Aus dem Adolf-Butenandt-Institut
Institut der Ludwig-Maximilians-Universität München
Lehrstuhl Molekularbiologie im Biomedizinischen Centrum (BMC)
Vorstand: Prof. Dr. rer. nat. Peter B. Becker

Deciphering the regulatory principles of histone H3K36 methylation in *Drosophila*

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Muhunden Nallappa Jayakrishnan

aus
Erode (Indien)

2025

Mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuer:	Prof. Dr. Peter B. Becker
Zweitgutachter:	Prof. Dr. Robert Schneider
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	04. August 2025

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dekanat Medizinische Fakultät
Promotionsbüro

Eidesstattliche Versicherung

Jayakrishnan, Muhunden Nallappa

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

Deciphering the regulatory principles of histone H3K36 methylation in Drosophila

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 14.08.2025

Ort, Datum

Muhunden Jayakrishnan

Unterschrift Doktorandin/Doktorand

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dekanat Medizinische Fakultät
Promotionsbüro

Erklärung zur Übereinstimmung der gebundenen Ausgabe der Dissertation mit der elektronischen Fassung

Jayakrishnan, Muhunden Nallappa

Name, Vorname

Hiermit erkläre ich, dass die elektronische Version der eingereichten Dissertation mit dem Titel:

Deciphering the regulatory principles of histone H3K36 methylation in Drosophila

in Inhalt und Formatierung mit den gedruckten und gebundenen Exemplaren übereinstimmt.

München, 14.08.2025

Ort, Datum

Muhunden Jayakrishnan

Unterschrift Doktorandin/Doktorand

There is no law except the law that there is no law.

- John Archibald Wheeler

Preface

Part of this dissertation has been published in a research article, titled "**Genomic context-dependent histone H3K36 methylation by three *Drosophila* methyltransferases and implications for dedicated chromatin readers**", in Nucleic Acid Research with PMID 38813825. This includes most of the data presented in the results. The corresponding figures reproduced here are in accordance with the Creative Common BY License and are attributed properly throughout. I performed all the experiments and the bioinformatic analyses except the Microscale Thermophoresis data, which was provided by our collaborators Magdalena Havlová and Václav Veverka at the Charles University, Prague, Czech Republic.

I. Table of Contents

Deciphering the regulatory principles of histone H3K36 methylation in Drosophila

II. List of Figures

III. Abbreviations

ac	Acetylation
ATAC-seq	Assay for Transposase-Accessible Chromatin sequencing
ATP	adenosine triphosphate
°C	degree Celsius
ChIP-seq	Chromatin Immunoprecipitation sequencing
CUT&RUN	Cleavage under targets and release using nuclease
DAPI	4',6-diamidino-2-phenylindole
DCC	dosage compensation complex
(c)DNA	complementary DNA
DRB	5,6-dichloro-1-β-D-ribofuranosyl benzimidazole
EDTA	ethylenediaminetetraacetate
EGTA	ethyleneglycol-bis-(2-aminoethyl)-N,N,N,N-tetraacetate
ENCODE	Encyclopedia of DNA Elements
HDM	Histone Demethylase
Hi-C	'all-vs-all' chromosome conformation capture
HMT	Histone Methyltransferase
H2A/H2B/H3/H4	Histone H2A/H2B/H3/H4
IF	Immunofluorescence staining
K	Lysine
Kbp/Kb	Kilo-base pairs
KD	Knockdown
KMT	Lysine Methyltransferases
LLPS	Liquid-Liquid Phase Separation
me1/2/3	mono / di / tri -methylation
mESC	mouse Embryonic Stem Cells
MRE	MSL Recognition Elements
MSL	Male Specific Lethal
MNase	Micrococcal Nuclease
NGS	Next Generation Sequencing
NSL	Non-Specific Lethal
PBS	Phosphate Buffered Saline
PCH	Pericentric Heterochromatin
PCR	Polymerase Chain Reaction
PPPS	Polymer-Polymer Phase Separation
PTM	Post Translational Modification
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT	Room Temperature
RT-qPCR	Reverse Transcription quantitative PCR
SAM	S-Adenosyl methionine

SC	Supercluster
STARR-seq	Self-Transcribing Active Regulatory Region sequencing
TE	Transposable Elements
TPM	Transcripts per million
TSS	Transcription Start Sites
TTS	Transcription Termination Sites

IV. Summary

Histone Post Translational Modifications (PTMs) are dynamic signatures that play a central role in genome function by influencing chromatin organization and transcriptional regulation. Their precise deposition and regulation are critical for development, and their misregulation is frequently linked to disease. While many histone modifications have been extensively studied, a major challenge remains in understanding how they interact within the chromatin landscape to influence gene activity and other biological processes.

Methylation of histone H3 at lysine 36 (H3K36me1/2/3) has been implicated in transcriptional regulation and chromatin organization, with K36me3 in particular correlating with active transcription. This modification is highly conserved across species, accumulating along transcribed gene bodies. Despite this conservation, the precise relative distribution of the K36me1-3 states and their function remain unclear in certain model organisms such as *Drosophila*. Additionally, while the enzymes responsible for K36 methylation have been identified, their respective contributions to the establishment of K36me1, K36me2, and K36me3 remain debated.

To address these gaps, ChIP-seq protocols were systematically optimized to define the genomic landscapes of K36me1/2/3 in *Drosophila* which revealed that each methylation state preferentially associates with distinct chromatin environments, highlighting their specialized roles. A key unresolved question is how K36 methyltransferases (HMTs) Set2, NSD and Ash1 contribute to the deposition of these modifications. Using an RNAi approach, it was observed that each HMT predominantly catalyzes methylation in specific chromatin domains with minor overlaps: Set2 deposits K36me3 at highly transcribed genes; NSD catalyzes K36me2/3 at constitutive heterochromatin in addition to weakly expressed euchromatic genes, while Ash1 deposits K36me1 at regions with enhancer signatures. Further, despite functioning largely independent of each other, evidence suggests that HMTs may indirectly influence each other's activity and localization.

K36 methylation has been implicated in both transcriptional regulation and histone modification crosstalk. To investigate these roles, RNA-seq and ChIP-seq datasets were integrated to analyze transcriptional changes and potential interactions with other chromatin marks. The results revealed a poor correlation between K36me alterations and transcriptional changes, challenging assumptions about its direct role in gene regulation. However, all three K36 HMTs were found to counteract the facultative heterochromatic mark H3K27me3 at specific loci, highlighting a broader regulatory interplay between these modifications.

Finally, the functional consequences of K36 methylation often depend on reader proteins. While many of these readers have been characterized for K36me3, their affinities for K36me1/2 remain unexplored. Examining the impact of K36 HMT loss on two key readers, MSL3 and JASPer,

revealed that K36me2 can partially compensate for K36me3 loss to maintain reader binding. Additionally, disruption of MSL3 binding alters H4K16 acetylation distribution and the organization of the dosage-compensated X-chromosome territory.

Together, these findings provide new insights into the regulatory logic of H3K36 methylation and its functional interactions within the chromatin landscape.

V. Zusammenfassung

Posttranslationale Histonmodifikationen (PTM) sind dynamische chemische Signaturen, die eine zentrale Rolle bei der Genomfunktion spielen, indem sie die Chromatinorganisation und die Transkriptionsregulation beeinflussen. Ihre genaue Platzierung und Regulierung ist entscheidend für die Entwicklung eines Organismus, und ihre Fehlregulierung ist häufig mit Krankheiten verbunden. Obwohl viele Histonmodifikationen bereits eingehend untersucht wurden, besteht eine große Herausforderung darin zu verstehen, wie sie innerhalb entlang der Chromosomen zusammenwirken und die Genaktivität und andere biologische Prozesse beeinflussen.

Die Methylierung von Histon H3 an Lysin 36 (H3K36me1/2/3) wurde mit der Transkriptionsregulation und der Chromatinorganisation in Verbindung gebracht, wobei insbesondere K36me3 mit aktiver Transkription korreliert. Diese Modifikation ist innerhalb der Eukaryonten hoch konserviert und wird entlang der transkribierten Genkörper angereichert. Trotz dieser Konservierung bleiben die genaue Verteilung und Funktion der H3K36-Methylierung in bestimmten Modellorganismen wie *Drosophila* unklar. Darüber hinaus wurden zwar die für die K36-Methylierung verantwortlichen Enzyme identifiziert, ihr jeweiliger Beitrag zur Bildung von K36me1, K36me2 und K36me3 bleibt jedoch umstritten.

Um diese Lücken zu schließen, wurden in dieser Dissertation ChIP-seq-Protokolle systematisch optimiert, um die genomischen Profile von K36me1/2/3 in *Drosophila*-Zellen zu kartieren. Dabei zeigte sich, dass jeder Methylierungszustand präferenziell mit unterschiedlichen Chromatinumgebungen assoziiert ist, was ihre spezialisierte Rolle hervorhebt. Eine wichtige ungelöste Frage ist, wie die K36-Methyltransferasen (HMTs) Set2, NSD und Ash1 zur Platzierung dieser Modifikationen beitragen. Mithilfe eines RNAi-Ansatzes wurde festgestellt, dass jede HMT vorwiegend in spezifische Chromatinbereichen methyliert, wobei es zu geringen Überschneidungen kommt: Set2 lagert K36me3 an stark transkribierten Genen ab; NSD katalysiert K36me2/3 an konstitutivem Heterochromatin und an schwach exprimierten euchromatischen Genen, während Ash1 K36me1 an Regionen mit Enhancer-Signaturen platziert. Obwohl die HMTs weitgehend unabhängig voneinander funktionieren, finden sich Hinweise, dass sie sich gegenseitig in ihrer Aktivität und Lokalisierung beeinflussen können.

Die K36-Methylierung wird sowohl mit der Transkriptionsregulierung als auch mit Wechselbeziehungen zwischen verschiedenen Histonmodifikationen in Verbindung gebracht. Es wurden RNA-seq- und ChIP-seq-Datensätze integriert, um Transkriptionsänderungen und potenzielle Einflüsse auf andere Chromatinmarkierungen zu untersuchen. Die Ergebnisse zeigten eine schwache Korrelation zwischen K36me-Veränderungen und geänderter Transkriptionsaktivität, was Annahmen über seine direkte Rolle bei der Genregulation in Frage stellt. Es wurde jedoch festgestellt, dass alle drei K36-HMTs der Methylierung des H3K27 an bestimmten Genorten entgegenwirken, einer Markierung des fakultativen Heterochromatins, was auf ein breiteres regulatorisches Zusammenspiel zwischen diesen Modifikationen hinweist.

Schließlich hängen die funktionellen Folgen der K36-Methylierung häufig von ‚Reader'-Proteinen ab. Während einige dieser Leseproteine für K36me3 schon charakterisiert wurden, sind ihre Affinitäten für K36me1/2 noch unerforscht. Die Untersuchung der Auswirkungen des Verlusts von K36 HMT auf zwei wichtige Leseproteine, MSL3 und JASPer, ergab, dass K36me2 den Verlust von K36me3 teilweise kompensieren kann, und so zur stabilen Rekrutierung an K36-methylierte Bereiche beiträgt. Darüber hinaus verändert die Störung der MSL3-Bindung die Verteilung der H4K16-Acetylierung und die Organisation des dosis-kompensierten X-Chromosomen-Gebiets.

Zusammengenommen bieten diese Ergebnisse neue Einblicke in die regulatorische Logik der H3K36-Methylierung und ihre funktionellen Interaktionen im Chromatin von *Drosophila*.

1. Introduction

In the 1870s, humanity's gaze was fixed on the cosmos. Peering into telescopes, astronomers had observed the Great White Spot, a vast and raging storm on Saturn for the first time. Meanwhile, Mars had just unveiled its long-hidden companions, Deimos and Phobos, as they were brought to light by human ingenuity. The Great Comet of 1874, a dazzling celestial visitor, captured the imagination of astronomers and the public alike. Observatories around the world mobilized to study this brilliant spectacle, which streaked across the night sky as a reminder of the vastness of the universe. It was a time when science was grappling with phenomena on a grand, almost incomprehensible scale.

But not all frontiers of discovery were vast. In a quiet laboratory, far removed from the sweeping motions of the heavens, Walther Flemming was peering into the microscopic world of living cells. Using recently developed aniline dyes, he uncovered a mysterious substance within the nucleus that absorbed color intensely, revealing itself with striking clarity under the microscope. Flemming called this material "chromatin," from the Greek *chroma*, meaning color.

While the comet's brilliant tail spanned millions of kilometers, chromatin existed on a scale almost unimaginably small—measured in mere nanometers. While astronomers marveled at the vastness of space and the distant moons of Mars, Flemming's work revealed a universe of its own within the confines of a single cell. Yet this microscopic material held secrets of life itself, carrying the genetic and epigenetic instructions that guide the development and function of every organism. Flemming's discovery marked the beginning of a journey to understand the architecture of life at its smallest scales, just as the comet inspired awe for the universe at its largest.

1.1 Hierarchical organization of chromatin in the nucleus

Chromatin is a highly organized and dynamic complex of DNA and proteins that plays a central role in the regulation of gene expression, DNA replication, and genome integrity within eukaryotic cells. This complex structure allows the vast amount of genetic information stored in DNA to be efficiently packaged into the confined space of the cell nucleus, while also ensuring that specific regions of the genome remain accessible to DNA-templated processes needed for different cellular processes.

At the most basic level, chromatin is composed of DNA wrapped around histone proteins to form nucleosomes, which are the fundamental repeating units of chromatin. Each nucleosome core particle consists of approximately 146 base pairs of DNA coiled around an octamer of histone proteins, comprising two copies each of histones H2A, H2B, H3, and H4. The DNA-histone interaction is stabilized by the electrostatic attraction between the negatively charged DNA backbone and the positively charged residues on the histones.

The nucleosome itself is not the terminal structure; rather, it forms higher-order structures through further interactions and folding (Figure 1).

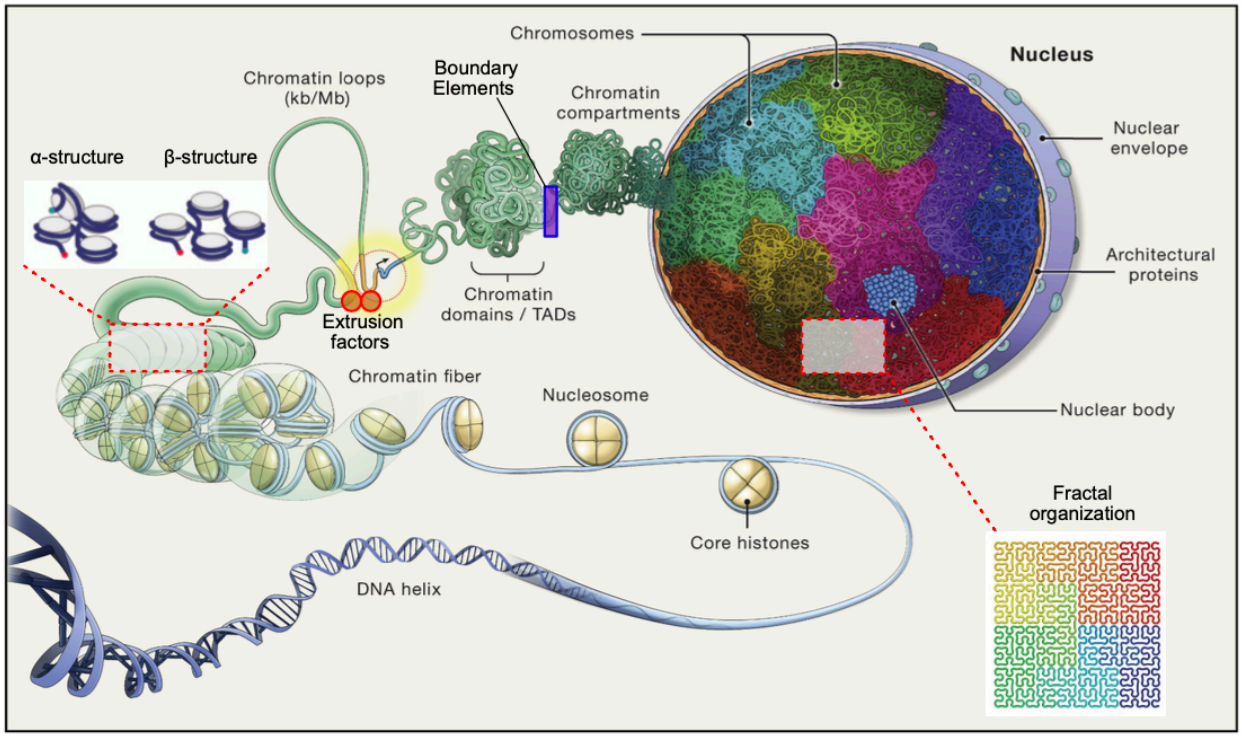


Figure 1. Hierarchical organization of chromatin across multiple length-scales: Adapted from Misteli, 2020.

Nucleosomes are connected by linker DNA, which can be bound by histone H1 to form the chromatosome. These nucleosomal arrays can then stack into structures analogous to protein secondary structures, defined by two characteristic structural motifs existing at the tetranucleosome level; a tetrahedral configuration and a rhomboid configuration, which are distributed nonrandomly throughout the genome (Ohno et al. 2019). Interactions at this level dictate the local stiffness / persistence length of chromatin (Cui and Bustamante 2000). Beyond this length-scale, chromatin forms higher-order structures at scales of thousands and even millions of base pairs, such as flexible loops and self-associating topological domains (TADs). These polymeric structures do not exist in isolation, but fold iteratively onto each other over multiple length scales to form a ‘fractal globule’, representing entire chromosomal territories (Lieberman-Aiden et al. 2009; Mirny 2011). While the ‘fractal’ collapsing of nucleosomal polymer onto itself is non-specific and random in nature, order is introduced by other chromatin-associated factors by several mechanisms that exist at different length scales. For example, molecular motors can bind and extrude chromatin until they encounter ‘boundary elements’ defined by specific DNA sequences. Order can also be introduced by phase separation of the nucleosome polymer, where specific factors bridging different chromatin molecules can facilitate the segregation from the nucleoplasmic mixture to form a more concentrated phase, resulting in nuclear subcompartments with defined molecular composition (Erdel and Rippe 2018). This

intricate layering of organization not only facilitates the dynamic interactions of chromatin but also underscores the complexity of its structural architecture across multiple length scales.

1.2 Nucleosome landscape is dynamic

The nucleosome landscape is not static but continuously undergoing dynamic fluctuations. The DNA wrapped around nucleosomes can be transiently unwrapped and the histone octamers themselves can be physically slid or evicted from DNA. Further, nucleosomes can be also modified in various ways; by incorporation of histone variants, by selective removal of a single copy of the H2A-H2B dimer to generate hexasomes or by covalent post-translational modifications (PTMs) of the histone proteins (Figure 2). The resulting heterogeneity of chromatin is due to the concerted effort of a wide array of chromatin modifying proteins, primarily ATP-driven nucleosome remodelers as well as epigenetic 'modifier' enzymes (more discussed in next section).

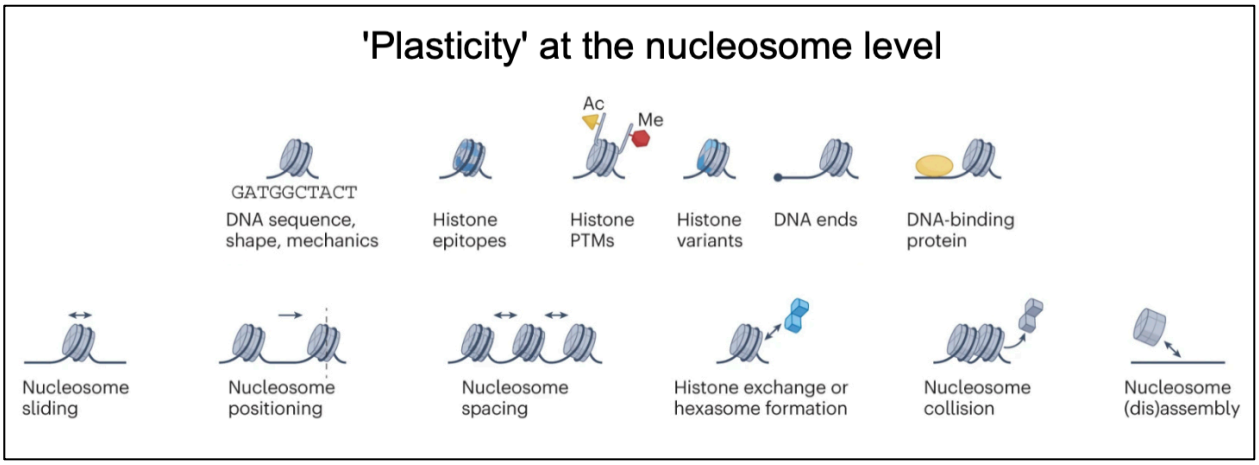


Figure 2. Different (but interconnected) mechanisms by which nucleosomes can affect chromatin structure and function. Features shown in the top row can act as inputs to enzymes which can result in physically remodelled chromatin (bottom row). Adapted from Eustermann, Patel et al., 2024.

These activities ultimately result in generation of local stretches of chromatin which vary in 'accessibility', allowing specific regions to be more or less permissive to the transcriptional machinery and regulatory factors. Regions of chromatin that are less compacted and more accessible to RNA polymerases are known as euchromatin. Conversely, heterochromatin is more densely packed and generally transcriptionally inactive, often associated with structural proteins that maintain its compact state. The features of chromatin discussed above are highly pliable and adopt different states as cells traverse the Waddington landscape —a conceptual model depicting changes in cell states as a marble rolling down a free-energy landscape of bifurcating valleys, where each path represents a possible cell fate— during normal cellular processes such as cell differentiation as well as during abnormalities like cancer.

1.3 Chromatin post-translational modifications: Foundations of the ‘histone code’ hypothesis

One of the key mechanisms discussed in the context of defining ‘active’ and ‘inactive’ states of chromatin are histone PTMs. Histones can be covalently modified by the addition of various post-translational modifications, which is a key mechanism of epigenetics – a layer of regulatory information above the DNA sequence. The presence of these modifications, many of which are evolutionary conserved in eukaryotes, orchestrate a wide range of chromatin-related processes. Studies spanning several decades have characterized over 25 different types of modifications occurring at more than 125 amino acids located on the canonical histones (Figure 3).

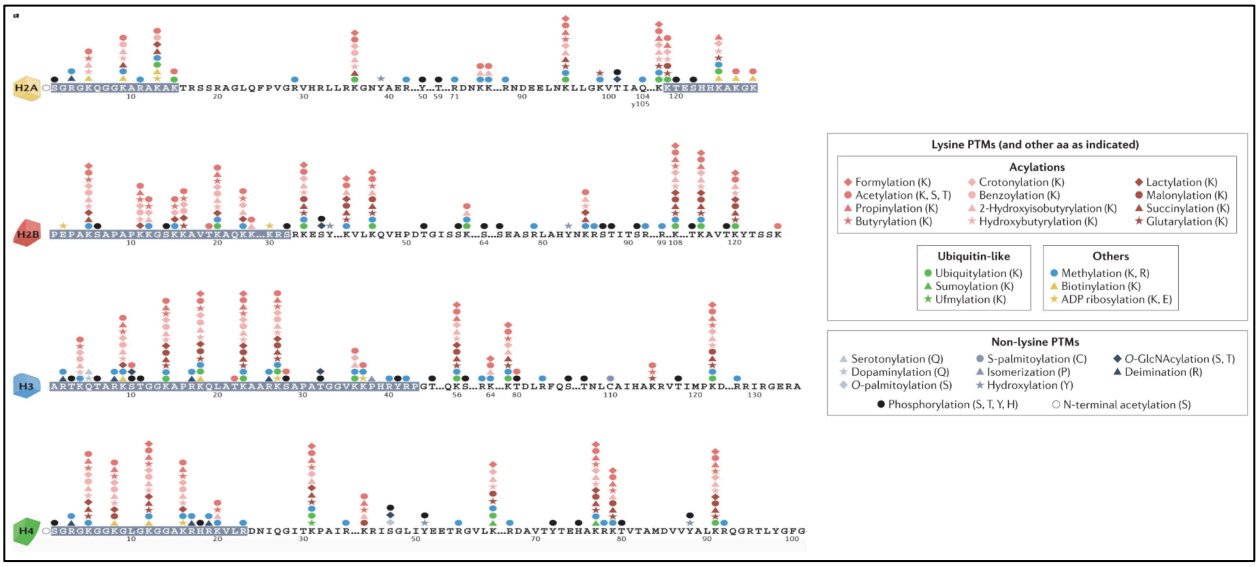


Figure 3: A comprehensive summary of known histone PTMs on the four canonical histones. Adapted from Milan-Zambrano, Burton et al., 2022.

While a subset of these modifications, particularly lysine acetylations and methylations, have been well characterized, the processes underlying regulation as well as the function of the majority of these modifications are still poorly understood. Importantly, these modifications do not occur in isolation but rather in particular combinations depending on the cellular and genomic context, leading to the idea of the existence of a ‘histone code’ dictionary (Jenuwein and Allis 2001). Proponents of the ‘histone code’ hypothesis argue that all the information regarding the current cellular state is coded within the histone modification profile. In support of this idea, recent studies combining systems biology and explainable artificial intelligence approaches have demonstrated that histone code by itself is self-sufficient to accurately predict the cell identity (Xia et al. 2020). In simpler words, a person perusing through a complete histone code dictionary should be able to exactly pinpoint the state of the cell based on a snapshot of the cell’s histone modification makeup (Figure 4).

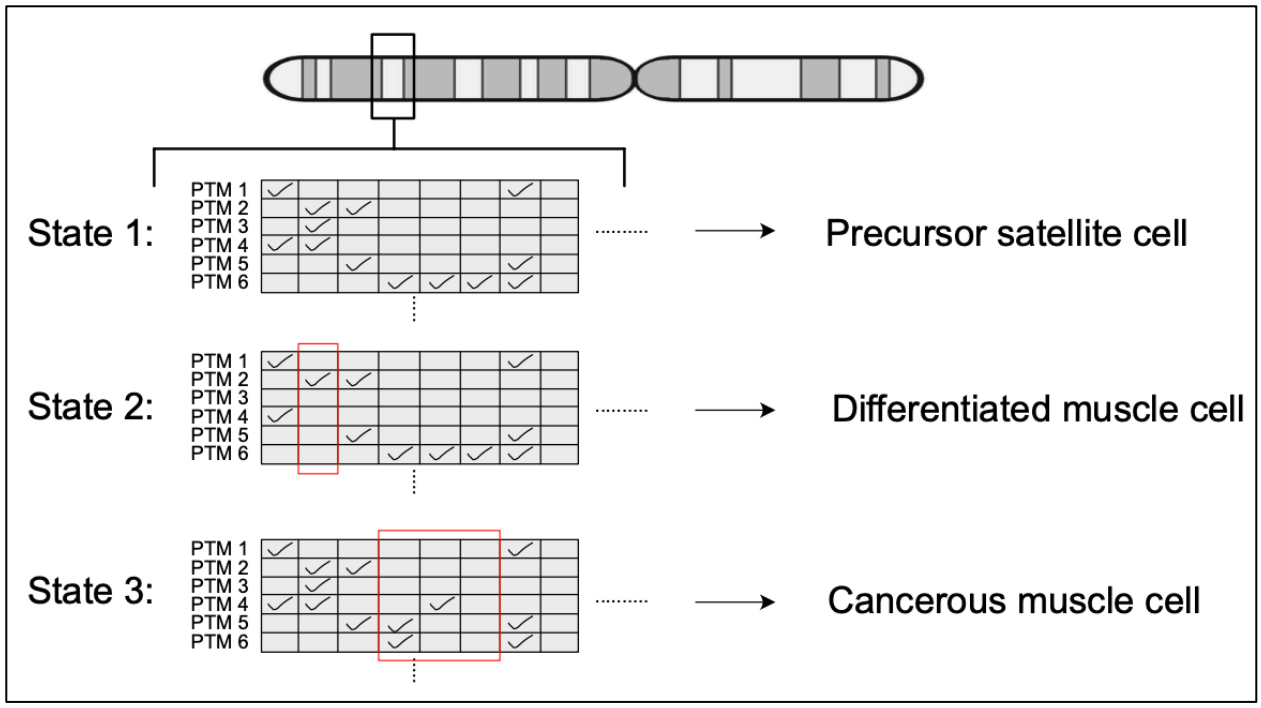


Figure 4: A schematic illustrating a hypothetical 'histone code' dictionary depicting PTM signatures as a function of genomic coordinates. Red boxes highlight the differences between a given cell state from the parent one. Chromosome schematic was generated using Biorender.com

This histone code is determined by a dynamic interplay between enzymes that add the modifications ('writers'), remove the marks ('erasers') and proteins that interact with the marks ('readers'), which are discussed in greater detail in subsequent sections.

1.4 *Drosophila* as a model organism to study chromatin

Before delving into the universe of histone PTMs, a case must be made for the model system used in this study, the mighty fruit fly. Fruit flies (*Drosophila melanogaster*) are a well-suited model system to study chromatin biology. Many fundamental principles of chromatin organization, such as the discovery of constitutive heterochromatin proteins, were originally elucidated in the fly (Tschiersch et al. 1994). A significant fraction of genes encoding the human epigenetic machinery, often deregulated in diseases, are conserved in flies (Chintapalli, Wang, and Dow 2007), making it an excellent model system. An extensive suite of genetic tools, such as balancer chromosomes, are available in flies which facilitates rapid and stable genetic manipulation. Perturbations of chromatin factors often result in distinct phenotypes such as modified eye color, wing morphology etc., which facilitated systematic screens. The small genome size of the fly (~160 million base pairs relative to 3.2 billion base pairs of the human genome, 20-fold difference), allows greatly reduced costs associated with Next Generation Sequencing (NGS) experiments, which are routinely employed in epigenetics. Despite the compact genome, flies share many features (such as 'enhancers', elements that activate gene transcription from a distance) with mammals allowing generalizability of findings. Importantly, paralogous genes in the mammalian genome introduce a

lot of redundancies, which can complicate systematic analysis of gene function. This is much less of an issue for the *Drosophila* genome with much fewer gene duplications.

1.5 Histone modification landscape in *Drosophila*: emergence of 'chromatin states'

In the 2000s, rapid improvements to (NGS) technologies lowered experimental costs and made it more accessible to labs worldwide. This enabled large-scale profiling of transcriptome (RNA-seq), DNA accessibility in chromatin (DNAse hypersensitivity-seq and ATAC-seq) and histone modifications status (ChIP-seq) in a tissue- and developmental state-specific manner across many different model organisms. These large datasets formed the basis for large research consortia attempting to decipher the regulatory code of chromatin structures, such as modENCODE for *Drosophila / C.elegans* and ENCODE for humans (Gerstein et al. 2010; Roy et al. 2010; ENCODE 2012). These datasets were then fed as sequence inputs to mathematical Hidden Markov Models (HMM), which collapse information to emit probabilistic HMM 'states', leading to identification of 'chromatin states' that demarcate the genome. Each chromatin state segments genomic regions that are marked by a relatively uniform combination of histone modifications (Figure 5).

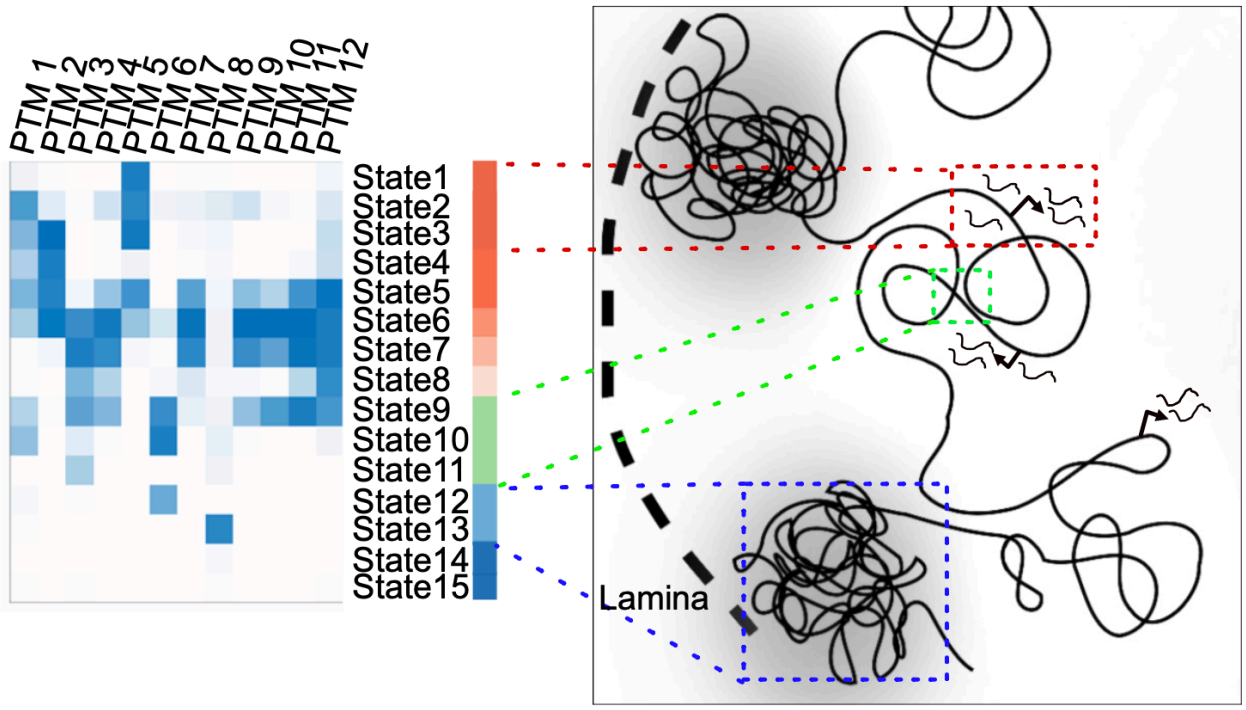


Figure 5: Illustration of the 'Chromatin States' concept. Each chromatin state is typified by a unique PTM signature. HMM models predict the most likely state for a given genomic region, which is strongly correlated to the spatial characteristics of the locus. Adapted and modified from Davis, 2023 as permitted by the Creative Commons Attribution CC-BY 4.0 License.

The seminal modENCODE *Drosophila* study used an HMM model that partitioned the genome into nine chromatin states with each state correlated to distinct transcriptional and spatial features (Kharchenko et al. 2011). State 1, which appeared to be marked by di-or trimethylation of histone H3 at lysine 4 (H3K4me2/3), was strongly correlated with gene promoters, while State 2 carried marks like H3K36me3 and correlated with gene transcription. State 3, consisting of acetylation of H3 at lysine 27 (H3K27ac), was defined as enhancers which occurred predominantly in introns and intergenic regions. In contrast to the previous 'active' states, numerous 'silent' states were also identified. State 6 was composed of H3K27me3, the same lysine described to be acetylated in State 3, and represented developmentally silenced genes, highlighting how different modifications on the same residue can result in different transcriptional outcomes. State 7 domains flanked the centromere and were heavily decorated with H3K9me3 while State 9 encompassed passive chromatin stretches which don't enrich for any prominent modification. It's important to note the chosen number of emission states was partly arbitrary and using an HMM with larger number of output 'chromatin states' results in identification of finer genomic regions, which include more peculiar and atypical combinations of histone modifications, such as co-occurring heterochromatic (H3K9me3) and euchromatic (H3K36me3) marks, which this thesis will demonstrate to be functionally relevant (see sections 3.3 and 3.12).
In addition to 'signature modifications', each chromatin state was correlated to numerous other histone modifications and chromatin factors. The description of chromatin states in the different model organisms sparked a flurry of scientific inquiry. Which protein factors contribute to the establishment of the chromatin states? What restricts the boundaries of these states? Are co-occurring modifications coexisting in single cells and are interdependent/coregulated? Are these marks merely correlated to the functional state (active or inactive) or are they causally implicated? How do chromatin states vary across development and disease? These questions laid the foundation for a new era of chromatin research, including the work described in this thesis.

1.6 Modifications of H3K36 and associated functions

An example of a heavily modified histone tail residue that will be prominent during the course of this thesis is the lysine residue located on the 36th position of the H3 tail (H3K36), referenced in the previous section as a gene body mark. As this thesis focuses on exploring the principles of K36 methylation, and therefore, it will introduce various histone modification concepts with a particular emphasis on K36 methylation.

This residue, which is proximal to the DNA entry/exit sites of the nucleosome, can be modified by adding of up to three methyl groups (mono-, di- or trimethyl) or an acetyl group. Among these, the trimethylated state (H3K36me3) is historically the most well studied modification. This mark correlates to active transcription and is enriched towards the Transcription Termination Site (TTS) in all eukaryotes (Kolasinska-Zwierz et al. 2009). This methylation is involved in signalling the transcribed state of chromatin to localize a wide variety of concerted processes that occur during transcriptional elongation. For instance, evidence for H3K36me3-mediated splicing regulation

comes from studies that altered the levels of H3K36me3 by either overexpression or downregulation of the corresponding methyltransferase, which resulted in splicing defects (Luco et al. 2010). H3K36me3 is also intimately linked to modulation of nucleosome turnover (LeRoy et al. 2019), partly due to crosstalk with deacetylation pathways (Kim et al. 2016; Martin et al. 2017) which has been suggested to be important for proper suppression of alternative/cryptic intragenic promoters (Carrozza et al. 2005). Interestingly, H3K36me3 has also been linked to methylation of nucleic acids. The mark recruits RNA methylation machinery to add *N*6-methyladenosine (m6A) to nascent RNAs (Huang et al. 2019) while also contributing to establishment of DNA methylation (Weinberg et al. 2019; Dukatz et al. 2019). Erosion of H3K36me3-mediated DNA methylation during aging of human cells leads to production of cryptic transcripts, again confirming that H3K36me3 plays a central role in multiple transcription-linked processes (McCauley et al. 2021). The presence of K36me3 on active regions of the region also appears to be critical to counteract spreading of repressive modifications (Schmitges et al. 2011; Alabert et al. 2020). Apart from being an important regulator of co-transcriptional processes, H3K36me3 is also linked to maintenance of genome integrity as evidenced by faulty DNA repair in cells lacking H3K36me3 (Jha and Strahl 2014; Carvalho et al. 2014; Daugaard et al. 2012).

Interestingly, recent studies have identified an additional function for H3K36 methylation, one that's radically different from aforementioned processes. Histone methylation, specifically H3K6me3 can act as a reservoir for the metabolite *S*-adenosyl methionine (SAM), which is an important cofactor for all methyltransferases. By removing methyl groups from the metabolic flux and thereby reducing intranuclear SAM concentrations, H3K36me3 can act as a 'buffer' against aberrant methylation. Interestingly, of all tested histone lysine methylations, H3K36 methylation appears to be the most prominent reservoir, which may be linked to affinity and the abundance of the corresponding methyltransferases (Ye et al. 2017; Ye and Tu 2018).

Historically, trimethylated H3K36 was the most prominently studied H3K36 modification while the di- and mono-methylation states were considered to be intermediates of the trimethylation reaction without any specific functions. However, over the years with the development of more sensitive approaches to investigate each mark, this view has gradually changed as researchers discovered that mono- and dimethylation of particular lysines may encode another layer of regulatory information. Depending on the specific case, these different methylation levels may have different functions and may reside in different types of chromatin. The earliest such example was described for H3K4 tri- and di-methylation where the tri-methylation was found exclusively on active genes while di-methylation was more widespread (Santos-Rosa et al. 2002). Only much later, subsequent studies discovered that H3K4me1 was present at enhancers, in contrast to H3K4me2/3 (Heintzman et al. 2007). Another example are the different methylation levels of H4K20, which are distributed in strikingly different parts of the genome. H4K20me1 is generally correlated to active chromatin and is broadly distributed throughout the genome. H4K20me2 is linked to DNA repair, while H4K20me3 marks repressed repetitive elements and transposons (Corvalan and Coller 2021). In contrast to the above examples, H3K9me1/2/3 are almost restricted to heterochromatin (Pinheiro et al. 2012).

The complexity of H3K36 methylation is akin to that of H4K20 or H3K4 methylation; mono-, di- or trimethylated H3K36 are found in distinct genomic compartments. Recently, H3K36me2 has received a lot of attention as it was found to be critical for establishment of DNA methylation in mammals (Weinberg et al. 2019). Importantly, H3K36me2 overrides H3K36me3 for recruitment of the DNA methylation machinery, suggesting a unique function for H3K36me2. H3K36me2 has also been implicated in other silencing pathways such as constitutive heterochromatin (Lindehell et al. 2021), and it may have specialized functions at certain active regulatory regions (Sun et al. 2023).

The functions of H3K36me1 and H3K36 acetylation are much less understood. H3K36me1 predominantly co-occurs with H3K36me3 in murine cells, while it is distributed differently from H3K36me3 in *Drosophila* to form its own chromatin state (Shipman et al. 2024; Kharchenko et al. 2011). For K36ac, *in vivo* profiling of the genome-wide distribution of mark in yeast shows that the mark is antagonistic to K36me; H3K36ac is enriched at promoters of active genes while K36me is enriched on gene bodies. Further, in yeast (and presumably in mammals as well), the mark is deposited by the Gcn5-containing SAGA complex (Pai et al. 2014; Morris et al. 2007). It's unclear whether H3K36ac merely counteracts H3K36me or if it recruits specific chromatin readers to regulate transcription and is a topic of future studies.

1.7 Establishing the chromatin modification landscape: ‘writers’ and ‘erasers’ of the histone code

1.7.1 Contribution of writers to complexity of chromatin states

Histone modifications are catalyzed by families of histone ‘writer’ enzymes. Commonly studied families include acetyltransferases (such as GCN5 mentioned previously), which add acetyl groups, while histone methyltransferases (henceforth abbreviated as HMTs) add methyl groups. In addition to their catalytic domains, writer enzymes also carry additional functional domains to facilitate interactions with DNA/RNA/proteins to modulate the transcriptional output.

For each histone lysine residue, multiple enzymes possess the ability to add a varying number of methyl groups which is dictated by critical residues present within the catalytic domains (Zhang et al. 2003). This apparent enzymatic ‘redundancy’ allows context-dependent activities (such as depositing methylation at a promoter versus an enhancer) as well as selective formation of domains consisting of different methylation statuses (me1, me2, me3).

A number of different factors contribute to the complexity arising from the concerted activities of the different HMTs acting on the same lysine. For instance, the differences in genomic distribution as well as the relative abundances of the different methylation levels (-me1/2/3) are strongly linked to the enzymatic properties of the corresponding methyltransferases. From an enzyme kinetics perspective, epigenetic writers can be broadly grouped into distributive or processive enzymes. Distributive enzymes, like the H3K79 HMT Dot1, go through multiple rounds of association and

dissociation to their substrate ('hit-and-run' mechanism), resulting in generation of intermediate modification states (Frederiks et al. 2008; Stulemeijer et al. 2015). This may be attributed to lower affinity of the enzyme to the partly modified nucleosomes as well as the structural necessity to dissociate to allow shuttling of the new SAM molecule to the active site. In contrast, processive enzymes, like the H3K9 HMT G9a, catalyze the addition of multiple modification groups within a single binding event, resulting in only the appearance of the final state without the intermediate levels (Patnaik et al. 2004). Evidently, in case of the distributive enzymes as well as processive enzymes specialized to generate only -me1/2, the intermediates can serve as substrates for other possibly distinct methyltransferases, allowing complex interdependencies (consider the example of H4K20 methylation; reviewed in Husmann & Gozani, 2019)

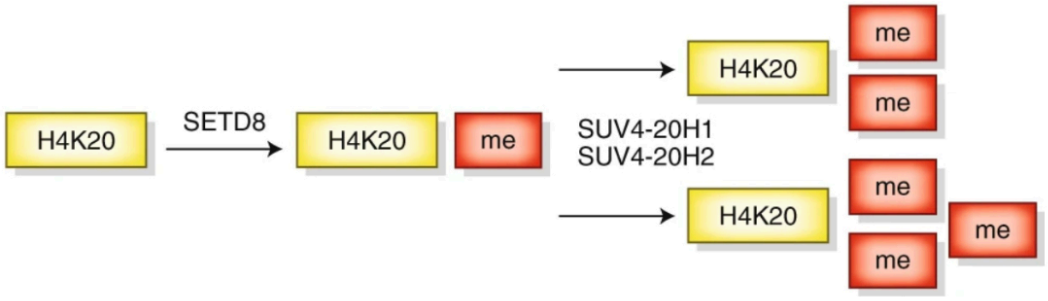


Figure 6: H4K20 methylation as an example for complex coordination between multiple HMTs. Adapted from Husmann & Gozani, 2019.

1.7.2 H3K36 histone methyltransferases in mammals and flies

Several writers of K36 methylation have been described to date, but their interdependencies as well as their contributions to K36me1/2/3 are still debated due to conflicting observations, often showing variable behaviour depending on the model and experimental system.

In mammals, SETD2 is assumed to be the sole trimethyltransferase while NSD1-3 and ASH1L contribute to K36me1/2 (Figure 7, reviewed in Husmann & Gozani, 2019). Other putative HMTs include SETD3 and SETMAR which may methylate histones in certain developmental contexts. However, more recent studies have shown that the specificity of the HMTs can be regulated in a context-dependent manner, evidenced by an *in vivo* pool of NSD1-3-dependent (and SETD2-independent) K36me3 at certain regulatory elements (Barral et al. 2022). These observations are at odds with numerous *in vitro* assays showing no trimethylation activity by the NSD family. This suggests that tissue-specific cofactors may be able to tune the methylation potential of HMTs.

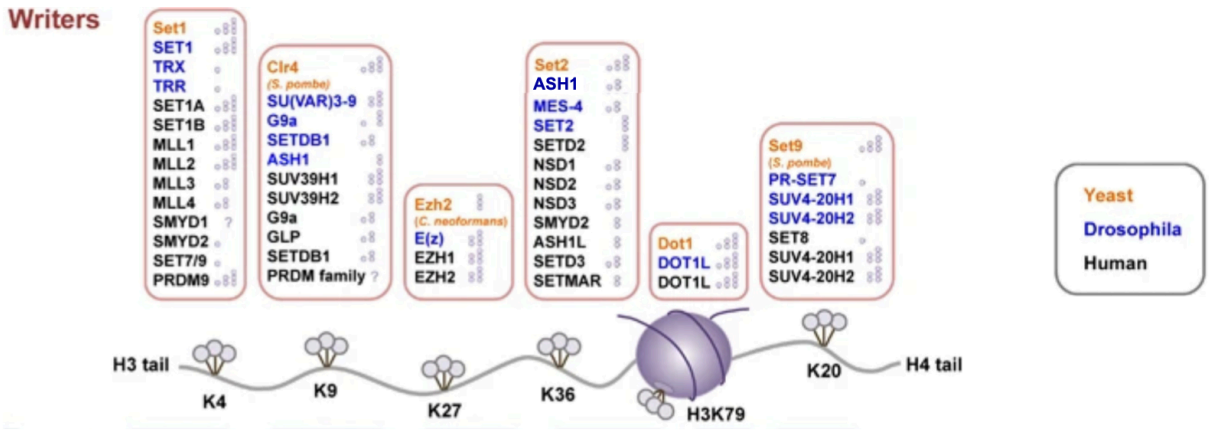


Figure 7: Histone H3 methyltransferases in three model organisms. The bars next to the names summarize the known product specificities of the HMTs. Adapted and modified from Hyun et al., 2017 as permitted by the Creative Commons Attribution CC-BY-NC-SA 4.0 License.

In *Drosophila*, the redundancy is much lower with only three evolutionarily conserved HMTs, namely Set2, NSD (or dMes-4) and Ash1. Set2 (orthologous to mammalian SETD2) catalyzes K36me3 co-transcriptionally through its interaction with Ser2P-Pol II. NSD (orthologous to mammalian NSD1/2/3) has been characterized predominantly as the major mono- and dimethyltransferase (Bell et al. 2007). Ash1 (orthologous to mammalian ASH1L) is described as a specialized dimethyltransferase whose role is developmentally restricted to a few hundred genes to counteract Polycomb repression (Klymenko and Muller 2004; Tanaka et al. 2007; Huang et al. 2017). In line with the dominant model of cooperative function in *Drosophila*, NSD was proposed to act as a pioneer enzyme that first 'pre-modifies' nucleosomes with K36me1/2, which then serves as a substrate for Set2 to catalyze K36me3 during the passage of the RNA polymerase. According to this model, the two enzymes act in a linear pathway (Bell et al. 2007; Bell et al. 2008; Lhoumaud et al. 2014). Ash1 is postulated to function autonomously of the Set2-NSD tandem (Huang et al. 2017; Schmahling et al. 2018).

This simple model of NSD-Set2 pathway in *Drosophila* is at odds with several recent independent observations. As described before, K36me3, me2 and me1 are differently distributed, implying distinct functions for each state (Kharchenko et al. 2011; Lindehell et al. 2021). The transcriptional and developmental outcomes of *NSD* deletion differ significantly from those of Set2 deficiency (Lindehell et al. 2021; Dorafshan et al. 2019). Flies lacking Set2 die during late larval stage, while *NSD*-null flies remain fertile and viable, although with minor defects. This disparity suggests that the two enzymes do not function in a strictly linear pathway to produce the trimethylated state. Interpreting these organismal phenotypes is challenging due to potential cell type-specific and developmental variations. Further, mammalian SETD2 doesn't require K36me1/2 to function which deviates from the reported behaviour of the fly Set2 (Weinberg et al. 2019).

The enzymatic properties of K36 methyltransferases have not been systematically examined, but are very likely subject to regulation by other factors. For example, processivity of yeast Set2 can

be boosted by the presence of monoubiquitinated K123, and interactions with specific cofactors can alter the relative K36me2:K36me3 activity (reviewed in Molenaar & van Leeuwen, 2022). These findings indicate that simple *in vitro* methylation assays on nucleosomes are insufficient to fully characterize the contributions of HMTs to K36me1/2/3, which may be only apparent *in vivo*.

1.7.3 KMTs are not only HMTs!

If one assumes that HMT function is solely exerted through methylation of the corresponding histone lysine, then replacing this lysine with a non-modifiable amino acid should phenocopy HMT removal. Interestingly, some effects observed following the deletion of K36 HMTs have not been replicated in histone replacement studies using non-methylatable variants (McKay et al. 2015). The cryptic transcription phenotype commonly associated with K36me3 loss is not recapitulated in flies carrying a lysine-to-arginine substitution at H3K36 (K36R) (Meers et al. 2017). Likewise, while the *Set2* deletion causes imbalanced expression of X-linked genes in male flies, this imbalance is absent in H3.2K36R flies (Lindehell et al. 2021). Similarly, another study demonstrated that the excessive Polycomb repression of HOX genes seen with Ash1 loss does not occur in K36R flies (Dorafshan et al. 2019). These findings suggest that certain effects attributed to H3K36 methylation might actually stem from the methylation of proteins other than histones, reinforcing that histones represent only a subset of targets of these more general Lysine Methyltransferases (KMTs) (reviewed in Husmann & Gozani, 2019). Some of the known non-histone methylation targets of histones include components of the cytoskeleton, such as tubulin (Park et al. 2016).

1.7.4 Contribution of erasers to complexity of chromatin states

Epigenetic writers are in a constant state of tug-of-war with ‘erasers’, enzymes that specialize in removal of the histone modifications. Similar to writers, erasers also show varying substrate specificities and reversal potencies (Figure 8).

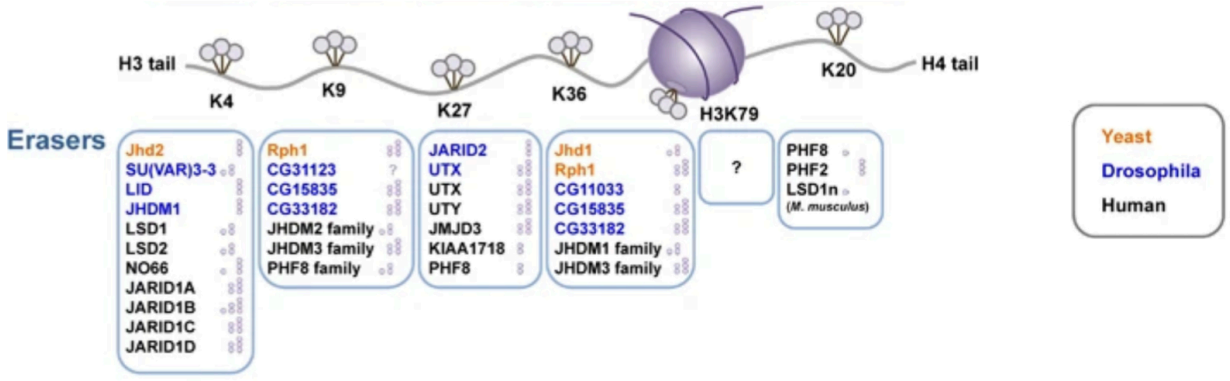


Figure 8: Histone H3 Demethylases in three model organisms. The bars next to the names summarize the known product specificities of the HDMs. Adapted from Hyun et al., 2017 as permitted by the Creative Commons Attribution CC-BY-NC-SA 4.0 License.

For example, LSD1, the first described demethylase, was shown to selectively act on H3K4me2 to progressively generate mono- and unmethylated states, but was not capable to destabilize H3K4me3 (Shi et al. 2004). In *Drosophila*, there are three proteins that show demethylase activity towards K36me, Kdm2 (ortholog of mammalian LSD1), Kdm4A and Kdm4B belonging to the Jumanji family (reviewed in Holowatyj et al., 2015). The enzymology of these proteins has been poorly characterized, but they likely differ in genomic localization and developmental functions. For instance, Kdm4A predominantly localizes in constitutive heterochromatin (Colmenares et al. 2017) while Kdm2 may have a function at counteracting Ash1 at Polycomb domains (Lagarou et al. 2008; Zheng et al. 2014).

1.8 Reading out the histone code: 'readers' of the histone code

Chromatin modifications can exert their functions either directly by affecting the physical properties of chromatin, or indirectly by specialized proteins that recognize the specific modification. The direct effect has been mostly described in the context of histone acetylation and phosphorylation, which weaken the stabilizing electrostatic interactions between positively charged histones and negatively charged DNA, leading to 'loosening' of chromatin. An often-discussed example in this context is the acetylation of the H4K16 residue (H4K16ac), a mark which is enriched on the X chromosome in male *Drosophila* melanogaster and is essential for male viability. Biochemical and computational studies have shown that this mark leads to reduced inter-nucleosomal interactions (Zhang, Erler, and Langowski 2017), which likely results in more efficient passage of the elongating RNA Pol II leading to transcriptional upregulation (Larschan et al. 2011).

In contrast, histone methylations generally do not strongly influence the underlying chromatin structure but rather serve as signals to recruit reader proteins. Reader proteins are typically modules within larger chromatin effector complexes and function by tethering epigenetic regulators to chromatin regions marked by specific combinations of histone modifications (Soshnev, Josefowicz, and Allis 2016). A wide arsenal of protein domains/motifs allow to recognize their target modifications (Figure 9A). Bromodomains and PHD fingers read out acetylated lysines (Marmorstein and Zhou 2014), while methylations are read out by PHD, WD40, chromo, tudor, PWWP and related domains (Hyun et al. 2017).

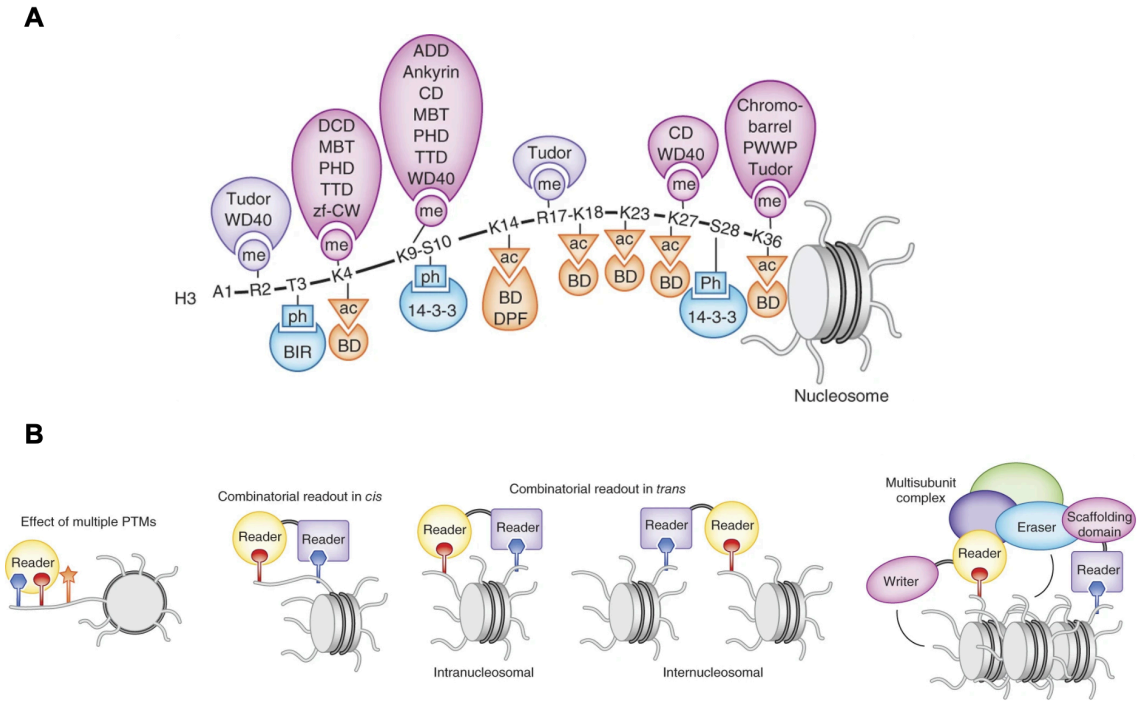


Figure 9: A. Schematic illustrating example protein domains that confer ability to 'read' specific PTMs. **B.** Possible modes of reader involvement in histone PTM crosstalk. Adapted from Musselman et al., 2012

1.8.1 Readers mediate histone-modification crosstalk

Some reader complexes can 'sense' multiple marks simultaneously, either on the same histone tail *in cis* or on neighbouring nucleosomes *in trans*. Such combinatorial readouts are essential for histone-modification crosstalk, where the presence of one mark can either antagonize or promote the deposition and sensing of nearby PTMs (Figure 9B).

A classic antagonistic example is the repulsion of the H3K9me3 reader HP1 by phosphorylation of the adjacent serine residue (H3S10ph). Similarly, recognition of K36me3 by the PRC2 reader PHF1 inactivates K27me3 deposition activity. On the other end of the spectrum, the affinity of H4K5ac to the reader Brdt is enhanced by the co-recognition of neighbouring H4K8ac (reviewed in Musselman et al., 2012).

The influence of histone modifications can extend beyond the nucleosome tail it is located on, primarily through effector complexes that have multiple readers. An example of this is the Rpd3s deacetylase complex which is targeted to gene bodies to repress spurious intragenic transcription (Carrozza et al. 2005). Remarkably, structural studies have shown that the resulting deacetylation patterns are dictated by the combination of modifications present on the nucleosome. Simultaneous recognition of two K36me3 marks (by Eaf3) orients the deacetylase subunit towards the histone H4 tail, while co-recognition of unmethylated H3K4 (by Rco1) and H3K36me3 within

a single copy of H3 promotes deacetylation of the H3 tail (Guan et al. 2023), illustrating how PTMs can exert their influence in trans through multi-subunit reader complexes.

1.8.2 K36me3 readers: Mixed bag of transcriptional outcomes

For K36me3, the resulting downstream effects can be either repressive or activating depending on the nature of the reader protein that is bound. For example, the previously mentioned Rpd3s complex globally directs a repressive network of factors to suppress cryptic transcription by deacetylation (Carrozza et al. 2005). On the other hand, the *Drosophila* Dosage Compensation Complex (DCC) is a quintessential activating complex containing the K36me3 reader MSL3, which positively regulates the expression of X-chromosomal genes in male flies (Figure 10) (Sural et al. 2008). In addition to the reader, the DCC also consists of the DNA binding protein MSL2, scaffolding protein MSL1, helicase MLE, the non-coding RNAs roX1/2 and the acetyltransferase MOF which imparts the 'activating' function. Interestingly, this complex provides an example where the search space of the reader is restricted by other subunits of the complex, in this case by the DNA binding protein MSL2, which selectively binds DNA motifs on the X-chromosome.

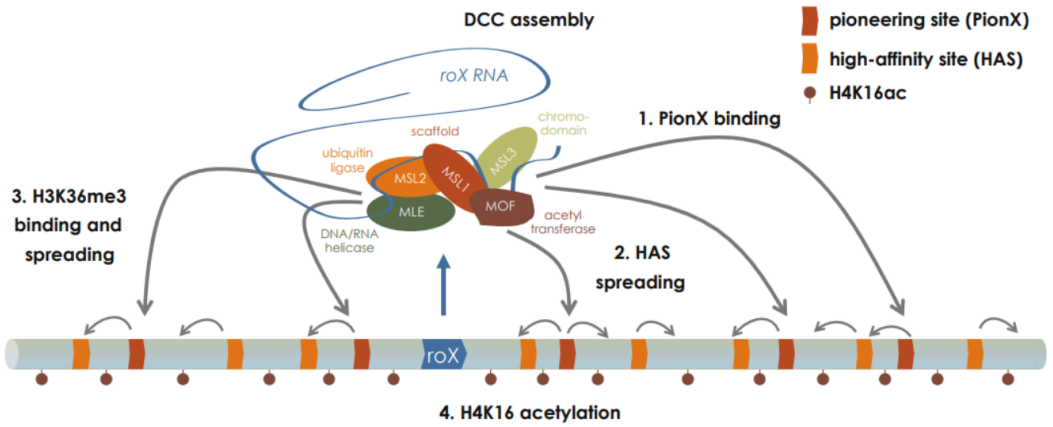


Figure 10: Hierarchy of Dosage compensation complex (DCC) function: Genetic targeting by MSL2 is followed by epigenetic 'spreading' mediated by MSL3 culminating in MOF-dependent H4K16ac. Reproduced with permission from Dr. Khairunnadiya Prayitno (Prayitno 2019)

Another example of a chromatin reader with multifaceted effects on transcription is JASPer (the fly ortholog of human LEDGF). In *Drosophila*, JASPer associates with JIL-1 kinase (JJ-complex) to assist the DCC in promoting X-linked gene expression, and it also plays independent roles on autosomes and heterochromatin (Albig et al. 2019). LEDGF is also involved in maintaining genome integrity via resolution of R-loops, a function likely conserved in JASPer (Jayakumar et al. 2024). JASPer's involvement in multiple chromatin pathways influences transcription in a highly context-dependent manner.

1.8.3 Sensitivity of readers to -me1, -me2, and -me3 states

A question that remains ambiguous in the field is the redundancy between mono-, di- and trimethylated states in recruiting a given reader. Not all known Kme3 readers were systematically investigated regarding their affinities to lower methylation states when they were originally discovered. Advancements in chemical synthesis of modified nucleosomes in conjunction with high throughput binding assays and mass spectrometry have contributed to our understanding regarding sensitivity of the binding pocket to methylation state. For example, barcoded designer nucleosomes containing a library of histone modifications can be exposed to a recombinant reader of choice resulting in a quantitative read out of reader affinity to all nucleosomes in the library (Dann et al. 2017). This barcoded approach has been utilized to show that DNMT3A preferentially binds H3K36me2 while DNMT3B prefers H3K36me3, further reinforcing the idea that each mark likely orchestrates two distinct but interdependent DNA methylation pathways (Hananya, Koren, and Muir 2024).

Other studies used a reverse approach, where a particular (di)nucleosome histone modification was exposed to crude cell extracts, thus interrogating the affinity of the entire proteome to a given modification (Lukauskas et al. 2024; Sankaran et al. 2016). Applying this approach to 55 dinucleosome variants, Lukauskas et al., catalogued a large-scale database of ~2000 chromatin readers, which provided insights into the extent of reader 'redundancies'. Interestingly, different lysine methylation readers showed different trends (Figure 11).

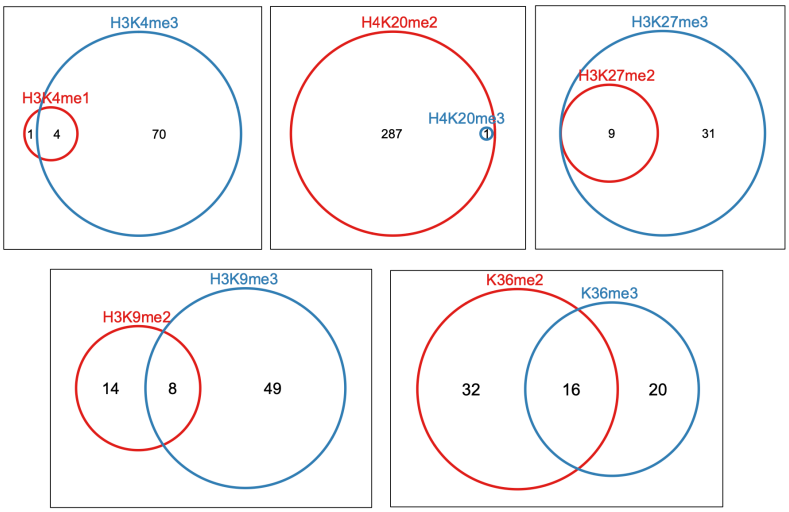


Figure 11: Overview of histone lysine methylation reader specificities. Data extracted from Lukauskas et al., 2024 and Sankaran et al., 2016.

For H3K27 methylation, many -me3 readers failed to recognize K27me2 and no protein showed selectivity towards the dimethylated state. On the other hand, several H3K9me readers appear to distinguish the di- and trimethylated state. Using a similar approach, Sankaran et al. demonstrate significant overlap between readers recruited by K36me2/3, but specific readers for each state

possibly exist. These observations hint at the possibility that reader redundancy might be coupled to the specific position of the lysine on the histone tail.

1.8.4 Physical effects of K36me on chromatin

It's worthwhile noting that while methylation may not have a strong physical effect on chromatin fiber organization in general, it may modulate the fine structure of intra-nucleosomal-DNA contacts. For example, a recent study using sensitive biophysical approaches showed that both H4K5/8/12/16ac as well as H3K36me3 nucleosomal arrays adopted looser chromatin configurations through different mechanisms; acetylated nucleosomes reduced inter-nucleosomal interactions while H3K36me3 increased nucleosomal breathing (i.e., transient unwrapping of DNA) within constituent mononucleosomes (Konrad, Vanderlinden, and Lipfert 2022; Lin et al. 2024). This increased nucleosomal breathing may have important implications for pioneer transcription factors to bind their target sequence effectively (Mondal, Felipe, and Kolomeisky 2023). Thus, methylations may also have a position-dependent physical effect on chromatin structure and transcription.

1.8.5 Histone Variants: Versatile twists on classic chromatin players

The final layer of complexity of chromatin structures is imparted by histone variants. While canonical histones are strongly conserved across different species, a lot more diversity has been observed in gene families encoding histone isoforms. Histone variants have highly controlled temporal and tissue-specific expression patterns, suggesting specialized roles. For example, the mammalian H2A variant H2A.X rapidly marks damaged chromatin, while variant H2A.B is exclusively expressed in brains and testes. A whole network of dedicated chaperones, such as HIRA for H3.3 and HJURP for CENPA, regulate the deposition and eviction of these variants. Importantly, some variants get incorporated during chromatin assembly of newly replicated DNA, while others can be exchanged in a replication-independent manner. The post-translational modification landscape of these variants is often unclear, but modifications have been reported on residues identical to those on canonical histones (such as methylation of H3K36 shared by H3.1-H3.3) as well as distinct ones (such as Ser31 phosphorylation present only on H3.3). A tantalizing possibility is the existence of readers that recognize particular histone modifications only on variant histones, supported by the discovery of H3.3K36me3-specific reader ZMYND11 which has roles in transcriptional repression (Wen et al. 2014).

Variants can also directly affect physical properties of the nucleosome, such as octamer stability as well as the flexibility of DNA ends. A comprehensive discussion of the roles of the different histone isoforms in mammalian genome regulation is beyond the scope of this thesis but is summarized extensively in recent reviews (Wong and Tremethick 2025; Martire and Banaszynski 2020). *Drosophila* has proven to be a very useful tool to study histone variants due to established methods for genetic replacement of the histone locus as described before (McKay et al. 2015). In the context of H3, three variants have been discovered (replication-dependent H3.2, replication-

independent H3.3 and centromeric variant Cid). These manipulations, which involve complete deletion of histone genes, have allowed to tease apart the impact of K36me3 on H3.2 versus H3.3. For instance, replacing H3.2K36 with non-methylatable K36R in mutants have revealed that ~80% of the bulk K36me3 is located on the H3.2 variant (Lindehell et al. 2021). While H3.3 appears to have a more specialized role in regulation of fly lifespan (Brown et al. 2024), gene repression (Salzler et al. 2023) and male germline developmental (Sakai et al. 2009), it may be functionally redundant with H3.2 in other pathways, like X-chromosome dosage compensation (Lindehell et al. 2021).

1.9 Scope of the study

This thesis aims to bridge several gaps in knowledge regarding the role of H3K36 methylation in *Drosophila*, focusing on its distribution, functional interactions, and impact on transcription and chromatin regulation (see graphical summary in Figure 12).

While the distribution of K36me1/2/3 in mammalian systems has been characterized relatively well, it is less clear in *Drosophila*. Comparison between existing datasets is complicated by large variations in (often suboptimal) protocols. To tackle this, I extensively optimize the ChIP-seq protocol to ask the following question: **Which genomic features are highlighted by each of K36me1, K36me2, and K36me3?**

Recent studies have provided conflicting results regarding the roles of H3K36 HMTs in K36me deposition. These studies frequently employed low resolution assays (like Western Blot or ChIP-qPCR), which may obscure the full complexity of HMT activity. Often, the function of a given HMT is studied in isolated contexts, which may miss effects such as HMT cross-regulation and how patterns depend on global variables such as transcription. To avoid this, I adopt a systems biology approach to probe HMT function in a unified experimental setting to explore the following: **How do Set2, NSD and Ash1 collaborate to shape the K36 methylation landscape in *Drosophila*? Are they governed by simple global rules (one enzyme - one function hypothesis) or do they demonstrate genomic context-dependent principles?**

Histone modifications do not exist in isolation but are often coupled to neighbouring modifications through positive or antagonistic interactions. This forms the basis of the next question: **Is there any evidence for context-dependent crosstalk of K36me1/2/3 with other histone modifications?**

While K36me3 has been described to correlate with active transcription, it's unclear if the mark directly promotes transcription. Further, how K36me1/2 affects transcription is unexplored. Previous studies have characterized the transcriptional defects occurring in HMT-null flies, however these observations weren't directly correlated with K36me1/2/3 changes. To this end, I

integrated RNA-seq and ChIP-seq datasets to address the following aim: **Are the transcriptional alterations emerging upon loss of K36 HMTs correlated to the altered patterns of K36me?**

For the major part, H3K36 methylation exerts its function through reader proteins. Many of these readers were originally described in the context of recognizing K36me3 but were not systematically investigated regarding their affinity towards K36me1/2. I tackle this problem by utilizing the RNAi system to quantify the behaviour of two readers, MSL3 (as part of the Dosage Compensation Complex) and JASPer (as part of the JJ-complex) in conjunction with orthogonal biochemical assays to evaluate the last question: **Are the functions of the DCC and the JJ-complex compromised in the absence of K36 HMTs? If yes, how and to what extent?**

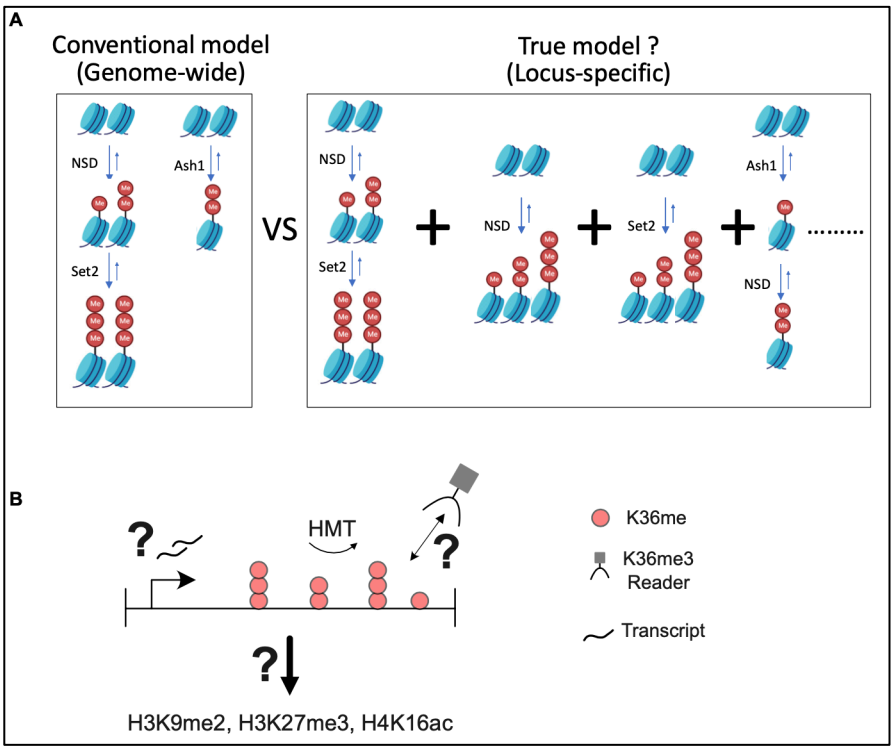


Figure 12: Illustration of the main aims of this thesis. **A.** Possible outcomes from proposed experiments regarding HMT contributions to K36me1/2/3. **B.** Exploration of potential roles for K36me in reader binding, histone modification crosstalk and transcription. Schematics were generated using Biorender.com

2. Materials and Methods

2.1 Primary Antibodies

Antibody (Rb = Rabbit; Ms = Mouse; Gp = Guinea Pig; Rt = Rat)	**Western**	**IF**	**ChIP**	**Supplier/Reference**
Rb α-H3K36me1	1:5000	1:800	2 µL	Abcam-ab9048
Rb α-H3K36me2	1:5000	1:800	2 µL	Abcam-ab9049
Rb α-H3K36me3	1:5000	1:800	2 µL	Abcam-ab9050
Rb α-H4K16ac		1:500	2 µL	Millipore 07-329
Ms α-H3K9me2	1:5000	1:1000	2 µL	Abcam-ab1220
Ms α-Hp1a		1:25	250 uL (culture supernatant)	C1A9 (Gerland et al. 2017)
Rb α-H3K27me3			2 µL	Millipore 07-449
Rb α-H4K20me1		1:1000		Abcam-ab9051
Rb α-H3K27ac			2 uL	Diagenode pAB-0174-050
Rb α-H2AV	1:500			Börner and Becker, 2016
Gp α-MSL2	1:2000	1:500		Albig et al., 2019
Gp α-MSL3			3 µL	Albig et al., 2019
Rt α-MSL3		1:10		

Gp α-JASPer	1:2000		3 μL	Albig et al., 2019
Rb α-NSD	1:1000		5 μL	Bell et al., 2007
Ms α-NSD	1:50	1:10		Bell et al., 2007
Ms α-Set2	1:50			Bell et al., 2007
Rb α-RBP1-S2ph (ePol)			2 μL	Abcam-ab5095
Ms α-Lamin	1:1000			Gift from H Saumweber

2.2 Secondary antibodies

Western

Antigen	Conjugated dye	Dilution	Supplier
Goat, guinea pig, mouse, rabbit, rat IgG	IRDye 680RD	1:20000	LI-COR Biosciences
Goat, guinea pig, mouse, rabbit, rat IgG	IRDye 800CW	1:20000	LI-COR Biosciences

Immunofluorescence

Antigen	Conjugated dye	Dilution	Supplier
Rabbit Ig	AF 488	1:800	Jackson Immunoresearch
Rat Ig	AF 647	1:800	Thermo Fisher
Guinea Pig Ig	Rhodamine Red (RRX)	1:250	LI-COR Biosciences
Mouse Ig	Cy3	1:500	Jackson Immunoresearch

Mouse Ig	AF 647	1:500	Thermo Fisher

2.3 Cell culture

S2 and Kc167 cells were obtained from *Drosophila* Genomic Resource Center (DGRC) and were cultured in Schneider's *Drosophila* Medium (Thermo Fisher), supplemented with 100 units/mL penicillin, 0.1 mg/mL streptomycin (Sigma-Aldrich) and 10% heat-inactivated Fetal Bovine Serum (FBS, Sigma-Aldrich) at 26°C.

D. virilis 79f7Dv3 cells (36) were cultured in Schneider's *Drosophila* Medium (Thermo Fisher), supplemented with 5% heat-inactivated FBS (Sigma-Aldrich), 100 units/ml penicillin and 0.1 mg/ml streptomycin (Sigma-Aldrich) at 26°C.

2.4 RNAi treatments

RNA interference (RNAi) against target genes was performed either in large scale (for ChIP-seq) or small scale (Western Blots, Immunofluorescence staining, RNA extraction).

Large scale

10^6 S2 cells were seeded per well of a 6-well dish (Cellstar) (two wells per RNAi condition). Cells were washed once in serum-free medium followed by treatment with 10 μg dsRNA/well/RNAi target in 1 ml serum-free medium. Two different dsRNAs were used for each target to improve RNAi efficiency. Cells were incubated with mild agitation for 10 min at room temperature (RT) and further 50 min at 26°C. 2ml of complete growth medium was added and cells were incubated for 5 days at 26°C. On day 5, cells were resuspended and counted using CASY cell counter (OMNI Life Sciences). 2×10^7 cells per RNAi condition were transferred to a 75 cm^2 flask (Greiner) for a second round of dsRNA treatment (80 μg dsRNA/flask/RNAi target) in 8 ml serum-free media and incubated as mentioned above. On day 10, cells were counted and processed for ChIP-seq. For Ash1, to improve knockdown efficiency, three rounds of knockdowns were performed for 12 days as described by (Dorafshan et al. 2019).

Small scale

3×10^5 S2 cells were seeded per well of a 12-well dish (Cellstar) and allowed to adhere. Cells were then washed once in serum-free medium followed by treatment with 3–4 μg dsRNA/well/RNAi target in 300 μl serum-free medium. Cells were incubated for 10 min at RT with slight agitation on a shaker and further 50 min at 26°C in the incubator. 600 μl of complete growth medium were added and cells were incubated for 5 days at 26°C. On day 5, cells were resuspended and counted using CASY cell counter. 1.5×10^6 cells per RNAi condition were transferred to a 6-well

dish for a second round of dsRNA treatment (10 μg dsRNA/flask/RNAi target) in 1 ml serum free media and incubated as mentioned above. On day 10, cells were counted and processed for Western Blot or Immunofluorescence.

2.5 DRB treatment

A 150 mM stock solution of 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) was prepared by dissolving 10 mg powder (Sigma Aldrich; Catalog # D1916) in 200 μL of Dimethyl Sulfoxide (DMSO; Sigma), aliquoted and stored in -20°C. For ChIP, $6x10^7$ S2 cells were collected in 15 ml Falcon tubes and treated with 100 μM final concentration of DRB. To avoid precipitation, DRB was first prediluted with 4 volumes of DMSO. Falcon tubes were gently rotated for 6 hrs at 26°C. After 6 hrs, the cells were subjected to fixation according to the ChIP protocol below.

2.6 1,6 Hexanediol Treatment

For IF experiments involving 1,6 Hexanediol (1,6-HD; Sigma #240117) treatment, S2 cells were seeded on poly-l-lysine- coated (Sigma #P8920, 0.01% (w/v) final concentration) coverslips and allowed to adhere for 1 h in wells of a 6-well plate. Appropriate amount of 1,6-HD powder was directly dissolved in full media to generate a 10% weight/volume (w/v) solution. 3 ml of the solution was pipetted into the wells containing the coverslips and incubated for 2 mins. After 2 mins, cells were quickly rinsed with full media before being subjected to fixation according to IF protocol below.

2.7 dsRNA generation and associated T7 primers

Double-stranded (ds) RNA for RNAi treatments were generated by *in vitro* transcription (NEB T7 E2050s transcription kit) from PCR products generated by the following forward and reverse primers:

Name	Forward sequence (5' to 3')	Reverse sequence (5' to 3')
GST (amplified from pGEX-4T-1 GE Healthcare)	TTAATACGACTCACTATAGGGAGAAT GTCCCCTATACTAGGTTA	TTAATACGACTCACTATAGGGAGAACGC ATCCAGGCACATTG

GFP (amplified from pDest8_MT1/N-eGFP-MLE)	TTAATACGACTCACTATAGGGTGCTCAGGTAGTGGTTGTCG	TTAATACGACTCACTATAGGGCCTGAAGTTCATCTGCACCA
Set2 #1	TTAATACGACTCACTATAGGGAGAAAATCCTTGATTCCAAGCAA	TTAATACGACTCACTATAGGGAGAAGTGGTTTCTACATTTTCGT
Set2 #2	TTAATACGACTCACTATAGGGAGACACGGCTTGAGATTGCTACA	TTAATACGACTCACTATAGGGAGACATGGACATGCTTTTGTTGG
NSD #1	TTAATACGACTCACTATAGGGAGACGCGAATTCCTGAGCACGGACGCGCACTC	TTAATACGACTCACTATAGGGAGACGCTCTAGATGGACACACGCTGTTGTTGCTGTTT
NSD #2	TTAATACGACTCACTATAGGGAGACCCTCCTCTGTGAGCATCGA	TTAATACGACTCACTATAGGGAGAACAACGTTTTCGTACGTCTGG
Ash1 #1	TTAATACGACTCACTATAGGGAGACTTTGTGGCCAGGACCAATCAA	TTAATACGACTCACTATAGGGAGACAGGCAAGGGATCGTGCTCGGT
Ash1 #2	CTAATACGACTCACTATAGGGAGGCAGTGCCATGGAGACCC	CTAATACGACTCACTATAGGGAGCAACACCCAGCAGCGTCC

2.8 Immunofluorescence microscopy (IF) and analysis

IF microscopy of S2 cells was performed as follows. Approximately 200-400 thousand treated cells were seeded on 12 mm round poly-l-lysine coated coverslips and allowed to adhere for 1 h. After a quick rinse with regular media to remove unadhered cells, adhered cells were fixed in 500

μL of ice-cold 1x Phosphate-Buffered Saline (1x PBS; 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) + 2% formaldehyde (Merck Millipore) for 7.5 min on ice and washed twice with cold PBS. If needed, coverslips were stored for an extended duration (1-2 weeks) in 100% MeOH at -20°C (post-fixation). Post-fixed cells were extensively washed with PBS before subsequent steps. Permeabilization was performed by adding 500 μL of ice-cold PBS + 0.25% Triton X-100 (Sigma) + 1% formaldehyde and incubated for 7.5 min on ice. Prior to immunostaining, samples were blocked in 3% Bovine Serum Albumin (BSA; Sigma) in PBS for 1 hr at RT. Coverslips were transferred onto clean parafilm, placed into a humidified chamber containing wet tissue paper, and 40 μL of appropriately diluted primary antibody solution (see antibodies table) (PBS, 0.1% Triton-X, 1.2% Normal Donkey Serum (NDS; Merck Millipore)) was gently added onto the inverted coverslip. Primary antibody staining was performed overnight at 4°C. Coverslips were washed twice with room temperature PBS the next morning and secondary antibody staining was performed for 90 mins in room temperature. Nuclei were stained by incubating coverslips in 1 ml of 0.2 μg/mL DAPI (Sigma-Aldrich) solution for 2 min at RT. Cells were washed extensively with PBS and subsequently with deionized water (to remove possible precipitation of salts onto the coverslip) and mounted on slides with 9 μL of VectaShield (Vector Laboratories H-1000) and sealed with nail polish. Slides were protected from light by wrapping in foil and/or storing in opaque containers in 4°C and warmed up to RT 30 mins before imaging.

Confocal images were acquired on a Leica TCS SP8 with a 63x/1.4NA oil-immersion objective. Image stacks were recorded at 100 Hz scan speed with a pixel size of 350 nm and z-step size of 300 nm. Pinhole was set to 1 AU (580 nm reference wavelength). Fluorescence signals were recorded sequentially to avoid channel crosstalk. Further image processing and maximum intensity projections were done in Fiji (Schindelin et al. 2012). CellProfiler (Stirling et al. 2021) was used to quantify histone modification immunostaining signals within DAPI defined regions and further plotted on R.

2.9 Western blot

2–3×10^6 RNAi-treated cells were pelleted and lysed in 5× Laemmli Buffer (250 mM Tris-HCl pH 6.8, 50% glycerol (v/v) 10% SDS (w/v) 0.05% Bromophenol Blue (w/v) 0.5 M DTT) diluted to 1x at a concentration of 25,000 cells/μl. Samples were denatured at 95°C for 10 mins on a thermomixer. 6–8 μl of lysate per sample was electrophoresed on SDS ServaGel TGPrimer (14% gel for histone modifications, 8% gel for other proteins) for 1.5–2 h at 180 V along with protein ladder (Serva Triple Colour Protein Standard III). Whenever possible, samples were loaded as duplicates (technical replicates) to minimize variability. Proteins were transferred to AmershamTM ProtranTM 0.45 μM Nitrocellulose Blotting Membrane for 1.5 h at 300–400 mA in either standard transfer buffer (20% MeOH, 25 mM Tris, 192 mM Glycine) or high molecular weight transfer buffer for methyltransferases (10% MeOH, 0.037% SDS, 25 mM Tris, 192 mM glycine) in the cold room at 4°C. Membranes were blocked with 3% BSA for 1h at RT. The membrane was incubated with primary antibody (see table) overnight at 4°C in 3% BSA PBS on a shaker. On the next day, the blots washed thrice with PBS-T (1x Phosphate-Buffered Saline, 0.1% Tween-20) and incubated

with secondary antibody in PBS-T for 1 h in RT in opaque boxes to minimize photobleaching. Images were acquired and quantified using the LICOR Odyssey CLx. Post imaging, the blots were placed in between absorbent paper to stored long-term shielded from light. For all quantifications, the signal was verified to be linear in the range of loading used (see Figure 17). However, due to blot-to-blot variability, the normalized signals should be interpreted as an estimate rather than an absolute quantification of protein levels.

2.10 RNA isolation, cDNA synthesis and RT-qPCR

Total RNA was extracted from 1.5×10^6 cells using the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions including the optional on-column DNA digestion step. cDNA was produced from 500 ng of total RNA using Superscript III First Strand Synthesis System (Invitrogen, Cat. No 18080–051, random hexamer priming) in accordance with the manufacturer's recommended protocol. cDNA was diluted 100-fold and qPCR reactions were assembled using Fast SYBR Green Mastermix (Applied Biosystem, Cat. No 4385612) and ran on a Lightcycler 480 II (Roche) instrument. Primer efficiencies were calculated via serial dilutions and verified to be in the range 1.95-2.05. Primer sequences for Ash1 and 7sk (control) were obtained from previously published sources (Huang et al. 2017) and (Muller et al. 2020).

2.11 Chromatin immunoprecipitation after MNase treatment (MNase ChIP-seq) with spike-in

Spike-in ChIP-seq on MNase-digested chromatin in combination with mild sonication was performed as in Albig et al, 2019. S2 cells after RNAi ($\sim1.2\times10^8$ cells) or DRB ($\sim6\times10^7$ cells) treatment were harvested in falcon tubes and cross-linked for 8 min by adding 1.1 ml 10 × fixing solution (50 mM HEPES pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 10% (w/v) methanol-free formaldehyde) per 10 ml culture at RT on a rotating wheel. The reaction was stopped by adding freshly prepared glycine (VWR) at 125 mM final concentration and incubating for 10 min on ice. Cells were washed twice in PBS before subsequent steps. If needed, the cells were snap frozen for storage in -75°C up to a few months. Prior to nuclei isolation, S2 cells were spiked with 5% (relative cell number) 79f7Dv3 fixed cells (processed as described for S2 cells). Cells were then resuspended in PBS supplemented with 0.5% (v/v) Triton X-100 and cOmplete EDTA-free Protease Inhibitor Cocktail (PIC; Roche), volume was adjusted to $\sim7\times10^7$ cells/ml and transferred to 1.5 ml Eppendorf tubes. Nuclei were extracted by incubation for 15 min at 4°C with end-over-end rotation. Nuclei were collected by centrifuging at 4°C for 10 min at 2000 × *g*, pelleted and washed once in cold PBS. For chromatin fragmentation, nuclei were spun down at 4°C for 10 min at 2000 × *g*, resuspended in cold RIPA (10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 1% (v/v) Triton-X 100, 0.1%(v/v) SDS, 0.1% (v/v) Sodium deoxycholate) supplemented with PIC, PhosSTOP (Roche) and 2 mM $CaCl_2$ at 7×10^7 cells/ml. These lysates were digested in 1 ml aliquots by adding 0.6 U MNase (Sigma Aldrich), resuspended in EX-50 (50 mM KCl, 10 mM HEPES pH 7.6, 1.5 mM $MgCl_2$, 0.5 mM EGTA, 10% glycerol) at 0.6 U/µl, and incubated at 37°C

for 35 min with slight agitation. The reaction was stopped by adding 10 mM EGTA and placing it on ice. To further solubilize visible chromatin aggregates, digested chromatin was mildly sheared further with Covaris AFA S220 using 12 × 12 tubes at 50 W peak incident power, 20% duty factor and 200 cycles per burst for 8 min at 5°C. To remove any residual debris, sheared chromatin was centrifuged at 4°C for 15 min at 15 000 *g* and the resulting pellet was discarded. For long-term storage, the supernatant was flash-frozen and stored in -75°C. For initial pilot experiments, it is recommended to test the MNase digestion efficiency from a small aliquot of chromatin (after decrosslinking as proteinase treatment as below) before proceeding to next steps. Prior to the IP reaction, ~ 30 μL of 1:1 mixture of Protein A and Protein G (GE Healthcare) beads were washed thrice with RIPA. 1 ml chromatin was precleared with 70 μL of bead mix at 4°C for 1 hr to eliminate proteins with non-specific binding to the beads. 200-250 μL of precleared chromatin was diluted to 500 μL by addition of RIPA for IP reactions, while a 50 μL aliquot of chromatin was set aside as input fraction. Purified polyclonal antibodies or unpurified monoclonal antibodies (available as culture supernatant) were then pre-coupled to beads (30 μL mix/ IP reaction) for 3-4 hrs at 4°C. Resulting bead-antibody complexes were added to diluted chromatin and incubated overnight at 4°C in a rotating wheel. Note that all antibodies used in this study were verified to ChIP efficiently in the complete RIPA buffer.

The next day, beads were spun down and washed five times with 500 μL RIPA (including PIC) by incubating with end-over-end rotation for 5 min at 4°C. For DNA recovery, beads were spun down, resuspended in 100 μL TE Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Subsequently, both IP and input samples were grouped and RNA was digested with 50 μg/ml RNaseA (Sigma-Aldrich) for 30 min at 37°C, and after addition of 0.5% (m/v) SDS (Serva), proteins were digested with 0.5 µg/ml Proteinase K (Sigma-Aldrich) for 3 h at 68°C with agitation. DNA was purified with 1.8× AMPure XP beads (Beckmann Coulter). Libraries were prepared using NEBNext Ultra II DNA Library with a starting ChIP DNA amount of 3–6 ng according to manufacturer's instructions. All libraries were sequenced on an Illumina NextSeq1000 sequencer at the Laboratory of Functional Genomic Analysis (LAFUGA, Gene Center Munich, LMU). About 20 million paired-end reads were sequenced per sample for each of the ChIP samples.

2.12 High-shear ChIP-seq for MSL2

While MNase-ChIP is ideal for profiling histone modification and chromatin binding proteins, it is less suitable for profiling direct binding sites for Transcription Factors. For this, the ChIP-seq protocol required certain modifications as detailed below. All steps involving crosslinking and nuclei isolation are identical to the MNase-ChIP protocol above. However, instead of using MNase digestion and mild shearing to solubilize chromatin, nuclei are directly lysed using Covaris AFA S220 system using 12 × 12 tubes with the following settings: 100W Peak power, 20% Duty Factor, 200 cycles/burst, 10 mins at 5°C. All subsequent steps were identical to the MNase-ChIP protocol.

2.13 Data analysis methods

All code with detailed comments used to analyze the data is available in the following GitHub link:

https://github.com/MuhundenJ/Jayakrishnan_2023

For the sake of brevity, only the key data analysis methods are described below. Most sequencing data generated as part of the project is available on the GEO portal with the accession ID GSE253391.

2.13.1 Read processing

ChIP-seq sequence reads were demultiplexed by JE demultiplexer (Girardot et al. 2016) using the barcodes from the Illumina Index read files. Demultiplexed files were aligned to either *D. melanogaster* reference genome (BDGP6, release 104) or independently to *D. virilis* genome (droVir3. Feb 2006) using Bowtie2 (Langmead and Salzberg 2012) version 2.28.0 (parameter "--end-to-end --very-sensitive --no-unal --no-mixed --no-discordant -X 500 -I 10") and filtered for quality using samtools 1.6 (Li, Handsaker, et al. 2009) with a MAPQ score cutoff of -q 10. For transposons, a custom genome containing repetitive regions was used as before (Albig et al. 2019) at the alignment step. Tag directories and input-normalized coverage files were generated using *Homer* (Heinz et al. 2010) with the parameter -totalReads set to the number of reads mapped to *D. virilis* genome for spike-in normalization. Input-normalized, scaled *D. melanogaster* coverage per base pair files were visualized using the Integrative Genomics Viewer (Robinson et al. 2023). Replicate coverages were first analyzed independently to confirm similarities in HMT-dependency patterns after which they were averaged for subsequent analyses. Resizing of coverages to fixed window sizes of mean signal was performed with bedops and bedmap (Quinlan 2014).

Published RNA-seq reads were obtained from respective sources using sra prefetch and processed according to (Albig et al. 2019). Single-end parameters were used for read alignment by STAR for Huang et al., 2017 and paired-end settings for Depierre et al., 2023.

2.13.2 ChIP-seq peak calling and annotation

Broad domains of modified H3K36me1/2/3, H327me3 and H3K9me2 were called using MACS2 v2.1.2 (Zhang et al. 2008) bdgpeakcall function using parameters -l 1000 (-c 3 for K36me1/2/3; -c 0.8 for K27me3; -c 2.0 for K9me2). For MSL2, focal peaks were detected with HOMER findPeak with the parameters -size 150 -F 15 -style factor -L 6. Manipulation (filtering, merging etc.) of peak sets was performed with BEDTOOLS2 v2.28.0. Genomic annotation of peaks was done using HOMER annotatePeaks.pl script.

2.13.3 Data analysis and plotting

Data were analysed in R (R Core Team, 2020) using tidyverse libraries. ChromoMaps were generated using the R chromoMap package (Anand and Rodriguez Lopez 2022). Clustered heatmaps were made using R package 'ComplexHeatmap' (Gu, Eils, and Schlesner 2016). Chromatin State annotations were derived from modENCODE (Kharchenko et al. 2011). Gene annotations were obtained from FlyBase GTF annotations BDGP6 release 104. Only genes associated with a unique FlyBase (FBgn) ID (N=17.8k) were used for subsequent analyses. Detailed explanation for certain analyses can be found within provided R scripts in the GitHub link above.

2.13.4 Differential binding analysis

To identify significant differential K36me1/2/3 regions across RNAi conditions, csaw v1.24.3 (Lun and Smyth 2016) was used. Reads for each unique experimental sample (3 replicates per RNAi condition) were counted into sliding windows of 250 bp across the entire genome. Windows were filtered using filterWindowsControl function using the corresponding Input profiles as controls. To calculate normalization factors derived from the spike-in, normFactors function was applied on *D. virilis* bins containing high ChIP signal. These high-signal bins in the spike-in genome shouldn't vary across the different *D.melanogaster* RNAi conditions, thus any systematic differences in the signal across the libraries reflect technical biases that can be removed by normFactor scaling. Differential binding was assessed for using the quasi-likelihood (QL) framework in the edgeR package v3.32.1 with robust=TRUE for glmQLFit. The design matrix was constructed using a layout specifying the RNAi treatment as well as the experimental batch. Proximal tested windows were merged into regions of maximum 3 kbp by clusterWindows with a cluster level FDR target of 0.05.

2.13.5 Data transformation

Z-score data transformation for average signal within 5-kbp genomic windows (Figure 3) was performed as previously described (Chaouch et al. 2021). Average feature coverage instead of RPKM was used in calculations. A consensus peak set representing all genomic regions containing at least one K36 modification in any RNAi condition was further used to filter windows before Z-score representation in Figure 3. Proportional change in average gene signal used in Figure X was defined as ((HMT RNAi – Control RNAi) / Control RNAi).

2.13.6 External Datasets

Dataset	Source
MBD-R2	modENCODE (GSE27802)

Ash1 ChIP (S2 cells)	Huang et al., 2017
H3K36me2 ChIP in Ash1 RNAi (S2 cells)	Huang et al., 2017
RNAseq in Ash1 RNAi (S2 cells)	Huang et al., 2017
Chromatin States	Kharchenko et al., 2011
H3K36me3 modENCODE (S2 cells)	modENCODE (GSE20785)
H3K36me3 Native-ChIP (Fly)	Chaouch et al., 2021
H3K36me3 modified MNase ChIP (S2 cells)	Albig et al., 2019
Su(var)3-7	modENCODE (GSE23487)
H3K9me2	modENCODE(GSE20792)
H4K20me1	modENCODE(GSE27743)
H3K27ac	modENCODE(GSE20779)
H3K4me1	modENCODE(GSE32826)
Nurf	modENCODE(GSE20829)
H2B-ub	modENCODE(GSE20773)
H3K79me3	modENCODE(GSE45090)
Kdm4A	modENCODE(GSE32839)
Kdm2	modENCODE(GSE45061)
MRG15	modENCODE(GSE25367)
Beaf-32	modENCODE(GSE20760)

SMC3	modENCODE(GSE45054)
CTCF	modENCODE(GSE32750)
H3K27me3	modENCODE(GSE20781)
H3.3	Henikoff et al., 2009
SNR1	Hendy et al., 2022
ISWI	Hendy et al., 2022
RNAseq in NSD/Set2 RNAi (S2 cells)	Depierre et al., 2023
WT S2 cells RNAseq	Albig et al., 2019
Fly HMT KO RNAseq	Lindehell et al., 2021
FlyAtlas Tissue Expression Data	Chintapalli et al., 2007
H3K36me2/3 in HMT KO (C3H10T cells)	Weinberg et al., 2019
RNAseq of WT C3H10T cells	Weinberg et al., 2019
H3K36me3 in HMT dTAG cells (mESC)	Sun et al., 2023

Genomic coordinates of datasets previously aligned to genome version dm3 were transformed to newer dm6 using liftOver tool (Hinrichs et al. 2006).

3. Results

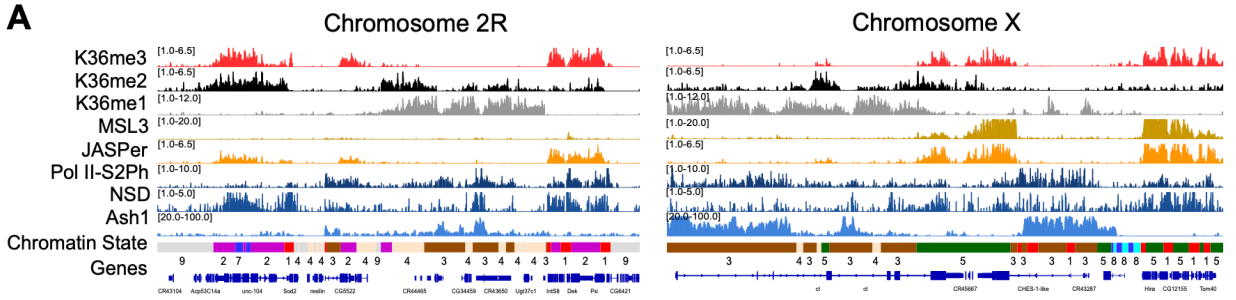
Principles of establishment of H3K36 methylation in *Drosophila*

3.1 Characterization of steady-state distribution of H3K36 methylation in S2 cells

3.1.1 K36me1/2/3 and associated HMTs are enriched in distinct chromatin states

To study the genomic distribution of the histone H3K36 methylation states, I generated high-resolution MNase chromatin immunoprecipitation (MNase ChIP-seq) profiles in male *Drosophila* S2 cells. As I aimed to compare the profiles of the modifications with corresponding HMTs and K36me3 readers, ChIP profiles for the HMT NSD and readers JASPer and MSL3 were also generated. As no ChIP-grade antibodies exist for HMTs Set2 and Ash1, I used an antibody for the elongating RNA polymerase II marked by serine-2 phosphorylation (ePol) as a proxy for Set2 and a published Ash1 ChIP-seq profile (Huang et al. 2017). As a first assessment, the ChIP profiles were visualized on the IGV genome browser at two representative genomic regions (one on the autosomes and the other on the X-chromosome) which reveal distinct domains of K36me3/2/1 correlated to various degrees to the different HMTs (Figure 13A).

A



B

	Fraction genome coverage (%)	Bonnet *et al* abundance Late embryo (copies/1C genome)	Density of mark (per n nucleo-somes)	Feller *et al* abundance S2 cells (% of H3K36)
me1	12.4	124,000	9 in 10	2
me2	23.3	64,000	1 in 4	1.1
me3	18.1	6,000	1 in 28	0.3

Figure 13: K36me1/2/3 are distributed differently throughout the genome. **A.** Genome browser profiles for representative *Drosophila* chromosomes 2R and X of MNase-ChIP of K36me1/2/3, the K36me3 reader proteins JASPer and MSL3 and the HMTs NSD and Ash1 and ePol (proxy of Set2). The 9-state ChromHMM (modENCODE) is color-coded and explained in Figure 14. **B.** Genomic coverage of K36me1/2/3 peaks and estimated densities based on Bonnet et al., 2019. Figure adapted from Jayakrishnan et al., 2024.

To understand how widely these marks are distributed, I annotated broad ChIP-seq 'peaks' and calculated the percentage of the mappable genome spanned by these peaks (Figure 13B). This analysis revealed that K36me2/3 roughly occupy ~20% of the genome while K36me1 is more localized, covering ~12%. It's important to note here that this approach will underestimate occupancy (and overestimate density) for factors concentrated on unmappable regions (i.e. repetitive sequences), which is shown to be relevant to the study below. By correlating the occupancy data to absolute quantification of abundance of the marks in S2 cells (utilizing published mass-spectrometry data, (Bonnet et al. 2019; Feller et al. 2015)), it can be inferred that K36me1 domains are highly dense while K36me3 are deposited in a sparse manner, with K36me2 showing an intermediate density.

To better understand the differences in the genomic distribution, the relative enrichments of the profiled factors within previously defined Chromatin States (Kharchenko et al. 2011) was calculated and visualized as a clustered heatmap (Figure 14A).

These nine states reflect different functional categories of chromatin, as highlighted by the enrichment of various other representative ChIP-seq profiles generated by the modENCODE project. K36me3 is enriched prominently on promoter state 1 and transcribed chromatin state 2, but also unexpectedly shows a modest enrichment within heterochromatic state 7. K36me2 overlapped partially with K36me3, but showed a much stronger enrichment at heterochromatin state 7. In contrast, states enriched for K36me2/3 show very low signal for K36me1, with K36me1 instead marking enhancer-like states 3 and 4. The profiles of the HMTs partly mirror the distributions of the K36 modifications. ePol (and likely Set2) is enriched on states 1-3 but is absent from heterochromatin, consistent with all knowledge about the transcriptionally inert environment of heterochromatin. Curiously, NSD appeared to show a dual eu-/heterochromatic distribution, localization both on promoters and gene bodies, but also in heterochromatic gene deserts. I verified this dual distribution using immunofluorescence (IF) staining for NSD, which showed a dispersed euchromatic signal along with strong speckles adjacent to (but not superposed on) DAPI-dense chromocenters (Figure 14B). The possibility of nonspecific IF staining was ruled out by performing an RNAi against NSD, which eliminated both eu- and heterochromatin signals, confirming the dual distribution of NSD.
Ash1, in contrast to the other K36 HMTs, was almost exclusively present at states 3-4 enriched for K36me1. Lastly, the readers JASPer and MSL3 also showed the expected patterns of enrichment; with MSL3 enriched at X-chromosomal state 5 while JASPer showing a broader enrichment resembling K36me3 distribution.

To get insights into the genomic features underlying each mark, I intersected the previously called domains with intergenic/genic annotations (Figure 14C). The strongest K36me3 peaks predominantly contained TSS and exons, confirming its 3'-biased distribution. K36me2 was majorly intergenic with a subset of peaks on introns. In contrast, K36me1 was almost exclusively present on introns. These differences highlight the possibility of involvement of the K36me1/2/3 marks in distinct pathways linked to transcription.

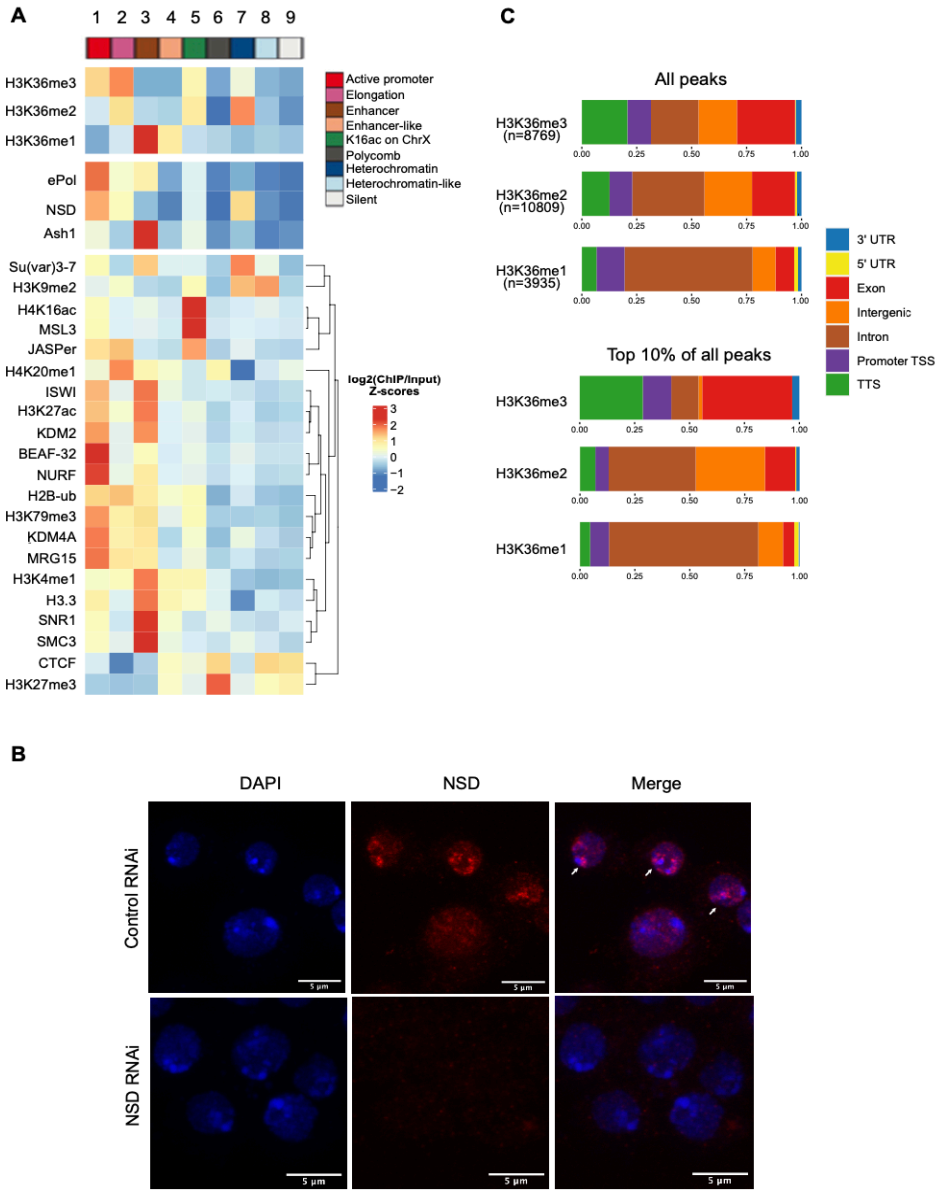


Figure 14: K36me1/2/3 enrich in distinct chromatin states and are correlated to certain HMT occupancies. **A.** Chromatin State enrichment (9-state ChromHMM, (Kharchenko et al., 2011)) for K36me1/2/3 and associated factors. Published ChIP-seq/CUT&RUN profiles of histone PTMs and other chromatin proteins were clustered to highlight differences between chromatin states. **B.** Immunofluorescence images of S2 cells subjected to RNAi against either GST (control) or NSD. Staining for DNA, NSD and merged images are shown. Arrowheads indicate chromocenter-proximal speckles of NSD. Scale bar is 5 μm. **C.** Genomic features marked by either all K36me1/2/3 domains (top) or filtered for the strongest 10% (bottom). Figure adapted from Jayakrishnan et al., 2024.

3.1.2 K36me3 at heterochromatin is a result of improved solubilization in MNase-ChIP protocol

The presence of K36me3 on heterochromatin was puzzling, which hitherto hasn't been reported in *Drosophila*. I reasoned that previous studies may have missed out on this either due A) aligning of ChIP-seq reads to older genome versions which missed many sequences near heterochromatic regions or B) incomplete solubilization of chromatin prior to immunoprecipitation step. To better understand the differences, I compiled several published K36me3 profiles generated using different ChIP-seq protocols as well as a K36me3 CUT&RUN profile generated in the group of Dr. Catherine Regnard and (re-)aligned all samples to the newer reference genome. Visualization of these profiles (normalized to their corresponding inputs wherever available) on IGV at the pericentric heterochromatic (PCH) region of chromosome 3L containing strong K36me2/K9me2 signatures, confirmed that the modified MNase-ChIP protocol employed in this study was superior in detecting heterochromatic K36me3 signal compared to all other methods (Figure 15A).

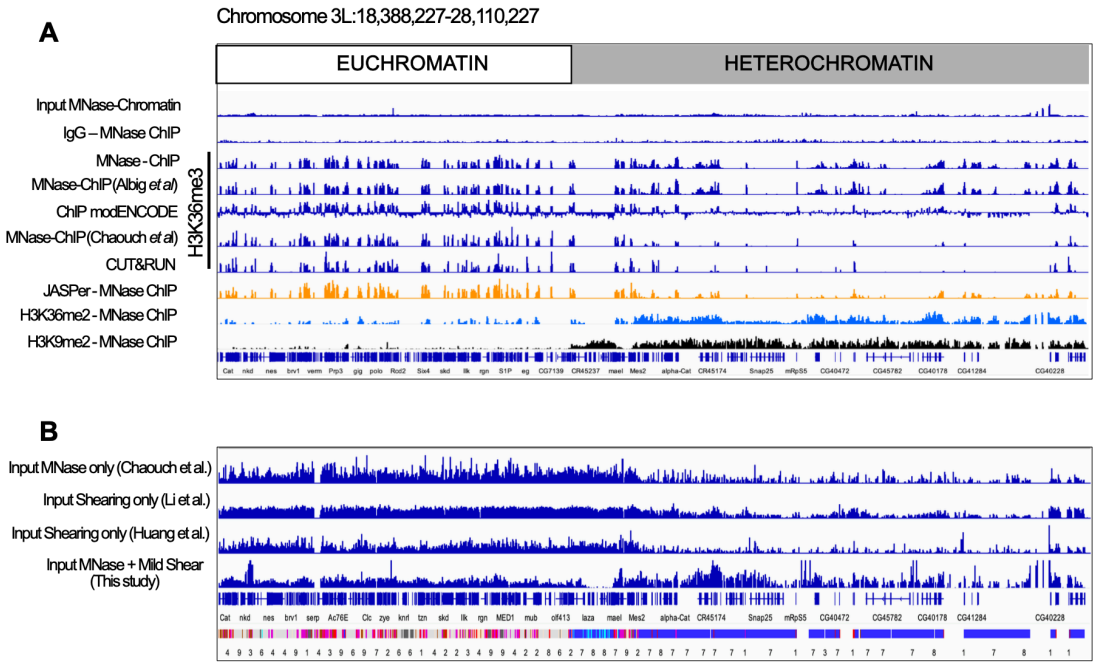


Figure 15: A. Genome browser profiles highlighting preferential solubilization of euchromatin in different methods. Comparison of different datasets: MNase-ChIPs: this work and Albig et al., 2019 (include mild shearing); CUT&RUN: generated in the Regnard group; MNase-ChIP Chaouch et al., 2021 : native MNase-ChIP without crosslinking or shearing; ChIP modENCODE: Bioruptor shearing. IgG ChIP as well as chromatin input serve as negative controls. Colocalization of K36me3 signal with reader JASPer (MNase-ChIP) confirms specific signal. H3K36me2 and H3K9me2 coverages provided for reference to highlight heterochromatin.
B. Genome browser profiles of chromatin inputs generated by MNase only (Chaouch et al., 2021), Shearing only (Huang et al., 2017, Li et al., 2025) and MNase + mild shearing (this study) demonstrating uneven solubilization in former methods. Figure adapted from Jayakrishnan et al., 2024.

In addition to K36me3, the K36me3 reader JASPer was also detected. Importantly, performing ChIP using a nonspecific IgG antibody raised from the same species (rabbit) shows no enrichment, ruling out that these signals are produced by general 'stickiness' of heterochromatin. Upon examination of the various ChIP protocols, which use either MNase-only or mechanical shearing-only, I realized that the key difference possibly lies in the solubilization step where the protocol from our lab (Albig et al., 2019 and this study) use a mild but focused acoustic shearing using Covaris in addition to the standard MNase treatment, thereby combining the advantages of both methods. The chromatin input generated from our protocol is more uniform compared to other protocols (Figure 15B) and input normalization doesn't accurately correct for this bias. Incomplete solubilization of heterochromatin in standard ChIP protocols, which can lead to bias in conclusions, has been reported before (Becker et al. 2017; Nicetto et al. 2019).

3.1.3 Broad K36me1 domains contain a specific class of enhancers and segregate in 3D space

The observation that K36me1 strongly differs from K36me2/3 in chromatin state enrichment and displayed enhancer-like features was intriguing. To investigate this further, I sought to analyze the distribution of K36me1/2/3 in reference to enhancers. *Drosophila* enhancers were recently classified into two categories in STARR-seq assays (housekeeping vs. developmental) based on the type of promoters they act on (Hendy et al. 2022; Zabidi et al. 2015). Using these enhancer classes, heatmaps of the reference ATAC-seq (representing accessibility) and ChIP-seq signals (of enhancer marks H3K4me1 and H3K27ac) of the 20-kb region surrounding the enhancer centers of two classes were generated (Figure 16A). This showed that developmental enhancers (which are frequently distal from their target promoters) were strongly enriched for K36me1 in addition to the classical enhancer marks, while housekeeping enhancers (which are often proximal to target promoter) were rather enriched for K36me3 (and K36me2, data not shown).

As mentioned previously, enhancers contribute to particular spatial genome organization. To understand how differences in '1-dimensional' K36me1/3 chromatin states correlate to differences in '3-dimensional' folding of the genome, published Hi-C data were visualized on Juicebox (Figure 16B). I observed that K36me1 domains are strongly correlated to TADs containing developmental enhancers which do not overlap with K36me3-marked TADs. Remarkably, the spatial boundaries of the developmental enhancer-containing TADs coincide exactly with the modification boundaries of the K36me1, suggesting that K36me1 may be instructive to and/or follows underlying genome topology. Taken together, these observations argue that K36me1 forms a unique developmental enhancer-like chromatin state that spatially segregates from actively transcribed genes containing K36me3.

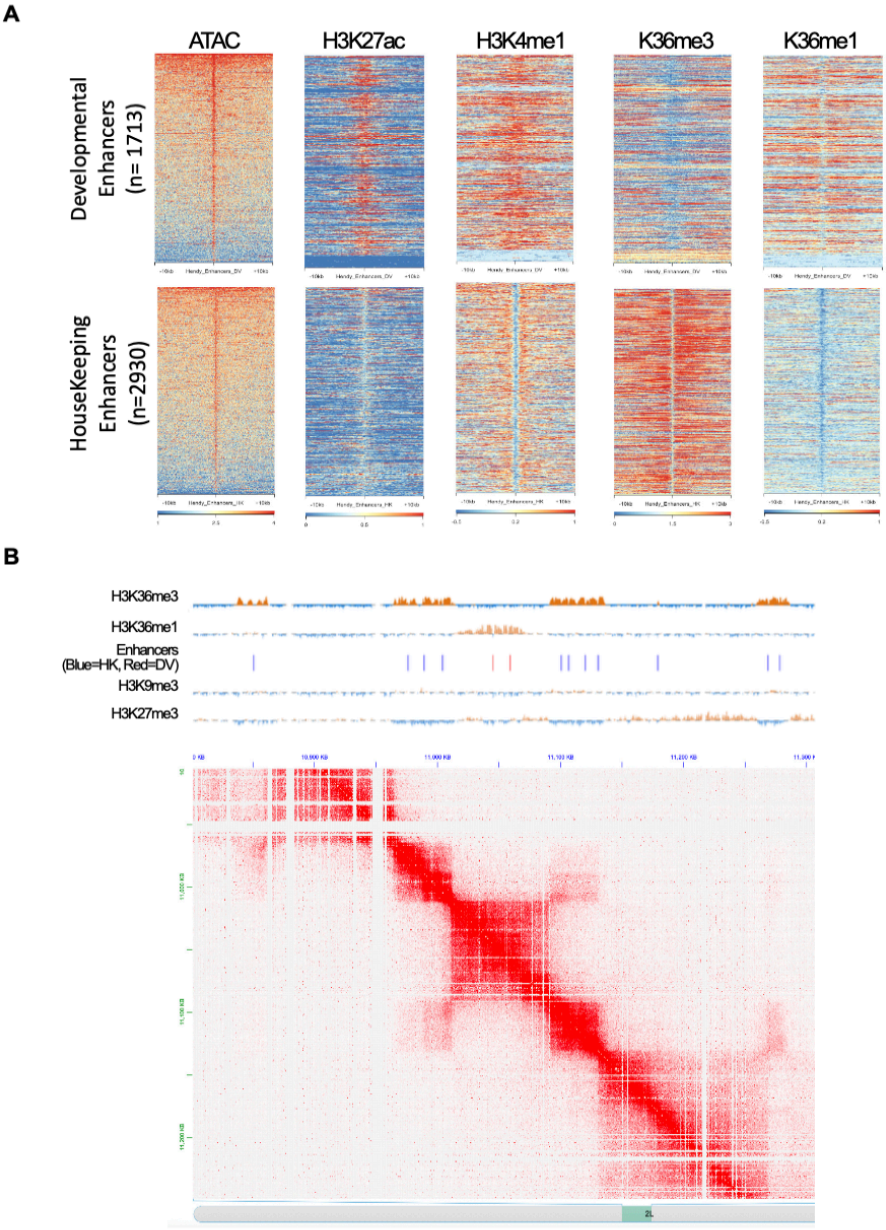
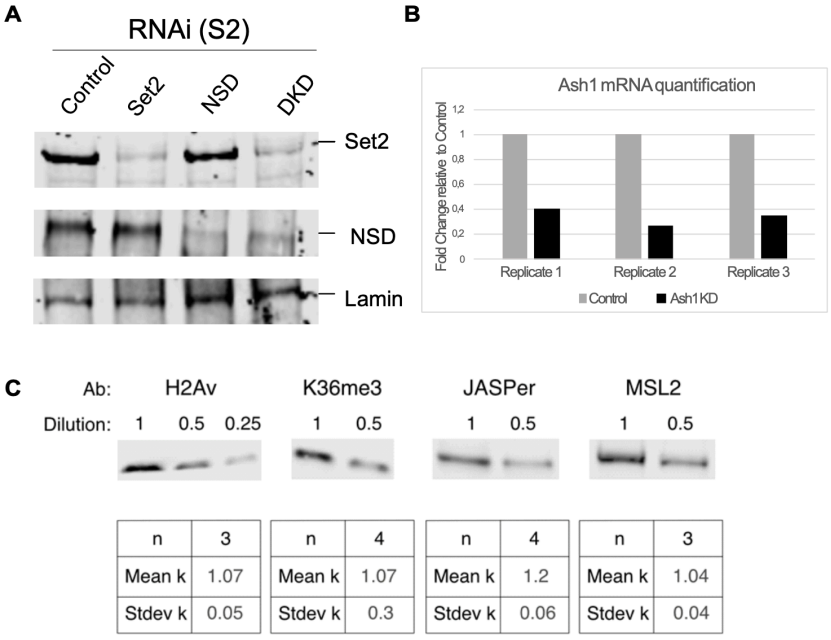


Figure 16: A. Heatmaps of enhancer hallmarks ATAC-seq (ref), H3K27ac, H3K4me1 (modENCODE) and K36me1/2/3 of 20-kb region centered around housekeeping and developmental enhancers as defined by Zabidi et al., 2015. **B.** Hi-C heatmap demonstrating distinct spatial chromatin domains formed by K36me1 and K36me3. Enhancers (housekeeping/HK and developmental/DV), H3K9me3 and H3K27me3 profiles are provided for reference.

3.2 Contribution of H3K36 HMTs to different K36 methylation states

3.2.1 Bulk changes of K36me1/2/3 occurring upon HMT RNAi

To understand the extent to which each HMT contributes to bulk levels of K36me1/2/3, I depleted each HMT by RNAi treatment for 7-10 days individually and in combination and measured K36me1/2/3 using Western blotting as well as IF. I first assessed the extent of knockdown of the target proteins i.e., the HMTs by Western blot analysis, which revealed a strong reduction (~80-90%) of Set2 and NSD upon corresponding RNAi (Figure 17A). Due to the poor quality of Ash1 antibody, I resorted to RT-qPCR to quantify the Ash1 mRNA levels after RNAi, which confirmed a 75% reduction (Figure 17B).



n	3
Mean k	1.07
Stdev k	0.05

n	4
Mean k	1.07
Stdev k	0.3

n	4
Mean k	1.2
Stdev k	0.06

n	3
Mean k	1.04
Stdev k	0.04

Figure 17: A. Representative Western blot documenting the depletion of HMTs in whole cell extracts from S2 cells treated with indicated dsRNA. α-Lamin serves as loading control. DKD: combined depletion of Set2 and NSD. **B.** RT-qPCR quantification of mRNA extracted from S2 cells treated with Ash1 dsRNA. **C.** Representative Western blots as well as linearity quantification using dilutions of cell extract. 'n' indicates number of replicates, 'k' denotes the value of slope in y=kx where 'y' is normalized Western blot signal and 'x' is the dilution factor. Value of 'k' close to 1 with low variability (std. deviation) indicates good performance of the antibody. Figure adapted from Jayakrishnan et al., 2024.

I first aimed to quantify the levels of K36me1/2/3 using Western blot. Depending on the nature of the antibody as well as the blotting protocol, the quantitative nature of Western blotting can vary. For proper quantification, the signal produced by the antibody staining scales linearly with the amount of cell lysate loaded. In other words, if the Western blot data is fitted with the equation 'y=kx', where 'y' is normalized Western blot signal and 'x' is the protein lysate dilution, then 'k' should be close to 1 for linear scaling. To assess this, I performed Western blot on lysate dilution series for several antibodies (Figure 17C). As shown by representative blots, the value of 'k' for

several antibodies was close to 1, although variation between experiments was fairly high for K36me3. Based on this, I assert that in this experimental setup, Western Blot is **semi-**quantitative.

I then probed for K36me3/2/1 levels after individual RNAi or Set2+NSD combined RNAi (DKD) or Set2+NSD+Ash1 combined RNAi (TKD) (Figure 18A, B).

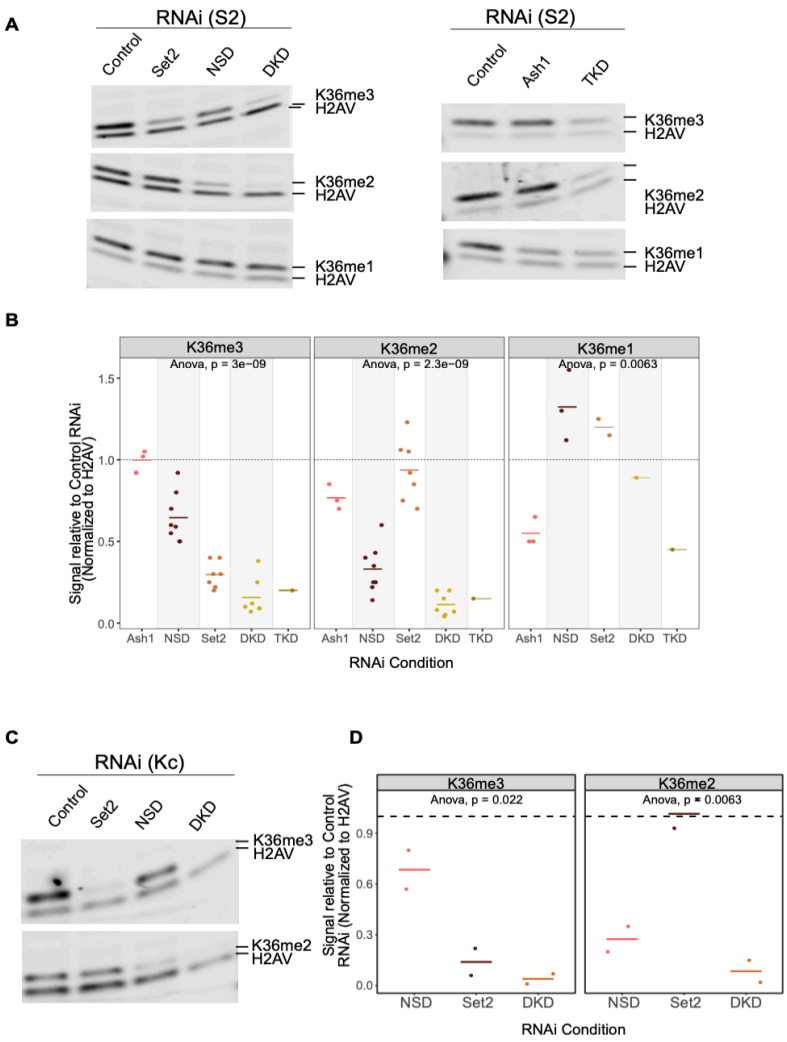


Figure 18: A. Representative Western blots of K36me1/2/3 used for quantification in panel B. Histone H2AV serves as loading control. **B.** Quantification of K36me1/2/3 levels in whole cell extracts from S2 cells treated with RNAi against indicated factor (DKD=Set2+NSD; TKD=Set2+NSD+Ash1). Values were normalized to H2AV signals on the same membrane and are represented as fraction relative to Control (GST) RNAi, which was run on same blot. Each dot represents an independent biological replicate. Calculated ANOVA p-values (null hypothesis: difference between means = 0) are presented for each antibody. **C-D:** K36me2/3 Western blot for female Kc cells treated with dsRNA against indicated factor and associated quantification. Figure adapted from Jayakrishnan et al., 2024.

Around 70% of K36me3 was lost upon Set2 RNAi alone, while NSD RNAi had only around 20% effect. Interestingly, a clear additive effect was observed upon DKD, suggesting that Set2 and NSD lie in distinct pathways. In other words, Set2 and NSD likely affect different pools of K36me3. Focusing on K36me2, I observed that NSD knockdown produced the strongest reduction (~70%), followed by Ash1 with a 20% reduction, while Set2 had no effect. Importantly, despite the strong K36me2 reduction observed upon NSD RNAi, K36me3 was only slightly affected. This observation is at odds with the model that NSD-dependent K36me2 is a necessary intermediate for Set2-dependent trimethylation.

To generalize the conclusions, the experiments were repeated in the female Kc cell line, which recapitulated the results (Figure 18C, D). Lastly, K36me1 was largely unaffected in all knockdowns involving Set2 and NSD, but was 50% lower in Ash1 RNAi. Of note, it was likely that the observed reduction was an underestimation as some lots of the K36me1 antibody were previously reported to have some low reactivity to unmodified nucleosomes in Western blotting experiments (Egelhofer et al. 2011).

As an orthologous approach to Western blotting, IF assays were performed in the same conditions and nuclear signals were quantified for two replicates (Figure 19A-C) and confirmed the major results obtained by the Western blot approach. K36me3 was strongly affected by Set2 RNAi, K36me2 predominantly by NSD (and Ash1 to a lower extent) while K36me1 was dependent exclusively on Ash1. It's unclear why an NSD-dependent reduction of K36me3 is not observed in IF despite the Western quantifications, but one possible explanation is that pan-nuclear quantification does not capture very local changes. Of note, the second replicate (quantified in Figure 19C) was of poorer quality due to high background signal, but nevertheless recapitulate the strongest effects from the first replicate shown in Figure 19B.

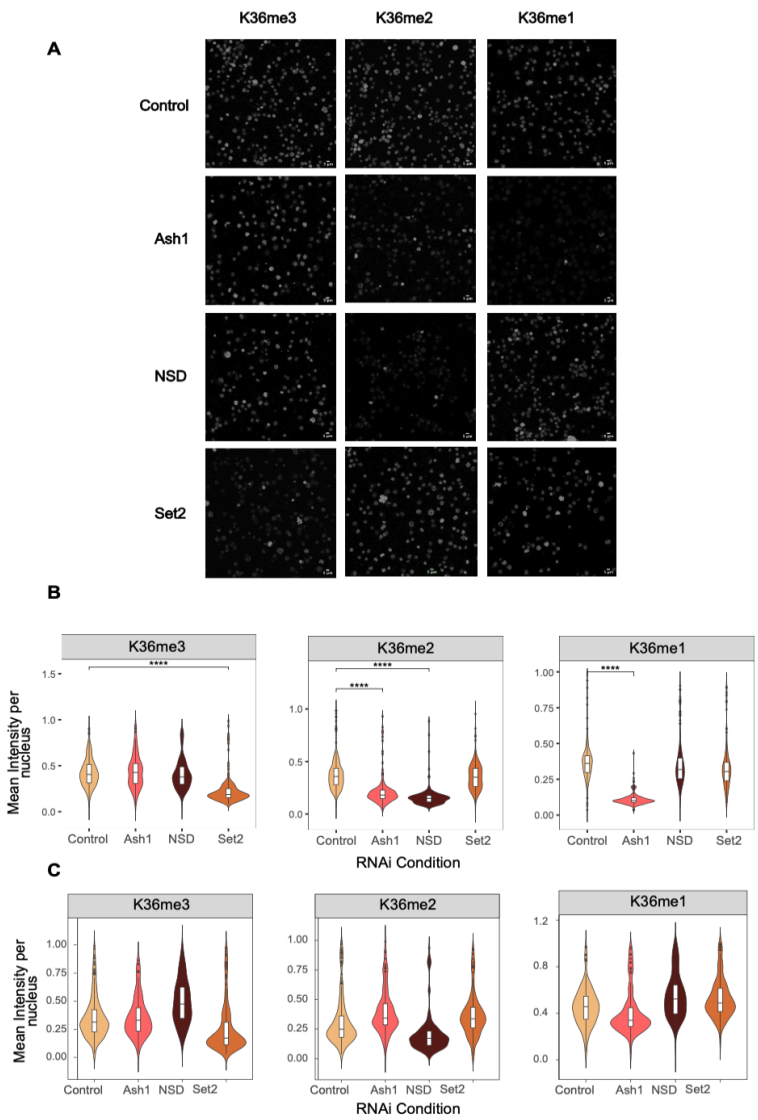


Figure 19: A. Representative IF microscopy images for K36me1/2/3 in S2 cells treated with RNAi against GST (Control), Ash1, NSD or Set2. The scale bar is 5 µm. **B-C.** Quantification of nuclear signal from IF images (n = ~500 nuclei from each biological replicate, separately quantified). ANOVA followed by post-hoc Tukey HSD was performed to identify groups with significantly different mean relative to GST RNAi (**** denotes a significance value of $P < 0.001$) for first replicate. Figure adapted from Jayakrishnan et al., 2024.

Overall, these observations regarding contributions of the different HMTs to K36me1/2/3 are at odds with the conventional model that K36me1/2 are obligate intermediates for Set2-dependent trimethylation. Rather, each HMT contributes majorly to one modification. The deposition pathways for each of K36me1/2/3 may be disconnected, a hypothesis that will be tested below.

3.2.2 K36 HMTs regulate distinct transcriptional pathways: Evidence from mutant fly brain transcriptomes

To further support the idea that activities of each HMT are largely uncoupled, published transcriptome datasets of fly larval brains deleted for each HMT were reanalyzed (Figure 20). If Set2, NSD and Ash1 lie in the same regulatory pathway, one would expect similar transcriptomic changes upon their deletion. By looking at the overlap of up- and downregulated genes upon various knockouts, I could observe that each HMT largely affects disjoint groups of genes, arguing that the HMTs likely have non-overlapping functions.

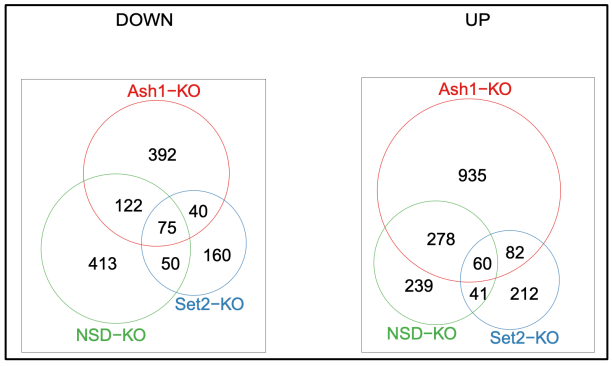


Figure 20: Venn diagram representing overlap between significantly up-/downregulated genes ($p<0.05$; $log2FC < -0.32$ or $log2FC > 0.26$ (i.e., increased or decreased by at least 20%) from HMT knock-out (KO) published fly larval brain RNA-seq data (Huang et al. 2017; Lindehell et al. 2021). Figure adapted from Jayakrishnan et al., 2024.

3.3 Chromosome locus-specific H3K36 methylation changes upon HMT depletions

3.3.1 Correlative analyses of genome-wide changes

Bulk approaches, like Western and IF, are not capable of detecting changes at specific genomic loci. To assess the specific patterns of K36me1/2/3 changes occurring upon the RNAi, MNase-ChIP for each modification in the different RNAi conditions was performed. I adopted a statistical approach using the tool 'csaw' to identify genomic regions that show significantly altered K36me1/2/3 signals upon corresponding RNAi with respect to the control RNAi (ref). To visualize these regions in a chromosome-wide manner, I generated chromoMaps representing 2-kb bins that provide information regarding spatial patterns (i.e., magnitude and direction of change) of methylation changes in each RNAi condition (Figure 21B-D). For representative purposes, only chromosome 3L is shown. The distribution of each mark in control RNAi condition is shown above the chromoMaps, while gene annotations and K9me2 peaks demarcate eu- and heterochromatin respectively. Each chromoMap is supplemented with a paired scatter plot showing the correlations of Z-score values (which represent the direction and magnitude of change), thus allowing quantitative comparisons of the contributions of different HMTs for a given genomic locus (see schematic Figure Figure 21A).

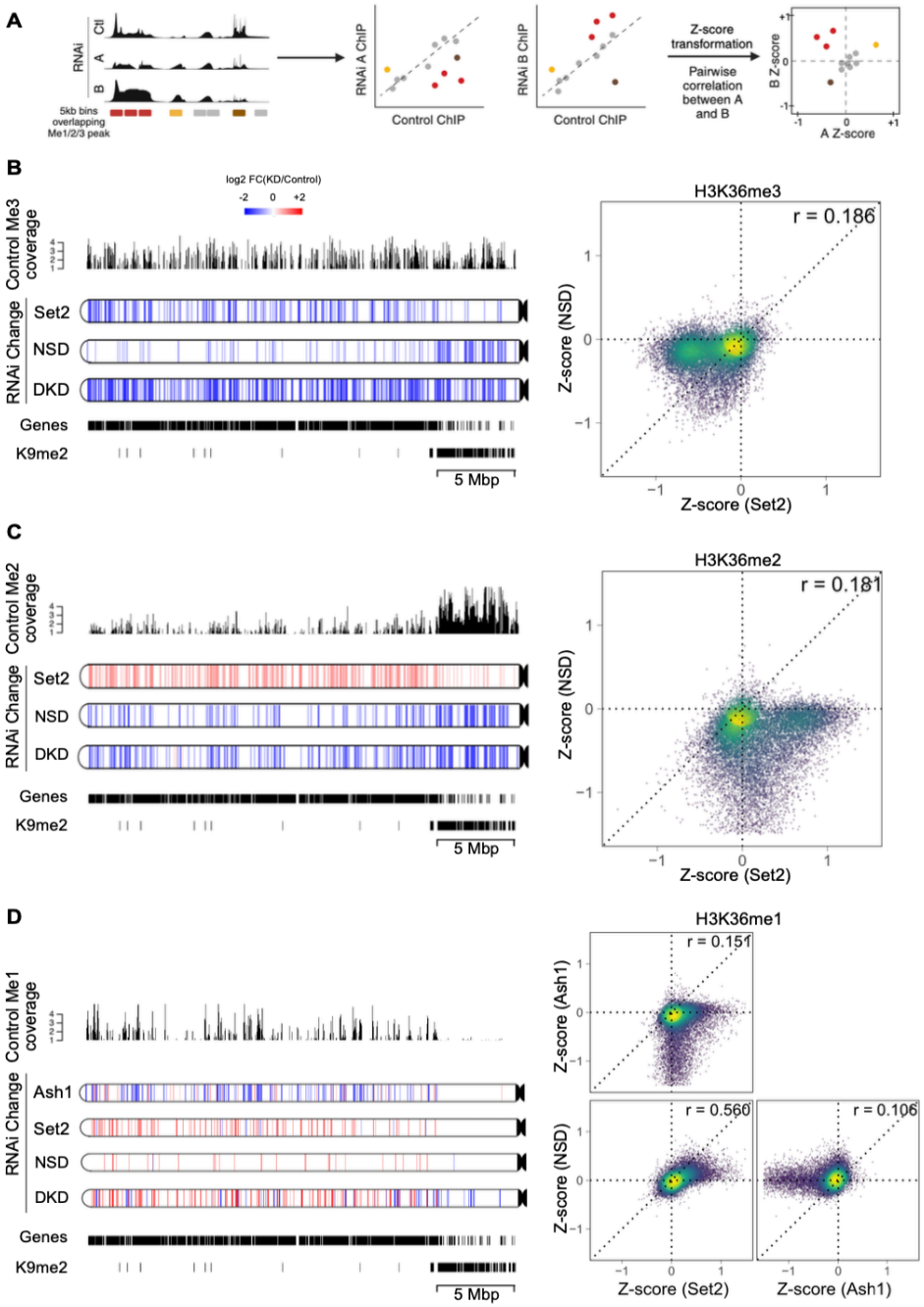


Figure 21: A. Pairwise Z-score analysis workflow schematic representing correlations among changes in ChIP signal upon RNAi of HMTs in genome-wide 5 kb bins. Bins not overlapping any K36me1/2/3 peaks in any RNAi condition were excluded. Schematic drawings were generated with BioRender. **B–D**. Pairwise Z-score scatter plots for K36me1/2/3 upon RNAi of indicated factor (right). Local density of points is indicated by the color of the dots. Corresponding chromoMaps representing regions of significant difference (at least 50% change) in K36me3/2/1 signal for indicated RNAi conditions from csaw analysis for chromosome 3L at 2-kbp resolution (left). The color of the regions represents log2 ratio of number of TMM-adjusted read counts in RNAi condition relative to control condition. Control K36me3/2/1 signal tracks (overlayed above chromoMaps) along with K9me2 peaks and gene annotations aid interpretation of relative changes. Figure adapted from Jayakrishnan et al., 2024.

The chromoMaps for K36me3 show that Set2 RNAi leads to a reduction of predominantly euchromatic K36me3 consistent with the known localization of ePol on gene bodies. NSD RNAi leads to K36me3 reduction at gene-poor PCH and a subset of euchromatic bins (but still different from Set2-dependent regions as well as from the K9me2 nanodomains dispersed throughout euchromatin). These two distinct sets of genomic regions likely correspond to the two 'pools' of K36me3 dependent on either Set2 or NSD that was previously observed in Western blot assays. The corresponding scatterplots show that many points strongly lose K36me3 in response to Set2 RNAi, but respond benignly to NSD RNAi (which lie on the left side of the horizontal axis). Another population can be observed on the right side of the diagonal line, which demonstrates that many regions that contain NSD-dependent K36me3 also show partial sensitivity to Set2 RNAi. Putting these results in the context of chromoMaps, it can be inferred that Set2 deposits K36me3 on euchromatin largely independent of NSD, but also plays a part in shaping heterochromatic K36me3 which is predominantly catalyzed by NSD.

Interpretation of K36me2 changes is more complex as this mark may arise from both methylation (me0/1→ me2) and demethylation (me3 → me2) reactions. As discussed previously, K36me2 is strongly enriched at heterochromatin in control conditions. On one hand, heterochromatic K36me2 depends largely on NSD, which also deposits K36me3, as seen before. On the other hand, the changes at 'weak' domains of K36me2 on gene bodies is milder, with only some bins showing slight decreases. Interestingly, these euchromatic bins showing reduced K36me2 upon NSD RNAi are the same bins that also lose K36me3 in the same RNAi condition (and appear to not overlap Set2-dependent K36me3 bins). In contrast to NSD RNAi, Set2 RNAi shows a strong uniform increase of K36me2 within euchromatin. These observations are quantified with the corresponding scatterplots, which confirm bidirectional response of K36me2 upon RNAi of either NSD or Set2. The strong increase observed upon Set2 RNAi is possibly due to a combination of two effects: i) demethylation of Set2-dependent K36me3 to yield K36me2 and ii) a mechanistically unexplored activation of NSD to generate more K36me2 in absence of Set2. Since the dramatic increase is not reproduced in the combined depletion of NSD and Set2, the second hypothesis is more likely.

Changes in K36me1 are the most straightforward to interpret. K36me1-rich domains are lost upon Ash1 RNAi, while RNAi of other HMTs leads to mild increases in K36me1-poor domains, as it is a likely product of demethylation of higher methylation states.

Taken together, these results show that each HMT dominantly exerts its methylation activity in dedicated spatial domains marked by its characteristic modification (Set2 at K36me3-rich euchromatic domains, NSD at K36me2-rich PCH and subset of euchromatic domains, Ash1 at K36me1-rich euchromatic domains), and HMT RNAi sometimes results in complex patterns likely due to a mix direct methylation and indirect demethylation or activation effects.

3.3.2 Insights into NSD targeting

The observations so far have established that NSD is a K36me2/3 methyltransferase that works mostly independently to sculpt the K36 modification landscape at PCH (in addition to a few euchromatic regions). NSD has previously been implicated at heterochromatin based on immunostaining (Lindehell et al. 2021) or HP1α pulldown assays (Alekseyenko et al. 2014). Both of these studies put forward the hypothesis that NSD targeting to heterochromatin is intertwined to the Su(var)3-9/K9me2/ HP1α regulatory axis. To study if there is any interdependency between the two pathways, I now examined i) how NSD-dependent K36me2/3 is affected upon Su(var)3-9 RNAi and ii) if Su(var)3-9-dependent K9me2 is affected upon NSD RNAi. These reciprocal RNAi perturbations would allow the establishment of whether (and how) these pathways are related.

Firstly, the bulk changes in the histone modifications upon RNAi were assayed using Western blot (Figure 22A). This confirmed a strong reduction of K36me2 upon NSD RNAi and K9me2 upon Su(var)3-9 RNAi. However, these changes were not coupled to any changes in the reciprocal modification, suggesting that the pathways are disconnected.

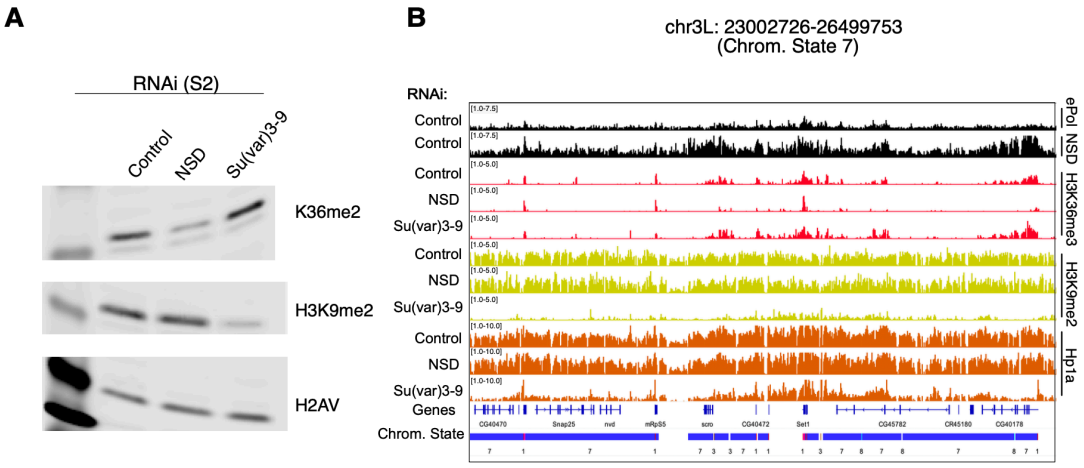


Figure 22: A. Western blots for K36me2, H3K9me2 and H2AV (Control) after RNAi of indicated factors. **B.** Genome browser view of a heterochromatic region for K36me3, H3K9me2 and HP1α (IP target shown on right side) after RNAi of indicated factors. Chromatin state as well as NSD and ePol profiles included for reference.

I then resorted to MNase-ChIP to inspect the differences at a high resolution. The ChIP results echo the observations from Western blotting (Figure 22B). Despite the loss of heterochromatic K9me2/HP1α by Su(var)3-9 depletion, NSD-dependent K36me3 is unaffected in the same condition. These patterns are true for euchromatic NSD targeting as well. Similarly, despite strong reduction of NSD-dependent K36me3 upon NSD RNAi, neither K9me2 or HP1α are affected, suggesting that heterochromatic K36 and K9 methylation operate through separate pathways. Overall, these observations suggest that other genetic (specific transposon sequences) or epigenetic (other heterochromatic marks such as H4K20me3) factors are implicated in NSD recruitment.

3.3.3 Does Ash1 have a function at enhancers?

The correlative analyses in unperturbed cells had established that K36me1 domains harbor developmental enhancers and these domains appear to exclusively depend on Ash1, based on the RNAi perturbations. Thus, Ash1 RNAi provides a tractable system to study the possible role of K36me1 in regulation of enhancers. Using H3K27ac as a readout of enhancer activity, I sought to characterize if there are any changes in the enhancer landscape upon Ash1 depletion. Despite strong reduction of K36me1 upon Ash1 RNAi, no difference in K27ac was observed both at housekeeping or developmental enhancers, suggesting that the acetyltransferase CBP that places the H3K27ac mark is not modulated by Ash1 (Figure 23).

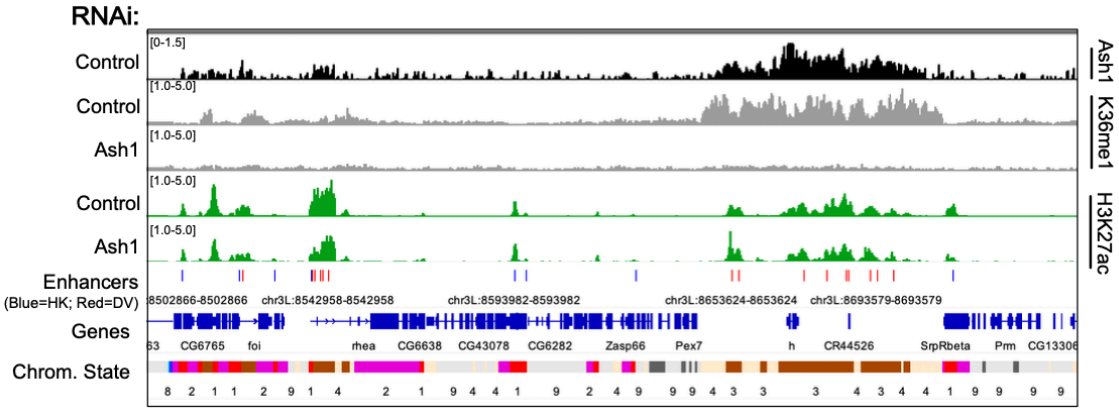


Figure 23: Genome browser view of K36me1 and H3K27ac (IP target shown on right side) after RNAi of indicated factors. Chromatin state, Enhancers (HK=housekeeping; DV=developmental) as well as Ash1 profile included for reference.

It is important to note that this result doesn't preclude the possibility that Ash1 may affect enhancer function at another level, for example by regulating spatial organization of enhancer-promoter contacts or recruitment of cofactors, etc. As the ChIP methodology is not well suited to study enhancer function (as opposed to targeted assays like STARR-seq), I decided not to pursue this further.

3.4 A gene-centric view of H3K36 methylation landscape

I decided to focus on the patterns of K36 methylation at a genic level, as such an analysis would facilitate detailed comparisons with interesting genomic features, like transcription or gene function. To do this, all annotated genes that overlapped at least one domain of K36me1, K36me2, or K36me3 were selected. This selection produced around 10,500 genes enriched for at least one K36 methylation state. ChIP-seq signals were averaged over the entire gene body across all RNAi conditions and the data was subjected to unbiased hierarchical clustering. This resulted in 12 clusters which vary as a continuum in steady-state distribution of K36me1/2/3 as well as HMT dependency (Figure 24A). To avoid overinterpretation and to allow broad elucidation of the data, I manually merged the clusters into three larger superclusters (SCs) as indicated (cluster no. 12

was excluded from all subsequent analyses as it mostly comprised genes with very low K36me1/2/3 but had minor overlaps with adjacent 'stronger' domains).

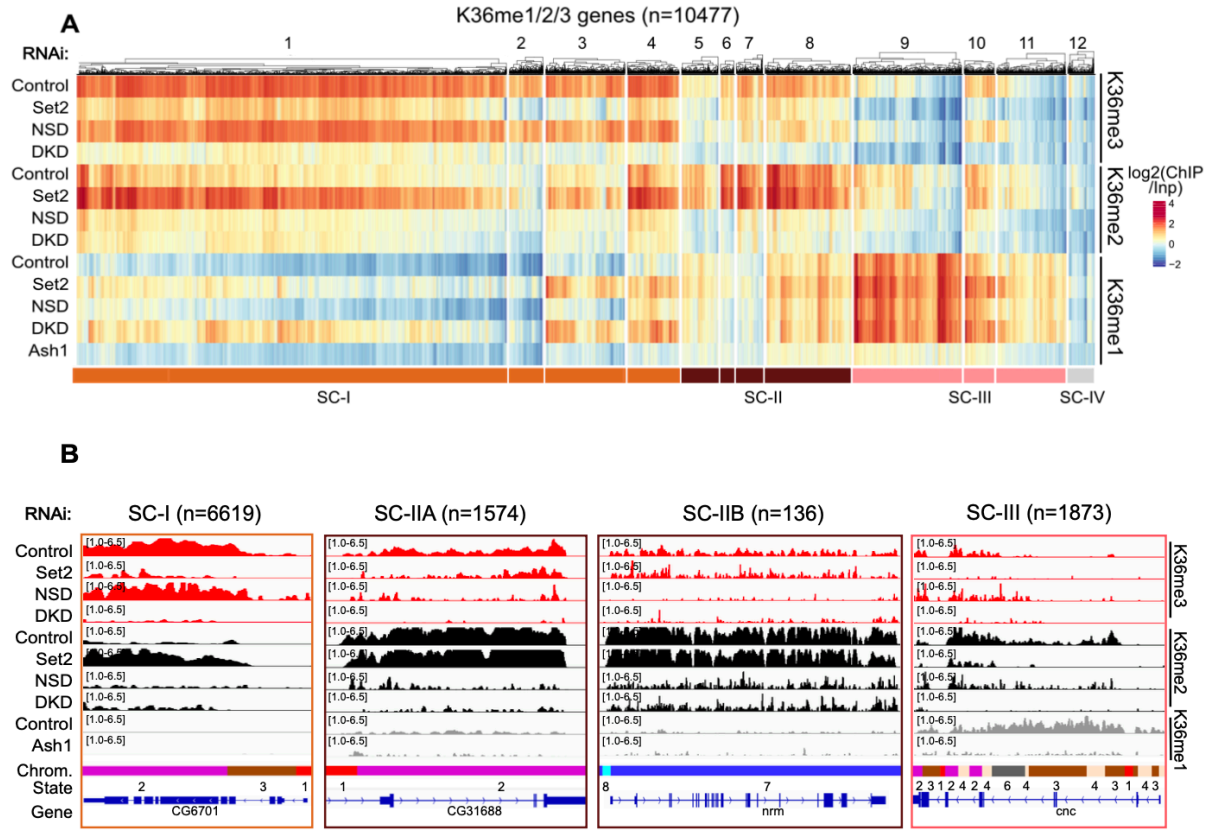


Figure 24: A. Hierarchically clustered heatmaps of gene body-averaged ChIP signal for K36me1/2/3 indicated on the right and RNAi condition indicated on the left. Only genes overlapping at least one of K36me1/2/3 peaks in any RNAi condition (*n* = 10477) were included in the heatmap. Clusters are numbered 1–12 as indicated above the heatmap and were further manually grouped to define Superclusters (SC) I-IV **B.** Genome browser profiles of representative genes from each supercluster along with the number of grouped genes. Supercluster II was further classified into Euchromatic (IIA) or Heterochromatic (IIB) based on overlap with H3K9me2 peaks (quantified in Figure 25). RNAi condition and ChIP target indicated on the left and right respectively. Figure adapted from Jayakrishnan et al., 2024.

3.4.1 Set2 is the major driver of K36 methylation patterns in Supercluster-I genes

SC-I genes are characterized by very strong levels of K36me3, low levels of K36me2 and absence of K36me1 in unperturbed S2 cells. K36me3 at these genes is exclusively dependent on Set2, paralleling the observations from the chromoMaps (Figure 24B; SC-I). These genes also carry low but non-zero amounts of K36me2, which show a slight sensitivity to NSD RNAi. Interestingly, at these genes K36me2 is strongly gained upon Set2 RNAi. This effect is lost upon the combined depletion of both NSD and Set2, strongly implicating augmented K36me2 deposition by NSD in the absence of Set2. These genes appear to accumulate K36me1 upon depletion of Set2, as they are likely intermediates of turnover reaction by demethylases.

To summarize, Set2 deposits K36me3 highly efficiently at these genes mostly without the need for prior methylation by NSD. The dimethylation activity of NSD is fairly low within this supercluster in control conditions, but may be boosted in the absence of Set2.

3.4.2 NSD predominantly controls K36me2/3 in Supercluster-II genes

In contrast to SC-I genes, SC-II genes are characterized by very high levels of K36me2, moderate K36me3 and variable amounts of K36me1 in unperturbed S2 cells. This genic K36me2 is sensitive only to NSD RNAi, confirming that NSD is the major dimethyltransferase at these genes. To verify if these genes all reside in heterochromatin (based on the chromoMap results), the fraction of SC-II genes that lie in K9me2 domains was calculated (Figure 25). Surprisingly, only 8% (136/1710) of these genes are heterochromatic, confirming that NSD methylates many euchromatic genes in addition to PCH (note the euchromatic 'blue' bins in K36me2/3 chromoMaps upon NSD RNAi).

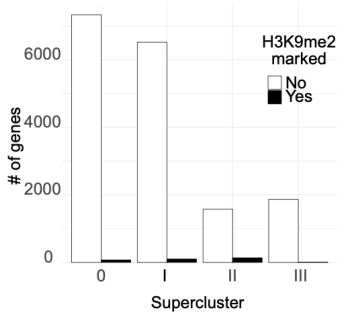


Figure 25: Quantification of eu-/heterochromatic genes for each supercluster-specific genes based on overlap with K9me2 peaks. Figure adapted from Jayakrishnan et al., 2024.

These euchromatic genes (SC-IIA) were separated from the heterochromatic genes (SC-IIB) for all following analyses/discussions (Figure 24B; SC-IIA, B). K36me3 at these genes are partly sensitive to Set2 RNAi in addition to NSD RNAi, with DKD eliminating the signal almost entirely. This additivity argues that Set2 and NSD cooperate but independently deposit K36me3 at these genes. Lastly, low levels of K36me1 are dependent on Ash1 in these genes.

3.4.3 Ash1-driven K36me1 exemplifies Supercluster-III genes

The most distinguishing feature of this supercluster is the presence of strong levels of K36me1, which depends only on Ash1 (Figure 24B; SC-III). Various clusters within this supercluster have moderate levels of K36me2/3 which depend on NSD and Set2 to varying degrees, suggesting that NSD/Set2 may possibly act on K36me1-deposited by Ash1. The sensitivity of K36me2 to NSD RNAi (and Set2 RNAi in some cases) at those genes was puzzling, as Ash1 has previously been described as a dimethyltransferase. As I hadn't directly profiled K36me2 in Ash1 RNAi in this study, I utilized a previously published dataset (Huang et al. 2017). By visualizing K36me2 changes upon Ash1 RNAi within each of the superclusters, I confirmed that K36me2 within SC-III is sensitive to Ash1 RNAi as well (Z-score < 0) (Figure 26A, B).

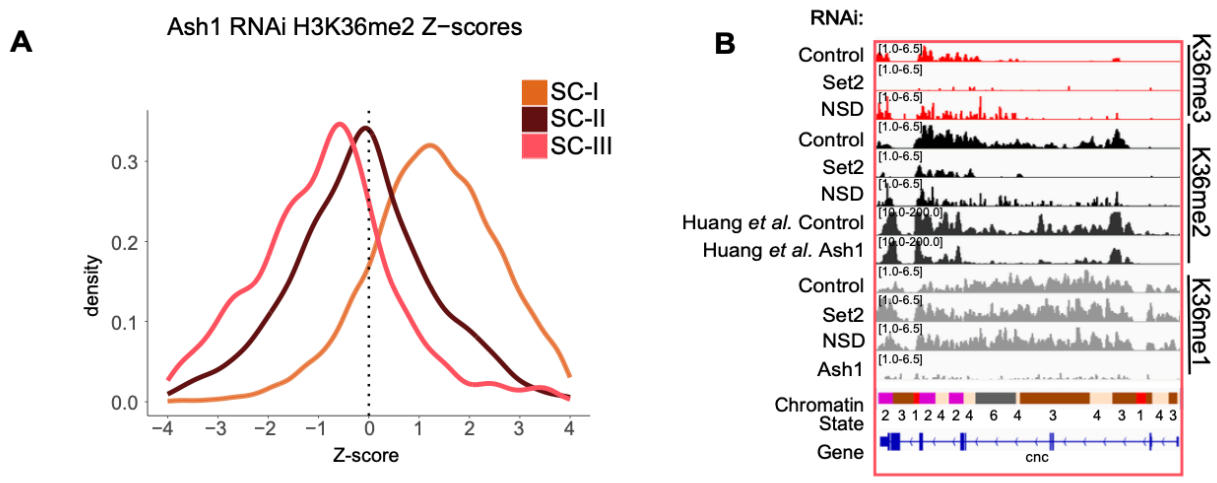


Figure 26: A. Density plot of *Z*-scores representing change in K36me2 signal upon Ash1 RNAi for superclusters defined in Figure 24. **B**. Genome browser profiles of SC-III representative region demonstrating effect of Ash1 RNAi on K36me2. RNAi condition and ChIP target are indicated on the left and right respectively. Figure adapted from Jayakrishnan et al., 2024.

Of note, SC-I shows a marked increase of K36me2 in the same condition, reminiscent of the increased NSD-dependent K36me2 observed upon Set2 RNAi within SC-I.

In summary, while Ash1 appears to be the unique monomethyltransferase within this supercluster, multiple HMTs cooperate to establish K36me2/3 at a subset of these genes.

Taken together, these results embody the complex regulatory landscape of genic K36 methylation. The dependency of genic methylation on the different HMTs is gradual, but each supercluster is typified by a dominant modification which predominantly depends on one methyltransferase.

3.5 Intragenic distribution of K36me1/2/3 vary across superclusters

H3K36me3 has been frequently described as a 3'-enriched mark (Pokholok et al. 2005) and has been linked to slowing down of RNA polymerase as it approaches Transcription Termination Sites (TTS). However, the detailed profiling of all modifications suggested that the intragenic distribution may differ depending on the context. To address this, cumulative plots representing K36me1/2/3 signal over scaled gene bodies for each supercluster were generated (Figure 27).

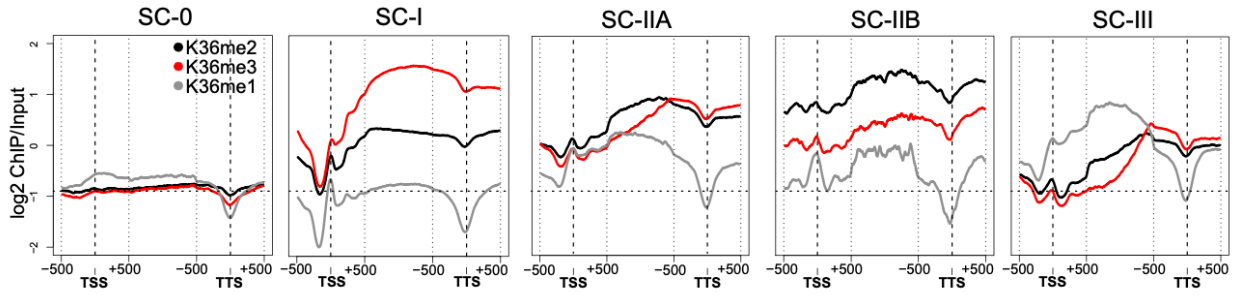


Figure 27: Gene body K36me1/2/3 cumulative plots for each supercluster. 1-kb regions centered around TSS and TTS are unscaled, while the rest of the gene body was scaled to 500 bins. SC-0 genes lack any detectable H3K36 methylation and serve as a reference for zero signal/baseline (also represented by the horizontal dotted line). Genes smaller than 1500 bp were excluded from the analysis. Figure adapted from Jayakrishnan et al., 2024.

SC-0 was defined as the set of all genes that lack any K36 methylation (and thus were not included in the above clustered heatmap), and serves as a useful baseline reference. Based on this, I confirmed that K36me3 within SC-I genes show the classically described 3' skew. This 3' bias was also observed for K36me3, but not K36me2 within SC-IIA, hinting at different deposition kinetics/modes for the two marks. In contrast to the above cases, the distribution of K36me2/3 at SC-IIB genes is relatively uniform and comparable to up-/downstream regions. This observation fits with genome browser views that these genes are embedded in large blocks of K36me2/3 at PCH. Finally, K36me1 at SC-III genes shows a 5' bias antagonistic to K36me2/3. One possible simple explanation for this bias is that the observed 3' K36me2/3 are obtained by further methylation of preexisting K36me1, resulting in its apparent depletion. This difference in distribution also lends support to the idea that multiple HMTs (with different preferences to TSS versus TTS) act on SC-III genes.

3.6 Set2 and NSD deposit K36me at distinct transposon families

K36 HMTs were previously suggested to regulate transposable element (TE) expression in *Drosophila* (Lindehell, Schwartz, and Larsson 2023). To verify how K36me2/3 changes at TE upon depletion of HMTs, I assembled a 'custom genome' containing the different *D.mel* repeat families across RNAi conditions and performed a hierarchical clustering analogous to the gene-based analysis above (Figure 28).

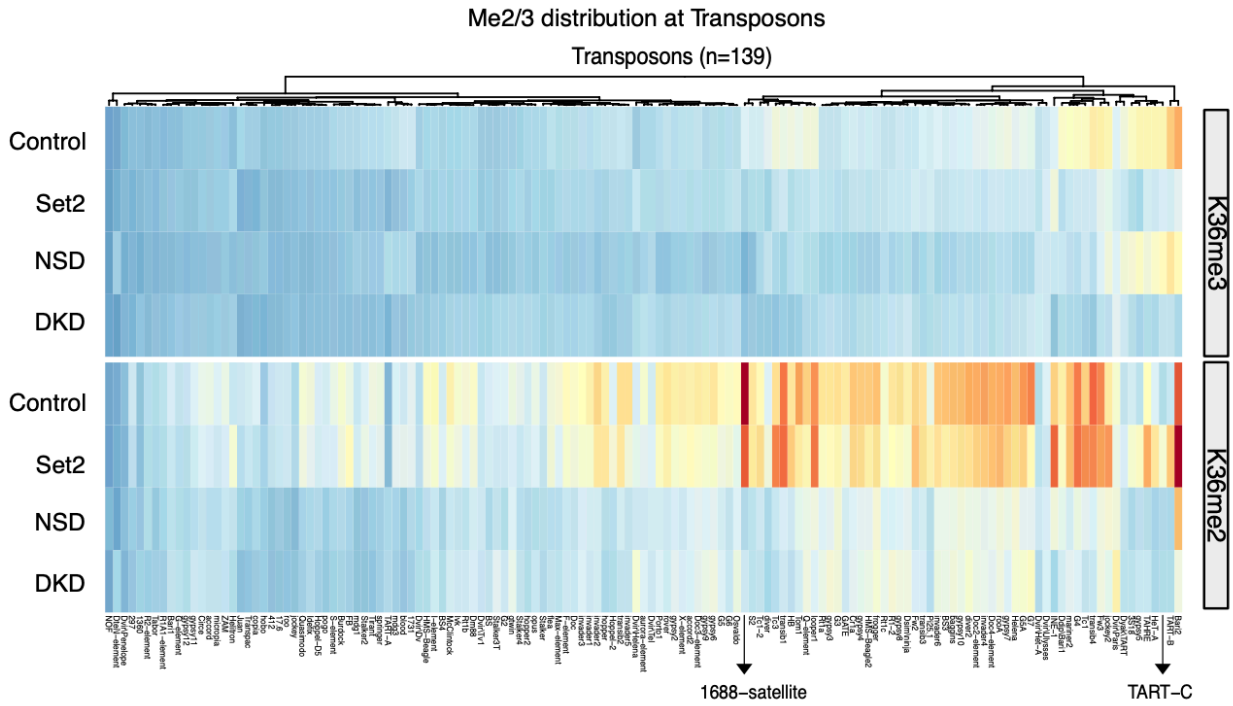


Figure 28: Clustered heatmap representing average K36me2/3 ChIP signal at 139 transposon families (names indicated on the right). RNAi target indicated on the left. High to low ChIP signal is mapped linearly to the Red to Blue color scale. TART-C is a representative example of telomeric transposons while 1.688-satellite repeats are enriched on heterochromatic arm of X-chromosome. Figure adapted from Jayakrishnan et al., 2024.

Around 50% of TE families were only weakly methylated across all conditions. Telomeric transposons appeared to be the major TE type that were exclusively methylated by Set2, which are characterized by high K36me3/low K36me2, similar to SC-I genes. Many transposon families that are characterized by NSD dependence, marked by high K36me2 and low K36me3, are probably embedded in pericentric heterochromatin (Lindehell, Schwartz, and Larsson 2023) as highlighted in the chromoMaps of Figure 21. These observations imply that many K36me2/3-linked gene regulatory pathways may be active on transposons as well.

3.7 Gene superclusters defined by K36 methylation patterns correlate to distinct genic features and functions

To better describe the superclusters defined by K36 methylation changes upon HMT depletion, I correlated the superclusters to several informative features. Firstly, it was checked if each HMT is principally enriched in their respective superclusters, with SC-0 serving as an internal negative control (Figure 29A).

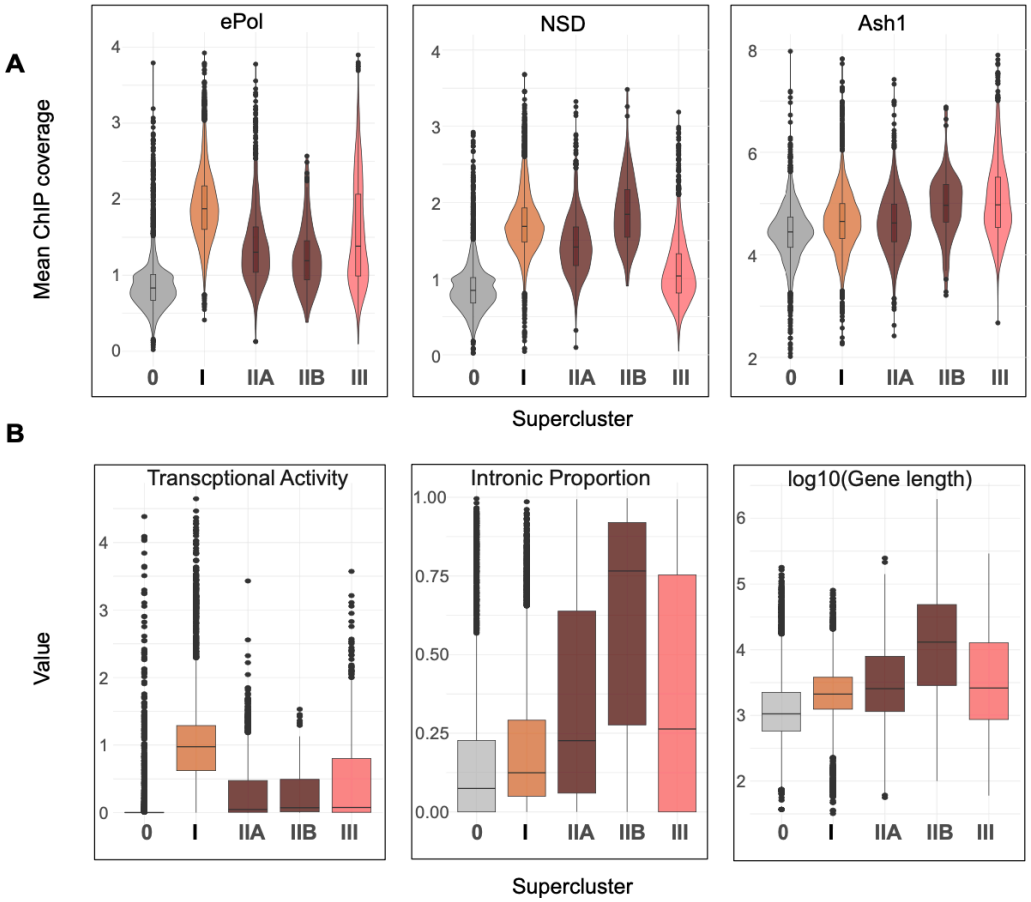


Figure 29: A. Violin plots representing average ChIP signal for 2-kb windows around HMT gene body peaks for NSD, ePol and Ash1 within SC 0-III. SC-0 represents randomly sampled genes (n = 3000) lacking any detectable K36 methylation and serves as reference for zero signal/baseline. **B.** Boxplots showing the proportion of introns (sum of length of introns/total gene length), transcriptional activity (log_{10}(RNA-seq TPM)) and log_{10}(gene length) for indicated superclusters. Figure adapted from Jayakrishnan et al., 2024.

ePol (serving as a proxy for Set2) showed the largest enrichment in SC-I, followed by SC-III, consistent with the predominant localization of the Set2-dependent K36me3/2. ePol was present at lower levels in SC-IIA and SC-IIB, where genes strongly depend on NSD for K36me2/3. Next, NSD showed the strongest signal with SC-IIB genes concordant with its strong heterochromatic localization. Interestingly, SC-I genes had a higher enrichment of NSD compared to SC-IIA genes, which is opposite to the abundances of NSD-dependent K36me2 observed at these superclusters. In other words, despite attracting higher levels of NSD, SC-I genes fail to reach the same amount of K36me2 as compared to SC-IIA genes. This suggests that targeting and activity of NSD might be uncoupled. NSD activity may be dampened at SC-I genes (possibly due to competition with Set2 for shared cofactors) or boosted at SC-IIA genes by specific cofactors. Lastly, using the published Ash1 profile, a very mild enrichment at SC-III genes could be observed, with other superclusters displaying a similar distribution as the background. Of note, the observed slight SC-IIB enrichment appears artefactual as these signals do not disappear upon

Ash1 RNAi (data not shown). These observations argue that while HMTs tend to localize predominantly to their main supercluster, they can also be mapped on other genes, but their activities may vary in a locus-specific manner.

The properties of genes belonging to each supercluster were then examined (Figure 29B). SC-I genes tend to be short, mostly exonic and highly expressed, consistent with the abundant ePol. On the other hand, SC-II and SC-III genes tend to be much longer, contain more introns and are expressed at much lower levels. Genes of the latter type are often differentially expressed in tissues and have developmental functions. To verify this notion, I sought to classify genes as 'developmental' or 'housekeeping' based on the degree of expression variability across 25 different fly tissues as described by FlyAtlas (Chintapalli, Wang, and Dow 2007). Plotting the distribution of the mean expression-normalized variance allowed the definition of an arbitrary threshold (Figure 30A).

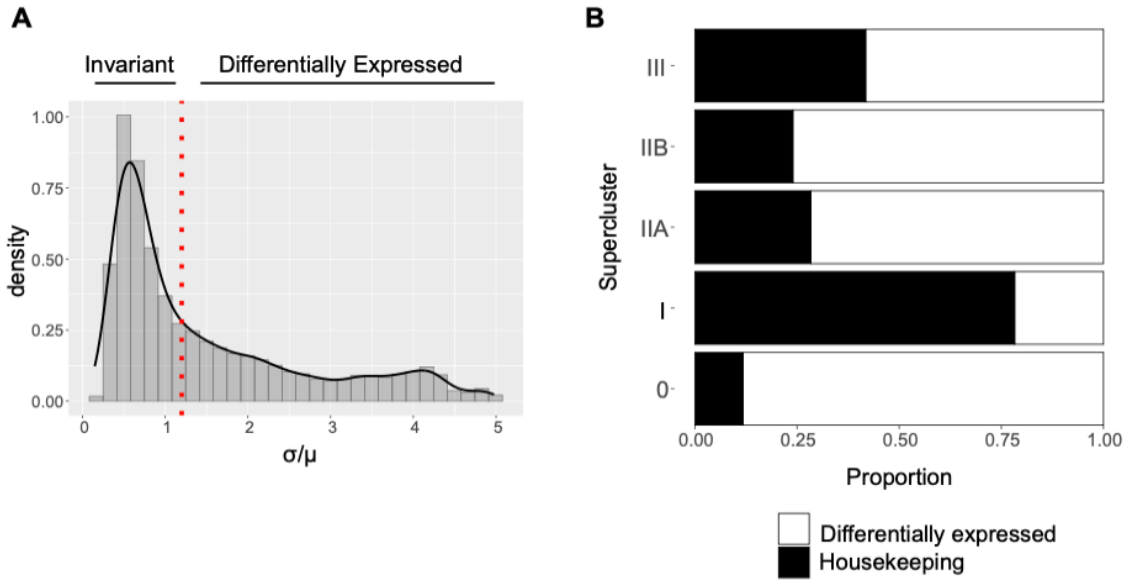


Figure 30: A. Distribution of σ/μ values (σ = expression standard deviation, μ = expression mean; calculated across 25 FlyAtlas Tissues). The red line represents the arbitrary threshold distinguishing tissue invariant and differentially expressed genes. **B.** Proportion of tissue-specific/invariant genes for SC 0-III genes based on variance threshold defined in panel A. Figure adapted from Jayakrishnan et al., 2024.

Of note, genes with very low expression in all 25 tissues would have low variance values and would be classified as 'housekeeping' based on the threshold. Such genes were subsequently reassigned to the 'developmental' category based on the rationale that they might be highly expressed in niche cell types not profiled in this study. Calculating the proportion of housekeeping versus developmental genes in each supercluster confirmed my hypothesis: SC-I genes are largely expressed in a tissue-invariant manner while SC-II and SC-III genes are expressed in a more specialized manner (Figure 30B).

To get a more detailed picture regarding the functions of these genes, a Gene Ontology analysis was performed and related GO-terms were grouped using Semantic Similarity Indexing (RHReynolds 2022) (Figure 31).

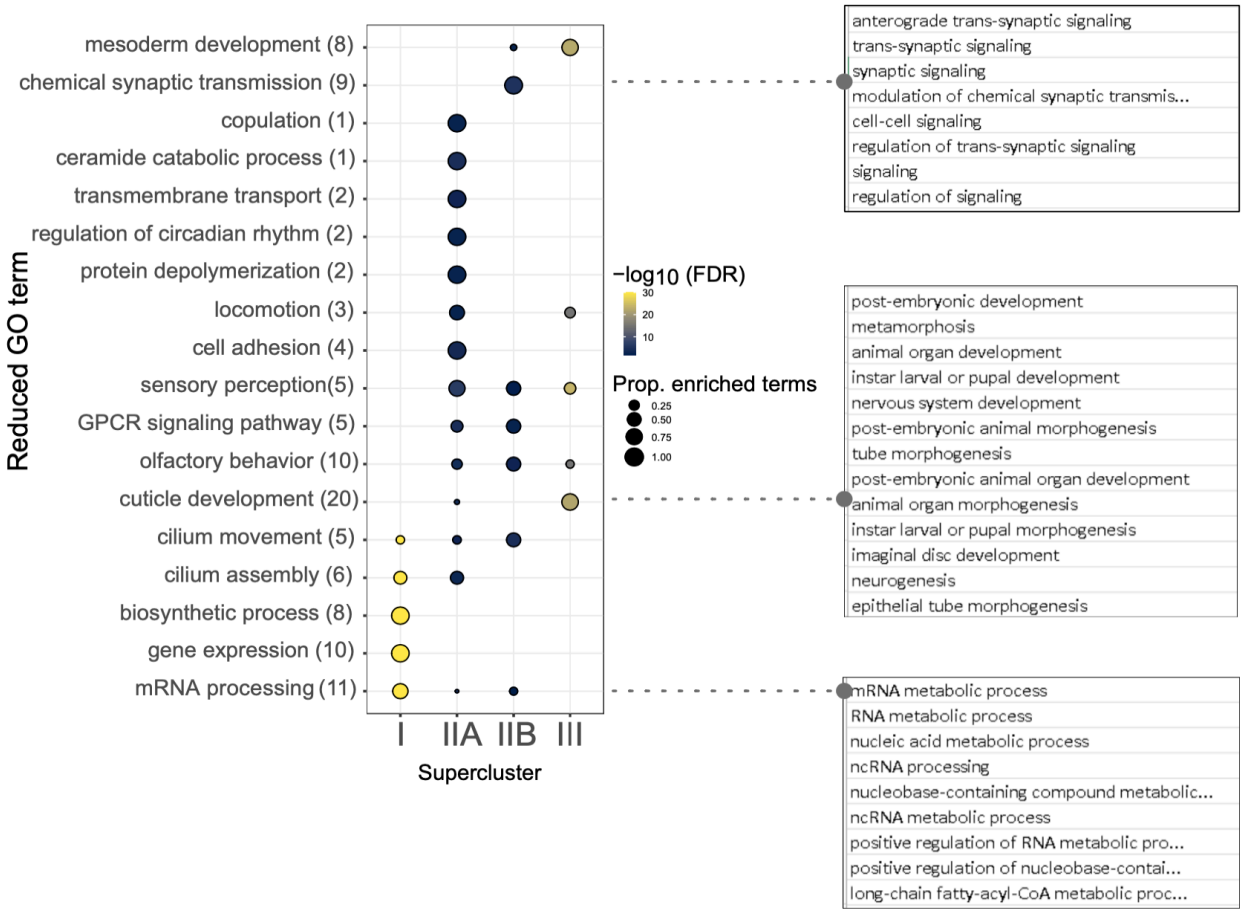


Figure 31: Gene function GO-term analysis for SC I-III genes. Obtained GO terms were grouped based on Semantic Similarity Analysis (ReduceGO (RHReynolds 2022)) and were defined by a 'parent' GO term indicated on the left. The associated number within brackets indicates the number of 'children' GO-terms grouped under the same category. The size of the dots indicates the proportion of terms contributed by genes belonging to the indicated supercluster relative to all supercluster genes (. i.e., rows should sum to 1). Example 'children' GO terms are shown for selected 'parent' GO terms.

GO terms for SC-I genes were related to housekeeping functions, like 'transcription' and 'synthesis of biomolecules'. Significant GO-terms for SC-IIA and SC-IIB genes overlapped partially and differed from SC-I, and were largely related to neuronal function (olfaction, sensory perception etc.). In contrast to the above, SC-III genes appeared to be involved in fly morphogenesis.

So far, one recurrent observation is the difference in intron/exon composition between the genes of the different supercluster, which may be linked to the specific features of the HMT. For example, K36me3 has been previously described to be enriched on exons, presumably due to slowdown

of ePol due to transcription-coupled splicing (Kolasinska-Zwierz et al. 2009). More recently, K36 methylation states have also been described at introns (Delandre et al. 2022). Since gene body-averaged approaches as used in the previous heatmap (Figure 24) do not distinguish intron from exon signals, I repeated a subset of the analyses by quantifying K36me1/2/3 within introns and exons for SC I-III genes across the different knockdowns (Figure 32). Features spanning regions smaller than 500 bp (~3 nucleosomes) were excluded from the analyses.

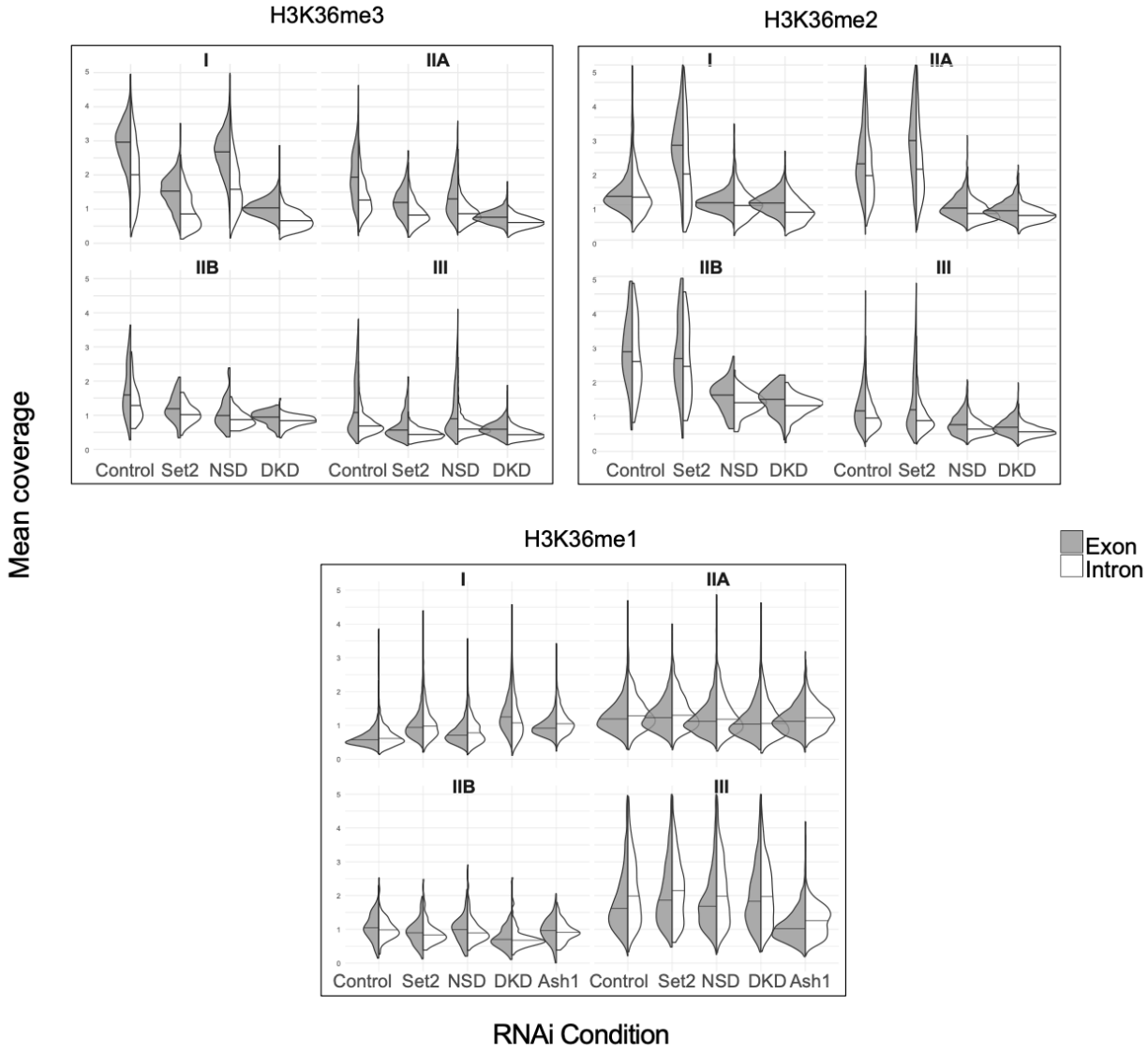


Figure 32: Split violin plots representing changes in distribution of K36me3/2/1 ChIP signal in introns (white) and exons (gray) of SC I-III genes in Control and HMT RNAi. Figure adapted from Jayakrishnan et al., 2024.

K36me3 at SC-I genes confirms the strong exon enrichment, and lower but significant signal on introns. Both populations were largely sensitive to Set2 RNAi only. K36me2, on the other hand, was present at much lower levels on both introns and exons in control conditions, strongly

increased upon Set2 RNAi and was reduced further upon NSD RNAi. SC-IIA genes provide a more interesting case. Despite having much higher intronic proportion, K36me3 was still preferentially enriched on exons. Comparison of the distribution in Set2 RNAi and NSD RNAi confirms that there is no division of labor at exons and introns, that is each HMT contributes to K36me3 at both introns and exons, which is lost cumulatively in the DKD. In contrast, K36me2 showed a much lower exon bias (similar distribution curves for introns and exons) and most of the K36me2 was lost only upon NSD depletion. These analyses suggest that K36me3 in general likely follows an exon bias, whether deposited by Set2 at SC-I genes or by Set2/NSD at SC-IIA genes. Conversely, K36me2 is rather uniformly distributed over exons and introns. Lastly, K36me1 is largely present on exons of SC-III genes and dependent on Ash1.

3.8 Do methyltransferases affect each other's localization?

So far, my results show the emergence of complex K36me1/2/3 patterns upon depletion of HMTs, which can be attributed to a mix of direct methylation and indirect effects, like demethylation and feedback regulation. Another aspect of this apparent complexity may be the relocalization of individual HMTs in absence of other enzymes. To assess such a scenario, a preliminary experiment (N=1) was performed where the localization of ePol (proxy of Set2) in NSD RNAi and NSD in Set2 RNAi was profiled and visualized adjacent to the previously described clusters (Figure 33).

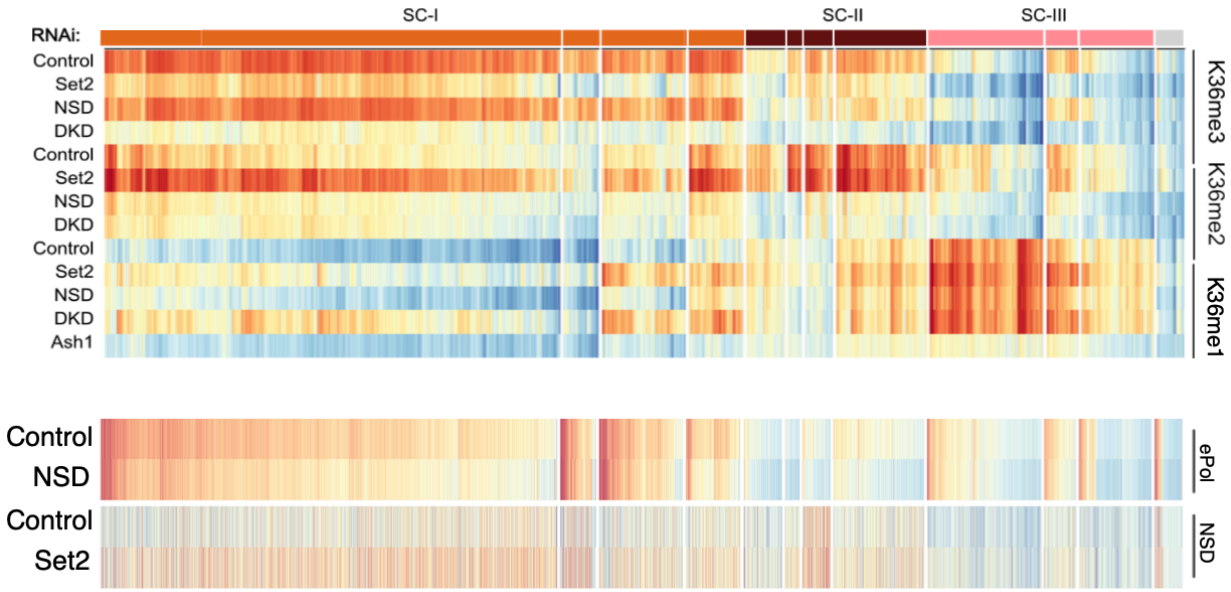


Figure 33: Heatmap of gene body-averaged signals of ePol (Set2) and NSD occupancy after NSD or Set2 RNAi, respectively. Heatmap was clustered and visualized with same color scale according to main K36me1/2/3 heatmap from Figure 24A. Figure adapted from Jayakrishnan et al., 2024.

This exercise confirmed my original hypothesis; depletion of one HMT led to the redistribution of the other. Upon NSD RNAi, ePol was partly displaced from chromatin on several gene clusters belonging to superclusters I and II (curiously, SC-III was less affected). This would suggest that

some of the K36me3 decreases observed upon NSD RNAi may result from indirect loss of the passage of elongating polymerase complexes, an idea more consistent with a chromatin adaptor-like role for NSD to recruit Set2. Of note, this observation doesn't necessarily rule out direct K36me3 deposition by NSD, considering that RNAi of Set2 or NSD individually only partially reduces K36me3, while a combined depletion completely eliminates it, arguing for parallel deposition pathways.

In contrast to the above, profiling NSD upon depletion of Set2 shows an opposite effect, with many clusters showing a uniform increase in occupancy of NSD. In hindsight, this observation ties up nicely with the previous finding that K36me2 is dramatically increased upon Set2 RNAi which was attributed to increased activity of NSD. Now, it would appear that this increased K36me2 may be also contributed to by higher levels of NSD at these regions, alongside a possible feedback activation. Why does Set2 depletion lead to more NSD on chromatin? The answer is unclear, but may possibly be linked to the PWWP domain of NSD. Unlike the two other K36 HMTs, NSD is an atypical 'writer-reader' methyltransferase which could result in a K36 methylation maintenance function. The specificity of the 'reader' domain of NSD is unknown, but may recognize certain modifications appearing upon Set2 RNAi to establish a positive feedback loop of K36me2 maintenance. Nevertheless, the preliminary observations lend support to the idea that HMTs affect each other's localization to some extent, a phenomenon that is rarely discussed in the literature when interpreting HMT-histone modification relationships.

3.9 Acute chemical inhibition of transcriptional elongation partly recapitulates Set2 RNAi patterns

I wanted to explore alternate ways to perturb the K36 methylation profile to reinforce the main conclusions from the HMT depletion experiments. To achieve this, the molecular inhibitor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) was used, which inhibits activating phosphorylations of the RNA polymerase II C-terminal domain in the timescale of hours. Since the distribution/activity of Set2 is coupled to the elongating polymerase (ePol), the prediction would be that this inhibition would disrupt chromatin targeting of Set2, without side effects that may occur upon depletion of the enzyme (such as methylation of non-chromatin substrates, consumption of metabolites, cell cycle effects over the 7-10 day RNAi period, etc.). With this purpose in mind, an acute 6-hr DRB treatment was performed and subsequently ePol and K36me2/3 was profiled and the patterns visualized adjacent to the previously described clustered heatmap (N=1) (Figure 34).

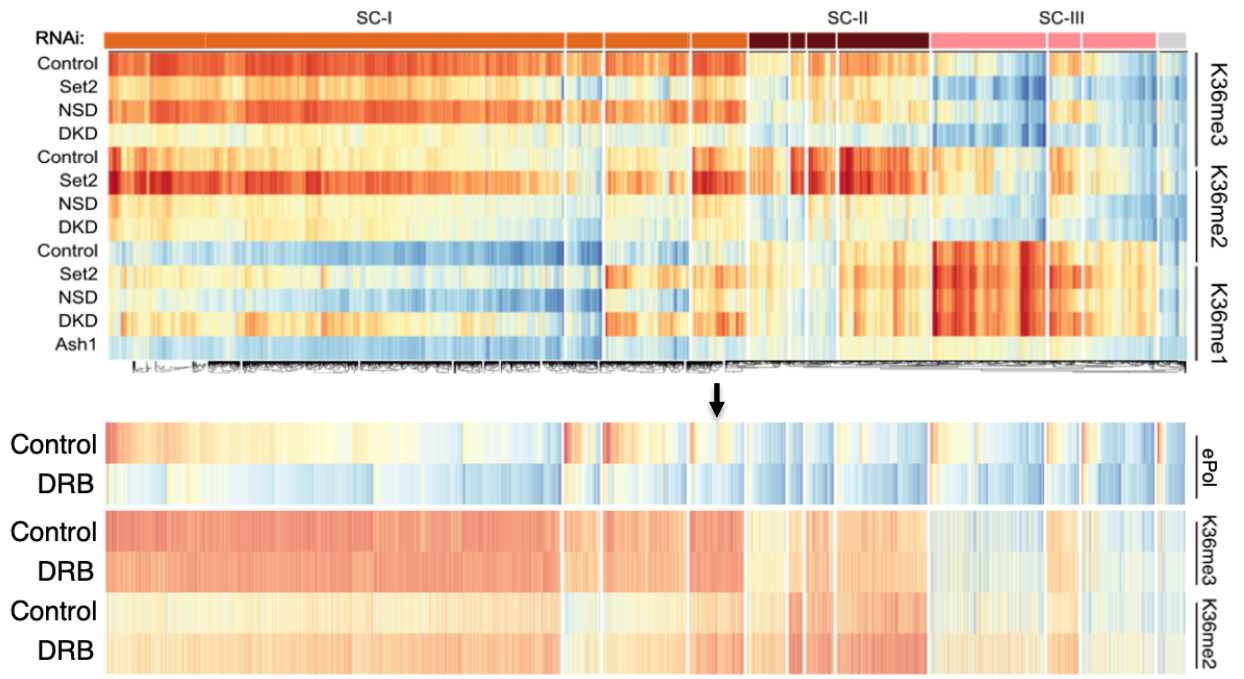


Figure 34: Heatmap of gene body averaged signals of ePol and K36me2/3 occupancy after Control or DRB treatment. Heatmap was clustered and visualized with same color scale according to main K36me1/2/3 heatmap from Figure 24A. The arrow indicates a representative cluster discussed in the main text. Figure adapted from Jayakrishnan et al., 2024.

A strong reduction of ePol can be observed in all clusters upon DRB treatment relative to the control DMSO treatment, confirming successful inhibition of transcriptional elongation. Interestingly, K36me3 was only mildly affected, with many genes showing no change at all. This suggests that the actions of demethylases at active genes are limited, as suggested by recent studies which attribute a more heterochromatic role for KDM4A (Colmenares et al. 2017). This would also mean that the reduction in K36me3 signal that was observed upon HMT depletion is likely to be mainly driven by dilution over cell divisions rather than active demethylation. Unlike K36me3, K36me2 shows a fairly strong increase upon DRB inhibition similar to Set2 RNAi (although to a lower degree). This K36me2 increase occurred even within clusters not showing any changes in K36me3 (see example cluster indicated by arrow in Figure 34), suggesting that the K36me2 is not generated by demethylation of preexisting K36me3 and could be linked to increased NSD chromatin binding. This ‘decoupling’ of the K36me3 and K36me2 pathways supports the idea that Set2/ePol machinery restricts NSD (and K36me2) at these highly transcribed euchromatic genes.

3.10 Summary of K36me deposition pathways and HMT relationships

The below schematic attempts to summarize the findings described so far. The three HMTs dominantly exert their methylation activities in distinct domains with minor overlaps (Figure 35A).

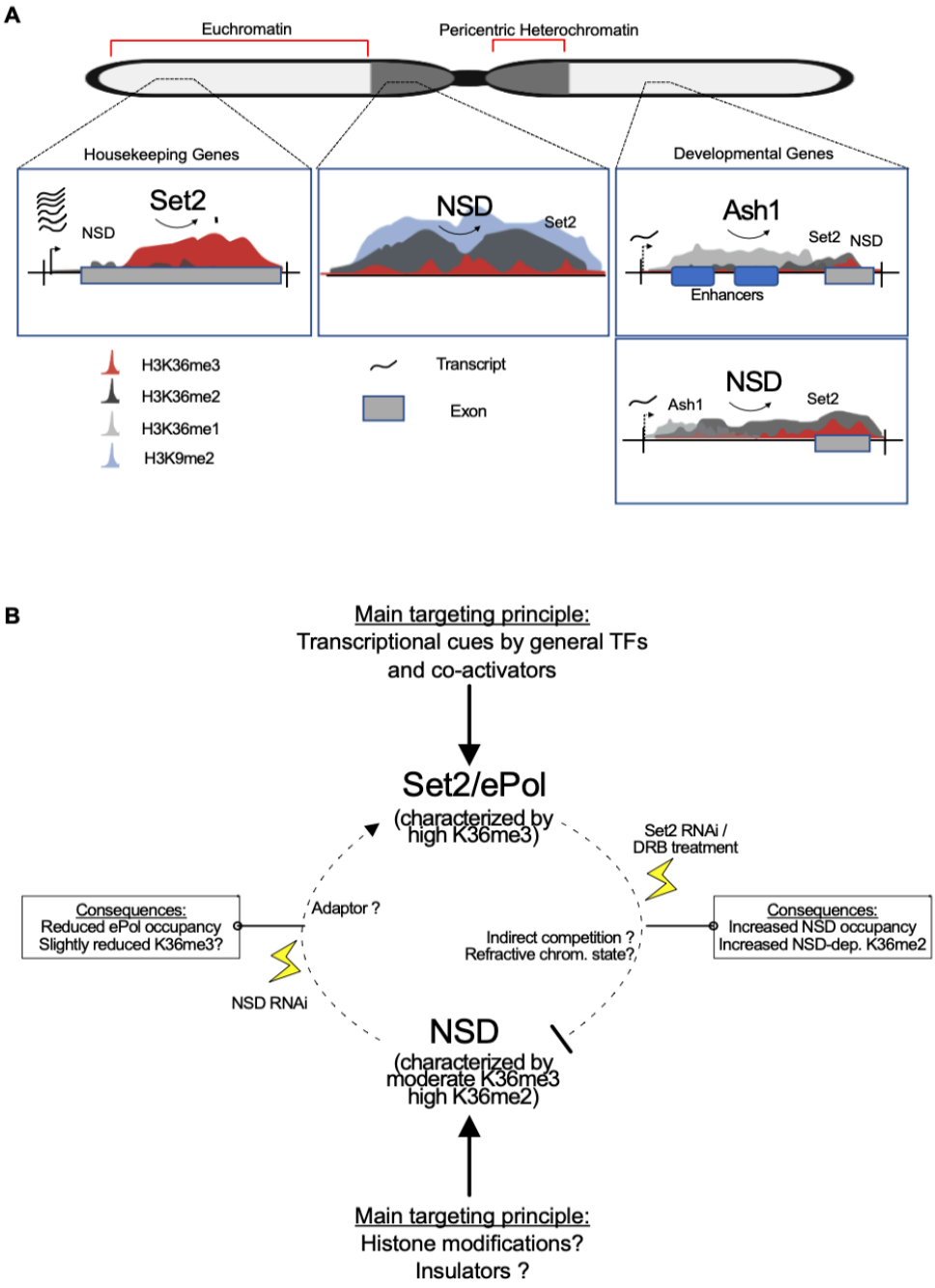


Figure 35: A. Graphic model depicting the context-specific contributions of H3K36 HMTs towards generating K36me1/2/3. Schematic was generated using Biorender.com. Figure adapted from Jayakrishnan et al., 2024. **B.** Flowchart summarizing the relationship between Set2 and NSD HMTs.

Set2 and NSD K36me deposition pathways are likely dictated by independent cues, but partially affect each other through a positive/negative dual feedback mechanism (Figure 35B). NSD promotes Set2/ePol localization, and loss of NSD can manifest in reduced additional contribution of Set2 towards K36me3. On the other hand, Set2/ePol negatively influences NSD

targeting/activity, as perturbation by either Set2 RNAi or DRB results in increased NSD binding to chromatin coupled with augmented K36me2 deposition. More sophisticated experiments in the future may help tease apart the targeting determinants as well as the exact mechanism of how Set2 and NSD affect each other.

3.11 Some features of Set2/NSD methylation are conserved in mouse cells

Recent studies in mice have revealed the presence of NSD-dependent K36me3 domains (Barral et al. 2022), suggesting that the NSD family of proteins (NSD1/2/3) retained the trimethylation activity during the course of evolution. To determine whether the genic methylation patterns observed in *Drosophila* are also conserved in mammals, I reanalyzed two recently published mouse datasets (Weinberg et al. 2019; Sun et al. 2023).

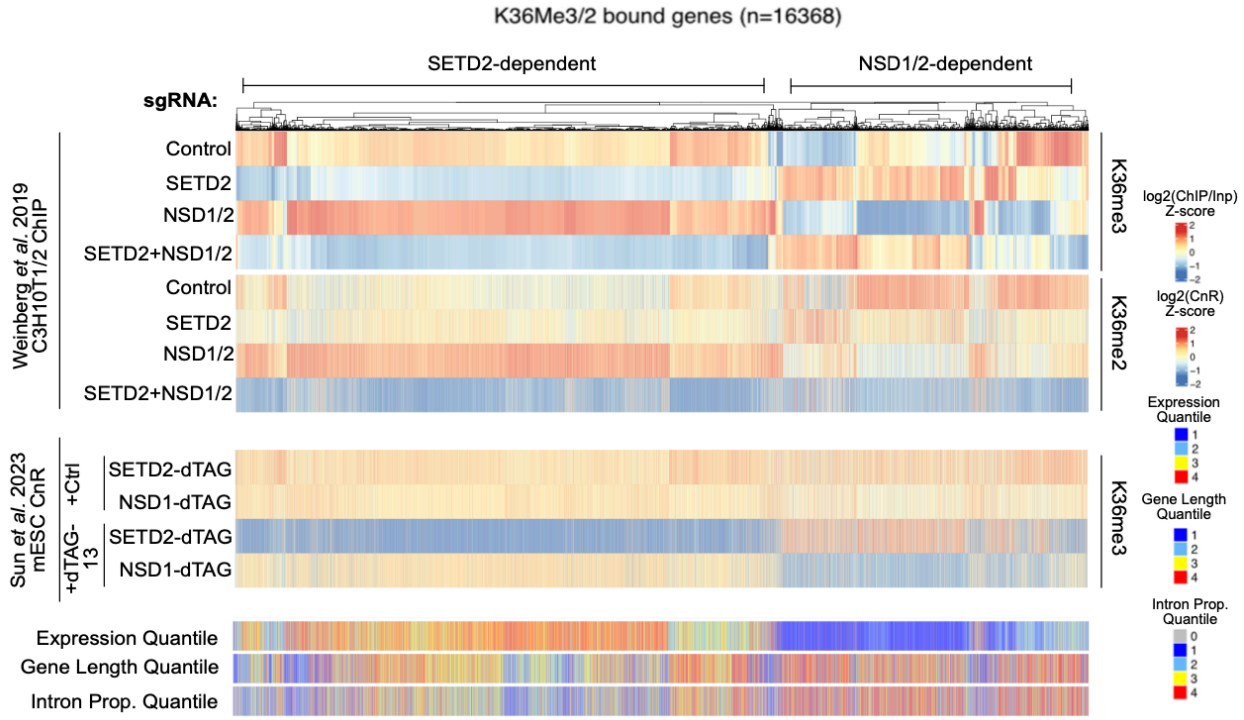


Figure 36: Top: Reanalysis of primary data from Weinberg et al., 2019. Heatmap representing gene body-averaged H3K36me3/2 ChIP signal (indicated on right) in Control or HMT knockout background (indicated on left) mouse C3H10T1/2 cells. Genes not overlapping any of K36me2/3 peaks in control cell line (n=16368) were excluded before clustering. Values were standardized for each gene (by subtracting mean and dividing by standard deviation) across KO conditions separately for K36me2/3 prior to clustering. **Middle:** Reanalysis of primary data from Sun et al., 2023. Heatmap representing gene body-averaged K36me3 CUT&RUN signal in untreated or dTAG-13 treated HMT degron-tagged mESC cells. Values were standardized across treatment conditions and are ordered based on the hierarchically clustered top heatmap. **Bottom:** Heatmaps representing Quantiles of Expression, Gene Length and Intron Proportion ordered according to the top heatmap (Quantile value 4=highest, 1=lowest; For Intron Proportion the additional quantile category '0' was added for genes lacking introns). Expression data is from Weinberg et al., 2019. Figure adapted from Jayakrishnan et al., 2024.

As a first step, clustering of murine genes associated with K36me2/3 peaks was performed using H3K36me2/3 ChIP data from mouse C3H10T HMT knockout (KO) cells (Weinberg et al. 2019) and the resulting clusters were correlated with H3K36me3 CUT&RUN data from mESC HMT-dTAG degron cells (Sun et al. 2023). In addition, features such as Gene Expression, Gene Length, and Intronic Proportion Quantiles were also visualized (Figure 36).

Two distinct patterns were observed in the HMT KO/dTAG-13-treated cells, with SETD2 depositing H3K36me3 independently of NSD1/2 at highly expressed, short, exon-enriched genes, while NSD1/2 catalyzes H3K36me2/3 at poorly expressed, long genes with many introns. Importantly, the TKO me2/3 patterns do not point at a simple additive effect of the individual knockouts, suggesting the involvement of other mammalian H3K36 HMTs (such as SETD5 and/or NSD3) (Sessa et al. 2019) or changes in the localization or activity of demethylases. Crucially, these patterns are consistently observed across two distinct perturbation strategies (long-term knockout vs Acute Inducible Degradation) and two mapping techniques (ChIP vs. CUT&RUN), which supports the robustness of this phenomenon. Based on these results, I surmise that the interplay between two K36me2/3 deposition pathways spearheaded by Set2/SETD2 and NSD/NSD1-3 has been maintained during evolution.

Consequences of altered H3K36me landscape upon HMT depletion

Until this point, the main aims revolved around discovering how the genomic K36me profiles were established by each HMT. In the subsequent chapters, the focus shifts to exploring the downstream effects that follow the depletion of HMTs. Specifically, I will focus broadly on three topics linked to K36 methylation as listed below:

1) Transcriptional (de-)regulation

2) Antagonism of Polycomb (K27me3) domains

3) Spatial and genomic (re-)distribution of chromatin readers

3.12 Are K36me marks ‘activating’?

Prior reports have documented that presence of K36me3 on genes bodies is generally correlated to active transcription (de Almeida et al. 2011; Kolasinska-Zwierz et al. 2009). Despite this, it’s unclear if K36me3 (and K36me1/2) causally promote transcription or if they are merely correlated to it. Several recent studies argue that K36 methylation is not directly involved in transcription regulation, but indirectly affects transcription by recruiting chromatin readers (which may be activating or repressive in nature) and/or by crosstalk with other histone modifications. I anticipate

that the transcriptional regulation by K36me will strongly depend on the nature and combination of K36me readers bound at a given gene locus.

Firstly, to verify the claim that K36me3 is correlated to active transcription, the relationship between mean K36me3/2/1 signal on gene bodies and corresponding RNA-seq Transcripts Per Million (TPM) values in unperturbed S2 cells were visualized (Figure 37).

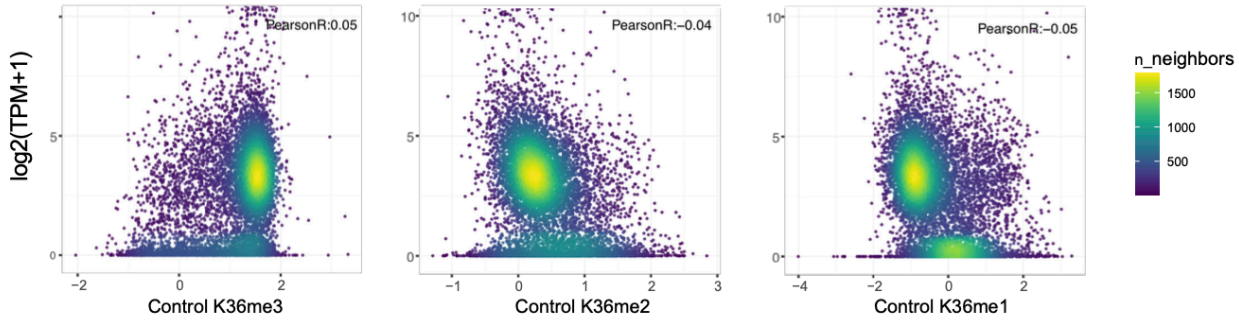


Figure 37: Scatter plots showing lack of correlation between S2 cell gene expression (log2 (RNAseq TPM)) and genic K36me3/2/1 ChIP signal. Overlayed color indicates density of points in the neighbourhood. Only genes shown in Figure 24A were used. Figure adapted from Jayakrishnan et al., 2024.

The high K36me3 density at expressed genes did not scale with TPM values (represented by the vertical cloud). K36me3 may thus be considered as a binary indicator of ON/OFF status of a given gene rather than a mark that promotes transcription. On the other hand, K36me1/2 was very low on expressed genes. So far, while studies have characterized the altered transcriptome upon loss of HMTs, they haven't systematically correlated these observations to changes in K36me, which is now being attempted. To explore this, I utilized published S2 cells RNA-seq datasets for Ash1 RNAi (Huang et al. 2017) and Set2, NSD RNAi (Depierre et al. 2023). Log2 fold-change values in gene expression were calculated for each HMT RNAi condition with respect to their respective controls and visualized as a boxplot for each supercluster (Figure 38A). To emphasize, these superclusters represent groups of genes exhibiting relatively consistent changes in K36me following HMT RNAi. If K36me exerts a uniform influence on gene transcription, HMT RNAi is expected to result in a unidirectional effect (either consistent up- or downregulation) on gene expression.

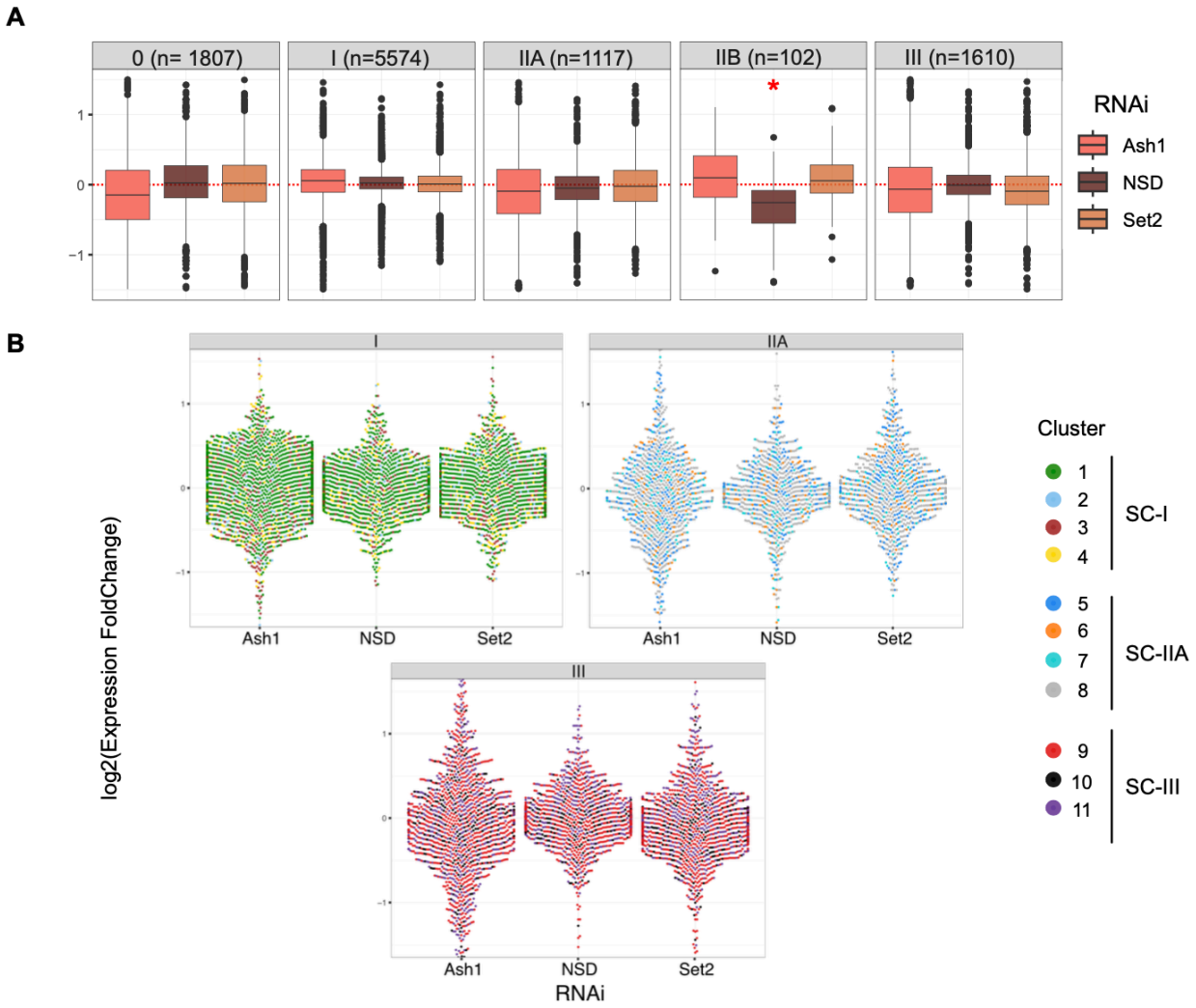


Figure 38: A. Boxplots representing RNA-seq transcriptional change (log2(RNAi/Control Expression)) upon indicated RNAi for SC 0-III genes. n denotes the number of genes within each class. Only one biological replicate per sample was available from source data. Two-sided one sample t-tests were performed for each 'box', i.e., RNAi - supercluster combination (null hypothesis: group expression change = 0), and significance adjusted for multiple testing. ' * ' represents the only group that passed the fold-change threshold. **B.** Bee swarm plots representing transcriptional change upon HMT RNAi for individual genes (dots) belonging to indicated superclusters. Color of dots indicate the gene cluster annotation (1-11) as also shown in Figure 24A. Figure adapted from Jayakrishnan et al., 2024.

Examination of the boxplots reveals that, although outliers are present, the majority of genes (represented by the interquartile range) within all gene clusters do not exhibit a consistent upward or downward trend in response to HMT RNAi. This observation suggests that changes in K36 methylation are weak predictors of gene expression changes. An exception is observed in SC-IIB (NSD-dependent heterochromatic) genes, where approximately 80% of the genes consistently show decreased expression following NSD RNAi. This aligns with previous research demonstrating that NSD positively regulates the expression of specific heterochromatic genes (Saha et al. 2020). Moreover, transcriptional defects caused by RNAi targeting a specific HMT (as indicated by outlier points representing differentially expressed genes) are not confined to the

supercluster, where the HMT primarily facilitates methylation, and can even impact genes lacking K36me (i.e., SC-0), indicating potential indirect effects. Each supercluster is composed of clusters which display subtle differences from one another (for example, the presence or absence of intermediate K36me1 between clusters 2 and 3), which may introduce heterogeneity in transcriptional response. Hence, cluster identity rather than the supercluster may correlate better with transcriptional alterations. To confirm this, I generated beeswarm plots for selected superclusters to visualize transcriptional changes (Figure 38B). These plots, which resemble violin plots but display individual data points (genes) overlaid with cluster-specific colors, reveal that the colored points are fairly evenly distributed across the beeswarm. This observation further supports the weak correlation between cluster identity and transcriptional change.

An alternative hypothesis is that the ratio of K36me2 to K36me3, rather than their absolute levels, may underlie transcriptional changes, as this ratio has been previously associated with transcriptional elongation. To evaluate this, the top 10% most upregulated and downregulated SC-I genes under Set2 RNAi conditions were selected along with a comparable number of unchanged genes as a control. These genes were then examined for the trends in me2:me3 ratio changes upon Set2 RNAi using line plots (Figure 39).

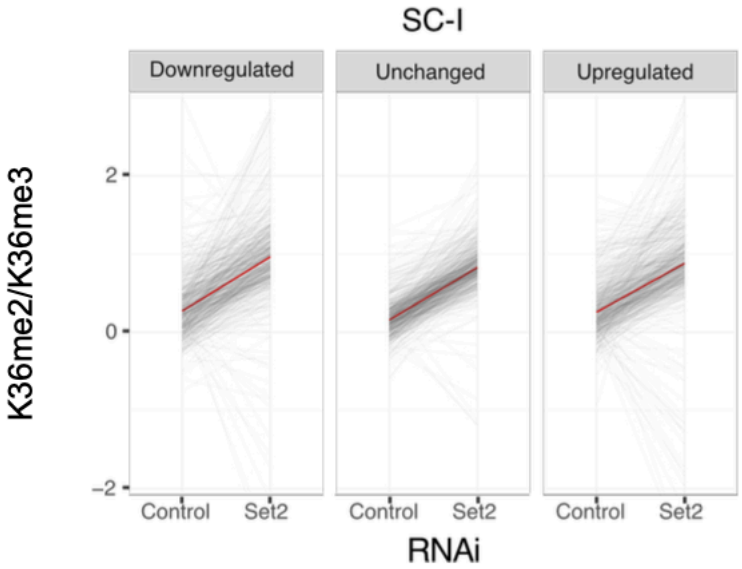


Figure 39: Lineplots demonstrating change in K36me2:K36me3 ratios for individual SC-I genes upon Set2 RNAi for 10% most downregulated (log2FC < -0.21; n=599), 10% most expression-unchanged (log2FC ~ 0; n=547) and 10% most upregulated (log2FC > 0.24; n=584) genes. Red line connects median ratios in Control and Set2 RNAi conditions. Figure adapted from Jayakrishnan et al., 2024.

This analysis reveals that the me2:me3 ratios increase markedly across all three gene sets, regardless of the transcriptional change observed, indicating that these ratios are also weak predictors of transcriptional changes. Similarly, analysis of NSD-dependent transcriptional changes relative to NSD-dependent me2:me3 ratios yields the same conclusion (data not shown). Taken together, these findings indicate that neither the absolute levels of K36me2/3 nor the

me2:me3 ratios account for the transcriptional changes observed upon HMT depletion. The RNAi approach often leads to incomplete depletion of K36me3 and the emergence of lower K36me states, which may compensate functionally for K36me3 (see section 3.14). As a result, the transcriptional impact of RNAi may be mitigated. To assess the reliability of prior observations using an independent dataset, I revisited the fly larval brain HMT knockout RNA-seq data from Lindehell et al., and performed an analogous analysis to that shown in Figure 38A. These knockout experiments are expected to completely abolish K36me deposition by the corresponding HMTs. Due to the substantial differences between the transcriptomes of S2 cells and fly brains, the analysis was restricted to housekeeping genes, which show highly correlated expression across both systems (Figure 40A) and are primarily enriched in supercluster I (Figure 40B). Given that housekeeping genes typically rely on conserved regulatory mechanisms across tissues, it is reasonable to assume that these genes are methylated at H3K36 by Set2 in the fly larval brain. The impact of Set2 knockout on the expression of approximately 3,500 housekeeping genes was then assessed (Figure 40C).

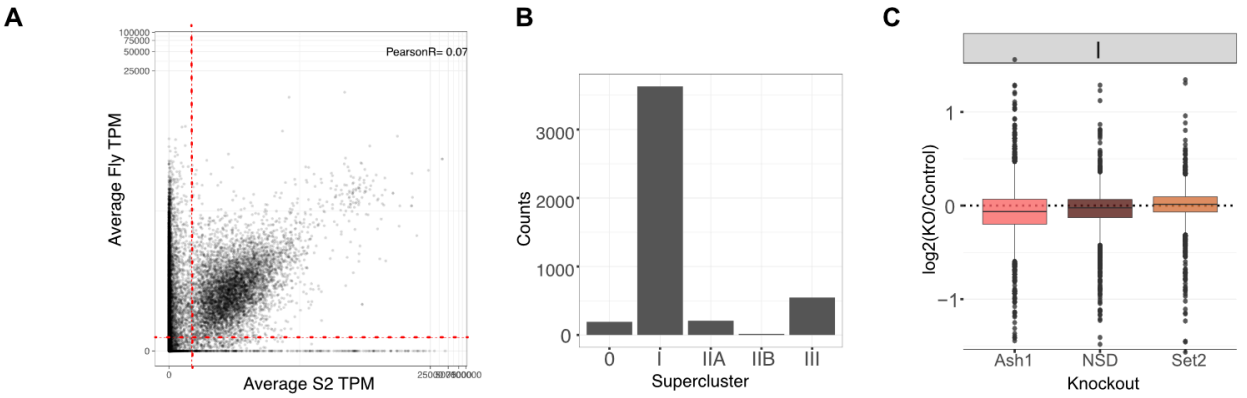


Figure 40: A. Correlation between OregonR fly and S2 gene expression (TPM) values. Red dotted line indicates the thresholds used to select for genes showing reasonably correlated expression in both systems (n=~3400). **B.** Classification of selected genes into SC 0-III. **C.** Boxplots representing RNA-seq transcriptional change (log2(KO/Control Expression)) upon knockout of indicated factor in fly brains for selected genes. Figure adapted from Jayakrishnan et al., 2024.

This analysis confirmed that the loss of K36me has minimal effects on transcription. Taken together, these results argue that the impact of K36me on transcription is not generalizable and the regulatory effect likely depends on the local chromatin context.

3.13 K36me as a 'barrier' restricting the spreading of facultative heterochromatin

One widely accepted hypothesis for the function of K36 methylation is its role in antagonizing the facultative heterochromatin mark K27me3. In *Drosophila*, earlier studies established that Ash1 is the sole methyltransferase whose loss results in classical homeotic transformations, indicative of excessive HOX gene repression by the Polycomb complex (Dorafshan et al. 2019). However,

more recent research has demonstrated that depletion of Set2 or NSD leads to increased K27me3 levels at specific genomic loci (Depierre et al. 2023). Notably, these studies did not directly link K27me3 changes to alterations in K36 methylation, which I sought to address through a focused dataset.

To investigate this, K27me3 MNase-ChIP profiles following HMT RNAi treatment were generated in S2 cells. Due to the availability of only a single replicate for each condition, I could not employ statistical tools like csaw to identify significantly altered ChIP regions. Instead, I used a z-score-based method (as illustrated in Fig. 21A) to isolate bins with the most significant increases or decreases, applying stringent filters. This analysis revealed that while most K27me3-marked regions remain stable, a notable subset of bins exhibited gains in K27me3 across all three HMT RNAi conditions (Figure 41A). Given that K27me3 reductions are likely indirect, I focused on bins showing strong increases. Importantly, these strongly increasing bins had no detectable signal in control conditions, suggesting that they are formed *de novo* upon HMT depletion (see representative profiles in Figure 41C). Further, these newly formed domains displayed substantially lower K27me3 levels than unchanged domains (data not shown), suggesting incomplete deposition of K27me3 by the Polycomb complex, potentially due to residual K36me (or ancillary repelling factors like ongoing transcription) following RNAi.

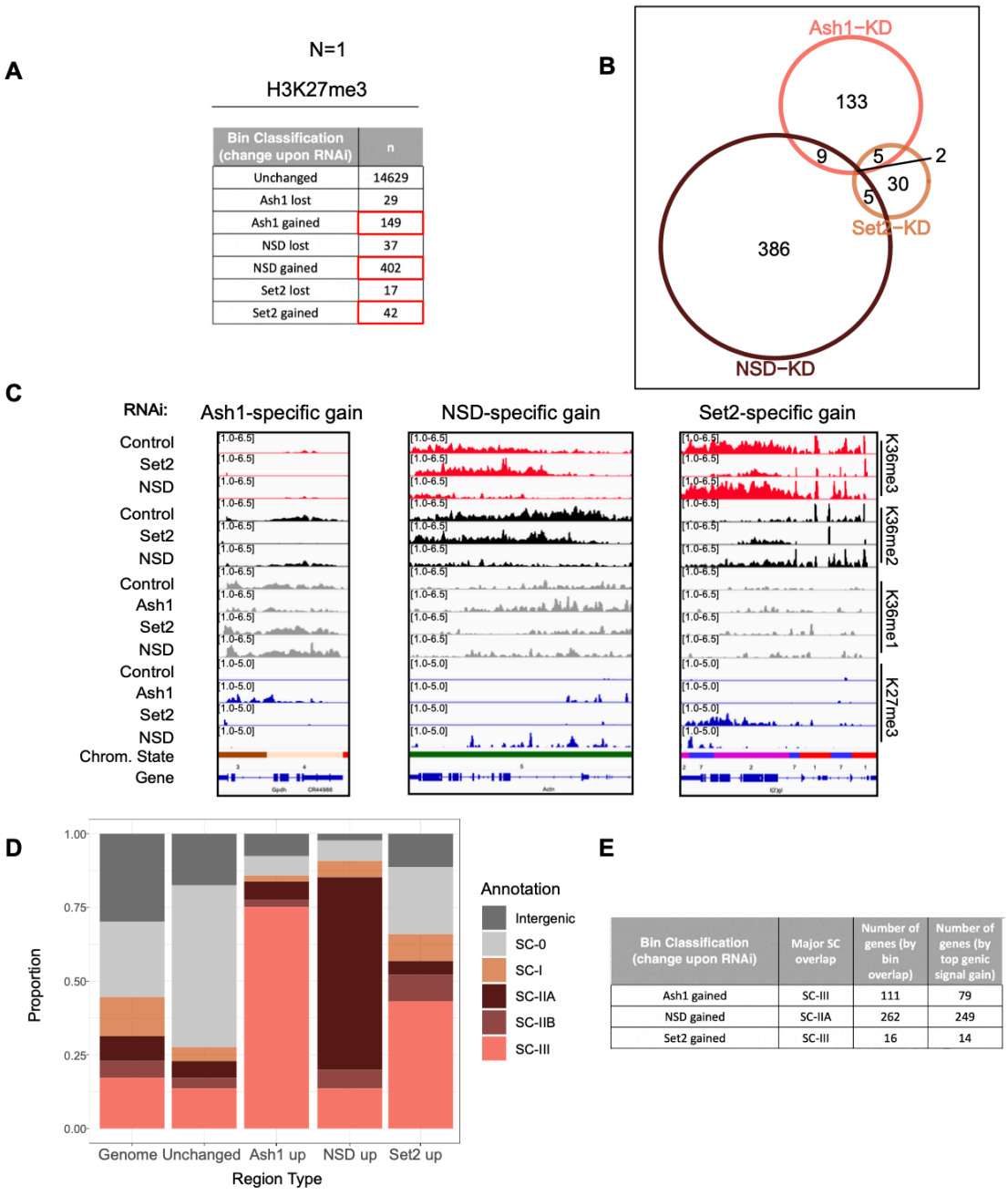


Figure 41: A. Table summarizing changes in H3K27me3 distribution in 5-kb bins in pilot (N=1) experiment. Bins were classified as 'Lost/Down', 'Gained/Up', or 'Unchanged' upon RNAi of indicated factor based on custom thresholds. n= Number of bins. **B.** Venn Diagrams representing overlap between K27me3 bins identified as 'Gained/Up' in indicated RNAi conditions **C.** Genome browser profiles for K36me1/2/3 and K27me3 for representative regions that form *de novo* K27me3 domains upon HMT RNAi. **D.** Stacked bar plots representing classification of bins where K27me3 was gained ('up') upon RNAi of indicated HMT (or unchanged in any condition 'Unchanged') as 'Intergenic' or belonging to SC 0-III based on overlap of genomic coordinates. The 'Genome' bar serves as reference indicating the proportion of genome overlapping each annotation. **E.** Table representing the major gene supercluster overlapped by the 'Gained' K27me3 bins upon different RNAi (as inferred from above panel D) as well as the number of genes falling into each category defined by two different approaches. Figure adapted from Jayakrishnan et al., 2024.

The bins with increased K27me3 across the three RNAi conditions revealed limited overlap, indicating that the three methyltransferases antagonize Polycomb deposition at distinct genomic sites (Figure 41B). Representative genome browser tracks (Figure 41C) illustrate an Ash1 RNAi-specific increase in K27me3 in a region also containing Set2-dependent K36me2/3. Despite substantial depletion of K36me2/3 following Set2 RNAi, K27me3 levels remained unaffected in this region, suggesting that Ash1-dependent K36me1 is essential for counteracting K27me3 deposition in this context.

To better characterize these regions, I categorized bins based on their genomic context—intergenic or part of superclusters (SC 0–III)—and visualized their distribution in bar plots (Figure 41D). This analysis confirmed that bins gaining K27me3 after Ash1 or NSD RNAi were predominantly located within genes in SC-III and SC-IIA, respectively. Notably, NSD RNAi did not lead to K27me3 encroachment into NSD-dependent heterochromatic genes (SC-IIB) or intergenic regions, indicating that NSD-dependent K36me2/3 is not essential for maintaining the separation between facultative and constitutive heterochromatin. Interestingly, bins gaining K27me3 in a Set2-dependent manner were enriched in SC-III genes distinct from those affected by Ash1 RNAi, suggesting a unique subset of SC-III genes relying on Set2-dependent K36me3 (such as cluster 10). Interestingly, despite substantial NSD-dependent K36me2 at these genes, NSD RNAi induces new K27me3 domains on only a very small number of SC-III genes. The number of affected genes is summarized in Figure (Figure 41E). Importantly, only a subset of supercluster genes showed K27me3 gains, implying that Polycomb repression is not the default outcome of K36me loss.

The absence of K27me3 gain in SC-I genes is intriguing and may result from the presence of intermediate K36me2 which can inhibit PRC2 enzymatic activity, however even genes that show near-complete loss of K36me2/3 upon Set2 RNAi fail to gain K27me3 (data not shown). Alternatively, other factors such as nascent RNA may also repel K27me3 at these genes even in the absence of K36me2/3. Alternatively, these housekeeping genes may lack factors required to recruit Polycomb components, which typically assemble into ‘Polycomb bodies’, nuclear foci comprised of facultative heterochromatin (Pirrotta and Li 2012).

As a secondary measure of K27me3 gain, I assessed transcriptional repression, given the repressive role of this mark. Using previously described RNA-seq datasets, transcriptional changes in genes gaining K27me3 under a given RNAi condition were compared to expression changes of other genes within the same supercluster (Figure 42).

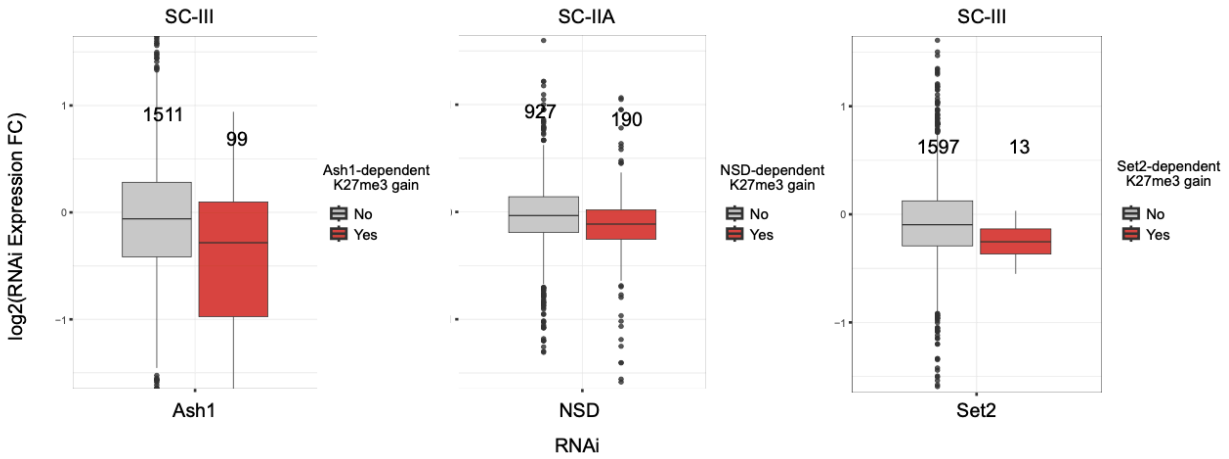


Figure 42: Expression change boxplots for genes belonging to corresponding majority supercluster (.i.e., as inferred from Figure 41 D, E) that gain K27me3 (red box) upon RNAi of indicated HMT. Expression changes for genes that belong to the same supercluster but do not gain K27me3 (gray box) serve as reference. Numbers displayed above the boxes correspond to number of filtered genes with valid (non-zero reads) RNA-seq log2 Fold Change values. Figure adapted from Jayakrishnan et al., 2024.

Genes gaining K27me3 showed a modest tendency toward downregulation compared to those without K27me3 gain. Collectively, these results indicate that all three methyltransferases independently counteract K27me3 deposition, primarily at subsets of developmental genes. While these findings align with a mechanism involving direct inhibition of PRC2 by H3K36 methylation, the possibility that these HMTs counteract PRC2 through alternative (non-histone) substrates cannot be excluded and warrants further investigation.

3.14 Depletion of H3K36 methyltransferases results in redistribution and reorganization of H3K36me3 reader complexes

3.14.1 K36me2 is partly redundant with K36me3 for reader binding

Depleting HMTs results in altered genic K36 methylation patterns with different combinations of residual K36me1/2/3, offering a unique opportunity to examine the dependence of reader protein interactions with chromatin on these marks. To investigate this, I analyzed the redistribution of exemplary K36me3 reader proteins, JASPer, a PWWP domain protein that binds genome-wide and mediates the localization and stability of the H3S10 kinase JIL-1 (Albig et al. 2019) and the chromodomain protein MSL3, which binds specifically on the X chromosome and mediates dosage compensation (Sural et al. 2008). I focused on supercluster I and II genes, which account for the majority of reader binding events. Before engaging in ChIP experiments, the overall levels of JASPer and MSL3 (inferred from other MSL complex proteins) were verified to be unchanged upon the RNAi treatments, confirming that the complexes need not remain chromatin-bound for stability (Figure 43A).

Firstly, I generated heatmaps of JASPer binding after HMT depletion and visualized it along with the clustered heatmap of K36me1/2/3 (Figure 43B). While JASPer binding decreased in clusters 2 and 3 following reductions in K36me3, the decline was notably less pronounced in clusters 1 and 4, where K36me2 levels strongly increased upon Set2 depletion. Conversely, NSD depletion caused a concomitant loss of both K36me2 and K36me3, leading to a significant reduction in JASPer binding, most prominently in clusters 6, 7, and 8. Genomic profiles of representative genes in clusters 1, 3, and 6 illustrate these scenarios (Figure 43C). These observations suggest that JASPer may recognize both K36me2 and K36me3.

To further investigate this possibility, I calculated proportional changes to quantify changes in reader binding and K36me2/3 signals across clusters 1, 3, and 6 (Figure 43D). Cluster 1 genes showed only a mild reduction in JASPer binding (~20% median decrease) despite a ~50% median loss in K36me3. This ‘resistance’ of binding could potentially be due to gains in K36me2 that compensate for loss of K36me3. In cluster 3, however, a stronger reduction in K36me3 (~70%) led to a ~55% decrease in JASPer binding, with only limited compensation by K36me2. At cluster 6 genes, NSD knockdown resulted in coordinated loss of both K36me2 and K36me3, leading to proportional reductions in JASPer binding (~60% for both; Figure 6C, right panel). Interestingly, JASPer binding at cluster 6 genes increased when Set2 was depleted, likely due to increased K36me2. These findings collectively suggest that JASPer binds to both K36me2 and K36me3.

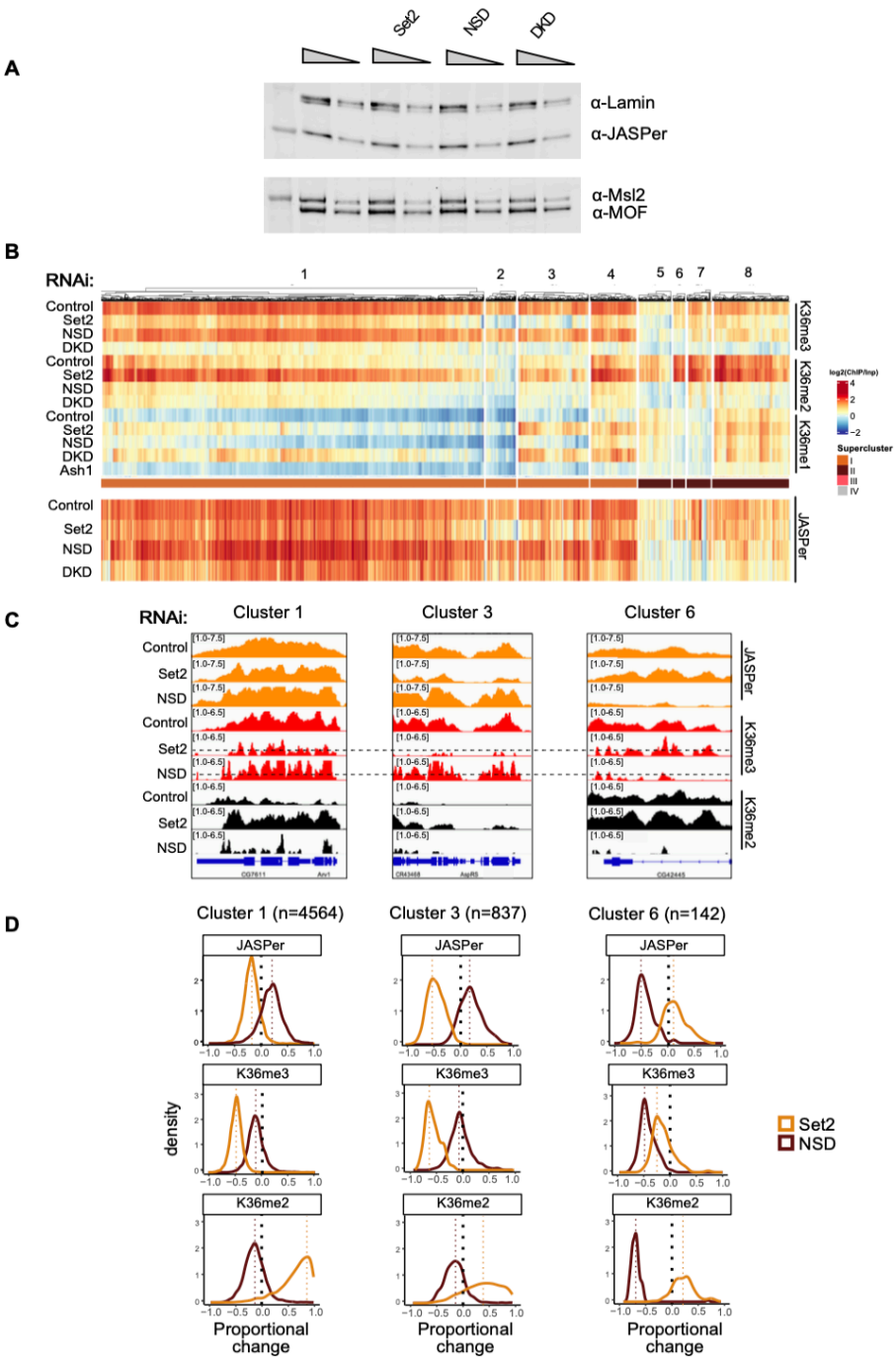


Figure 43: A. Western blot analysis for JASPer, Lamin, MSL2 and MOF in whole cell extracts of RNAi-treated cells. Two dilutions (1x and 0.5x) were loaded per sample. **B.** Heatmap of gene body average JASPer ChIP signal upon indicated RNAi condition, ordered according to the heatmap from Figure 24A. SC-III was excluded as it has relatively few genes bound by JASPer. **C.** Genome browser profiles for three distinct genomic loci highlighting differential response of JASPer to Set2 RNAi and NSD RNAi within indicated clusters. The dotted line in K36me3 ChIP in Set2/NSD RNAi represents an arbitrary threshold distinguishing 'resistant' and 'sensitive' clusters. **D**. Proportional change (RNAi-Control)/Control) density plots representing direction and magnitude of change in gene body average ChIP signal for K36me2/3 along with reader JASPer for indicated representative clusters. Figure adapted from Jayakrishnan et al., 2024.

To quantitate this phenomenon at a gene level, changes in JASPer binding as a function of K36me3 change in response to Set2 depletion were visualized as a scatterplot across supercluster I genes (Figure 44A). Genes that exhibit a proportional reduction in JASPer binding relative to the loss of K36me3 following Set2 RNAi are expected to cluster near the y=x line in the scatter plot. In contrast, genes where K36me2 compensates for the loss of K36me3 display an offset to the left of this line. For SC-I genes, the trendline deviates significantly from the y=x line, indicating that many genes experience a milder reduction in JASPer binding despite substantial decreases in K36me3 levels. Notably, numerous genes either maintain stable JASPer binding or even exhibit slight increases in binding (proportional changes ≥ 0) despite experiencing up to a 40% decrease in K36me3 levels, as marked by the red dotted line.

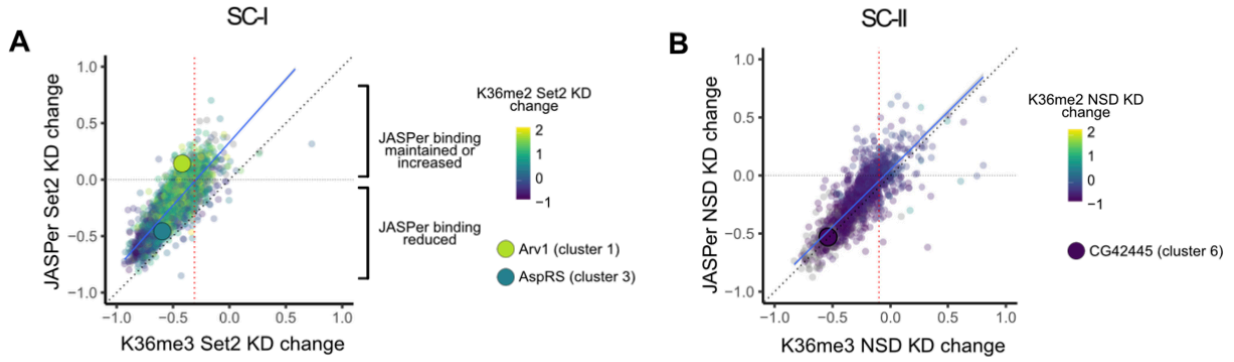


Figure 44: A. Scatter plot comparing the changes of JASPer binding (y-axis) with the changes of K36me3 (x-axis) upon depletion Set2 for SC-I genes. Each point represents a gene. The color of the points indicates the accompanying change of K36me2. The solid blue line denotes the population trendline. The red dashed line denotes the offset value of the trendline from the y=x line (black dashed diagonal line) and represents the magnitude of compensation by K36me2. Representative genes for clusters 1 and 3 as shown in Figure 43C are highlighted. **B.** Same as panel A but for assessment of NSD RNAi effect at SC-II genes with a representative gene from cluster 6 highlighted. Figure adapted from Jayakrishnan et al., 2024

This offset trend correlates strongly with increases in K36me2 levels, as illustrated by the color coding of individual data points. Genes that exhibit higher offsets generally show robust gains in K36me2, in support of a compensatory mechanism. Conversely, genes that fail to retain JASPer binding tend to show minimal or no increases in K36me2. Highlighted in this scatter plot are previously described representative genes from clusters 1 and 3, reinforcing the distinct dynamics of JASPer binding within these clusters under Set2 depletion conditions.

This functional compensation by K36me2 upon loss of K36me3 observed for SC-I genes appears markedly reduced for SC-IIA genes upon depletion of NSD, which typically results in a simultaneous reduction of both K36me2 and K36me3. This prediction is confirmed by the visualization of JASPer binding changes in response to K36me3 reductions under NSD RNAi (Figure 44B). In SC-IIA genes, JASPer binding is lost in a proportional manner with K36me3 depletion, indicating the absence of K36me2-mediated compensation. This coordinated loss of

both marks further emphasizes the differential regulation and dependency of JASPer binding on K36 methylation states across different superclusters.

Our correlative analyses support the notion of redundancy between K36me2 and K36me3 in retaining JASPer chromatin binding. Motivated by these observations, the intrinsic biochemical affinity of the recombinant reader protein for modified (K36me0/2/3) nucleosomes was measured using Microscale Thermophoresis (Figure 45). These experiments, performed by my collaborators Magdalena Havlová and Václav Veverka, show that JASPer exhibited a slight preference for K36me2 (KD = 0.32 ± 0.02 μM) over K36me3 (KD = 0.60 ± 0.04 μM), while its affinity for unmodified nucleosomes was substantially lower (KD = 4.00 ± 1.54 μM), confirming K36me2's compensatory role.

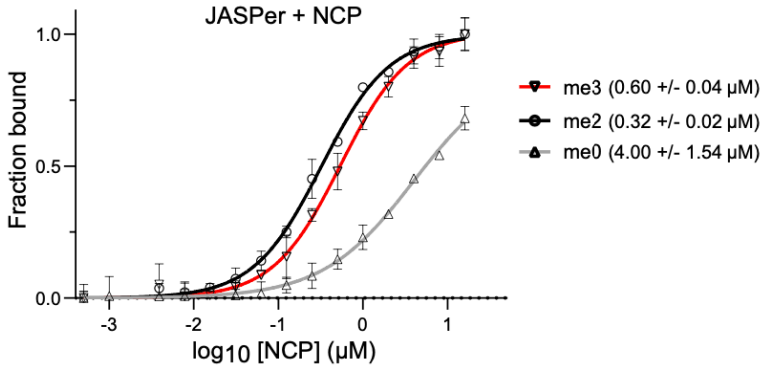


Figure 45: Equilibrium binding of recombinant JASPer to unmodified or -me2/3 nucleosomes, determined using microscale thermophoresis (MST). Error bars represent the standard deviation from the mean values obtained from $n = 2$ experiments. Calculated dissociation constants are indicated within brackets. Figure adapted from Jayakrishnan et al., 2024.

I also examined whether these binding principles applied to the chromodomain protein MSL3. Towards this end, the ChIP analyses were restricted to X-chromosomal genes (Figure 46).

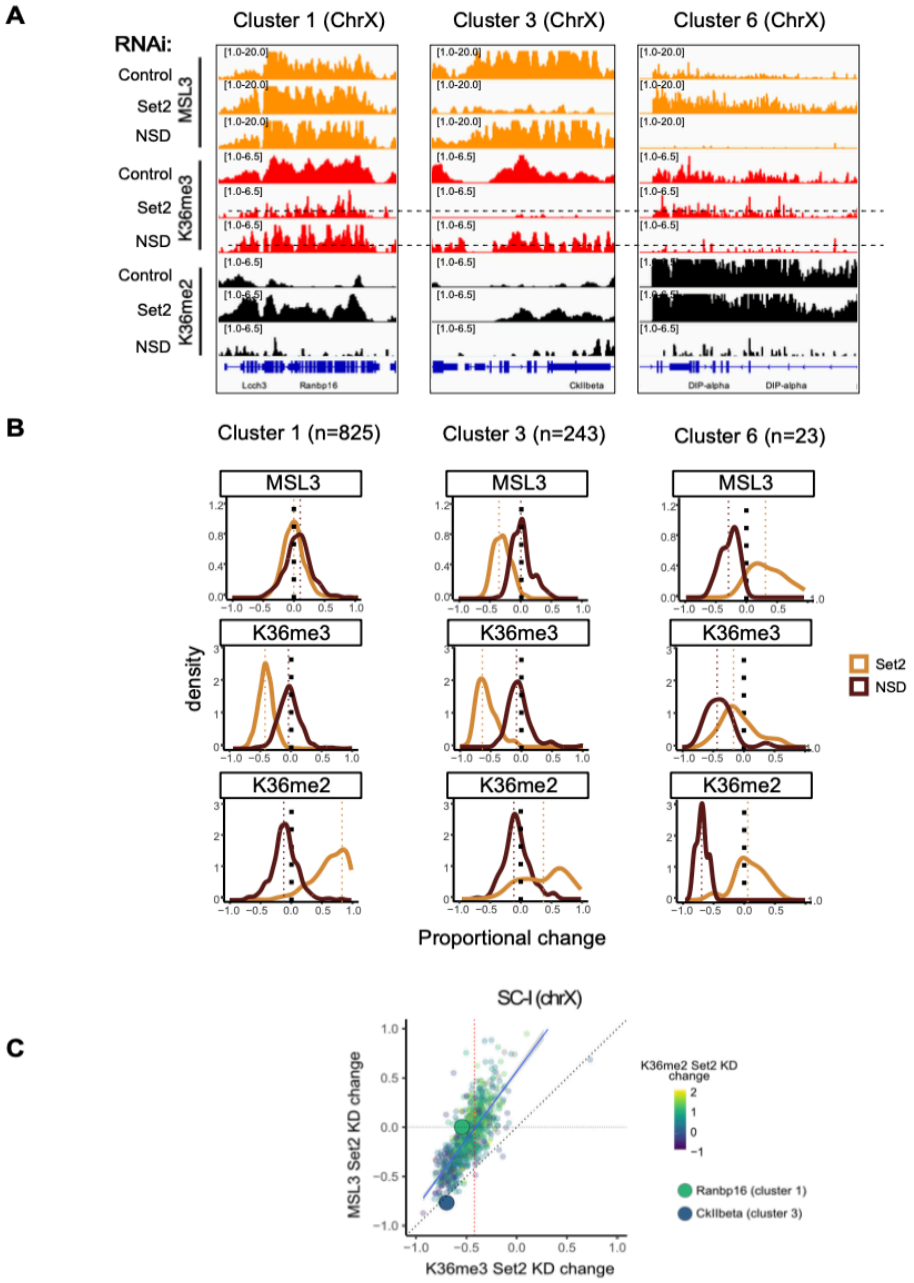


Figure 46: A. Genome browser profiles for three distinct genomic loci highlighting differential response of MSL3 to depletion of Set2 and NSD within indicated clusters. The dotted line in K36me3 ChIP represents an arbitrary threshold distinguishing 'resistant' and 'sensitive' clusters. **B.** Proportional change ((RNAi-Control)/Control) density plots representing direction and magnitude of change in gene body-average ChIP signal for K36me2/3 along with reader MSL3 for indicated representative clusters. **C.** Scatter plot representing the change of MSL3 (y-axis) as a function of K36me3 change (x-axis) upon depletion of Set2 for SC-I X-chromosomal genes. Each point represents a gene. The color of the points represents the values of K36me2 under those conditions. The solid blue line denotes the population trendline. The red dashed line denotes the offset value of the trendline from the y=x line (black dashed diagonal line) and represents the magnitude of compensation by K36me2. Representative genes for clusters 1 and 3 as shown in panel A are highlighted. Figure adapted from Jayakrishnan et al., 2024.

The density plots and scatterplots for MSL3 exhibited an even higher degree of compensation by K36me2 than for JASPer. For instance, MSL3 binding at cluster 1 genes remained stable despite around 50% reduction of K36me3, while clusters 3 and 6 showed significant but dampened reductions in MSL3 binding upon Set2 or NSD knockdowns, respectively. These patterns mirrored those of JASPer, suggesting that MSL3 also binds both K36me2 and K36me3. Altogether, these results indicate that both JASPer and MSL3 dissociate from chromatin when the combined levels of K36me2 and K36me3 fall below a critical threshold.

The persistence of relatively high combined levels of K36me2/3 at certain genes, particularly those in clusters 1 and 4, despite substantial Set2 depletion raised an important question: which factors determine this 'resistance'? By contrast, genes in other clusters exhibit a near-complete loss of K36me2/3 methylation under the same conditions. We consider that the retention of residual K36me2/3 at specific loci is heavily influenced by genomic context, including transcription-associated nucleosome turnover rates and the dynamics of methylation versus demethylation.

Supporting this hypothesis, the analyses revealed that elongating RNA polymerase II (ePol) is significantly more abundant at cluster 2 and 3 genes, which display much lower combined levels of K36me2/3 ('sensitive genes'), compared to genes in clusters 1 and 4 ('resistant genes'; Figure 47). This suggests that higher transcriptional activity at sensitive genes may accelerate nucleosome turnover, thereby facilitating the loss of K36me2/3-bearing nucleosomes. In addition, the putative K36-specific demethylase KDM2, but not KDM4A, was observed to be enriched at sensitive cluster 2 and 3 genes. The preferential association of KDM2 with these loci could further exacerbate the loss of K36 methylation by actively removing K36me2/3 marks, reinforcing their sensitivity to Set2 depletion.

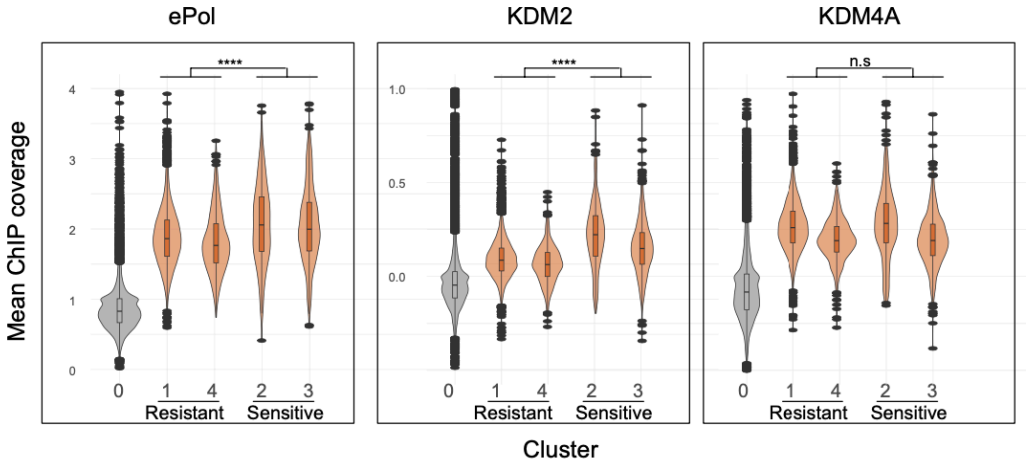


Figure 47: Violin plots representing gene body-average ChIP signal for HMT (ePol) and HDMs (KDM2 and KDM4A) for indicated SC-I clusters showing different K36me2/3 turnover kinetics (resistant and sensitive) along with 'negative' control Cluster 0 genes as a baseline signal. One-sided t-test was performed to identify factors selectively enriched in the 'Sensitive' clusters relative to the 'Resistant' clusters (**** denotes a significance value of $p < 0.001$, n.s=Not Significant). Figure adapted from Jayakrishnan et al., 2024.

These findings point to a nuanced regulatory landscape where the resilience of K36me reader binding depends on a complex interplay between methyltransferase activity, transcription-driven turnover of nucleosomes, and demethylase-mediated removal of K36me2/3 marks. The distinct behavior of resistant and sensitive genes underscores the critical role of genomic context in shaping the stability of K36me patterns under conditions of HMT perturbation.

3.14.2 Loss of MSL3-mediated 'epigenetic' targeting of MOF from the X chromosome results in partial relocalization of H4K16ac to autosomal NSL-bound sites

In the previous experiments, I had established that perturbation of K36me2/3 by RNAi of HMTs led to substantial loss of MSL3 binding from the X chromosome. This confirms that MSL3 is targeted to chromatin largely through recognition of K36me2/3 via its chromodomain, as previously suggested. MSL3, recognizes active chromatin and concentrates the DCC onto gene bodies to enable MOF to deposit H4K16ac. Accordingly, I explored whether K16ac is affected upon HMT depletions.

Before examining K16ac, I first wanted to verify that the 'genetic' targeting aspect of the DCC, mediated by MSL2, was unaffected upon HMT RNAi. Previous studies from the lab have shown that standard ChIP protocols do not distinguish the binding of different DCC subunits and are less suited for examining the contributions of MSL2 to chromatin interactions (Straub et al. 2013). To profile the direct occupancy of MSL2, I used a modified high-shear ChIP-seq protocol, that disrupts the integrity of the DCC, allowing to monitor the crosslinking of each individual DCC subunit. This revealed that MSL2's 'spread-out' localization on gene bodies is lost upon shearing and is instead restricted exclusively to the accessible High Affinity Sites containing the MRE (see methods, Figure 48A). I only examined control conditions (GFP RNAi) and combined Set2/NSD RNAi to assess if MSL2 binding is affected, reasoning that if no changes were observed with combined RNAi, then it can be safely assumed that single depletions of Set2 and NSD also do not affect MSL2. Examining the signal of MSL2 around ~500 MSL2 peaks, shows that MSL2 largely maintains robust binding, consistent with its known ability to directly bind DNA (Figure 48B). While a slight reduction in MSL3 can be observed in the regions flanking the peaks, MSL3 behaves identical to MSL2 at the peak centers, suggesting that MSL3 occupancy at these peaks occurs indirectly via MSL2, independent of its nucleosome binding properties,

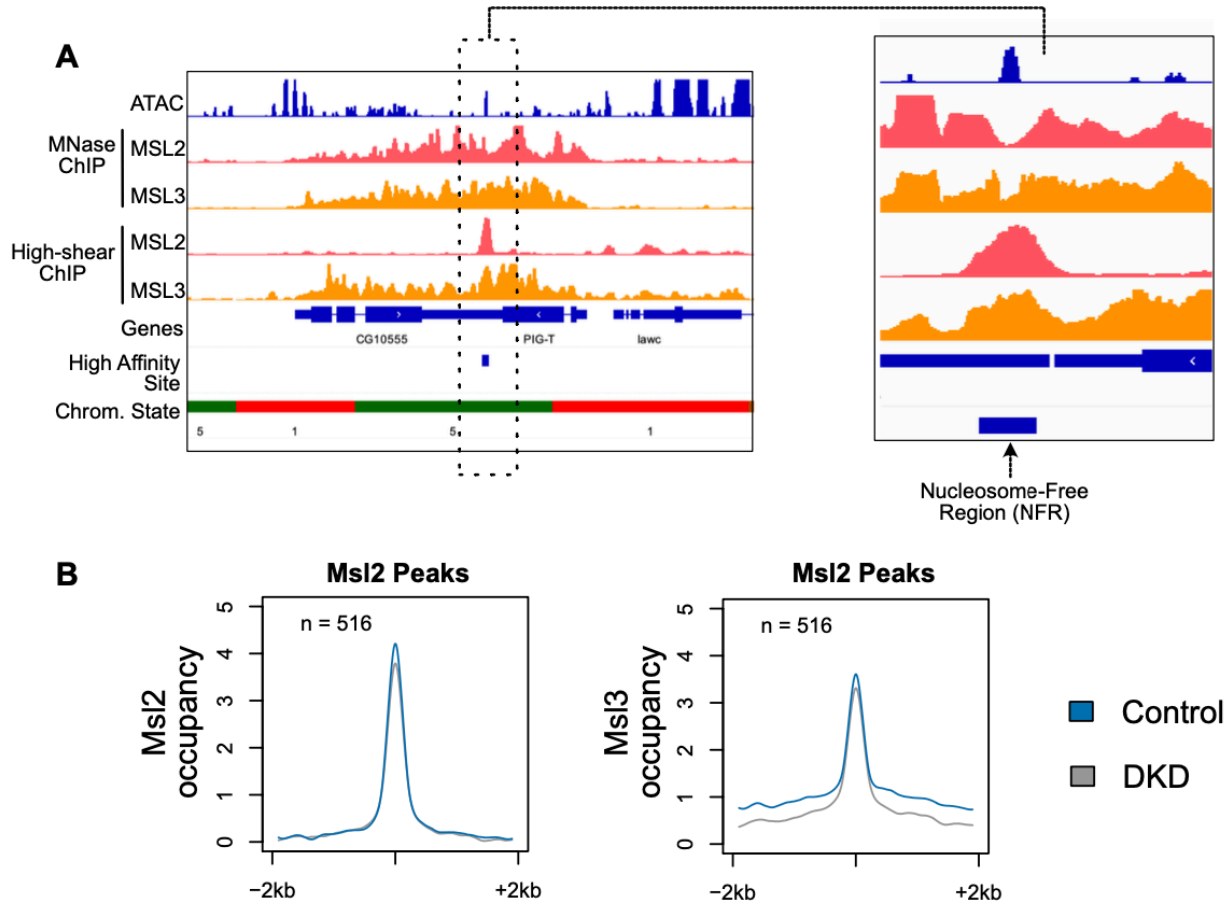


Figure 48: A. Genome browser views for MSL2 and MSL3 profiled using either MNase ChIP or high-shear ChIP. The zoomed view provides a clearer view of the NFR at the ATAC-seq accessible site along with neighbouring nucleosomes in MNase-ChIP data. High Affinity Site and chromatin state annotations are provided for reference. **B.** Cumulative plots representing averaged MSL2/3 high-shear ChIP signal for 4kb regions centered around MSL2 peaks after indicated RNAi treatment.

These results confirm that only the nucleosome binding, but not the DNA binding of the DCC, are affected upon HMT depletion. Therefore, any changes observed for K16ac can be attributed mainly to altered ‘epigenetic’ targeting via K36me3/2 of the complex.

MNase-ChIP for K16ac upon HMT depletion confirms that K16ac is reduced at subsets of X-chromosomal gene bodies, although the reduction is relatively modest compared to the loss of MSL3 (Figure 49A). This may be due to residual MSL3/DCC bound at these sites, which may be sufficient to maintain acetylation of multiple nucleosomes in the vicinity. Further, K16ac is known as a relatively stable chromatin mark due to lack of any known deacetylases.

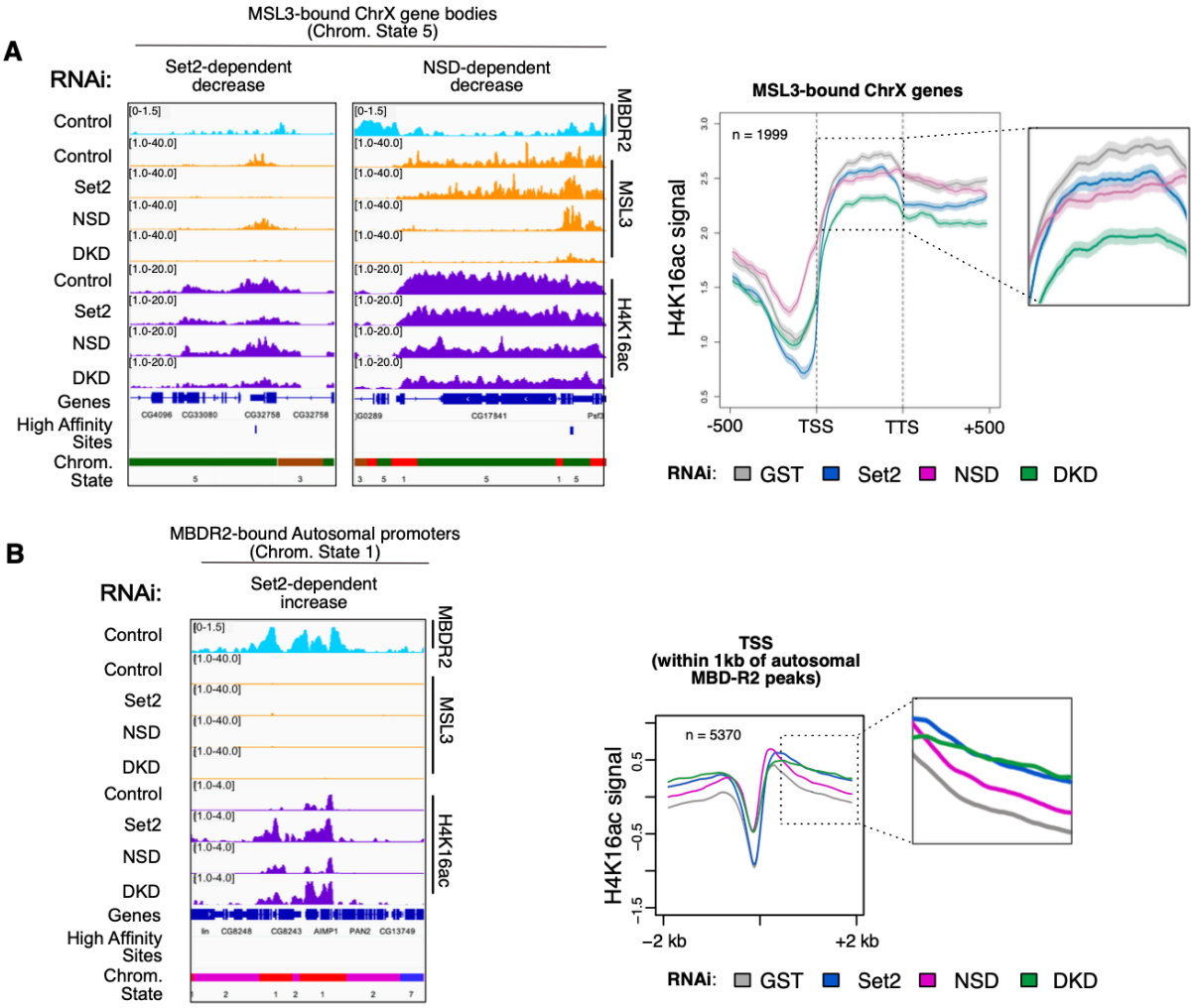


Figure 49: A. Genome browser views of X-chromosomal regions demonstrating MSL3/H4K16ac decreases in response to either Set2 or NSD RNAi (left). MBD-R2 profile serves as a reference for NSL-bound regions. Cumulative plots representing reduction in gene body-averaged H4K16ac signal in response to indicated HMT RNAi for all X-chromosomal MSL3-bound genes (right; n=1999). **B.** Genome browser views of autosomal regions demonstrating H4K16ac increase in response to Set2 (left). MBD-R2 profile serves as a reference for NSL-bound regions. Cumulative plots representing average H4K16ac signal reduction within 4 kb region centered around NSL-proximal TSS (n=5370) upon indicated HMT RNAi (right).

Of note, this cumulative plot likely underestimates the reduction as the effect is averaged over all MSL3 bound genes, while HMT RNAi affects only a defined subset. Regardless, this observation argues for impaired DCC function specifically due to perturbation of MSL3 targeting. The gains of H4K16ac on autosomes can be rationalized as follows. Because MOF is no longer sequestered on the X chromosome, a larger pool of MOF is now available for assembly of the other MOF-containing complex NSL, which is targeted to promoters marked by specific Ohler and DRE DNA sequence motifs (Feller et al. 2012). It's been previously shown that the two complexes compete for MOF, as ectopic overexpression of MSL2 in females (which leads to *de novo* assembly of the DCC) relocates MOF from autosomes to the X-chromosome (Prestel et al. 2010). To assess an

increased MOF activity in the context of the NSL complex, I examined H4K16ac signal at promoters of autosomal genes proximal to MBDR2 binding sites, a subunit of the NSL complex. This revealed a slight but noticeable gain upon RNAi of HMTs (Figure 49B). Of note, the H4K16ac signal on autosomes is still far lower (approximately 5-fold) on autosomes compared to the X-chromosome, as indicated by the value of the genome browser scale. Lastly, promoters distal to these MBDR2 sites, which carry very low levels of K16ac, were largely unaffected (data not shown). These observations add support to the long-standing hypothesis that MOF exists in a dynamic equilibrium between the genome-wide NSL and X-chromosomal MSL complexes and that disruption of targeting principles leads to redistribution of their catalytic subunit, MOF (Figure 50).

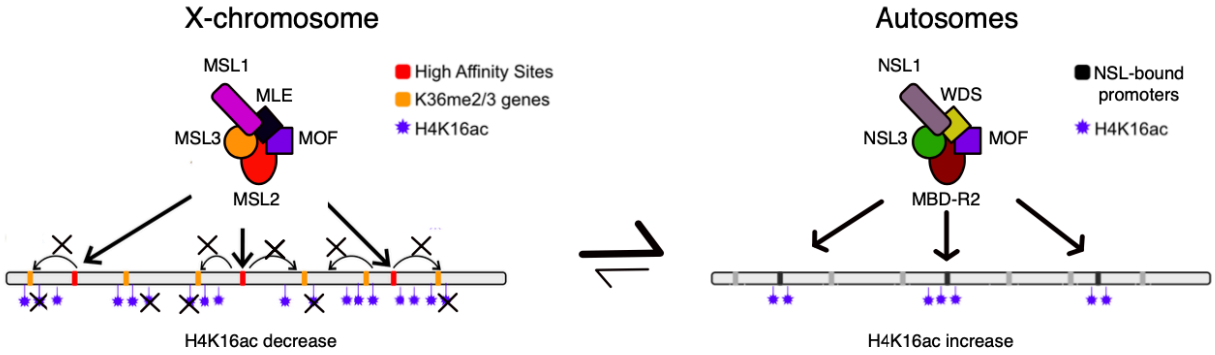


Figure 50: Graphical summary representing shift in H4K16ac balance towards NSL-bound regions upon H3K36 HMT depletion.

This dynamic relocation of epigenetic effector complexes illustrates the intricate interplay between histone marks and their readers, where perturbations can result in complex and context-dependent outcomes.

3.14.3 Spatial reorganization of MSL complex in response to HMT RNAi

As previously described, an X chromosome forms a distinct chromosomal territory. It's dosage-compensated nature is evident from a coherent IF staining of DCC subunits as well as H4K16ac. Alterations in DCC composition or levels, or loss of ancillary recruitment factors results disrupt the coherent MSL staining of the chromosome territory. While the prior ChIP experiments confirm that Set2 and NSD contribute to MSL3 targeting and/or DCC activity, it's unclear if the spatial organization of the territory is affected. To address this, I performed IF for selected DCC components upon H3K36 HMT depletion (Figure 51).

RNAi:

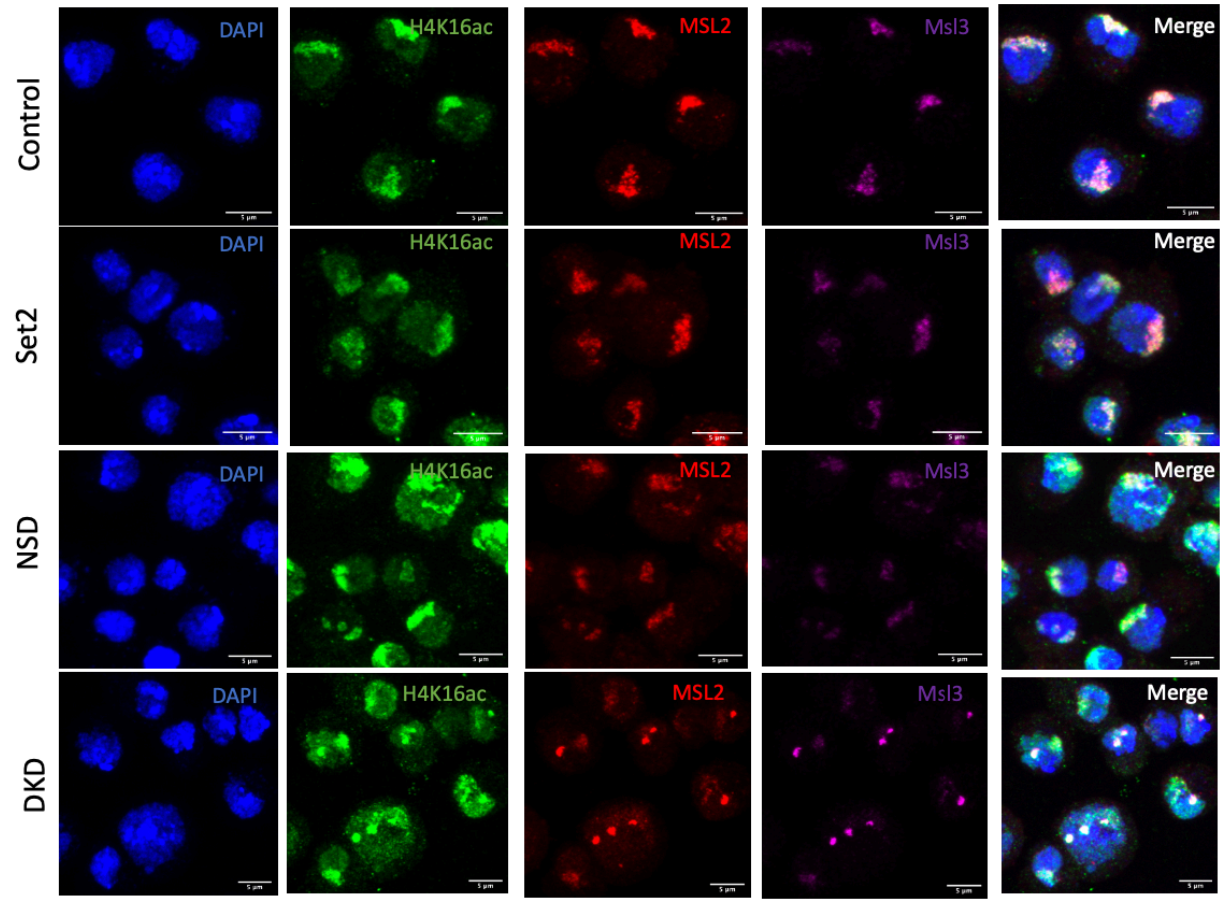


Figure 51: Representative IF confocal images of DNA (DAPI), MSL2, MSL3 and H4K16ac visualize their redistribution upon HMT depletion. Scale bar denotes 5 μm.

Upon individual depletion of either Set2 or NSD no dramatic changes in territory organization can be observed, although a slight qualitative decrease in signal intensity is consistent with the ChIP results. However, upon their combined depletion (DKD), a dramatic reorganization of the MSL-stained territory into spherical 'blobs' can be observed. This phenotype is quite penetrant with ~50% of the cells displaying it, while the rest of the cells almost completely lose territory staining. To rule out experimental artefacts, these results were replicated over a wide range of conditions (coverslip surface coats, fixation and permeabilization conditions, two different S2 cell clones) and proved reproducible (data not shown). This phenotype has not been previously reported in the context of DCC, as perturbation of DCC targeting results either in uniform dilution of the DCC into the nucleoplasm or in formation of multiple speckles randomly distributed in the nucleus (Villa et al. 2021; Jagtap et al. 2023). DKD cells show a mild cell cycle defect observed during culturing, necessitating a control to verify if the phenotype is dependent on abnormal proliferation. To this end, I also performed RNAi treatment for Prset7, a histone methyltransferase which predominantly deposits H4K20me1 and is essential for proper cell cycle progression (Li et al. 2016) (Figure 52). Of note, H4K20me1 was previously implicated in targeting of MSL3 based on structural and cellular studies (Kapoor-Vazirani and Vertino 2014; Kim et al. 2010).

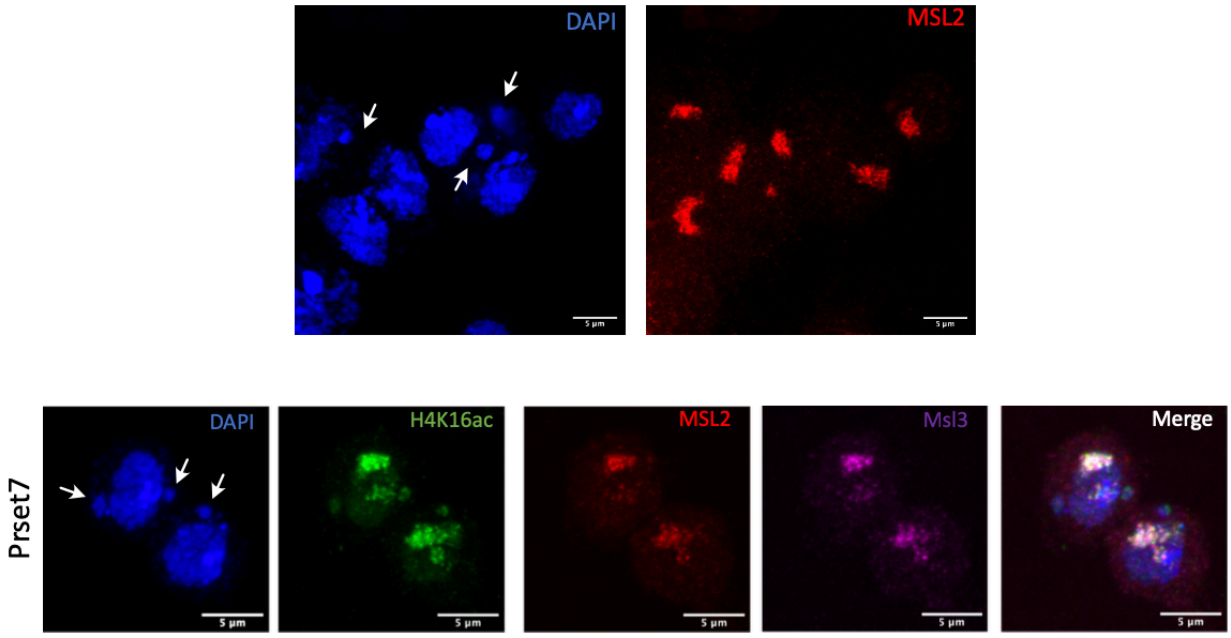


Figure 52: Representative IF confocal images of DNA (DAPI), MSL2, MSL3 and H4K16ac spatial redistribution upon Prset7 RNAi. Scale bar denotes 5 μm. White arrows highlight hallmarks of proliferation defects such as lagging chromosomes and micronuclei. Note that a separate control (GST) wasn't provided as this experiment was performed alongside the experiment shown in Figure 51.

Prset7 RNAi leads to strong cell cycle defects, evidenced by very poor proliferation upon culturing with visible micronuclei in DAPI staining (indicated by white arrows). Despite these effects, robust territories were observed, suggesting that the DKD phenotype was likely not due to cell cycle defects.

Spherical structures within the cell have been frequently been attributed to liquid-liquid phase-separated (LLPS) droplets due to the intrinsic nature of liquids to minimize surface area/tension (Alberti, Gladfelter, and Mittag 2019). To assess the physical properties of MSL blobs, I performed a 1,6-hexanediol (HD) treatment on Control and DKD cells (Figure 53). Successful 1,6-HD treatment was verified by reduction in HP1α staining, which has been reported previously to undergo LLPS (Strom et al. 2017).

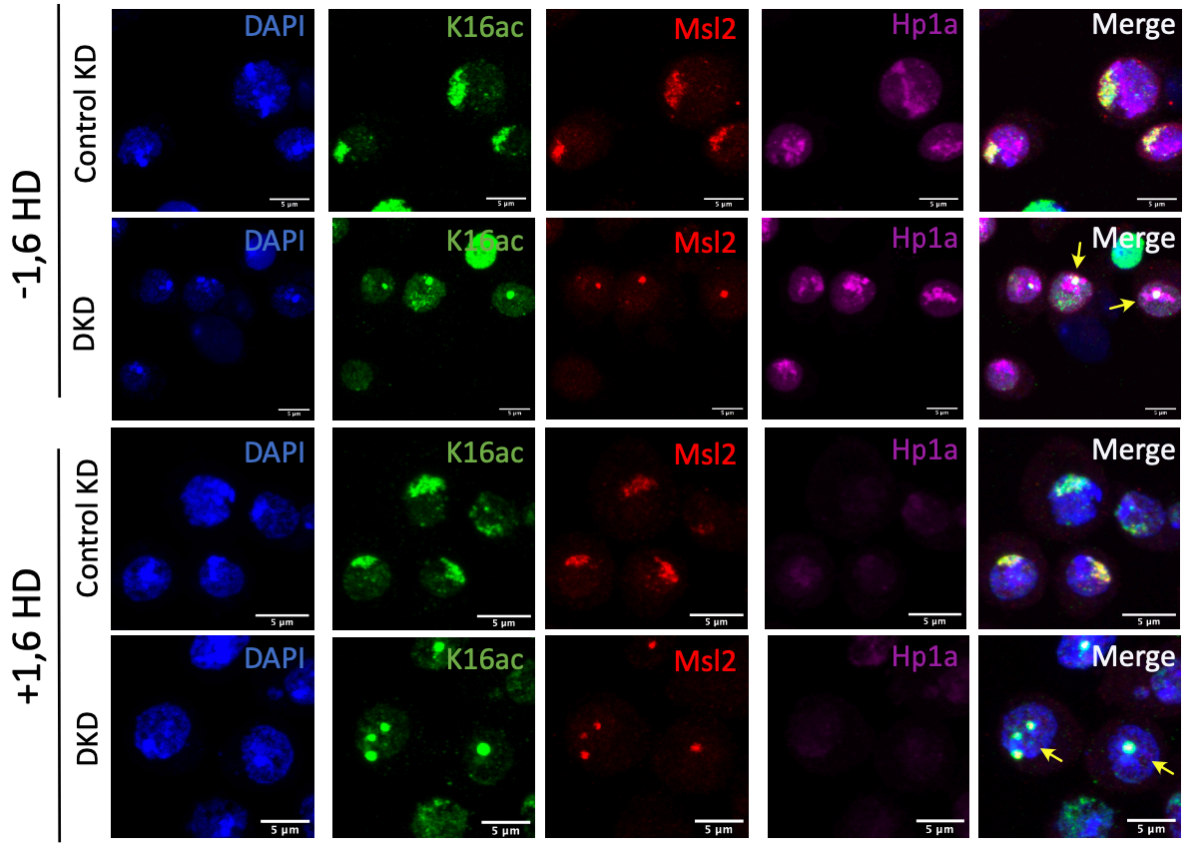


Figure 53: Representative IF confocal images of DNA (DAPI), MSL2, H4K16ac and HP1α spatial redistribution upon Control or DKD RNAi treatments followed by +/- 1,6 Hexanediol treatment. Scale bar denotes 5 *µ*m. Yellow arrows highlight proximity of MSL 'blobs' to DAPI dense/ HP1α-bound chromocenters.

This treatment had no effect on the blobs, ruling out the possibility of LLPS. It has been previously shown that under certain perturbations (for example, deletion of the *roX2* locus), components of the DCC relocate to certain heterochromatic repetitive DNA sequences resembling the MRE (Joshi and Meller 2017). However, while the blobs appear adjacent to heterochromatic HP1α/DAPI dense regions (yellow arrows in merge panel), they don't fully overlap, suggesting that these blobs are likely still on the euchromatic part of the X-chromosome.

Overall, these observations argue that alterations in MSL binding upon DKD are coupled with changes in spatial organization of the dosage-compensated territory. The resulting structures, which are likely still on euchromatin, are not due to LLPS, but are more likely to be polymer-polymer phase-separated (PPPS) condensates, as suggested by a recent study (Valsecchi et al. 2021). These findings highlight how alterations in K36me2/3 not only impact local nucleosome-level interactions but also reshape micron-scale chromosomal territories, underscoring the profound influence of histone modifications on the multiscale organization of chromatin architecture and its functional dynamics.

4. Discussion and future directions

In this dissertation work, I deconstructed the various K36me deposition pathways mediated by the three *Drosophila* H3K36 HMTs Set2, NSD and Ash1, part of which is published (Jayakrishnan et al. 2024). Previous observations were extended to show that H3K36 modifications are not restricted to actively transcribed gene bodies, but also label distinct chromatin compartments, including enhancers and heterochromatin. Functionally, the three distinct methylation signatures (K36me1/2/3) appear to be linked to different regulatory pathways. The comprehensive dissection of the changing K36me landscapes upon various HMT depletion scenarios provided an opportunity to study downstream consequences on transcription, histone modification crosstalk, reader binding specificity (or the lack thereof) and subnuclear organization.

4.1 Technical considerations for ChIP-seq: Old method, new challenges

In this study, I extensively optimized and performed comparative analyses to verify the quality of ChIP-seq datasets. Two different ChIP-seq protocols were employed, each suited better for specific applications, as discussed below. Regarding profiling histone modifications and chromatin proteins, two specific modifications were critical:

a) Counteracting solubility issues with dual 'digestion' steps: Most ChIP protocols solubilize chromatin through either mechanical sonication or MNase digestion, with each method having its own bias. Based on reanalysis of published chromatin input datasets, I found that these methods poorly solubilize heterochromatin (which is likely lost in the insoluble chromatin fraction). In contrast, the in-house method uses a combination of MNase and mild shearing, which produced the most uniform chromatin input profiles. These observations were partly recapitulated experimentally, as a higher fraction of aggregated chromatin was observed when performing only MNase digestion instead of combining MNase with mild shearing. Such recovery biases have been reported before (Becker et al. 2017; Nicetto et al. 2019), and found to distort the relative signals of marks shared between euchromatin and heterochromatin (as this solubilization bias selectively suppresses reads emerging from heterochromatin).

b) Incorporation of 'spike-in' scaling for global changes: In this study, spike-in *D.virilis* cells served as an external reference for changes in *D.melanogaster* genome. I had initially anticipated that this shouldn't change the results much, as most marks interrogated in this study cover a maximum of 25% (for K36me2) of the genome, resulting in sufficient unchanged 'background' representation as a basis for signal normalization. However, over the course of the study, I realized that MNase-ChIP is highly selective: despite the fact that K36me2 marks only ~20% of the genome, 80-90% of the sequence reads emerge from these regions. This results in insufficient background reads, which is crucial when normalizing to total number of reads. In addition to the standard ChIP-seq preprocessing

pipeline, we also used the statistical tool 'csaw' (Lun and Smyth 2016) which was developed based on the principles of edgeR for RNA-seq analysis (see schematic in Figure 54). This tool allows the estimation of a 'normalization factor', a measure of systematic differences in regions expected to have the same signal in all libraries. In line with this, the magnitude of the scaling factor calculated for K36me2 upon NSD RNAi using only 'background' non-differential binding regions is ~1.1 while using the spike-in K36me2 peaks results in a much larger scaling factor of ~1.5, suggesting that the effect is severely underestimated by the former method. This can result in dramatically different interpretations of the biological outcome.

Based on these tests, it may be concluded that inclusion of spike-in is important when a) the mark is widely distributed, with possibly large pools of the factor also part of unmappable regions and b) the genome size is fairly small, resulting in insufficient representation of the 'background' regions in the sequencing reads.

Of note, progress has been steadily made to make ChIP-seq more quantitative (reviewed in Patel et al., 2024). Some of the suggested modifications could be potentially incorporated to improve the current analysis. For instance, the aforementioned study advises using a mixed (spike-in + target) reference genome for read mapping instead of aligning to individual genomes, since closely related genomes (such as mouse and human) may inflate the number of spike-in reads by up to 15%, affecting accurate estimation of the scaling factor. Further, usage of exogenous spike-ins only allows calculation of a single scaling factor, which is then used to globally transform the entire coverage, which may mask true locus-specific changes. For such problems (which resemble complex K36me contributions), methods like SNP-ChIP (Vale-Silva, Markowitz, and Hochwagen 2019) are more appropriate, which may allow calculation of locus-specific scaling factors.

In contrast to the above points, I noticed that high-intensity acoustic shearing performed better than MNase-based protocols for profiling direct targets of DNA binding proteins, such as MSL2. Transcription factors often bind linker DNA between nucleosomes and/or evict nucleosomes to generate accessible sites (nucleosome-free regions or NFRs), which can be prone to over-digestion by MNase. Further, our MNase digestion protocol incorporates only mild-shearing, which maintains protein complex integrity. Under those conditions, all subunits of a complex are mapped to the same chromosomal sites. In contrast, the focused acoustic shearing may result in disruption of a protein complex, so that the constituent subunits are mapped individually according to their chromatin crosslinking potential. Thus, the choice of ChIP-seq protocol should be carefully considered depending on the biological question.

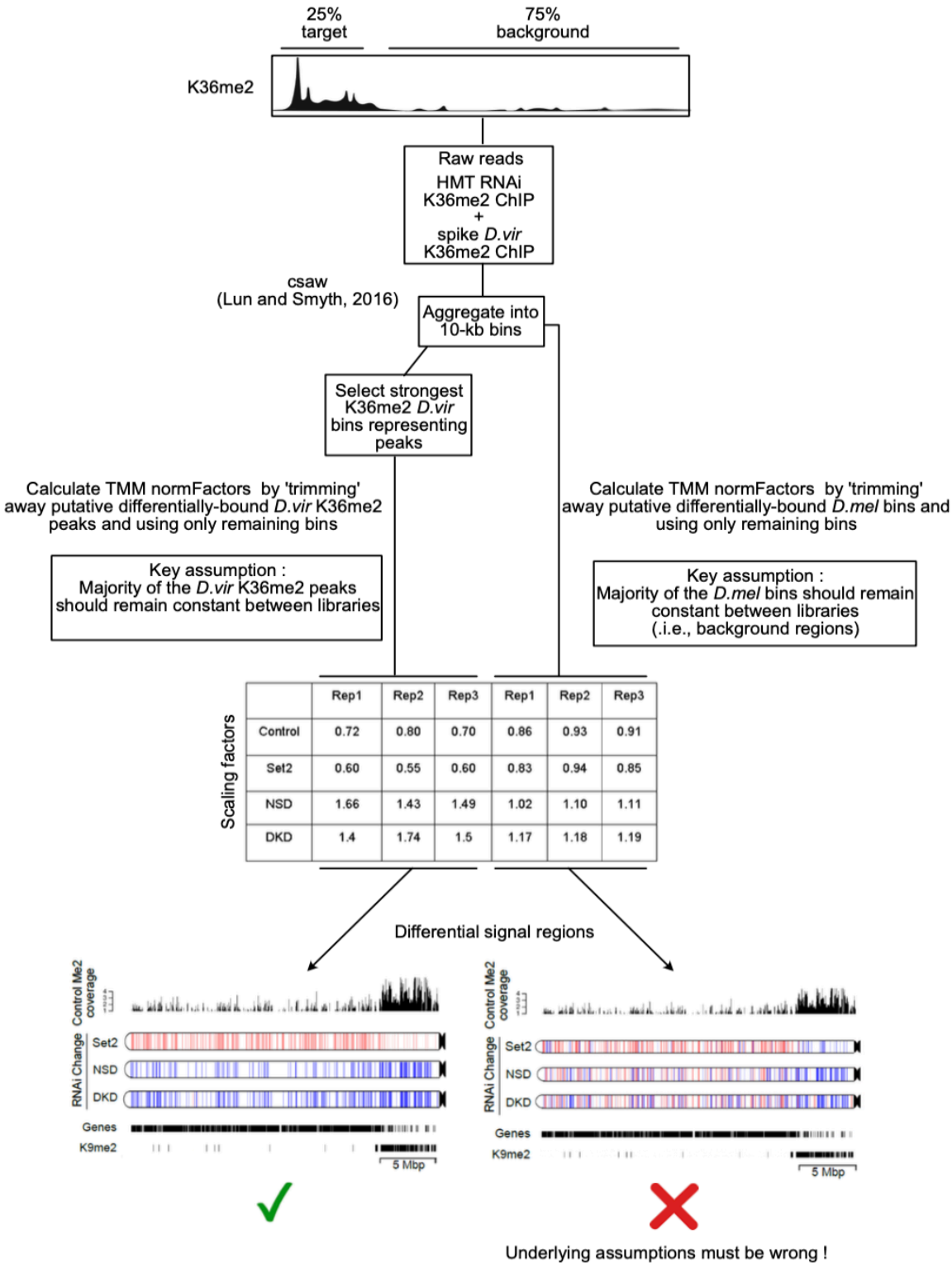


Figure 54: Schematic illustration of the csaw pipeline along with a table of scaling factors obtained for K36me2 ChIPs for indicated replicate experiments. chromoMaps of chromosome 3L generated with two different sets of scaling factors are shown to highlight importance of spike-in inclusion to assess global effects.

4.2 A revised model for K36me deposition and function in *Drosophila*

Previous studies, which were based on bulk (Western blot) or low-resolution ChIP-qPCR assays, had suggested a simple linear pathway involving Set2 and NSD, in which NSD marks gene bodies with K36me2, which then serves as substrate for Set2 to generate K36me3 (Lhoumaud et al. 2014; Bell et al. 2007). Ash1 was assumed to work independently of the Set2/NSD tandem. In contrast, my genome-wide data-driven approach argues for a more complex relationship between the HMTs. Depending on the local genomic context, the degree of interdependence varies. Set2 works largely independently of NSD at highly transcribed housekeeping genes, but cooperates with NSD at certain heavily dimethylated developmental genes. NSD also appears to dominate K36me2/3 patterns at heterochromatin, a fact that was previously underappreciated due to technical issues with ChIP-seq. These observations argue against a 'pioneering' function for NSD to promote Set2-dependent K36me3. In agreement, a recent study biochemically reconstituted the entire mammalian elongating RNA Pol II machinery (which also included SETD2) and observed K36me3 signal on recombinant chromatin templates, confirming the sufficiency of the ePol complex in promoting transcription-coupled H3K36me3 activity (Markert, Soffers, and Farnung 2025).

NSD may have two different functions at heterochromatin and euchromatin. At a subset of euchromatin genes, it may act as a 'maintenance' methyltransferase, where it possibly preserves the active chromatin signature by a read-write mechanism (recollect that NSD has PWWP domains in addition to its catalytic domain). This is supported by the study of PWWP domains of the mammalian NSD2/3, which show a stronger affinity for K36me2 than for K36me3 (Sankaran et al. 2016). This may facilitate appropriate tissue-specific activation of these genes. This proposed function may be partly linked to its K27me3-counteracting activity that was characterized as part of the study. My preliminary analysis suggests that genes marked highly with NSD-dependent K36me2/3 are particularly active in neuronal and testes cells, consistent with recent reports regarding involvement of mammalian NSD1 in the neurodevelopmental disorder Sotos syndrome (Hamagami et al. 2023; Sun et al. 2023).

The function of NSD at heterochromatin is even less understood. Unlike NSD-dependent methylation at euchromatin, which is restricted to genes, heterochromatic NSD-methylation appears as large contiguous intergenic blocks with only a few embedded genes. These blocks likely harbor transposable elements, suggesting a role for NSD in transposon regulation. Further credibility for the hypothesis that K36me2 controls transposon expression comes from overexpression of histone H3.3K27M, which leads to redistribution of H3K36me2 away from pericentromeric heterochromatin to euchromatin, resulting in derepression of heterochromatic transposons. However, this increased transposon expression in the oncogenic mutant histone background is reversed by depletion of Ash1 but not depletion of NSD, suggesting that Ash1 may have a role in transposon regulation as well (despite no detectable K36me1 or Ash1 occupancy) (Chaouch et al. 2021). These observations are concordant with recent findings regarding shared roles for all three H3K36 HMTs in transposon regulation (Lindehell, Schwartz, and Larsson 2023).

In my transcriptome analyses, changes in K36me3 were uncorrelated to transcriptional changes elicited by Set2 RNAi, adding support to the view that Set2 doesn't directly affect transcription. Interestingly, ~100 NSD-dependent protein-coding heterochromatic genes showed a clear downregulation in response to NSD RNAi, an effect that's not observed in any other scenario, which strongly argues that NSD has an activating role at heterochromatin (but not at euchromatin!) in line with a recent study (Saha et al. 2020). This study also demonstrated that heterochromatic genes that selectively lose gene expression upon NSD RNAi also show the same response to depletion of the helicase dADD1 but not HP1α, arguing that NSD is likely part of a heterochromatic complex not related to the Su(var)3-9-H3K9me3-HP1α axis. Possible candidates include other Su(var)s, for example the H4K20me3 methyltransferase Su(var)4-20 which should be tested in future studies.

Teasing apart the multifaceted role of NSD is challenging and future studies should investigate this by utilizing point mutations/deletions to selectively disrupt the targeting, catalytic and auxiliary interaction domains of NSD.

Lastly, it was determined that Ash1 is responsible for almost all observable K36me1 in S2 cells. Evidently, Set2 and NSD rapidly progress through intermediate K36me1 during their catalytic cycle towards generating K36me2/3. In addition to verifying its *bona fide* role as a Polycomb antagonist, substantial Ash1 (and K36me1) was found at certain developmental enhancers, the function of which is unclear. The preliminary data suggest that CBP-dependent H3K27ac is unaffected in Ash1 RNAi, suggesting that if Ash1 has a role in enhancer regulation, it's probably downstream of CBP. An informative experiment might be to assay Ash1 function in a STARR-seq experiment, which will allow the quantitative detection of even subtle contributions. I also observed a strong correlation between K36me1 and a subset of TADs, but a conservative stance on this observation is that K36me1 spreading is merely contained within preexisting boundaries.

4.3 Conservation of the various *in vivo* K36me deposition pathways in mammalian systems

During the course of this study, a study was published which examined K36me1/2/3 in mouse mesenchymal cells involving individual knockouts of SETD2, NSD1-3 and ASH1L (Shipman et al. 2024). This provided a unique resource to compare the findings in the *Drosophila* model system, which revealed several interesting similarities and differences. For instance, active gene body enrichment was observed for K36me3 in both model systems, which was largely deposited by SETD2, independent of other methyltransferases. However, the study did not discuss any evidence regarding heterochromatic K36me2/3 in murine cells, contrasting the description of *Drosophila* NSD. Instead, K36me2 in mice is present more in intergenic regions around regulatory elements and is collectively deposited by NSD1/2 (broad domains) and NSD3/ASH1L (focal peaks). One study reported that the enhancer association of NSD1 depends on H3K18ac

recognition through its PWWP domain and importantly, that the catalytic activity of NSD1 is dispensable for its regulatory function, strongly suggesting that H3K36 HMTs may have 'moonlighting' functions (Sun et al. 2023). Other studies in mouse ESCs have verified the existence of heterochromatic K36me2/3 nanodomains (deposited by SETDB1 and NSD1/2) dispersed within euchromatin which mask hidden enhancers, something that appears to have not been discussed in the Shipman et al., study. With the availability of a large number of curated ChIP-seq datasets along with improved reference genomes, recent studies have discovered K36me2/3 as a mark that is present quite abundantly at previously unmappable heterochromatin even in mammalian genomes. In one study, the authors developed a computational pipeline, which interrogated the fraction of repetitive centromeric reads in ~41,000 ChIP-seq samples, and found H3K36me3 as one of the 8 significantly enriched post-translational modifications (Corless et al. 2023). In another landmark study, realigning ChIP-seq datasets to the complete 'Telomere-To-Telomere' (T2T) human genome showed that after K9me3 and K27me3, K36me3 gained the highest number of peaks by comparison to mapping to the old hg38 genome (Gershman et al. 2022). It is plausible that this heterochromatic K36me3/2 may be controlled by NSD in the mammalian system as well, which is supported by a recent study (Goto et al. 2024). Also, in contrast to findings in *Drosophila*, K36me1 overlapped the other K36me modifications extensively and was also jointly deposited by NSD1-3/ASH1L on gene bodies.

Recently, *in vivo* studies have leveraged artificial tethering approaches (using dCas9 fusions) to systematically target chromatin factors to defined chromatin landing sites, which allows to systematically study the impact of histone modifications without confounding variables. One such study targeted SETD2 (along with a panel of several other HMTs) to a reporter locus and observed the deposition of K36me3, confirming its role as an independent K36 trimethyltransferase which can deposit methylation without interaction with ePol (Policarpi et al. 2024). Interestingly, this study also discovered that it had a mild repressive effect in the context of CTCF motifs, which partly depended on DNA methylation. As *Drosophila* lacks DNA methylation machinery, these experiments need to be extended to *Drosophila* to verify the exact nature of the Set2-dependent K36me3.

Overall, these interspecies comparisons argue that the mechanism of Set2-dependent K36me3 was maintained during evolution, but orthologs of NSD and Ash1 functionally diverged to various degrees.

4.4 Some insights into HMT enzymatic features from *in vivo* data

My study also provides some hints about the enzymatic mechanisms of the HMTs. For instance, the data argue that none of the three HMTs require prior methylation at K36 (with lower methylation state) to exert their function, but rather they can directly act on unmethylated nucleosomes. This contrasts prior observations from *in vitro* studies that conclude pre-methylation is required for *Drosophila* Set2-dependent K36me3 (Bell et al. 2007). However, these observations do not preclude the possibility that Set2 and NSD may also have the capacity to act

on already methylated nucleosomes. In support of this, controlled *in vitro* methylation assays on chemically modified nucleosomes showed that SETD2 can act on K36me0/1/2 and NSD2 can work on K36me0/1, suggesting relaxed specificity with regards to the modification status of H3K36 (Li, Trojer, et al. 2009). Such activity on premodified K36 sites may occur *in vivo* at a subset of genes where Set2 and NSD activity at Ash1-dependent K36me1 domains was observed. As a cautionary note, these early *in vitro* results of single HMT proteins must now be repeated in the context of physiological enzyme complexes (such as shown in Markert et al., 2024), which may significantly alter the preference/activity of these HMTs.

One can also predict the mode of catalysis (processivity versus distributivity) of HMTs based on *in vivo* data (Frederiks et al. 2008; Stulemeijer et al. 2015). For example, reduction in the levels of a distributive enzyme should result in the appearance of lower methylation states. This was observed in my study at Set2-dependent clusters, where K36me1 appeared upon depletion of Set2. On the other hand, NSD RNAi results in coordinated reduction of both K36me2/3 (i.e., relative K36me1/2/3 levels remain the same), arguing for some degree of processivity. However, these patterns can be affected by demethylases as well which can preclude accurate classification. Future studies should utilize single molecule imaging approaches to deconvolve the behavior of the HMTs to quantify discrete kinetics of recruitment (productive versus non-productive interaction with nucleosomes), activation (is adding the first methyl group the rate-limiting step?), catalysis (does the reaction follow first-order or second-order kinetics?) and disengagement (how often does the HMT prematurely terminate the reaction? What is the stability of the HMT-product complex?), which has been recently used to study processivity of histone ubiquitylation by PRC complexes (Teslenko A. 2024).

4.5 Indirect communication between Set2 and NSD

While the results show that the main recruitment mechanisms of Set2 and NSD are largely independent, there is some evidence for crosstalk between these pathways. One striking observation in my analyses is the apparent ‘hyperactivation’ of NSD-dependent K36me2 upon Set2 depletion, which primarily occurs at genes where Set2 is predominantly active (SC-I). These genes, despite having moderate levels of NSD (comparable to occupancy at NSD’s ‘main target’ SC-IIA genes), have far lower K36me2 when Set2 is present, suggesting that NSD’s activity is dampened by the presence of Set2. One possible explanation is that NSD and Set2 locally compete for shared cofactors (for example, through differing affinities to the metabolites SAM/SAH). This may occur within the confines of a defined nuclear subcompartment, such as ‘transcription factories’. A ‘local’ competition would also conveniently explain why this effect is restricted to only Set2 and NSD, while HMTs at other subcompartments (like Su(var)3-9 in heterochromatic domains or E(z) in Polycomb domains) are unaffected. This hypothesis can possibly be investigated through super-resolution imaging of Set2 and NSD to see if they partitioning into foci rich for nascent RNA and ePol characteristic of transcriptional factories.

Further, looking at other gene body modifications, such as Dot1-mediated K79me3, might provide some clues if competition for metabolites or co-factors is relevant. In addition to a presumed catalytic activation, NSD's chromatin association is also apparently strengthened in the absence of Set2. The basis of this observation is unclear but may arise due to Set2 actively excluding NSD directly (for example, by physical impediment by ePol) or indirectly (by generating a refractory chromatin state). It's also possible that loss of Set2 triggers demethylation pathways to generate K36me2, which may be recognized by NSD (through its PWWP domains) to establish a feedback loop.

On other hand, NSD appears to positively influence ePol association to chromatin, in line with some prior studies that postulated that NSD accompanies the insulator BEAF-32 to promote transcription at some genes (Lhoumaud et al. 2014). Future studies need to systematically interrogate how Set2 and NSD affect the distribution and activities of each other in a more controlled manner.

4.6 K36me2 and K36me3 – Functionally redundant in all contexts?

Our experiments examining the distribution of two exemplary K36me3 reader proteins, JASPer and MSL3, upon various perturbation scenarios show that both proteins likely bind K36me2 in addition to K36me3. It's unclear if this relaxed specificity is a product of directed selection (i.e., confers evolutionary advantage), but one can imagine scenarios where it might prove useful. K36me2 may compensate for K36me3 in conditions where the latter can fluctuate, for example possibly during heat stress (which is accompanied by rapid transcriptional shutdown) or in tissues with high expression levels of demethylases, which may shift the equilibrium towards lower methylation states. By maintaining reader binding, active chromatin signatures may be preserved during these transient fluctuations and ensure robustness to the system. These results do not preclude the possibility that other histone modifications may also contribute to reader binding alone or in combination, which may be investigated using high-throughput methods using modified nucleosome libraries (Dann et al. 2017; Hananya, Koren, and Muir 2024). Interestingly, despite the strong affinity of JASPer to K36me2-nucleosomes *in vitro*, many heterochromatic domains are observed *in vivo* where K36me2 alone is insufficient to recruit the readers. This would argue that certain factors can counteract K36me2 and repel JASPer, thus providing an interesting avenue for future research. Lastly, over a dozen K36me readers exist in *Drosophila* and it remains to be determined how pervasive the K36me2/K36me3 redundancy is. Previous mammalian pulldown studies have documented a large overlap between K36me2 and K36me3 readers (Sankaran et al. 2016).

4.7 Defective dosage compensation and revelations into histone methylation-acetylation crosstalk

Chronologically, one of the first objectives of this thesis project was to determine the epigenetic determinants of MSL3 targeting in the context of dosage compensation, which led to a screen for

different methyltransferases and adaptor proteins using a microscopy approach. This exercise led to the discovery of the bizarre 'blob' phenotype upon the combined depletion of Set2 and NSD. This observation motivated the full-fledged characterization of *Drosophila* K36 methylation chromatin landscape, presented in this thesis. The role of K36me in MSL3 targeting has been addressed before in several studies (Morales et al. 2005; Larschan et al. 2007; Buscaino, Legube, and Akhtar 2006; Sural et al. 2008) using either Set2 mutations/depletions or expression of MSL3 mutants lacking its chromodomain. Theoretically, the chromodomain mutants of MSL3 used in Morales et al., should produce the same 'blob' phenotype as combined depletion of Set2/NSD, but as the resolution in this study was low, direct comparison is difficult. Reproducing these experiments were also technically challenging due to the confounding presence of endogenous MSL3 in addition to the overexpression construct. The basis of the blob phenotype is ambiguous, but appears to be reproducibly observed in many technical variations (RNAi duration, various immunostaining protocols etc.). Of course, this doesn't rule out the possibility that this phenotype arises due to the loss of methylation of non-histone proteins by Set2 and NSD, in which case the phenotype won't be reproduced with MSL3 chromodomain mutants. Hence, these experiments need to be validated with MSL3 variants (preferably in a clean background, like MSL3-deficient flies) to rule out biological artefacts. The striking round appearance of the X-chromosome bearing residual MSL proteins has never been observed before despite extensive manipulation of the different components of the MSL complex. The blob structures are resistant to 1,6-hexanediol, indicative of PPPS. It's unclear if this structure is associated with a remodelling of the underlying X-chromosome structure, which can be addressed using FISH and/or 3C-based approaches. Previous studies have shown that High Affinity Sites (HAS) tend to cluster in nuclei (Schauer et al. 2017; Pal et al. 2019) independent of the MSL complex and possibly mediated by CLAMP (Jordan and Larschan 2021), positing that CLAMP-chromatin interactions may be affected in the HMT depletions.

Our system also facilitated the study of the crosstalk between histone methylation and acetylation. Prior studies have described this crosstalk in several model systems, primarily in the context of K36me3-mediated recruitment of a histone deacetylase complex indicative of an antagonistic relationship between the two types of marks. In contrast, my results demonstrate that H4K16ac is slightly diminished on the X-chromosome upon reduction of K36me2/3 (as a consequence of defective dosage compensation), provide an example of a synergistic relationship between methylation and acetylation, as posited by the prevalent reader-writer model. We also corroborate previous studies which describe that MOF eviction from the X-chromosome (through perturbation of either genetic or epigenetic targeting) also results in partial redistribution of K16ac to lower affinity secondary autosomal NSL sites, highlighting the fine balance between two competing complexes. (Bell et al. 2007; Prestel et al. 2010). It has to be emphasized that the prevalent model of K36me-H4K16ac synergy at dosage compensated X-chromosomes is being challenged by a recent study which utilized *Set2*-null, unmodifiable H4K16R and H3K36R mutant flies (Salzler et al. 2024). This study demonstrated a poor correlation between X-linked genes with reduced gene expression in these genetic backgrounds, with some of the trends also observed in female flies, necessitating a more thorough investigation of the link between K36me3 and H4K16ac in regulating male dosage compensation.

Some interesting, but less explored findings suggest that prior acetylation may also influence methylation patterns (Feller et al. 2015). One intriguing observation from Feller et al., is the estimation that K36me2 and K36me3 are ~50 and 30% lower, respectively, in female cells compared to male cells using quantitative mass spectrometry. Ablation of MOF in male cells results in a lowering of K36me2/3 abundances towards resembling females (i.e., a reduction), suggesting that the DCC may promote K36me2/3 through unknown mechanisms. Interestingly, this MOF-dependent 'activation' is also visible in human HeLa cells, suggesting a possibly conserved crosstalk. Since K36me2 shows the most striking sex difference in *Drosophila*, it's possible that NSD may be influenced by H4K16ac. The nature of this effect may be through direct allosteric stimulation (akin to activation of yeast Dot1 K79 methylation activity by H4K16ac (Valencia-Sanchez et al. 2021)) or indirect (possibly improved/altered targeting), which should be pursued in future studies.

4.8 Other limitations of this study

While this study extensively characterized the effect of HMT depletions on histone modifications, it's important to remember that these methyltransferases likely also have non-histone targets, which can complicate interpretation of transcriptomes or organism-level phenotypes. For example, SETD2 has been shown to methylate alpha-tubulin as well as actin, suggesting it has an important role in cytoskeletal organization (Park et al. 2016; Seervai et al. 2020).

Our RNAi treatments typically result in strong reduction of the target proteins but do not eliminate them entirely. Potentially this could lead to non-uniform retention of the residual target protein at selected genomic regions. While there is confidence in the robustness of my observations (as similar patterns are evident in numerous other published datasets, including mouse knockout cells), it might be helpful to verify these findings using knockout cells or flies.

I also performed correlative analyses between ChIP-seq and bulk RNA-seq datasets to infer how K36me1/2/3 influence transcription. However, standard RNA-seq may not capture subtle differences in transcriptional defects, which may involve aspects of production of nascent RNA, RNA stability and processing, etc. To address this sensitively, techniques like PRO-seq, SLAM-seq or GRO-seq coupled with spike-ins (to address global changes) may be used to substantiate some of the observations presented in this thesis.

5. Acknowledgements

First and foremost, I would like to thank Prof. Peter Becker – my doctoral supervisor. In addition to his timely advice and tailored mentorship, he also gave me the freedom to steer the project in the direction of my interests deep into the 'uncharted waters' of methyltransferase biology. Despite this topic being outside the lab's comfort zone of 'Dosage Compensation', he supported my way of storytelling and in the end my manuscript has only one Supplementary Figure dedicated to the topic that I was hired to work on. He also played an important role in my transformation from an 'infinite-energy conquer-the-world' type of researcher into a more pragmatic one. At several points during my PhD, I was deep down in scientific rabbit holes and he always gently nudged me back into refocusing on priorities. In hindsight, having my main project wrapped up mid-way through my PhD greatly alleviated stress in the later stages of the doctorate and gave me more freedom to explore high-risk projects.

I would also like to thank Dr. Catherine Regnard, who was the more 'hands-on' mentor during this thesis work. She also had my back in pushing the boundaries of the project and her attention to detail was very important in many occasions. I also need to thank her for initiating the collaboration with Dr. Vaclav Veverka which allowed us to clinch the conclusive proof regarding JASPer binding to K36me2/3.

I would like to thank all members of the Becker lab, especially Silke and Aline for their technical support. I would also like to thank all members of the Molecular Biology department as well as the Bioinformatics Core (Dr. Tobias Straub and Dr. Tamas Schauer) for generously sharing reagents, scripts and advice.

The two doctoral programs I was part of, IMPRS and IRTG, played very important roles in my holistic development as a researcher and a person for which I'm very grateful. My entire PhD journey started with IMPRS who helped me find Peter's lab. As a 23-year-old who was directly jumping into a PhD from an integrated Bachelors/Masters, IMPRS' structured program with lectures, workshops and activities eased my transition from being a student into a (~65%) full-time scientist. The folks at the IMPRS office (Dr. Hans-Joerg Schaeffer, Dr. Ingrid Wolf and Dr. Marta Cipinska) were attentive to the needs of every student. They were also very generous with relocation support, conference funds and allowances for technical workshops for which I'm very thankful. IRTG was another important cornerstone of my PhD experience. IRTG brought together chromatin researchers in Munich which gave me a sense of scientific belonging. Dr. Elizabeth-Schroeder Reiter, the coordinator of the program, helped me navigate many bureaucratic hurdles during the PhD journey and constantly tried to improve the experience of every IRTGer. I'll forever cherish the memory of the IRTG retreats which offered an unrivalled combination of fun and camaraderie, fostering bonds among everyone in a way unmatched by any other program.

I also extend my gratitude to my former research mentors, especially Dr. Kundan Sengupta, Dr. Roopali Pradhan and Dr. Eric Schirmer, for nurturing my growth as a budding scientist.

I would like to also thank my friends Gizem, Fotios, Janet, Anna, Vera, Nikas, Nitya, Payal, Ankita, Shubham and Divanshu for a lot of great memories and support over the years. I'm especially thankful to my partner Avani Koparkar who rejoiced for me during the 'highs' and had my back during the 'lows' of my PhD.

Lastly, I would like to thank my parents, who put their unwavering faith in me to pursue a career of my own choosing.

6. References

Alabert, C., C. Loos, M. Voelker-Albert, S. Graziano, I. Forne, N. Reveron-Gomez, L. Schuh, J. Hasenauer, C. Marr, A. Imhof, and A. Groth. 2020. 'Domain Model Explains Propagation Dynamics and Stability of Histone H3K27 and H3K36 Methylation Landscapes', *Cell Rep*, 30: 1223-34 e8.

Alberti, S., A. Gladfelter, and T. Mittag. 2019. 'Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates', *Cell*, 176: 419-34.

Albig, C., C. Wang, G. P. Dann, F. Wojcik, T. Schauer, S. Krause, S. Maenner, W. Cai, Y. Li, J. Girton, T. W. Muir, J. Johansen, K. M. Johansen, P. B. Becker, and C. Regnard. 2019. 'JASPer controls interphase histone H3S10 phosphorylation by chromosomal kinase JIL-1 in Drosophila', *Nat Commun*, 10: 5343.

Alekseyenko, A. A., A. A. Gorchakov, B. M. Zee, S. M. Fuchs, P. V. Kharchenko, and M. I. Kuroda. 2014. 'Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs', *Genes Dev*, 28: 1445-60.

Anand, L., and C. M. Rodriguez Lopez. 2022. 'ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes', *BMC Bioinformatics*, 23: 33.

Barral, A., G. Pozo, L. Ducrot, G. L. Papadopoulos, S. Sauzet, A. J. Oldfield, G. Cavalli, and J. Dejardin. 2022. 'SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers', *Mol Cell*, 82: 816-32 e12.

Becker, J. S., R. L. McCarthy, S. Sidoli, G. Donahue, K. E. Kaeding, Z. He, S. Lin, B. A. Garcia, and K. S. Zaret. 2017. 'Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes', *Mol Cell*, 68: 1023-37 e15.

Bell, O., T. Conrad, J. Kind, C. Wirbelauer, A. Akhtar, and D. Schubeler. 2008. 'Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster', *Mol Cell Biol*, 28: 3401-9.

Bell, O., C. Wirbelauer, M. Hild, A. N. Scharf, M. Schwaiger, D. M. MacAlpine, F. Zilbermann, F. van Leeuwen, S. P. Bell, A. Imhof, D. Garza, A. H. Peters, and D. Schubeler. 2007. 'Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila', *EMBO J*, 26: 4974-84.

Bonnet, J., R. G. H. Lindeboom, D. Pokrovsky, G. Stricker, M. H. Celik, R. A. W. Rupp, J. Gagneur, M. Vermeulen, A. Imhof, and J. Muller. 2019. 'Quantification of Proteins and Histone Marks in Drosophila Embryos Reveals Stoichiometric Relationships Impacting Chromatin Regulation', *Dev Cell*, 51: 632-44 e6.

Brown, J. C., B. D. McMichael, V. Vandadi, A. Mukherjee, H. R. Salzler, and A. G. Matera. 2024. 'Lysine-36 of Drosophila histone H3.3 supports adult longevity', *G3 (Bethesda)*.

Buscaino, A., G. Legube, and A. Akhtar. 2006. 'X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3', *EMBO Rep*, 7: 531-8.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. 2005. 'Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription', *Cell*, 123: 581-92.

Carvalho, S., A. C. Vitor, S. C. Sridhara, F. B. Martins, A. C. Raposo, J. M. Desterro, J. Ferreira, and S. F. de Almeida. 2014. 'SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint', *Elife*, 3: e02482.

Chaouch, A., J. Berlandi, C. C. L. Chen, F. Frey, S. Badini, A. S. Harutyunyan, X. Chen, B. Krug, S. Hebert, A. Jeibmann, C. Lu, C. L. Kleinman, M. Hasselblatt, P. Lasko, M. Shirinian, and N. Jabado. 2021. 'Histone H3.3 K27M and K36M mutations de-repress transposable elements through perturbation of antagonistic chromatin marks', *Mol Cell*, 81: 4876-90 e7.

Chintapalli, V. R., J. Wang, and J. A. Dow. 2007. 'Using FlyAtlas to identify better Drosophila melanogaster models of human disease', *Nat Genet*, 39: 715-20.

Colmenares, S. U., J. M. Swenson, S. A. Langley, C. Kennedy, S. V. Costes, and G. H. Karpen. 2017. 'Drosophila Histone Demethylase KDM4A Has Enzymatic and Non-enzymatic Roles in Controlling Heterochromatin Integrity', *Dev Cell*, 42: 156-69 e5.

Corless, Samuel, Noor-Pratap Singh, Nezha S. Benabdhalla, Jasmin Böhm, Alexander M. Simon, Vojtěch Dolejš, Simon Anders, Ana Banito, and Sylvia Earhardt. 2023. 'The bromodomain inhibitor JQ1 is a molecular glue targeting centromeres', *bioRxiv*.

Corvalan, A. Z., and H. A. Coller. 2021. 'Methylation of histone 4's lysine 20: a critical analysis of the state of the field', *Physiol Genomics*, 53: 22-32.

Cui, Y., and C. Bustamante. 2000. 'Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure', *Proc Natl Acad Sci U S A*, 97: 127-32.

Dann, G. P., G. P. Liszczak, J. D. Bagert, M. M. Muller, U. T. T. Nguyen, F. Wojcik, Z. Z. Brown, J. Bos, T. Panchenko, R. Pihl, S. B. Pollock, K. L. Diehl, C. D. Allis, and T. W. Muir. 2017. 'ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference', *Nature*, 548: 607-11.

Daugaard, M., A. Baude, K. Fugger, L. K. Povlsen, H. Beck, C. S. Sorensen, N. H. Petersen, P. H. Sorensen, C. Lukas, J. Bartek, J. Lukas, M. Rohde, and M. Jaattela. 2012. 'LEDGF (p75) promotes DNA-end resection and homologous recombination', *Nat Struct Mol Biol*, 19: 803-10.

de Almeida, S. F., A. R. Grosso, F. Koch, R. Fenouil, S. Carvalho, J. Andrade, H. Levezinho, M. Gut, D. Eick, I. Gut, J. C. Andrau, P. Ferrier, and M. Carmo-Fonseca. 2011. 'Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36', *Nat Struct Mol Biol*, 18: 977-83.

Delandre, Caroline , John P.D. McMullen, Jonas Paulsen, Philippe Collas, and Owen J. Marshall. 2022. 'Eight principal chromatin states functionally segregate the fly genome into developmental and housekeeping roles', *bioRxiv*.

Depierre, D., C. Perrois, N. Schickele, P. Lhoumaud, M. Abdi-Galab, O. Fosseprez, A. Heurteau, R. Margueron, and O. Cuvier. 2023. 'Chromatin in 3D distinguishes dMes-4/NSD and Hypb/dSet2 in protecting genes from H3K27me3 silencing', *Life Sci Alliance*, 6.

Dorafshan, E., T. G. Kahn, A. Glotov, M. Savitsky, M. Walther, G. Reuter, and Y. B. Schwartz. 2019. 'Ash1 counteracts Polycomb repression independent of histone H3 lysine 36 methylation', *EMBO Rep*, 20.

Dukatz, M., K. Holzer, M. Choudalakis, M. Emperle, C. Lungu, P. Bashtrykov, and A. Jeltsch. 2019. 'H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction', *J Mol Biol*, 431: 5063-74.

Egelhofer, T. A., A. Minoda, S. Klugman, K. Lee, P. Kolasinska-Zwierz, A. A. Alekseyenko, M. S. Cheung, D. S. Day, S. Gadel, A. A. Gorchakov, T. Gu, P. V. Kharchenko, S. Kuan, I. Latorre, D. Linder-Basso, Y. Luu, Q. Ngo, M. Perry, A. Rechtsteiner, N. C. Riddle, Y. B. Schwartz, G. A. Shanower, A. Vielle, J. Ahringer, S. C. Elgin, M. I. Kuroda, V. Pirrotta, B. Ren, S. Strome, P. J. Park, G. H. Karpen, R. D. Hawkins, and J. D. Lieb. 2011. 'An assessment of histone-modification antibody quality', *Nat Struct Mol Biol*, 18: 91-3.

ENCODE. 2012. 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489: 57-74.

Erdel, F., and K. Rippe. 2018. 'Formation of Chromatin Subcompartments by Phase Separation', *Biophys J*, 114: 2262-70.

Feller, C., I. Forne, A. Imhof, and P. B. Becker. 2015. 'Global and specific responses of the histone acetylome to systematic perturbation', *Mol Cell*, 57: 559-71.

Feller, C., M. Prestel, H. Hartmann, T. Straub, J. Soding, and P. B. Becker. 2012. 'The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset', *Nucleic Acids Res*, 40: 1509-22.

Frederiks, F., M. Tzouros, G. Oudgenoeg, T. van Welsem, M. Fornerod, J. Krijgsveld, and F. van Leeuwen. 2008. 'Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states', *Nat Struct Mol Biol*, 15: 550-7.

Gerland, T. A., B. Sun, P. Smialowski, A. Lukacs, A. W. Thomae, and A. Imhof. 2017. 'The Drosophila speciation factor HMR localizes to genomic insulator sites', *PLoS One*, 12: e0171798.

Gershman, A., M. E. G. Sauria, X. Guitart, M. R. Vollger, P. W. Hook, S. J. Hoyt, M. Jain, A. Shumate, R. Razaghi, S. Koren, N. Altemose, G. V. Caldas, G. A. Logsdon, A. Rhie, E. E. Eichler, M. C. Schatz, R. J. O'Neill, A. M. Phillippy, K. H. Miga, and W. Timp. 2022. 'Epigenetic patterns in a complete human genome', *Science*, 376: eabj5089.

Gerstein, M. B., Z. J. Lu, E. L. Van Nostrand, C. Cheng, B. I. Arshinoff, T. Liu, K. Y. Yip, R. Robilotto, A. Rechtsteiner, K. Ikegami, P. Alves, A. Chateigner, M. Perry, M. Morris, R. K. Auerbach, X. Feng, J. Leng, A. Vielle, W. Niu, K. Rhrissorrakrai, A. Agarwal, R. P. Alexander, G. Barber, C. M. Brdlik, J. Brennan, J. J. Brouillet, A. Carr, M. S. Cheung, H. Clawson, S. Contrino, L. O. Dannenberg, A. F. Dernburg, A. Desai, L. Dick, A. C. Dose, J. Du, T. Egelhofer, S. Ercan, G. Euskirchen, B. Ewing, E. A. Feingold, R. Gassmann, P. J. Good, P. Green, F. Gullier, M. Gutwein, M. S. Guyer, L. Habegger, T. Han, J. G. Henikoff, S. R. Henz, A. Hinrichs, H. Holster, T. Hyman, A. L. Iniguez, J. Janette, M. Jensen, M. Kato, W. J. Kent, E. Kephart, V. Khivansara, E. Khurana, J. K. Kim, P. Kolasinska-Zwierz, E. C. Lai, I. Latorre, A. Leahey, S. Lewis, P. Lloyd, L. Lochovsky, R. F. Lowdon, Y. Lubling, R. Lyne, M. MacCoss, S. D. Mackowiak, M. Mangone, S. McKay, D. Mecenas, G. Merrihew, D. M. Miller, 3rd, A. Muroyama, J. I. Murray, S. L. Ooi, H. Pham, T. Phippen, E. A. Preston, N. Rajewsky, G. Ratsch, H. Rosenbaum, J. Rozowsky, K. Rutherford, P. Ruzanov, M. Sarov, R. Sasidharan, A. Sboner, P. Scheid, E. Segal, H. Shin, C. Shou, F. J. Slack, C. Slightam, R. Smith, W. C. Spencer, E. O. Stinson, S. Taing, T. Takasaki, D. Vafeados, K. Voronina, G. Wang, N. L. Washington, C. M. Whittle, B. Wu, K. K. Yan, G. Zeller, Z. Zha, M. Zhong, X. Zhou, Encode Consortium mod, J. Ahringer, S. Strome, K. C. Gunsalus, G. Micklem, X. S. Liu, V. Reinke, S. K. Kim, L. W. Hillier, S. Henikoff, F. Piano, M. Snyder, L. Stein, J. D. Lieb, and R. H. Waterston. 2010. 'Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project', *Science*, 330: 1775-87.

Girardot, C., J. Scholtalbers, S. Sauer, S. Y. Su, and E. E. Furlong. 2016. 'Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers', *BMC Bioinformatics*, 17: 419.

Goto, N., K. Suke, N. Yonezawa, H. Nishihara, T. Handa, Y. Sato, T. Kujirai, H. Kurumizaka, K. Yamagata, and H. Kimura. 2024. 'ISWI chromatin remodeling complexes recruit NSD2 and H3K36me2 in pericentromeric heterochromatin', *J Cell Biol*, 223.

Gu, Z., R. Eils, and M. Schlesner. 2016. 'Complex heatmaps reveal patterns and correlations in multidimensional genomic data', *Bioinformatics*, 32: 2847-9.

Guan, H., P. Wang, P. Zhang, C. Ruan, Y. Ou, B. Peng, X. Zheng, J. Lei, B. Li, C. Yan, and H. Li. 2023. 'Diverse modes of H3K36me3-guided nucleosomal deacetylation by Rpd3S', *Nature*, 620: 669-75.

Hamagami, N., D. Y. Wu, A. W. Clemens, S. A. Nettles, A. Li, and H. W. Gabel. 2023. 'NSD1 deposits histone H3 lysine 36 dimethylation to pattern non-CG DNA methylation in neurons', *Mol Cell*, 83: 1412-28 e7.

Hananya, N., S. Koren, and T. W. Muir. 2024. 'Interrogating epigenetic mechanisms with chemically customized chromatin', *Nat Rev Genet*, 25: 255-71.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren. 2007. 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nat Genet*, 39: 311-8.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38: 576-89.

Hendy, O., L. Serebreni, K. Bergauer, F. Muerdter, L. Huber, F. Nemcko, and A. Stark. 2022. 'Developmental and housekeeping transcriptional programs in Drosophila require distinct chromatin remodelers', *Mol Cell*, 82: 3598-612 e7.

Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, and W. J. Kent. 2006. 'The UCSC Genome Browser Database: update 2006', *Nucleic Acids Res*, 34: D590-8.

Huang, C., F. Yang, Z. Zhang, J. Zhang, G. Cai, L. Li, Y. Zheng, S. Chen, R. Xi, and B. Zhu. 2017. 'Mrg15 stimulates Ash1 H3K36 methyltransferase activity and facilitates Ash1 Trithorax group protein function in Drosophila', *Nat Commun*, 8: 1649.

Huang, H., H. Weng, K. Zhou, T. Wu, B. S. Zhao, M. Sun, Z. Chen, X. Deng, G. Xiao, F. Auer, L. Klemm, H. Wu, Z. Zuo, X. Qin, Y. Dong, Y. Zhou, H. Qin, S. Tao, J. Du, J. Liu, Z. Lu, H. Yin, A. Mesquita, C. L. Yuan, Y. C. Hu, W. Sun, R. Su, L. Dong, C. Shen, C. Li, Y. Qing, X. Jiang, X. Wu, M. Sun, J. L. Guan, L. Qu, M. Wei, M. Muschen, G. Huang, C. He, J. Yang, and J. Chen. 2019. 'Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally', *Nature*, 567: 414-19.

Hyun, K., J. Jeon, K. Park, and J. Kim. 2017. 'Writing, erasing and reading histone lysine methylations', *Exp Mol Med*, 49: e324.

Jagtap, P. K. A., M. Muller, A. E. Kiss, A. W. Thomae, K. Lapouge, M. Beck, P. B. Becker, and J. Hennig. 2023. 'Structural basis of RNA-induced autoregulation of the DExH-type RNA helicase maleless', *Mol Cell*, 83: 4318-33 e10.

Jayakrishnan, Muhunden, Magdalena Havlová, Václav Veverka, Catherine Regnard, and Peter B Becker. 2024. 'Genomic context-dependent histone H3K36 methylation by three Drosophila methyltransferases and implications for dedicated chromatin readers', *Nucleic Acids Research*.

Jayakumar, S., M. Patel, F. Boulet, H. Aziz, G. N. Brooke, H. Tummala, and M. M. Pradeepa. 2024. 'PSIP1/LEDGF reduces R-loops at transcription sites to maintain genome integrity', *Nat Commun*, 15: 361.

Jenuwein, T., and C. D. Allis. 2001. 'Translating the histone code', *Science*, 293: 1074-80.

Jha, D. K., and B. D. Strahl. 2014. 'An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair', *Nat Commun*, 5: 3965.

Jordan, W., 3rd, and E. Larschan. 2021. 'The zinc finger protein CLAMP promotes long-range chromatin interactions that mediate dosage compensation of the Drosophila male X-chromosome', *Epigenetics Chromatin*, 14: 29.

Joshi, S. S., and V. H. Meller. 2017. 'Satellite Repeats Identify X Chromatin for Dosage Compensation in Drosophila melanogaster Males', *Curr Biol*, 27: 1393-402 e2.

Kapoor-Vazirani, P., and P. M. Vertino. 2014. 'A dual role for the histone methyltransferase PR-SET7/SETD8 and histone H4 lysine 20 monomethylation in the local regulation of RNA polymerase II pausing', *J Biol Chem*, 289: 7425-37.

Kharchenko, P. V., A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, D. Linder-Basso, A. Plachetka, G. Shanower, M. Y. Tolstorukov, L. J. Luquette, R. Xi, Y. L. Jung, R. W. Park, E. P. Bishop, T. K. Canfield, R. Sandstrom, R. E. Thurman, D. M. MacAlpine, J. A. Stamatoyannopoulos, M. Kellis, S. C. Elgin, M. I. Kuroda, V. Pirrotta, G. H. Karpen, and P. J. Park. 2011. 'Comprehensive analysis of the chromatin landscape in Drosophila melanogaster', *Nature*, 471: 480-5.

Kim, D., B. J. Blus, V. Chandra, P. Huang, F. Rastinejad, and S. Khorasanizadeh. 2010. 'Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain', *Nat Struct Mol Biol*, 17: 1027-9.

Kim, J. H., B. B. Lee, Y. M. Oh, C. Zhu, L. M. Steinmetz, Y. Lee, W. K. Kim, S. B. Lee, S. Buratowski, and T. Kim. 2016. 'Modulation of mRNA and lncRNA expression dynamics by the Set2-Rpd3S pathway', *Nat Commun*, 7: 13534.

Klymenko, T., and J. Muller. 2004. 'The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins', *EMBO Rep*, 5: 373-7.

Kolasinska-Zwierz, P., T. Down, I. Latorre, T. Liu, X. S. Liu, and J. Ahringer. 2009. 'Differential chromatin marking of introns and expressed exons by H3K36me3', *Nat Genet*, 41: 376-81.

Konrad, S. F., W. Vanderlinden, and J. Lipfert. 2022. 'Quantifying epigenetic modulation of nucleosome breathing by high-throughput AFM imaging', *Biophys J*, 121: 841-51.

Lagarou, A., A. Mohd-Sarip, Y. M. Moshkin, G. E. Chalkley, K. Bezstarosti, J. A. Demmers, and C. P. Verrijzer. 2008. 'dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing', *Genes Dev*, 22: 2799-810.

Langmead, B., and S. L. Salzberg. 2012. 'Fast gapped-read alignment with Bowtie 2', *Nat Methods*, 9: 357-9.

Larschan, E., A. A. Alekseyenko, A. A. Gortchakov, S. Peng, B. Li, P. Yang, J. L. Workman, P. J. Park, and M. I. Kuroda. 2007. 'MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism', *Mol Cell*, 28: 121-33.

Larschan, E., E. P. Bishop, P. V. Kharchenko, L. J. Core, J. T. Lis, P. J. Park, and M. I. Kuroda. 2011. 'X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila', *Nature*, 471: 115-8.

LeRoy, G., O. Oksuz, N. Descostes, Y. Aoi, R. A. Ganai, H. O. Kara, J. R. Yu, C. H. Lee, J. Stafford, A. Shilatifard, and D. Reinberg. 2019. 'LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells', *Sci Adv*, 5: eaay3068.

Lhoumaud, P., M. Hennion, A. Gamot, S. Cuddapah, S. Queille, J. Liang, G. Micas, P. Morillon, S. Urbach, O. Bouchez, D. Severac, E. Emberly, K. Zhao, and O. Cuvier. 2014. 'Insulators recruit histone methyltransferase dMes4 to regulate chromatin of flanking genes', *EMBO J*, 33: 1599-613.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.

Li, Y., R. L. Armstrong, R. J. Duronio, and D. M. MacAlpine. 2016. 'Methylation of histone H4 lysine 20 by PR-Set7 ensures the integrity of late replicating sequence domains in Drosophila', *Nucleic Acids Res*, 44: 7204-18.

Li, Y., P. Trojer, C. F. Xu, P. Cheung, A. Kuo, W. J. Drury, 3rd, Q. Qiao, T. A. Neubert, R. M. Xu, O. Gozani, and D. Reinberg. 2009. 'The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate', *J Biol Chem*, 284: 34283-95.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science*, 326: 289-93.

Lin, Y. Y., P. Muller, E. Karagianni, N. Hepp, F. Mueller-Planitz, W. Vanderlinden, and J. Lipfert. 2024. 'Epigenetic Histone Modifications H3K36me3 and H4K5/8/12/16ac Induce Open Polynucleosome Conformations via Different Mechanisms', *J Mol Biol*, 436: 168671.
Lindehell, H., A. Glotov, E. Dorafshan, Y. B. Schwartz, and J. Larsson. 2021. 'The role of H3K36 methylation and associated methyltransferases in chromosome-specific gene regulation', *Sci Adv*, 7: eabh4390.
Lindehell, H., Y. B. Schwartz, and J. Larsson. 2023. 'Methylation of lysine 36 on histone H3 is required to control transposon activities in somatic cells', *Life Sci Alliance*, 6.
Luco, R. F., Q. Pan, K. Tominaga, B. J. Blencowe, O. M. Pereira-Smith, and T. Misteli. 2010. 'Regulation of alternative splicing by histone modifications', *Science*, 327: 996-1000.
Lukauskas, S., A. Tvardovskiy, N. V. Nguyen, M. Stadler, P. Faull, T. Ravnsborg, B. Ozdemir Aygenli, S. Dornauer, H. Flynn, R. G. H. Lindeboom, T. K. Barth, K. Brockers, S. M. Hauck, M. Vermeulen, A. P. Snijders, C. L. Muller, P. A. DiMaggio, O. N. Jensen, R. Schneider, and T. Bartke. 2024. 'Decoding chromatin states by proteomic profiling of nucleosome readers', *Nature*, 627: 671-79.
Lun, A. T., and G. K. Smyth. 2016. 'csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows', *Nucleic Acids Res*, 44: e45.
Markert, J. W., J. H. Soffers, and L. Farnung. 2025. 'Structural basis of H3K36 trimethylation by SETD2 during chromatin transcription', *Science*, 387: 528-33.
Marmorstein, R., and M. M. Zhou. 2014. 'Writers and readers of histone acetylation: structure, mechanism, and inhibition', *Cold Spring Harb Perspect Biol*, 6: a018762.
Martin, B. J., K. L. McBurney, V. E. Maltby, K. N. Jensen, J. Brind'Amour, and L. J. Howe. 2017. 'Histone H3K4 and H3K36 Methylation Independently Recruit the NuA3 Histone Acetyltransferase in Saccharomyces cerevisiae', *Genetics*, 205: 1113-23.
Martire, S., and L. A. Banaszynski. 2020. 'The roles of histone variants in fine-tuning chromatin organization and function', *Nat Rev Mol Cell Biol*, 21: 522-41.
McCauley, B. S., L. Sun, R. Yu, M. Lee, H. Liu, D. S. Leeman, Y. Huang, A. E. Webb, and W. Dang. 2021. 'Altered Chromatin States Drive Cryptic Transcription in Aging Mammalian Stem Cells', *Nat Aging*, 1: 684-97.
McKay, D. J., S. Klusza, T. J. Penke, M. P. Meers, K. P. Curry, S. L. McDaniel, P. Y. Malek, S. W. Cooper, D. C. Tatomer, J. D. Lieb, B. D. Strahl, R. J. Duronio, and A. G. Matera. 2015. 'Interrogating the function of metazoan histones using engineered gene clusters', *Dev Cell*, 32: 373-86.
Meers, M. P., T. Henriques, C. A. Lavender, D. J. McKay, B. D. Strahl, R. J. Duronio, K. Adelman, and A. G. Matera. 2017. 'Histone gene replacement reveals a post-transcriptional role for H3K36 in maintaining metazoan transcriptome fidelity', *Elife*, 6.
Mirny, L. A. 2011. 'The fractal globule as a model of chromatin architecture in the cell', *Chromosome Res*, 19: 37-51.
Mondal, A., C. Felipe, and A. B. Kolomeisky. 2023. 'Nucleosome Breathing Facilitates the Search for Hidden DNA Sites by Pioneer Transcription Factors', *J Phys Chem Lett*, 14: 4096-103.
Morales, V., C. Regnard, A. Izzo, I. Vetter, and P. B. Becker. 2005. 'The MRG domain mediates the functional integration of MSL3 into the dosage compensation complex', *Mol Cell Biol*, 25: 5947-54.
Morris, S. A., B. Rao, B. A. Garcia, S. B. Hake, R. L. Diaz, J. Shabanowitz, D. F. Hunt, C. D. Allis, J. D. Lieb, and B. D. Strahl. 2007. 'Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification', *J Biol Chem*, 282: 7632-40.
Muller, M., T. Schauer, S. Krause, R. Villa, A. W. Thomae, and P. B. Becker. 2020. 'Two-step mechanism for selective incorporation of lncRNA into a chromatin modifier', *Nucleic Acids Res*, 48: 7483-501.
Nicetto, D., G. Donahue, T. Jain, T. Peng, S. Sidoli, L. Sheng, T. Montavon, J. S. Becker, J. M. Grindheim, K. Blahnik, B. A. Garcia, K. Tan, R. Bonasio, T. Jenuwein, and K. S. Zaret.

2019. 'H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification', *Science*, 363: 294-97.
Ohno, M., T. Ando, D. G. Priest, V. Kumar, Y. Yoshida, and Y. Taniguchi. 2019. 'Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs', *Cell*, 176: 520-34 e25.
Pai, C. C., R. S. Deegan, L. Subramanian, C. Gal, S. Sarkar, E. J. Blaikley, C. Walker, L. Hulme, E. Bernhard, S. Codlin, J. Bahler, R. Allshire, S. Whitehall, and T. C. Humphrey. 2014. 'A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice', *Nat Commun*, 5: 4091.
Pal, K., M. Forcato, D. Jost, T. Sexton, C. Vaillant, E. Salviato, E. M. C. Mazza, E. Lugli, G. Cavalli, and F. Ferrari. 2019. 'Global chromatin conformation differences in the Drosophila dosage compensated chromosome X', *Nat Commun*, 10: 5355.
Park, I. Y., R. T. Powell, D. N. Tripathi, R. Dere, T. H. Ho, T. L. Blasius, Y. C. Chiang, I. J. Davis, C. C. Fahey, K. E. Hacker, K. J. Verhey, M. T. Bedford, E. Jonasch, W. K. Rathmell, and C. L. Walker. 2016. 'Dual Chromatin and Cytoskeletal Remodeling by SETD2', *Cell*, 166: 950-62.
Patnaik, D., H. G. Chin, P. O. Esteve, J. Benner, S. E. Jacobsen, and S. Pradhan. 2004. 'Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase', *J Biol Chem*, 279: 53248-58.
Pinheiro, I., R. Margueron, N. Shukeir, M. Eisold, C. Fritzsch, F. M. Richter, G. Mittler, C. Genoud, S. Goyama, M. Kurokawa, J. Son, D. Reinberg, M. Lachner, and T. Jenuwein. 2012. 'Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity', *Cell*, 150: 948-60.
Pirrotta, V., and H. B. Li. 2012. 'A view of nuclear Polycomb bodies', *Curr Opin Genet Dev*, 22: 101-9.
Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young. 2005. 'Genome-wide map of nucleosome acetylation and methylation in yeast', *Cell*, 122: 517-27.
Policarpi, C., M. Munafo, S. Tsagkris, V. Carlini, and J. A. Hackett. 2024. 'Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications', *Nat Genet*, 56: 1168-80.
Prayitno, Khairunnadiya. 2019. 'The role of roX RNA in Dosage Compensation during Drosophila melanogaster embryogenesis', Ludwig Maximilians University of Munich.
Prestel, M., C. Feller, T. Straub, H. Mitlohner, and P. B. Becker. 2010. 'The activation potential of MOF is constrained for dosage compensation', *Mol Cell*, 38: 815-26.
Quinlan, A. R. 2014. 'BEDTools: The Swiss-Army Tool for Genome Feature Analysis', *Curr Protoc Bioinformatics*, 47: 11 12 1-34.
RHReynolds. 2022. 'Reduce GO Terms by Semantic Similarity Indexing'. https://github.com/RHReynolds/rutils.
Robinson, J. T., H. Thorvaldsdottir, D. Turner, and J. P. Mesirov. 2023. 'igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV)', *Bioinformatics*, 39.
Roy, S., J. Ernst, P. V. Kharchenko, P. Kheradpour, N. Negre, M. L. Eaton, J. M. Landolin, C. A. Bristow, L. Ma, M. F. Lin, S. Washietl, B. I. Arshinoff, F. Ay, P. E. Meyer, N. Robine, N. L. Washington, L. Di Stefano, E. Berezikov, C. D. Brown, R. Candeias, J. W. Carlson, A. Carr, I. Jungreis, D. Marbach, R. Sealfon, M. Y. Tolstorukov, S. Will, A. A. Alekseyenko, C. Artieri, B. W. Booth, A. N. Brooks, Q. Dai, C. A. Davis, M. O. Duff, X. Feng, A. A. Gorchakov, T. Gu, J. G. Henikoff, P. Kapranov, R. Li, H. K. MacAlpine, J. Malone, A. Minoda, J. Nordman, K. Okamura, M. Perry, S. K. Powell, N. C. Riddle, A. Sakai, A. Samsonova, J. E. Sandler, Y. B. Schwartz, N. Sher, R. Spokony, D. Sturgill, M. van Baren, K. H. Wan, L. Yang, C. Yu, E. Feingold, P. Good, M. Guyer, R. Lowdon, K. Ahmad, J. Andrews, B. Berger, S. E. Brenner, M. R. Brent, L. Cherbas, S. C. Elgin, T. R. Gingeras,

R. Grossman, R. A. Hoskins, T. C. Kaufman, W. Kent, M. I. Kuroda, T. Orr-Weaver, N. Perrimon, V. Pirrotta, J. W. Posakony, B. Ren, S. Russell, P. Cherbas, B. R. Graveley, S. Lewis, G. Micklem, B. Oliver, P. J. Park, S. E. Celniker, S. Henikoff, G. H. Karpen, E. C. Lai, D. M. MacAlpine, L. D. Stein, K. P. White, and M. Kellis. 2010. 'Identification of functional elements and regulatory circuits by Drosophila modENCODE', *Science*, 330: 1787-97.

Saha, P., D. T. Sowpati, M. Soujanya, I. Srivastava, and R. K. Mishra. 2020. 'Interplay of pericentromeric genome organization and chromatin landscape regulates the expression of Drosophila melanogaster heterochromatic genes', *Epigenetics Chromatin*, 13: 41.

Sakai, A., B. E. Schwartz, S. Goldstein, and K. Ahmad. 2009. 'Transcriptional and developmental functions of the H3.3 histone variant in Drosophila', *Curr Biol*, 19: 1816-20.

Salzler, H. R., V. Vandadi, B. D. McMichael, J. C. Brown, S. A. Boerma, M. P. Leatham-Jensen, K. M. Adams, M. P. Meers, J. M. Simon, R. J. Duronio, D. J. McKay, and A. G. Matera. 2023. 'Distinct roles for canonical and variant histone H3 lysine-36 in Polycomb silencing', *Sci Adv*, 9: eadf2451.

Salzler, H. R., V. Vandadi, J. R. Sallean, and A. G. Matera. 2024. 'Set2 and H3K36 regulate the Drosophila male X chromosome in a context-specific manner, independent from MSL complex spreading', *Genetics*, 228.

Sankaran, S. M., A. W. Wilkinson, J. E. Elias, and O. Gozani. 2016. 'A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin', *J Biol Chem*, 291: 8465-74.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides. 2002. 'Active genes are tri-methylated at K4 of histone H3', *Nature*, 419: 407-11.

Schauer, T., Y. Ghavi-Helm, T. Sexton, C. Albig, C. Regnard, G. Cavalli, E. E. Furlong, and P. B. Becker. 2017. 'Chromosome topology guides the Drosophila Dosage Compensation Complex for target gene activation', *EMBO Rep*, 18: 1854-68.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. 2012. 'Fiji: an open-source platform for biological-image analysis', *Nat Methods*, 9: 676-82.

Schmahling, S., A. Meiler, Y. Lee, A. Mohammed, K. Finkl, K. Tauscher, L. Israel, M. Wirth, J. Philippou-Massier, H. Blum, B. Habermann, A. Imhof, J. J. Song, and J. Muller. 2018. 'Regulation and function of H3K36 di-methylation by the trithorax-group protein complex AMC', *Development*, 145.

Schmitges, F. W., A. B. Prusty, M. Faty, A. Stutzer, G. M. Lingaraju, J. Aiwazian, R. Sack, D. Hess, L. Li, S. Zhou, R. D. Bunker, U. Wirth, T. Bouwmeester, A. Bauer, N. Ly-Hartig, K. Zhao, H. Chan, J. Gu, H. Gut, W. Fischle, J. Muller, and N. H. Thoma. 2011. 'Histone methylation by PRC2 is inhibited by active chromatin marks', *Mol Cell*, 42: 330-41.

Seervai, R. N. H., R. K. Jangid, M. Karki, D. N. Tripathi, S. Y. Jung, S. E. Kearns, K. J. Verhey, M. A. Cianfrocco, B. A. Millis, M. J. Tyska, F. M. Mason, W. K. Rathmell, I. Y. Park, R. Dere, and C. L. Walker. 2020. 'The Huntingtin-interacting protein SETD2/HYPB is an actin lysine methyltransferase', *Sci Adv*, 6.

Sessa, A., L. Fagnocchi, G. Mastrototaro, L. Massimino, M. Zaghi, M. Indrigo, S. Cattaneo, D. Martini, C. Gabellini, C. Pucci, A. Fasciani, R. Belli, S. Taverna, M. Andreazzoli, A. Zippo, and V. Broccoli. 2019. 'SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring', *Neuron*, 104: 271-89 e13.

Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, and Y. Shi. 2004. 'Histone demethylation mediated by the nuclear amine oxidase homolog LSD1', *Cell*, 119: 941-53.

Shipman, G. A., R. Padilla, C. Horth, B. Hu, E. Bareke, F. N. Vitorino, J. M. Gongora, B. A. Garcia, C. Lu, and J. Majewski. 2024. 'Systematic perturbations of SETD2, NSD1, NSD2, NSD3, and ASH1L reveal their distinct contributions to H3K36 methylation', *Genome Biol*, 25: 263.

Soshnev, A. A., S. Z. Josefowicz, and C. D. Allis. 2016. 'Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome', *Mol Cell*, 62: 681-94.

Stirling, D. R., M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, and A. Goodman. 2021. 'CellProfiler 4: improvements in speed, utility and usability', *BMC Bioinformatics*, 22: 433.

Straub, T., A. Zabel, G. D. Gilfillan, C. Feller, and P. B. Becker. 2013. 'Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq', *Genome Res*, 23: 473-85.

Strom, A. R., A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen. 2017. 'Phase separation drives heterochromatin domain formation', *Nature*, 547: 241-45.

Stulemeijer, I. J., D. De Vos, K. van Harten, O. K. Joshi, O. Blomberg, T. van Welsem, M. Terweij, H. Vlaming, E. L. de Graaf, A. F. Altelaar, B. M. Bakker, and F. van Leeuwen. 2015. 'Dot1 histone methyltransferases share a distributive mechanism but have highly diverged catalytic properties', *Sci Rep*, 5: 9824.

Sun, Z., Y. Lin, M. T. Islam, R. Koche, L. Hedehus, D. Liu, C. Huang, T. Vierbuchen, C. L. Sawyers, and K. Helin. 2023. 'Chromatin regulation of transcriptional enhancers and cell fate by the Sotos syndrome gene NSD1', *Mol Cell*, 83: 2398-416 e12.

Sural, T. H., S. Peng, B. Li, J. L. Workman, P. J. Park, and M. I. Kuroda. 2008. 'The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome', *Nat Struct Mol Biol*, 15: 1318-25.

Tanaka, Y., Z. Katagiri, K. Kawahashi, D. Kioussis, and S. Kitajima. 2007. 'Trithorax-group protein ASH1 methylates histone H3 lysine 36', *Gene*, 397: 161-8.

Teslenko A., Fierz B. . 2024. 'Single-molecule analysis reveals the mechanism of chromatin ubiquitylation by variant PRC1 complexes', *bioRxiv*.

Tschiersch, B., A. Hofmann, V. Krauss, R. Dorn, G. Korge, and G. Reuter. 1994. 'The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes', *EMBO J*, 13: 3822-31.

Vale-Silva, L. A., T. E. Markowitz, and A. Hochwagen. 2019. 'SNP-ChIP: a versatile and tag-free method to quantify changes in protein binding across the genome', *BMC Genomics*, 20: 54.

Valencia-Sanchez, M. I., P. De Ioannes, M. Wang, D. M. Truong, R. Lee, J. P. Armache, J. D. Boeke, and K. J. Armache. 2021. 'Regulation of the Dot1 histone H3K79 methyltransferase by histone H4K16 acetylation', *Science*, 371.

Valsecchi, C. I. K., M. F. Basilicata, P. Georgiev, A. Gaub, J. Seyfferth, T. Kulkarni, A. Panhale, G. Semplicio, V. Manjunath, H. Holz, P. Dasmeh, and A. Akhtar. 2021. 'RNA nucleation by MSL2 induces selective X chromosome compartmentalization', *Nature*, 589: 137-42.

Villa, R., P. K. A. Jagtap, A. W. Thomae, A. Campos Sparr, I. Forne, J. Hennig, T. Straub, and P. B. Becker. 2021. 'Divergent evolution toward sex chromosome-specific gene regulation in Drosophila', *Genes Dev*, 35: 1055-70.

Weinberg, D. N., S. Papillon-Cavanagh, H. Chen, Y. Yue, X. Chen, K. N. Rajagopalan, C. Horth, J. T. McGuire, X. Xu, H. Nikbakht, A. E. Lemiesz, D. M. Marchione, M. R. Marunde, M. J. Meiners, M. A. Cheek, M. C. Keogh, E. Bareke, A. Djedid, A. S. Harutyunyan, N. Jabado, B. A. Garcia, H. Li, C. D. Allis, J. Majewski, and C. Lu. 2019. 'The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape', *Nature*, 573: 281-86.

Wen, H., Y. Li, Y. Xi, S. Jiang, S. Stratton, D. Peng, K. Tanaka, Y. Ren, Z. Xia, J. Wu, B. Li, M. C. Barton, W. Li, H. Li, and X. Shi. 2014. 'ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression', *Nature*, 508: 263-8.

Wong, L. H., and D. J. Tremethick. 2025. 'Multifunctional histone variants in genome function', *Nat Rev Genet*, 26: 82-104.

Xia, B., D. Zhao, G. Wang, M. Zhang, J. Lv, A. S. Tomoiaga, Y. Li, X. Wang, S. Meng, J. P. Cooke, Q. Cao, L. Zhang, and K. Chen. 2020. 'Machine learning uncovers cell identity regulator by histone code', *Nat Commun*, 11: 2696.

Ye, C., B. M. Sutter, Y. Wang, Z. Kuang, and B. P. Tu. 2017. 'A Metabolic Function for Phospholipid and Histone Methylation', *Mol Cell*, 66: 180-93 e8.

Ye, C., and B. P. Tu. 2018. 'Sink into the Epigenome: Histones as Repositories That Influence Cellular Metabolism', *Trends Endocrinol Metab*, 29: 626-37.

Zabidi, M. A., C. D. Arnold, K. Schernhuber, M. Pagani, M. Rath, O. Frank, and A. Stark. 2015. 'Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation', *Nature*, 518: 556-9.

Zhang, R., J. Erler, and J. Langowski. 2017. 'Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction', *Biophys J*, 112: 450-59.

Zhang, X., Z. Yang, S. I. Khan, J. R. Horton, H. Tamaru, E. U. Selker, and X. Cheng. 2003. 'Structural basis for the product specificity of histone lysine methyltransferases', *Mol Cell*, 12: 177-85.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. 2008. 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biol*, 9: R137.

Zheng, Y., F. N. Hsu, W. Xu, X. J. Xie, X. Ren, X. Gao, J. Q. Ni, and J. Y. Ji. 2014. 'A developmental genetic analysis of the lysine demethylase KDM2 mutations in Drosophila melanogaster', *Mech Dev*, 133: 36-53.