019) SMARCA1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. *Nat Commun* 10:1335
288. Sadic D, Schmidt K, Groh S, Kondofersky I, Ellwart J, Fuchs C, Theis FJ, Schotta G (2015) Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep* 16:836–850
289. Wolf G, Rebollo R, Karimi MM, et al (2017) On the role of H3.3 in retroviral silencing. *Nature* 548:E1–E3
290. Navarro C, Lyu J, Katsori A-M, Caridha R, Elsässer SJ (2020) An embryonic stem cell-specific heterochromatin state promotes core histone exchange in the absence of DNA accessibility. *Nat Commun* 11:5095
291. Groh S, Milton AV, Marinelli LK, et al (2021) Morc3 silences endogenous retroviruses by enabling Daxx-mediated histone H3.3 incorporation. *Nat Commun* 12:5996
292. Sundaram V, Choudhary MNK, Pehrsson E, et al (2017) Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat Commun* 8:14550
293. Britten RJ, Davidson EH (1971) Repetitive and Non-Repetitive DNA Sequences and a Speculation on the Origins of Evolutionary Novelty. *Q Rev Biol* 46:111–138
294. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 2010 427 42:631–634
295. Fueyo R, Judd J, Feschotte C, Wysocka J (2022) Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol* 23:481–497
296. van de Lagemaat LN, Landry J-R, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 19:530–536
297. Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D (2013) TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* 23:452–461
298. Chuong EB, Elde NC, Feschotte C (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351:1083–1087
299. Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E (2020) Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc Natl Acad Sci* 117:7905–7916
300. Buttler CA, Chuong EB (2022) Emerging roles for endogenous retroviruses in immune epigenetic regulation*. *Immunol Rev* 305:165–178

301. Cañadas I, Thummalapalli R, Kim JW, et al (2018) Tumor innate immunity primed by specific interferon-stimulated endogenous retroviruses. *Nat Med* 24:1143–1150
302. Srinivasachar Badarinarayan S, Shcherbakova I, Langer S, Koepke L, Preising A, Hotter D, Kirchhoff F, Sparrer KMJ, Schotta G, Sauter D (2020) HIV-1 infection activates endogenous retroviral promoters regulating antiviral gene expression. *Nucleic Acids Res* 48:10890–10908
303. Rehwinkel J, Gack MU (2020) RIG-I-like receptors: their regulation and roles in RNA sensing. *Nat Rev Immunol* 20:537–551
304. Chiappinelli KB, Strissel PL, Desrichard A, et al (2016) Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* 164:1073
305. Srinivasachar Badarinarayan S, Sauter D (2021) Switching Sides: How Endogenous Retroviruses Protect Us from Viral Infections. *J Virol* 95:e02299-20
306. Schmidt N, Domingues P, Golebiowski F, Patzina C, Tatham MH, Hay RT, Hale BG (2019) An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc Natl Acad Sci* 116:17399–17408
307. Gehre M, Buccitelli C, Diaz N, Korbel J, Noh K-M (2019) Efficient strategies to detect genome editing and integrity in CRISPR-Cas9 engineered ESCs. *bioRxiv* 1–16
308. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8:2281–2308
309. Ma M, Zhuang F, Hu X, Wang B, Wen X-Z, Ji J-F, Xi JJ (2017) Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Res* 27:578–581
310. Savic N, Ringnalda FC, Lindsay H, et al (2018) Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *eLife*. <https://doi.org/10.7554/eLife.33761>
311. Aird EJ, Lovendahl KN, St. Martin A, Harris RS, Gordon WR (2018) Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun Biol* 1:54
312. Shan Y, Liang Z, Xing Q, et al (2017) PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. *Nat Commun* 8:1–14
313. Grosswendt S, Kretzmer H, Smith ZD, Kumar AS, Hetzel S, Wittler L, Klages S, Timmermann B, Mukherji S, Meissner A (2020) Epigenetic regulator function through mouse gastrulation. *Nature* 584:102–108
314. Zou Z, Ohta T, Miura F, Oki S (2022) ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res* 50:W175–W182
315. Van Oss SB, Cucinotta CE, Arndt KM (2017) Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. *Trends Biochem Sci* 42:788–798

316. Deng L, Ren R, Liu Z, et al (2019) Stabilizing heterochromatin by DGCR8 alleviates senescence and osteoarthritis. *Nat Commun* 10:3329
317. Kamal A, Arnold C, Claringbould A, et al (2022) GRaNIE and GRaNPA: Inference and evaluation of enhancer-mediated gene regulatory networks applied to study macrophages. 2021.12.18.473290
318. Bots M, Medema JP (2008) Serpins in T cell immunity. *J Leukoc Biol* 84:1238–1247
319. Dennis EA, Norris PC (2015) Eicosanoid storm in infection and inflammation. *Nat Rev Immunol* 15:511–523
320. Yoshida H, Lareau CA, Ramirez RN, et al (2019) The cis-Regulatory Atlas of the Mouse Immune System. *Cell* 176:897–912.e20
321. Le Blanc K, Mougiakakos D (2012) Multipotent mesenchymal stromal cells and the innate immune system. *Nat Rev Immunol* 12:383–396
322. Danko CG, Hyland SL, Core LJ, Martins AL, Waters CT, Lee HW, Cheung VG, Kraus WL, Lis JT, Siepel A (2015) Identification of active transcriptional regulatory elements from GRO-seq data. *Nat Methods* 12:433–438
323. Wang Z, Chu T, Choate LA, Danko CG (2019) Identification of regulatory elements from nascent transcription using dREG. *Genome Res* 29:293–303
324. Chu T, Rice EJ, Booth GT, et al (2018) Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat Genet* 50:1553–1564
325. Berest I, Arnold C, Reyes-Palomares A, et al (2019) Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Rep* 29:3147–3159.e12
326. Zhang Y, Gao S, Xia J, Liu F (2018) Hematopoietic Hierarchy – An Updated Roadmap. *Trends Cell Biol* 28:976–986
327. Miller SA, Damle M, Kingston RE (2020) H3K27me3 is dispensable for early differentiation but required to maintain differentiated cell identity. 2020.06.27.175612
328. Penke TJR, McKay DJ, Strahl BD, Matera AG, Duronio RJ (2016) Direct interrogation of the role of H3K9 in metazoan heterochromatin function. *Genes Dev* 30:1866–1880
329. Penke TJR, McKay DJ, Strahl BD, Matera AG, Duronio RJ (2018) Functional Redundancy of Variant and Canonical Histone H3 Lysine 9 Modification in *Drosophila*. *Genetics* 208:229–244
330. Su I -hsin, Basavaraj A, Krutchinsky AN, Hobert O, Ullrich A, Chait BT, Tarakhovsky A (2003) Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat Immunol* 4:124–131
331. Burr ML, Sparbier CE, Chan KL, et al (2019) An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer Cell* 36:385–401.e8

332. Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD (2019) Epigenetic therapy in immune-oncology. *Nat Rev Cancer* 19:151–161
333. Jansz N, Faulkner GJ (2021) Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol* 2021 22:1–22
334. Onomoto K, Onoguchi K, Yoneyama M (2021) Regulation of RIG-I-like receptor-mediated signaling: interaction between host and viral factors. *Cell Mol Immunol* 18:539–555
335. Rawling DC, Jagdmann GEJr, Potapova O, Pyle AM (2020) Small-Molecule Antagonists of the RIG-I Innate Immune Receptor. *ACS Chem Biol* 15:311–317
336. O'Hara R, Banaszynski LA (2022) Loss of heterochromatin at endogenous retroviruses creates competition for transcription factor binding. 2022.04.28.489907
337. Liang X, Potter J, Kumar S, et al (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* 208:44–53
338. Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Babraham Bioinforma <http://www.bioinformatics.babraham.ac.uk/projects/>
339. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359
340. Afgan E, Baker D, Batut B, et al (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46:W537–W544
341. Sidoli S, Bhanu NV, Karch KR, Wang X, Garcia BA (2016) Complete workflow for analysis of histone post-translational modifications using bottom-up mass spectrometry: From histone extraction to data analysis. *J Vis Exp*. <https://doi.org/10.3791/54112>
342. Yuan ZF, Sidoli S, Marchione DM, Simithy J, Janssen KA, Szurgot MR, Garcia BA (2018) EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *J Proteome Res* 17:2533–2541
343. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26:966–968
344. Zhang Q, McKenzie NJ, Warneford-Thomson R, et al (2019) RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat Struct Mol Biol* 26:237–247
345. Zhang Q, Agius SC, Flanigan SF, Uckelmann M, Levina V, Owen BM, Davidovich C (2021) PALI1 facilitates DNA and nucleosome binding by PRC2 and triggers an allosteric activation of catalysis. *Nat Commun* 12:4592
346. Martinez AM, Kim A, Yang WS (2020) Detection of Ferroptosis by BODIPY™ 581/591 C11. In: Vancurova I, Zhu Y (eds) *Immune Mediat. Cancer Methods Protoc*. Springer US, New York, NY, pp 125–130

347. Bibel M, Richter J, Lacroix E, Barde YA (2007) Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* 2:1034–1043
348. Takahashi T, Lord B, Schulze PC, Fryer RM, Sarang SS, Gullans SR, Lee RT (2003) Ascorbic Acid Enhances Differentiation of Embryonic Stem Cells Into Cardiac Myocytes. *Circulation* 107:1912–1916
349. Wang Y, Li Y, Guo C, Lu Q, Wang W, Jia Z, Chen P, Ma K, Reinberg D, Zhou C (2016) ISL1 and JMJD3 synergistically control cardiac differentiation of embryonic stem cells. *Nucleic Acids Res* 44:6741–6755
350. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
351. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930
352. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44:W160–W165
353. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550
354. Conway JR, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33:2938–2940
355. Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J Integr Biol* 16:284–287
356. Wu T, Hu E, Xu S, et al (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2:100141
357. Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N (2014) Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* 15:1–17
358. Zhang Y, Liu T, Meyer CA, et al (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137
359. Feng J, Liu T, Qin B, Zhang Y, Liu XS (2012) Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 7:1728–1740
360. Histone ChIP-seq Data Standards and Processing Pipeline – ENCODE. <https://www.encodeproject.org/chip-seq/histone/>.
361. Stark R, Brown G (2022) DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data. <https://doi.org/10.18129/B9.bioc.DiffBind>
362. Yu G, Wang L-G, He Q-Y (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31:2382–2383

- 363. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ (2013) Software for Computing and Annotating Genomic Ranges. *PLOS Comput Biol* 9:e1003118
- 364. Bunina D, Abazova N, Diaz N, Noh K-M, Krijgsveld J, Zaugg JB (2020) Genomic Rewiring of SOX2 Chromatin Interaction Network during Differentiation of ESCs to Postmitotic Neurons. *Cell Syst* 10:480-494.e8
- 365. Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT (2016) Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* 11:1455–1476