

RESEARCH ARTICLE



Emergent properties of the lysine methylome reveal regulatory roles via protein interactions and histone mimicry

Gareth Pollin ^{a,b}, Young-In Chi^{a,b}, Angela J. Mathison^{a,b}, Michael T. Zimmermann^{a,c,d}, Gwen Lombark ^{a,b,e} and Raul Urrutia ^{a,b,c}

^aLinda T. and John A. Mellows Center for Genomic Sciences and Precision Medicine (Mellows Center), Medical College of Wisconsin, Milwaukee, WI, USA; ^bDivision of Research, Department of Surgery, Medical College of Wisconsin, Milwaukee, WI, USA; ^cDepartment of Biochemistry, Medical College of Wisconsin, Milwaukee, WI, USA; ^dClinical and Translational Sciences Institute, Medical College of Wisconsin, Milwaukee, WI, USA; ^eDepartment of Pharmacology and Toxicology, Medical College of Wisconsin, Milwaukee, WI, USA

ABSTRACT

Aims: Epigenomics has significantly advanced through the incorporation of Systems Biology approaches. This study aims to investigate the human lysine methylome as a system, using a data-science approach to reveal its emergent properties, particularly focusing on histone mimicry and the broader implications of lysine methylation across the proteome.

Methods: We employed a data-science-driven OMICS approach, leveraging high-dimensional proteomic data to study the lysine methylome. The analysis focused on identifying sequence-based recognition motifs of lysine methyltransferases and evaluating the prevalence and distribution of lysine methylation across the human proteome.

Results: Our analysis revealed that lysine methylation impacts 15% of the known proteome, with a notable bias toward mono-methylation. We identified sequence-based recognition motifs of 13 lysine methyltransferases, highlighting candidates for histone mimicry. These findings suggest that the selective inhibition of individual lysine methyltransferases could have systemic effects rather than merely targeting histone methylation.

Conclusions: The lysine methylome has significant mechanistic value and should be considered in the design and testing of therapeutic strategies, particularly in precision oncology. The study underscores the importance of considering non-histone proteins involved in DNA damage and repair, cell signaling, metabolism, and cell cycle pathways when targeting lysine methyltransferases.

ARTICLE HISTORY

Received 28 June 2024

Accepted 25 November 2024

KEYWORDS

Lysine methylome; histone mimicry; post-translational modification; methylation; molecular dynamics; epigenetics; methyltransferases; BRD4

1. Introduction

Emergent properties in biological systems refer to the novel and complex behaviors or functions that arise from the interactions of simpler components or processes. Epigenetic modifications are an example of emergent properties that control gene regulation and nuclear organization through molecular processes other than directly modifying the DNA code, mediated by epigenomic regulators [1]. These epigenetic molecular processes play essential roles in diverse cellular outcomes such as DNA replication [2], cellular differentiation [3], development [4], and diseases such as cancer [5]. The coordinated action of these epigenetic components exemplifies how complex regulatory networks and emergent properties can arise from the interactions of simpler molecular processes. Notably, at the core of the mechanisms underlying the epigenetic code are the types and combinations of post-translational modifications (PTMs) in histones and nucleic acids, known as marks [6].

Methylation of histones can occur on both arginine and lysine residues, with both able to exist in mono- or dimethylated states, while lysine can also be tri-methylated

[7,8]. Most lysine histone methyltransferases belong to the SET-domain-containing protein family, except for DOT1L [9,10]. The effect of lysine residue methylation depends on its position and methylation state. For example, methylation at H3K4, H3K36, and H3K79 is generally associated with gene activation, while methylation at H3K9, H3K27, and H4K20 is often repressive [11]. Histones are regulated through different combinations of marks that alter the post-translational modifications (PTMs) at various residues to control gene expression. For instance, the phosphorylation of H3S10 fluctuates during the cell cycle and is required for faithful chromosome dynamics and transcription [12,13]. Phosphorylation at H3S10 prevents the addition of repressive methylation marks to the adjacent H3K9 by inhibiting the binding of euchromatic histone lysine methyltransferase (EHMT) writers and methylation readers such as chromobox homologs (CBX) [14]. Notably, misregulation of histone methyltransferases can contribute to various cancers. For instance, chromosomal translocation of the MLL1 gene, an H3K4 methyltransferase, is observed in around 70% of infant leukemia patients, while overexpression of EHMT2 (H3K9) and EZH2

CONTACT Raul Urrutia  rurrutia@mcw.edu; Gwen Lombark  glombark@mcw.edu  Linda T. and John A. Mellows Center for Genomic Sciences and Precision Medicine (Mellows Center), Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, WI 53226, USA

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/17501911.2024.2435244>

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Article highlights

- Systematic mapping and analysis of the current human lysine methylome reveals its widespread impact on cellular regulation, with methylation present on 15% of the human proteome, affecting critical cellular pathways.
- Exploring the interactome formed by 13 distinct histone methyltransferases reveals emerging properties of these molecular systems and highlights how KMTs might interact with components of larger and more complex biological systems.
- Exploring histone mimicry candidates implicates protein-protein regulatory mechanisms by lysine methyltransferases and highlights how histone mimicry contributes to complex protein-protein interactions regulated by lysine methyltransferases.
- Uncovering the role of lysine methyltransferases in cancer and cell essentiality through histone mimicry implicates the regulation of key proteins and pathways that are important for cell survival and cancer development.
- Molecular dynamics simulation characterizes how methylation by histone mimicry impacts the structure and biomolecular function of methylated targets and demonstrates how BRD4 methylation can disrupt the reading function of the bromodomain through a potential allosteric regulation site.

(H3K27) has been detected in multiple cancer types [15,16]. Targeting lysine histone methyltransferases for cancer therapy has emerged as a promising approach due to their physiological significance [17]. However, some lysine methyltransferases methylate numerous non-histone proteins in a process referred to as ‘histone mimicry’ [18]. Histone mimicry is a broad concept often focused on pathogen-related histone mimicry, where pathogens exploit host cellular machinery by mimicking histone-like sequences to disrupt normal cellular processes by capturing the histone-modifying enzyme [19–23]. However, in this study, we focus on endogenous histone mimics (hereafter referred to as just histone-mimics), which are non-histone proteins that contain sequences resembling histone motifs. These endogenous histone mimics interact with histone-modifying enzymes, leading to changes in protein-protein interactions, protein stability, and the functionality of the proteins containing these mimic sites [24]. Given the ancient origins and high conservation of histones across eukaryotes [25], and the well-established nature of histone-based chromatin regulation [26], it is reasonable to infer that endogenous histone mimics in non-histone proteins emerged later in evolutionary history. These mimics likely evolved as an adaptive mechanism, allowing non-histone proteins to utilize the already established epigenetic machinery [27]. Endogenous histone mimicry introduces a simple regulatory mechanism that has significant impacts on downstream key pathways and processes, altering the overall biological system [28]. This phenomenon represents an emergent property where the modification of a few residues can drive substantial changes in cellular functions and pathways. Given that histone mimicry on endogenous proteins remains largely understudied in lysine methylation, further investigation is essential to fully understand its implications and to optimize the use of inhibitors targeting histone-modifying enzymes in cancer and other diseases [27]. While non-histone substrates have been rigorously explored for other PTMs such as ubiquitination, phosphorylation, and arginine methylation, investigations into lysine methylation of non-histone substrates are still relatively understudied. However,

there has been a growing body of research on this topic recently [29].

Lysine methylation has been shown to play important roles in protein regulation. For example, P53 has its localization restricted to the nucleus, and its binding decreased to only target gene promoters when methylated by SET7/9 [30] and SMYD2 [31], respectively. Additionally, an increasing body of evidence suggests the involvement of non-histone lysine methylation in the regulation of cancer-related proteins such as RB1, VEGFR1, and PARP1 [32–34]. This broad spectrum of lysine methylation across various proteins is collectively known as the lysine methylome. The lysine methylome encompasses the complete set of lysine methylation marks across the human proteome. This intricate interplay between lysine methylation and other cellular components highlights the emergence of novel regulatory mechanisms, illustrating the lysine methylome’s role in the emergence of complex properties in biological systems. Despite the increase of literature surrounding lysine methylation in regulating key cancer and rare disease-related proteins, the lysine methylome is still substantially under-studied with many substrates and their respective function left unverified. Recently, we reported that the H3K9me2 writers, EHMT1 and EHMT2, form distinct protein-protein networks during the cell cycle and hypothesized targets for histone mimicry [18]. The current study aims to comprehensively explore the lysine methylome as a molecular system and derive emergent properties from this analysis. Through our investigations, we systematically provide data on lysine methyltransferases and potential targets of histone mimicry based on substrate specificity and known interactions. This study underscores the value of studying writers, readers, and erasers of the histone code at a systems biology level. Lastly, due to the rapid emergence of drugs targeting these epigenomic regulators, the widespread mechanism of histone mimicry suggests the potential for additional effects, beyond the regulation of the histone code. Thus, this new knowledge bears a significant impact for better understanding of epigenetic mechanisms and potential implications for pharmacotherapeutics.

2. Methods

2.1. Assembly and analyses of the lysine methylome

Methylation data were obtained from large-scale proteomics experiments according to the Phosphosite database (www.phosphosite.org/), which is a repository for various types of post-translational modifications [35]. Initially, the dataset was harmonized to include only human lysine modifications. Subsequently, the data was filtered to lysine mono-methylation, di-methylation, and tri-methylation. This process resulted in 5484 modifications suitable for further analysis. This data was subsequently used to generate heatmaps using the package ComplexHeatmap [36]. Gene ontology enrichment analysis was conducted using EnrichR [37] with the Pathways enrichment database MSigDB Hallmark 2020 [38] and CORUM [39]. Enrichment results were exported to tables for subsequent analysis. The enrichment tables were imported into R for further

processing. Visualization of the results was achieved by generating plots using the ggplot2 package [40]. All lysine methylation data can be downloaded from <https://www.phosphosite.org/staticDownload> by selecting the corresponding data file and the exact version of the data used in this study can be found in table S1.

2.2. Gene Networks

Candidate proteins were inputted into Cytoscape (v3.9.1) [41] using the String database [42] query to retrieve protein-protein physical interactions. String database edges (interactions) for the lysine methylome or histone mimicry networks were included in the network if they had a high confidence score above 0.7 [42,43]. The histone mimicry candidate network included manually curated BioGRID interactions between the lysine methyltransferase and respective candidates as specified in the figure legends. Nodes (proteins) are color-coded based on some criteria of interest, as specified in the figure legends. Nodes without interactions (singletons) were removed.

2.3. Motif Histone-Mimicry Analysis

A literature search was carried out for previous substrate analysis for the human methyltransferases as well as previously identified histone and non-histone targets to curate a list of known short linear motifs (SLiMs) [44–53]. All the sequences collected were then entered through MEME (Multiple Em for Motif Elicitation) to discover recurring, fixed-length patterns that have the ability to be methylated [54]. The pattern was then used as the sequence pattern through Scansite 4.0 to identify all potential proteins containing the SLiMs [55]. These proteins were then cross-referenced from known interactors from the bioGRID database to identify potential candidates for histone mimicry [56]. Interactome data can be downloaded from <https://thebiogrid.org/by> searching the protein of interest (I.E EHMT1) and selecting download curated data for this protein then selecting a suitable format for the file.

2.4. Gene essentiality and functionality analysis

We obtained gene dependency scores for 17,386 genes across 1,086 cancer cell lines from the DepMap data portal (<https://depmap.org/portal/>). These scores were generated using high-throughput CRISPR screening techniques [57]. Essentiality was defined as a dependency score < -1 in the corresponding cell lines. Specifically, DepMap data can be downloaded from https://depmap.org/portal/data_page/?tab=allData by selecting the file CRISPRGeneDependency.csv. We then used R-studio to visualize essentiality highlighting our refined set of stringent candidates for histone mimicry. The analysis center in the Genomic Data Commons data portal (GDC, <https://portal.gdc.cancer.gov/>) was used to identify which of our defined histone mimicry candidates have been denoted as cancer census genes. The GDC cancer projects TCGA-BRCA, TCGA-ACC and TCGA-UCEC were used to examine copy

number variance and mutation prevalence in CDC73, BRD4 and NACA [58].

2.5. Protein stability and structural perturbation measurement of the bromodomain-containing protein 4 (BRD4) and bromodomain 2 (BD2) domain in complex with an acetylated lysine-containing peptide

The previously determined NMR structure of the human BRD4-BD2 (333–460) in complex with a peptide encompassing the human FOXO3A acetyl lysine-containing region (237–252) was used as a prototype (PDB code 6MNL). For modified structure generation, methyl groups were added to both K404 and K406 residues based on the known standard geometry of the either mono- or di-methylated lysines. We assessed the impact of modifications on protein stability by comparing folding energy ($\Delta\Delta G_{\text{fold}}$) using FoldX [59]. The energy-minimized structures were calculated after introducing each set of modifications to the native structure. Subsequently, we assessed structure perturbations at the global and local levels by measuring the positional displacement of backbone atoms between the native and modified proteins (global) or only the atoms near the modified residues (local). For local structure perturbation, residues within a 10 Å radius of the modification sites were identified from the energy-minimized structures using PyMol (Molecular Graphics System, Schrödinger, LLC). The root mean square deviation (RMSD) of the backbone atoms between the wild-type and mutant proteins was then calculated using Coot [60]. For global structure perturbation, entire backbone atoms were used for RMSD calculation between the structures.

2.6. Molecular dynamics (MD) simulation, atom displacements, and time-dependent interaction energy calculations

MD simulations were performed using the CHARMM36 all-atom force field [61] implemented in the Discovery Studio suite version 21.1 (Dassault Systèmes BIOVIA) with a 2 fs time step. Molecular models were simulated using a simplified implicit solvent model with a distance-dependent dielectric constant of 80 and a pH of 7.4. Energy minimization was performed for 5,000 steps with the steepest descent algorithm, followed by an additional 5,000 steps of conjugate gradient minimization to relax the protein structure obtained from the stressed crystal environment. Each system of 10 replicates of the model was independently heated to 300 K over 200 ps and equilibrated for 500 ps followed by a 10 ns production simulation under the NPT (constant Number of atoms, Pressure, and Temperature) ensemble by changing the initial seed (100 ns total). Structures during the unconstrained dynamics simulation were recorded every 10 ps to give a total of 1000 frames for analysis. The modified proteins were subjected to the same energy minimization and MD simulation protocol as the native protein. Trajectories were aligned to the initial wild-type conformation before analysis. RMSD and RMSF values for all atoms at the residue level were calculated using Discovery Studio tools and Microsoft Excel algorithms. Additional analyses were performed in R, utilizing the bio3d

package [62]. Molecular visualizations were generated using PyMOL. The free energy of molecular interaction was measured using the protocol implemented in Discovery Studio. This was done using the MD simulation trajectories and by selecting the protein and the interaction groups of interest, either acetylated lysine K242 or the entire peptide. Non-bonded interactions were monitored and dynamic interaction energies (van der Waals and electrostatic energies) for each replicate were calculated from the MD trajectories using the CHARMM36 force field and the implicit distance-dependent dielectric solvent model and averaged. These measurements were made for all 10 replicates and averaged for comparison with the native protein.

3. Results

3.1. Systematic mapping and analysis of the current human lysine methylome reveals its widespread impact in cellular regulation

In this study, we systematically mapped the distribution of lysine methylation marks across various proteins to develop the lysine methylome and investigate the potential role of histone mimicry, along with its implications for the regulation of cellular function. We generated an up-to-date lysine methylome using high-throughput mass spectrometry datasets that include data from a mixture of cell types and tissues, covering both normal and cancerous contexts. This approach enables a broad analysis of lysine methylation across different human biological systems [35]. This meta-analysis revealed a strong bias toward mono-methylation, indicated by the abundance of mono-methylated proteins, which comprised 79% of the lysine methylome. In addition, groups of proteins marked with mono-, di-, and tri-methylation were largely distinct to one of the groups (Figure 1(a)). In total, we found 2,769 unique proteins in humans that have been previously identified to be marked with lysine methylation (Table S1). According to the neXtProt Human Proteome Project (HPP) database 18,397 proteins are currently reported, suggesting that approximately 15% of the known proteome contains evidence of lysine methylation [63]. In stark contrast 15,894 unique proteins have been marked by phosphorylation at a serine residue, marking an estimated 86% of the current proteome. A further 13,490 (73%) proteins are phosphorylated at a threonine, and 11,849 (64%) proteins are marked with tyrosine phosphorylation. Additionally, we identified 98,019 ubiquitination sites on lysine residues across 11,793 (64%) proteins [35]. The stark contrast between phosphorylation and ubiquitination compared to lysine methylation highlights the current limitations of available studies for methylation-derived datasets. A key limitation is the enrichment step, where affinity reagents often exhibit bias for specific residues surrounding the methylated lysine, affecting the comprehensiveness of the data. While such approaches have been effectively applied to other modifications like phosphorylation, acetylation, and arginine methylation, there is a notable scarcity of antibodies targeting specific types of lysine methylation, such as mono-, di-, and tri-methylation. While there have been some recent advances in the development of targeted

antibodies and alternative enrichment methods, substantial progress is still needed to overcome these challenges [64–67]. Of these 2,769 proteins, we found 1,073 (39%) mapped with nuclear localization according to the cellular component gene ontology Nucleus (GO:0005634), highlighting the significant enrichment of lysine methylation within the nucleus. This high percentage of enrichment may suggest that lysine methylation plays a critical role in regulating nuclear processes. Given the importance of these processes in both normal cellular function and disease, particularly cancer, we must fully understand the lysine methylome's emergent properties to uncover its broader implications in nuclear regulation. To investigate the emergent properties of proteins that undergo lysine methylation, we performed a MsigDB enrichment analysis, revealing novel diverse biological functions of the lysine methylome, including DNA damage and repair, signaling pathways, metabolism, and the cell cycle (Figure 1(b)). Specifically, mono-methylation is enriched in all four groups reflecting the abundance observed in Figure 1(a). In contrast, di-methylation is enriched in DNA damage and repair, and cell cycle, while trimethylation is enriched in DNA damage and repair, cell cycle, and metabolism (Figure 1(b)). Performing a network analysis, we found 213 proteins with multiple lysine methylation marks revealing a moderately interconnected structure with 577 edges, indicating a network where the interaction of methylation marks on different proteins creates a complex regulatory web. This web likely drives the emergent properties of the lysine methylome by integrating individual signals to coordinate the regulation of complex molecular systems, that are essential for maintaining cellular homeostasis (Figure 1(c)). Of these 213 proteins, we found that 108 contained both Kme1 and Kme2, 59 between Kme1 and Kme3, 11 for Kme2 and Kme3, and finally, 35 proteins that contained all 3 variations of lysine methylation throughout the protein. From the current lysine methylome collection, we discovered that several proteins exhibit evidence of a methylation-regulated 'molecular switch' with 2 or all 3 lysine methylation states observed on a single residue. Notably, we observed 21/108 (17%) proteins within the Kme1/Kme2 group containing both variations of the lysine modification at the same residue. Furthermore, we found 6/59 (10%) for the Kme1/Kme3 group and 3/11 (27%) for the Kme2/Kme3 group. As for the proteins that contained all 3 variations of the lysine methylation modification, we found that 25/35 (71%) contained either 2 or 3 of the lysine modification marks at a single residue. In total, we identified 55 proteins with two or more methylation states at the same lysine residue. These subsets of proteins, much like histones, may operate under a dynamic regulatory mechanism where their function is dependent on the specific methylation state present on the lysine residue. For example, lysine methylation on histone H3 at lysine 9 (H3K9) illustrates this versatility: mono-methylation of H3K9 is linked to transcriptionally active euchromatin, while di- and tri-methylation lead to gene repression in heterochromatin regions [11]. Similarly, these non-histone proteins marked by different methylation states may undergo functional changes based on the degree of methylation, allowing for fine-tuned regulation of cellular processes. This finding reinforces the concept of lysine methylation acting as a versatile switch for regulating diverse protein

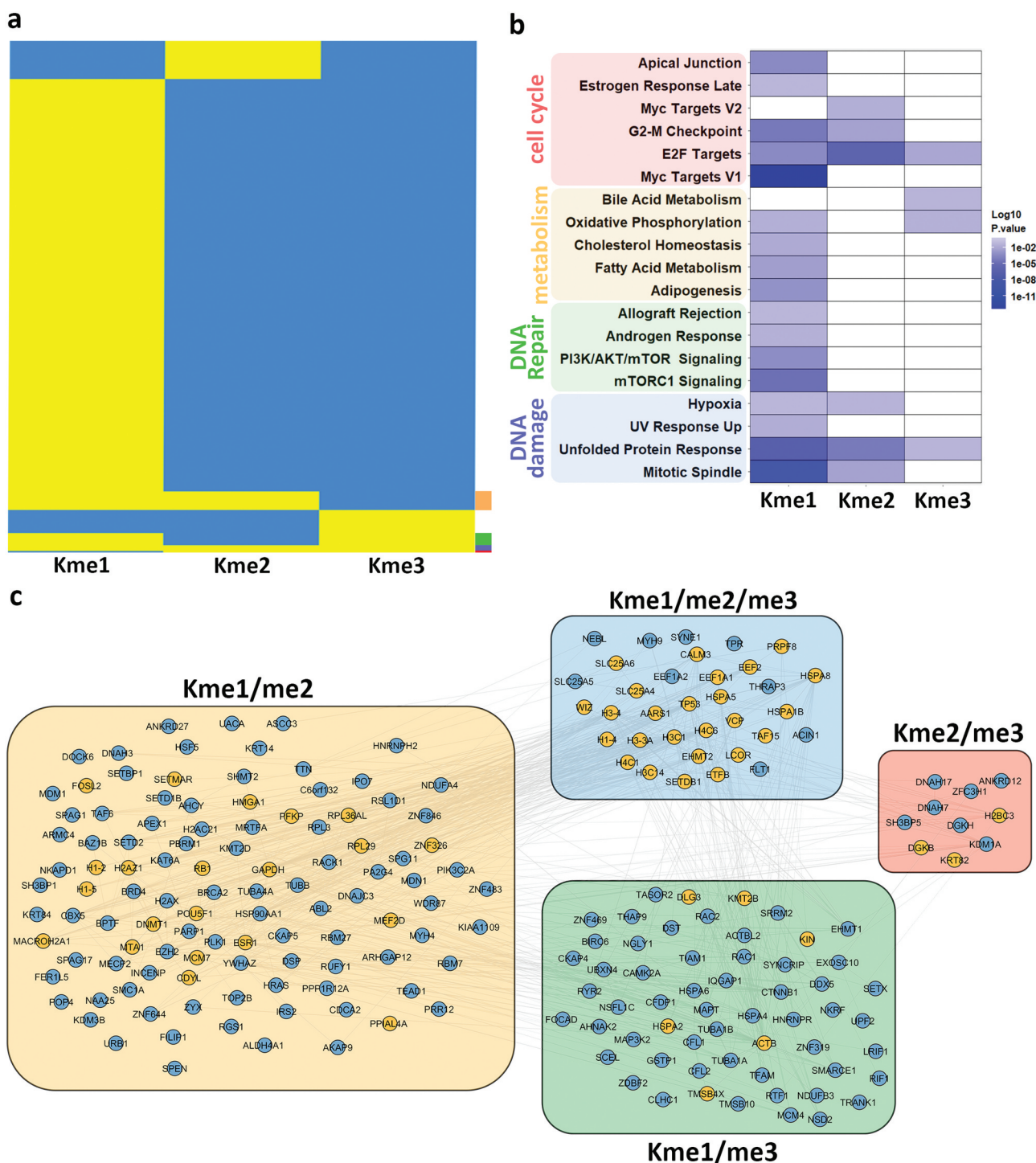


Figure 1. Deriving functional inferences from the lysine methylome. (a) Heatmap illustrating the distribution of 5485 lysine methylation marks across 2869 proteins, according to the PhosphoSite database v6.7.1.1. Yellow represents the presence of lysine methylation mark on the protein while blue represents lack of the methylation mark. Areas of overlap are highlighted as follows: orange for Kme1 (lysine monomethylation) and Kme2 (lysine dimethylation), green for Kme1 and Kme3 (lysine trimethylation), blue for Kme1, Kme2, and Kme3, and red for Kme2 and Kme3. (b) MsigDBundefinedDBundefined hallmark gene ontology analysis depicting the molecular functions of proteins tagged with mono-, di-, and trimethylation at lysine residues. (c) Overlapping groups are highlighted in panel (a), displaying a protein network that shows interactions among proteins with one or more lysine methylation marks. Blue nodes represent proteins that contain variations of lysine methylation throughout the entire protein. Orange nodes indicate proteins with at least one residue previously identified with different variations of lysine methylation.

functions in key biological pathways. Notably, these proteins are enriched in pathways that are heavily regulated by post-translational modifications including both histone and non-histone proteins. Specifically, we observed involvement in Chromatin Organization with proteins such as RB1, Histone

H1.5, MTA1, Histone H3.1t, Histone H3.2, Histone H2B type 1-B, Histone H3.1, MACROH2A1, ESR1, and Histone H3.3; Protein Folding including the heat-shock proteins HSPA8, HSPA5, HSPA2, and HSPA1B; and lastly Regulation of Gene Expression with proteins such as RB1, DNMT1, MTA1,

SETDB1, Histone H3.1, HMGA1, ZNF326, P53, MACROH2A1, ESR1, POU5F1, and HSPA1B. This comprehensive and systematic analysis of the lysine methylome provides not only an updated resource but also a framework for understanding how lysine methylation, through its marking of diverse proteins and pathways, can contribute to complex regulatory behaviors. Further research into the lysine methylome and the emergent properties identified in this study will be essential for elucidating its role in orchestrating cellular processes.

3.2. Exploring the interactome formed by 13 distinct histone methyltransferases reveals emerging properties of these molecular systems

We conducted an evidence-based analysis by leveraging literature based substrate specificity peptide studies for several lysine methyltransferases, including SUV39H1, SUV39H2 [68], SETD2 [51], SMYD2 [52], SUV420H1, SUV420H2 [53], and NSD1 [47]. Building on this foundation, we extended the analysis to include methyltransferases which have been previously known to methylate non-histone targets. By examining both the histone and non-histone targets of these enzymes, we derived the common short linear motifs (SLiMs) present across all known targets at the time this study was initiated. This approach allowed us to broaden our investigation to include EHMT1, EHMT2 [50], and the KMT2 family members KMT2A, KMT2B, KMT2C, and KMT2D [48] (Figure 2(a)). By utilizing our evidence-based derived SLiMs, we performed a peptide search for proteins containing each of the composite KMT-based SLiMs, focusing on residues that were the least variable across different motif variations. For SUV420H2 and SUV420H1, we used the motifs [K][IVLMK][LVFI] and [RY][K][VILM][LFI], respectively. For the SUV39 and EHMT proteins, we used the [R][K] motif due to the dominant role these residues appear to play. Next, we used [LF][K] for SMYD2, [IVFY][K] for SETD2, and [VIL][K] for NSD1. Based on the limited methylation-based literature for the KMT2 proteins, we used the motif [T][K] based on the reoccurrence of these residues in their histone substrates H3K4 and K305 of P53 [48]. We utilized ChromoHub v.2.0 (<https://chromohub.thesgc.org/static/ChromoHub.html>) to group the lysine methyltransferases under investigation based on sequence similarity, as determined through phylogenetic tree analysis [69]. This analysis resulted in four groups: Group 1 consisted of EHMT1, EHMT2, SUV39H1, SUV39H2, and NSD1; Group 2 included KMT2A, KMT2B, KMT2C, and KMT2D; Group 3 comprised SMYD2 and SETD2; and Group 4 included SUV420H1 and SUV420H2 (Figure S1). To increase the stringency of our candidates for histone mimicry we filtered large-scale proteomic datasets for each of the methyltransferases to identify interactors containing their respective modifiable SLiM [56]. We identified a total of 802 distinct proteins that hereafter will be referred to as the motif-associated interactome (Table S2). In particular, we found 36 candidates for SUV420H2 and only 3 for SUV420H1. For SUV39H2, we identified 69 interactors, while for SUV39H1, we identified a total of 179. Additionally, we found 31 interactors for SMYD2, 103 for SETD2, and 72 for NSD1. For the KMT2 proteins, we identified 149 for KMT2A, 68 for KMT2B, 76 for KMT2C, and 89 for KMT2D. Finally, we identified 125 interactors for EHMT1 and 213 for EHMT2. Next, we investigated the

significant enrichment for protein complexes within the motif-associated interactome. This analysis is centered on the potential complexes formed by the motif-associated interactome, without presuming the presence of the respective lysine methyltransferases in these complexes. CORUM enrichment analysis revealed two consistent repressive complexes within the motif-associated interactome, specifically the C-terminal binding protein (CtBP) complex and nucleosome remodeling and deacetylation (NuRD) complex (Figure 2(b)). KMT2A was the only KMT2 family protein that showed significant enrichment for the NuRD complex and was related to the CtBP complex through the shared HDAC1 and HDAC2 proteins. In contrast, SUV39H1 showed enrichment for the CtBP complex and shared the HDAC proteins with the NuRD complex. The motif-associated interactome for EHMT1 and EHMT2 showed significant enrichment for both the NuRD and CtBP complexes. Moreover, both the EHMT2 family (EHMT1 and EHMT2) and the SUV39H1 and SUV39H2 methyltransferases from group 1 showed significant enrichment for the RBL2 complex. The enrichment of these complexes as targets of lysine methylation may allude to an activation/assembly or deactivation/disassembly mechanism that is regulated through lysine methylation. For the unique complexes, we found that the motif-associated interactome for the KMT2 families included their respective MLL complexes. Interestingly, the SUV39 proteins showed significant enrichment for the Chaperonin Containing TCP-1 (CCT) complex, with all members of this complex containing a modifiable motif and interacting with SUV39 KMTs (Figure 2(b)). To our knowledge, this is the first description of KMTs being associated with a role in de novo protein folding, potentially through complex formation or regulation of a chaperonin complex by lysine methylation. However, further rigorous research is required to confirm the extent of this association. In summary, the analysis of the motif-associated interactome uncovered that KMTs interact with a variety of proteins known to form significant functional and biologically relevant complexes. The identification of modifiable SLiMs within these interacting proteins suggests that KMTs may play a pivotal role in regulating or modulating the assembly, disassembly, or activity of these complexes through lysine methylation. This finding highlights an emergent property of KMTs, wherein their involvement extends beyond direct substrate methylation to influencing broader protein complex dynamics.

3.3. Exploring histone mimicry candidates implicates protein-protein regulatory mechanisms by lysine methyltransferases

For internal validation, we combined the motif-associated interactome with the lysine methylome to match a total of 70 proteins to their respective KMT based on the lysine methylation status and protein-protein interactions that contain modifiable SLiMs (Table S3). This list of 70 proteins was considered our final refined set of histone mimicry candidates. We took our histone mimicry candidates and generated a gene network for visual representation (Figure 3(a)). In Group 1 (EHMT1, EHMT2, SUV39H1, SUV39H2, and NSD1), we found a total of 48 unique histone mimicry candidates. Of these,

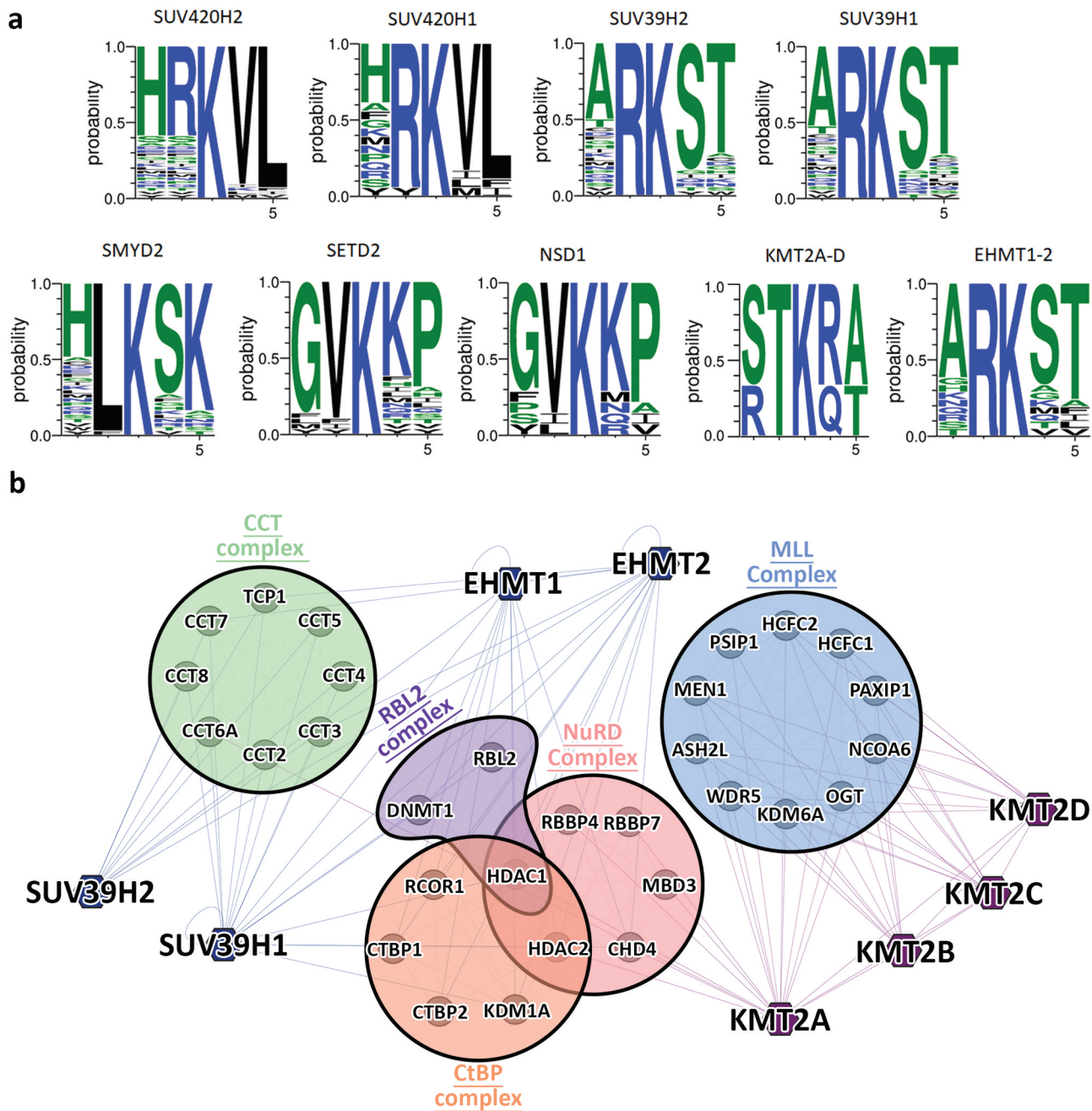


Figure 2. Substrate specificity linear analysis highlights potential functional associations and protein complex interactions of lysine methyltransferase candidates. (a) Linear motifs of known substrate candidates for 13 KMTs were curated through a literature search, as described in the methods, to generate a SLiM for each KMT. The sequence logo representing KMT2 was used for KMT2A, KMT2B, KMT2C, and KMT2D, highlighting conserved substrate recognition patterns among these methyltransferases. (b) Protein network showing lysine methyltransferases (KMTs) EHMT1, EHMT2, SUV39H1, SUV39H2, KMT2A, KMT2B, KMT2C, and KMT2D, and their motif-associated interactors. The interactors, identified through CORUM as enriched in protein complexes, are linked to the KMTs based on their specific SLiMs. KMTs are color-coded to their respective groups: group 1 (blue) includes EHMT1, EHMT2, SUV39H1, and SUV39H2, while group 2 (purple) comprises KMT2A, KMT2B, KMT2C, and KMT2D. The network highlights interactions between KMTs and proteins with modifiable SLiMs associated with each specific KMT. Protein complexes are represented by colored circles with the names of the complexes listed next to them, the enriched complexes include RBL2 (purple), CtBP (orange), MLL (blue), and NuRD (red), which are known for their roles in transcriptional repression, as well as the chaperone complex CCT (green), involved in protein folding. The analysis focuses on SLiM-associated interactors and does not include KMTs in the protein complex enrichment analysis.

43.75% (21/48) were identified as candidates for more than one of these lysine methyltransferases, likely due to the large similarity in protein sequence and substrate specificity, especially between EHMT1 and EHMT2, as well as SUV39H1 and SUV39H2, both of which regulate H3K9 methylation. For Group 2 (KMT2A, KMT2B, KMT2C, and KMT2D), we identified 7 proteins. Among these, only NUCKS1 and TAF6 were not

shared by more than one KMT2 methyltransferase. Group 3 (SETD2 and SMYD2) was associated with 20 histone mimicry candidates. Of these, only P53 was found interacting with both SMYD2 and SETD2, showing distinct partners despite both being regulators of H3K36. This was expected from our SLiM analysis due to the large variation in the modifiable residue sequences (Figure 2(a)). Lastly, Group 4 (SUV420H1 and

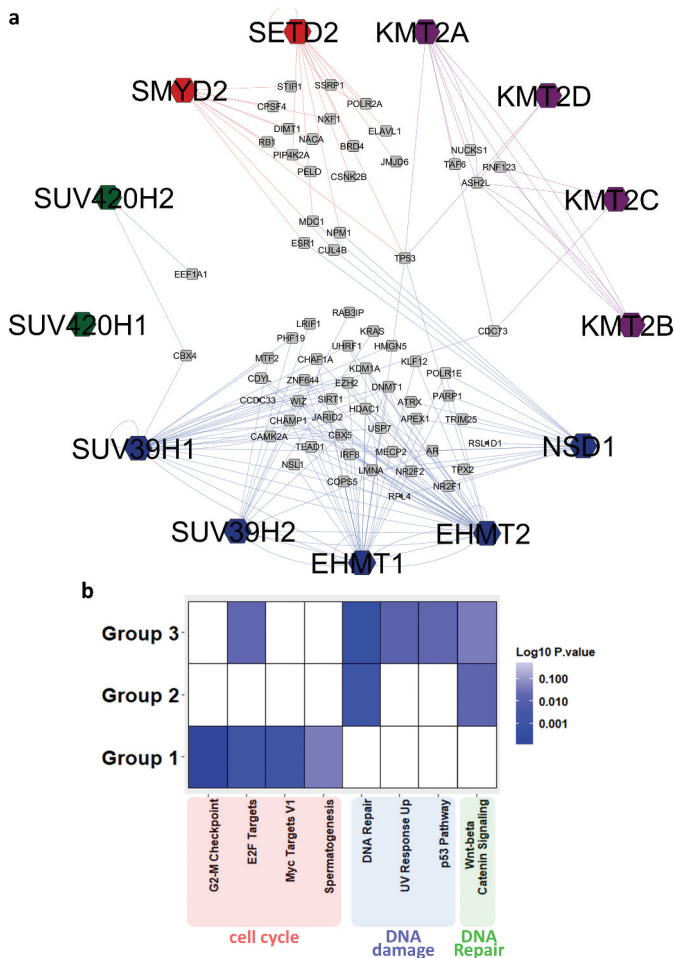


Figure 3. Network of histone mimicry candidates are associated with key biological pathways. (a) Network visualization of histone mimicry candidates generated using Cytoscape. The network displays four groups of lysine methyltransferases (KMTs) investigated in this study: group 1 (blue) includes EHMT1, EHMT2, SUV39H1, SUV39H2, and NSD1; group 2 (purple) consists of KMT2A, KMT2B, KMT2C, and KMT2D; group 3 (red) includes SETD2 and SMYD2; and group 4 (green) comprises SUV420H1 and SUV420H2. The network illustrates the overlap of histone mimicry candidates among the KMT groups, with some candidates shared across different groups. Edges are color-coded to indicate the association of histone mimicry candidates with specific KMT groups, highlighting the common targets among them. (b) MSigDB enrichment analysis of histone mimicry candidates reveals significant biological relevance in cell-cycle, DNA damage, and DNA repair pathways according to interactions found within the group KMTs.

SUV420H2) contained only two histone mimicry candidates: EEF1A1 and CBX4, with the latter being shared with SUV39H1. Interestingly, the largest shared targets from different groups were between NSD1 (Group 1) and SMYD2 (Group 3), both of which are H3K36 modifiers despite differences in their protein sequences. These histone mimicry candidates included MDC1, NPM1, ESR1, and CUL4B. CDC73 contained a methylated TK motif and interacted with both KMT2C and KMT2A while also containing a further methylated H3K9-like motif 'ARKT' and interacting with SUV39H1. P53 was the only protein observed to be a candidate for histone mimicry by the methyltransferases within Groups 1, 2, and 3. Performing a MSigDB ontology analysis for all candidates found in Groups 1–3 (Group 4 was excluded due to a lack of candidates) revealed distinct

biological relevance between the 3 groups of histone mimicry candidates. Interestingly, Group 1 was more involved in cell cycle molecular systems, while Groups 2 and 3 showed enrichment in DNA damage and DNA repair-related systems (Figure 3(b)). Understanding the distinct interactions within the identified histone mimicry candidates, reveals the potential regulatory influence of lysine methylation, which may act as molecular switches on shared targets, modulating key biological processes such as the cell cycle and DNA repair. These processes emerge as critical properties of histone mimicry, highlighting the broader impact of the investigated lysine histone-methyltransferases through non-histone methylation, impacting protein dynamics within biological systems. Further investigation into these mechanisms could provide valuable insights into understanding the full extent of their regulation in biological systems.

3.4. Uncovering the role of lysine methyltransferases in cancer and cell essentiality through histone mimicry

Recognizing the critical roles of histone mimicry candidates in regulating cell cycle and DNA damage/repair pathways, which are central to cancer biology, we focused our analysis on identifying histone mimicry proteins essential for cell survival and frequently mutated in cancer. To derive these cancer-related functional insights, we utilized our histone mimicry candidates (Table S3) and analyzed their essentiality and cancer census status. First, we investigated the systemic impact of histone mimicry, utilizing the DepMap dataset. This dataset contains gene dependency scores of over 17,000 genes across 1,086 cancer cells obtained from CRISPR screening. Setting a gene dependency maximum score of -1 , we identified 14 histone mimicry proteins as essential for cell survival, namely, *RPL4*, *CPSF4*, *TAF6*, *CDC73*, *NXF1*, *COP55*, *SSRP1*, *NACA*, *CHAF1A*, *RSL1D1*, *BRD4*, *TPX2*, *CSNK2B* and *NSL1*. All of these genes were found to be essential in over 500 different datasets, underscoring their critical role across multiple cell lines and cancer types (Figure 4(a)). Furthermore, using the GDC cancer portal, we identified a further 13 proteins that were found to be mutated across 6,389 cancer cases and annotated as cancer census genes, highlighting that their proteins are likely implicated in cancer development and progression (Figure 4(b)). Three genes were identified again, *CDC73*, *NACA*, and *BRD4* along with 10 additional genes including *TP53*, *KRAS*, *ATRX*, *SETD2*, *RB1*, *AR*, *ESR1*, *EZH2*, *LMNA* and *BPM1*. Further analysis of the 24 histone mimicry candidates, which include both proteins essential for cell survival and those annotated as cancer census genes, revealed distinct functional groups, as represented in Figure 4(c). This included *BRD4*, *TP53*, *RB1*, *KRAS*, *CDC73*, *TPX2*, *NSL1*, *CSNK2B*, and *COP55*, all of which are instrumental in regulating cell cycle progression and maintaining genomic stability. We also found essential cellular DNA repair machinery components, such as *NPM1*, *TAF6*, *LMNA*, and *ATRX*. Moreover, the chromatin remodelers *SSRP1*, *EZH2*, and *CHAF1A* were also identified among these candidates. In addition, *NACA*, *RSL1D1*, *RPL4*, *NXF1*, and *CPSF4* emerged as key players in ribosome biogenesis and RNA processing. Our analysis also predicted hormone

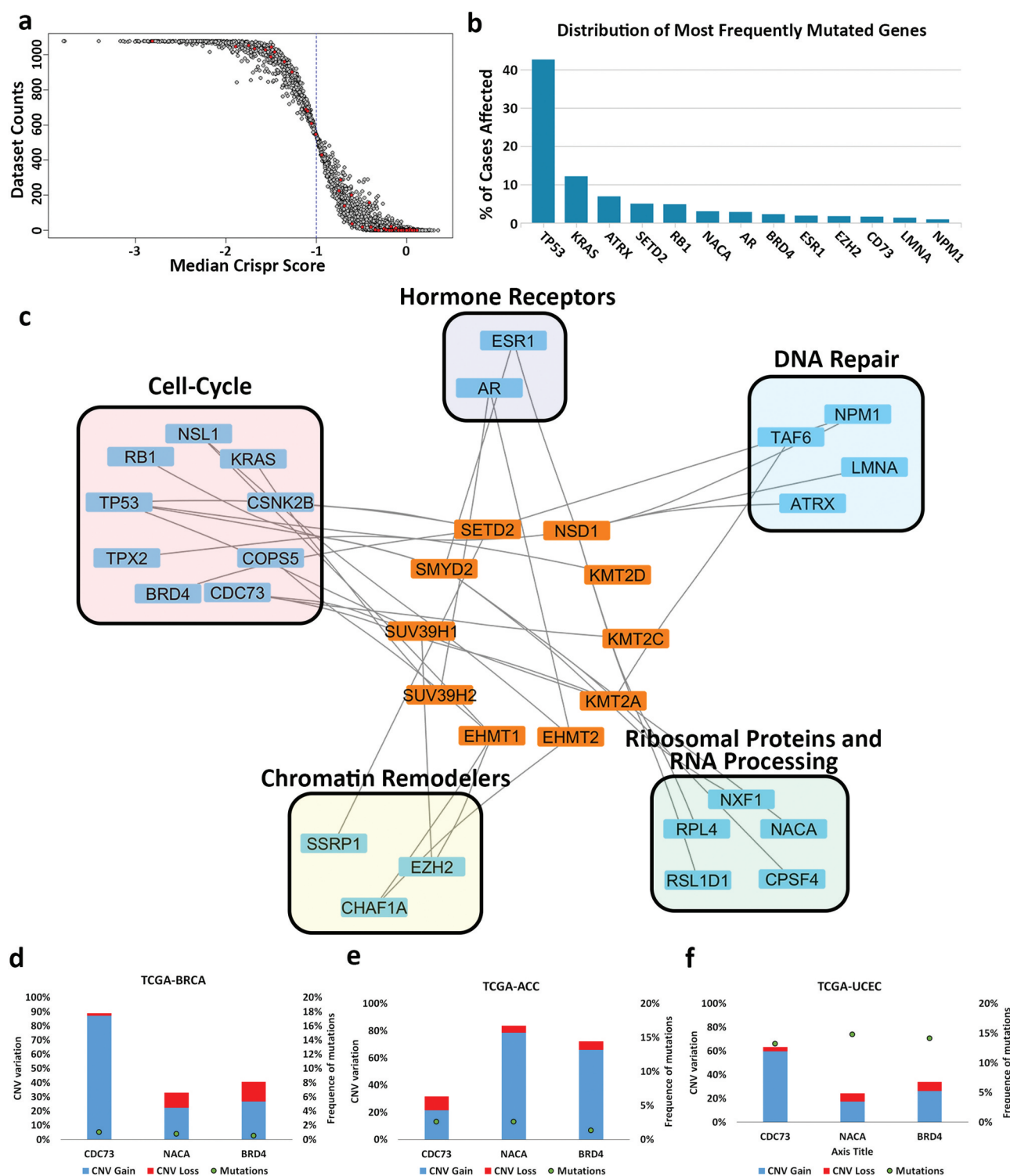


Figure 4. Internally validated candidates of lysine histone mimicry show relevance in cell survivability and cancer. (a) Graph illustrating essentiality through gene dependency scores for 17,386 genes across 1086 cancer cell lines from the DepMap data portal, highlighting the 70 internally validated candidates of histone mimicry in red. Genes with a media crispr score below -1 are classified as essential for cell survival. (b) GDC data portal identified 13 validated candidates of histone mimicry defined as cancer-census (genes with causal impacts in human cancer) and found to be frequently mutated across different cancers. (c) Gene network depicting 10 of the KMTs (orange) which are potential writers for the remaining 23 internally validated candidates relevant for cell survivability and/or cancer. The histone mimicry candidates are grouped based on their molecular functions, illustrating functional relationships and potential pathways involved in cancer biology. Graphs show the CNV gains, losses and mutations for CDC73, NACA, and BRD4, which are essential and cancer-related proteins, across three types of cancer (d) Breast cancer, (e) Adrenocortical carcinoma, and (f) Uterine corpus endometrial carcinoma.

receptors ESR1 and AR as candidates for histone mimicry. Thus, the regulation of these previously detected methylation marks extends KMT functionality beyond transcriptional regulation through histone methylation to non-canonical

pathways that regulate normal cell growth and key cancer-related proteins. Given that CDC73, BRD4, and NACA were found to be essential and mutated in approximately 3%, 4%, and 5% of the 6,389 cancer cases analyzed, we conducted

further investigations (Figure 4(a,b)). To that end, we utilized the Genomic Data Commons (GDC) Data Portal of harmonized cancer datasets to investigate the largest gains of copy number variation (CNV) across all represented cancers. This analysis revealed an 87% CNV gain for *CDC73* in invasive breast carcinoma (Figure 4(d)), whereas *BRD4* and *NACA* exhibited a CNV gain of 65% and 78%, respectively, in adrenocortical carcinoma (Figure 4(e)). Finally, using the Uterine Corpus Endometrial Carcinoma project dataset, we found *NACA* was mutated in around 14% of cases, followed by *BRD4* and *CDC73* which were observed in just over 13% and 12% of cases, respectively (Figure 4(f)). These findings underscore the emergent properties of KMTs through histone mimicry, implicating their functions in regulating essential and cancer-related proteins through a systemic biological regulatory approach. Therefore, understanding the potential regulation of these proteins through methylation could be useful in exploring alternative methods for targeting these essential cancer-related proteins, as well as comprehending the full impact of inhibiting their respective KMT in cancer-related therapies.

3.5. Molecular dynamics simulation characterizes how methylation by histone mimicry impacts the structure and biomolecular function of methylated targets

To uncover potential regulatory mechanisms that may be of importance in cancer, we focused on the methylation sites that have been previously reported [35] and the histone mimicry candidates *BRD4* and *CDC73*, due to their essential role in cell growth and cancer. Initially, we aimed to characterize the methylation status of the proteins *BRD4* and *CDC73*, focusing on the associated SLiMs and the likely KMTs involved in their methylation. To derive functional inferences, we employed evolutionary analysis to assess the structural and functional importance of the methylated lysine residues, evaluating their conservation and potential impact on protein function and/or structure. Based on the specificity analysis from Schuhmacher et al. [51], the di-methylation reported by large-scale mass-spec datasets at residues K404 and K406 of bromodomain 2 (BD2) in *BRD4*, which contains the motif 'IKSKL,' is potentially mediated by SETD2. This is due to previously described interactions and the presence of the recognizable 'IK' pair of residues. However, while SETD2 is primarily known for its tri-methylation activity, it has also been observed to catalyze de novo mono- and di-methylation events that can lead to a tri-methylated lysine mark [70]. Therefore, further investigation of these sites is warranted to assess their tri-methylation status and gain a comprehensive understanding of the methylation dynamics at these residues. The authors also acknowledge the similarity to the 'KSK' motif of the dimethylation KMT SMYD2, suggesting it as another potential candidate for methylating these sites (Figure 2(a)). Similarly, we predict that the monomethylation observed at the K283 residue of *CDC73* is likely catalyzed by one of the KMT2 proteins, due to the similarity of the K283 SLiM (RTKQP) to the H3K4 SLiM 'RTKQT.' The K331 residue of *CDC73* is likely a substrate for SUV39H1 or SUV39H2, based on previously identified interactions and the similarity between the SLiM

'ARKTQ' of *CDC73* K331 and the H3K9 motif 'ARKST.' Alternatively, given that the methylation at K331 is mono-methylated, it is more plausible that the EHMT proteins, which also target H3K9, are involved. However, while the datasets used in this study revealed no previous interactions between *CDC73* and either EHMT1 or EHMT2, their interaction has been discussed in the literature [18,71] (Figure 2(a)). Both the K283 and K331 residues fall within the hPAF1 binding domain of *CDC73*, suggesting that methylation at these residues may play a role in the assembly or disassembly of the PAF1 complex [72]. Thus, to predict the functional and structural relevance of these sites, we performed 3D protein conservation analysis and observed that both K404 and K406 are exposed and highly conserved within *BRD4*, suggesting that they likely play an important functional role for this protein (Figure S2A). In contrast, while exposed, the K283 and K331 residues of *CDC73* are located within highly variable and intrinsically disordered regions of the protein. (Figure S2B). Thus, it is not clear if these sites would impact the function and/or structure of *CDC73*. However, these variable regions of *CDC73* May result in species-specific functions and thus would require further investigation.

Subsequently, we chose to further explore the BD2 domain of *BRD4*, as a case demonstration of the impact of histone mimicry methylation, due to the availability of a high-resolution experimental structure of this domain in a complex with an acetylated peptide, (Foxo3a-K242ac/K245ac), allowing us to define changes in reading activities upon methylation [73]. By incorporating the experimentally validated methylation marks at K404 and K406 into the *BRD4* protein in silico, our analysis aims to elucidate how methylation at these specific lysine residues can alter the structural and dynamic properties of the BD2 domain, potentially affecting how this modification influences *BRD4* and its reader domain's functional interactions. Figure 5(a) shows the structure of the BD2 domain, with the inset focused on the region near residues K404 and K406. This domain contains four α -helical tufts, referred to as α Z, α A, α B, and α C, as well as two interhelical loops, ZA and BC. These structures together create a binding region for acetyl-lysine-containing peptides docking at one extremity of the helical bundle [74]. The ZA loop connects the α A and α Z helices, while the BC loop binds α B and α C, forming a hydrophobic pocket. This pocket likely stabilizes the left-handed four-helix bundle [75]. A highly conserved asparagine residue in the BC loop positions itself to create a hydrogen bond with the amide nitrogen and the acetyl-lysine carbonyl oxygen [76]. Utilizing the PDB 6MNL, we performed molecular dynamics on *BRD4*, in which we observed that the methylated lysine residues 404 and 406 are found on opposing sides of the α A helix and are structurally and dynamically important for the function of the BD2 domain (Figure 5(b)). We found that these two residues are surrounded by negatively charged or polar residues (Figure 5(a), inset). The side chain of K404 is orientated toward the α Z helix forming an intra-helical salt bridge with E408. Conversely, the side chain of K406 is pointing toward the α B helix forming a salt bridge with D421 on the α B helix while also making important interactions with E411 and Y412 in the folding region between the α A and α B helices. Methylation of K404 and K406 would alter the local electrostatic landscape and weaken these non-

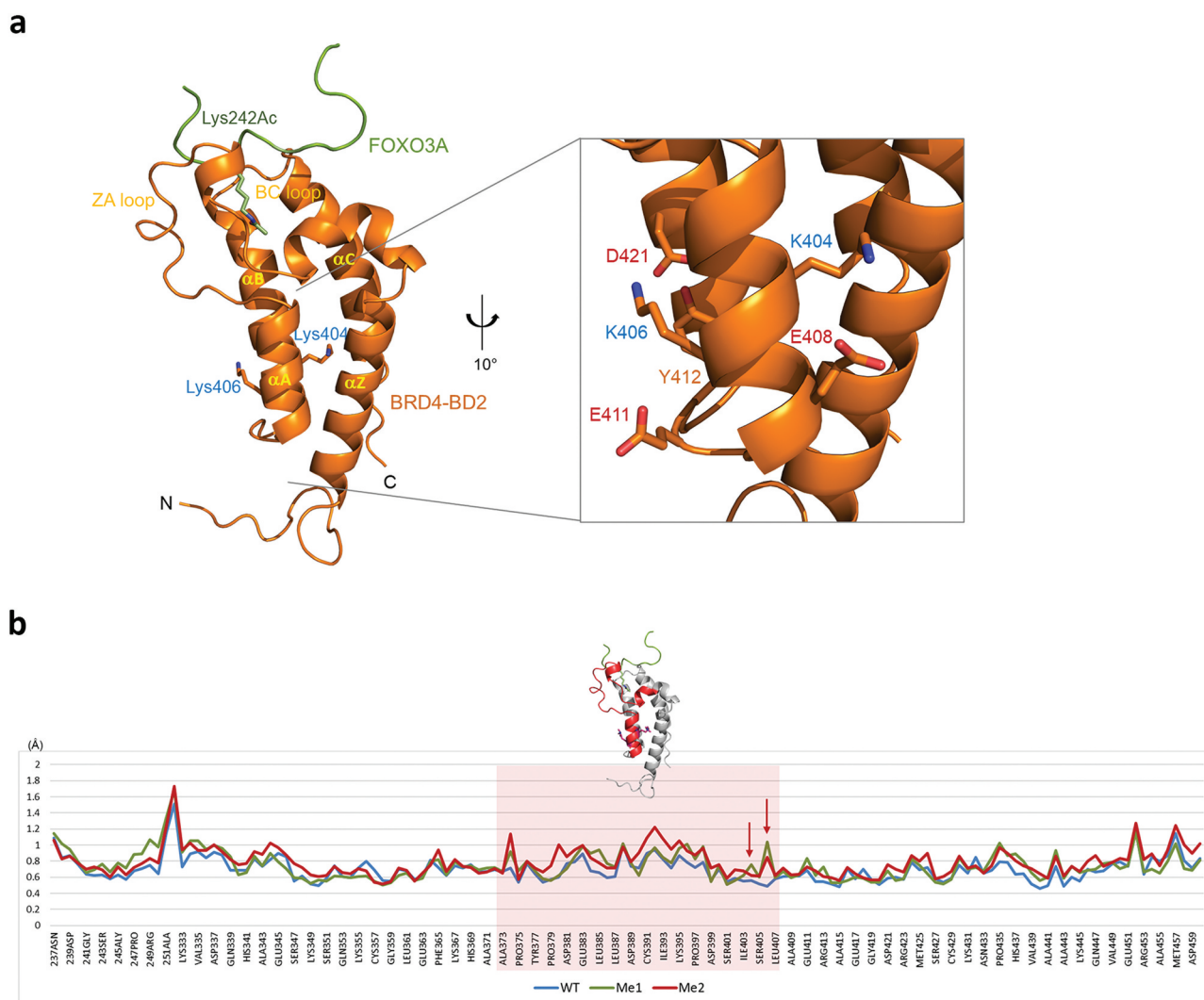


Figure 5. Molecular architecture of the BRD4-BD2 domain is. (a) A ribbon representation of the BRD4-BD2 domain in complex with a FOXO3A acetylated lysine (AcLys)-containing peptide (PDB 6MNL). K404 and K406 of BRD4 and the acetylated K424 of FOXO3A are shown as ball-and-sticks. In the magnified window, the surrounding negatively charged or polar residues are also highlighted. (b) RMSF plots for the BRD4-BD2 domain (native and modified proteins) in complex with an AcLys-containing peptide. The most prominent alterations in the dynamic motions are observed in the region that extends to the AcLys binding pocket (red shaded box). The affected region is shown in red in the ribbon representation of the structure. The positions of the methylated lysine residues are indicated by arrows.

covalent interactions in the region. In turn, these modifications could affect the functional protein-protein and/or protein-DNA interactions made by BD2 of BRD4 [77]. Indeed, our MD analysis with either mono- or di-methylation on both residues revealed that methyl-induced polarization causes a disturbance in the local structure and dynamic motions (how the protein's flexibility and conformational changes over time) compared to the wild-type (Figure 5(b) and Table S4). The alterations in its dynamic fluctuations were more prominent in the region that includes the modification sites and extends into the acetylated lysine binding pocket as highlighted in red (Figure 5(b)). The overall correlation coefficients of the RMSF of the WT replicates showed a relatively high correlation coefficient of 0.9, which aligns with typical observations in similar experiments comparing wild-type or variants [78,79]. The alterations in its dynamic fluctuations were more prominent in the region that includes the modification sites and extends into the acetylated lysine binding pocket as highlighted in red (Figure 5(b)). The overall correlation coefficients of the RMSF plots between the native

and the modified proteins are lower, ranging from 0.72 to 0.86 for the Di- and Mono-methylated proteins, respectively. The differences are more prominent in the region highlighted compared to the outside, and they are more salient between the di-methylated protein (red line) and the wild-type (blue line). As a result, the acetylated lysine (AcLys) interaction free energy went down by the methylated proteins (columns E and F of Supplementary Table S4). Although the interaction with AcLys-only was reduced slightly less for the di-methylated protein, the interaction with the entire peptide went down more for the di-methylated protein than the mono-methylated protein. As a result, its time-dependent interaction with a FOXO3A peptide containing an acetylated lysine residue (PDB 6MNL) was marginally reduced (Table S4). These findings demonstrate how endogenous histone mimicry, through lysine histone-modifying enzymes methylating BRD4 at specific residues, likely alters the dynamics of the BD2 domain ultimately impacting its reading functions. This occurs by destabilizing local structures, such as the inter-helical arrangement and the hydrophobic

pocket formed by the ZA and BC loops. While our results demonstrate significant changes in BRD4's structure and function due to methylation, further investigation is needed to understand how non-histone methylation might influence the traditional histone methylation activities of lysine methyltransferases. In conclusion, our molecular dynamics simulations demonstrate the emergent properties of lysine methylation in the regulation of an essential cancer-related protein. These findings highlight the potential broader contributions of the lysine methylome to the regulation of protein dynamics, underscoring the importance of further studies to elucidate these complex mechanisms in cancer biology.

4. Discussion

The innovative nature of the current study stems from defining and deriving emergent properties from the human lysine methylome, the writers of that methylome, and the regulation of proteins by histone mimicry, using a systems biology approach. In this study, we define the emergent properties of the lysine methylome as the complex and far-reaching effects that arise when considering lysine methylation within the broader context of the cellular system. While methylation on a single protein is a straightforward molecular modification, its impact becomes significantly more intricate when viewed across the entire proteome. The lysine methylome, which comprises methylated lysine residues across 15% of the known proteome, exhibits properties that extend beyond the sum of individual modifications. These properties emerge from the collective and often synergistic effects of methylation on multiple proteins, influencing intricate cellular pathways and processes in ways that cannot be predicted by examining a single protein in isolation. By regulating interactions between methylated proteins, their binding partners, and subsequent cellular cascades, the lysine methylome orchestrates higher-order biological functions (Figure 1(b)). This systems-level perspective underscores the dynamic role of lysine methylation in cellular regulation, revealing its significance in controlling complex molecular systems. In fact, our major findings provide the most up-to-date bioinformatics-driven systems biological meta-analysis of the human lysine methylome, offering valuable insights into its potential impact on cellular regulation. Furthermore, this study underscores the importance of understanding these emergent properties, highlighting the need for the development of resources to enhance specificity and for further research to elucidate the lysine methylome's role in cellular function. In addition, we characterize histone mimicry as an emerging regulatory system formed by 13 distinct KMTs, selected based on the availability of substrate specificity analysis or well-described studies on non-histone substrates. Considering the novelty of these discoveries, it becomes important to discuss their significance and biomedical relevance.

First, this study highlights the complex and interconnected nature of the lysine methylome as a molecular system. Analyzing 2,869 proteins comprising the lysine methylome, we found significant enrichment in critical cellular pathways, such as cell cycle, DNA repair, metabolism, and signaling pathways (Figure 1(b)). This is notable, as many of these methylation marks have been largely unexplored. Given their involvement in essential pathways, these modifications could be key to controlling molecular systems.

Methylation marks on non-histone proteins have been shown to regulate protein-protein [80] and protein-DNA binding interactions [30,81], alter sub-cellular localization [82], influence post-translational modification cross-talk [83], activate and stabilize proteins [30,84,85], and signal for protein degradation [86]. Consequently, a deeper understanding of lysine methylation could unveil novel strategies for therapeutic intervention and provide insights into the regulation of cellular processes, particularly in the context of cancer. One notable example of lysine methylation regulating key cancer pathways is SMYD2-mediated methylation of K810 on the tumor suppressor retinoblastoma protein (Rb), which enhances cell cycle progression in cancer cells [32]. Furthermore, cancer-related substrates, such as transcriptional regulators p53 [87] and GATA4 [88], as well as signaling kinases MAPKAPK3 [89] and AKT1 [90], are reported to substantially undergo regulation by lysine methylation. Thus, the current study expands upon these examples, underscoring the growing role of KMTs in regulating complex cellular processes and maintaining cellular homeostasis.

Subsequently, we defined a motif-associated interactome in humans, comprising 802 KMT-interacting proteins with the potential to be methylated through modifiable SLiMs associated with their respective KMTs, using a linear sequence-based approach (Figure 2(a) and Table S2). To our surprise, even though histone methylation has been known since 1964, its extended repertoire of regulatory functions has yet to be fully integrated and analyzed within a systems biology framework [91–94]. Fortunately, our analysis of the motif-associated interactome extends current knowledge by revealing that four transcriptional regulatory complexes are enriched with modifiable SLiMs and interact with the KMTs investigated in this study (Figure 2(b)). This emergent property suggests that methylation activity or interactions by the KMTs may regulate the activity or formation of these complexes. Furthermore, we discovered that all members of the CCT complex contained modifiable SLiMs and interacted with SUV39H1 and SUV39H2, extending the influence of lysine methylation into other molecular systems beyond transcriptional regulation (Figure 2(b)). Given the critical role of KMTs in cellular regulation and cancer, the development of inhibitors targeting these enzymes has become a major area of interest in cancer therapeutics. Tazemetostat, the only currently FDA-approved KMT inhibitor, targets EZH2 in the Polycomb repressive complex 2 (PRC2) and is used to treat certain epithelioid sarcoma and follicular lymphoma patients [95,96]. While Tazemetostat decreases all methylated forms of H3K27, leading to de-repression of PRC2 target genes, its effects on non-histone substrates like ROR α , STAT3, and GATA4 remain understudied [96,97]. UNC0638, an EHMT1 and EHMT2 inhibitor, which is still in its preclinical research phase, reduces H3K9 methylation and displays anti-proliferative and anti-cancer effects [83–85]. However, despite substantial evidence indicating that KMTs can target non-histone substrates and play critical roles in the molecular dynamics of these proteins, similar to Tazemetostat, research on the therapeutic effects of inhibiting KMTs remains limited. Understanding these effects is essential, as it could uncover new regulatory functions of lysine methylation and broaden the therapeutic potential of the rapidly emerging class of KMT inhibitors.

As a next step, we cross-referenced the lysine methylome and motif-associated interactome datasets, identifying 70 lysine-

methylated proteins with candidate KMTs from our analysis (Figure 3(a)). We found that 14 of these proteins are essential and 13 are classified as cancer census genes, highlighting the critical role of these proteins in maintaining cellular homeostasis. Notably, we identified three proteins BRD4, CDC73, and NACA that are essential for cell survival and classified as cancer census genes, underscoring the potential role of KMTs in regulating cancer-related pathways beyond histone methylation (Figure 4(a,b)). We found that BRD4 contained lysine methylation at two sites, K404 and K406. Both sites are within the acetylation-reading domain of BRD4, BD2, known to bind H4K12ac, as well as non-histone proteins, such as acetylated FOXO3A. Interestingly, the K404 site has also been observed to be ubiquitinated, suggesting that this lysine may be involved in a post-translational molecular switch potentially driving different regulatory mechanisms. For CDC73, we observed marks of methylation and ubiquitination at both K283 and K331 sites of CDC73, alluding to another potential molecular switch mechanism. However, K283 also undergoes SUMOylation. Moreover, K283 is located within the Paf1 complex binding domain and is near two tyrosine sites, Y290 and Y293, which, when dephosphorylated by SHP2, which results in CDC73 converting from a tumor suppressor to an oncogenic driver [98]. These modifications demonstrate the important regulatory mechanism of this region and suggests potential crosstalk between methylation and phosphorylation, a phenomenon often observed with histones [99].

Lastly, utilizing an established model of the BRD4 BD2 domain in complex with the acetylated FOXO3A peptide, we demonstrated how both mono- and di-methylation at K404 and K406 can alter the functional and structural dynamics of BRD4 (Figure 5(a,b)). Our simulations revealed that mono- and di-methylation at these sites disturbs the local structure and dynamic motions, impacting the acetyl-lysine binding pocket without being directly within it (Figure 5(b) and Table S4). This potential allosteric regulatory mechanism induced by lysine methylation at K404 and K406 offers important knowledge when considering future therapeutic options. As most inhibitors of BRD4 target both BD1 and BD2 [100], regulating this region with lysine methylation at sites within BD2 has the potential to disrupt the reading activity of BRD4. This would include altering the acetylated FOXO3A/BRD4 interaction, which is known to upregulate genes like CDK6, promoting drug resistance in cancers like breast cancer [73]. However, further research is needed to fully elucidate and validate this mechanism. Thus, in summary, the current study extends our knowledge of the emergent properties of the molecular system formed from histone mimicry and the landscape of interactions regulated by lysine methylation. In addition to mechanistic relevance, the data described here has particular relevance to therapeutics and should be taken into consideration when evaluating the additional effects of KMT inhibitors.

5. Conclusion

In conclusion, this study significantly enhances our understanding of the lysine methylome and its emergent properties within the broader context of cellular systems. By defining the intricate interplay between lysine methylation, histone mimicry, and the regulatory networks involving key proteins across various

cellular pathways, we have highlighted the complex and dynamic role of lysine methylation in cellular regulation. The findings underscore the importance of considering the lysine methylome as a comprehensive system that extends beyond individual protein modifications, influencing critical processes such as DNA repair, cell cycle progression, and cellular signaling. This systems biology approach reveals the potential for developing targeted therapeutic strategies, particularly in the context of precision oncology, by identifying key regulatory pathways that can be modulated through KMT inhibitors. The study also emphasizes the need for further research to explore the broader impact of lysine methylation on non-histone proteins and to validate the potential therapeutic implications of these findings. Overall, this work provides a foundational framework for future investigations into the regulatory mechanisms of the lysine methylome and its relevance to disease and therapy.

Disclosure statement

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Funding

Research reported in this manuscript was supported by NIH [grant numbers R01DK52913 (to Raul Urrutia and Gwen Lomber) and R01CA247898 (to Gwen Lomber)]; Advancing a Healthier Wisconsin Endowment (to Gwen Lomber and Raul Urrutia); the Linda T. and John A. Mellowes Endowed Innovation and Discovery Fund (to Raul Urrutia); The Joel and Arlene Lee Endowed Chair for Pancreatic Cancer Research (to Gwen Lomber); and Markus Family Funds for Discovery and Innovation Family Funds to Mellowes Center for Genomic Sciences and Precision Medicine.

Author contributions

Gareth Pollin: Data curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing.

Angela J. Mathison: Supervision, Writing – review and editing.

Michael T. Zimmermann: Writing – review and editing.

Gwen Lomber: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing.

Raul Urrutia: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing.

Data availability statement

All data from this study are included in the supplementary tables and are also accessible through the following repositories:

Phosphosite: <https://www.phosphosite.org/homeActionDepMap>; <https://depmap.org/portal/>

BioGRID: <https://thebiogrid.org/>

ORCID

Gareth Pollin  <http://orcid.org/0000-0002-2787-6673>

Gwen Lomber  <http://orcid.org/0000-0001-5463-789X>

Raul Urrutia  <http://orcid.org/0000-0002-1640-6780>

References

Papers of special note have been highlighted as either of interest (*) or of considerable interest () to readers.**

- Feinberg Andrew P. The key role of epigenetics in human disease prevention and mitigation. *N Engl J Med*. 2018;378(14):1323–1334. doi: [10.1056/NEJMr1402513](#)
- Budhavarapu VN, Chavez M, Tyler JK. How is epigenetic information maintained through DNA replication? *Epigenet Chromatin*. 2013;6(1):32. doi: [10.1186/1756-8935-6-32](#)
- Srinageshwar B, Maiti P, L DG, et al. Role of epigenetics in stem cell proliferation and differentiation: implications for treating neurodegenerative diseases. *Int J Mol Sci*. 2016;17(2):199. doi: [10.3390/ijms17020199](#)
- Skinner MK. Role of epigenetics in developmental biology and transgenerational inheritance. *Birth Defects Res C Embryo Today*. 2011;93(1):51–55. doi: [10.1002/bdrc.20199](#)
- Miranda-Gonçalves V, Lameirinhas A, Henrique R, et al. Metabolism and epigenetic interplay in cancer: regulation and putative therapeutic targets. *Front Genet*. 2018;9. doi: [10.3389/fgene.2018.00427](#)
- Bou Kheir T, Lund AH. Epigenetic dynamics across the cell cycle. *Essays Biochem*. 2010;48(1):107–120. doi: [10.1042/bse0480107](#)
- Bedford MT, Clarke SG. Protein arginine methylation in mammals: who, what, and why. *Mol Cell*. 2009;33(1):1–13. doi: [10.1016/j.molcel.2008.12.013](#)
- Rice JC, Allis CD. Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr Opin Cell Biol*. 2001;13(3):263–273. doi: [10.1016/S0955-0674\(00\)00208-8](#)
- Min J, Feng Q, Li Z, et al. Structure of the catalytic domain of human dot1l, a non-set domain nucleosomal histone methyltransferase. *Cell*. 2003;112(5):711–723. doi: [10.1016/s0092-8674\(03\)00114-4](#)
- Sugeedha J, Gautam J, Tyagi S. SET1/mll family of proteins: functions beyond histone methylation. *Epigenetics*. 2021;16(5):469–487. doi: [10.1080/15592294.2020.1809873](#)
- Hyun K, Jeon J, Park K, et al. Writing, erasing and reading histone lysine methylations. *Exp Mol Med*. 2017;49(4):e324. doi: [10.1038/emm.2017.11](#)
- Duan Q, Chen H, Costa M, et al. Phosphorylation of h3s10 blocks the access of h3k9 by specific antibodies and histone methyltransferase. *J Biol Chem*. 2008;283(48):33585–33590. doi: [10.1074/jbc.M803312200](#)
- Williams MM, Mathison AJ, Christensen T, et al. Aurora kinase b-phosphorylated hp1α functions in chromosomal instability. *Cell Cycle*. 2019;18(12):1407–1421. doi: [10.1080/15384101.2019.1618126](#)
- Grzenda A, Leonard P, Seo S, et al. Functional impact of aurora a-mediated phosphorylation of hp1γ at serine 83 during cell cycle progression. *Epigenetics Chromatin*. 2013;6:21. doi: [10.1186/1756-8935-6-21](#)
- Casciello F, Kelly GM, Ramarao-Milne P, et al. Combined inhibition of g9a and ezh2 suppresses tumor growth via synergistic induction of il24-mediated apoptosis. *Cancer Res*. 2022;82(7):1208–1221. doi: [10.1158/0008-5472.CAN-21-2218](#)
- Wu M, Shu H-B. MLL1/wdr5 complex in leukemogenesis and epigenetic regulation. *Chin J Cancer*. 2011;30(4):240–246. doi: [10.5732/cjc.011.10055](#)
- Aziz N, Hong YH, G KH, et al. Tumor-suppressive functions of protein lysine methyltransferases. *Exp Mol Med*. 2023;55(12):2475–2497. doi: [10.1038/s12276-023-01117-7](#)
- Pollin G, Assuncao T M D, Doria Jorge S, et al. Writers and readers of h3k9me2 form distinct protein networks during the cell cycle that include candidates for h3k9 mimicry. *Biosci Rep*. 2023;43(10):BSR20231093. doi: [10.1042/BSR20231093](#)
- This study utilizes protein modeling and Short Linear Motifs (SLiMs) to predict potential histone mimicry candidates through computational analysis, providing a significant advancement in understanding protein interactions during the cell cycle.**
- Avgousti DC, Herrmann C, Kulej K, et al. A core viral protein binds host nucleosomes to sequester immune danger signals. *Nature*. 2016;535(7610):173–177. doi: [10.1038/nature18317](#)
- Silmon de Monerri NC, Kim K. Pathogens hijack the epigenome. *Am J Pathol*. 2014;184(4):897–911. doi: [10.1016/j.ajpath.2013.12.022](#)
- Marazzi I, Ho JSY, Kim J, et al. Suppression of the antiviral response by an influenza histone mimic. *Nature*. 2012;483(7390):428–433. doi: [10.1038/nature10892](#)
- Avgousti DC, Della Fera AN, Otter CJ, et al. Adenovirus core protein vii downregulates the DNA damage response on the host genome. *J Virol*. 2017;91(20):e01089–17. doi: [10.1128/JVI.01089-17](#)
- Kee J, Thudium S, M RD, et al. Sars-cov-2 disrupts host epigenetic regulation via histone mimicry. *Nature*. 2022;610(7931):381–388. doi: [10.1038/s41586-022-05282-z](#)
- Tarakhovsky A, Prinjha RK. Drawing on disorder: how viruses use histone mimicry to their advantage. *J Exp Med*. 2018;215(7):1777. doi: [10.1084/jem.20180099](#)
- This paper clearly defines viral and endogenous histone mimicry, which is essential for our research as it aims to uncover more about the endogenous aspect of histone mimicry and its regulatory roles at the systems biology level.**
- Talbert PB, Henikoff S. Histone variants—ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol*. 2010;11(4):264–275. doi: [10.1038/nrm2861](#)
- Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet*. 2016;17(8):487–500. doi: [10.1038/nrg.2016.59](#)
- Zhang X, Huang Y, Shi X. Emerging roles of lysine methylation on non-histone proteins. *Cell Mol Life Sci*. 2015;72(22):4257–4272. doi: [10.1007/s00018-015-2001-4](#)
- This paper highlights the critical roles of non-histone protein lysine methylation in human tumorigenesis, underscoring the importance of further investigating endogenous histone mimicry in cancer biology.**
- Rungtatanawanich W, Ballway JW, Wang X, et al. Post-translational modifications of histone and non-histone proteins in epigenetic regulation and translational applications in alcohol-associated liver disease: challenges and research opportunities. *Pharmacol Ther*. 2023;251:108547. doi: [10.1016/j.pharmthera.2023.108547](#)
- Hamamoto R, Saloura V, Nakamura Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat Rev Cancer*. 2015;15(2):110–124. doi: [10.1038/nrc3884](#)
- This paper highlights the critical roles of non-histone protein lysine methylation in human tumorigenesis, underscoring the importance of further investigating endogenous histone mimicry in cancer biology.**
- Chuiikov S, Kurash JK, R WJ, et al. Regulation of p53 activity through lysine methylation. *Nature*. 2004;432(7015):353–360. doi: [10.1038/nature03117](#)
- Huang J, Perez-Burgos L, J PB, et al. Repression of p53 activity by smyd2-mediated methylation. *Nature*. 2006;444(7119):629–632. doi: [10.1038/nature05287](#)
- Cho H-S, Hayami S, Toyokawa G, et al. RB1 methylation by smyd2 enhances cell cycle progression through an increase of rb1 phosphorylation. *Neoplasia*. 2012;14(6):476. doi: [10.1593/neo.12656](#)
- Kunizaki M, Hamamoto R, Silva FP, et al. The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by smyd3. *Cancer Res*. 2007;67(22):10759–10765. doi: [10.1158/0008-5472.CAN-07-1132](#)
- Piao L, Kang D, Suzuki T, et al. The histone methyltransferase smyd2 methylates parp1 and promotes poly(adp-ribosylation) activity in cancer cells. *Neoplasia*. 2014;16(3):257–264. doi: [10.1016/j.neo.2014.03.002](#)
- Hornbeck PV, Zhang B, Murray B, et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*. 2015;43(D1):D512–D520. doi: [10.1093/nar/gku1267](#)
- Gu Z. Complex heatmap visualization. *iMeta*. 2022;1(3):e43. doi: [10.1002/imt2.43](#)
- Xie Z, Bailey A, V KM, et al. Gene set knowledge discovery with enrichr. *Curr Protocol*. 2021;1(3):e90. doi: [10.1002/cpz1.90](#)
- Liberzon A, Birger C, Thorvaldsdóttir H, et al. The molecular signatures database (msigdb) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–425. doi: [10.1016/j.cels.2015.12.004](#)

39. Tsitsiridis G, Steinkamp R, Giurgiu M, et al. CORUM: the comprehensive resource of mammalian protein complexes–2022. *Nucleic Acids Res.* **2023**;51(D1):D539–D545. doi: [10.1093/nar/gkac1015](https://doi.org/10.1093/nar/gkac1015)
40. Wickham H. Data analysis. In: Wickham H, editor. *ggplot2: elegant graphics for data analysis*. Cham: Springer International Publishing; **2016**. p. 189–201. doi: [10.1007/978-3-319-24277-4_9](https://doi.org/10.1007/978-3-319-24277-4_9)
41. Otasek D, Morris JH, Bouças J, et al. Cytoscape automation: empowering workflow-based network analysis. *Genome Biol.* **2019**;20(1):185. doi: [10.1186/s13059-019-1758-4](https://doi.org/10.1186/s13059-019-1758-4)
42. Szklarczyk D, Kirsch R, Koutrouli M, et al. The string database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**;51(D1):D638–D646. doi: [10.1093/nar/gkac1000](https://doi.org/10.1093/nar/gkac1000)
43. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**;13(11):2498–2504. doi: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
44. Daks A, Vasileva E, Fedorova O, et al. The role of lysine methyltransferase set7/9 in proliferation and cell stress response. *Life (Basel).* **2022**;12(3):362. doi: [10.3390/life12030362](https://doi.org/10.3390/life12030362)
45. Dhayalan A, Kudithipudi S, Rathert P, et al. Specificity analysis-based identification of new methylation targets of the set7/9 protein lysine methyltransferase. *Chem Biol.* **2011**;18(1):111–120. doi: [10.1016/j.chembiol.2010.11.014](https://doi.org/10.1016/j.chembiol.2010.11.014)
46. Hoekstra M, Chopra A, Willmore WG, et al. Evaluation of jumoni c lysine demethylase substrate preference to guide identification of *in vitro* substrates. *Star Protocol.* **2022**;3(2):101271. doi: [10.1016/j.xpro.2022.101271](https://doi.org/10.1016/j.xpro.2022.101271)
47. Kudithipudi S, Lungu C, Rathert P, et al. Substrate specificity analysis and novel substrates of the protein lysine methyltransferase nsd1. *Chem Biol.* **2014**;21(2):226–237. doi: [10.1016/j.chembiol.2013.10.016](https://doi.org/10.1016/j.chembiol.2013.10.016)
48. Li Y, Zhao L, Tian X, et al. Crystal structure of mll2 complex guides the identification of a methylation site on p53 catalyzed by kmt2 family methyltransferases. *Structure.* **2020**;28(10):1141–1148.e4. doi: [10.1016/j.str.2020.07.002](https://doi.org/10.1016/j.str.2020.07.002)
49. Nishioka K, Chuikov S, Sarma K, et al. Set9, a novel histone h3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev.* **2002**;16(4):479–489. doi: [10.1101/gad.967202](https://doi.org/10.1101/gad.967202)
50. Rathert P, Dhayalan A, Murakami M, et al. Protein lysine methyltransferase g9a acts on non-histone targets. *Nat Chem Biol.* **2008**;4(6):344–346. doi: [10.1038/nchembio.88](https://doi.org/10.1038/nchembio.88)
51. Schuhmacher MK, Beldar S, Khella MS, et al. Sequence specificity analysis of the setd2 protein lysine methyltransferase and discovery of a setd2 super-substrate. *Commun Biol.* **2020**;3(1):511. doi: [10.1038/s42003-020-01223-6](https://doi.org/10.1038/s42003-020-01223-6)
52. Weirich S, Schuhmacher MK, Kudithipudi S, et al. Analysis of the substrate specificity of the smyd2 protein lysine methyltransferase and discovery of novel non-histone substrates. *Chembiochem.* **2020**;21(1–2):256–264. doi: [10.1002/cbic.201900582](https://doi.org/10.1002/cbic.201900582)
53. Weirich S, Kudithipudi S, Jeltsch A. Specificity of the suv4–20h1 and suv4–20h2 protein lysine methyltransferases and methylation of novel substrates. *J Mol Biol.* **2016**;428(11):2344–2358. doi: [10.1016/j.jmb.2016.04.015](https://doi.org/10.1016/j.jmb.2016.04.015)
54. Bailey TL, Johnson J, Grant CE, et al. The meme suite. *Nucleic Acids Res.* **2015**;43(Web Server issue):W39–W49. doi: [10.1093/nar/gkv416](https://doi.org/10.1093/nar/gkv416)
55. Obenaus JC, Cantley LC, Yaffe MB. Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* **2003**;31(13):3635–3641.
56. Oughtred R, Stark C, Breitkreutz B-J, et al. The biogrid interaction database: 2019 update. *Nucleic Acids Res.* **2019**;47(Database issue): D529–D541. doi: [10.1093/nar/gky1079](https://doi.org/10.1093/nar/gky1079)
57. Dempster JM, Rossen J, Kazachkova M, et al. Extracting biological insights from the project achilles genome-scale crispr screens in cancer cell lines. **2019 Jul 31**. doi: [10.1101/720243](https://doi.org/10.1101/720243)
58. Grossman Robert L, Heath Allison P, Ferretti Vincent, et al. Toward a shared vision for cancer genomic data. *N Engl J Med.* **2016**;375(12):1109–1112. doi: [10.1056/NEJMp1607591](https://doi.org/10.1056/NEJMp1607591)
59. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol.* **2002**;320(2):369–387. doi: [10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4)
60. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **2004**;60(Pt 12 Pt 1):2126–2132. doi: [10.1107/S0907444904019158](https://doi.org/10.1107/S0907444904019158)
61. Huang J, Rauscher S, Nawrocki G, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* **2017**;14(1):71–73. doi: [10.1038/nmeth.4067](https://doi.org/10.1038/nmeth.4067)
62. Grant BJ, Rodrigues APC, ElSawy KM, et al. Bio3d: an r package for the comparative analysis of protein structures. *Bioinformatics.* **2006**;22(21):2695–2696. doi: [10.1093/bioinformatics/btl461](https://doi.org/10.1093/bioinformatics/btl461)
63. Adhikari S, Nice EC, Deutsch EW, et al. A high-stringency blueprint of the human proteome. *Nat Commun.* **2020**;11:5301. doi: [10.1038/s41467-020-19045-9](https://doi.org/10.1038/s41467-020-19045-9)
64. Cornett EM, Ferry L, Defossez P-A, et al. Lysine methylation regulators moonlighting outside the epigenome. *Mol Cell.* **2019**;75(6):1092–1101. doi: [10.1016/j.molcel.2019.08.026](https://doi.org/10.1016/j.molcel.2019.08.026)
65. Berryhill CA, Hanquier JN, Doud EH, et al. Global lysine methylome profiling using systematically characterized affinity reagents. *Sci Rep.* **2023**;13:377. doi: [10.1038/s41598-022-27175-x](https://doi.org/10.1038/s41598-022-27175-x)
66. Wu Z, Connolly J, Biggar KK. Beyond histones – the expanding roles of protein lysine methylation. *FEBS J.* **2017**;284(17):2732–2744. doi: [10.1111/febs.14056](https://doi.org/10.1111/febs.14056)
67. Moore KE, Carlson SM, Camp ND, et al. A general molecular affinity strategy for global detection and proteomic analysis of lysine methylation. *Mol Cell.* **2013**;50(3):444–456. doi: [10.1016/j.molcel.2013.03.005](https://doi.org/10.1016/j.molcel.2013.03.005)
68. Weirich S, Khella MS, Jeltsch A. Structure, activity and function of the suv39h1 and suv39h2 protein lysine methyltransferases. *Life (Basel).* **2021**;11(7):703. doi: [10.3390/life11070703](https://doi.org/10.3390/life11070703)
69. Liu L, Zhen XT, Denton E, et al. ChromoHub: a data hub for navigators of chromatin-mediated signalling. *Bioinformatics.* **2012**;28(16):2205–2206. doi: [10.1093/bioinformatics/bts340](https://doi.org/10.1093/bioinformatics/bts340)
70. Park IY, Powell RT, Tripathi DN, et al. Dual chromatin and cytoskeletal remodeling by setd2. *Cell.* **2016**;166(4):950–962. doi: [10.1016/j.cell.2016.07.005](https://doi.org/10.1016/j.cell.2016.07.005)
71. Ropa J, Saha N, Chen Z, et al. PAF1 complex interactions with setdb1 mediate promoter h3k9 methylation and transcriptional repression of *hoxa9* and *meis1* in acute myeloid leukemia. *Oncotarget.* **2018**;9(31):22123–22136. doi: [10.18632/oncotarget.25204](https://doi.org/10.18632/oncotarget.25204)
72. Lamoliatte F, Caron D, Durette C, et al. Large-scale analysis of lysine sumoylation by sumo remnant immunoaffinity profiling. *Nat Commun.* **2014**;5(1):5409. doi: [10.1038/ncomms6409](https://doi.org/10.1038/ncomms6409)
73. Liu J, Duan Z, Guo W, et al. Targeting the *brd4/foxo3a/cdk6* axis sensitizes akt inhibition in luminal breast cancer. *Nat Commun.* **2018**;9(1):5200. doi: [10.1038/s41467-018-07258-y](https://doi.org/10.1038/s41467-018-07258-y)
74. Vollmuth F, Blankenfeldt W, Geyer M. Structures of the dual bromodomains of the p-tefb-activating protein *brd4* at atomic resolution. *J Biol Chem.* **2009**;284(52):36547–36556. doi: [10.1074/jbc.M109.033712](https://doi.org/10.1074/jbc.M109.033712)
75. Zaware N, Zhou M-M. Bromodomain biology and drug discovery. *Nat Struct Mol Biol.* **2019**;26(10):870–879. doi: [10.1038/s41594-019-0309-8](https://doi.org/10.1038/s41594-019-0309-8)
76. Wang Z-Q, Zhang Z-C, Wu Y-Y, et al. Bromodomain and extraterminal (bet) proteins: biological functions, diseases, and targeted therapy. *Sig Transduct Target Ther.* **2023**;8(1):1–26. doi: [10.1038/s41392-023-01647-6](https://doi.org/10.1038/s41392-023-01647-6)
77. Yang M, Liu K, Chen P, et al. Bromodomain-containing protein 4 (*brd4*) as an epigenetic regulator of fatty acid metabolism genes and ferroptosis. *Cell Death Dis.* **2022**;13(10):912. doi: [10.1038/s41419-022-05344-0](https://doi.org/10.1038/s41419-022-05344-0)
78. Fuglebakk E, Echave J, Reuter N. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics.* **2012**;28(19):2431–2440. doi: [10.1093/bioinformatics/bts445](https://doi.org/10.1093/bioinformatics/bts445)
79. Chi Y-I, Stodola TJ, Assuncao TMD, et al. Structural bioinformatics enhances the interpretation of somatic mutations in *kdm6a* found in human cancers. *Comput Struct Biotechnol J.* **2022**;20:2200. doi: [10.1016/j.csbj.2022.04.028](https://doi.org/10.1016/j.csbj.2022.04.028)

80. Tsusaka T, Kikuchi M, Shimazu T, et al. Tri-methylation of atf7ip by g9a/glp recruits the chromodomain protein mpp8. *Epigenetics Chromatin*. 2018;11(1):56. doi: [10.1186/s13072-018-0231-z](https://doi.org/10.1186/s13072-018-0231-z)
81. Zhang W, Wu X, Shi T, et al. Regulation of transcription factor yin yang 1 by set7/9-mediated lysine methylation. *Sci Rep*. 2016;6:21718. doi: [10.1038/srep21718](https://doi.org/10.1038/srep21718)
82. Fang L, Teng H, Wang Y, et al. SET1A-mediated mono-methylation at k342 regulates yap activation by blocking its nuclear export and promotes tumorigenesis. *Cancer Cell*. 2018;34(1):103–118.e9. doi: [10.1016/j.ccell.2018.06.002](https://doi.org/10.1016/j.ccell.2018.06.002)
83. Chen K, Liu J, Liu S, et al. Methyltransferase setd2-mediated methylation of stat1 is critical for interferon antiviral activity. *Cell*. 2017;170(3):492–506.e14. doi: [10.1016/j.cell.2017.06.042](https://doi.org/10.1016/j.cell.2017.06.042)
84. Subramanian K, Jia D, Kapoor-Vazirani P, et al. Regulation of estrogen receptor alpha by the set7 lysine methyltransferase. *Mol Cell*. 2008;30(3):336–347. doi: [10.1016/j.molcel.2008.03.022](https://doi.org/10.1016/j.molcel.2008.03.022)
85. Ivanov GS, Ivanova T, Kurash J, et al. Methylation-acetylation interplay activates p53 in response to DNA damage. *Mol Cell Biol*. 2007;27(19):6756–6769. doi: [10.1128/MCB.00460-07](https://doi.org/10.1128/MCB.00460-07)
86. Elkouris M, Kontaki H, Stavropoulos A, et al. SET9-mediated regulation of tgfbeta signaling links protein methylation to pulmonary fibrosis. *Cell Rep*. 2016;15(12):2733–2744. doi: [10.1016/j.celrep.2016.05.051](https://doi.org/10.1016/j.celrep.2016.05.051)
87. Huang J, Dorsey J, Chuikov S, et al. G9a and glp methylate lysine 373 in the tumor suppressor p53. *J Biol Chem*. 2010;285(13):9636–9641. doi: [10.1074/jbc.M109.062588](https://doi.org/10.1074/jbc.M109.062588)
88. He A, Shen X, Ma Q, et al. PRC2 directly methylates gata4 and represses its transcriptional activity. *Genes Dev*. 2012;26(1):37–42. doi: [10.1101/gad.173930.111](https://doi.org/10.1101/gad.173930.111)
89. Reynoird N, K MP, Stellfeld T, et al. Coordination of stress signals by the lysine methyltransferase smyd2 promotes pancreatic cancer. *Genes Dev*. 2016;30(7):772–785. doi: [10.1101/gad.275529.115](https://doi.org/10.1101/gad.275529.115)
90. Yoshioka Y, Suzuki T, Matsuo Y, et al. SMYD3-mediated lysine methylation in the ph domain is critical for activation of akt1. *Oncotarget*. 2016;7(46):75023–75037. doi: [10.18632/oncotarget.11898](https://doi.org/10.18632/oncotarget.11898)
91. Lehning NA, Morrison BE. Nonhistone lysine methylation as a protein degradation signal. *J Chem*. 2022;2022:1969299. doi: [10.1155/2022/1969299](https://doi.org/10.1155/2022/1969299)
92. Murray K. The occurrence of epsilon-n-methyl lysine in histones. *Biochemistry*. 1964;3:10–15. doi: [10.1021/bi00889a003](https://doi.org/10.1021/bi00889a003)
93. Biggar KK, S-C LS. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat Rev Mol Cell Biol*. 2015;16(1):5–17. doi: [10.1038/nrm3915](https://doi.org/10.1038/nrm3915)
94. Carlson SM, Gozani O. Nonhistone lysine methylation in the regulation of cancer pathways. *Cold Spring Harb Perspect Med*. 2016;6(11):a026435. doi: [10.1101/cshperspect.a026435](https://doi.org/10.1101/cshperspect.a026435)
95. Straining R, Eighmy W. Tazemetostat: ezh2 inhibitor. *J Adv Pract Oncol*. 2022;13(2):158–163. doi: [10.6004/jadpro.2022.13.2.7](https://doi.org/10.6004/jadpro.2022.13.2.7)
96. Julia E, Salles G. Ezh2 inhibition by tazemetostat: mechanisms of action, safety and efficacy in relapsed/refractory follicular lymphoma. *Future Oncol*. 2021;17(17):2127–2140. doi: [10.2217/fon-2020-1244](https://doi.org/10.2217/fon-2020-1244)
97. Park SH, Fong K-W, Mong E, et al. Going beyond polycomb: ezh2 functions in prostate cancer. *Oncogene*. 2021;40(39):5788–5798. doi: [10.1038/s41388-021-01982-4](https://doi.org/10.1038/s41388-021-01982-4)
98. Takahashi A, Tsutsumi R, Kikuchi I, et al. SHP2 tyrosine phosphatase converts parafibromin/cdc73 from a tumor suppressor to an oncogenic driver. *Mol Cell*. 2011;43(1):45–56. doi: [10.1016/j.molcel.2011.05.014](https://doi.org/10.1016/j.molcel.2011.05.014)
99. Lee J-S, Smith E, Shilatifard A. The language of histone crosstalk. *Cell*. 2010;142(5):682–685. doi: [10.1016/j.cell.2010.08.011](https://doi.org/10.1016/j.cell.2010.08.011)
100. To KKW, Xing E, Larue RC, et al. BET bromodomain inhibitors: novel design strategies and therapeutic applications. *Molecules*. 2023;28(7):3043. doi: [10.3390/molecules28073043](https://doi.org/10.3390/molecules28073043)