System-Wide Investigation of Hydroxymethylcytosine Recognition and Distribution in the Mammalian Genome

Dissertation

Submitted for the degree of Doctor of Natural Sciences
(Dr. rer. nat.)
to the
Faculty of Chemistry and Chemical Biology
Technical University of Dortmund

presented by

Lena Engelhard

from Wuppertal

Dortmund, 2025

„Without the love of research, mere knowledge and intelligence cannot make a scientist.“

— Irène Joliot-Curie

This work was prepared from January 2021 to February 2025 in the group of Prof. Daniel Summerer of the Faculty of Chemistry and Chemical Biology at the Technical University Dortmund.

The research projects presented in this work are funded by TU Dortmund, the Deutsche Forschungsgemeinschaft (503990008), and the European Research Council (10100794).

Abstract

The epigenetic cytosine modifications 5-methylcytosine (mC) and 5-hydroxymethylcytosine (hmC) are key regulatory elements of mammalian genomes, occurring within CpG dinucleotides in either strand-symmetric or strand-asymmetric combinations. While reader proteins for symmetrically and hemi-modified CpG dyads have been identified, the recognition of asymmetrically modified CpGs and the potential regulatory implications of their symmetry remain unexplored. In this work, mammalian nuclear proteins binding to hmC-containing CpG dyads were identified and characterized. A protocol for generating asymmetrically modified DNA probes was established and applied in pull-down assays coupled with mass spectrometry-based proteomics. Comparative enrichment studies were performed using promoter probes bearing symmetric or asymmetric C, mC, and hmC modifications, allowing for direct assessment of reader profiles in the same sequence, tissue and experimental contexts. In human and mouse nuclear extracts, numerous tissue-specific readers of hmC-modified sequences were identified, falling into distinct, probe-specific subgroups. These include transcription factors and chromatin regulators such as MYC and MAX, which read hmC in a sequence-dependent manner, and RFX5, which selectively discriminated between hmC symmetry states in CpG dyads. The presumably prevalent asymmetric CpG dyad hmC/mC exhibited a distinct reader protein profile, supporting the hypothesis that hmC symmetry information provides unique regulatory outputs. To complement these proteomics analyses, selective enrichment and sequencing of native double-stranded DNA fragments containing hmC/mC CpG dyads was implemented, based on an evolved methyl-CpG-binding domain protein with enhanced binding specificity. This approach enabled genome-wide mapping of hmC/mC sites, offering a valuable tool for investigating their roles in chromatin biology. Together, these findings provide a comprehensive framework for the study of symmetric and asymmetric hmC-containing CpG dyads and lay the groundwork for elucidating their functional contributions to transcriptional regulation, development and disease.

Zusammenfassung

Die epigenetischen Cytosin-Modifikationen 5-Methylcytosin (mC) und 5-Hydroxymethylcytosin (hmC) sind wichtige regulatorische Elemente des Säugergenoms und kommen innerhalb von CpG-Dinukleotiden entweder in strangsymmetrischen oder strangasymmetrischen Kombinationen vor. Während Leserproteine für symmetrisch- und halb-modifizierte CpG-Dyaden identifiziert wurden, sind die Erkennung asymmetrisch-modifizierter CpGs und damit die potenziellen regulatorischen Auswirkungen ihrer Symmetrie noch unerforscht. In dieser Arbeit wurden nukleare Säugetierproteine identifiziert und charakterisiert, die an hmC-haltige CpG-Dyaden binden. Es wurde ein Protokoll zur Erzeugung asymmetrisch modifizierter DNA-Sonden erstellt, welches in Pull-down-Analysen in Verbindung mit massenspektrometriebasierter Proteomik angewendet wurde. Vergleichende Anreicherungsstudien wurden mit Promotor-Sonden durchgeführt, die symmetrische oder asymmetrische C-, mC- und hmC-Modifikationen aufwiesen, wodurch eine direkte Bewertung der Leserprofile in derselben Sequenz, demselben Gewebe und denselben experimentellen Kontexten möglich war. In nuklearen Extrakten von Mensch- und Mausgeweben wurden zahlreiche gewebespezifische Leser von hmC-modifizierten Sequenzen identifiziert, die in unterschiedliche, sondenspezifische Untergruppen fallen. Dazu gehören Transkriptionsfaktoren und Chromatin-Regulatoren wie MYC und MAX, die hmC sequenzabhängig lesen, sowie RFX5, das selektiv zwischen hmC-Symmetriezuständen in CpG-Dyaden unterscheidet. Die vermutlich vorherrschende asymmetrische CpG-Dyade hmC/mC wies ein inividuelles Leserproteinprofil auf, was die Hypothese stützt, dass die Symmetrieinformation von hmC einzigartige regulatorische Ergebnisse zur Folge hat. Zur Ergänzung dieser Proteomanalysen wurde eine selektive Anreicherung und Sequenzierung von nativen doppelsträngigen DNA-Fragmenten, die hmC/mC-CpG-Dyaden enthalten, durchgeführt, basierend auf einem weiterentwickelten Methyl-CpG-bindenden Domänenprotein mit verbesserter Bindungsspezifität. Dieser Ansatz ermöglichte eine genomweite Kartierung von hmC/mC-Stellen und bietet ein wertvolles Werkzeug zur Untersuchung ihrer Rolle in der Chromatinbiologie. Zusammen bieten diese Ergebnisse einen umfassenden Rahmen für die Untersuchung symmetrischer und asymmetrischer hmC-haltiger CpG-Dyaden und legen den Grundstein für die Aufklärung ihrer funktionellen Beiträge zur Transkriptionsregulation, Entwicklung und Krankheit.

Acknowledgements

This thesis marks the end of a remarkable chapter in my life, and I find myself looking back with immense gratitude, not just for the science, but for the people. Although authored by me, this work was shaped by so many others. I have been incredibly fortunate to meet and work with amazing people and I want to thank each of you, though words alone can never do justice to my appreciation.

First and foremost, I want to thank Prof. Daniel Summerer, for giving me the opportunity to explore the fascinating world of epigenetic DNA modifications and their interactions with proteins. Thank you for trusting me with this project, for giving me the freedom to contribute my own ideas and for your support throughout all phases, scientifically and beyond. Your encouragement, guidance and the space you allowed for creativity made this journey very fulfilling.

I am sincerely grateful to Prof. Boris Pfander, who kindly agreed to act as the second examiner of this thesis.

A heartfelt thank you goes to my thesis advisory committee, Prof. Peter t'Hart and Dr. Jochen Imig, for your advice, your time, and your thoughtful insights throughout this project. The TAC meetings always left me with motivation and perspective and I deeply appreciate your openness and support.

I also wish to thank the collaborators who made this project what it is. To Dr. Petra Janning, thank you for your support and expert analysis in mass spectrometry, without which this project would not have been possible. To Prof. Rasmus Linser and Tye Gonzales, thank you for the insights in structural modeling and helping us interpret protein-DNA interactions. And to Dr. Sidney Becker and Marlon Zambrano, thank you for the discussions and analysis of the NGS data.

I feel incredibly lucky to have been part of the Summerer group. From scientific exchanges to retreats, group events and shared lunches, these experiences made many of you more than just coworkers. Thank you to everyone I had the pleasure to work with over the years: Dr. Anne Jung, Dr. Benjamin Buchmüller, Brinja Kosel, Dr. Damian Schiller, Dr. Jan Wolffgramm, Katrin Bigler, Kotryna Keliotyte, Lejla Maksumic, Sayan Sil, Dr. Shubhendu Palei, Simone Eppmann, Dr. Sudakshina Banerjee, Dr. Tzu-Chen Lin and Zeyneb Vildan Çakıl. Your support, kindness and encouragement brightened even the longest days. Also, a big thank you to our neighbors from the Dehmelt lab, Arya Sachan, Carolin Gierse and Ricarda Lüttig, for the chats and good vibes.

A very special thank you goes to Zeyneb Vildan Çakıl, your unwavering support throughout the proteomics project was invaluable. I am thankful for your dedication, your hands-on help with the revision experiments, your feedback while proof-reading this thesis and most importantly, your constant encouragement and friendship along the way. I would also like to thank Dr. Damian Schiller for your

ideas, enthusiasm and teamwork during the DNA enrichment project, as well as Simone Eppmann for your indispensable help with cloning and purification and for all kinds of advice.

My gratitude also goes to Naomi Zink, Lina Agorski, and Sara Miethe, it was a pleasure and an honor to supervise your bachelor theses. I learned from you as much as you (hopefully) learned from me.

I am deeply grateful to Dr. Lucia Sironi and Christa Hornemann for coordinating the IMPRS-LM program providing all the opportunities, workshops, travel support and for building such a wonderful community. Thank you for your incredible support and dedication.

Moreover, I would like to thank the CCB administrative team, Alicia Wachholz, Ilona Severin, Maria Sergani, Martina Reibner, and Petra Alhorn for your constant support in all the practical and bureaucratic matters.

One of the most unexpected and precious gifts of this journey has been the friendships I've made along the way. Zeyneb, Sudakshina, Ricarda, thank you for becoming not only close friends but also becoming my safe space through all the highs and lows. I honestly cannot imagine having done this without you and I feel very lucky and grateful to have found you.
Ein ganz besonderer Dank gilt all meinen Freunden außerhalb des Labors, die während all der Zeit immer für mich da waren. Insbesondere danke ich Katrin, Dunya, Ranya, sowie meinen Master-Mädels Lara und Michelle. Eure Freundschaft, euer Verständnis und eure Unterstützung haben mir den Weg sehr erleichtert und ich danke euch von Herzen für alles.
Am Ende möchte ich vor allem meiner Familie danken. Mama, danke für deine bedingungslose Unterstützung, dein Vetrauen und die Kraft, die du mir immer dann gegeben hast, wenn ich sie brauchte. Arne und Papa, ich danke euch für eure Unterstützung und euren Rückhalt, auf den ich mich immer verlassen konnte. Ohne euch wäre dieser Weg nicht möglich gewesen.

Thank you all for being part of this journey.

Publications

Part of this work has been published in:

Engelhard, L., Cakil, Z.V., Eppmann, S.,Gonzalez, T., Linser, R., Janning, P. Summerer, D. (2025) Mammalian Proteome Profiling Reveals Readers and Antireaders of Strand-Symmetric and -Asymmetric 5-Hydroxymethylcytosine-Modifications in DNA. *bioRxiv* 2025.06.27.661915; https://doi.org/10.1101/2025.06.27.661915

Other publications:

Lin, T. C., **Engelhard, L.**, Söldner, B., Linser, R., Summerer, D. (2024) Light-Activatable MBD-Readers of 5-Methylcytosine Reveal Domain-Dependent Chromatin Association Kinetics In Vivo. *Adv. Sci. 2024* 2307930 11(11), e2307930; https://doi.org/10.1002/advs.202307930

Chen, M., Blum, D., **Engelhard, L.**, Raunser, S., Wagner, R., Gatsogiannis, C. (2021) Molecular architecture of black widow spider neurotoxins. *Nat Commun. 2021* Nov 29;12(1):6956. https://doi.org/10.1038/s41467-021-26562-8

Dudziak, A., **Engelhard, L.**, Bourque, C., Klink, BU., Rombaut, P., Kornakov, N., Jänen, K., Herzog, F., Gatsogiannis, C., Westermann, S. (2021) Phospho-regulated Bim1/EB1 interactions trigger Dam1c ring assembly at the budding yeast outer kinetochore. *EMBO J. 2021* Sep 15;40(18):e108004. https://doi.org/10.15252/embj.2021108004

Contents

List of Figures

List of Tables

Abbreviations

A adenine

BCA bicinchoninic acid

BER base excision repair

bHLH basic helix-loop-helix

BS bisulfite sequencing

bZIP basic leucine zipper

C cytosine

caC 5-carboxycytosine

CDS coding sequence

ChIP chromatin immunoprecipitation

Cy3 Cyanine 3

Cy5 Cyanine 5

DBD DNA-binding domain

DNA deoxyribonucleic acid

DNMT DNA methyltransferase

EMSA Electrophoretic Mobility Shift Assay

ESC embryonic stem cell

FAM 6-Carboxyfluorescein

fC 5-formylcytosine

FDR false discovery rate

G guanine

HEK human embryonic kidney

hmC 5-hydroxymethylcytosine

LC liquid chromatography

LFQ label-free quantification

LINE1 long Interspersed Nuclear Element-1

MBD methyl-CpG binding domain

MBP maltose binding protein

mC 5-methylcytosine

MECP2 methyl-CpG-binding protein 2

MS mass spectrometry

MUC1 mucin 1

NGS next generation sequencing

NuRD nucleosome remodeling and deacetylase

PCA principal component analysis

PRC polycomb repressor complex

PTM post-translational modification

RFX regulatory Factor binding to the X-box

RNA ribonucleic acid

SAM S-adenosylmethionine

SELEX Systematic Evolution of Ligands by EXponential Enrichment

SP1 specificity protein 1

T thymine

TDG thymine DNA glycosylase

TET ten-eleven translocation dioxygenase

TF transcription factor

TSS transcription start site

TTS transcription termination site

VEGFA vascular endothelial growth factor A

1 Introduction

1.1 Epigenetic regulation through cytosine modifications

Genetic information in all living organisms is stored in DNA as a sequence of the four nucleobases A, T, C and, G (Alberts et al., 2022). This sequence encodes genes for building and maintaining the organism, which are transcribed and translated into proteins that carry out essential cellular functions. However, not all genes are needed at all times. Selective gene expression is reflected by exhibition of distinct cell phenotypes, despite nearly every cell in an organism containing the same DNA sequence. Moreover, during development, only a subset of genes is active at a given stage, leading to the formation of specialized tissues. Similarly, in disease states, aberrant activation or repression of genes leads to disruption of cellular balance causing pathological outcomes.

The mechanisms determining when, where and to which extent genes are expressed exceed the DNA sequence itself. Conrad H. Waddington introduced the term epigenetics in 1942 to describe the field implicated in understanding "causal interactions between genes and their products, which bring the phenotype into being" (Waddington, 1942). Today, the term 'epigenetics' refers to several key regulatory mechanisms, including DNA modifications, histone modifications, chromatin remodeling, and non-coding ribonucleic acids (ncRNAs), which collectively cause dynamic and heritable changes in gene expression (**Figure 1.1a**).

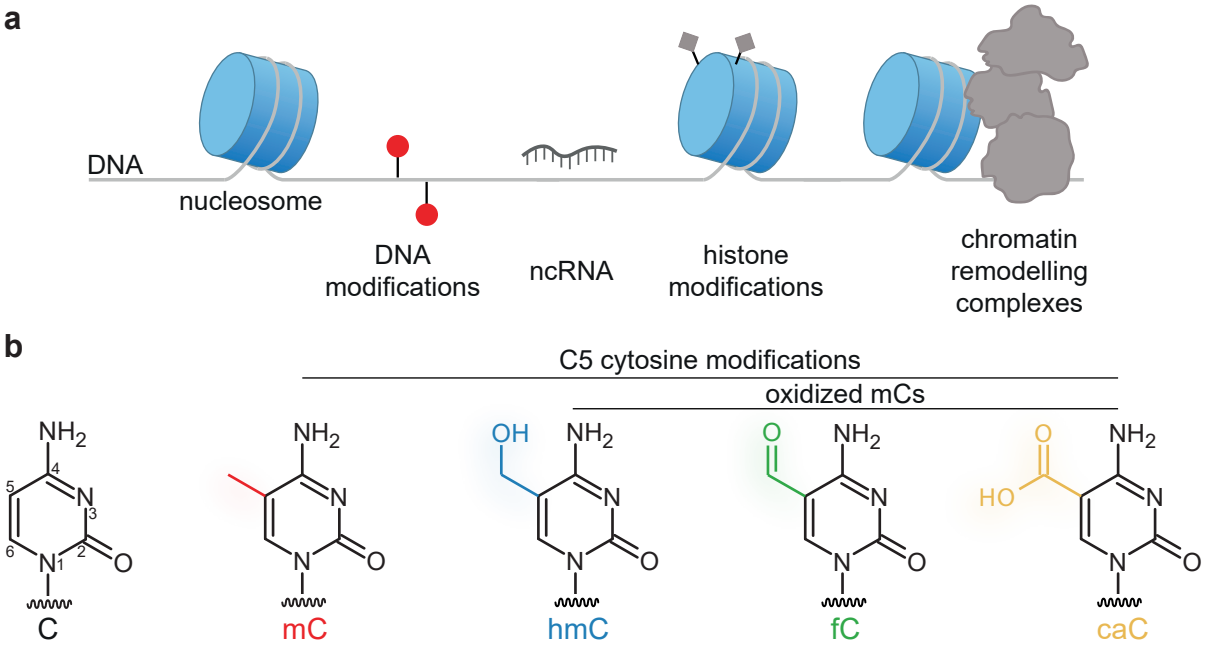


Figure 1.1: Epigenetic regulatory mechanisms. (a) Schematic representation of the key epigenetic regulatory mechanisms comprising DNA modifications, ncRNA, histone modifications, and chromatin remodeling complexes. **(b)** C and its 5'-modified derivatives mC, hmC, fC, and caC. The latter three are oxidation products of mC.

Methylation of cytosine at the 5th carbon (5-methylcytosine, or 5mC, hereafter mC) is the most prevalent DNA modification in the mammalian genome (Law and Jacobsen, 2010). It is often referred to as the "fifth base" owing to its central role in gene regulation (Greenberg and Bourc'his, 2019). Building on previous biochemical work, Wyatt ultimately confirmed mC as an additional component to the genetic system in 1950 (Wyatt, 1950). Today, mC is acknowledged as a key epigenetic signal. Occurrence of mC within promoter regions is commonly associated with transcriptional repression of the corresponding gene. In recent years, the field has broadened beyond mC by discovery of its oxidized derivatives 5-hydroxymethylcytosine (5hmC, hereafter hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009), 5-formylcytosine (5fC, hereafter fC), and 5-carboxylcytosine (5caC, hereafter caC) (He et al., 2011; S. Ito et al., 2011; Pfaffeneder et al., 2011) (**Figure 1.1b**). Among these, hmC has emerged as a particularly intriguing DNA modification due to its relative stability and tissue-specific enrichment, especially in the brain.

In the following sections, the chemistry and biology of mC and its oxidized derivatives will be explored, with a special focus on hmC. This includes an examination of how the dynamic methylation landscape is shaped, the genomic distribution of these modified bases, and their functional implications.

1.1.1 mC and its oxidative derivatives hmC, fC and caC

mC is the most prevalent and well-studied DNA modification in mammals and is predominantly found at CpG (cytosine-guanine sequences connected by a phosphate backbone) dinucleotides, which are key genomic elements for gene regulation (Z. D. Smith and Meissner, 2013). The human genome contains around 28 million CpG sites (Lister et al., 2009), that each exist in either a modified or unmodified state, giving rise to a dynamic regulatory network. Indeed, 60 - 80 % of all CpG sites of somatic cells in the mammalian genome are methylated (Z. D. Smith and Meissner, 2013). CpG dinucleotides are not uniformly distributed among the genome as the majority of the genome is CpG depleted, containing fewer than one-quarter of CpG sites that would be expected by random sequence composition. However, regions referred to as CpG islands (CGIs) have the expected CpG density. They are 0.5-2 kb in length, GC rich and usually located at gene promoter regions (Bird, 1987). CGIs are often found unmethylated assumingly supporting active gene transcription. Notably, there is an inverse correlation between CpG density and mC levels as high CpG density regions tend to stay unmethylated and low density regions are rather methylated (Bird, 1980). Among the different nucleobase modifications, cytosine modifications at the C5 position, namely mC and its oxidized derivatives, are particularly prominent in mammals due to their implications in transcriptional regulation, development, and disease. Regarding the oxidized forms of mC, hmC has gained considerable interest due to its potential role as a stable epigenetic mark and as a candidate for the "sixth base" of genome. The subsequent oxidation products fC and caC were generally viewed as intermediates of the active DNA demethylation pathway. Notably, the global methylation landscape is not static but it is dynamically regulated through the action of specialized enzymes that write, erase, and interpret these cytosine modifications. The following sections will examine the key enzymes that shape this epigenetic code.

The writers of DNA methylation

The DNMTs family catalyzes the transfer of a methyl group from SAM to cytosine, primarily at CpG dinucleotides and are essential for establishment and maintenance of DNA methylation (**Figure 1.2a**). The evolutionary history of DNMTs revealed that eukaryotes acquired DNMTs by horizontal gene transfer from bacterial cytosine methyltransferases, which are part of the restriction-methylation system (T. H. Bestor, 1990) and have since diverged in structure and function (Zilberman, 2008). Eukaryotic

DNMTs share a conserved catalytic domain essential for DNA binding, SAM recognition and catalysis (X. Cheng et al., 1993; Kumar et al., 1994). Six families have been reported (DNMT1, DNMT2, DNMT3, DNMT4, DNMT5, and DNMT6) (Huff and Zilberman, 2014; Ponger and Li, 2005), with DNMT1 and DNMT3 being the major enzymes responsible for maintenance and de novo methylation, respectively (Greenberg and Bourc'his, 2019).

Methylation maintenance by DNMT1. DNMT1 was the first eukaryotic DNA methyltransferase to be discovered in 1988 and is essential for the maintenance of DNA methylation patterns during cell division (T. Bestor et al., 1988). It preferentially recognizes hemimethylated CpG sites (Hermann et al., 2004) following DNA replication and restores symmetrical methylation, thereby supports the accurate transmission of epigenetic information across cell generations (Yoder, Soman, et al., 1997). Knockout studies of DNMT1 have shown its critical involvement in transposon silencing, X-inactivation and imprinting (G.-L. Xu et al., 1999). DNMT1 uses a base-flipping mechanism by rotating the target cytosine out of the DNA helix to allow methyl group transfer. The preference for hemimethylated CpG sites is driven by its interaction with UHRF1, a cofactor that recognizes hemimethylated DNA and recruits DNMT1 to replication forks (Bostick et al., 2007), hence DNMT1 is poorly active on unmethylated DNA. Structural studies have shown that flanking sequences significantly influence DNMT1 activity (J. Song et al., 2012). An about 100-fold difference in methylation rates at NNCGNN sites was observed depending on the flanking bases at position ±1 and ±2. These preferences are attributed to structural alterations in DNA and base-stacking interactions that stabilize or destabilize the active conformation of DNMT1 (Adam et al., 2020).

De novo methylation by DNMT3A and DNMT3B. DNMT3A and DNMT3B were discovered in 1998 and, in contrast to DNMT1, are capable of catalyzing de novo DNA methylation independent of DNA replication, enabling methylation at previously unmethylated sites during embryonic development, differentiation, and diseases (Okano et al., 1999). Their biological importance is highlighted by mutations in DNMT3B causing Immunodeficiency, Centromeric instability, and Facial anomalies (ICF) syndrome (G.-L. Xu et al., 1999). DNMT3A and DNMT3B were initially thought to lack sequence specificity and that specific interactions originated from interactions with transcription factors (Hervouet et al., 2009, 2018), such as c-MYC (Brenner et al., 2005). Subsequent studies, however, have demonstrated that both enzymes display flanking sequence preferences around CpG sites (L. Gao et al., 2020). For instance, DNMT3A preferentially methylates sites with C at the +1 position, while DNMT3B prefers G or A. Both enzymes exhibit a T preference at the -2 position. In addition to CpG methylation, DNMT3A and DNMT3B can also methylate non-CpG sites, particularly CpA, followed by CpC (for DNMT3A) and CpT (for DNMT3B). Remarkably, methylation efficiency at CpG sites with unfavorable flanking contexts can be lower than that at non-CpG sites in favorable contexts (L. Gao et al., 2020). Similar results were obtained in other studies (Mallona et al., 2021; S.-Q. Mao et al., 2020). Like DNMT1, DNMT3A and DNMT3B utilize a base-flipping mechanism to access the cytosine base. Structurally, DNMT3 enzymes form heterotetrameric complexes with DNMT3L, which is a non-catalytic cofactor that stimulates de novo methylation (Jia et al., 2007; Z.-M. Zhang et al., 2018). Two DNMT3A/3B subunits are arranged centrally with DNMT3L molecules on the edges, enabling cooperative binding to two CpG sites at 12 bp spacing (L. Gao et al., 2020), which enhances activity and specificity. More recently, it was reported that DNMT3 activity can also be stimulated by UHRF1, which was previously thought to primarily interact with DNMT1 (Yamaguchi et al., 2024).

Other DNMTs. While DNMT1 and DNMT3 enzymes govern epigenetic regulation in mammals, other DNMTs have not been as extensively studied or contribute to methylation-related processes in various

organisms. DNMT2, also known as tRNA aspartic acid methyltransferase 1 (TRDMT1), is not a DNA methyltransferase but mediates tRNA methylation at position C38 and is involved in RNA stability, stress responses, and cell proliferation in humans, mice, plants, and insects (Jeltsch et al., 2017; Schaefer et al., 2010). DNMT5 is a recently characterized ATP-dependent maintenance methyltransferase identified in the fungus *Cryptococcus neoformans* (Huff and Zilberman, 2014). It is the only DNMT to comprise an SNF2 ATPase domain, requiring ATP for DNA methylation (Catania et al., 2020; Dumesic et al., 2020). This domain acts in a chaperone-like, enzyme-remodeling manner leading to high-fidelity methylation of hemimethylated DNA (Dumesic et al., 2020; J. Wang et al., 2022). DNMT5 is vital for epigenetic memory in fungi and is critical for virulence in pathogenic species, despite the absence of a de novo methyltransferase in these organisms (Catania et al., 2020). Other fungal DNMTs, such as DNMT4, are implicated in DNA methylation related processes known as Repeat-Induced Point Mutation (RIP) and Methylation Induced Premeiotically (MIP), which contribute to silence or mutate transposable elements (Malagnac et al., 1997). DNMT6 has been found in several unicellular eukaryotes such as dinoflagellates and algae (Ponger and Li, 2005) and has yet to be functionally characterized (Hoguin et al., 2023).

Active DNA demethylation by TET Oxidation and TDG-Mediated Repair

TET enzymes and oxidative demethylation. Despite mCs chemical stability from its carbon-carbon bond and genetic stability via DNMT1 maintenance during DNA replication, DNA methylation is reversible by restoration of unmodified C via passive and active demethylation (**Figure 1.2b**).

Passive demethylation describes the loss of mC through DNA replication by lack of efficient methylation maintenance leading to dilution of mC over multiple rounds of cell divisions. Active demethylation is, in contrast, replication independent and subject to enzymatic mechanisms. The TET family of dioxygenases catalyzes the iterative oxidation of mC to hmC, then to fC, and finally to caC (He et al., 2011; S. Ito et al., 2011; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). fC and caC are then excised by TDG, and the resulting abasic sites are repaired through BER, which finally restores unmodified C (He et al., 2011; Maiti and Drohat, 2011; Weber et al., 2016). The TET family, comprising TET1, TET2 and TET3, was first characterized in 2010 (S. Ito et al., 2010). They are iron(II)/α-ketoglutarate (Fe(II)/α-KG) dependent dioxygenases, comprised of a conserved double-stranded β-helix (DSBH) domain that assembles Fe(II), α-KG and mC for oxidation and a cysteine-rich domain for stabilization (L. Hu et al., 2013) (**Figure 1.2c**). Studies on substrate preference have revealed that TET2 preferentially oxidizes mCpG over mCpA and mCpC, likely due to impaired base stacking (L. Hu et al., 2013). Both, TET1 and TET2, display a substrate preference of mC > hmC > fC consistent with their kinetic conversion rates and the energetics of hydrogen abstraction required for each oxidation step (L. Hu et al., 2015; S. Ito et al., 2011) TET processivity can be assessed at three levels. While TET enzymes are not physically processive, meaning that they do not oxidize all CpG sites on one DNA molecule, they may act chemically processive and depending on reaction conditions either release the substrate after each oxidation step or oxidize mC through to caC without dissociating (Crawford et al., 2016; Tamanaha et al., 2016). Genetically, processivity varies across the genome and correlates with chromatin accessibility and the presence of transcription factor binding sites (H. Wu et al., 2014). TET1 is enriched at CpG islands, active promoters and bivalent promoters that carry both H3K4me3 and H3K27me3. This preference is suspected to be mediated by the CXXC domain of TET, which preferentially binds CpG-rich regions (Y. Xu et al., 2012). For TET2, which lacks the CXXC domain, recruitment is partially mediated via its interaction partner IDAX (Ko et al., 2013.). The recruitment to specific genomic regions can also be facilitated by TET interaction partners. In mouse embryonic stem cells (mESCs), TET proteins interact with NANOG, PRDM14, PRC2 and LIN28A (Costa et al., 2013; Neri et al., 2013; Okashita et al., 2014;

Pantier et al., 2020; Zeng et al., 2016). In AML cells, TET2 is recruited by WT1 (Y. Wang et al., 2015), while PU.1 mediates TET2 recruitment during monocyte-to-osteoclast differentiation (de la Rica et al., 2013). TET3 has been shown to interact with REST in the mouse retina (Perera et al., 2015) and during fibroblast-to-adipocyte transdifferentiation, TET-dependent demethylation occurs around PPARγ and CTCF binding sites, which may be mediated by interaction with these transcription factors (Dubois-Chevalier et al., 2014; Fujiki et al., 2013). Hepatic TET3 was shown to be recruited by FOXA2 to the promoters of the fetal version of HNF4α, ultimately linking it to type-2 diabetes (D. Li et al., 2020).

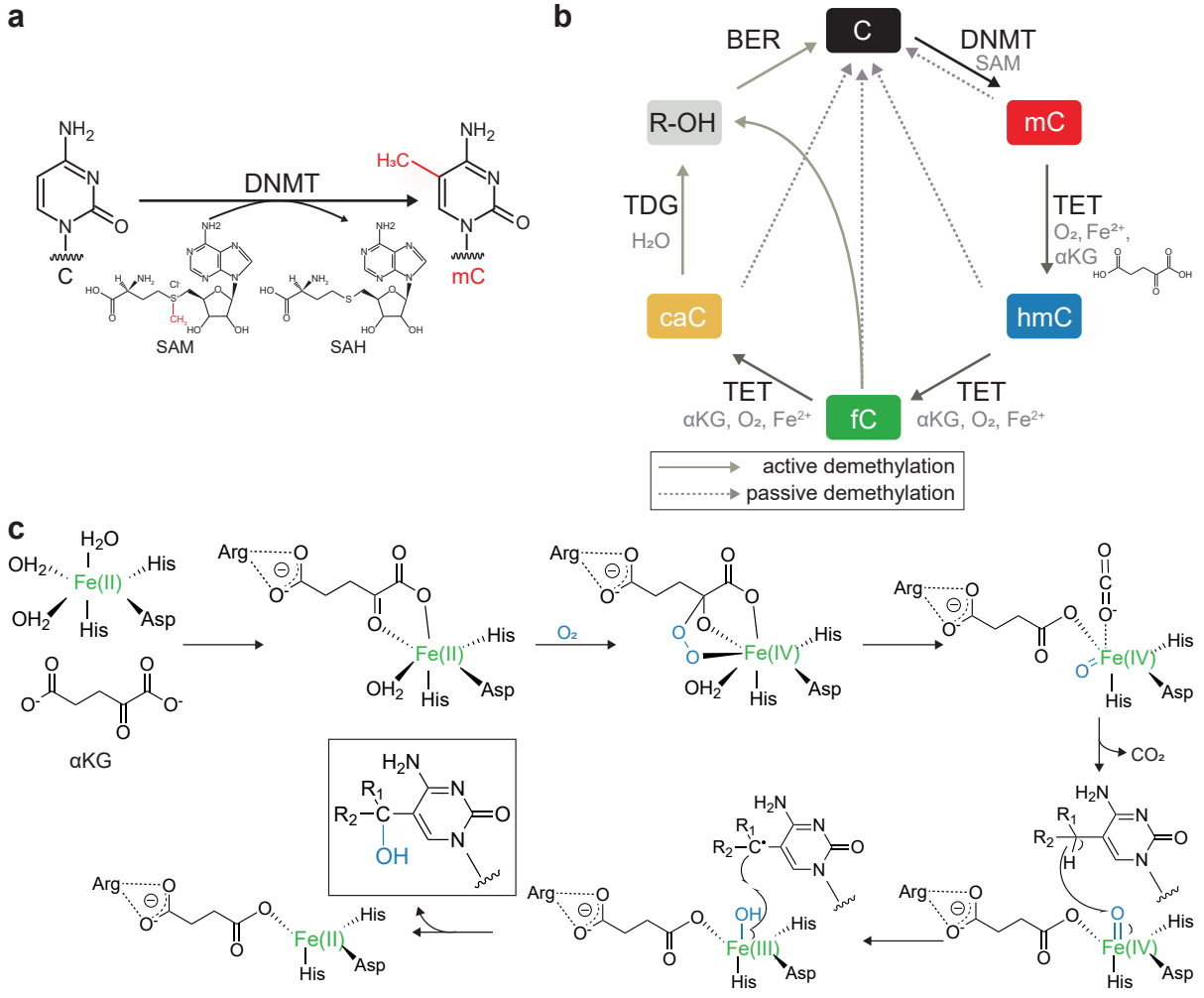


Figure 1.2: Establishment and removal of mC. (a) DNA methylation is established by DNMT enzymes by transferring a methyl group from SAM to cytosine generating mC. **(b)** After establishing, mC can be removed either by active or passive demethylation. Passive demethylation occurs through replication-dependent dilution, while active demethylation is catalyzed by TET enzymes that oxidize mC stepwise to hmC, fC, and caC. fC and caC can be excized by TDG and subsequently repaired via BER. **(c)** Mechanistically, TET enzymes require Fe(II) and α-KG as cofactors. A conserved arginine of TET coordinates the binding of the cofactors. Addition of molecular oxygen results in oxidative decarboxylation of α-KG generating the reactive Fe(IV)=O intermediate, which abstracts a hydrogen atom from the methyl group of cytosine, forming a radical intermediate that is then hydroxylated to yield the oxidized cytosine derivative.

TDG and BER. The final step of active demethylation includes the removal of fC and caC by TDG (He et al., 2011; Maiti and Drohat, 2011). TDG excises these bases and generates an abasic site, which is then subject to the BER machinery. AP endonuclease I cleaves the backbone to generate a single strand break, polymerase β inserts an unmodified cytosine and DNA ligase 3 together with XRCC1 ligate the nick to restore the DNA duplex (Weber et al., 2016) (**Figure 1.3**). In mESCs, the genomic distribution of

TDG is similar to that of TET1, being enriched at active promoters and enhancers (Neri et al., 2015). This co-localization likely results from the direct interaction with TET enzymes (Müller et al., 2014; Weber et al., 2016) and its recruitment by accumulation of fC and caC (Iurlaro et al., 2013; Spruijt et al., 2013).

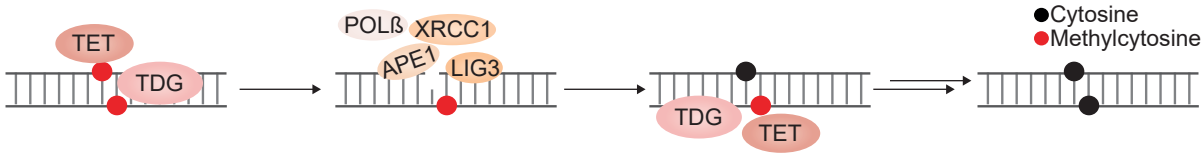


Figure 1.3: The TDG-BER pathway. TET oxidizes mC to fC or caC, which are then recognized and excised by TDG. The resulting abasic site is then repaired by BER. First, AP endonuclease I (APE1) mediated backbone cleaving at the abasic site, followed by insertion of unmodified cytosine by DNA polymerase β beta and subsequent nick ligation by DNA ligase III (LIG3) in complex with XRCC1. Usually this repair occurs sequentially, first on one DNA strand, followed by the second strand.

1.1.2 Detection and distribution of oxidized cytosine modifications

While mC is generally regarded as a relatively stable DNA modification, its oxidized derivatives hmC, fC, and caC show highly dynamic and tissue-dependent distributions. They are particularly variable in the brain and ESCs, indicating their involvement in context-dependent regulatory processes. Understanding their biological functions requires quantification of their abundance and determination of their genomic localization. MS-based approaches have been used to measure global levels of cytosine modifications across tissues but they lack spatial resolution. This limitation has been resolved by the development of sequencing technologies capable of mapping cytosine modifications across the genome. In this section, an overview of these sequencing methods will be provided, followed by an examination of the prevalence and genomic distribution of oxidized cytosine modifications.

Methods for mapping cytosine modifications

A plethora of sequencing methods has emerged to map the genomic distribution of modified Cs, which can be separated into non–base-resolution and base-resolution techniques.

Non–base-resolution methods enrich DNA fragments that contain specific modifications followed by high throughput sequencing to identify their approximate genomic locations. These include enrichments with modification-specific antibodies, such as methylated DNA immunoprecipitation (MeDIP), hydroxymethylated DNA immunoprecipitation (hMeDIP), as well as fC and caC DIP strategies ((Mohn et al., 2009; Nestor and Meehan, 2014; L. Shen et al., 2013) (**Figure 1.4a**). Additionally, chemical labeling strategies, like GLIB-seq and CMS-seq, hMe-Seal, fC-Seal, and fC-DP were applied for enriching DNA fragments that contain the modification of interest (Pastor et al., 2011; Raiber et al., 2012; C.-X. Song et al., 2011, 2013). While these methods aid in large-scale pattern detection, they have low spatial resolution and have limitations such as off-target reactivity of antibodies, CpG density related bias, and limited quantitative accuracy.

Base-resolution methods provide single-nucleotide resolution yielding the exact position and frequency of modified cytosines. Most are based on bisulfite sequencing, in which sodium bisulfite treatment deaminates unmodified cytosines, fC, and caC, converting them to uracil, which is then read as thymine after PCR amplification, while mC and hmC are protected and read as cytosine (Booth et al., 2012; He et al., 2011) (**Figure 1.4b**). Differentiating between the oxidized bases requires chemical or enzymatic treatments. For instance, TAB-seq specifically identifies hmC and M.SssI methylase-assisted bisulfite sequencing allows for identification of fC and caC (H. Wu et al., 2014; Yu et al., 2012), while oxBS-seq, fCAB-seq, caCAB-seq and CLEVER seq allow to distinguish between mC, hmC, fC, and caC

either directly or by comparison with untreated bisulfite sequencing data (Booth et al., 2012; X. Lu et al., 2015; C. Zhu et al., 2017). While these methods offer high accuracy, they are usually high in costs due to a lack of pre-enrichment steps.

Additionally, enzymatic methods have been employed, which are based on cytosine-modification sensitive restriction enzymes (Petterson et al., 2014; Sérandour et al., 2016; Sun et al., 2015). Furthermore, third-generation sequencing platforms are advancing such as Nanopore and SMRT sequencing, which can directly detect modifications without chemical pre-treatments (Flusberg et al., 2010; Laszlo et al., 2013).

The existing methods mostly lack strand-specific information, which is becoming particularly relevant in the study of modified CpG dyads. Recent developments such as in third generation sequencing technologies, have started to address this limitation by enabling the detection of asymmetric cytosine modification patterns. A more detailed discussion of these technologies and their potential will be provided in **Subsection 1.3.2**.

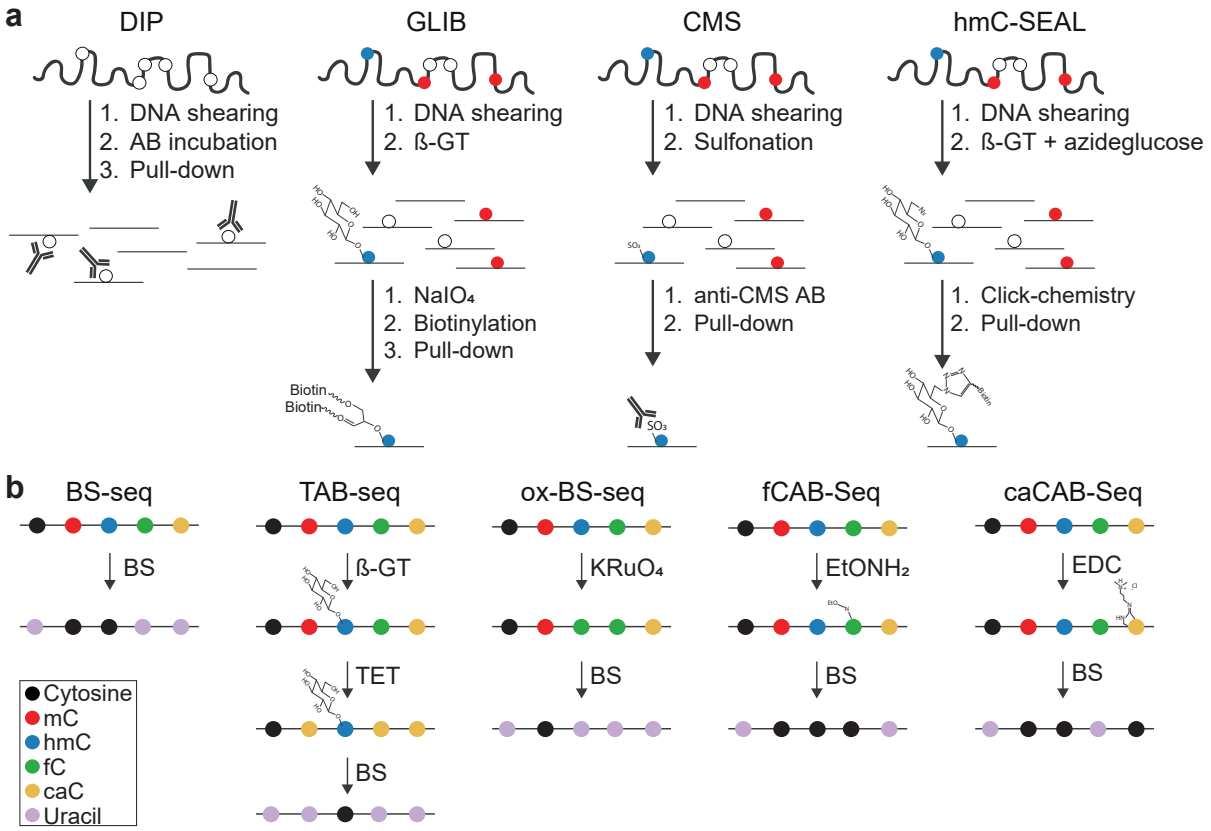


Figure 1.4: Sequencing methods to assess cytosine modifications. **(a)** Non-base resolution enrichment-based sequencing methods include DNA immunoprecipitation (DIP) sequencing with modification-specific antibodies (AB), GLIB-sequencing which involves glucosylation of hmC, sugar oxidation and biotin-based affinity enrichment, CMS-sequencing, where hmC is chemically converted to CMS for antibody enrichment, and hmC-SEAL, which uses β-glucosyltransferase to transfer an azide-modified glucose to hmC, followed by click-chemistry and affinity purification. **(b)** Base-resolution sequencing methods include bisulfite sequencing (BS). In classical bisulfite sequencing C, fC, and caC are deaminated to uracil and read as thymine, while mC and hmC are inert to bisulfite treatment and are read as cytosine. In TAB-sequencing, hmC in protected by glucosylation before TET is added to oxidize mC to caC. hmC remains unconverted and is read as cytosine. In oxBS, hmC is selectively oxidized to fC with $KRuO_4$ resulting in only mC being read as cytosine. fCAB- and caCAB-sequencing specifically modify fC or caC to protect them from deamination enabling their base-resolution identification.

.

Tissue and genomic distribution of oxidized cytosine modifications

Tissue distribution. Quantitative analyses across mammalian tissues have shown that while mC levels remain relatively stable, the levels of its oxidized derivatives are highly variable between tissues (Bachman et al., 2014, 2015; Globisch et al., 2010; S. Ito et al., 2011). **Table 1.1** summarizes the abundance of these modifications across representative tissues.

hmC is most abundant in the central nervous system, reaching up to 0.7 % of cytosines in neurons, which corresponds to up to 40 % of mC levels in specific neuronal contexts (Globisch et al., 2010; S. Ito et al., 2011; Kriaucionis and Heintz, 2009). Intermediate hmC levels are observed in organs like liver and kidney, while rapidly proliferating tissues such as spleen and thymus show only trace levels (Bachman et al., 2014; Globisch et al., 2010; S. Ito et al., 2011). ESCs exhibit enhanced but variable hmC levels, reflecting active DNA demethylation dynamics (Globisch et al., 2010; Szwagierczak et al., 2010), consistent with reports on two waves of hmC accumulation during zygote development (D. Bai et al., 2025). In contrast, fC and caC are typically detected at much lower levels due to their rapid excision by TDG, although they become more abundant upon TDG depletion (L. Shen et al., 2013).

Table 1.1: Genomic prevalence of cytosine modifications. Indicated as percentage of all cytosines. a: Gama-Sosa et al., 1983, b: Globisch et al., 2010, c: W. Li and Liu, 2011, d: Kriaucionis and Heintz, 2009, e: Booth et al., 2012

Tissue, Cell	mC	hmC	fC and caC
Brain, Neurons	4 - 5 %[a, b], 0.6 - 1.5 %[c]	0.4 - 0.7 %[b, c, d]	< 0.02 %
Kidney, Heart, Liver	4 - 5 %[b]	0.1 - 0.4 %[b, c]	ND
Thymus, Spleen	4 - 5 %[b]	< 0.06 %[b, c]	ND
ESCs	3 - 4 %	0.2 - 1.0 %[e]	< 0.01 - 0.02 %

Genomic distribution of hmC. In ESCs, hmC is enriched at bivalent promoters, that are poised for activation during differentiation (Pastor et al., 2011; H. Wu et al., 2011; Yu et al., 2012) and at gene bodies, especially of low or moderately expressed genes, but it is low at promoters of these genes. It is also enriched at distal regulatory elements (Stroud et al., 2011; Williams et al., 2011; H. Wu et al., 2011; Yu et al., 2012). hmC levels are typically low at highly expressed gene promoters and hmC also tends to be depleted at CpG-rich promoters, especially those marked by H3K4me3 (Williams et al., 2011; H. Wu et al., 2011; Y. Xu et al., 2011), but is enriched in regions of intermediate CpG density and intermediate methylation (Stadler et al., 2011). Moreover, hmC levels are low at TF binding sites (± 100 bp) but enriched in flanking regions, suggesting a potential role in shaping local chromatin accessibility (Yu et al., 2012).

In neurons, hmC is depleted at transcription start sites (TSS) regardless of gene expression level and CpG content but it accumulates strongly within gene bodies, where its levels are positively correlated with gene expression (Mellén et al., 2012; C.-X. Song et al., 2011). hmC is also found at intron-exon boundaries and active enhancers. These patterns suggest that in neurons, hmC may have a role in splicing and may support stable gene expression (Lister et al., 2013; Ruzov et al., 2011; Wen et al., 2014). Additionally, hmC accumulates with age in neuronal genomes and marks active or poised enhancers across cell types (Sérandour et al., 2012; Stroud et al., 2011). TET2-mediated oxidation of mC at enhancers may prime or maintain regulatory activity by recruiting activators or excluding repressor complexes (Hon et al., 2014).

Genomic distribution of fC and caC. fC and caC are much less abundant than hmC, due to the high efficiency of TDG-mediated removal during active demethylation. They accumulate under TDG depletion, revealing their presence in ESCs at distal regulatory elements, enhancers and bivalent promoters

(L. Shen et al., 2013; H. Wu et al., 2014). They are further enriched in CGIs and Pol II binding sites, where fC enrichment correlates with active transcription and H3K4me3 marks (Raiber et al., 2012). Upon TDG knockdown, significant increases in fC and caC at these elements highlight their dynamic turnover (L. Shen et al., 2013). Interestingly, regions with higher fC and caC levels relative to hmC, like OCT4 and SOX2 binding sites, suggest high processivity of TET enzymes and selective turnover, further suggesting oxidized cytosines to be functionally distinct regulatory marks (L. Shen et al., 2013).

1.1.3 Biological functions and implications of cytosine modifications

DNA methylation is involved in regulating gene expression by modulating the accessibility of DNA to transcription factors and other DNA-binding proteins. It plays a critical role in X chromosome inactivation by silencing one of the two X chromosomes in female mammals which ensures dosage compensation (Riggs and Pfeifer, 1992). Moreover, it is important in genomic imprinting by marking the maternal or paternal allele to enable monoallelic expression (J. R. Mann et al., 2000). Furthermore, it is essential for suppression of germline-specific genes in somatic cells to prevent inappropriate gene activation outside of the germline (P. A. Jones, 2012). DNA methylation is also important beyond the regulation of protein-coding genes, as it is essential for genomic integrity, which is achieved by silencing of repetitive elements, including transposable elements and endogenous retroviruses (Yoder, Walsh, and Bestor, 1997). Despite the importance of stable methylation maintenance, temporal and spatial dynamics of DNA methylation are tightly regulated. Two major waves of global methylation reprogramming occur during development. First, in primordial germ cells (PGCs), where the epigenetic state is resetted by largely erasing the methylation and second, after fertilization, where methylation patterns are again erased and then re-established during preimplantation development (Messerschmidt et al., 2014; Seisenberger et al., 2013). This warrants the removal of epigenetic patterns and is the basis for the establishment of lineage-specific arrangements. Beyond these drastic reprogramming events, changes in DNA methylation are usually more stable and occur during lineage specification and cellular differentiation. DNA methylation levels represent a dynamic steady-state which is determined by the local activities of DNMTs and demethylation processes (Jeltsch and Jurkowska, 2014). Moreover, site-specific transcription factors play a huge role in managing these processes by either excluding DNA methylation from their binding sites or promoting active DNA demethylation (Hahn et al., 2019; Stadler et al., 2012). Aberrant DNA methylation is implicated in a wide range of diseases, including cancer (H. Shen and Laird, 2013). While global hypomethylation may lead to genomic instability, hypomethylation of oncogenes and hypermethylation of tumor suppressor genes can contribute to oncogenesis (P. A. Jones and Baylin, 2002; Vilain et al., 1999; Wilson et al., 2007). Furthermore, mutations in the methylation machinery and TETs are associated with developmental and neurological disorders (Beck et al., 2020; Hutnick et al., 2009). Aberrant DNA methylation patterns are not only implicated in cancer but also in a variety of other conditions. These include neurodevelopmental disorders such as Rett syndrome, Rubinstein-Taybi syndrome, and Fragile X syndrome (Amir et al., 1999; Petrif et al., 1995; Sutcliffe et al., 1992), neurodegenerative diseases like Alzheimer's and Huntington's disease (Gräff and Mansuy, 2009), cardiovascular conditions such as atherosclerosis (Lund et al., 2004), as well as obesity, psychiatric disorders, immune dysfunctions, and the aging process (Renthal and Nestler, 2009; Suarez-Alvarez et al., 2012; Szulwach et al., 2011; X. Wang et al., 2010). As previously mentioned, there are multiple epigenetic mechanisms and DNA methylation is part of a huge epigenetic regulatory network encompassing histone modifications, ncRNAs and chromatin remodeling complexes. These systems interact with each other, for instance, by proteins that bind to DNA modifications and subsequently recruit histone modifying enzymes as is the case for MBD proteins (also see **Subsection 1.2.4**). While methylation is generally associated with gene repression, the implications of oxidized mCs are less well studied.

However, hmC, fC, and caC may carry distinct biological functions. Especially hmC is recognized as a stable and functional mark and found abundant in neuronal tissues where it correlates with active gene expression and enhancer activity (Pastor et al., 2011). Its association with gene bodies of actively transcribed genes and poised enhancers suggests it to play a role in transcriptional activation, in contrast to mC (Stroud et al., 2011). It was also reported to be involved in maintaining pluripotency in stem cells and guiding different trajectories (H. Wu et al., 2011). Prospects of the hmC mark are further discussed in **Subsection 1.3.4**. fC and caC were typically seen as transient bases that are recognized and removed by TDG. Given their relative higher abundance in ESCs, reports on higher stability of fC than initially expected (M. Su et al., 2016), and enrichment at developmental gene promoters (Santiago et al., 2014), they are assumed to be involved in developmental processes. Further insights into the oxidized variants can be retrieved by examining reader proteins that specifically bind to them. These interactions will be discussed in the following chapter.

1.1.4 Essential points

- Epigenetic regulatory mechanisms comprise DNA modifications, histone modifications, non-coding RNAs, and chromatin remodeling complexes.
- Methylation of cytosine at position 5 (mC) is the predominant DNA modification, also known as the fifth base. It can be oxidized to hmC, fC, and caC by TET enzymes.
- DNA methylation is established de novo by DNMT3A and DNMT3B and maintained during replication by DNMT1. It can be removed passively through replication-dependent dilution, or actively via TET enzyme-mediated oxidation of mC to hmC, fC, and caC. The latter two are recognized and excised by TDG and unmodified cytosine is restored by the base excision repair pathway.
- Genome-wide detection of DNA modifications uses enrichment-based techniques, bisulfite conversion-based single-nucleotide resolution methods, or third-generation sequencing that directly detects modified bases.
- The abundance and dynamics of these cytosine modifications vary by tissue, genomic context, and developmental context. While mC is generally stable, hmC, fC, and caC show variable abundance. hmC is enriched in the brain, and all three oxidized derivatives are more abundant in embryonic stem cells. Also the genomic distribution of cytosine modifications differs. hmC is enriched at bivalent promoters, gene bodies, and distal regulatory elements, but depleted at CpG-rich promoters and promoters of highly expressed genes. In contrast, fC and caC are enriched at enhancers, distal regulatory elements, bivalent promoters, and CpG islands.
- Functionally, mC is primarily associated with transcriptional silencing and genome stability. hmC is linked to gene regulation, and may serve as a stable epigenetic mark in certain contexts. fC and caC, beyond being intermediates in demethylation, are increasingly recognized as functional marks involved in transcriptional regulation, enhancer activity, and developmental gene expression.

1.2 Interpretation of epigenetic cytosine modifications by proteins

Cytosine modifications are tightly regulated by the opposing enzymatic activities of DNMT and TET enzymes. Although mC and hmC are relatively stable and abundant, fC and caC are also measurably present. Remarkably, these modified cytosines show distinct genomic distributions and play crucial roles in gene regulation. However, the mechanisms underlying their biological impact are still being uncovered. This raises the central question of how these chemical modifications of cytosine translate into functional outcomes. Cytosine modifications can alter the mechano-physical properties of the DNA double helix and influence how DNA is recognized and interpreted by proteins. Both the base read-out and the shape read-out mechanisms are sensitive to structural variations in DNA. In this chapter, the effects of cytosine modifications on DNA structure and on protein binding behavior are explored, emphasizing their significance as regulatory elements in epigenetic gene regulation.

1.2.1 Influence of cytosine modifications on the DNA double helix structure and the major groove

mC is generally associated with an increase in the thermal stability of DNA, which is concluded from a higher melting temperature (T_m) compared to unmodified cytosine. Oxidation of mC to hmC reverses this effect, leading to decreased T_m values and, in some cases, even to greater destabilization than unmodified cytosine itself (Lercher et al., 2014; López et al., 2012). Despite extensive research, the impact of cytosine modifications on DNA structure is still highly debated. Globally, modified cytosines do not seem to disrupt the canonical B-DNA conformation. However, locally, several studies have shown that sequence-dependent structural alterations can appear near the modification site (Hardwick et al., 2017; Renciuk et al., 2013; Wanunu et al., 2011). Cytosine methylation has been shown to change the geometry of base pair steps, which leads to an increased local curvature and decreased flexibility attributed to steric hindrance from the methyl group in the major groove (Pérez et al., 2012). Additionally, mC influences the stability of nucleosomes, thereby affecting local chromatin structure and the accessibility of TFs. Furthermore, groove geometry is influenced by cytosine modifications (**Figure 1.5a**). mC can slightly widen the major groove while narrowing the minor groove (Machado et al., 2015) as opposed to hmC, fC and caC, which lead to increased local DNA flexibility and alterations in the minor groove geometry (Fu et al., 2019; Wanunu et al., 2011). hmC was reported to not induce major changes in chromatin structure relative to mC (Battistini et al., 2021). Assumingly, the increasing polarity of oxidized cytosine modifications leads to a reduction in DNA rigidity. The shift from hydrophobic mC to hydrophilic hmC and fC may contribute to destabilization of the duplex and thereby affect the structural dynamics, such as flexibility, rigidity and the solvent shell state which alters the local electrostatic environment (Wanunu et al., 2011). The electron withdrawing groups in fC and caC (CHO and COOH) leading to increased N3 acidity and therefore weakened hydrogen bonding and reduced base pair stability, likely facilitate the selective recognition by TDG (Dai et al., 2016). Cytosine modifications can also influence base stacking and hydrogen bonding interactions. While mC enhances base stacking due to increased molecular polarizability (Sowers et al., 1987), hmC, fC, and caC exhibit similar base stacking (Szulik et al., 2015). However, their impact on hydrogen bonding between base pairs varies. mC and hmC do not alter hydrogen bonding patterns, but fC and caC can form intranucleobase hydrogen bonds through their ketone and amino groups, which can lead to destabilization of the duplex (Fu et al., 2019; Helabad et al., 2014; Szulik et al., 2015).Cytosine modifications exposed in the major groove significantly alter its physicochemical signature (**Figure 1.5b-d**). These signature comprises hydrogen bond donors and acceptors, methyl groups, and nonpolar hydrogen atoms, which allow for distinct protein interactions via direct or water mediated hydrogen bonds and hydrophobic interactions (Rohs et al.,

2010; Seeman et al., 1976). For instance, guanine can be recognized by arginine via bidentate hydrogen bonding and thymine and cytosine are distinguished via hydrophobic interactions (Garvie and Wolberger, 2001; Harrison and Aggarwal, 1990). Consequently, the influence of cytosine modifications on DNA structure and major groove signature has important implications for protein-DNA interactions. Selective recognition of modified cytosines by proteins will be discussed in the upcoming sections.

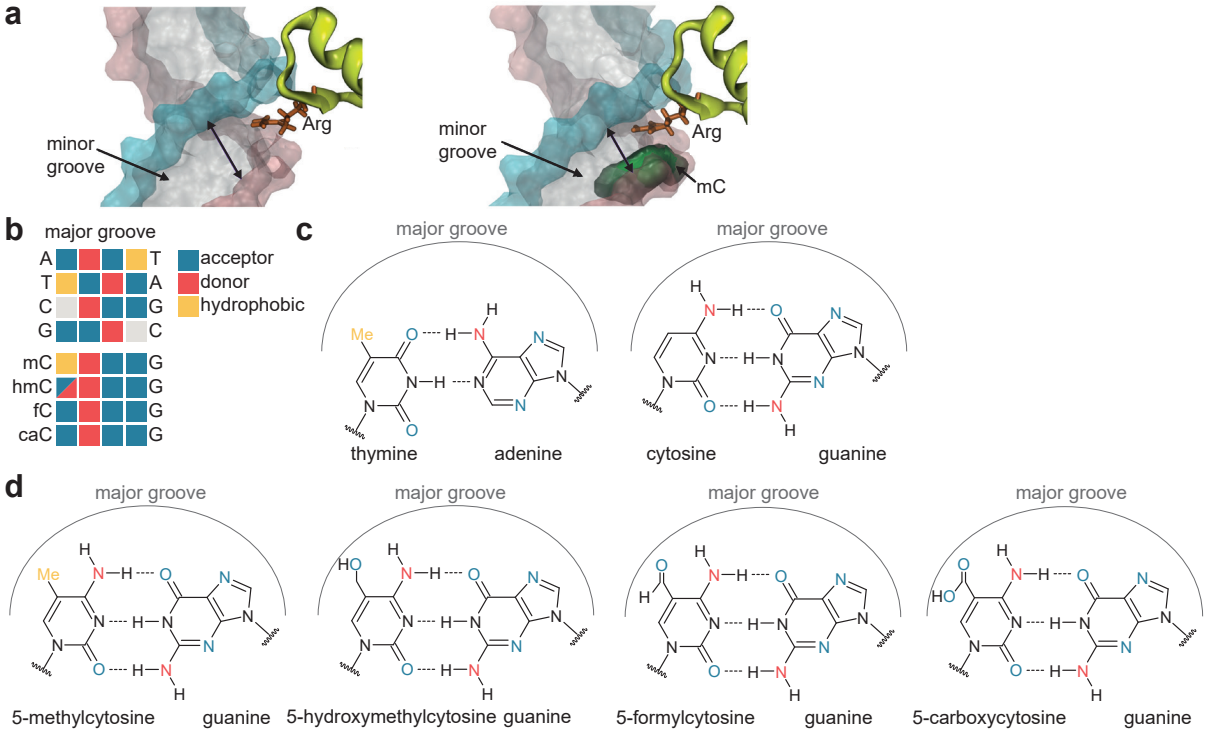


Figure 1.5: Influence of C modifications on minor and major grooves. (a) CpG methylation leading to a narrowed minor groove and widened major groove, enhancing the affinity of the DNA to the histone octamer. Figure (a) is adapted and modified from S. Li et al., 2022 CC BY 4.0, full license in appendix A.4. **(b)** Physicochemical signatures of base pairs in the DNA major groove showing H-bond acceptors (blue), H-bond donors (red), and hydrophobic groups (yellow) for different base pairs. **(c)** Standard Watson–Crick bp are shown and the major groove site is indicated. **(d)** Unique major groove signatures of modified cytosines are shown. Figures (b)-(d) are inspired by Chiu et al., 2023.

1.2.2 Transcription factors interpret the genomic information

TFs are DNA-binding proteins that play a fundamental role in gene regulation. They recognize specific DNA sequences, so called motifs, and they influence transcription by modulating chromatin accessibility and recruiting further proteins. Although many potential motifs have been predicted in silico, most of these sites are not bound in vivo, indicating complex and context dependent binding. This observation was named "futility theorem" (Wasserman and Sandelin, 2004). TFs constitute the first step in interpreting the DNA sequence, acting as master regulators and selector genes which control processes including cell type specification, development and immune responses (T. I. Lee and Young, 2013; H. Singh et al., 2014).

TFs rarely act alone but often bind cooperatively as homodimers (e.g. bZIP and bHLH families), heterodimers, or higher-order structures, enhancing DNA binding affinity, stability, and regulatory specificity. Moreover, these interactions can influence sequence preferences (Jolma et al., 2015; Slattery et al., 2011). TFs further compete with nucleosomes, which can lead to displacement or repositioning of nucleosomes (e.g. FOXA1) (Adams and Workman, 1995; Iwafuchi-Doi et al., 2016; Polach and Widom, 1996). To do this, they often recruit ATP-dependent chromatin remodelers (Swinstead et al., 2016). The

ability of TFs to bind nucleosomal DNA is also influenced by the rotational positioning of the binding site, as for the Yamanaka factors (POU5F1, SOX2, KLF4, and MYC) (Soufi et al., 2015). Therefore, TF binding specificity is determined not only by the DNA sequence but also by chromatin context and protein–protein interactions.

Most TFs contain a DBD, which was derived from a small set of common ancestors (Fulton et al., 2009; S. A. Lambert et al., 2018). The major families comprise C2H2 zinc fingers (ZFs), homeodomains, bHLH, bZIPs, and nuclear hormone receptors (NHRs) (P. F. Johnson and McKnight, 1989). Most mammalian TFs have at least one characterized DBD (Fulton et al., 2009), although novel domains continue to be found, such as the recent discovery of polycomb-like proteins preference for motifs containing CG dinucleotides (H. Li et al., 2017). In addition to DBDs, many TFs comprise effector domains that recruit cofactors, remodelers, or RNA polymerase directly (Frietze and Farnham, 2011). Classical activator sequences are often unstructured low-complexity regions, short linear motifs or DBDs that interact with cofactors (Brayer and Segal, 2008; Garza et al., 2009). However, many TFs lack a dedicated effector domain, especially in the bZIP and bHLH families. Instead, they act by steric mechanisms blocking other proteins from their binding sites (Akerblom et al., 1988).

The classification of TFs into activators and repressors is now perceived as being too simple as many TFs are capable of both activating and repressing gene expression depending on binding partners and local sequence context (Frietze and Farnham, 2011; Rosenfeld et al., 2006). For example, MAX has a repressor function when binding DNA as a heterodimer with MNT or MXD1, but operates as an activator upon heterodimerization with MYC (Amati and Land, 1994). Most human TFs recruit cofactors which then act as coativators or corepressors (Reiter et al., 2017). These are often multi-subunit complexes, which influence chromatin state through remodeling, histone modification, or higher-order structural changes (Frietze and Farnham, 2011). Different chromatin remodeling complexes exhibit preferences for specific DNA sequences and additionally add DNA-sequence specificity to TFs (Rippe et al., 2007). These complexes will be discussed in **Subsection 1.2.3**.

The human genome encodes over 1,600 TFs, corresponding to approximately 8 % of all protein-coding genes (S. A. Lambert et al., 2018) TF sequences, their regulatory regions and their physiological roles are highly conserved across metazoans, although a specific TF is also capable of regulating different genes in different cell types (Carroll, 2008; Gertz et al., 2012). Conservation of physiological roles is observed e.g. in HOX and RFX TFs, playing roles in body patterning and regulation of cilia genes (Bürglin, 2011; Choksi et al., 2014). Other TFs do evolve under diversifying selection and change their motifs or binding partners such as KRAB-ZFs, which target transposable elements (Arendt et al., 2016; Schmitges et al., 2016). Around half of the TF families show tissue-specific expression indicating specific physiological functions. Examples are SOX2 and POU3F2, which are expressed almost exclusively in the cerebral cortex, and GATA4 and TBX20 in cardiac muscle (S. A. Lambert et al., 2018).

1.2.3 Chromatin regulators determine the chromatin state

Central to chromatin regulation is the recruitment of multi-subunit chromatin-modifying and -remodeling complexes by TFs. TFs bind to DNA and guide these complexes to specific genomic loci through protein-protein interactions. There, these complexes modulate chromatin accessibility, histone modifications, and DNA modifications (Becker and Workman, 2013; Clapier et al., 2017). Chromatin regulators comprise histone modifiers, histone chaperones and chromatin remodelers (Lai and Pugh, 2017). Chromatin modifiers can be broadly separated into ATP-dependent and ATP-independent groups.

ATP-dependent chromatin remodelers. ATP-dependent remodelers are classical chromatin remodeling complexes, which use ATP hydrolysis to slide, evict, or restructure nucleosomes (Saha et al., 2002; Whitehouse et al., 2003). Thereby, they regulate access to the transcription machinery, as nucleosomes can impede RNA polymerase II (RNAPII). The four major families of ATP-dependent remodelers are named after their central ATPase and encompass switching defective/sucrose nonfermenting (SWI/SNF), imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), and INOsitol requiring 80 (INO80). Their function, however, is largely determined by attendant proteins that often contain domains such as DNA-binding motifs, PHD fingers, bromodomains, chromodomains, and TF-interaction sites (Becker and Workman, 2013; Flaus, 2006). For instance, SWI/SNF complexes are often recruited by TFs like p53 or Oct4, and often contain bromodomains that recognize acetylated histones (Awad and Hassan, 2008; D. Lee et al., 2002; Lei et al., 2020). In contrast, ISWI remodelers, bind unmodified H4 tails via SANT and SLIDE domains and help organize nucleosome spacing during transcription and replication by interacting with TFs. Attendant proteins have specialized domains such as PHD, bromodomains and additional DNA binding motifs (Boyer et al., 2004; Clapier, 2002). CHD/NuRD complexes combine ATP-dependent remodeling via CHD3/4 with HDAC activity (**Figure 1.6a**) and are often recruited via methyl-CpG-binding proteins (MBD2/3) and repressive TFs like REST. Attendant proteins have DNA binding domains and further domains such as PHD, BRK, CR1-3 and SANT. While some CHD remodelers promote transcription, others, like Mi-2/NuRD, are associated with transcriptional silencing, linking DNA methylation to histone deacetylation (Kehle, 1998; Torchy et al., 2015). INO80/SWR1 complexes, not only remodel nucleosomes but also exchange histone variants such as H2A.Z. They are frequently involved in DNA damage response or transcriptional activation (Papamichos-Chronakis et al., 2006).

ATP-independent chromatin regulators. Non-ATPase complexes like the Polycomb Repressive Complexes (PRC1 and PRC2) regulate long-term silencing through histone modifications (e.g., H3K27me3) and chromatin compaction. They are essential for maintaining cellular identity in higher eukaryotes (Schuettengruber et al., 2017). The PRC2 core is composed of the histone methyltransferases EZH1/2 together with SUZ12 or EED (Margueron and Reinberg, 2011) and different attendant proteins associate with core PRC2, modulating the recruitment and enzymatic activity of the complex (Laugesen et al., 2019). The PRC2 complex for instance catalyzes H3K27 trimethylation, which in turn recruits canonical PRC1 (cPRC1) (Boyer et al., 2006). However, PRC1 can also be recruited independently by KDM2B and MAX-mediated DNA interactions resulting in non canonical PRC1 (ncPRC1) (Farcas et al., 2012). PRC1 compacts chromatin and monoubiquitinates H2A via its RING-PCGF core that functions as a E3 ubiquitin ligase (McGinty et al., 2014; H. Wang et al., 2004). Six variants of PRC1 exist (PRC1.1 - 1.6), which are distinguished by their PCGF paralogues and associated accessory proteins (Z. Gao et al., 2012) (**Figure 1.6b**) .

Other non-ATPase complexes include histone chaperones, which are no true remodelers. However, they influence nucleosome assembly and chromatin organization during replication and transcription and hence may influence DNA accessibility. In chromatin assembly factor-1 (CAF-1), CHAF1b is the histone chaperone that complexes with CHAF1a and p48, which facilitates nucleosome assembly by delivering newly synthesized H3 and H4 dimers to replicating DNA. Besides, CHAF1b is also implicated in DNA repair, replication, and in maintaining cell fate decisions in some processes such as neuronal differentiation (Volk and Crispino, 2015). Other histone chaperones, like FACT (facilitates chromatin transcription) and Spt6 (suppressor of Ty), play essential roles during transcription elongation by mediating nucleosome disassembly and assembly (Duina, 2011).

Interestingly, chromatin complexes rarely operate in isolation. For example, SWI/SNF, NuRD, and

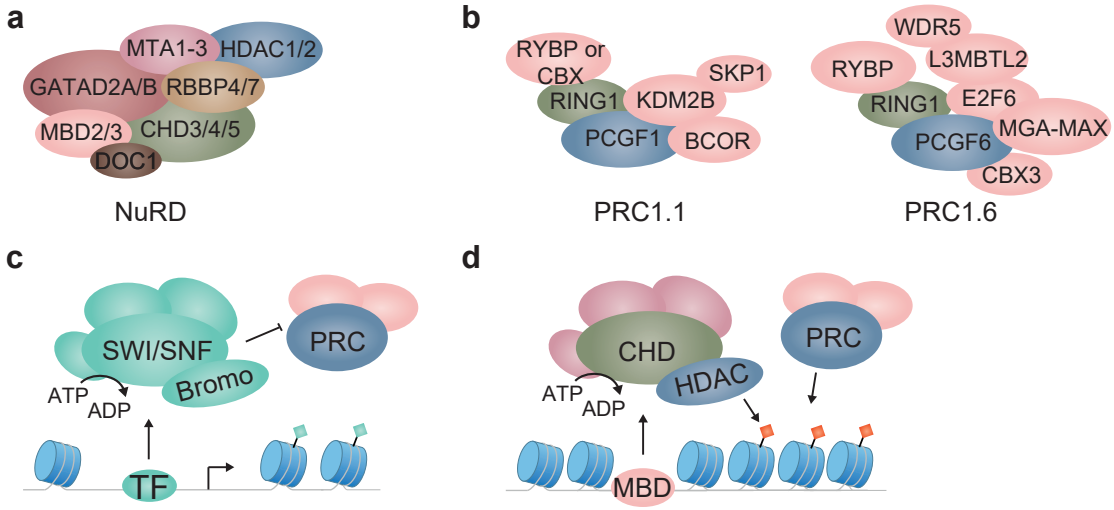


Figure 1.6: Chromatin regulator complexes. **(a)** Schematic representation of the mammalian NuRD chromatin remodeling complex. **(b)** Schematic representation of two ncPRC1 complexes PRC1.1 and PRC1.6. **(c)** SWI/SNF remodeling complexes can be recruited by TFs or local chromatin state, such as histone acetylation. They typically generate open chromatin and counteract PRCs. Green marks on nucleosomes depict histone acetylation. **(d)** The NuRD remodeling complex can be recruited by MBDs and leads to nucleosome invasion and histone deacetylation, antagonizing SWI/SNF. This can recruit PRCs. Red marks on nucleosomes depict H3K27me3.

PRC complexes are part of a regulatory network that involves other chromatin regulators and transcription factors. NuRD and SWI/SNF act antagonistically on common regulatory elements (Hainer et al., 2015; Yildirim et al., 2011) (**Figure 1.6c, d**). While NuRD promotes the formation of repressive chromatin through nucleosome invasion, histone deacetylation, and subsequent PRC recruitment, SWI/SNF can open chromatin and oppose PRC-mediated silencing, suggesting a dynamic interplay between chromatin modulating enzymes rather than a static chromatin state (Bracken et al., 2019). Thus, recruitment of chromatin complexes by transcription factors provides a scaffold linking DNA sequence recognition to chromatin structure and the epigenetic landscape.

1.2.4 Reader proteins recognize cytosine modifications to mediate physiological outcomes

The presence of modified cytosines alters the chemical landscape of the DNA helix, especially in the major groove, which influences the binding of regulatory proteins. Remarkably, even small chemical changes can modulate both base readout, involving hydrophobic contacts in the major groove, and shape readout, involving electrostatic interactions in the minor groove. The modifications can either promote or inhibit DNA-protein interactions by affecting direct side chain contacts or altering the local DNA structure (see also **Subsection 1.2.1**). However, the impact of these modifications is highly dependent on context and sequence and cannot be explained by a set of general rules. Hence, data-driven approaches are needed to uncover these protein-DNA interactions. Readers are a class of proteins that specifically recognize and bind modified cytosines and thereby mediate the downstream outcomes of these epigenetic marks. Unlike writers and erasers (see also **Subsection 1.1.1**) that establish and remove the marks, respectively, readers interpret the modifications, thereby influencing gene regulation. For instance, hmC has been proposed to be an activating signal at enhancers by either directly recruiting transcriptional activators or by preventing the binding of transcriptional repressors that typically interact with mC (Pfeifer et al., 2020). Identifying the proteins that recognize these modifications is essential to understand their downstream effects. In this section, reader proteins for mC, hmC, fC, and caC will be introduced and their functional implications will be discussed.

Readers of mC

Canonical mC readers comprise proteins containing a MBD or SET and RING-associated (SRA) domain. They recognize methylated CpGs (mCpGs) in a sequence-independent manner (Baubec et al., 2013). Other factors, such as homeobox proteins and zinc fingers can recognize mC in sequence-dependent contexts indicating a broader variety of mC readers (**Table 1.2**).

MBD proteins. Since the discovery of MECP2 as a mCpG binding protein in independent DNA sequence contexts (R. R. Meehan et al., 1989), other MBD protein family members were found comprising MBD1-6, that all share a conserved MBD domain. MBDs are key readers of mCpGs interacting with both DNA strands (Du et al., 2015). Despite the core MBD being conserved, they exhibit differential selectivity for cytosine modifications. MBD1, 2, 4 and MECP2 all preferentially bind mC over the other cytosine modifications, with MECP2 showing the highest overall affinity towards mC among the MBD family (Buchmuller et al., 2020). In contrast, MBD3, MBD5, and MBD6 are not capable to bind mC, and their MBD domains are likely evolved for protein-protein interactions (Baymaz et al., 2014). Interestingly, MBD3 shares 70 % amino acid sequence similarity with MBD2, but due to two amino acid substitutions its mCpG binding ability is abolished (Saito and Ishikawa, 2002). mC recognition occurs through a pair of Arg residues by stacking and hydrogen-bonding interactions (Ho et al., 2008; K. Liu et al., 2018) (**Figure 1.7a**). Among the MBDs, MECP2 plays an especially important role in the brain, where it is expressed at levels comparable to histones (Skene et al., 2010). Functionally, MBDs are implicated with transcriptional repressive complexes and proteins. While MBD2 is a part of the transcriptional repressor complex NuRD, MBD1 interacts with the histone methyltransferase SETDB1 and the CAF-1 complex, contributing to heterochromatin formation (Minkovsky et al., 2014; Reese et al., 2003), and MECP2 recruits co-repressors such as NCOR/SMRT and the histone deacetylase Sin3A (P. L. Jones et al., 1998; Lyst et al., 2013). Taken together, recruitment of these MBDs to mCpGs leads to transcriptional silencing of methylated regions (Du et al., 2015). MBD4 has a distinct role as a DNA glycosylase, excising thymine from G/T mismatches which can arise from spontaneous deamination of mC, thereby preserving genetic and epigenetic integrity (Bellacosa and Drohat, 2015; Pfeifer, 2006). Importantly, mutations in the MBD of MECP2 lead to altered mCpG binding and are causative of the neurodevelopmental disorder Rett syndrome underlining the importance of mC readers in maintaining development and cellular function (Amir et al., 1999; R. Meehan et al., 1992).

SRA proteins. The SRA-domain containing proteins UHRF1 and UHRF2 recognize (hemi)methylated DNA in a sequence independent context. UHRF1 is essential for maintenance methylation during replication, recruiting DNMT1 to hemimethylated CpGs (Bostick et al., 2007; Rothbart et al., 2012). UHRF1 contains a binding pocket to accommodate mC which is flipped out of the DNA duplex (**Figure 1.7b**), while two loops read the other bases of the CpG dyad by reaching through the gap. Despite UHRF2 being structurally similar to UHRF1, it is less well understood and cannot maintain mC levels when UHRF1 is lost (J. Zhang et al., 2011). While UHRF2 is dispensable for global mC levels it contributes to mC maintenance at specific loci (Y. Liu et al., 2017). Very recently it was reported to play a role in site-specific maintenance of DNA methylation during germ cell development, especially at retrotransposons (A. Bender et al., 2025).

Non-canonical mC readers. In addition to MBD and SRA proteins the list of newly identified mC interactors is expanding. These include domains such as homeobox (including SIX), winged-helix (including forkhead boxes and RFX proteins), C_2H_2 zinc-finger proteins (including Krüppel-like zinc-fingers, BTB/POZ (e.g. ZBTB33/KAISO)), bZIP and T-box (L. Bai et al., 2021; Bartke et al., 2010; Buck-Koehntop

and Defossez, 2013; Spruijt et al., 2013; Yin et al., 2017) (**Table 1.2**). Many of these transcription factors and chromatin-associated proteins interact with methylated DNA in a sequence-dependent manner by recognizing specific motifs which enhance or disrupt their binding upon methylation. Some interactions might be due to the structural resemblance of mC to thymine in the major groove, while others depend on the altered electrostatic surface. Usually, mC readers act as transcriptional repressors. mC readers can be cell-type specific as a result of their expression levels, PTMs or chromatin context. For instance, the MBD2-NuRD complex interacts with methylated DNA in neural progenitor cells and mouse brain but not in mESCs. Similarly, DLX proteins were found as specific mC interactors in brain (Spruijt et al., 2013).

Anti-mC readers. Some proteins preferentially bind to unmethylated CpGs and are repelled by mC, thereby acting as anti-readers. Among them are CXXC domain-containing proteins, which maintain gene activity at CpG rich promoters (Blackledge et al., 2010). Moreover components of the PRC1.1 complex, the INO80 chromatin remodeling complex and members of the bZIP, zinc finger containing and MAD families (L. Bai et al., 2021; Spruijt et al., 2013). For instance, the binding of the bZIP domain containing protein CREB to its CRE recognition motif is impaired by methylation leading to reduced gene expression (Kitsera et al., 2017). Also MAX was reported to lose binding affinity upon DNA methylation, while it bound to other cytosine modification derivatives (D. Wang et al., 2017).

This distinct profile of canonical and non-canonical mC readers as well as anti-readers indicates a complex and nuanced landscape of methylation recognition underscoring that introduction of a methyl group leads to recruitment or repulsion of proteins. Reader binding can be determined by DNA sequence context, PTMs and cell-type specific factors.

Readers of hmC

In contrast to mC, only few proteins have been identified to be specific hmC readers (**Table 1.2**). From proteomics screens, a limited number of overlapping proteins were identified that were discovered as hmC readers across different cell types. These include UHRF1, NEIL1, WDR76, and THY28 (Spruijt et al., 2013). Some of them are known for other biological functions, such as UHRF1 being implicated in binding hemimethylated CpGs and NEIL1 as a DNA repair enzyme that recognizes oxidized guanines. WDR76 was reported in contexts of DNA damage response as well as in destabilization of RAS (Gallina et al., 2015; Jeong et al., 2019). Also the ribosomal protein RPL26 and the mismatch repair protein MSH6 were identified as hmC readers (Iurlaro et al., 2013). However, proteomics screens identify hmC readers in specific sequence contexts which leaves the possibility that further hmC readers exist in different contexts. Different MBD family proteins have been reported to also bind hmC. Among them, MBD3 was suggested to bind (Yildirim et al., 2011), however, this was not confirmed in follow up studies (Iurlaro et al., 2013; Spruijt et al., 2013; J. Xiong et al., 2016). MECP2 was shown to have affinity towards hmC, but it seemed to bind mC more strongly (Buchmuller et al., 2020; Mellén et al., 2012, 2017; Spruijt et al., 2013). UHRF1 and UHRF2 were proposed as specific hmC readers (Frauer et al., 2011; Spruijt et al., 2013). Structural studies of UHRF2 binding to hmC have revealed that it binds symmetric and hemihydroxymethylated (mC/hmC) sites with greater affinity than hemimethylated (C/mC) sites (Zhou et al., 2014). hmC is inserted into the UHRF2-SRA pocket by a base flipping mechanism in the same ways as UHRF1 recognizes mC (Arita et al., 2008) (**Figure 1.7c**). The SOS response-associated peptidase (SRAP) domain family constitutes another group of hmC interactors including C3ORF37/HMCES (hydroxymethylcytosine binding, ESC-specific) (Aravind et al., 2013; Spruijt et al., 2013). This family is generally associated with DNA damage response and was reported to convert oxidized mC bases

to unmodified cytosine by an autopeptidase-coupled nuclease activity (Kweon et al., 2017). However, the preference of HMCES for hmC could not be consistently reproduced (Mohni et al., 2019). The zinc finger proteins SALL1 and SALL4 were identified to bind hmC in ESCs (J. Xiong et al., 2016). They were reported to preferentially bind in enhancer regions and the long isoform of SALL4 (SALL4A) has a hmC binding zinc-finger cluster. Further, SALL4A cooperated, but not directly interacted, with TET2 leading eventually to DNA demethylation at enhancer regions. TCF4 is a transcription factor that demonstrated higher affinity towards hmC modified E-box motifs (Khund-Sayeed et al., 2016), suggesting that sequence specific transcription factors can recognize hmC depending on the sequence context. Many more proteins were identified to bind hmC. Several known mC readers like ZBTB33/KAISO, UHRF1, MECP2 and MBD4 were reported to also bind hmC, while some DNA glycosylases and helicases, such as NEIL1, NEIL3 and HELLS (LSH), HARP and RECQL preferentially bound oxidized mCs (Frauer et al., 2011; Spruijt et al., 2013). In the brain, ZHX1/2, THAP11, and RCF1-5 were proposed for hmC specificity. The preference for hmC was further observed in different species and contexts, including families, such as AP2 and RFX, and C_2H_2 zinc-finger proteins, such as YY1, KLF, and ZBTB43 (L. Bai et al., 2021; Spruijt et al., 2013). Some of these preferentially bound unmethylated DNA in human and hmC-containing DNA in mouse (L. Bai et al., 2021). As for mC, also hmC can repel DNA binders, such anti-readers include, for instance, ZBTB2 (Lafaye et al., 2014). In TF-profiling studies, USF1 and USF2 were observed to preferentially bind to hmC and comparison with ChIP-seq peaks suggested their binding to hmC at weak enhancers (G. Song et al., 2021). Together, hmC might also serve the dual function of recruiting specific proteins and excluding other proteins (Jin et al., 2010). The exclusion of proteins might concern mC binders such as MBDs and KAISO, which would imply relief from the transcriptional repression maintained by mC. Given the different abundance of hmC in different tissues (**Subsection 1.1.2**), detection of hmC binders in tissue specific contexts will help to further reveal hmCs biological implications. As hmC is especially prevalent in the brain, some hmC binders might be differentially expressed in this tissue. Additionally, characterization of identified hmC readers will give further insights into its biological role.

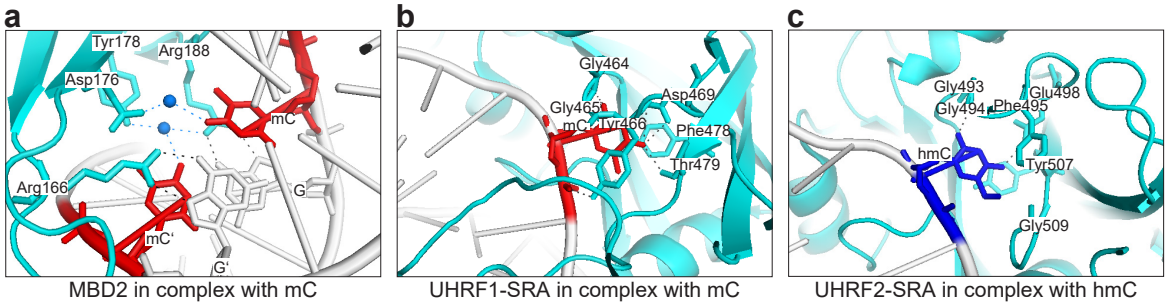


Figure 1.7: Crystal structures of readers in complex with DNA containing modified cytosines. **a** The MBD domain of MBD2 (cyan) bound to DNA (white) containing mCs (red). The interacting residues are labeled and shown in stick representation and hydrogen bonds are shown as dashed lines (black for direct interaction, blue for water-mediated interaction). PDB: 6CNP. **b** The SRA domain of UHRF1 (cyan) bound to DNA (white) with mC (red) flipped out of the duplex. The interacting residues are labeled and shown in stick representation and hydrogen bonds are shown as dashed lines. PDB: 3CLZ. **c** The SRA domain of UHRF2 (cyan) bound to DNA (white) with hmC (blue) flipped out of the duplex. The interacting residues are labeled and shown in stick representation and hydrogen bonds are shown as dashed lines. PDB: 4PW5.

Readers of fC

A greater number of putative readers were identified for fC compared to hmC, possibly due to increased polarity (Iurlaro et al., 2013; Spruijt et al., 2013). The formyl group of fC introduces unique chemical properties and is capable of forming reversible Schiff base linkages with lysine residues on proteins,

such as histones (Ji et al., 2017; F. Li et al., 2017). This has been shown in vitro but also in nucleosomal DNA, fC has been linked to increased nucleosome occupancy at enhancer regions possibly due to Schiff base formation. As these regions are also associated with elevated transcriptional activity, a regulatory function for fC at active chromatin sites is possible (Raiber et al., 2018). Multiple putative fC readers were identified in proteomics studies using pull-down approaches. Among them is TDG, which is involved in the active demethylation pathway (Spruijt et al., 2013, **Subsection 1.1.1**). Also p53, the "guardian of the genome", was proposes as an fC reader in this study, but its mode of action is yet unclear (Spruijt et al., 2013). Transcription factor families that were detected to bind fC include FOX, RFX, SIX, SCAN and MADs-box families, as well as CDKs (L. Bai et al., 2021; Iurlaro et al., 2013) (**Table 1.2**). Also components of the NuRD chromatin remodeling complex have also been reported to associate with fC (Iurlaro et al., 2013), suggesting fCs involvement in gene regulation and not merely as a passive intermediate in the active demethylation pathway. Interestingly, also DNMT1 was identified as a fC reading protein suggesting that it might have a more nuanced role in epigenetic regulation beyond maintenance methylation (Spruijt et al., 2013).

Table 1.2: Putative reader proteins of epigenetic cytosine modifications. A comprehensive list of proteins that bind to mC, hmC, fC, or caC roughly grouped into TF families or domains. It should be noted, that although some proteins were detected in human and some in mouse contexts, all protein names are written in capital for readability.

Family/Domain	mC	hmC	fC	caC
MBD	MBD1, MBD2, MBD4, MECP2	MBD4, MECP2		
SRA	UHRF1, UHRF2	UHRF1, UHRF2		
Homeobox	DLX1, DLX5, DLX6, HOMEZ, HOX2, HOX4, MEIS1, MEIS2, PBX1, PKNOX2, SIX4, ZFHX3, ZHX1	HDX, ZHX1, ZHX2	SIX4	CRX, HMBOX, MEIS1, MEIS2, MEIS3
FOX	FOXK1, FOXK2, FOXP1, FOXP4		FOXI3, FOXJ3, FOXK1, FOXK2, FOXP1, FOXP4	
RFX	RFX1, RFX2, RFX5, RFXANK, RFXAP		RFX1, RFX2, RFX3, RFX5, RFXANK, RFXAP	
Zinc finger	KLF2, KLF3, KLF4, KLF5, ZBTB40, ZHX1, ZNF384	KLF3, KLF7, KLF16, PRDM5, SALL1, SALL4, THAP11, YY1, ZBTB43, ZNF618	CNBP, ZFP3, ZNF24, ZSCAN3, ZSCAN12, ZSCAN21	CTCF, WT1, ZBTB7, ZNF187
bZIP	CEBPβ	ATF2		SMARCC1
T-box	TBX2, TBX3, TBX20		TBX2	
bHLHZ		TCF4, USF1, USF2		MAX, TCF4

Family/Domain	mC	hmC	fC	caC
others	AKSCAN3, CHD5, ESRRB, OGT, RBPJ, TET1	DHM1, ESRRB, HARP, HMCES, HMG1, HMG2, LSH, MGME1, MPG, MSH2, NEIL1, NEIL3, PNKP, PRP8, RBM14, RECQL, RFC, RPL26, TAF7, THY28, WDR76	CRSP2, CSDA, DNMT1, EHMT1, GTF2I, L3MBTL2, MPG, NRF1, p53, PURA, TDG, UBTF	BAF179, DNA-Pol β, DNMT1, NRF1, TDG, TET3, RNA-PolII

Readers of caC

The most established caC interactor is TDG (L. Zhang et al., 2012), which excises caC in the context of active DNA demethylation (He et al., 2011; Maiti and Drohat, 2011, **Subsection 1.1.1**). In this process, caC is flipped out of the helix and stabilized in a specific binding pocket (Hashimoto et al., 2013). However, TDG might not be a real epigenetic reader but it senses the caC modification and is therefore involved in chromatin dynamics. Also other DNA glycosylases and repair proteins have shown preferential interactions with oxidized mCs (hmC, fC, and caC), including the NEIL1 family, indicating the involvement of these bases in DNA demethylation pathways (Spruijt et al., 2013). Several TFs were reported to recognize caC (**Table 1.2**), among them WT1, which binds to its consensus sequence also in the presence of caC (Hashimoto et al., 2014). Moreover, MAX was shown to preferentially bind to its canonical E-box sequence when either C or caC are present (D. Wang et al., 2017). Another bHLH TF, TCF4, showed enhanced binding when caC flanked the E-box motif which was suggested to be mediated by electrostatic interactions between two Args with the negatively charged caC (Golla et al., 2014; J. Yang et al., 2019). bHLH TFs are known to recruit coactivator or corepressor complexes (de Martin et al., 2021), but also components of the Swi/Snf chromatin remodeling complex were directly identified as putative caC readers, among them BAF179 (Spruijt et al., 2013). While CXXC domains are usually repelled by cytosine modifications, it was reported that the CXXC domain of TET3 has a high affinity for caC suggesting a function as a regulator of caC removal (Jin et al., 2016). RNA PolII can be arrested by caC likely due to interactions between its glutamine rich regions and the negatively charged caC (L. Wang et al., 2015). Therefore, diminished RNApolII-mediated transcription elongation is observed at caC sites. Similarly, DNA pol β is affected by caC, while it is not affected by the other cytosine modifications. This is attributed to interactions between lysine residues in the polymerase active sites with caC (Howard et al., 2019).

1.2.5 Profiling methods for DNA-interacting proteins in epigenetic contexts

Understanding gene expression requires the determination of which proteins interact with chromatin. Gene regulation depends not only on the chromatin context but is also modulated by epigenetic marks. The identification of proteins that read or interpret these marks is therefore key to advance this understanding. As discussed in the previous section, many binders of mC, hmC, fC, and caC have been detected in the past years, however, other proteins might bind only in sequence- or tissue-dependent contexts and were therefore overseen. A plethora of methods were developed to profile these interactions ranging from classical DNA-binding profiling methods to more recent ones, that account for epigenetic modifications and context dependence, all of which will be examined in this section (**Table 1.3**).

In vitro high throughput assays

In vitro high-throughput technologies enable systematic assessment of TF sequence preferences by using large libraries of synthetic DNA sequences in parallel. Therefore they have helped to decode the intrinsic DNA binding specificities of multiple TFs. However, they often neglect the influence of epigenetic modifications, that can significantly alter TF-DNA interactions in vivo. Beyond this limitation, they provide a rich source to study sequence-specific preferences of TFs independent of chromatin context and cellular cofactors.

Classical high-throughput methods. Classical methods include protein binding microarrays (PBMs), which were used to determine the DNA-binding motifs of various TFs. PBMs usually involve applying purified, epitope-tagged TFs to a microarray containing double-stranded oligonucleotides, followed by detection with a fluorophore-conjugated antibody allowing the identification of DNA sequences preferentially bound by the applied TF (Berger et al., 2006). Although covering broad sequence contexts, there are several drawbacks including the requirement for high amounts of purified protein, making it difficult to study full-length proteins and information of PTM influence is lacking. Also the oligonucleotide length is limited, which constrains the resolution of complex binding architectures. Methylation-sensitive PBMs have been developed (S. Hu et al., 2013; I. K. Mann et al., 2013) and have identified novel methylated DNA binding proteins, most of which bound in a sequence-dependent manner. These studies, however, are limited by design as they either contain fully methylated or unmethylated probes in isolation, thus preventing competitive binding assessment or position-specific methylation effects. TF arrays constitute another method and have been used to detect which TFs in a protein extract can bind a known DNA sequence. For this purpose, TFs are immobilized on the array and a labeled DNA probe is added (X. Jiang et al., 2004). More recently, SMiLE-seq (selective microfluidics-based ligand enrichment followed by sequencing) was established, which utilizes a microfluidic platform to immunocapture full-length TFs and enrich for their DNA ligands from random sequence libraries (Isakova et al., 2017). Coupled with high-throughput sequencing, de novo DNA binding specificities and affinities of TFs and TF dimers can be determined, with improved throughput and sequence resolution compared to PBMs. However, it does not address methylation sensitivity directly.

High-throughput and methylation-sensitive SELEX. SELEX has been adapted to high-throughput (HT-SELEX), determining protein–DNA binding specificities of hundreds of TFs in parallel using barcoded DNA libraries and low nanogram levels of protein. HT-SELEX allows for binding motifs to emerge over iterative selection rounds from randomized oligonucleotide pools. SELEX is compatible with full-length TFs and those that require PTMs (Jolma et al., 2010). Moreover, SELEX was adapted to address methylation sensitivity. In methylation sensitive SELEX, DNA libraries are enzymatically methylated using M.SssI before each selection cycle allowing direct comparison of binding specificity between methylated and unmethylated DNA (Yin et al., 2017). This study revealed that 60 % of the examined TFs were affected by DNA methylation. Another method, EpiSELEX-seq, applies single-round SELEX by conducting an EMSA after TF-DNA incubation to assess binding to methylated and unmethylated DNA fragments simultaneously (Kribelbauer et al., 2017).

Affinity profiling via proximity ligation. A more recent methodology is Digital Affinity Profiling via Proximity Ligation (DAPPL), which enables multiplexed profiling of TF binding to random DNA libraries bearing different cytosine modifications (G. Song et al., 2021). By combining digital barcoding with proximity ligations, DAPPL enables quantitative, high-resolution read-outs of TF-DNA interac-

tions, allowing systematic mapping of differential binding. However, DAPPL retains some drawbacks that are intrinsic to in vitro assays. It does not account for chromatin context and cofactor-mediated effects that influence in vivo binding. Moreover, it depends on high quality purified recombinant proteins in sufficient quantity and activity, excluding proteins that are difficult to express and those that require PTMs. It also only detects direct protein-DNA interactions and can not capture indirect interactions or complex formations, therefore the results may not fully reflect the physiological binding landscapes.

Quantitative interaction proteomics

Quantitative interaction proteomics contributed significantly to expand the list of proteins that selectively interact with epigenetically modified DNA sequences in vitro. This approach combines affinity purification using DNA probes bearing specific modifications, such as mC and its oxidized derivatives, with high-resolution MS to identify and quantify interacting proteins in an unbiased, proteome-wide manner (L. Bai et al., 2021; Iurlaro et al., 2013; Spruijt et al., 2013; J. Xiong et al., 2016). For this purpose, biotinylated oligonucleotides carrying distinct epigenetic marks are immobilized on streptavidin-coated beads and incubated with nuclear protein extracts. Proteins that are specifically enriched on modified DNA are then enzymatically digested and subjected to LC-MS analysis (**Figure 1.8**). Different quantification strategies can be employed, including stable isotope labeling by amino acids in cell culture (SILAC) and tandem mass tags (TMT). These enable multiplexed and accurate relative quantification (Bartels et al., 2011; Mateus et al., 2020). Beyond, LFQ has been widely used due to its simplicity and broad compatibility. LFQ circumvents the need for metabolic labeling and is sensitive enough to detect subtle differences in protein binding affinities to differentially modified DNA. However, there are some drawbacks to quantitative proteomics, such as the lack of native chromatin context of the applied DNA probes. As a result, interactions depending on nucleosomal structure may be missed. Therefore, recombinant nucleosomes carrying specific histone modifications have been developed and used for affinity proteomics (Bartke et al., 2010). Moreover, detected proteins may be part of bigger complexes or bind indirectly via protein-protein interactions, making it difficult to uncover the direct DNA binders. This becomes further apparent by the difficulty to confirm the discovered modification-dependent interactions in vivo (H. Zhu et al., 2016). TF occupancy in vivo is determined by local protein concentration and DNA affinity (Segal and Widom, 2009), hence complementary biophysical approaches are needed to validate these interactions. These have included isothermal calorimetry, fluorescence polarization measurements or EMSA. However, they are low throughput and require the laborious expression and purification of recombinant proteins. Despite these limitations, quantitative proteomics continues to be an essential tool to identify potential reader proteins for epigenetic DNA modifications. Advances towards high-throughput adaptations have been made as well, including 96-well format screens (Hubner et al., 2015).

ChIP-based genomic approaches

In vivo interactions between proteins and DNA can be mapped by high-throughput sequencing built on ChIP.

ChIP-seq. ChIP-seq combines antibody-based enrichment of crosslinked chromatin fragments with NGS (D. S. Johnson et al., 2007). This allows for genome wide mapping of TF-binding and when combined with epigenetic maps it allows for comparison of TF occupancy relative to local DNA modifications. The use of ChIP-bisulfite sequencing enables a direct correlation of TF binding with DNA modifications at single base pair resolution (Brinkman et al., 2012).

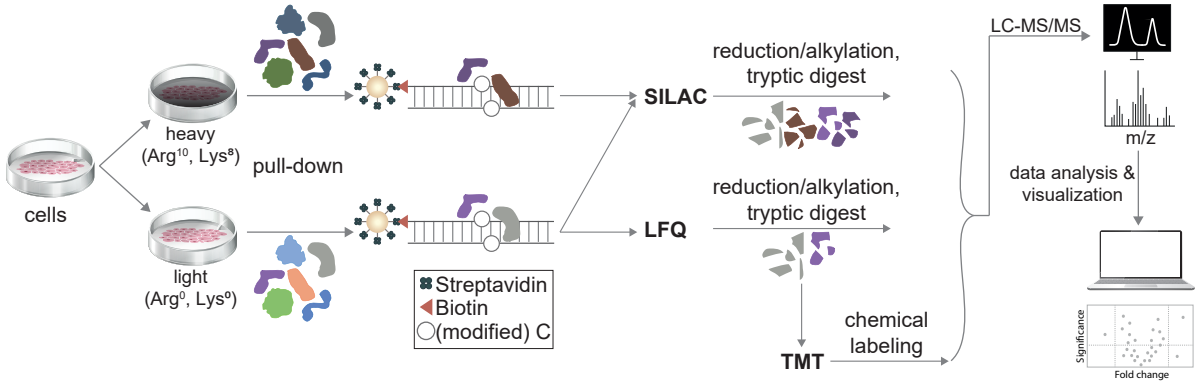


Figure 1.8: Workflow of quantitative interaction proteomics. Quantification can be achieved either by SILAC, TMT, or label-free methods. Proteins are first extracted from cells, for SILAC, cells are additionally grown in isotopically labeled (heavy) medium. Isolated proteins are subjected to a pull-down assay, typically using a biotin-labeled DNA probe that is attached to streptavidin-coated beads. After enrichment, bound proteins are reduced, alkylated and digested into peptides. For TMT quantification, the peptides are chemically labeled at this step. The resulting peptide mixtures are analyzed LC-MS/MS to identify and quantify the proteins.

CUT&RUN and CUT&Tag. Several alternative genomic sequencing based methods have been developed, such as Cleavage Under Targets and Release Using Nuclease (CUT&RUN)-seq (Skene and Henikoff, 2017), an in situ method, which likely maintains protein-DNA interactions in their natural state. A specific antibody binds the protein of interest and a Protein A-fused MNase is subsequently recruited to this site and cuts out the protein-bound DNA to release the DNA fragment. Cleavage Under Targets and Tagmentation (CUT&Tag) (Kaya-Okur et al., 2019) uses a Protein A-fused Tn5 transposase, which is recruited to the site bound by the protein of interest, and then cuts and simultaneously inserts sequencing adapters. These methods results in lower background than ChIP-seq, however, prior knowledge and high quality antibodies are still needed.

Global chromatin proteomics approaches

Global proteome profiling is a complementary approach that provides identification of nuclear proteins under physiological conditions.

ChIP-MS. Combining ChIP with MS uncovers protein complexes associated with specific chromatin marks or TF-bound regions (J.-P. Lambert et al., 2009). However, it still relies on prior knowledge of the TF and requires highly specific antibodies. Moreover, it yields averaged interaction profiles across all bound sites and therefore lacks locus-specific resolution.

Chromatin Proteomics. Chromatin fractionation followed by MS enables unbiased profiling of chromatin bound proteins without the need for specific antibodies (van Mierlo and Vermeulen, 2021). It allows to study protein function and complex formation in their in vivo context. By prior chromatin enrichment cell-type and disease-specific chromatin bound proteomes can be identified and it is particularly suited to detect low-abundant factors (Kustatscher et al., 2014).

Locus-specific chromatin approaches

Several approaches have been developed to investigate proteins associated with specific genomic loci, which can be interesting in view of epigenetic DNA modifications, when studying CpG island or CpG-rich enhancers in a modified vs unmodified state.

Hybridization-based enrichment workflows. Proteomics of Isolated Chromatin Segment (PICh) uses sequence specific desthiobiotinylated locked nucleic acid (LNA) probes to target genomic regions by hybridization, enabling the isolation of genomic DNA and its associated proteins with subsequent identification of these proteins (Déjardin and Kingston, 2009). Originally, it was used for telomeric regions and could be combined with SILAC for quantitative analysis. (Gauchier et al., 2019).
Hybridization Capture of Chromatin Associated Proteins for Proteomics (HyCAPP) offers a similar concept to PICh but using standard oligonucleotide probes without LNA chemistry (Guillen-Ahlers et al., 2016). While still enabling locus-specific enrichment and subsequent proteome profiling, affinity and specificity of this approach is lower compared to PICh.

Table 1.3: Experimental methods for profiling protein-DNA interactions. Summary of key methods that were used or are promising to detect proteins that recognize and bind epigenetically modified DNA. Evaluation is based on throughput, ability to incorporate epigenetic context (synthetic vs native), quantitative readout, in vivo applicability and locus-specific resolution.

Method	Throughput	Epigenetic Context	Quantitative	In Vivo	Locus Specific
PBM	Medium	Limited	Semi-quantitative	No	No
SELEX/Methyl-SELEX	High	Yes, with modified oligos	Semi-quantitative	No	No
DAPPL	High	Yes, various modifications	Yes, digital readout	No	No
Affinity Purification + MS	Medium–High	Yes, modified DNA probes	Yes, LFQ, SILAC, TMT	No	No
ChIP-(bisulfite)-seq	Medium–High	Yes, co-mapping/base resolution	No	Yes	No (genome wide)
CUT&RUN	Medium	Yes, native chromatin	No	Yes	No (genome wide)
ChIP-MS	Low–Medium	Yes, histone mark/TF context	Yes	Yes	No (averaged loci)
Chromatin Proteomics	Medium	Yes, global chromatin state	Yes	Yes	No (global)
PICh / HyCAPP	Low	Yes, native chromatin	Yes, SILAC	Yes	Yes
dCas9-based Capture	Medium	Yes, native chromatin	Yes, MS/proximity label	Yes	Yes

CRISPR/dCas9-based targeting. CRISPR-based technologies further advanced locus-specific proteomics using catalytically inactive Cas9 (dCas9) (Sander and Joung, 2014). One approach, CAPTURE, uses a biotinylated dCas9 that is targeted by a guide RNA to isolate the locus of interest along with its associated proteins by affinity enrichment (X. Liu et al., 2017). Another approach uses proximity labeling with biotin ligases, like BirA, which are fused to dCas9 facilitating locus specific biotinylation (Schmidtmann et al., 2016), however, these methods suffer from off-target binding and dCas9 background as well as possible perturbations of the locus-specific proteome by dCas9 (Wierer and Mann, 2016). Despite these drawbacks, these approaches are promising for examining the locus-specific proteome when combined with proteomics analyses.

Together, these technologies provide complementary approaches to identify protein-DNA interactions. However, no method offers complete resolution of the interplay between epigenetically modified DNA and TFs (**Table 1.3**). Quantitative interaction proteomics remains a valuable tool for identification of novel readers of modified DNA. Meanwhile, evolving CRISPR-based methods hold promise to aid in understanding the locus-specific functional relevance.

1.2.6 Essential points

- The cytosine modifications mC, hmC, fC, and caC alter the chemical and structural properties of DNA. By modifying the electrostatic and hydrogen-bonding signatures of the major groove, they influence how DNA-binding proteins interact with DNA.
- Transcription factors play a pivotal role in interpreting the DNA sequence and epigenetic modifications and they additionally recruit chromatin remodeling complexes and other chromatin regulators, thereby building a complex regulatory network that connects DNA modifications with histone modifications and chromatin state.
- A growing number of reader proteins for mC, hmC, fC, and caC have been identified, showing that each modification is recognized by a distinct set of proteins. While MBD proteins represent the most prominent mC readers, many sequence-specific binders have been found to preferentially bind to one or multiple of the modified cytosine bases, adding further complexity to the epigenetic landscape.
- A plethora of methods is available to profile protein-DNA interactions. Quantitative interaction proteomics appears as the most powerful one for identifying novel epigenetic reader proteins. Despite many current approaches focusing on DNA sequence recognition, locus-specific strategies, especially those based on CRISPR-dCas9 combined with proteomics, have the potential to uncover how reader proteins engage at specific genomic loci.

1.3 Advances in understanding CpG dyad symmetry and hydroxymethylcytosine

1.3.1 Symmetry and asymmetry in CpG dyads

CpG dinucleotides constitute a palindromic motif, meaning the sequence is identical when read in 5'-3' direction on either strand (**Figure 1.9a**). This inherent symmetry allows CpG dinucleotides to act as dyads that form two-stranded units of epigenetic information, in which the cytosines on either strand can independently exist in either their unmodified form or as one of the four epigenetic derivatives mC, hmC, fC, or caC. Although the canonical view assumes that CpG dyads are either fully unmethylated (C/C) or symmetrically methylated (mC/mC), several mechanisms frequently disrupt this symmetry. Asymmetry in CpG dyads can arise from replication-dependent dilution of modifications, de novo methylation by DNMT3A and DNMT3B acting only on a single strand, errors in the maintenance methylation machinery and imprecise maintenance activity (Charlton et al., 2018) or from TET protein processivity, which either oxidize cytosines on only one strand or stall intermediate forms. These processes together yield a theoretical permutation of 25 combinations with 15 unique dyad combinations when strand equivalence is considered (**Figure 1.9b**). Among these, five combinations are symmetric, while ten are asymmetric, each of which may represent distinct epigenetic states with unique biochemical and regulatory outcomes. In vitro studies have shown, that TET proteins are capable of binding a variety of CpG dyad combinations including mC paired with all modified Cs (Crawford et al., 2016). Moreover, the oxidation rate of mC by TET2 was shown to be largely unaffected by the modification state of the complementary C, suggesting TETs to be capable of processing symmetric and asymmetric CpG dyads with comparable efficiency (Crawford et al., 2016). This indicates that all possible CpG dyad modification combinations are biochemically viable substrates in vivo. Notably, BER, the final step of active DNA demethylation, proceeds in a strand-specific manner as well. The oxidized strand is usually fully repaired before the complementary strand is processed, which allows for stable asymmetric demethylation without potentially harmful DNA intermediates (Weber et al., 2016). This further indicates the presence of hemimodified CpG states. Together, this challenges the traditional assumption of CpG dyad symmetry and underlines the importance of considering strand-specific cytosine modification states for interpreting the regulatory potential of the epigenetic code. Indeed, it has been shown that the symmetry of CpG dyads affects how reader proteins recognize and bind DNA (Buchmuller et al., 2020).

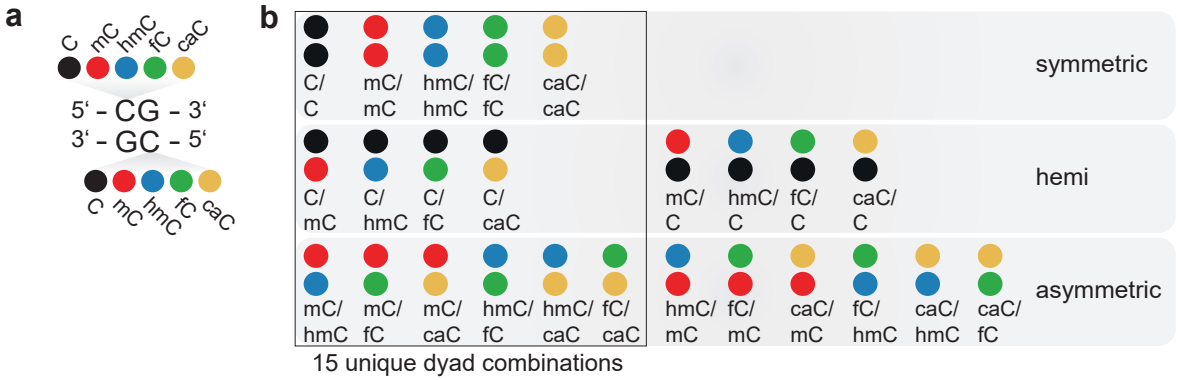


Figure 1.9: Cytosine modifications in the CpG dyad. (a) CpG (Cytosine-guanine sequences connected by a phosphate backbone) is a palindromic sequence, being identical when read in 5'-3' direction on either strand. In dsDNA, CpG dinucleotides exist as dyads, in which the cytosines on either strand can independently exist in either their unmodified form or as one of the four epigenetic derivatives mC, hmC, fC, or caC. **(b)** Biological activities of DNMT and TET enzymes yield a theoretical permutation of 25 modification combinations in CpG dyads. 15 of these are unique dyad combinations when strand equivalence is considered (black rectangle). 5 symmetric combinations, 4 hemi-modified, and 6 asymmetrical combinations are feasible in CpG dyads.

1.3.2 Advances in duplex sequencing technologies and abundance of asymmetric CpG dyads

While a plethora of methods exist to detect and quantify epigenetic cytosine modifications (**Subsection 1.1.2**, determination of combinatorial modification states of CpG dyads has remained a challenge. This limitation arises from the fact that most cytosine modification mapping technologies rely on DNA denaturation and single-strand sequencing, thereby losing the information of the complementary strand, which makes it impossible to determine the modification states of both Cs of the CpG dyad.

In traditional BS, the information of the modification on the opposite strand is lost due to PCR after bisulfite conversion, therefore methylation patterns from both strands of a single duplex molecule could not be captured. To overcome this, hairpin bisulfite PCR was established, where a short hairpin oligonucleotide is ligated to restriction enzyme digested DNA fragments (Laird et al., 2004). By physically linking the Watson and Crick strands, strand pairing can be preserved during bisulfite treatment and amplification to later reconstruct the original duplex methylation state (**Figure 1.10a**). These applications showed varying levels of CpG dyad asymmetry across genomic regions and cell types (Arand et al., 2012; H. Xie et al., 2011). High symmetric methylation patterns were found in around 30 % of CGIs and 10 % of Alu repeats, while other regions showed more hemimethylation. Adaptions of this approach enabled simultaneous monitoring of mC and hmC. For instance, hairpin bisulfite with oxidative bisulfite sequencing (HPoxBS) combines hairpin BS with oxBS to allow discriminating between mC and hmC on both DNA strands (Giehr et al., 2018). However, these techniques are technically demanding, low throughput and not easily scalable to single cell resolution (Giehr et al., 2018; Laird et al., 2004; L. Zhao et al., 2014).

More recently, alternative approaches have emerged, such as the bisulfite-free Mhemi-seq method that uses the methylation-dependent restriction enzyme MspJI and can distinguish between full, hemi- and unmethylated states (X. Xiong et al., 2024). Functionally, it revealed that hemimethylation can inhibit TF binding to varying degrees, including strand-specific effects. However, it is limited to specific sequence motifs and does not include hmC. Another MspJI-based approach is Dyad-seq, which overcomes the challenges of the traditional hairpin methods (**Figure 1.10b**). Dyad-seq combines enzymatic modification detection (MspJI for mC, AbaSI for hmC), with base conversion chemistry to quantitatively assess all possible mC/hmC combinations genome wide at CpG dyads (Chialastri et al., 2024). By using H-M-Dyad-seq and H-H-Dyad-seq it was shown that hmC is more frequently paired with

mC than with itself indicating a higher abundance of asymmetrically modified hmC dyads in mESCs. Follow-ups comprised scDyad-seq and scDyad&T-seq which allow for single-cell profiling of CpG dyad states together with transcriptomics to offer insights into global and maintenance methylation patterns.

Recently, The SCoTCH-seq (Strand-Coupled Tandem Cytosine Hydroxymethylation and methylation sequencing) has been introduced, which is a quantitative, base-resolution approach to simultaneously detect C, mC, and hmC on both strands of the same DNA fragment (Hardwick et al., 2025). It uses a hairpin-based strategy to preserve strand pairing throughout a couple of enzymatic conversions and base-modification specific protection reactions, which enables the identification of all nine possible CpG dyad modifications containing either C, mC, or hmC. By using this method on the mouse genome, they revealed that 98% of hmC is found in asymmetrically modified CpG dyads. hmC was most frequently paired with C (hmC/C made up 3.6 % of CpG dyads), followed by hmC/mC with 2.0 % each. Different forms of asymmetric hmC occur at distinct genomic elements such as different enhancer types.

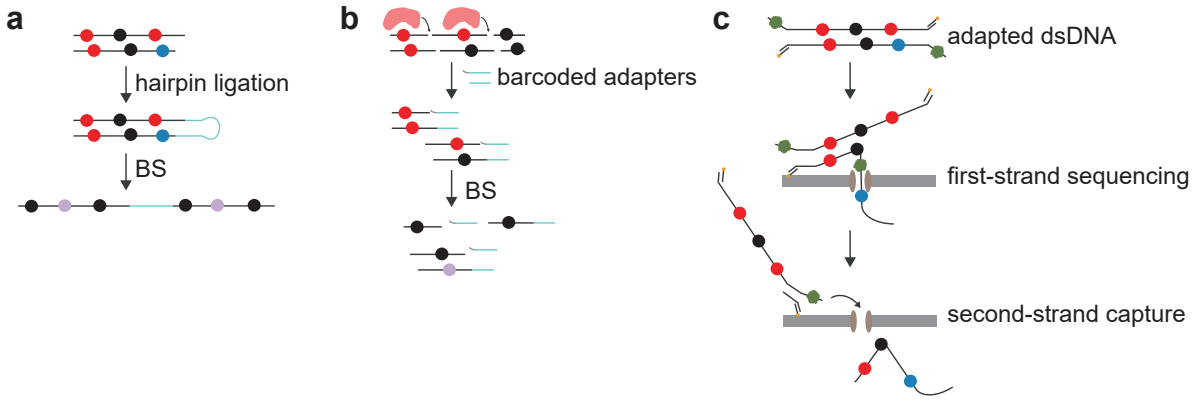


Figure 1.10: Strand-specific or duplex-sequencing approaches. (a) In Hairpin-bisulfite sequencing, both strands of the DNA duplex are covalently linked with a hairpin adapter to retain the strand information after bisulfite sequencing. **(b)** Dyad-seq uses MspJI, which generates double-stranded DNA fragments with defined methylation-dependent cutting sites. Adapters are ligated and the DNA is treated with bisulfite and sequenced preserving information on methylation. **(c)** Nanopore duplex sequencing uses adapted dsDNA molecules which allow the two strands to enter the pore successively. Subsequently, the strands can be bioinformatically paired.

Another promising advance is nanopore duplex sequencing. The latest Oxford Nanopore flowcells (R10.4.1HD) enable both strands of a dsDNA molecule to pass through the pore, thereby generating duplex reads that can be bioinformatically paired, which allows for a direct readout of cytosine modification states on both strands (**Figure 1.10c**). According to a recent study, symmetric mC/mC accounted for 52.5 % of CpG dyads, while unmodified C/C made up 23.5 % (Halliwell et al., 2025). The hemimethylated dyads C/mC and C/hmC were less common with 6.3 % and 2.0 %, respectively and symmetric hmC/hmC made up 3.0 %. However, mC/hmC comprised 12.8 % of all CpG dyads, making them twice as prevalent as hemimethylated C/mC and over four times and six times as prevalent as symmetric hmC/hmC and C/hmC, respectively (**Figure 1.11a**). Strikingly, hmC was paired with mC in 72.3 % of the cases (**Figure 1.11b**). The relative abundances of hmC-containing dyads are contrasting the ones reported with the SCoTCH-seq method, which could indicate tissue-specific distributions as SCoTCH-seq was performed on E14 mESCs and nanopore duplex sequencing on mouse cerebellum tissue. In total, these findings challenge the assumption of CpG modification to be largely symmetric and highlight the potential relevance of CpG asymmetric dyads. Albeit, also the nanopore duplex sequencing still faces limitations, as duplex-paired reads only accounted for 32 % of all reads and also high sequencing costs impede the application. Together these technologies pave the way for quantifying and mapping combinatorial CpG dyad modifications.

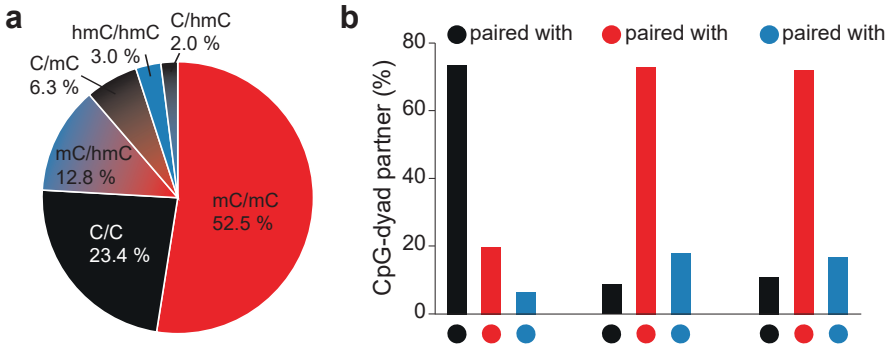


Figure 1.11: Putative prevalence of modified CpG dyads. (a) CpG-context cytosine base-calls from all duplex reads according to Halliwell et al., 2025. 52.5 % of CpG dyads were found to be symmetrically methylated, followed by 23.4 % of dyads being unmodified. The third most prevalent CpG dyads was hmC/mC with a prevalence of 12.8 %. **(b)** In a CpG-dyad, C was found to be predominantly paired with another C, mC was predominately paired with another mC, while hmC was predominately paired with mC. Data used for figure preparation are from Halliwell et al., 2025.

1.3.3 Functional implications of asymmetric modifications for protein recognition

Interpretation of modified CpG dyads by reader proteins has traditionally focused on symmetrically modified CpGs. As outlined in **Subsection 1.2.4**, many DNA-binding proteins engage with the CpG duplex in a bivalent manner and make contacts on both strands. Most proteome profiling approaches have neglected asymmetric modifications, which, regarding hmC, are more prevalent than their symmetric counterparts, as recent mapping studies suggest (Halliwell et al., 2025). Only a few proteins, such as DNMT1 and UHRF1, were known to recognize hemimethylated CpGs in vivo (T. Li et al., 2018). More recently, TF profiling studies by Song et al. (G. Song et al., 2021) have identified numerous TFs with enhanced binding to hemimodified dyads. Specifically, 18 and 11 proteins showed enhanced affinity for C/mC and C/hmC, respectively. Among them were FOXC2, MEIS1-3, TBX2-3, HOMEZ, RFX2 and TFAP2 for C/mC and ZBTB7B, HMBOX1, FOXP4 and HOMEZ for C/hmC. Further distinct binders were seen for C/fC and C/caC. A lot of proteins were also found to be repressed by the hemimodifications, which were largely overlapping for all four hemimodified dyads. These included E2F1, USF1 and USF2. More proteins were repressed for C/hmC dyads than for other hemimodified dyads. Beyond these screens, emerging technologies offer quantitative perspectives to model TF binding in the context of CpG methylation. For instance, Methyl-Spec-seq allows to assess the effects of methylation independently on each DNA strand, making it suitable for studying asymmetric dyads (Zuo et al., 2017). This method enables a precise prediction of binding outcomes under different epigenetic states such as strand-specific inhibition or facilitation of TF-DNA interactions. Although only reported for mC, in principle, this method can be extended to hmC. Despite these advances, the protein interactomes of asymmetric CpG duplex modifications, especially with the most prevalent mC/hmC dyad, remain largely unexplored.

1.3.4 hmC as a potential sixth base

hmC has emerged as a prospective candidate for the designation of the sixth base of the genome. Initially considered as a transient intermediate in the active demethylation pathway, it is now recognized as distinct and stable epigenetic modifications with a wide range of biological functions. Its prevalence, tissue-specific distribution, stability and emerging roles in gene regulation indicate hmC to be a unique regulatory mark in its own right. As outlined in the previous chapters, hmC is increasingly found to be prevalent in specific tissues and developmental stages (**Subsection 1.1.2**). Unlike the other oxidized derivatives fC and caC, hmC is relatively stable and persists at elevated levels in certain cell types,

further supporting the hypothesis that it may function independently of the DNA demethylation cycle (Bachman et al., 2014). Functionally, hmC has been shown to influence chromatin states and gene expression. In contrast to mC, which usually represses transcription through recruitment of specific readers, like MBDs, hmC is not recognized by several of these mC-binding proteins (Buchmuller et al., 2020; Hashimoto et al., 2012), suggesting that the conversion of mC to hmC may relieve transcriptional silencing and therefore reverse the repressive effects of DNA methylation. Indeed, hmC is often found enriched at active enhancers, gene bodies and regulatory regions and is frequently associated with increased transcriptional activity in many cellular contexts (Sérandour et al., 2012). Albeit, the functional outcome of hmC is context dependent. In ESCs and different somatic tissues, gene body hmC is positively correlated with transcription, while in neuronal progenitors, hmC may have a repressive effect (J. Robertson et al., 2011; Tan et al., 2013; H. Wu et al., 2011). Moreover, hmC was found depleted in promoter regions of highly expressed genes contradicting the assumption of hmC to be an universal activator (J. Yang et al., 2020).

hmC may exert its functions by structural and protein-mediated mechanisms. Although studies suggest hmC to only cause subtle changes to the DNA helix, it may alter physical properties such as DNA flexibility and nucleosome positioning (Battistini et al., 2021). Primarily, it may act as a signal for recruiting specific reader proteins, which subsequently initiate chromatin remodeling or transcriptional outcomes. Interestingly, hmC is non-randomly distributed but enriched at CpG islands, transcription start sites and gene bodies with patterns that are different from mC, fC and caC (Ficz et al., 2011; H. Wu et al., 2011; Yu et al., 2012). This implies a regulatory potential that has not been elucidated yet. Additionally, hmC has been implicated in replication, DNA repair and genomic stability (Kafer et al., 2016; Nakatani et al., 2015; A. B. Robertson et al., 2014) as well as development, pluripotency and cellular differentiation (D. Bai et al., 2025; H. Wu et al., 2011). Altogether, this underscores its potential significance as an epigenetic mark. For instance, TET3-mediated hmC formation ensured transcriptional fidelity in lung development (F. Wu et al., 2023). Clinically, global loss of hmC was observed in cancer and reduced hmC levels are correlated with increased tumor aggressiveness and metastasis (Lian et al., 2012; LIAO et al., 2016; Y. Zhang et al., 2016). hmC disregulation was further implicated with other complex diseases including Alzheimer's and diabetes (Y. Yang et al., 2019; J. Zhao et al., 2025). These findings have led to the suggestion of using hmC as a diagnostic biomarker for various diseases and as a potential therapeutic target (W. Li et al., 2017; C.-X. Song et al., 2017).

While hmC is clearly involved in many essential cellular processes, its biological significance as an independent epigenetic signal is not fully understood yet. Its stability, targeted distribution, and proposed reader proteins suggest it to play distinct regulatory roles that go beyond its role as a demethylation intermediate. These functions may include gene regulation, chromatin organization, developmental functions and disease pathogenesis. Ongoing research will help to determine to which extent hmC serves as a causative epigenetic signal and justify its consideration as a sixth base in the mammalian genome.

1.3.5 Essential points

- CpG dyads can bear symmetric or asymmetric cytosine modifications due to TET processivity, failures in maintenance methylation or de novo activity on single strands, which can result in up to 15 distinct CpG dyad modification states.
- Emerging duplex sequencing technologies, including hairpin bisulfite sequencing, Dyad-seq and Nanopore duplex sequencing enable strand specific analysis of CpG dyad modifications at base pair resolution.
- Asymmetric mC/hmC dyads are the most frequent form of asymmetric CpG dyads and more prevalent than hemimodifed C/mC and C/hmC dyads.
- Hemimodified dyads were shown to interact with TFs in distinct ways, however, the full spectrum of protein interactions with asymmetric dyads, especially the most prevalent dyad mC/hmC, remains underexplored.
- hmC is increasingly recognized as a stable and functionally relevant DNA modification engaging in distinct biochemical behavior with potential unique readers.
- hmCs prevalence, stability and potential for specific protein interactions, positions it as a strong candidate for designation as the sixth base of the mammalian genome.

2 Aim

Knowledge gap. 5-hydroxymethylcytosine (hmC) is an epigenetic DNA modification with emerging relevance in gene regulation, yet its functional implications are still not completely understood. Previous studies on hmC reader proteins have focused mostly on symmetric (hmC/hmC) CpG contexts. Importantly, asymmetric CpG dyads have not been explored in proteomics-based studies. However, recent duplex sequencing studies indicate that hmC frequently occurs in asymmetric combinations, particularly in hmC/5-methylcytosine (mC) and hmC/cytosine (C) CpG dyads. Despite their apparent abundance, these dyads have not been studied yet in regards to protein interactions leaving their biological functions unexplored.

Rationale. This thesis presents the first proteomics-based investigation of hmC in asymmetric CpG dyads in mammals, with a particular focus on the hmC/mC modification. To achieve this, a pull-down strategy coupled to quantitative MS-based proteomics was used and extended by utilizing DNA probes carrying strand-specific hmC-modification combinations. This approach enables the identification of nuclear proteins that distinguish between symmetric and asymmetric hmC-containing DNA probes. Additionally, a targeted DNA enrichment strategy was developed using an evolved MECP2 variant with enhanced hmC/mC specificity to map the distribution of this asymmetric dyad in mouse embryonic stem cells.

Significance. By uncovering hmC-interacting proteins and validating selected candidates, this work provides a foundation for understanding the biological role of hmC-containing CpG dyads in chromatin dynamics and gene regulation. The identified interactors from the proteomics study serve as a basis for future functional studies, while the genome-wide map of hmC/mC dyads offers spatial context for these interactions. Together, these complementary approaches advance our ability to explore the regulatory potential of asymmetric hmC-containing CpG dyads in the mammalian genome.

3 Results

3.1 Generation of DNA probes containing modified CpG dyads

The employment of DNA pull-down studies to identify reader proteins of epigenetic DNA modifications is an established technique that was used in multiple studies as discussed in (**Subsection 1.2.5**). The discovery of such reader proteins of modified CpG dyads requires careful consideration of the DNA probe. The probe design must account for sequence context and length, number of CpGs, CpG islands, flanking sequences of CpGs, off-target Cs and their modifications and cost. This section will discuss these factors while suggesting different DNA probes that were examined regarding their functionality and practicality for the scope of this thesis. Finally, I will introduce the approach for generating asymmetrically modified DNA probes that was ultimately used in this study.

3.1.1 Requirements and considerations for CpG-modified DNA probes

The optimal DNA probe for studying asymmetrically modified CpG dyads would contain C modifications exclusively in CpG but not CpH contexts. Such a probe needs to be generated by Solid-Phase DNA Synthesis (Phosphoramidite Chemistry) (Merrifield, 1963) by synthesizing each DNA strand with the desired C modification in deliberate CpG sites. The strands must not be palindromic to ensure only one possible (asymmetric) product. The final DNA probe is obtained by hybridization of these two strands and, as a result, is asymmetrically modified in its CpG dyads with all other Cs being unmodified. However, this approach has two major limitations, which are probe length and cost. The sequence length is restricted to a maximum of 200 bp due to cumulative inefficiencies, error accumulations and purification challenges during the synthesis workflow. Additionally, the incorporation of modified Cs significantly increases the price and multiple CpG modifications are necessary to increase the signal-to-noise ratio in proteomics experiments as will be discussed in **Section 3.3**. Hence, using Solid-Phase synthesis to obtain a DNA probe with defined C modifications is only feasible for short sequences with limited numbers of CpGs. Spruijt et al., 2013 chose such a probe design for their pull-down study resulting in a short, 27 bp probe with highly artificial context (5'-AAG.ATG.ATG.AXG.AXG.AXG.AXG.ATG.ATG-3'; with X = C, mC or hmC) (**Table 3.1**). While it contains modifications only in the target Cs, unmodified CpAs in the reverse strand and minimal spacing between CpGs by only one A could be a constraint.
An opposing perspective on the choice of probe context was given by Iurlaro et al., 2013 by using natural probe contexts corresponding to the promoter regions of Pax6 and Fgf15 in their pull-down studies, which have a length of over 200 bp and contain modified Cs in CpG and CpH contexts (**Table 3.1**). Promoter sequences play important roles in epigenetic DNA regulation. Naturally, many transcription factors will bind there and although promoter probes only present a small excerpt of the vast landscape of DNA promoter sequences, they provide a natural context which may enable insights into DNA regulation regarding modification and sequence context which can be used as a starting point for further investigations. Depending on their length, these promoter probes cannot be synthesized due to size

limit and cost factors. Instead, they can be generated by PCR using a DNA polymerase that can incorporate the corresponding modified dCTPs. This straightforward and cost effective approach allows for quick and large production of epigenetically modified DNA, however, off target C modifications will occur as differentiation between CpG and CpH sites by the polymerase is not possible. Furthermore, PCR synthesized probes will by nature always be symmetrically modified.
Taken together, certain trade-offs must be considered when it comes to probe design primarily revolving around the discussion of cost, length and off target modifications. In this study, two different probe categories were evaluated for their use in pull-down studies comprising natural promoter context probes and randomized core CG probes.

Table 3.1: DNA probes used for pull-down-MS experiments.

Probe	Species	Length (bp)	CpG	CpG (%)	CpH fw	CpH rv	Locus
Arbitrary seq.[Spruijt et al., 2013]	NA	27	4	22.2	0	5	NA
Fgf15[Iurlaro et al., 2013]	m	248	14	11.3	57	48	chr7:144,449,397-144,449,644 (GRCm39)
Pax6[Iurlaro et al., 2013]	m	280	8	5.7	48	61	chr2:105,498,259-105,498,538 (GRCm39)
VEGFA	h	201	11	10.9	30	73	chr6:43,769,579-43,769,779 (GRCh38)
MUC1	h	193	6	6.2	40	48	chr1:155,193,255-155,193,447 (GRCh38)
LINE1	h	372	26	14.0	55	62	e.g. chrX:81,841,201-81,841,572 (GRCh38)
Sp1	m	241	17	14.1	67	71	chr15:102,314,552-102,314,792 (GRCm39)
8NX	NA	92	1	2.17	9	9	NA
12NX	NA	98	2	4.08	9	9	NA
16NX	NA	104	3	5.77	9	9	NA

Natural promoter context probes. Natural promoter context probes are derived from selected sequences within promoter regions of genes that are associated with disease or dysfunction upon hyper- or hypomethylation, ranging in length from 200 to 400 base pairs.

The first promoter context probe was chosen from the promoter region of a growth factor corresponding to Iurlaro et al., 2013. For this probe a part of the human VEGFA promoter region was chosen (relative position to the gene is illustrated in **Figure 3.1a**). VEGFA is a pivotal regulator of angiogenesis (Neufeld et al., 1999) and earlier studies indicated that promoter CpG methylation and overexpression of TET3 can suppress VEGFA expression (L. Bai et al., 2021; Gorenjak et al., 2020). Aberrant VEGFA promoter methylation has been associated with pathological outcomes, such as various types of cancer and inflammation (C. Lu et al., 2010; Marsit et al., 2010; Ping et al., 2013), making VEGFA a promising target for cancer chemotherapy (Dvorak, 2002). Indeed, targeted methylation of the VEGFA promoter with a Dnmt3a-3L sc variant has been accomplished to silence VEGFA expression (Siddique et al., 2013). The VEGFA probe used in this study is 201 bp long and contains 11 CpG dyads (**Table 3.1**).

A second promoter context probe was derived from the promoter region of MUC1 (relative position to MUC1 gene is illustrated in **Figure 3.1b**). MUC1 is aberrantly overexpressed in various cancers and its promoter is tumor associated (L. Chen et al., 1995; Yamada et al., 2008). The MUC1 promoter sequence used as probe in this study originated from Invivogen, 2025 and encompasses a region that is responsible for regulation of MUC1 expression (L. Chen et al., 1995), which is 193 bp long and contains 6 CpG dyads (**Table 3.1**).

As a third promoter context probe, part of the LINE1 promoter was chosen (relative position to LINE1 gene is illustrated in **Figure 3.1c**). LINE1 is heavily methylated repeat DNA (J. Bender, 1998). Methylation of its promoter presents the major mechanism of repressing retrotransposition of LINE1 elements (Hata and Sakaki, 1997; Woodcock et al., 1997). Demethylation of essential CpG sites leads to

increased levels of LINE1 transcription (Hata and Sakaki, 1997) and such hypomethylation of LINE1 is expected to promote instability of the genome (Florl et al., 1999), hence linking hypomethylation to tumorigenesis (Roman-Gomez et al., 2005). However, the molecular mechanisms underlying hypomethylation of LINE1 elements are not fully understood yet. The LINE1 probe sequence was retrieved from Roman-Gomez et al., 2005. It is 372 bp long and contains 26 CpG dyads (**Table 3.1**).

As a last promoter context probe the mouse Sp1 promoter was chosen (relative position to the Sp1 gene is illustrated in **Figure 3.1d**). Sp1 is a transcription factor that is highly upregulated during early embryogenesis (Kojima et al., 2014) and is involved in multiple processes such a cell cycle regulation, hormonal activation and angiogenesis (Tapias et al., 2008). Sp1 regulates multiple genes including the Sp1 gene itself and is proposed as an key mediator of cell cycle associated changes in gene expression (Tapias et al., 2008). The Sp1 probe has a length of 241 bp and has 17 CpG dyads (**Table 3.1**).

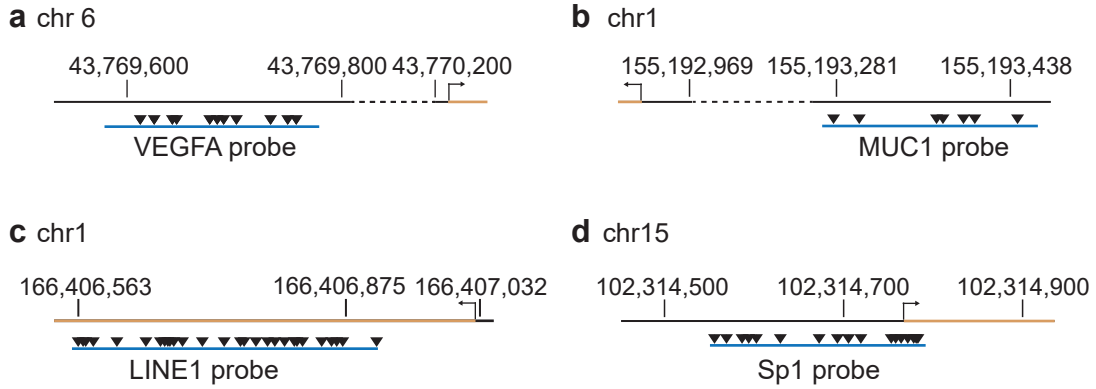


Figure 3.1: Schematic representation of the genomic locations of promoter context probes. DNA probes (**(a)** VEGFA, **(b)** MUC1, **(c)** LINE1 and **(d)** Sp1) are depicted in blue, with black triangles indicating CpG sites. The probes are positioned upstream of the gene, which is shown in yellow. Chromosomal locations and nucleotide positions are provided for reference.

Randomized core CG probes Randomized core CG probes are synthetic DNA probes with a fixed central CpG that is flanked by random 4-mers. Adjacent primer binding sequences next to the random region were introduced for PCR amplification. These adjacent sequences are designed in a way that off-target Cs are not modified during PCR but fully covered by the PCR primers. These orthogonal primer sequences were modified from Subramanian et al., 2018 ensuring that no G is present in the primer sequences to prevent off-target modified C incorporation. Probes can contain either one or multiple repeats of 5'-NNNNCGNNNN-3' and are named accordingly as 8NX (8N = 8 randomized nucleotides, X = C, mC, or hmC modification of central CpG), 12NX or 16NX (**Figure 3.2, Table 3.1**). The random flanking nucleotides enable the identification of both sequence- and modification-specific binders while overcoming the restrictions imposed by sequence specificity in binder identification. Indeed, 65,000 possible CpG sequence contexts arise with the 8NX probe and 16,777,216 and 4,294,967,296 for the 12NX and 16NX probes, respectively. It should be mentioned that the random nucleotides can also be C and therefore will be accordingly modified yielding off-target modifications of 0.375 CpGs on average to the central binding. The probes have a length of 92, 98 and 104 bp, respectively, and contain one, two or three CpG dyads (**Table 3.1**).

3.1.2 Generation of asymmetrically modified DNA probes

To generate strand asymmetric CpG modified DNA probes for this study, a PCR-based approach was developed. During the extension phase of PCR, the DNA polymerase moves along the template strand

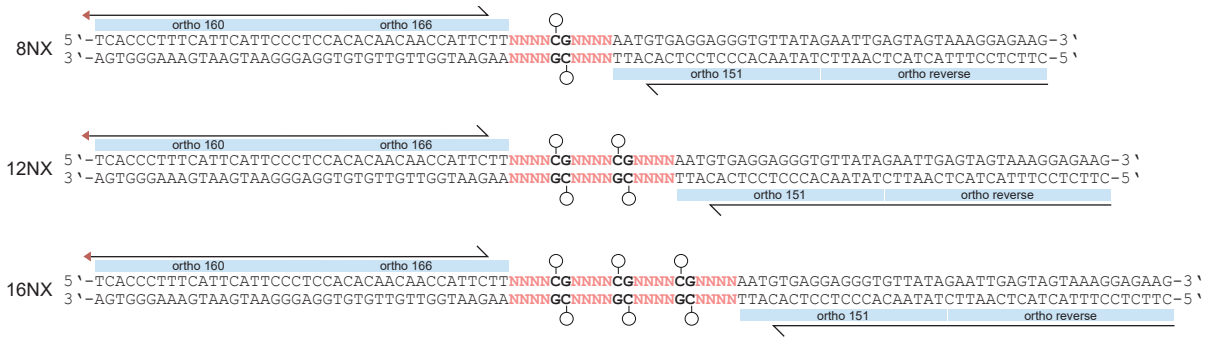


Figure 3.2: Randomized core CG probes. 8NX, 12NX and 16NX probes share the same sequence, differing only in the number of 5'-NNNNCGNNNN-3' repeats, where the central C can carry a modification represented by a circle. Random nucleotides (N) are shown in red. PCR primers cover the flanking sites and have been adapted and changed from Subramanian et al., 2018 and are named accordingly. The forward primer is biotinylated, as indicated by the red triangle.

and synthesizes a new complementary strand by incorporating nucleotides complementary to the template. By providing only modified 2'-Desoxycytidin-5'-triphosphats (dCTPs), all Cs in the final dsDNA PCR product will carry the corresponding modification. First, the efficiency of common high-fidelity polymerases including Taq Polymerase, Dreamtaq™ Polymerase and Phusion® Polymerase was examined regarding their capability to incorporate dCTP, 5-Methyl-2'-deoxycytidine-5'-triphosphate (mdCTP), 5-Hydroxymethyl-2'-deoxycytidine-5'-triphosphate (hmdCTP) and 5-Carboxy-2'-deoxycytidine-5'-triphosphate (cadCTP) using different DNA templates (**Figure 3.3a-c**). Taq Polymerase successfully integrated C, mC and hmC, though mC and hmC were integrated to a lesser extent. Additionally, it failed to integrate caC. Dreamtaq™ Polymerase also incorporated C, mC and hmC, albeit with slightly lower yields for mC and hmC (**Figure 3.3b**), but overall yielded more product than Taq Polymerase. Like Taq Polymerase, it was unable to integrate caC. Phusion Polymerase successfully integrated all tested modified dCTPs, though with significantly lower yields compared to unmodified dCTP. Based on these results, Dreamtaq™ Polymerase was selected for all subsequent PCRs for probe generation. A similar test was conducted for the 8NX probe, showing that Dreamtaq™ Polymerase can successfully incorporate mC and hmC into randomized CG core probes as well (**Figure A.1**).

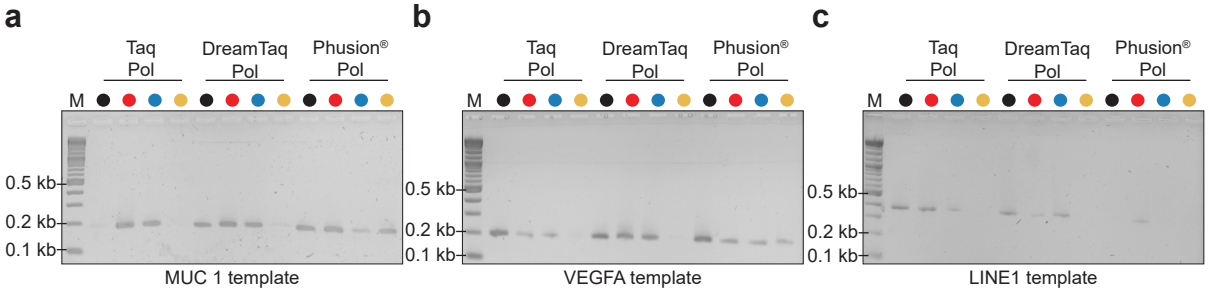


Figure 3.3: Incorporation of modified cytosines by PCR. Taq Polymerase, Dreamtaq™ Polymerase and Phusion® Polymerase were tested for their capabilities of integrating dCTPs (black circle), mdCTPs (red circle), hmdCTPs (blue circle) and cadCTPs (yellow circle) into the templates of **(a)** MUC1, **(b)** VEGFA and **(c)** LINE1. The PCR products have expected sizes of 197, 201 and 370 bp respectively. Samples were run on a 2 % agarose gel with a 1 kb Plus DNA Ladder (New England Biolabs; lane M).

PCR products inherently have symmetrically modified CpG dyads. To generate a DNA probe with asymmetrically modified CpG dyads, an alkaline denaturation approach followed by primer extension was applied. Alkaline denaturation is a technique that is widely used for producing ssDNA, usually in the context of identifying aptamers by SELEX (Oh et al., 2010; Sefah et al., 2010). Using a biotiny-

lated forward primer during PCR, the symmetrically modified PCR product could be immobilized to streptavidin-linked magnetic beads via its 5'-biotin (**Figure 3.4a**). Immobilization efficiency was evaluated by varying the incubation time of the MUC1 probe with the streptavidin beads. Results indicated that incubation of at least 1 h is sufficient, as longer incubation times or increased bead amounts did not improve efficiency. This was determined by analyzing the unbound fraction in the supernatant via agarose gel electrophoresis (**Figure 3.4 b**). After 15 min 15 % of DNA was not immobilized, whereas after 1 h or longer, approximately 10 % of DNA was not immobilized. Fluorescently labeled PCR primers were used to label the forward strand with FAM and the reverse strand with Cy5. Immobilization efficiency was then assessed by measuring Cy5 emission of the supernatant containing unbound DNA. Yield was determined by generating standard curves with selected concentrations of Cy5-labeled DNA (**Figure A.2a**). 1 h of DNA probe incubation with the beads yielded very high immobilization efficiency for MUC1 and VEGFA probe with less than 5 % unbound DNA, however LINE1 immobilization was not as successful with around 30 % unbound DNA (**Figure 3.4c**). Strand separation was achieved through alkaline denaturation with sodium hydroxide (NaOH). The biotinylated forward strand was subsequently removed using a magnet, retaining only the reverse strand (**Figure 3.4a**). A common critique of NaOH denaturation is that significant amounts of the biotinylated strand are also eluted due to streptavidin denaturation (Paul et al., 2009). To determine the optimal amount of NaOH that effectively denatures the strands without denaturing streptavidin and consequently removing the forward strand from the beads, a NaOH titration was performed. First, the sufficient NaOH amount needed to obtain the Cy5 strand was determined. Therefore, increasing amounts of NaOH were added stepwise to the immobilized DNA probe and the Cy5 emission of the eluted strand was detected. Results indicated that the majority of the Cy5 strand was already eluted with 15 mM NaOH (**Figure 3.4d**). To maximize the yield of ssDNA, 20 mM NaOH was used for subsequent denaturations. To address the question of streptavidin denaturation, FAM fluorescence was measured after NaOH incubation alongside Cy5 fluorescence to assess the presence of the forward strand in relation to the reverse strand (**Figure 3.4e**). Yield was determined by generating standard curves with selected concentrations of Cy5- and FAM-labeled DNA (**Figure A.2b, c**). The Cy5-labeled reverse strand was obtained with a yield of around 50 - 60 % for the VEGFA and MUC1 probe while the yield of LINE1 Cy5-labeled strand was much lower which can be linked to the low immobilization efficiency of LINE1 (**Figure 3.4c**). FAM fluorescence was very low for all three probes, and given the imperfect standard curves (**Figure A.2c**), it could either be background signal or negligibly small amount of forward strand, suggesting no denaturation of streptavidin (**Figure 3.4e**). In conclusion, 20 mM of NaOH is the optimal concentration for obtaining high yields of the reverse strand, free from contamination by the forward strand, which is significantly lower than the typically reported NaOH concentration in the literature of around 100-200 mM (Oh et al., 2010; Sefah et al., 2010).

After confirming the recovery of the reverse strand, the forward strand synthesis was performed to incorporate the modification of choice. Therefore, a biotinylated forward primer was hybridized to the reverse strand, followed by primer extension using Klenow Polymerase utilizing only the desired modified dCTPs. This resulted in a biotinylated DNA probe with strand (asymmetric) CpG modifications (see **Figure 3.5a** for complete procedure). The workflow was monitored by fluorescent labeling of the strands, as exemplified with the VEGFA probe in **Figure 3.5**. As described before, the forward strand was labeled with FAM and the reverse strand with Cy5. After alkaline denaturation with 20 mM NaOH, a Cy3-labeled forward primer was annealed to the reverse strand. Samples were collected after PCR (1), alkaline denaturation (2), primer hybridization (3) and primer extension (4) as visualized in **Figure 3.5a**. 50 ng of each sample were run on an agarose gel and then imaged using a fluorescence scanner. The PCR product was visible in lane 1 of both the FAM and Cy5 channels (**Figure 3.5b**). After alkaline denaturation, the FAM channel no longer showed a DNA band (lane 2), while the band

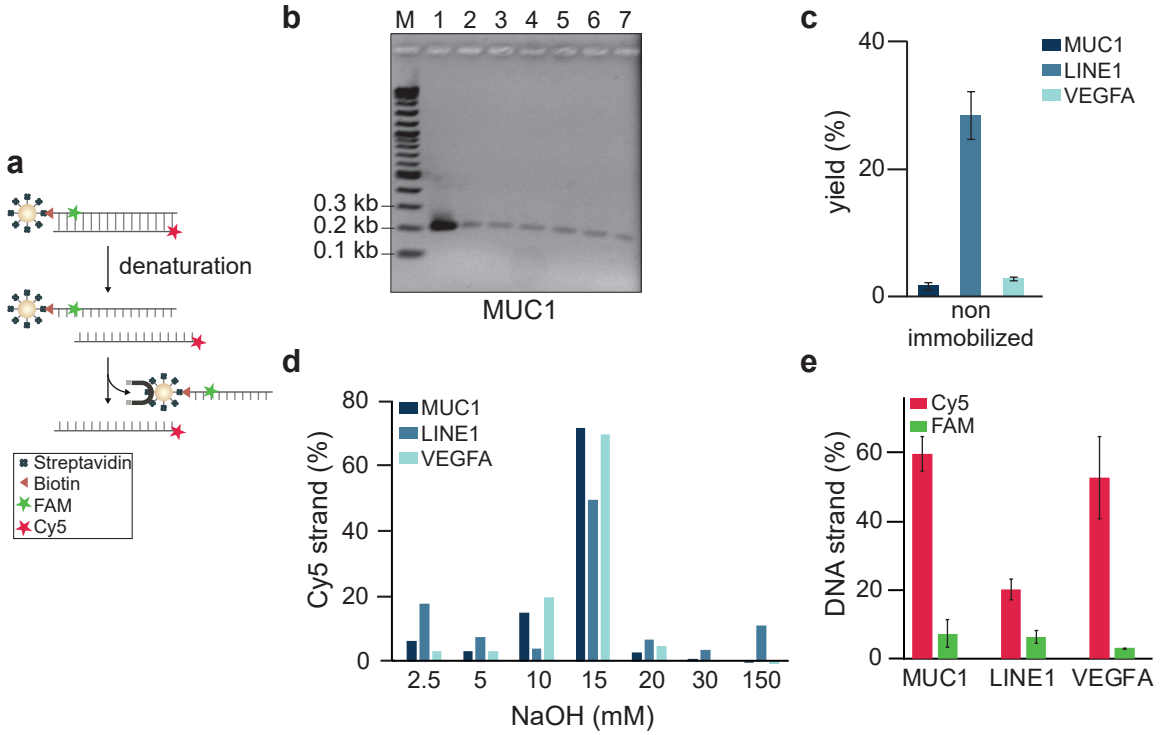


Figure 3.4: Strand separation by alkaline denaturation. **(a)** Schematic representation of the alkaline denaturation workflow. A 5′-biotin labeled PCR product is immobilized on magnetic streptavidin beads. After denaturation, the reverse strand is retained by removing the forward strand that is bound to the beads. For verification assays the forward strand was FAM-labeled (green star) and the reverse strand Cy5-labeled (red star). **(b)** MUC1 PCR product (lane 1) and supernatants with unbound DNA after varying incubation times of MUC1 with streptavidin magnetic beads (lane 2: 15 min, lane 3: 1 h, lane 4: 2 h, lane 5: 4 h, lane 6: 16 h, lane 7: 1 h and double amount of beads). 2 % agarose gel electrophoresis, 1 kb Plus DNA Ladder (New England Biolabs; lane M). **(c)** Non immobilized DNA probe after 1 h incubation with streptavidin beads, determined by Cy5 emission of unbound fraction. Yield was determined from standard curve (**Figure A.2**). **(d)** Yield of Cy5-labeled reverse strand after stepwise addition of NaOH corresponding to Cy5 standard curve (**Figure A.2**). **(e)** Yield of Cy5-labeled reverse strand (red) and FAM-labeled forward strand (green) after alkaline denaturation with 20 mM NaOH. Yield was determined from standard curves (**Figure A.2**).

was still present in the Cy5 channel, indicating successful strand separation and isolation of the reverse strand. Additionally, the ssDNA amount appeared to be higher as indicated by a stronger band, which is due to using the same mass of 50 ng and not the same molar concentration of DNA sample. After annealing the Cy3-labeled forward primer, its presence was detectable in the Cy3 channel in lane 3, along with excess primer running low in the same lane (**Figure 3.5b**). However, a slight FAM signal was also detected in lane 3, attributed to the overlap in emission spectra, which can also be seen in the control sample in lane 8 containing pure Cy3 PCR product (Cy3 forward primer, unlabeled reverse primer). Since no FAM signal was observed in the ssDNA fraction in lane 2, it can be assumed that no contamination by the FAM strand is present. Lanes 4 and 5 show the final primer extended product. Extension was either carried out with unmodified dCTPs (lane 4) or hmdCTPs (lane 5) to obtain either a product that equals the PCR product or a true asymmetric product. Lanes 6-9 are control PCR products that are either unlabeled or singly fluorescently labeled. After fluorescence detection, the gel was stained with ethidium bromide (EtBr) (**Figure 3.5c**), which confirmed that all dsDNA products (lane 1 and lanes 4-9) migrated at the same height and intensity, while lanes 2 and 3 showed ssDNA products that ran slightly different.

In theory, the PCR and the primer extended product should have the same properties as they are both synthesized by a polymerase (Dreamtaq^{TM} or Klenow), cleaned up with the same PCR clean-up

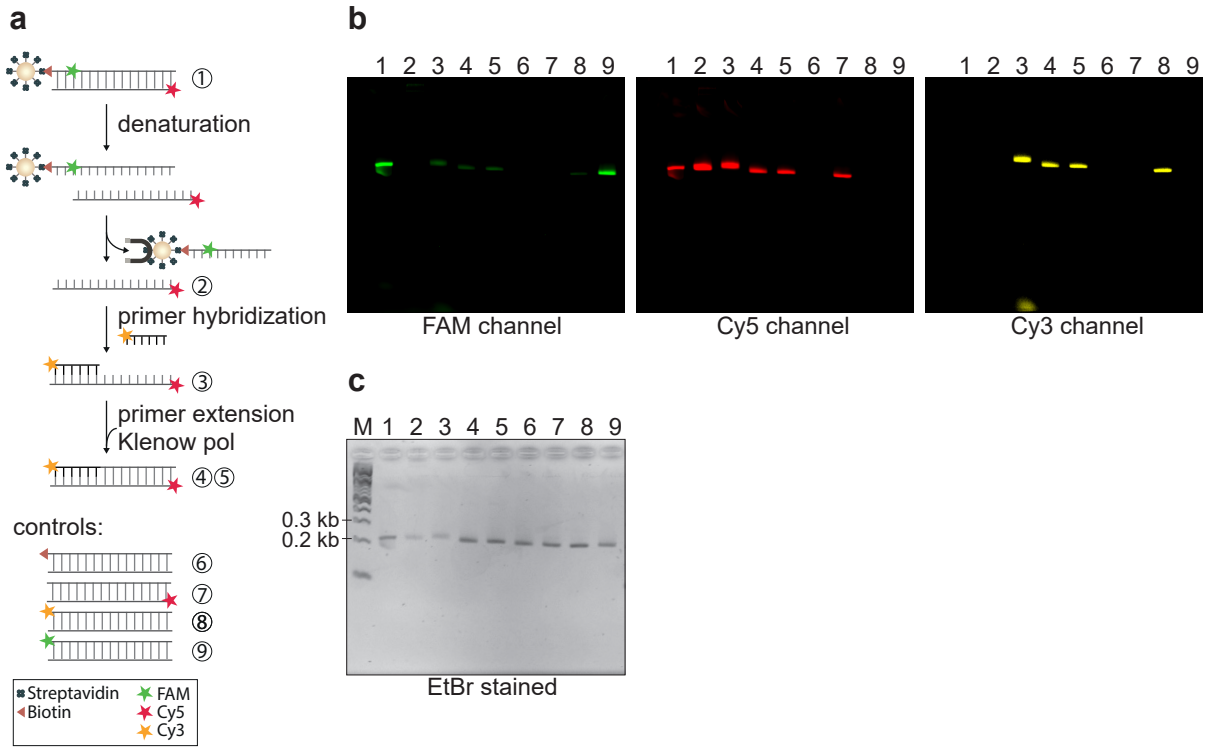


Figure 3.5: Asymmetric probe generation. (a) Schematic illustration of the asymmetric probe generation workflow with fluorescent labeling depicting the biotinylated PCR product immobilized on streptavidin beads (1), the reverse ssDNA after alkaline denaturation (2), the primer annealed product (3) and the final primer extended product obtained with dCTPs (4) or hmdCTPs (5). Additionally, all control PCR products with different fluorescent labeling are shown which are unlabeled (6), Cy5 labeled (7), Cy3 labeled (8) or FAM labeled (9). **(b)** Fluorescent scanning of an 2 % agarose gel containing the samples from **(a)**, with corresponding numbering. The gel was imaged in the FAM, Cy5 and Cy3 channel. **(c)** The same agarose gel was stained in EtBr to visualize all DNA bands including the 1 kb Plus DNA Ladder (New England Biolabs; lane M).

kit and migrate in the same way on the agarose gel. To further verify this, two additional steps were undertaken. First, all PCR as well as symmetric and asymmetric primer extended products were subjected to Sanger sequencing (**Figure A.3**). All sequencing traces were very clean, with no mutations or gaps, confirming that all modified Cs were incorporated and extension was completed. Second, restriction enzyme digests were carried out to see if the enzymes would behave in the same way with probes that were generated in different ways. As restriction enzymes can be modification sensitive, the asymmetric probe generation workflow was carried out with unmodified dCTPs in the Klenow extension step, yielding an unmodified primer extension product that should be equal to the initial PCR product. All three promoter context probes were subjected to the asymmetric probe generation assay and the four quality control steps of PCR product (1), ssDNA (2), primer hybridized product (3) and primer extended product (4) can be seen in **Figure 3.6a**. As expected, the dsDNA products in lane 1 and 4 migrated in the same way, while the full or partial ssDNA products in lanes 2 and 3 ran slightly higher and are less concentrated. Restriction enzymes were chosen as 6+ cutters that yield visible bands on an agarose gel. For MUC1 *MspI* digest resulted in bands that run at a height of 95, 66 and 32 bp. For LINE1 the *NheI* digest yielded bands in the height of 338 and 34 bp and digest of VEGFA with *HphI* yielded bands of 99, 52 and 50 bp showing that restriction enzymes engage in the same way with PCR and primer extended products. Please note that running the digested probes on an 2 % agarose gel did not resolve bands below 50 bp, however as there were no visible bands at the initial position of the full length probes, it is assumed that the probes were completely digested. Although sequencing and restriction digest results indicated that that PCR and primer extended products pose the same properties, it was decided

that all promoter context DNA probes are prepared with the asymmetric probe generation protocol even if the desired product is symmetric to make sure that pull-down results are comparable between symmetric and asymmetric probes and not biased by the probe generation workflow. Visualization of the four critical steps of the asymmetric probe generation protocol ((1) PCR product, (2) ssDNA, (3) primer hybridization, (4) primer extended product) was used as a quality control step each time before using these probes in downstream assays. The randomized core CG probes were prepared in this way as well and could be visualized as shown in **Figure 3.6b**. The overall yield of the primer extension product differed slightly between different probes depending on their complexity (**Figure 3.6c**). The yield for LINE1 probe generation was usually lower which most likely resulted from insufficient probe immobilization as was shown in **Figure 3.4c**. Since the VEGFA probe yielded the highest amount of product, this promoter context probe was used in the subsequent assays. This protocol is applicable to various DNA sequences, such as those from different species or loci that for technical reasons cannot be hybridized.

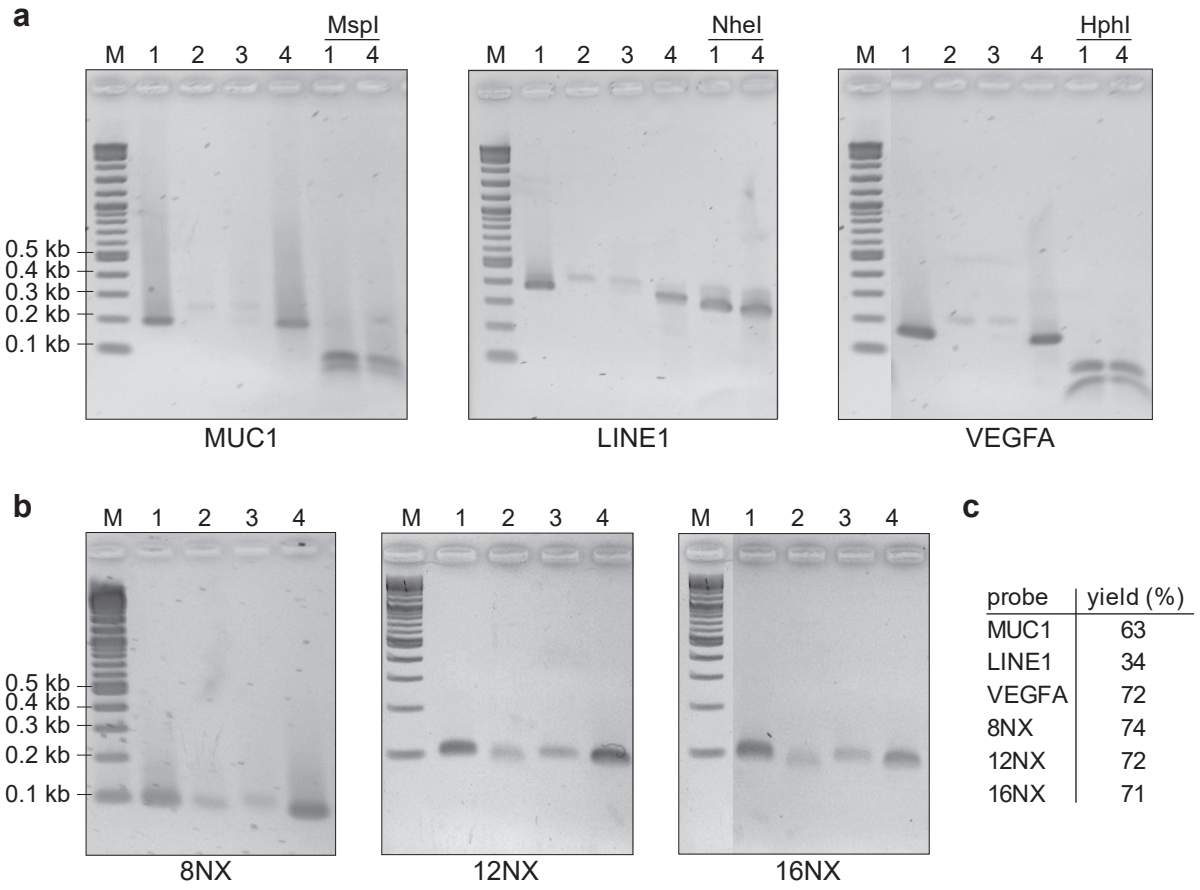


Figure 3.6: Comparison of PCR and primer extended probes.(a) Probe generation of MUC1, LINE1 and VEGFA was monitored on an agarose gel by applying samples from (1) PCR product, (2) ssDNA, (3) primer hybridized product and (4) primer extended product. dsDNA products (1) and (4) were subjected to restriction digest with *MspI*, *NheI* and *HphI*, respectively. **(b)** Asymmetric probe generation workflow monitored for the randomized core CG probes 8NX, 12NX and 16NX. Numbers 1-4 correspond to labeling in **(a)**.

Since the randomized core CG probes are shorter than promoter context probes and only contain one to three CpGs that need to be modified, they could as well be synthesized and hybridized offering a simpler and more elegant alternative to primer extension. However, a potential pitfall is that the random nucleotides might not perfectly match given that there are (at least) 65,000 different sequence contexts. Even though the thermodynamically most favorable product should be obtained by grad-

ual and slow cooling after denaturation, mismatches of single nucleobases are likely to persist. This could pose a challenge especially in an assay that sets out to find binding proteins as it may attract DNA repair enzymes and could introduce another layer of bias into a competitive binding assay. To check for hybridization mismatches, hybridized 8NC/C and 8NmC/mC, along with control probes, were digested with T7 endonuclease I (which preferentially cleaves CC mismatches) and mismatch endonuclease I (which preferentially cleaves TT mismatches). The control oligonucleotides were designed in a way that the hybridized probe contained either no mismatch or specific CC or TT mismatches (**Figure A.4**). Undigested probes were always loaded as controls, yielding a sharp band (**Figure 3.7**). Control probes without mismatches (0 CG mismatch and 0 AT mismatch) also showed sharp bands after digestion, indicating no off-target cleavage by the enzymes. However, the probe with eight defined CC mismatches was completely cleaved by T7 endonuclease I, and the probe with one defined TT mismatch was degraded by mismatch endonuclease I and partially degraded by T7 endonuclease I. The hybridized 8NC/C and 8NmC/mC probes showed a slight smear and an apparent lower concentration of fully double-stranded probe compared to the undigested probe, suggesting imperfect hybridization. Different sequence contexts likely led to various mismatches, making it impossible to obtain sharp bands. However, the slight smear indicated that hybridization may not work perfectly for these kinds of probes. Therefore, also the randomized CG core probes were generated with the asymmetric probe generation protocol as Klenow Polymerase ensures the correct nucleotide incorporation.

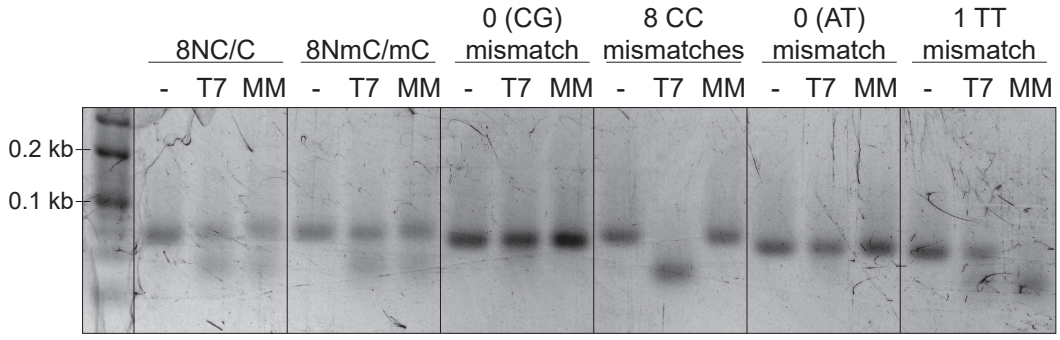


Figure 3.7: Mismatch assay for randomized core CG probes. 8NC/C and 8NmC/mC probes were generated by hybridization along with control probes that either did not contain mismatches (0 CG and 0 AT), eight defined CC mismatches (8 CC), or one defined TT mismatch (1 TT). Hybridized probes were digested with T7 endonuclease I (T7) or mismatch endonucleaseI (MM) and analyzed by 2 % agarose gel electrophoresis with a 1 kb Plus DNA Ladder (New England Biolabs; lane M).

3.1.3 Essential points

- The decision for asymmetrically modified DNA probes for pull-down studies is a deliberate trade-off between length, cost and off-target modifications. Promoter context DNA probes and randomized CG core probes have been discussed.
- An asymmetric probe generation protocol was established and validated in order to generate DNA probes for pull-down assays to identify natural reader proteins in the following steps. This protocol is applicable to various DNA sequences, such as those from different species or loci, that for technical reasons cannot be hybridized.

The Sp1 probe used in this chapter was designed by Zeyneb Vildan Çakıl. The 8NX probe used here is based on an original design by Dr. Damian Schiller, which I subsequently optimized for the current study. I gratefully acknowledge their contributions.

3.2 Extraction of nuclear proteins

In a quest that sets out to identify DNA-binding proteins it is reasonable to use only nuclear proteins as input. This ensures the reduction of nonspecific binding and background signals caused by the presence of cytoplasmic proteins, which are typically more abundant than nuclear proteins. To obtain only a specific subset of cellular proteins, subcellular fractionation can be employed, a method that emerged first around 80 years ago by Claude, 1946. Subcellular fractionation allows for the separation of cellular components due to their distinct characteristics, such as size, shape, density and solubility. Over time, protocols have been refined and optimized enabling the complete separation of nuclear and cytoplasmic fractions. Here, I employed a method that utilizes cell lysis in a low-salt buffer supplemented with detergent followed by a high-salt extraction of proteins from the nuclei, as previously described by Schreiber et al., 1989. The buffer conditions and pH values have been optimized as described by Dignam et al., 1983. The basic protocol was adapted from Y. Luo et al., 2014. Their approach was adhered to while omitting the sucrose density centrifugation step, relying instead on differential centrifugation. Nuclear proteins can be extracted from cultured cells or animal tissues using similar protocols, with variations primarily in the cell lysis step. Differences in protocols and purity of cellular fractions will be discussed in the following sections.

3.2.1 Extraction from cultured cells

Nuclear proteins were extracted from HEK293T cells. To use these proteins in binding studies it is of high importance that they remain functional during and after extraction, hence mild conditions need to be applied in the extraction process. As mentioned above, the protocol by Luo et al. was employed with slight adaptations in the buffers and workflow, utilizing hypotonic detergent cell lysis followed by high-salt extraction of nuclear proteins (Y. Luo et al., 2014) (**Figure 3.8a**). In the first step the cells were gently lysed to enable the separation of cytoplasm and nuclei. This was accomplished through a combination of osmotic shock in a low-salt buffer and subsequent homogenization using a detergent and mechanical disruption by repeatedly pipetting the swollen cells with a yellow pipette tip. IGEPAL® CA-630 was used as a nonionic detergent that disrupts cell membranes while preserving protein integrity. Instead of subjecting the lysed samples to sucrose density gradient centrifugation as described by Y. Luo et al., 2014, differential centrifugation was used. In the initial low-speed centrifugation step, nuclei were pelleted while the soluble cytoplasmic fraction remained in the supernatant. Although density gradient centrifugation is supposed to provide higher purity, differential centrifugation offers a faster alternative while still achieving adequate fraction separation. To minimize cytoplasmic protein contamination, the nuclear pellets were washed in detergent-free low-salt buffer. The nuclear pellets were then incubated in a high-salt extraction buffer containing 420 mM NaCl as optimized by Dignam et al., 1983 which caused the nuclei to shrink, facilitating the extraction of nucleic acid binding proteins through the nuclear pores, allowing them to solubilize in the buffer. To isolate the solubilized proteins, high-speed centrifugation was applied. The separation of cytoplasmic and nuclear fractions was validated by Western blotting using marker antibodies (**Figure 3.8b, c**). anti-β-Tubulin antibody was used as a cytoplasmic marker and anti-Lamin-A/C antibody as a nuclear marker and the presence of the corresponding proteins was assessed in different batches of extracts prepared as biological replicates. The absence of Lamins A and C in the cytoplasmic fraction confirmed that the lysis conditions were sufficiently mild to preserve nuclear integrity (**Figure 3.8c**). However, β-Tubulin was detected in the nuclear fraction suggesting that cell lysis was not completed which led to contamination of the nuclear fraction by cytoplasmic proteins (**Figure 3.8b**). This does not present a problem but rather biases the nuclear protein concentration determined by BCA assay, as the sample is not purely nuclear protein. However, it may pose a challenge

for reproducibility. Nonetheless, across different biological replicates the effects appeared consistent when comparing the ratio of β-Tubulin in the cytoplasmic and nuclear fractions. Moreover, the yield was comparable, ranging from 15 to 19 μg of proteins per 1 million cells (**Table 3.2**). More stringent lysis conditions may help to improve the purity but would also risk a loss of nuclear proteins in the cell lysis step. This protocol was used for nuclear protein extraction from HeLa and E14TG2a mESCs as well demonstrating its applicability across different cell lines (**Figure A.5**).

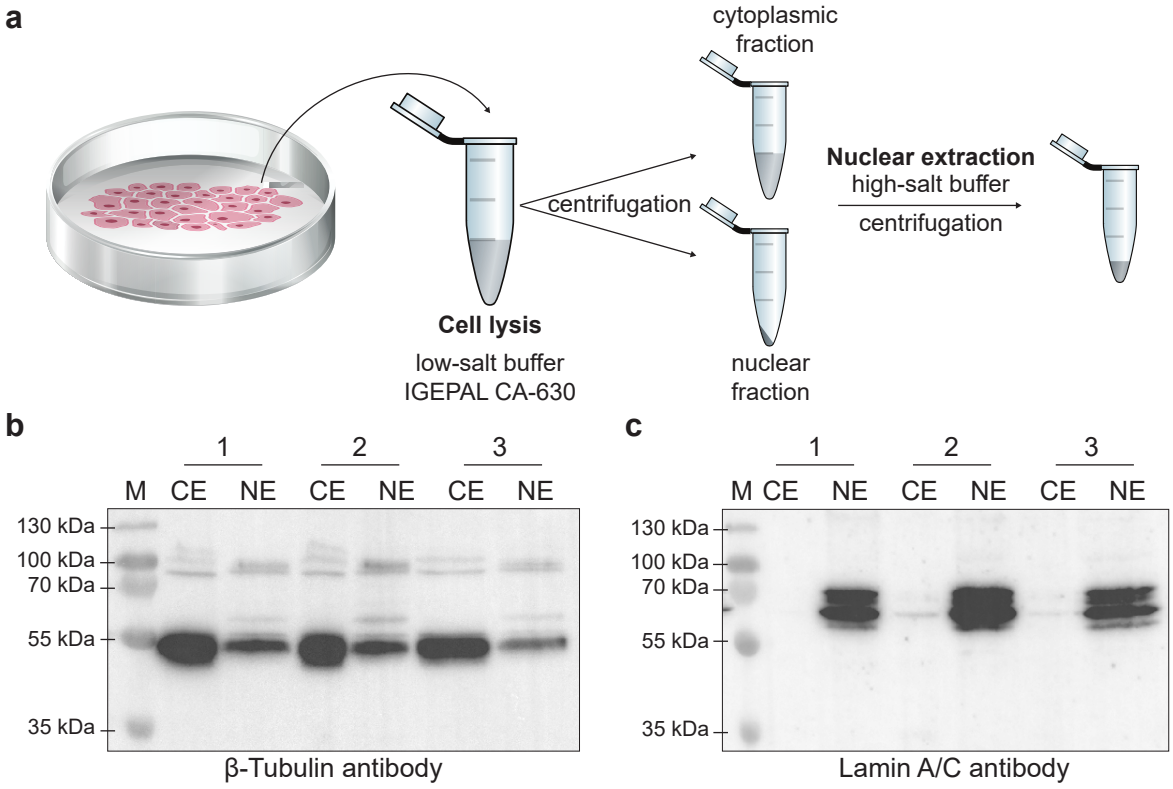


Figure 3.8: Nuclear protein extraction from HEK293T cells. (a) Schematic representation of the nuclear extraction workflow. Cultured cells were lysed in a low-salt buffer supplemented with detergent and cytoplasmic and nuclear fractions were separated by centrifugation. Addition of a high-salt buffer to the nuclei facilitated nuclear protein extraction. The schematic tubes are a modified version of a public domain vector graphic (original source: OpenClipart, SVG ID 55801, available at https://freesvg.org/opened-chemical-tube). **(b)** Western blot showing the presence of β-Tubulin in cytoplasmic extracts (CE) and nuclear extracts (NE) from three biological replicates (1-3). **(c)** Western blot showing the presence of Lamins A and C in CE and NE from three biological replicates (1-3). Samples were run on a 12 % SDS-PAGE with a Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder (lane M)), expected molecular weights: β-Tubulin 55 kDa, Lamins A and C: 63 and 70 kDa.

Table 3.2: Nuclear extraction yields from cultured HEK293T cells.

Replicate	Passage number	Number of cells	Nuclear extract concentration (µg/ml)	Mass (mg)	Yield (µg/1 mio. cells)
1	29	$160.05*10^6$	2.11	2.90	18.12
2	21	$142.925*10^6$	1.62	2.27	15.88
3	29	$191.2*10^6$	1.70	3.74	19.56

3.2.2 Extraction from animal tissues

To extract nuclear proteins from mouse brain tissue, the protocol was changed in the cell lysis step according to Lahiri and Ge, 2000. Utilizing a glass dounce homogenizer achieved the disruption of tissue and cells (**Figure 3.9a**). The required number of strokes needed for cell lysis was evaluated by comparing the presence of β-Tubulin and Lamins A and C in the cytoplasmic and nuclear fractions (**Figure 3.9b, c**). Increasing the number of strokes resulted in a decrease of β-Tubulin in the nuclear fraction suggesting improved cell lysis. Simultaneously, the amount of Lamins A and C also increased, likely due to more effective tissue disruption, making more cells available for lysis. Based on this effect, 15 strokes were selected as optimal. Although the protocol could have been further refined, minimizing mouse tissue waste was a priority given its value and the results were considered as sufficient. The overall yield per mg mouse tissue was in the upper range compared with Lahiri and Ge, 2000 (**Table 3.3**).

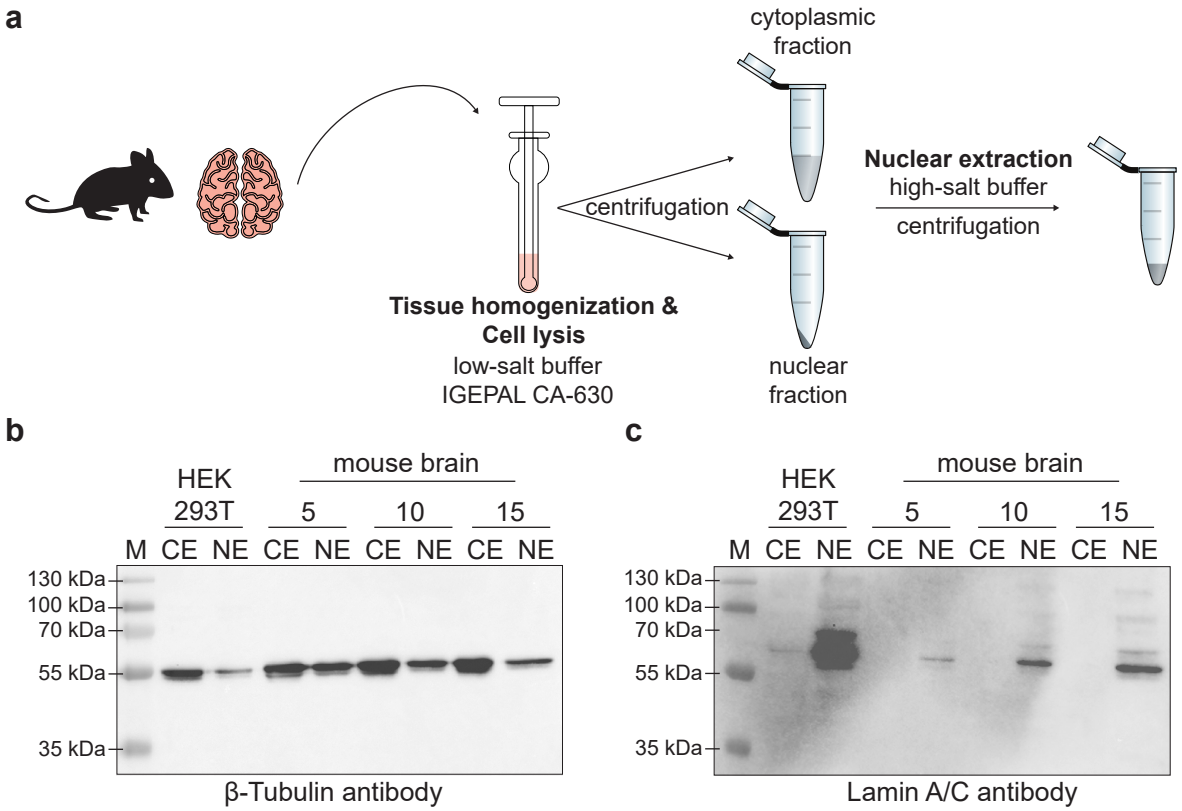


Figure 3.9: Nuclear protein extraction from mouse brain tissue. (**(a)** Schematic representation of the nuclear extraction workflow. Mouse tissue was disrupted and cells were lysed in a low-salt buffer with detergent by using a glass dounce homogenizer. Subsequent steps were identical to **Figure 3.8a**. The schematic tubes are a modified version of a public domain vector graphic (original source: OpenClipart, SVG ID 55801, available at https://freesvg.org/opened-chemical-tube) **(b)** Western blot showing the presence of β-Tubulin in cytoplasmic extracts (CE) and nuclear extracts (NE) after applying 5, 10 or 15 strokes with a dounce homogenizer. HEK293T extracts are shown for comparison. **(c)** Western blot showing the presence of β-Tubulin in CE and NE after applying 5, 10, or 15 strokes with a dounce homogenizer. HEK293T extracts are shown for comparison. Samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M), expected molecular weights: β-Tubulin 55 kDa, Lamins A and C: 63 and 70 kDa.

Table 3.3: Nuclear extraction yields from mouse brain tissue. Yield of nuclear proteins from mouse brain after 5, 10 or 15 strokes with a dounce homogenizer compared with yield from Lahiri and Ge, 2000.

Sample	tissue (mg)	Nuclear extract (mg)	Yield (µg/mg tissue)
Brain (5)	0.048	0.456	9.50

Sample	tissue (mg)	Nuclear extract (mg)	Yield (µg/mg tissue)
Brain (10)	0.050	0.201	4.02
Brain (15)	0.059	0.124	2.10
Caudate[Lahiri and Ge, 2000]	0.215	0.530	2.47

3.2.3 Essential points

- Adapting and changing existent protocols for nuclear protein extraction from cultured cells or animal tissues yielded sufficient amounts and purity of nuclear proteins as validated by marker antibodies.
- The nuclear protein extracts will be used in subsequent pull-down studies.

3.3 Pull-down of nuclear proteins using epigenetically modified DNA probes

Building on the development of DNA probes in **Section 3.1** and the establishment of nuclear protein extraction in **Section 3.2**, a workflow was set up to enable the identification of reader proteins of asymmetric duplex combinations. I intended to utilize a pull-down approach coupled to quantitative MS which is a widely used technique in the field of reader identification for epigenetically modified DNA (L. Bai et al., 2021; Iurlaro et al., 2013; Spruijt et al., 2013; J. Xiong et al., 2016). This chapter explores the establishment and applicability of the pull-down technique within the scope of this thesis.

3.3.1 Establishing a nuclear protein pull-down protocol

The pull-down approach relies on a 5'-biotinylated DNA probe carrying the desired modifications which is immobilized on magnetic streptavidin beads. Incubation with nuclear protein extracts and subsequent wash steps retain only bound proteins that are subsequently identified by Western blotting or MS (**Figure 3.10a**). Identification of proteins using Western blots requires denaturation-based elution of bound proteins from the beads typically being achieved by using heat and sodium dodecyl sulfate (SDS) and the availability of specific antibodies. During the initial establishment and optimization steps only DNA probes with unmodified or methylated Cs were used (symmetric PCR products C/C or mC/mC) and differences in their protein binding preferences were set out to be identified. The pull-down conditions were optimized with regards to protein binding conditions, aiming to ultimately enable the identification of new specific binders of other duplex combinations by MS.

Initially, it was verified if the screen could enrich known mC binders. Therefore pull-downs were performed with C or mC modified probes and Western blots were carried out to assess the presence of endogenous methyl-CpG binding domain protein 1 (MBD1) and methyl-CpG binding domain protein 2 (MBD2). Detecting endogenous proteins in Western blots after the pull-down turned out to be limited by sensitivity, likely due to their low abundance, as indicated by the weak bands observed in the nuclear extract input (**Figure 3.10b, c**). Consequently, a recombinant MBP-MBD2(aa145-225) fusion protein was employed as a benchmark. To assess the general performance of MBP-MBD2 its binding preferences towards differentially modified VEGFA probes were monitored via an EMSA (**Figure 3.10d**). The results were consistent with previous findings (Buchmuller et al., 2020), confirming that mC/mC is the preferred modification, followed by hmC/mC. Weak binding was observed for hmC/hmC modified VEGFA probes while no binding was detected with C/C and hmC/C modified VEGFA probes. Subsequently, MBP-MBD2 was used in the pull-down and incubated with immobilized DNA at varying concentrations. Presence of MBP-MBD2 was assessed by using a MBP-antibody, ruling out the detection of endogenous MBD2. This indicated specific binding to methylated DNA, however binding to unmethylated DNA was still observed (**Figure 3.10e**). Assuming the remaining C/C binding is caused by an absence of background proteins, it was tested how the pull-down of MBP-MBD2 performs in the presence of nuclear extracts. As shown in **Figure 3.10f**, specific binding of MBP-MBD2 to the methylated probe was observed with various concentrations of background nuclear proteins, and the effect of C/C vs. mC/mC binding was more enhanced, indicating that the background proteins facilitated more specific binding. However, a high concentration of nuclear protein background of 3.6 μm did not display any binding of MBP-MBD2 to mC/mC. Moreover, an aggregation of the streptavidin beads was observed at this concentration (**Figure 3.10g**), likely caused by a high local protein concentration at the DNA probe. As this NE concentration corresponds to 50 μg of proteins and constituted the final condition that was intended to be used in the pull-downs (corresponding to Iurlaro et al., 2013; Spruijt et al., 2013), the bead aggregation was counteracted by adding competitor DNA to the pull-down, with the

aim to trap unspecific DNA binders and therefore decrease the local protein concentration at the beads. Indeed, binding was improved and the beads did not aggregate when dA:dT competitor DNA was added (**Figure 3.10h, i**). As an alternative also a dI:dC competitor was tested but not further considered given its inferior performance (**Figure A.6a**). Further optimizations included the protein incubation time with the best result being overnight (**Figure A.6b**). Finally, the performance of PCR vs PE probes was tested. Both probe types showed the expected preferences, however the effects differed in the band intensities (**Figure A.6c**). As a consequence, both, symmetrically and asymmetrically modified probes, were always generated with the asymmetric probe generation protocol from **Subsection 3.1.2** to exclude any bias by means of probe generation. These initial experiments showed that the known mC/mC binder MBD2 can be captured by the methylated DNA probe in presence of an excess of nuclear proteins using the established pull-down conditions which are 3.6 μm of nuclear extracts that were supplemented with 1 nM MBP-MBD2. A Western blot was used as a proof of concept for every pull-down experiment being prepared in parallel to MS samples to validate the functionality of the probes and the workflow. An exemplary blot for a pull-down with the VEGFA probe is shown in **Figure 3.10j** which revealed that MBD2 was enriched only with the mC/mC and hmC/mC probes. This is perfectly in line with the binding preferences observed in the EMSA (**Figure 3.10d** and the aforementioned previous study (Buchmuller et al., 2020)).

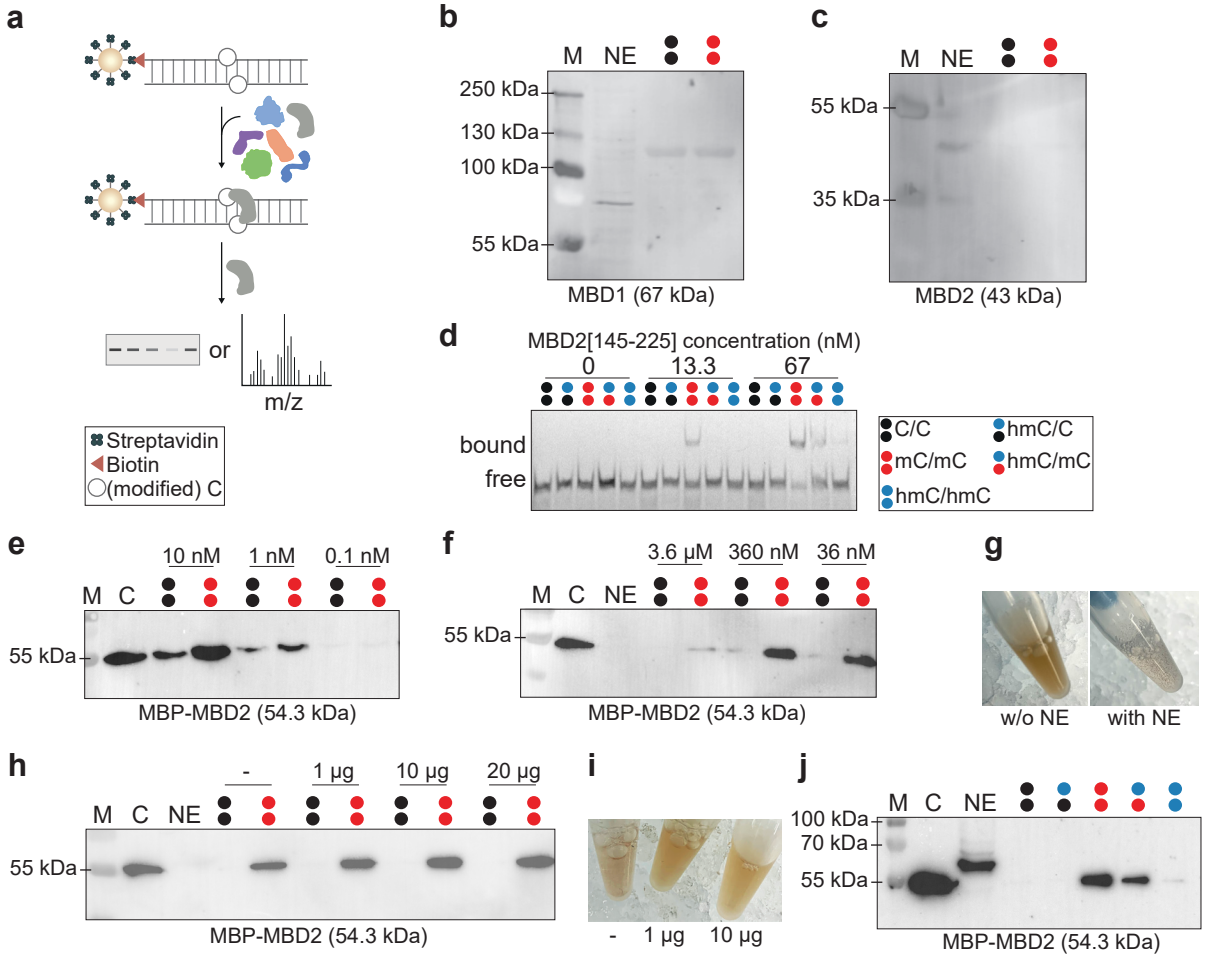


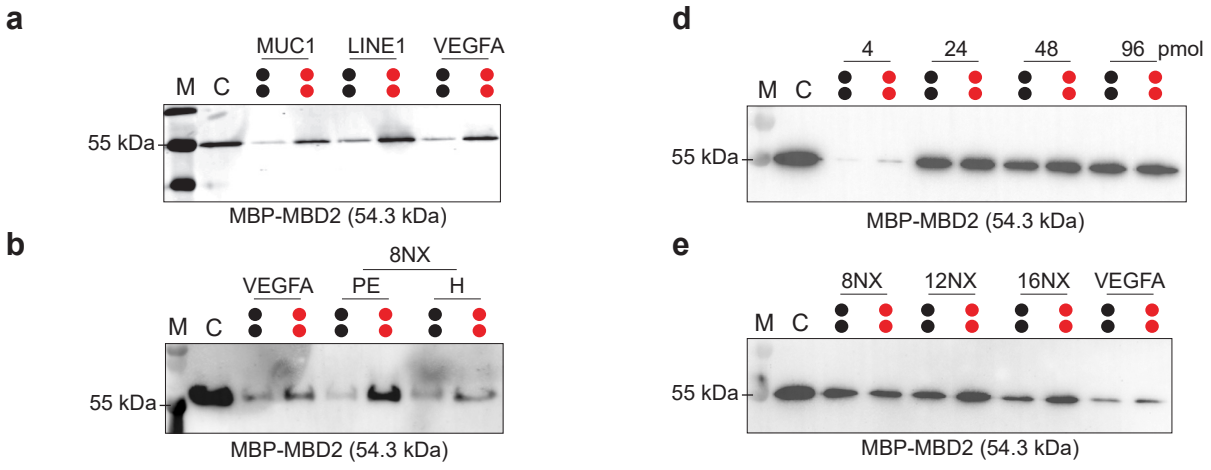
Figure 3.10: Establishing nuclear protein pull-downs with a MBP-MBD2 benchmark. (a) Schematic representation of the pull-down workflow. The modified DNA probe is immobilized on streptavidin beads via its 5′ biotin and subsequently incubated with nuclear protein extracts. Bound proteins are analyzed by Western blots or MS. **(b)** Western blot showing the presence of endogenous MBD1 in nuclear extracts (NE) and after pull-down with C/C (black circles) or mC/mC (red circles) VEGFA. **(c)** Western blot showing the presence of endogenous MBD2 in nuclear extracts (NE) and after pull-down with C/C (black circles) or mC/mC (red circles) VEGFA. **(d)** EMSA gel showing MBD2-VEGFA complex and free VEGFA probe. Samples were run on a 12 % gel. **(e)** Western blot showing binding of different concentrations of MBP-MBD2 to unmodified (black circles) or methylated (red circles) VEGFA probes. Pure MBP-MBD2 was run in lane C as control. **(f)** Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) VEGFA probes in the presence of varying concentrations of nuclear extracts. Pure MBP-MBD2 was run in lane C as control. **(g)** Appearance of beads without and with nuclear extract background during the pull-down. **(h)** Western blot showing binding of 1 nM MBP-MBD2 in the presence of 3.6 μm nuclear extracts to unmodified (black circles) or methylated (red circles) VEGFA probes with varying concentrations of dA:dT competitor DNA. Pure MBP-MBD2 was run in lane C as control. **(i)** Appearance of beads with increasing amounts of dA:dT after incubation of VEGFA probe with 1 nM MBP-MBD2 and 3.6 μm nuclear extracts. **(j)** Western blot showing the presence of MBP-MBD2 after pull-down with differentially modified VEGFA probes using the established conditions. Pure MBP-MBD2 was run in lane C as control. For Western blots the samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M), expected molecular weights: MBD1: 67 kDa, MBD2: 43 kDa, MBP-MBD2: 53 kDa)

3.3.2 Performance of different DNA probes in the pull-down assay

The pull-down must be sufficiently sensitive to reveal differences in enriched proteins across various modifications. This sensitivity depends on both, the DNA probe and the protein pool. Since nuclear extracts are a standard in enrichment studies with DNA (L. Bai et al., 2021; Iurlaro et al., 2013; Spruijt et al., 2013; J. Xiong et al., 2016), the main source of uncertainty lies in the DNA probe. Therefore, the different probe types introduced in **Subsection 3.1.1** were evaluated in Western blots to determine their ability to display significant differences within the pull-down assay.

Promoter context probes. As shown above in **Figure 3.10j** the VEGFA probe demonstrated the expected differences in selectivity of the established reader protein MBD2 in the pull-down assay. In addition, also the other promoter context probes revealed specific binding of MBP-MBD2 proving the workflow successful (**Figure 3.11a**). However, as discussed in **Subsection 3.1.2**, the VEGFA probe was chosen as the promoter context probe for following experiments due to easy handling and high PCR and primer extension yields. 500 ng of the VEGFA probe were used in the pull-down, corresponding to an amount of 4 pmol.

Randomized core CG probes. The randomized core CG probe 8NX also showed the expected differences in the pull-down, though the primer extended 8NX probe (PE) performed much better than the hybridized 8NX probe (H) (**Figure 3.11b**). This reaffirmed that hybridization of the probe might not be a suitable alternative, reinforcing the initial assumptions from **Subsection 3.1.2** and indicating that MBD2 binding might be impeded by hybridization mismatches. 500 ng of each probe was immobilized in accordance with the optimal VEGFA probe mass, explaining the stronger difference in enrichment between the VEGFA and 8NX probe, which arises from 8NX having a four times higher concentration as a result of its length (**Figure 3.11c**). Hence, next the same molar concentration of the 8NX and VEGFA probes was used. 4 pmol of the 8NX probe resulted in a weak signal indicating an insufficient number of binding sites for MBD2 (**Figure 3.11d**). Since the 8NX probe has only one CpG, while the VEGFA probe has 11 CpGs per molecule, $2.66*10^{13}$ more CpGs are present in the VEGFA probe when using 4 pmol of each probe ((**Figure 3.11c**)). Increasing the 8NX probe concentration on the other hand to adjust the CpG amounts eliminated modification preference differences presumably due to an excess of DNA that led to unspecific MBD2 binding (**Figure 3.11d**). Assuming that one CpG per molecule is insufficient, the amount of CpGs per probe was increased yielding the 12 and 16NX probes (compare **Subsection 3.1.1**). The pull-down was conducted with DNA probe concentrations adjusted to ensure that the number of total CpGs was the same. This was achieved by using 48 pmol 8NX, 24 pmol 12NX and 16 pmol 16NX (**Figure 3.11c, e**). With the 8NX probe again no difference between unmethylated and methylated DNA was observed given the high probe concentration to make up for the number of CpGs which presumably leads to saturation. The 12NX probe showed slight preference of MBD2 for methylated over unmethylated DNA, while this effect was even more pronounced with the 16NX probe. Thus, a trend was observed that increasing CpG sites render the binding more selective and enhance differences as unspecific binding is minimized. In total the probe concentration of randomized core CG probes is still higher than the VEGFA probe concentration which explains the stronger signal. To make the binding more specific nuclear extracts were added which helped to achieve stronger mC selectivity (**Figure 3.11f**). This confirmed that while higher probe concentration leads to unspecific MBD2 binding, specificity can be increased by addition of background proteins. The general limitation of NX pull-downs is the limited number of CpGs which was circumvented by higher probe concentration which may lead to the recruitment of more unspecific binders.



c

probe	seq. contexts (#)	CpGs (#)	molarity (pmol)	seq. coverage	total CpGs
8NX	65536	1	4	37214813	$2.4*10^{12}$
			24	223288879	$1.5*10^{13}$
			48	446577758	$2.9*10^{13}$
			96	893155518	$5.9*10^{13}$
12NX	16777216	2	24	872222	$2.9*10^{13}$
16NX	4294967296	3	4	568	$7.3*10^{12}$
			16	2271	$2.9*10^{13}$
VEGFA	1	11	4	NA	$2.9*10^{13}$

Figure 3.11: Evaluating different DNA probes in the pull-down assay. (a) Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) MUC1, LINE1 or VEGFA probes. Pure MBP-MBD2 was run in lane C as control. **(b)** Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) VEGFA or 8NX probes. 8NX was generated either by primer extension (PE) or hybridization (H). Pure MBP-MBD2 was run in lane C as control. **(c)** Overview of evaluated probes, their possible sequence contexts and number of CpGs. In gray different pull-down conditions are shown with the sequence coverages and total number of CpGs corresponding to the input amounts in pmol. **(d)** Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) 8NX probes. Amount of 8NX was titrated from 4 pmol up to 96 pmol. **(e)** Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) 8NX, 12 NX or 16NX probes. Total CpG amount was kept constant by using 48 pmol of 8NX, 24 pmol of 12NX and 16 pmol of 16NX probe. **(f)** Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) 8NX, 12NX and 16NX probes. The probe concentrations were kept as in **(e)** and 50 μg nuclear extracts were added (NE). Pure MBP-MBD2 and nuclear extracts were run as controls in lane C and NE. For Western blots the samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M).

3.3.3 Validation of the pull-down assay by MS

After determining the pull-down conditions and validating them with Western blots showing that the probes could selectively enrich MBD2 using this workflow, modification-specific MS studies were conducted to identify natural binders which aided to constitute a proteome-wide validation as opposed to solely relying on the MBD2 benchmark. On that account, the functionality of this setup was determined by comparing the proteomes identified for unmethylated (C/C) and methylated (mC/mC) probes. The intended DNA probe was incubated with nuclear extracts from HEK293T cells as shown in **Figure 3.10a**. A benchmark pull-down was always conducted with MBP-MBD2, whereas MBP-MBD2 was omitted in pull-downs designated for proteome analyses. Bound proteins were on-bead digested to peptides and prepared for MS. First, the VEGFA and 8NX probes were compared. The amount of CpGs was kept constant by using a higher probe concentration of 8NX according to **Figure 3.11c**. The benchmark pull-down confirmed the previous results (**Figure 3.12a**) and also indicated a higher MBD2 modification specificity for the VEGFA probe compared to 8NX where slight binding to the unmethylated probe was observed. The proteomics analysis revealed around 1080 proteins that bound to the 8NX probe while the VEGFA probe enriched on average 580 proteins from the nuclear extract pool (**Figure 3.12b**). When comparing proteins enriched with C/C and mC/mC it became evident that the 8NX probe, despite having more total binders, enriched fewer modification-specific binders than the VEGFA probe (**Figure 3.12c**). This might be attributable to the presence of only one CpG site per molecule and increased amounts of unspecific DNA binders due to a higher probe concentration, presenting further evidence that one CpG might not be sufficient in accordance with **Subsection 3.3.2**. Volcano plots were generated to compare the differential enrichment of proteins with C/C and mC/mC. While the 8NX probe enriched eight proteins significantly for C/C and only one protein for mC/mC (p-value < 0.05 and fold change (FC) > 1), the VEGFA probe enriched 67 proteins for C/C and 52 proteins for mC/mC (**Figure 3.12c - e**).

Proteins commonly enriched for C/C with both probes were the known unmethylated DNA binders SAM domain-containing protein 1 (SAMD1) and its interaction partners Lethal(3)malignant brain tumor-like protein 3 (L3MBTL3) (Stielow et al., 2021) as well as BCL-6 corepressor (BCOR) and other components of the PRC1.1 like Polycomb Group Ring Finger 1 (PCGF1) (Sugishita et al., 2021) that were also detected in previous pull-down MS studies (Spruijt et al., 2013). PRC1.1 complexes maintain transcriptional repression by recognizing unmethylated CGIs through Lysine Demethylase 2Bs (KDM2Bs) CXXC domain and subsequent recruitment of the other components including the core elements (PCGF1, Ring Finger Protein 1 (RING1) and S-Phase Kinase Associated Protein 1 (SKP1)) (Farcas et al., 2012; Wong et al., 2016). Ubiquitin-like with PHD and Ring Finger Domains 1 (UHRF1) was the only protein significantly enriched for mC/mC with both examined probes, presenting a known mC/mC binder which was also reported in previous pull-down studies (Iurlaro et al., 2013) (**Figure 3.12d, e**). Further known mC binders such as MBD2, methyl-CpG binding domain protein 4 (MBD4) and Regulatory Factor X5 (RFX5) (Du et al., 2015; Hendrich and Bird, 1998; Spruijt et al., 2013) were solely enriched with the VEGFA probe although the canonical RFX5 X-box binding motif was not present in the VEGFA probe (Steimle et al., 1995). The enriched proteins are in good agreement to findings of previous pull-down studies (L. Bai et al., 2021; Iurlaro et al., 2013; Spruijt et al., 2013). Interestingly, RFX5 did not bind to the 8NX probe at all, while MECP2, MBD1 and MBD4 did bind, however, a single CpG site appeared to be insufficient to show differential binding, or the high probe concentration may have led to unspecific binding, concealing the proteins' preferences. Some proteins bound exclusively to the 8NX and not to the VEGFA probe. Among them, CGG Triplet Repeat Binding Protein 1 (CGGBP1), a protein that binds CGG triplet repeats in RNA (Müller-Hartmann et al., 2000), which is an interactor of the chromatin regulator protein CCCTC-binding factor (CTCF) (Ong and Corces, 2014) and regulator of

genome-wide cytosine methylation in the CpG context (Agarwal et al., 2015). Remarkably, CTCF was not enriched for 8NC/C but for the VEGFA C/C probe, while VEGFA, in turn, did not enrich CGGBP1. Further, E2F Transcription Factor 6 (E2F6) and KDM2B were only enriched with 8NC/C, which are components of the polycomb repressive complex PRC1.6 (Ogawa et al., 2002). Also other components of PRC1.6 were enriched significantly with the 8N probe, however with a very small FC of around 0.5 (**Figure 3.12e**). This is in line with ChIP-seq analyses showing colocalization of the PRC1.6 components MAX gene-associated protein (MGA), Lethal(3)malignant brain tumor-like protein 2 (L3MBTL2), E2F6 and Polycomb Group Ring Finger 6 (PCGF6) (Stielow et al., 2018). PRC1.6 is sequence-based recruited due to its sequence-specific members MGA/Myc-associated factor X (MAX) and E2F6 (Stielow et al., 2018). E2F6 recognizes the motif TCCCGC (Gaubatz et al., 1998), while MGA/MAX bind to the E-box motif CANNTG (Hurlin, 1999). This sequence-specificity could explain why no enrichment of PRC1.6 members with VEGFA was observed in contrast to PRC1.1 enrichment that was observed with both probes (see above and **Figure 3.12d, e**), where KDM2B binds to unmethylated CGIs without sequence preference via its ZF-CXXC domain (Wong et al., 2016). Members of both, PRC1.1 and PRC1.6 were identified also in other pull-down studies (Spruijt et al., 2013), further indicating that PRC1.6 could be enriched in a sequence-specific manner where proteins were either rejected from the VEGFA probe due to absence of the binding motif or outcompeted by VEGFA specific binders.

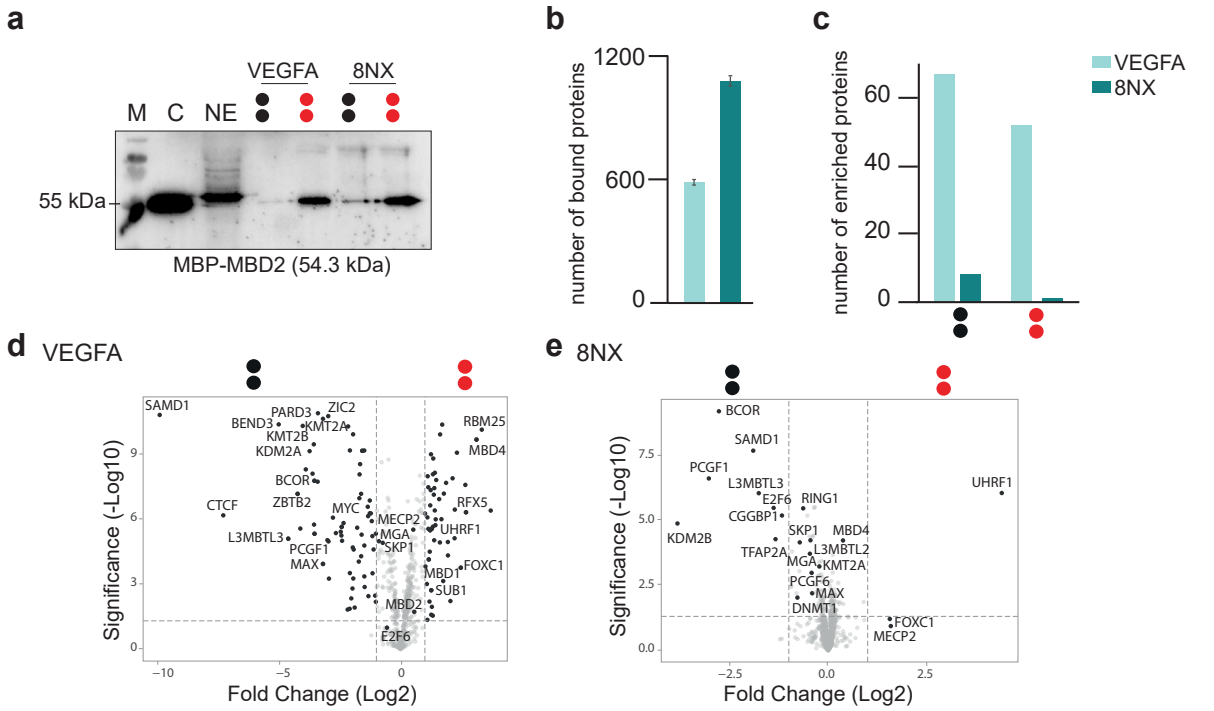


Figure 3.12: Comparison of VEGFA and 8NX probes by MS. (a) Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) VEGFA and 8NX probes. Samples were run on a 12 % SDS-PAGE with a Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder (lane M), a MBP-MBD2 control (lane C) and nuclear extract control (lane NE). **(b)** Bar diagram depicting the average total number of proteins detected after the pull-down by MS from ten pull-downs each. **(c)** Bar diagram depicting the number of significantly enriched proteins with unmodified and methylated VEGFA and 8NX probes. **(d)** Volcano plot comparing VEGFA C/C (black circles) and mC/mC (red circles) enriched proteins (p-value < 0.05 and log2 fold change >1). Corresponding protein names are listed in Table A.1. **(e)** Volcano plot comparing 8NX C/C (black circles) and mC/mC (red circles) enriched proteins (p-value < 0.05 and log2 fold change >1). Corresponding protein names are listed in Table A.2.

To further evaluate the influence of the number of CpG sites and the probe concentration, pull-down assays with the 16NX probe were conducted using two different probe concentrations. The first

concentration corresponded to the amount of VEGFA probe (4 pmol), while the second concentration was adjusted to match the number of CpG sites present with 4 pmol of the VEGFA probe. This led to a higher probe concentration, using 16 pmol of 16NX (compare **Figure 3.11c** VEGFA 4 pmol vs 16NX 4 and 16 pmol). The MBP-MBD2 Western blot revealed that a higher probe concentration resulted in more MBD2 being pulled down but also increased unspecific binding, as evidenced from background smear (**Figure 3.13a**). However, the binding appeared to be highly specific to mC and, although run on different gels, seemed improved compared to 8NX (also see **Figure 3.11f**). Comparison of the total number of identified proteins shows that a higher probe concentration is associated with an increased number of proteins as on average 404 proteins were enriched with 4 pmol and 682 with 16 pmol 16NX probe. (**Figure 3.13b**). This corresponds to the observations from the pull-downs with the 8NX and VEGFA probes (**Figure 3.12b**). Assumingly, a higher probe concentration facilitates unspecific DNA binding, which also explains the lower detection of proteins with the 16NX probe compared to the 8NX probe. An additional factor could be the lower coverage of the different sequence contexts in the 16NX probe due to the increased number of randomized nucleotides. Analysis of modification-specific enrichment revealed that the number of binders was comparable, with 19 and 14 proteins enriched for C/C over mC/mC and one and two proteins enriched for mC/mC over C/C for 4 and 16 pmol of 16NX, respectively (**Figure 3.13c**). From Volcano plots it became apparent that many C/C binders were consistent with the aforementioned binders of the VEGFA and 8NX probe. Among the mC/mC binders, MeCP2 and MBD4 were detected, however, they were not consistently enriched. While the pull-down with 4 pmol 16NX probe showed significant enrichment of MeCP2, the 16 pmol 16NX dataset only indicated a trend towards enrichment and vice versa for MBD4 (**Figure 3.13d, e**). Interestingly, the mRNA binder UAP56-interacting factor (UIF) (Forty-two-three Domain Containing 1 (FYTTD1) gene) was significantly enriched for mC/mC over C/C with 16 pmol of 16NX (**Figure 3.13e**). While it was also detected in the 8NX pull-down and in other studies (L. Bai et al., 2021; Spruijt et al., 2013), no significant enrichment was reported before. The protein was also not identified with 4 pmol of 16NX. UIF function is connected to mRNA export (Hautbergue et al., 2009), but reports on UIF interaction with CpG methylation are scarce and confined to a study on non-small cell lung cancer (NSCLC) in relation to smoking (L. Chen et al., 2025). In that study, UIF significantly bound to a methylated CpG probe and was proposed to directly regulate gene expression and therefore influence mRNA transcription (L. Chen et al., 2025).

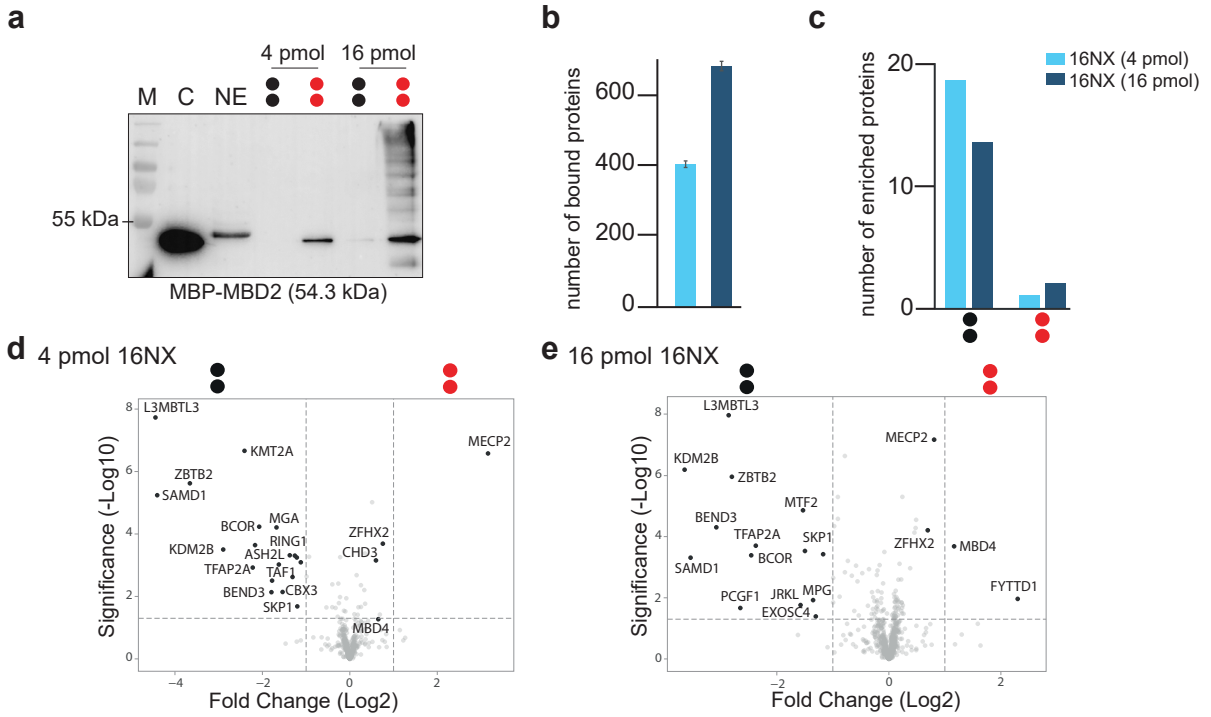


Figure 3.13: Comparison of different 16NX probe concentrations by MS. (a) Western blot showing binding of 1 nM MBP-MBD2 to unmodified (black circles) or methylated (red circles) 4 pmol or 16 pmol 16NX probes. Samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M), a MBP-MBD2 control (lane C) and nuclear extract control (lane NE). **(b)** Bar diagram depicting the average total number of proteins detected after the pull-down by MS from ten pull-downs each. **(c)** Bar diagram depicting the number of significantly enriched proteins with unmodified and methylated 16NX probes. **(d)** Volcano plot comparing 4 pmol 16NX C/C (black circles) and mC/mC (red circles) enriched proteins (p-value < 0.05 and log2 fold change >1). Corresponding protein names are listed in Table A.3. **(e)** Volcano plot comparing 16 pmol 16NX C/C (black circles) and mC/mC (red circles) enriched proteins (p-value < 0.05 and log2 fold change >1). Corresponding protein names are listed in Table A.4.

Significantly enriched proteins for unmodified or methylated DNA were compared in Venn diagrams. However, this comparison should be evaluated with caution as the pull-downs with the VEGFA and 8NX probes were conducted on different days than those with the 16NX probes and different batches of nuclear extracts were used. VEGFA enriched the highest number of specific binders although either the amount of CpG sites was kept constant with the 8NX and 16NX (16 pmol) probes or the probe concentration with 16NX (4 pmol). Only the four proteins SAMD1, L3MBTL3, Transcription factor AP-2-alpha (TFAP2A) and BCOR were enriched across all four sets of pull-downs for C/C (**Figure 3.14a** and **Table A.5**). Of the 67 proteins enriched with the VEGFA C/C probe, 53 were exclusively enriched by this probe. The amount of mC-specific binders was low for the randomized core CG probes and mainly limited to non-sequence-specific readers such as MBD proteins with either MECP2, MBD4 and UHRF1 (**Figure 3.14b** and **Table A.6**). mC binders were rather inconsistently identified across the different probes, showing minimal to no overlap resulting from the low identification depth which led to higher variability.

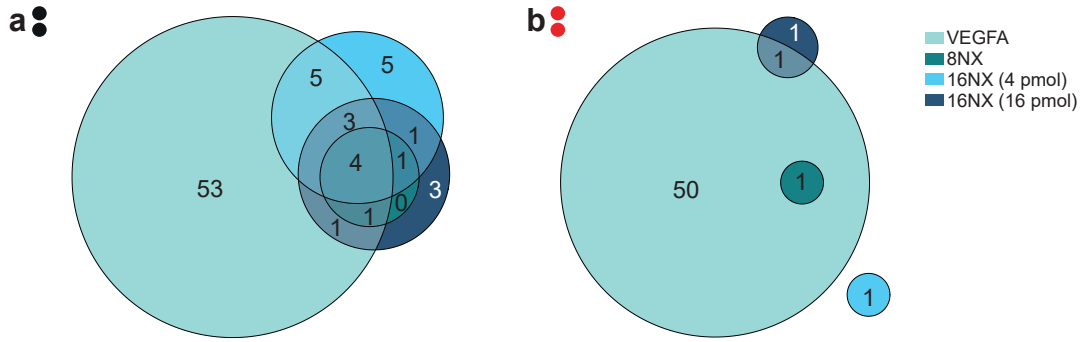


Figure 3.14: Overlap of significantly enriched proteins for unmodified and methylated DNA probes. (a) Venn diagram showing overlap and number of significantly enriched proteins for C/C with the VEGFA, 8NC and 16NC probes. **(b)** Venn diagram showing overlap and number of significantly enriched proteins for mC/mC with the VEGFA, 8NmC and 16NmC probes.

The low number of specific mC binders for the randomized core CG probes could also stem from the exclusion of sequence-specific binders due to the randomized regions which were underrepresented resulting in a low concentration of the individual sequences preventing effective enrichment. While the high amount of individual sequences led to increased binding opportunities resulting in a increased total number of detected proteins the modification specificity was lost attributed to a loss of weak or transient interactions. This led to a higher variability especially for proteins near the detection threshold. While the promoter context probes contain specific, high-affinity binding sites, randomized core CG probes lack such consistent sites, reducing strong and specific interactions. With more diverse sequences, non-specific binding increases leading to higher variability. Some proteins might have a broad DNA-binding affinity but require a specific sequence for strong binding. Hence, the subsequent reader identification experiments were conducted with the VEGFA probe.

3.3.4 Essential points

- A pull-down protocol was established allowing for the enrichment of a MBP-MBD2 benchmark protein with methylated DNA probes. MBP-MBD2 exhibits the expected binding behavior of probes with strand-specific asymmetric CpG modifications generated by the protocol introduced in Subsection 3.1.2.
- Further validation of the functionality of the pull-down assay was achieved by MS detection of known C and mC readers with different DNA probe types.
- Promoter context probes and randomized core CG probes were compared for their ability to significantly enrich nuclear proteins with the pull-down assay. A randomized sequence context generally led to more protein binding overall, but with reduced specificity.
- This ultimately resulted in the choice to use the VEGFA promoter probe for subsequent reader identification studies.

3.4 Proteomics studies for the discovery of potential readers

After successfully validating the pull-down workflow through the identification of known reader proteins for C and mC, and confirming the effectiveness of the MBP-MBD2 benchmark, the focus will now be shifted to novel reader protein discoveries. This section explores the identification of potential reader proteins for hmC-containing CpG dyads using MS. I will present comparative proteomics analyses employing human and mouse nuclear extracts alongside DNA promoter probes, which have been modified with C, mC or hmC in either a strand-symmetric or asymmetric way. This allows for a direct assessment of their differential interactomes.

The MS results will be presented in a structured manner, beginning with an evaluation of the functionality of the employed probes, as confirmed by benchmarking pull-downs being conducted in parallel to pull-downs designated for MS analyses. Once their reliability is proven, the MS data will be assessed qualitatively to evaluate reproducibility. A general comparison of enrichment patterns between probes will follow before moving on to one-on-one comparisons of enriched proteins with different probe modifications and in-depth discussion of specific proteins. The experimental setup consists of two biological replicates (referred to as biological experiments from here on) of HEK293T nuclear protein pull-downs, each conducted with five differentially modified VEGFA probes (C/C, hmC/C, mC/mC, hmC/mC, hmC/hmC), with five technical replicates per modified probe. A third biological experiment was conducted using only three differentially modified VEGFA probes (C/C, mC/mC and hmC/mC), each with five technical replicates. Additionally, one mouse brain nuclear protein pull-down experiment was performed, consisting of five differentially modified Sp1 probes (C/C, hmC/C, mC/mC, hmC/mC, hmC/hmC), again with five technical replicates per modification.

Proteins that were captured in these pull-downs were identified by analyzing their peptides by high-resolution nanoHPLC-tandem mass spectrometry (nanoHPLC-MS/MS) on an Orbitrap platform. This enabled proteome-wide identification by first separating the peptide mixtures and then measuring the mass-to-charge ratios of ionized fragments, with subsequently matching them to sequence databases. LFQ was used to compare proteins captured with differentially modified probes, which allowed for statistical discrimination of specific interactors from background, e.g. general DNA-binders, ultimately enabling the generation of candidate lists for further analysis.

3.4.1 Discovery of potential readers from HEK293T nuclear extracts

To study interactions of hmC-containing CpG dyads with human proteins, the proteins were extracted from HEK293T cells. HEK293T cells originate from HEK293 cells, which were derived from a human embryonic kidney (Russell et al., 1977). Due to this human-derived background, their proteins and metabolic pathways closely resemble those of human cells, making them especially useful for studying human proteins. Their human origin further means that their protein expression, PTMs and folding are similar to human physiology. HEK293T cells were derived from HEK293 cells by transfection with a plasmid encoding the SV40 large T-antigen, which is constitutively expressed (DuBridge et al., 1987). HEK293T cells are a convenient choice for large scale proteomics studies due to their rapid growth, ease of culture under standard conditions and high reproducibility. As they are widely used in research, a vast collection of protocols and data exists. However, one drawback is their cancer-like metabolism. As a transformed cell line, HEK293T cells may exhibit altered protein expression and metabolic pathways compared to normal cells.

HEK293T protein pull-down set-up

A total of three biological HEK293T protein pull-down experiments were conducted using five differentially modified VEGFA probes, each with five technical replicates. The nuclear extracts used are presented in **Section 3.2, Figure 3.8b, c, and Table 3.2**. In this figure, biological experiment 1 corresponds to replicate 3, experiment 2 to replicate 2, and experiment 3 to replicate 1. Symmetrically and asymmetrically modified DNA probes were prepared using the asymmetric probe generation protocol from **Section 3.1, Figure 3.5** and run on agarose gels for validation. The corresponding gels for biological experiments 1, 2 and 3 are depicted in **Figure 3.15a-c**. For biological experiment 1, the different samples collected throughout the workflow are displayed for C/C, while all other probes are shown only in their primer extended final state (**Figure 3.15a**, number 4). For biological experiments 2 and 3, all steps of the workflow are shown for each probe modification (**Figure 3.15b, c**). As demonstrated in **Section 3.3**, alongside the pull-downs conducted for MS, an additional pull-down was performed using the MBP-MBD2 benchmark to validate probe functionality and the general pull-down workflow. The Western blots corresponding to experiment 1, 2 and 3 are shown in **Figure 3.15d-f**. In each experiment, the VEGFA mC/mC probe enriched the highest amount of MBP-MBD2, followed by hmC/mC. Slight enrichment was also observed for hmC/hmC while C/C and hmC/C did not show enrichment detectable by Western blotting (with an exception for experiment 2 (**Figure 3.15e**) where weak bands were visible for C/C and hmC/C in the same intensity range as for hmC/hmC). However, the general selectivity of MBD2 was confirmed each time, making these modified VEGFA probes suitable for large scale proteomics experiments.

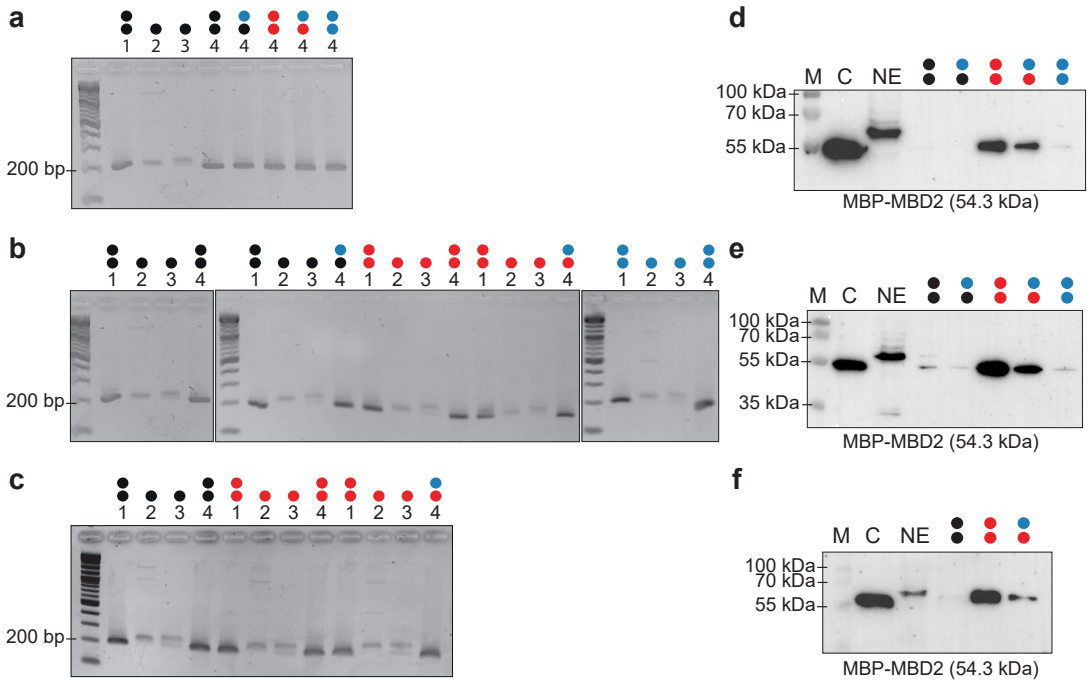


Figure 3.15: VEGFA probes used in HEK293T nuclear protein pull-downs. (a-c) Agarose gel analyses of all employed VEGFA probes used for proteomics pull-downs. Samples were collected during the process of generating (a)symmetric probes for experiments 1-3 and run on a 2 % agarose gel with a 1 kb Plus DNA ladder (New England Biolabs). The numbering corresponds to **Section 3.1 Figure 3.5** representing 1: the PCR product, 2: ssDNA, 3: primer hybridized product and 4: (asymmetric) primer extended product. Circles are colored according to the C modifications (C – black, mC – red and hmC – blue) depicting ssDNA (one circle) or dsDNA (two circles). **(d-f)** Western blots of experiments 1-3 show the presence of MBP-MBD2 after pull-down with differentially modified VEGFA probes in parallel to pull-downs conducted for MS. Pure MBP-MBD2 was run in lane C as control and nuclear extracts in lane NE. Samples were run on a 12 % SDS-PAGE with a Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder (lane M), expected molecular weight of MBP-MBD2 is 53 kDa.

Evaluation of sample and data quality

To assess the quality of the MS data, different quality controls were examined. First, the number of proteins was checked. Ideally, the number of detected proteins should be similar for different conditions in one biological experiment as only fine differences are expected for the different probe modifications but the majority of bound proteins are expected to be non-modification specific DNA binders. The number of quantified proteins was averaged over all 25 pull-downs for experiment 1 and 2 and over all 15 pull-downs for experiment 3 and is depicted with the corresponding standard deviation in **Figure 3.16a**. The variation within each set of experiments is extremely low, as indicated by the error bars. However, it is striking that significantly fewer proteins were quantified in experiment 1. While only 501 proteins were quantified in experiment 1 per pull-down, 966 and 881 proteins were detected in experiment 2 and 3 per pull-down, respectively. One possible explanation is that the cytoplasmic protein contamination in the NE batch used in this experiment was higher, reducing the apparent nuclear protein concentration and leading to fewer bound proteins overall. However, when comparing the nuclear extracts, the cytoplasmic contamination as identified by β-Tubulin was relatively low (**Figure 3.8b, 3**) and the nuclear protein concentration appeared consistent with the batches used in the other two experiments (**Figure 3.8c**). Importantly, since significant proteins are determined separately for each biological experiment, the variation in total protein numbers does not pose a problem as long as within each experiment the values remain within a comparable range, as confirmed by the low standard deviation.

Next, the log2-transformed LFQ-intensities were examined and displayed in a multi-scatter plot to estimate the degree of correlation between the individual samples (**Figure 3.16b**). If the data is consistent, similar intensity distributions between replicates and conditions are expected. In this context, a "condition" refers to a single probe modification. Each of the five probe modifications were analyzed with five technical replicates. The technical replicates within one condition showed strong positive correlation (tight clustering of points along the diagonal line) in all three experiments, indicating that the measurements were reproducible across technical replicates, suggesting good consistency in the pull-down MS process. The correlation of the technical replicates within one condition was higher than between the conditions indicating an effect of the different dyad modifications. Still, the conditions showed a relatively high correlation, suggesting that the protein binding behaviors were relatively similar across the modifications which was expected as many non-modification specific DNA binders were enriched. Ideally, the correlation within one condition corresponding to the technical replicates of one probe modification would be extremely high and also between groups correlation should be good depending on how strong they enrich different proteins.

Lastly, the overall variance of the dataset was visualized using a PCA plot (**Figure 3.16c**), highlighting patterns such as sample clustering and potential outliers, which were not immediately apparent in multi-scatter plots. This provided insights into whether technical replicates from different conditions clustered together as expected. Since PCA groups data points based on overall similarity, it served as a useful tool for assessing reproducibility, whereas dispersed replicates suggested technical inconsistencies. In experiment 1, outliers were observed, particularly for mC/mC replicate 1, hmC/mC replicate 1, and hmC/hmC replicate 5, while in experiment 2, C/C replicate 4 was identified as an outlier, suggesting possible technical errors or sample preparation issues. The PCA plot also provided insights into global differences between conditions. If conditions formed distinct clusters, this indicated strong biological differences in protein binding patterns, as seen in experiments 2 and 3. However, if conditions overlapped significantly, it suggested minimal differences in protein composition, which was particularly evident in experiment 1 for mC/mC (replicates 2–5), hmC/mC (replicates 2–5), and some hmC/hmC replicates. In experiments 1 and 2, PC1 (x-axis) and PC2 (y-axis) together explained less than 50 % of the total variance, indicating complex variation without a single dominant trend. When

PC1 and PC2 fail to capture the majority of variance, exploring higher principal components or alternative clustering methods becomes necessary. Because PCA did not reveal clear separation in experiment 1, hierarchical clustering was applied using a heatmap (**Figure 3.16d**). This analysis showed that replicates within each condition clustered well together, with only a few outliers that corresponded to those identified in the PCA plot. Interestingly, although mC/mC and hmC/mC replicates showed overlap in the PCA plot, the heatmap revealed that they were in fact distinct from one another. Heatmap analyses of experiment 2 and 3 are shown in **Figure A.7**.

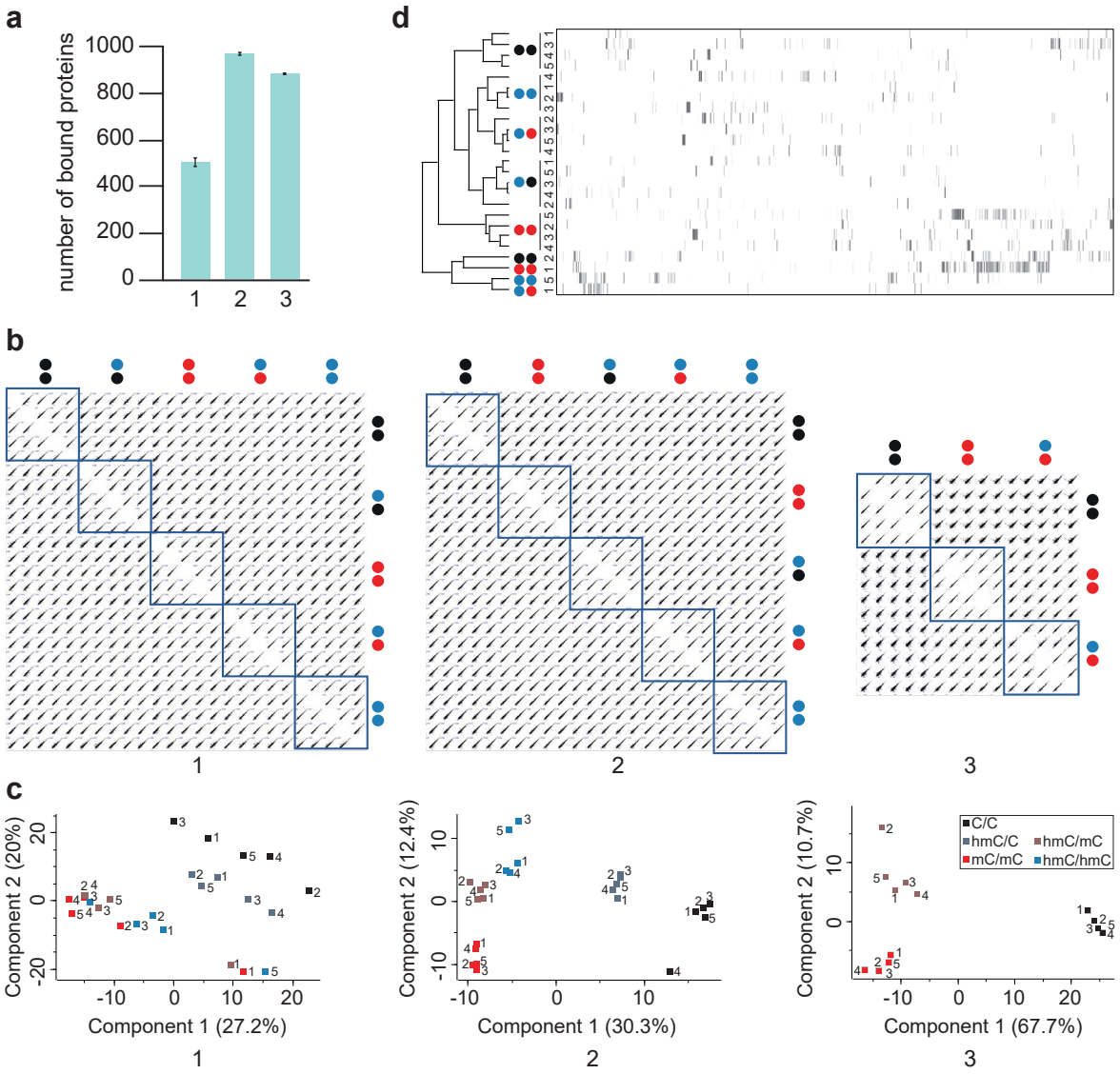


Figure 3.16: Data correlation and quality check. (a) Bar diagram representing the number of quantified proteins in biological experiments 1-3 as an average of all technical replicates. Standard deviation is shown as error-bars. **(b)** Multi scatter plots of LFQ intensities of biological experiments 1-3 showing the correlation of all technical replicates to each other sorted within the modification conditions. Correlations of technical replicates within one group are shown in blue squares. Modifications are depicted as circles, colors correspond to **Figure 3.15**. **(c)** PCA plots of LFQ intensities of biological experiments 1-3 showing variance across samples. Technical replicates are numbered accordingly and shown in one color per modification as indicated in the legend. **(d)** Hierarchical clustering of samples from experiment 1 visualized in a heatmap. Modifications are depicted as circles, colours correspond to **Figure 3.15** and technical replicates are numbered from 1-5.

Identification of potential readers from HEK293T nuclear extracts

To identify significant binders for different modifications, log2 LFQ intensities were grouped by probe modification and a t-test was performed with an FDR of 0.025 and a FC of 1.5. A protein qualified as a reader if it was significantly enriched in at least four out of five technical replicates in either experiment 1 or 2. While results from experiment 3 are also discussed, they are not fully comparable due to the limited number of modified probes examined. Therefore, experiment 3 was excluded from reader identification. The number of readers for every dyad modification was assessed by comparing all modified probes against one another and counting significantly enriched proteins for each condition (**Figure 3.17a**). A reader protein had to be enriched for at least one modified probe over another. This approach led to overlapping readers, which are discussed in detail in the following sections. The highest number of readers was identified for C/C, with 66 in total across experiments 1 and 2, and 72 in experiment 3. This was followed by hmC/C with 47 readers and mC/mC with 44 readers (33 in experiment 3). hmC/mC had fewer binders, with 22 in total (31 in experiment 3), while the lowest number was observed for hmC/hmC, with 20 binders, consistent with previous reports stating low amounts of hmC readers (Iurlaro et al., 2013; Spruijt et al., 2013). However, probes containing a single hmC showed increased reader numbers, highlighting the influence of the second C modification on reader recruitment (**Figure 3.17a**).

To explore differences in protein binding across various modifications, differentially enriched proteins between modified dyads were visualized using volcano plots. A volcano plot helps to distinguish proteins that show significant changes in abundance between conditions, thereby revealing potential biological differences. By simultaneously displaying the log2 fold change and -log10 p-value, these plots allow for identification of proteins that are both significantly and substantially different, while filtering out non-specific binders that show non-significant changes between the conditions. As a final quality control step, volcano plots were used to assess the differential enrichments of C/C and mC/mC in all three biological experiments, providing a visual summary of reproducibility and specificity in reader identification (**Figure 3.17b-d**).

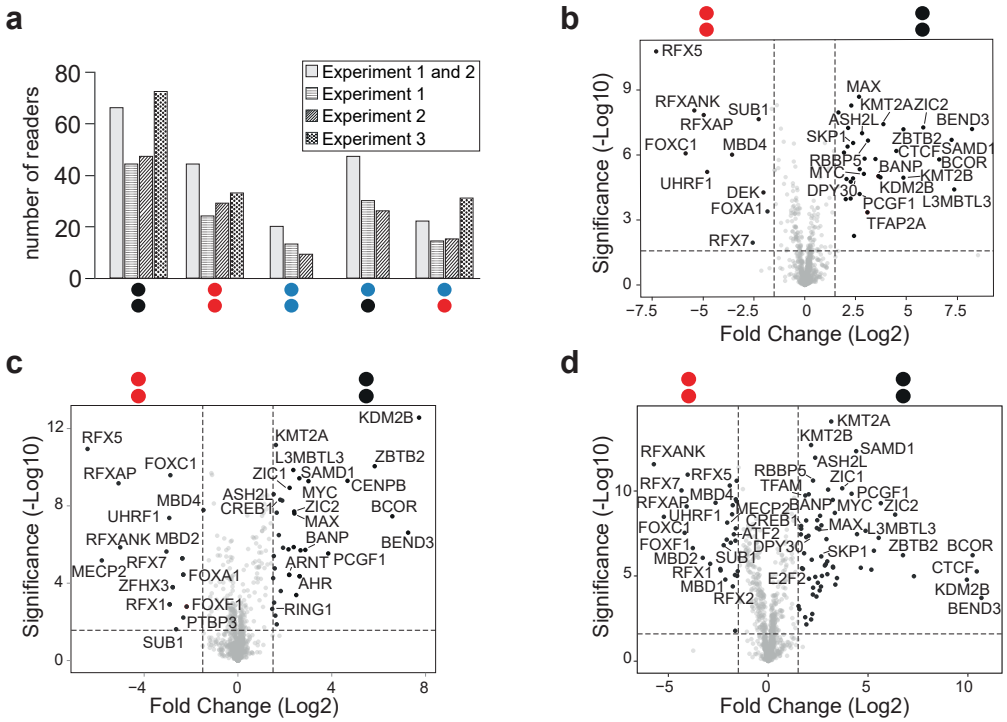


Figure 3.17: General comparison of HEK293T enrichments. **(a)** Bar diagram showing the distribution of reader numbers for the different modified dyads. All three biological experiments are shown with an additional bar summarizing experiment 1 and 2. **(b-d)** Volcano plots comparing C/C (black circles) and mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) of **(b)** experiment 1, **(c)** experiment 2 and **(d)** experiment 3. See Table A.7, Table A.8, Table A.9 for full protein lists.

Many proteins enriched for C/C and depleted by mC/mC are overlapping with results discussed in **Subsection 3.3.3**. The majority of C/C preferred proteins can be grouped into PRC and mixed lineage leukemia (MLL)-complexes. PRCs were already discussed in **Subsection 3.3.3**. Proteins of PRC1.1 such as PCGF1, BCOR andKDM2B were significantly enriched for C/C in all three biological experiments. Further PRC1.1 components like SKP1 were enriched only in experiment 1 and 3, while RING1 was only enriched in experiment 2. From PRC1.6, only MAX was significantly enriched in all three experiments. MAX was not only associated with the PRC1.6 complex but was also accompanied by consistent detection of its dimerization partner MYC. MAX was shown to play an essential role in MYC-dependent transcriptional regulation (D. Y. Mao et al., 2003). Proteins from the MLL1 complex, such as Histone-lysine N-methyltransferase 2A (KMT2A) and Absent, small, homeotic disks-2-like (ASH2L), were consistently enriched across all three experiments. In contrast, some other components of this complex like DumPY protein 30 (DPY30) and Retinoblastoma-binding protein 5 (RbBP5), were detected only in experiments 1 and 3. The MLL1 complex (also known as KMT2A complex) is a multi-protein histone methyltransferase that catalyzes methylation of lysine (Lys)4 of histone H3 (H3K4me). KMT2A is the catalytic subunit and contains a Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) domain, which directly mediates H3K4 methylation (Rea et al., 2000). It also comprises a CXXC domain that enables direct binding to unmethylated CpGs. The interacting proteins are essential for activating and stabilizing its enzymatic function (Rao and Dou, 2015). In agreement with the results from **Subsection 3.3.3**, SAMD1 and its interaction partner L3MBTL3 were consistently identified in all three experiments (Stielow et al., 2021). Moreover, Btg3-associated nuclear protein (BANP), a methylation-sensitive transcription factor connected to chromatin opening, was identified in all three experiments. It has been reported to bind CGIs, specifically recognizing the CGCG motif (Grand et al., 2021), however, this motif was not present

in the VEGFA probe. Next to BANP, another BEN domain-containing protein, BEND3, was enriched for C/C across all experiments. BEND3 has been reported to bind to its unmethylated consensus site CCACGC and regulate hundreds of bivalent promoters (Yakhou et al., 2023). The VEGFA probe lacks this consensus site. Lastly, also Zinc finger and BTB domain-containing protein 2 (ZBTB2) was consistently enriched. It was already proposed as a methylation-sensitive transcription factor that regulates cellular differentiation (Karemaker and Vermeulen, 2018).

On the other side, proteins that preferentially bound to mC/mC, comprise SUB1, also known as positive cofactor 4 (PC4). Activating and repressing functions have been ascribed to PC4 and its chromatin function is regulated by its phosphorylation (Mustafi et al., 2022). Other pull-down MS studies had also identified this protein to be enriched for mC/mC or fC/fC (L. Bai et al., 2021; Iurlaro et al., 2013) but it was never further mentioned for its preference of methylated DNA binding. Further, the RFX-family proteins RFX5, RFX7, Regulatory Factor X Associated Ankyrin Containing Protein (RFXANK), and Regulatory Factor X Associated Protein (RFXAP) were consistently enriched for mC/mC across all three experiments, and RFX1 was enriched in experiments 2 and 3. RFX5, RFXANK and RFXAP compose a subcomplex that binds the X1 box of the SXY module found in the promoters of genes encoding MHC class II molecules (Steimle et al., 1995). All three proteins were identified as mC interactors (Bartke et al., 2010; Spruijt et al., 2013). Interestingly, the X2 box of the SXY module is bound by cAMP-responsive element-binding protein 1 (CREB1) (Moreno et al., 1999), which was significantly depleted for mC/mC in experiments 2 and 3. In contrast to RFX5 and its binding partners, RFX7 has been poorly studied. Although it was suggested to have a tumor suppressor role (Fischer et al., 2020), there are no reports on its mC binding preferences. RFX7 was also not significantly enriched in other pull-down MS studies despite its detection (L. Bai et al., 2021). RFX1 has been shown to preferentially bind methylated DNA sequences (Y. Xu et al., 2006) and was also identified in previous pull-down MS studies (Iurlaro et al., 2013; Spruijt et al., 2013), however it was not always significantly enriched for mC (L. Bai et al., 2021). Consistent with **Subsection 3.3.3** and previous pull-down MS studies, UHRF1 was continuously enriched across all three experiments for mC/mC. UHRF1 acts as a key factor for DNMT1 maintenance methylation through recognition of mC (Arita et al., 2008; Avvakumov et al., 2008). Furthermore, several members of the forkhead box (FOX) transcription factor superfamily were found to be enriched for mC/mC. Notably, FOXC1 was consistently enriched across all three experiments, while FOXA1 was identified in experiments 1 and 2 and FOXF1 in experiments 2 and 3. FOX proteins have been reported to bind mC and fC (Iurlaro et al., 2013; Spruijt et al., 2013), though little is known about their direct interaction with methylated CpG sites. They play diverse roles in embryonic and adult development and are linked to chromatin remodeling and nuclear localization (Carlsson and Mahlapuu, 2002). FOXC1 binds strongly to the non CpG containing DNA motif GTAAATAAA (Pierrou et al., 1994) and regulates transcription through its N- and C-terminal activation domains and a phosphorylated inhibitory domain, suggesting regulation via kinase/phosphatase activity (Lines, 2002). FOXA1 functions as a pioneer factor (Cirillo et al., 2002), capable of accessing hypermethylated heterochromatin, however, there is no report of a direct recognition of methylated cytosines. FOXF1 is essential for lung development and repair (Ren et al., 2014). Its expression is induced by Sonic hedgehog signaling, promoting VEGF receptor expression and VEGF signaling in embryonic endothelial cells (Mahlapuu et al., 2001). The established mC readers of the MBD family (Du et al., 2015 and Subsection 1.2.4) were discovered as well, with MBD4 being identified across all three experiments and MBD2 and MECP2 in experiments 2 and 3. Although strongest enrichment differences are expected between C/C and modified DNA probes as most proteins are sensitive to DNA modifications, known readers such as MECP2 were not consistently enriched across all three experiments. This is not uncommon in pull-down proteomics experiments, where proteins may appear enriched in only a subset of replicates, even when biologically expected. Several technical factors can contribute to this, particularly stochastic sampling in MS.

In data-dependent acquisition (DDA) modes, peptides near the detection threshold may be selected in some runs but missed in others, leading to inconsistent protein detection. Low-abundance proteins or those with poorly ionizing peptides are especially affected by this. Additional sources of variability stem from minor differences in sample preparation such as nuclear extraction, the pull-down itself, or digestion efficiency, as well as stringent data filtering, where proteins detected in all runs may be excluded due to falling below thresholds for intensity, peptide count, or FDR. Missing value imputation may also lead to apparent dropouts if low-level signals fall below quantification limits, even when the protein is present. Finally, two proteins were detected in only one experiment each, namely DEK which was enriched for mC/mC in experiment 1 and Polypyrimidine Tract Binding Protein 3 (PTBP3) in experiment 2. The role of DEK in transcriptional regulation appears paradoxical. While some studies report that DEK is essential for maintaining transcriptionally inactive heterochromatin (Kappes et al., 2011), others show its accumulation in euchromatin, as shown by immunofluorescence and immunoprecipitation studies (H.-g. Hu et al., 2007; Sawatsubashi et al., 2010). Interestingly, DEK was suggested as a transcriptional co-activator by interacting with TFAP2A, which was identified as a C/C binder in experiment 1 (Campillos, 2003). PTBP3 is a RNA-binding protein involved in RNA splicing, translational activation, and mRNA stabilization (C. Xie et al., 2022). However, its binding preference for methylated DNA was not reported yet.

Taken together, the detection of expected proteins, including typical readers and antireaders of mC, and previously identified interactors, such as those from **Subsection 3.3.3** and L. Bai et al., 2021; Iurlaro et al., 2013; Spruijt et al., 2013, confirms the technical robustness and biological relevance of this approach. This provides a solid foundation for further identification of readers for hmC-containing dyads.

Readers and antireaders of hmC/hmC from HEK293T cells. hmC/hmC appears to be the dyad with the least amount of interactors as was previously reported (Iurlaro et al., 2013; Spruijt et al., 2013) and confirmed by the present pull-down MS study with only 20 readers in experiments 1 and 2, while C/C and mC/mC modified dyads had 66 and 44 readers, respectively (**Figure 3.17a**). The largest differences in readers are expected between symmetrically modified probes, which is why differences in enrichments between hmC/hmC and C/C and mC/mC were evaluated first.

Proteins that preferentially bound to C/C over hmC/hmC in both experiments were partially overlapping with those that were previously enriched over mC/mC (**Figure 3.18a, b and e**). These are binders that are clearly rejected by cytosine modifications, among them, the aforementioned members of the PRC1.1 (PCGF1, KDM2B and BCOR) and MLL1 complexes (KMT2A, ASH2L, DPY30, RBBP5 and HCFC2) were found as well as L3MBTL3, SAMD1, ZBTB2 and BEND3. 61 proteins were enriched for C/C over hmC/hmC in both experiments (**Figure 3.18e**) including Centromere protein B (CENP-B), which is a DNA-binding protein that recognizes and binds to a specific motif known as the CENP-B box at centromers. Notably, CpG methylation at these sites was reported to significantly reduce CENP-Bs binding affinity (Tanaka et al., 2005). Also Menin (MEN1) was enriched for C/C over hmC/hmC, which is a nuclear protein that interacts with multiple transcriptional regulators, particularly chromatin-modifying enzymes. Among these, MEN1 forms complexes with histone methyltransferases such as KMT2A (Yokoyama et al., 2004). Zinc finger transcription factors were enriched as well. ZIC1 has been implicated in the progression of various human cancers and it itself is frequently down regulated via promoter hypermethylation (L. J. Wang et al., 2009). However, direct interactions between ZIC1 and cytosine modifications have not been reported to date. Similar to ZIC1, ZIC2 is involved in chromatin regulation. It has been identified as a cofactor of the MBD3-NuRD complex, with which it co-occupies enhancer regions across the genome. Together, ZIC2 and Mbd3-NuRD contribute to the regulation of chromatin structure and gene expression in ESCs. Depletion of ZIC2 impairs lineage commitment, highlighting its role in pluripotency and differentiation (Z. Luo et al., 2015). Moreover,

ESRRA (Estrogen-Related Receptor Alpha), a nuclear receptor that regulates the transcription of key metabolic enzymes involved in glycolysis, the tricarboxylic acid (TCA) cycle, and oxidative phosphorylation (Huss et al., 2015) was consistently depleted by hmC/hmC. In addition to its metabolic role, ESRRA also promotes tumor progression through interactions with oncogenic factors such as MYC (Cai et al., 2013). Continuously depleted was also Aryl Hydrocarbon Receptor (AHR), a ligand-activated transcription factor known for mediating cellular responses to xenobiotic and endogenous ligands (Nguyen and Bradfield, 2008). Further studies have demonstrated its role in mast cells, where AHR regulates environmentally induced differentiation, growth, and cellular functions (Kawasaki et al., 2014). Moreover, FLYWCH1 was depleted, a C2H2-type zinc finger transcription factor. Although the molecular mechanisms governing its activity remain elusive, FLYWCH1 has been shown to participate in transcriptional regulation, particularly within the cardiovascular system (Asl et al., 2015). C2H2 zinc finger domains are traditionally associated with DNA binding or protein–protein interactions (Muhammad et al., 2018). Nuclear Factor I/X (NFIX), a member of the Nuclear Factor I family of transcription factors, was also consistently depleted by hmC/hmC. It binds DNA as a homodimer or heterodimer to palindromic consensus sequences (TTGGC(N5)GCCAA) and can also bind as a monomer to each half-site independently (Nowock et al., 1985). Moreover, ARNT and TFAM were consistently enriched for C/C over hmC/hmC in both experiments. Interestingly, in experiment 2, even RFX1 and RFX5 were significantly enriched for C/C over hmC/hmC indicating strong repulsion of RFX proteins from hmC.

Taking a closer look at hmC/hmC readers, 13 proteins were enriched for hmC/hmC over C/C (**Figure 3.18a, b, e**) and UHRF1 was the only protein enriched consistently over C/C in both experiments, which is an expected finding as UHRF1 can accommodate hmC via its SRA domain (Frauer et al., 2011). The remaining proteins were solely enriched in one experiment. Among them, FOXC1 was identified for reading hmC/hmC over C/C in experiment 1, which was already identified as a mC/mC reader in the previous section. Furthermore, RNA-Binding Motif Protein 14 (RBM14) was significantly enriched for hmC/hmC, an adapter protein involved in transcriptional regulation and epigenetic modification. It was identified as a factor binding to DNMT3B and shown to recruit DNMT3B to specific DNA regions of the TLR4 gene likely facilitating DNA methylation at this locus (Narabayashi et al., 2022). Whether the interactions of RBM14 with DNMT3B and DNA are direct or indirect is not yet clarified, but it may lead to complex formation with additional proteins. Moreover, RBM14 has been reported to suppress c-MYC expression via recruitment of HDACs highlighting its broader role in modulating chromatin state and gene expression (Kang et al., 2008). Also Cadherin EGF LAG seven-pass G-type receptor (CELSR1) was identified as a hmC/hmC preferred binder, which is surprising as it is localized in the plasma membrane, where it is a core component of the planar cell polarity pathway, crucial for regulating tissue polarity across epithelial cells in various organs and species (Tissir and Goffinet, 2013). This could be a contaminant artifact of the nuclear extraction preparation. Further, EXOSC5 was identified, which is a non-catalytic subunit of the RNA exosome complex, responsible for RNA processing and degradation (X. Guo et al., 2007). Epigenetic connections were not reported yet. Interestingly, BAZ1B was enriched in the hmC/hmC pull-down, which is a bromodomain-containing protein and regulatory subunit of the WICH (WSTF-ISWI) chromatin remodeling complexes. These complexes regulate chromatin accessibility during essential DNA-templated processes such as transcription, replication, and repair by organizing nucleosome positioning. BAZ1B plays a critical role in maintaining proper chromatin architecture and genomic function (Bozhenok, 2002; Oppikofer et al., 2017). Lastly, Splicing Factor 3A Subunit 3 (SF3A3) was identified for hmC/hmC in experiment 1, which is a core component of the precatalytic B complex of the spliceosome. It is essential for precursor mRNA splicing, participating in the regulation of alternative splicing events that contribute to transcriptomic complexity (Zhan et al., 2018). In experiment 2, MECP2 was found to be enriched for hmC/hmC over C/C, which was identified as a mC/mC reader in the previous section and is a well

established mC and hmC reader (Lagger et al., 2017). CHAF1A and CHAF1B were enriched with the hmC/hmC probe as well. CHAF1B is a master epigenetic regulator of transcription and is closely associated with cell state maintenance and self-renewal (Dean et al., 2023). CHAF1A and CHAF1B are components of the heterotrimeric CAF-1 complex, which plays a pivotal role in chromatin organization during DNA replication. The CAF-1 complex facilitates the deposition of histone H3–H4 tetramers onto newly synthesized DNA during S phase, thereby preserving chromatin structure and genomic integrity (Kaufman et al., 1995). While CAF-1 is well recognized for its roles in heterochromatin maintenance and cell identity, its direct contribution to gene expression regulation in coordination with DNA methylation is largely unexplored (Ng et al., 2019). Additionally, three RNA binders were enriched with the hmC/hmC probe. Among them, HNRNPLL was identified, a member of the heterogeneous nuclear ribonucleoprotein (hnRNP) family, a group of RNA-binding proteins predominantly involved in pre-mRNA processing, including splicing (Marasco and Kornblihtt, 2023). Although HNRNPLL has been shown to bind various long non-coding RNAs (lncRNAs) that influence gene expression (Gu et al., 2020), its role in direct epigenetic regulation through DNA binding has not been reported yet. Another RNA binder was RPP25L, which is a component related to the ribonuclease P/MRP complexes, which are involved in the maturation and processing of various RNA species. While its precise biological role is still being elucidated, RPP25L is thought to contribute to RNA metabolism and potentially influence gene expression at the post-transcriptional level (Köferle et al., 2022). The last RNA binding protein identified was SYF2, which is a core component of the spliceosome and is involved in the regulation of pre-mRNA splicing (Chanarat and Sträßer, 2013).

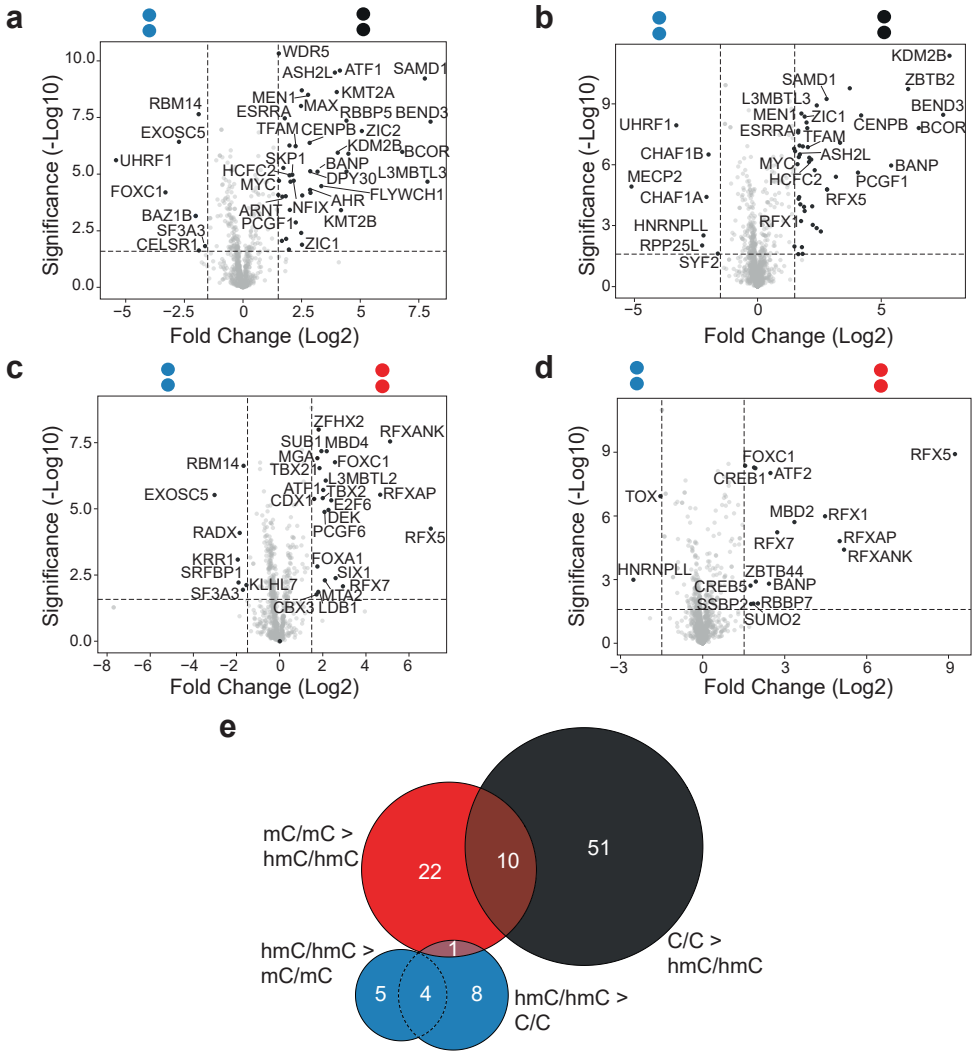


Figure 3.18: Readers and antireaders of hmC/hmC. (a) Volcano plot comparing hmC/hmC (blue circles) with C/C (black circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 1. **(b)** Volcano plot comparing hmC/hmC (blue circles) with C/C (black circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 2. **(c)** Volcano plot comparing hmC/hmC (blue circles) with mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 1. **(d)** Volcano plot comparing hmC/hmC (blue circles) with mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 2. See Table A.10, Table A.11, Table A.12, Table A.13 for full protein lists. **(e)** Readers and antireaders of hmC/hmC (blue) compared with C/C (black) and mC/mC (red) visualized in a Venn diagram. Corresponding protein names are listed in Table A.14.

33 proteins were enriched for mC/mC over hmC/hmC (**Figure 3.18c, d and e**). Specific mC/mC binders that were enriched in both experiments over hmC/hmC and also enriched over C/C comprised the aforementioned RFX proteins RFX5, RFXANK, RFXAP and RFX7 as well as FOXC1. Interestingly, FOXC1 was a common antireader of C/C indicating a binding preference of mC/mC > hmC/hmC > C/C. Moreover, in experiment 1, FOXA1 and SUB1 were enriched for mC/mC over the other two symmetric modifications. In experiment 1, the before discussed readers MBD4, DEK and ZFHX2 were enriched for mC/mC over hmC/hmC and MBD2, ZFHX3 and RFX1 were enriched in experiment 2. Previous studies found that ZFHX3 can either positively (Qi et al., 2008) or negatively (Mori et al., 2007) regulate gene expression via AT motifs, depending on the functional domains and binding partners involved (Berry et al., 2001; Sakata et al., 2014). Interestingly, some proteins that were previously depleted for mC/mC compared to C/C were now enriched for mC/mC over hmC/hmC, indicating a modification preference of these proteins of C/C > mC/mC > hmC/hmC. Among them were PRC1.6 members such as MGA, PCGF6, E2F6 and L3MBTL2 in experiment 1 and BANP and CREB1 in experiment 2. Other mC/mC enriched and hmC/hmC depleted proteins included several transcription factors, chromatin regulators, and proteins involved in genome maintenance, such as CBX3, an oncoprotein implicated in the regulation of gene expression in cancer, which is recruited to chromatin via histone H3 Lys9 trimethylation (H3K9me3) and is involved in establishing and maintaining heterochromatin (Machida et al., 2018; Y. Xu et al., 2020) and LIM domain-binding protein 1 (LDB1), a key component of a multimeric erythroid transcriptional complex, often referred to as the LDB1 complex. This complex binds to a composite E-box/GATA motif frequently identified in ChIP-Seq studies. LDB1 plays a crucial role in mediating long-range promoter–enhancer interactions that regulate gene expression during erythroid differentiation in mice (Love et al., 2014). Moreover, several T-box family transcription factors were detected, including T-box transcription factor (TBX)2 and TBX21. TBX2 mainly functions as a transcriptional repressor and is essential in developmental processes (Papaioannou and Silver, 1998), while TBX21 regulates the differentiation and function of various immune cell types (Lazarevic and Glimcher, 2011). Sine oculis homeobox 1 (SIX1) was identified as well, which is a developmental transcription factor that acts as both an activator and repressor of transcription. It regulates genes involved in organogenesis and differentiation (D. Liu et al., 2014; Patrick et al., 2013). Finally, Caudal-type homeobox 1 (CDX1), which is involved in the regulation of cell proliferation and differentiation, particularly within the intestinal epithelium (Silberg et al., 2000) was enriched over hmC/hmC, as well as Metastasis-associated tumor gene family 2 (MTA2), a component of the NuRD complex, that contributes to chromatin remodeling and transcriptional repression, further highlighting the prevalence of chromatin-associated factors among mC/mC-enriched proteins. Activating transcription factor (ATF)1 and ATF2 were enriched for mC/mC which belong to the activating transcription factor family and are involved in regulating gene expression through dimerization with other Activator protein (AP)-1 family members such as CREB1 (Kvietikova et al., 1995). These interactions influence numerous cellular processes including differentiation and stress responses. In experiment 2, Zinc Finger And BTB Domain Containing (ZBTB)44 and small ubiquitin-like modifier (SUMO)2 were identified. SUMO2, is known to regulate diverse cellular processes, including transcriptional activity and maintenance of genomic integrity (Gill, 2005). Also CREB5 was identified, which was reported to interact with the androgen receptor transcriptional machinery, including FOXA1. Notably, CREB5 has been shown to reprogram FOXA1 nuclear interactions, contributing to resistance against AR-targeted therapies (Hwang et al., 2022). Moreover, RBBP7 was enriched, which is a component of the NuRD complex. It is a WD40 repeat-containing protein that interacts with MTA proteins and histones and plays a crucial role in chromatin remodeling and transcriptional repression (Brasen et al., 2017). Finally, also the mitochondrial Single Stranded DNA Binding Protein (SSBP)2 was identified (Tiranti et al., 1995).

Nine proteins were identified as potential readers for hmC/hmC that are depleted by mC/mC

(**Figure 3.18c, d and e**). These included several factors involved in DNA replication, ribosome biogenesis, transcriptional regulation and protein degradation. Among them, in experiment 1, RBM14 was detected as well as RPA1 Related Single Stranded DNA Binding Protein, X-Linked (RADX), which is a ssDNA binding protein that localizes at replication forks, where it maintains fork stability (Adolph et al., 2021). Also KRR1 was enriched, a component of the small subunit processosome which is essential for the biogenesis of the ribosome (S. Singh et al., 2021). Moreover, Serum Response Factor Binding Protein (SRFBP)1 was detected, which has been implicated in histone modification and mitochondrial gene regulation. Overexpression of SRFBP1 was shown to induce histone H4 deacetylation, repress mitochondrial regulatory genes and subsequently alter mitochondrial function and membrane potential. These changes suggest a role for SRFBP1 in transcriptional regulation during aging, potentially with functional implications outside the nucleus (X. Zhang et al., 2016). Another hmC/hmC enriched protein was Kelch Like Family Member 7 (KLHL7), which is part of the BTB-CUL3-RBX1 (BCR) E3 ubiquitin ligase complex and is responsible for the substrate-specific recognition necessary for K48-linked ubiquitination, targeting proteins for proteasomal degradation (Kigoshi et al., 2011). Finally, Thymocyte Selection Associated High Mobility Group Box (TOX) was identified, a DNA-binding protein belonging to the high-mobility group box family, interacting with DNA in a structure-dependent, rather than sequence-specific, manner (O'Flaherty and Kaye, 2003). TOX functions as a chromatin-modifying scaffold, recruiting diverse chromatin remodeling complexes. It is centrally involved in an epigenetic regulatory cascade that governs T cell exhaustion (TEX) and immune cell differentiation (Khan et al., 2019).

To summarize, no hmC/hmC reader was consistently enriched over both, C/C and mC/mC, probes in both experiments. However, RBM14, EXOSC5 and SF3A3 were enriched over C/C and mC/mC in experiment 1 and HNRNPLL was a specific hmC/hmC reader and C/C and mC/mC antireader in experiment 2 (hmC/hmC overlap in **Figure 3.18e**).

In contrast, there were several antireaders for hmC/hmC that were enriched by both, C/C and mC/mC. Among them were RFX5 (mC/mC enriched in both experiments and C/C enriched in experiment 2), BANP (C/C enriched in both experiments and mC/mC enriched in experiment 2), SSBP2 (C/C enriched in experiment 1, mC/mC enriched in experiment 2), ATF1, E2F6, PCGF6 and LDB1 (C/C enriched in experiment 1, mC/mC enriched in experiment 2), RFX1 and CREB1 (C/C and mC/mC enriched in experiment 2) and MGA (C/C enriched experiment in 2, mC/mC enriched in experiment 1) (C/C and mC/mC overlap in **Figure 3.18e**).

Readers and antireaders of hmC/C from HEK293T cells. Pull-downs with hmC/C modified probes yielded 47 readers constituting the second highest reader number after C/C (66 readers) and being in the same range with mC/mC (44 readers) (**Figure 3.17a**).

Differences in proteins that were enriched or depleted with hmC/C compared to C/C were much smaller than the previously observed changes between symmetrically modified probes (**Figure 3.19a, b and g**). This can be ascribed to the similarities of C/C and hmC/C as they differ only in one additional hmC. The only protein that was consistently enriched in both experiments for hmC/C over C/C was UHRF1 underlining its hmC preference even when it was only present in one strand, which is in line with reports on UHRF1 binding only to the modification on one DNA strand (Arita et al., 2008). Proteins enriched for hmC/C over C/C in experiment 1 included EXOSC5, which was already enriched for hmC/hmC over C/C, and the RNA binder Signal Recognition Particle 68 (SRP68), a component of the SRP complex (J. H. Lee et al., 2021). In experiment 2, hmC/C enriched proteins included MECP2, Cell Division Cycle Associated 7 Like (CDCA7L), which is an interactor of MYC frequently found to be disregulated across multiple cancer types (A. Huang et al., 2005), and RBM45, which contains two RNA recognition motifs that bind RNA and ssDNA, particularly recognizing the GGACGG motif (X. Chen

et al., 2021), that is not present in the VEGFA probe. In addition to RNA binding, RBM45 is involved in the DNA damage response and in chromatin remodeling by influencing HDAC1 recruitment (Gong et al., 2017) and it is also known to bind m6A-modified RNA (Choi et al., 2022). In contrast, seven proteins were consistently enriched for C/C over hmC/C (**Figure 3.19a, b, g**) primarily comprising PRC1.1 proteins (BCOR, HCFC2, PCGF1 and KDM2B) as well as CENPB, ZBTB2 and NFIX. Other proteins enriched in only one of the two experiments included many of the aforementioned C/C readers. These proteins apparently solely bind to unmodified DNA and are repelled by modification of one of the two Cs in the CpG dyad.

Next, proteins recruited or repelled in the transition from hemi-hydroxymethylated dyads to fully hydroxymethylated dyads were determined. In total, 29 proteins were found significantly enriched for hmC/C over hmC/hmC (**Figure 3.19c, d, g**). Six of them were consistently enriched in both experiments being L3MBTL3, SAMD1, BCOR, BEND3, BANP and RBM45, which were primarily C/C readers indicating that hemi-hydroxymethylation does not completely abolish their binding. Strikingly, some of these proteins were depleted for hmC/C when comparing with C/C, but are enriched for hmC/C when comparing with hmC/hmC revealing their affinity of C/C > hmC/C > hmC/hmC (overlap of black and grey circles in (**Figure 3.19g**). Further proteins enriched for hmC/C in experiment 1 included ATF1, ZIC1 and ZIC2, LDB1, SIX1, MAX, ASH2L and RBBP5. Moreover, Synaptotagmin-binding, cytoplasmic RNA-interacting protein (SYNCRIP), an AU-rich RNA-binding protein involved in multiple aspects of mRNA regulation (Bannai et al., 2004), was enriched. Interestingly, SYNCRIP was reported to regulate the expression of DNMT3A, but not DNMT1 or DNMT3B (C. Li et al., 2024). Paired Box (PAX)1, a tumor suppressor that forms a complex with the PRC proteins WDR5 and SET1B, which promotes H3K4me3 histone modification and activates the expression of multiple phosphatases (PTPRR, DUSP1, DUSP5, DUSP6) in cervical cancer cells (P.-H. Su et al., 2019) was found to be enriched for hmC/C as well. Mitochondrial transcription factor A (TFAM) was also enriched, whose DNA-binding affinity at the HSP1 promoter was reported to be increased by C methylation (Dostal and Churchill, 2019). Lastly, NKX2-5 was enriched, a transcription factor crucial for cardiac development, which binds its cognate site TTAAGTG, that is not present in VEGFA, to regulate genes involved in cardiomyocyte differentiation and cardiac morphogenesis (Carlson et al., 2010). In experiment 2, SSBP2, ARNT, CDCA7L, TFPT, CREB1, CCAR2, PCGF1, KDM2B and AHR were enriched for hmC/C. NKX2-1, another homeobox transcription factor was also enriched. It exhibits cell-type-specific DNA binding that is not solely determined by sequence but shaped by cofactors like YAP/TAZ (in AT1 cells) and CEBPs (in AT2 cells). Its role in epigenetic state maintenance and cell fate restriction was previously highlighted (Little et al., 2021). Also NKX2-3 was enriched, which regulates tissue-specific gene expression during the differentiation and development of various organs (Holland, 2013). Runt-related transcription factor (RUNX)2 was enriched in experiment 2, which functions through its Runt domain, forming heterodimers with the cofactor CBF-β which enhances its DNA-binding to its target motif and supports its role in bone development and osteoblast differentiation (Tahirov et al., 2001). Interestingly, also RFX5 was enriched for hmC/C in experiment 2, which was already enriched for C/C over hmC/hmC indicating a strong repulsion of RFX5 by hmC/hmC. Of the eight proteins found to be enriched for hmC/hmC over hmC/C, only DACH1 and UHRF1 were consistently enriched in both experiments (**Figure 3.19c, d, g**). UHRF1 binding significantly stronger to hmC/hmC indicated its strong preference for hmC which could either indicate its preference for hmC in both strands or is due to higher overall number of hmC sites if two UHRF1 can bind per dyad. Either way increasing number of hmC in the CpG dyad facilitated UHRF1 binding as seen by comparison of C/C, hmC/C and hmC/hmC dyads. Other proteins that were enriched for hmC/hmC over hmC/C but not consistently across the experiments included FOXC1, BAZ1B (previously enriched for hmC/hmC over C/C), SRFBP1 (previously enriched for hmC/hmC over mC/mC) and DnaJ Heat Shock Protein Family (Hsp40) Member B3 (DNAJB3), a member

of the HSP40/DNAJ family of co-chaperones, which interacts with proteins that are key players in stress signaling pathways (Abubaker et al., 2013). Furthermore, CHAF1A and CHAF1B were enriched over hmC/C (previously hmC/hmC over C/C).

Comparison of hmC/C and mC/mC enriched proteins might be biologically not as relevant as multiple conversion steps lie between those two modified dyads. This however explains, why the binding profile of these probes strongly differed (**Figure 3.19e, f, g**). When comparing enriched and depleted proteins of those two probes there is close resemblance to C/C vs mC/mC enriched proteins.

To summarize the hmC/C readers, the only proteins that were consistently enriched for hmC/C and depleted for C/C, mC/mC and hmC/hmC were RBM45 and CDCA7L. Many proteins were not found solely as hmC/C readers but they resembled C/C readers (overlap of black and grey circles in **Figure 3.19g**). Therefore, most proteins that were enriched for hmC/C were also enriched with C/C and were depleted for mC/mC and hmC/hmC. Among them were consistently L3MBTL3, SAMD1, BCOR, BEND3 and BANP for which one C in the dyad seemed sufficient, however, they still preferred C/C. A lot more proteins fell into this category when looking at individual experiments.

There were no hmC/C antireaders identified that were significantly enriched for C/C, mC/mC and hmC/hmC over hmC/C. However, when C/C is excluded from this comparison, hmC/C antireaders, that read mC/mC and hmC/hmC, emerged, comprising FOXC1 and DACH1 (overlap of red and blue circles in **Figure 3.19g**), which were also depleted with C/C.

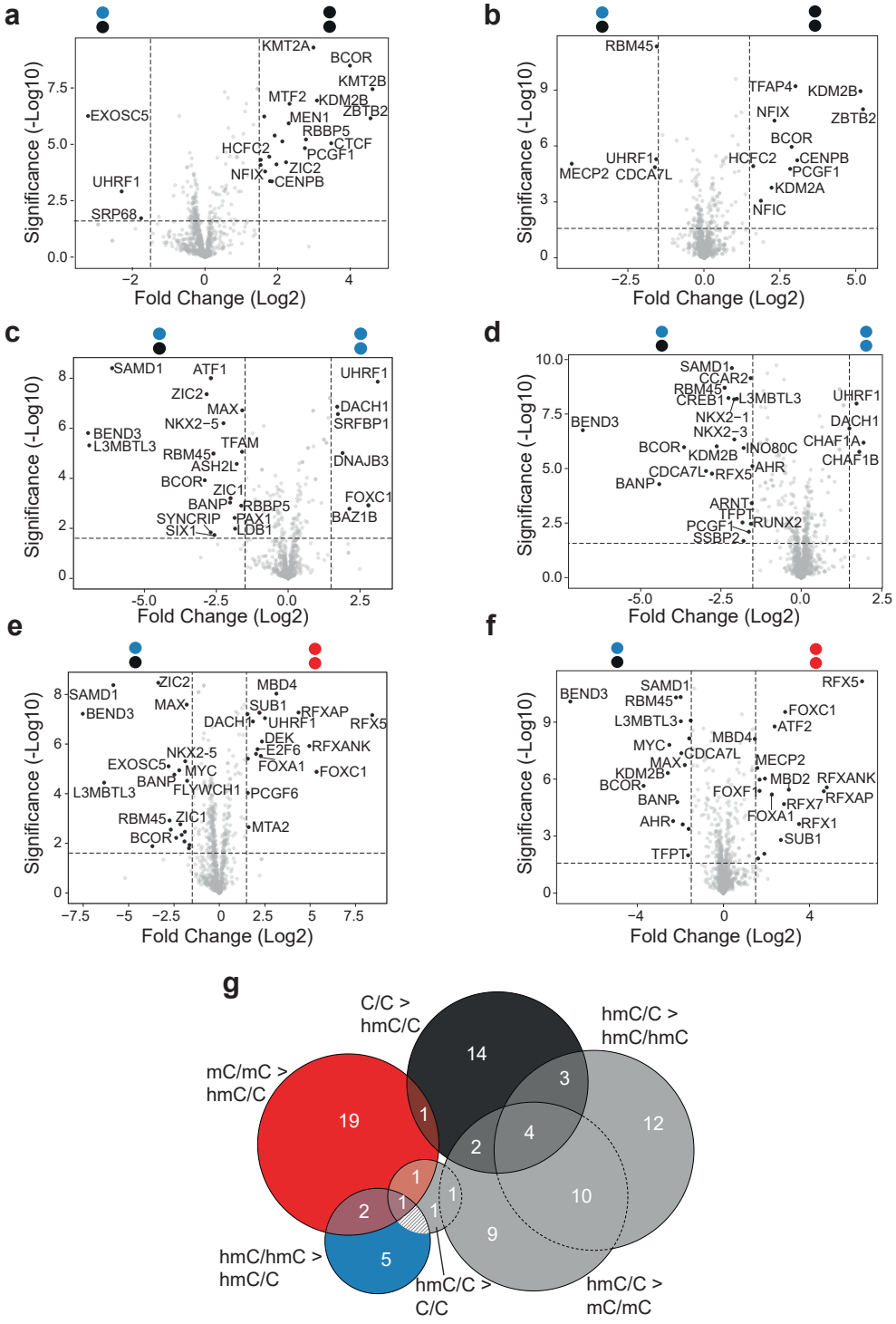


Figure 3.19: Readers and antireaders of hmC/C. (a) Volcano plot comparing hmC/C (blue and black circles) with C/C (black circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 1. **(b)** Volcano plot comparing hmC/C (blue and black circles) with C/C (black circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 2. **(c)** Volcano plot comparing hmC/C (blue and black circles) with hmC/hmC (blue circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 1. **(d)** Volcano plot comparing hmC/C (blue and black circles) with hmC/hmC (blue circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 2. **(e)** Volcano plot comparing hmC/C (blue and black circles) with mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 1. **(f)** Volcano plot comparing hmC/C (blue and black circles) with mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) in experiment 2. See Table A.15, Table A.16, Table A.17, Table A.18, Table A.19, Table A.20 for full protein lists. **(g)** Readers and antireaders of hmC/C (grey) compared with C/C (black), mC/mC (red) and hmC/hmC (blue) visualized in a Venn diagram. Corresponding protein names are listed in Table A.21.

Readers and antireaders of hmC/mC from HEK293T cells. hmC/mC might be the most prevalent asymmetrically modified CpG dyad (Halliwell et al., 2025). 22 readers were identified for hmC/mC, which is half as much as for mC/mC (44 readers) but in the same range with hmC/hmC (20 readers) (**Figure 3.17a**).

When comparing enriched proteins for hmC/mC with unmodified DNA, the difference in enriched proteins was large and the aforementioned C/C binders were again consistently enriched (**Figure 3.20a, b, i**). Additionally, MEN1, MYC and MAX were enriched in experiments 1 and 2. Further enriched proteins that occurred only in one of the experiments were ATF1, LDB1, NKX2-1, NKX2-3, TFPT, and the Euchromatic histone-lysine N-methyltransferase 1 (EHMT1) which catalyzes mono- and dimethylation of histone H3 Lys9 (H3K9me1/2), a key epigenetic mark associated with gene silencing. In addition to histone substrates, it also targets non-histone proteins, thereby extending their influence beyond classical chromatin regulation (Rathert et al., 2008). EHMT1 has been implicated in a wide range of cellular processes, including transcriptional repression and the regulation of DNA methylation patterns (Q. Jiang et al., 2020). Also RUVBL1 was depleted by hmC/mC, which is a highly conserved AAA+ ATPase, playing an important role in transcriptional regulation and chromatin remodeling. It modulates transcriptional activities of factors such as MYC (Wood et al., 2000) and is a core component of several chromatin remodeling complexes, including TIP60, INO80, and SWR1 (Huen et al., 2010). Moreover, INO80C was enriched, another part of the INO80 complex. Recent studies have demonstrated its ability to stabilize chromatin structure and impact transcription and epigenetic integrity, particularly in the context of pluripotency and DNA replication (Cheblal et al., 2020; Kunert et al., 2022). TBRG1, another, only recently identified, interactor of the INO80 complex was also enriched for C/C over hmC/mC, suggesting a potential role in the modulation of chromatin remodeling activity and genome maintenance (Lukauskas et al., 2024). Finally, CBFB, the non-DNA-binding subunit of the RUNX transcription factor complex, which stabilizes RUNX proteins and facilitates their DNA-binding capacity was identified. The CBFB–RUNX interaction is essential for normal development and hematopoietic lineage specification (Y. Ito et al., 2015). Notably, RUNX2 was enriched for C/C over hmC/mC as well.

Proteins consistently enriched for hmC/mC over C/C (including experiment 3, see Figure A.8a) were FOXC1, UHRF1, RFXANK, RFX5 and RFXAP, which were identified as mC/mC or hmC/hmC binders in previous comparisons. A similar picture arised for SUB1 and FOXA1 that were enriched in experiments 1 and 2, MBD4 which was enriched in experiment 1 and FOXF1, MECP2 and ZFHX3 that were enriched in experiments 2 and 3. Also CELSR1 and SYF2 were enriched, which previously preferred also hmC/hmC over C/C. PTBP3, that previously preferred mC/mC over C/C was also enriched with hmC/mC over C/C. Another enriched protein was Zinc finger CCCH-type antiviral protein 1 (ZC3HAV1), also known as ZAP, which is an interferon-stimulated gene product with broad antiviral activity. Upon induction by type I interferons or viral infection, ZAP selectively targets viral RNA through its zinc finger domains, which recognize CpG dinucleotide motifs, leading to RNA degradation or translational repression (MacDonald et al., 2007; Takata et al., 2017). This mechanism represents a form of epigenetic mimicry, where host proteins exploit CpG content—a hallmark of epigenetic regulation—for viral RNA recognition, linking innate immunity with sequence-specific RNA surveillance. Finally, RBM14, Non-POU domain-containing octamer-binding protein (NONO) and splicing factor, proline- and glutamine-rich (SFPQ) were enriched in experiment 3, which are key structural and functional components of paraspeckle nuclear bodies that are involved in the regulation of gene expression, particularly under stress or differentiation conditions (Bond and Fox, 2009). NONO and SFPQ belong to the Drosophila Behavior Human Splicing (DBHS)) family and are multifunctional nuclear factors known to act as both transcriptional co-repressors and co-activators, depending on the cellular context and interacting partners (Knott et al., 2016). Their functional versatility is largely attributed to their heterodimerization, particularly the well-characterized NONO–SFPQ complex, which plays a crucial role

in coordinating various aspects of gene regulation, including transcription, RNA splicing, and DNA repair (Schell et al., 2022). Notably, NONO was reported to exhibit methylation-dependent DNA binding activity, with a specific preference for the methylated motif CCGGTGAC (Park et al., 2013), linking its regulatory function directly to the epigenetic landscape of the genome. This motif is however absent in the VEGFA probe.

The comparison of hmC/mC and mC/mC enriched proteins reveals only a minimal difference at first glance likely attributed to their overall similarity and the difference of just a single oxidation step distinguishing them (**Figure 3.20c, d, i**, and **Figure A.8b** for experiment 3). Proteins consistently enriched for mC/mC over hmC/mC across all 3 experiments comprised RFX5, RFX7, RFXANK and RFXAP and, in experiments 1 and 2, MTA2 (mC/mC over hmC/mC and hmC/hmC), as well as MBD2 and RFX1 in experiments 2 and 3. Interestingly, RFX5 was previously preferred for C/C over hmC/hmC indicating a preference of mC/mC > hmC/mC > C/C > hmC/hmC. hmC/mC enriched proteins in experiment 1 solely included KRR1 (previously hmC/hmC over mC/mC), TFAP2A (previous C/C binder) and NOP16, which was suggested to be a histone mimetic that regulates histone H3K27 methylation and gene repression (Takashima et al., 2023). No significant enrichment of proteins for hmC/mC over mC/mC was observed in experiment 2. In experiment 3 some proteins were enriched, comprising, among others, SFPQ, NONO, RBM14, MYC, MAX, ZIC2, C14orf93 as well as TBP, which binds both CpG and non-CpG promoters effectively, suggesting it's a core part of the transcription initiation complex across promoter types. It can co-recruit TAFs for transcription initiation (Denissov et al., 2007). BEND7 was also enriched for hmC/mC. The previously detected BEN domain containing proteins, BEND3 and BANP, were C/C readers. Also FBXO11 was identified, a F-box protein, which is part of the SCF (SKP1–CUL1–F-box) E3 ubiquitin ligase complex that targets proteins for ubiquitylation and subsequent degradation by the proteasome (Cardozo and Pagano, 2004). Finally, PAX3 was identified, bearing a N-terminal DNA binding domain and a C-terminal transactivation domain. PAX3 recognizes the TCACGC/G DNA motif (Apuzzo and Gros, 2007).

Progressing towards the fully hydroxymethylated dyad after an additional oxidation step, the differences in enrichment were again expectedly small (**Figure 3.20e, f, i**). hmC/mC consistently enriched RFX5 and RFXAP, proteins that were depleted for hmC/mC when compared to mC/mC. Other hmC/mC over hmC/hmC enriched proteins in one experiment only were NOP16, SUB1, FOXA1, SIX1, FOXC1, SSBP2, RFX1 and RFXANK being mostly mC/mC readers. hmC/hmC only enriched one protein over hmC/mC in each experiment, which were UHRF1 and HNRNPLL, respectively. UHRF1 was enriched for hmC/mC over C/C underlining again UHRF1s hmC affinity.

hmC/mC and hmC/C differ only in the methylation status of one C in the dyad which might be a prevalent transition. The difference in significantly enriched proteins was much larger than the before observed differences between hmC/mC with its cognate symmetric dyads (**Figure 3.20g, h, i**). hmC/C consistently enriched the C/C binders L3MBTL3, SAMD1, MYC, MAX, BEND3, BANP and RBM45, while hmC/mC consistently enriched the mC/mC binders SUB1, FOXC1, FOXA1, RFX5 and RFXAP. A single hmC in the dyad seemed to not have a major influence on reader recruitment and the general instructions from C/C and mC/mC could still be implemented. However, the signals from the symmetrically modified probes might be attenuated as becomes apparent when comparing enriched proteins of hmC/C and hmC/mC with the corresponding symmetric probes. Interestingly, Bcl2-associated athanogene 2 (BAG2), a co-chaperone that acts as a nucleotide-exchange factor (NEF) to promote the release of ADP from HSP70 and HSC70 proteins thereby triggering client/substrate protein release (Rauch and Gestwicki, 2014; Takayama et al., 1999) was enriched for hmC/mC over hmC/C. This protein was only found to be enriched in these probes comparison and is unexpected as it is usually not found in the nucleus.

General hmC/mC readers could not be identified accounting to the high similarity of hmC/mC

with its cognate symmetric CpG dyads mC/mC and hmC/hmC. Rather a change in preference was seen for some established readers such as RFX5, RFXANK and RFXAP that were depleted when comparing mC/mC to hmC/mC but enriched for hmC/mC over the other probe modifications. Same is true for other mC/mC readers like SUB1, FOXC1 and FOXA1. NOP16 was enriched for hmC/mC over mC/mC, hmC/C and hmC/hmC but not over C/C. Some proteins were only enriched with hmC/mC but solely over one other probe modification such as FOXF1, ZC3HAV1, ZFHX3, SYF2, PTBP3 and CELSR1 that were enriched over C/C or BAG2 which was enriched over hmC/C and KRR1 and TFAP2A over mC/mC. However, these proteins were not consistently enriched over other modifications as well indicating their preference for hmC/mC but suggesting that they might not be distinguished hmC/mC readers.

Specific antireaders could also not be identified. Antireaders were rather clustered and preferred binding to C/C and hmC/C (gray and black circles in **Figure 3.20i**). ATF1 and LDB1 were depleted from hmC/mC and enriched with C/C, hmC/C and mC/mC but not with hmC/hmC (overlap of red, gray and black circles in **Figure 3.20i**).

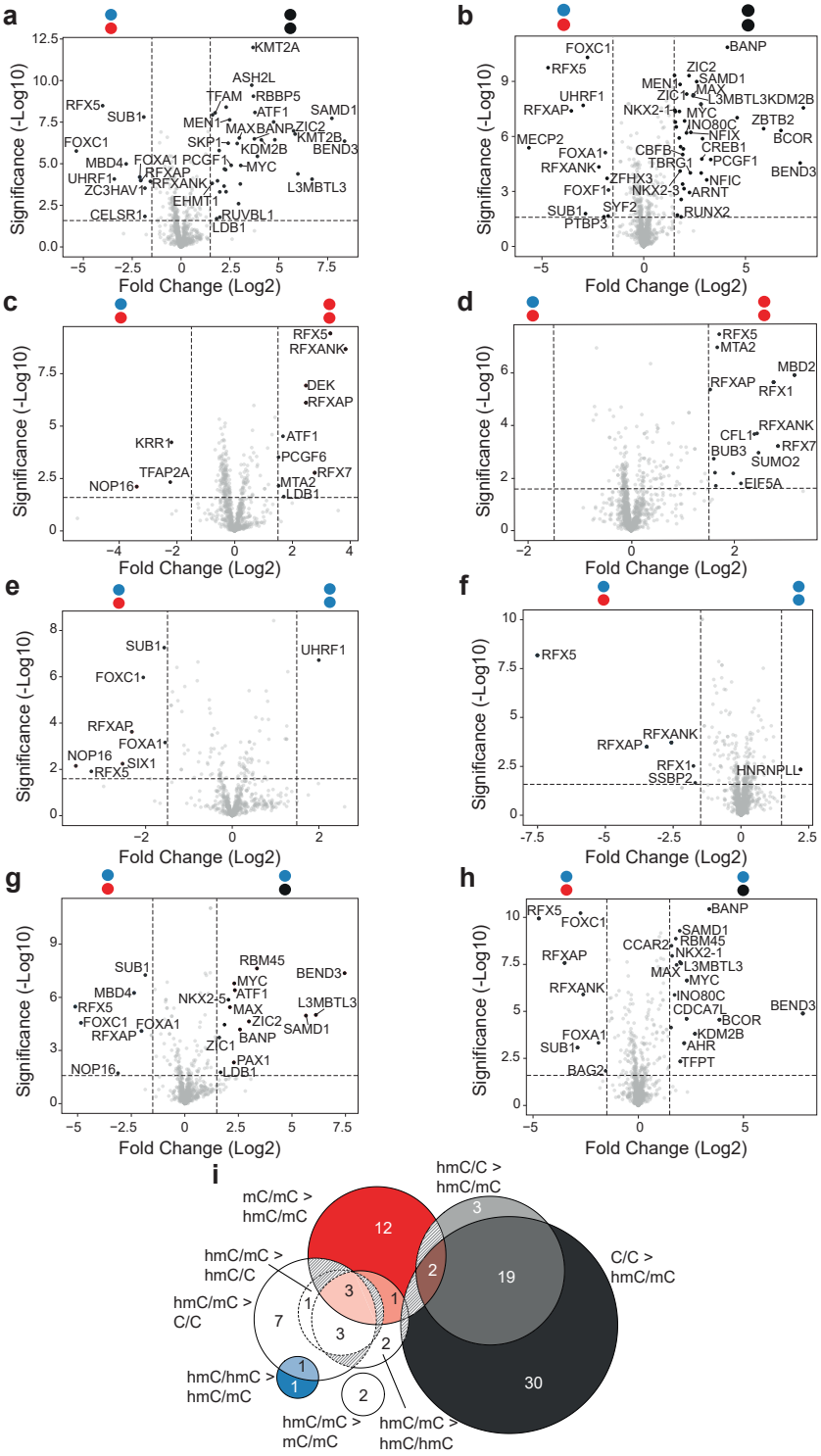


Figure 3.20: Readers and antireaders of hmC/mC. Volcano plots comparing enriched proteins (p-value < 0.025 and log2 fold change >1.5) between hmC/mC (blue and red circles) with C/C (black circles) **(a)** in experiment 1 and **(b)** in experiment 2, hmC/mC (blue and red circles) with mC/mC (red circles) **(c)** in experiment 1 and **(d)** in experiment 2, hmC/mC (blue and red circles) with hmC/hmC (blue circles) **(e)** in experiment 1 and **(f)** in experiment 2, hmC/mC (blue and red circles) with hmC/C (blue and black circles) **(g)** in experiment 1 and **(h)** in experiment 2. See Table A.22, Table A.23, Table A.24, Table A.25, Table A.26, Table A.27, Table A.28, Table A.29 for full protein lists. **(i)** Readers and antireaders of hmC/mC (white) compared with C/C (black), mC/mC (red), hmC/hmC (blue) and hmC/C (gray) visualized in a Venn diagram. Corresponding protein names are listed in Table A.30.

Overall comparison and summary of HEK293T reader proteins. Many proteins showed modification-specific DNA binding. While the difference of enriched proteomes was largest between symmetrically modified probes, also asymmetrically modified probes revealed distinguished binding profiles. The overall difference in enrichment was visualized in a heatmap portraying the most significantly enriched proteins (**Figure 3.21a, b**). Experiment 1 had less significantly enriched proteins in total than experiment 2, however, the stark contrast between the probe modifications becomes evident in both experiments. Especially C/C and mC/mC enriched proteins differ strongly from one another. hmC/hmC binders were partially enriched for C/C and mC/mC as well, while a lot of proteins enriched with mC/mC were depleted from hmC/hmC. hmC in an asymmetric context (hmC/C, hmC/mC) attracted mostly proteins that were enriched with the cognate symmetrically modified probes as can be seen by the similar enrichment profiles of hmC/C and C/C as well as hmC/mC with mC/mC and hmC/hmC. Nonetheless, each modified probe exhibited its own protein enrichment profile. The heatmap presentation further highlights the degree of similarity between the binding profiles. The technical replicates cluster corresponding to the modified probe underlining that reader enrichment indeed has a biological implication. C/C and hmC/C enriched proteins were closely clustered, as well as mC/mC and hmC/mC. hmC/hmC was hierarchically farthest away but still closer to the mC/mC and hmC/mC cluster than to the C/C and hmC/C cluster. Overall comparison of the identified readers of each modified probe with one another displays many overlaps (**Figure 3.21c**, heatmap for experiment 3 in Figure A.8c). The highest number of unique readers was seen for C/C with 31 readers, followed by mC/mC with 21 unique readers. hmC/C and hmC/hmC had 11 unique readers each, while hmC/mC only had three unique readers. However, these readers did not consistently bind to these probe modifications across all comparisons, as can be seen from the three unique hmC/mC readers ZC3HAV1 (hmC/mC over C/C), BAG2 (hmC/mC over hmC/C) and NOP16 (hmC/mC over hmC/hmC, hmC/C and mC/mC) that were not enriched over all four other modified probes. Overlaps of readers between different probe modifications indicated more flexible binding of those proteins and slow transitions of protein compositions upon modification changes. The comparative proteomics results indicated multiple affinity preferences of different proteins as for example for RFX5 and SSBP1 identified as readers for the four probe modifications C/C, hmC/C, hmC/mC and mC/mC. However, by pairwise comparison of individual probes, RFX5 appeared to prefer mC/mC > hmC/mC > C/C and hmC/C > hmC/hmC, while SSBP1 was continuously depleted by hmC/hmC but did not appear in comparisons lacking hmC/hmC. In both cases, hmC/hmC was the undesired probe modification making RFX5 and SSBP2 antireaders of this modification. Same was true for UHRF1 and MECP2 that were enriched with hmC/C, hmC/hmC, hmC/mC and mC/mC. UHRF1 therefore showed a preference of hmC/hmC > hmC/mC and hmC/C > C/C as well as a preference of mC/mC > hmC/C > C/C, while MECP2s preferences were mC/mC > hmC/C > C/C as well as hmC/hmC and hmC/mC and hmC/C > C/C suggesting these proteins to be strong antireaders of C/C. Other proteins also showed preferences for more than one probe modification. The largest overlap in readers was seen for C/C and hmC/C and for mC/mC and hmC/mC in agreement with the tight clustering of significantly enriched proteins from the heatmap.

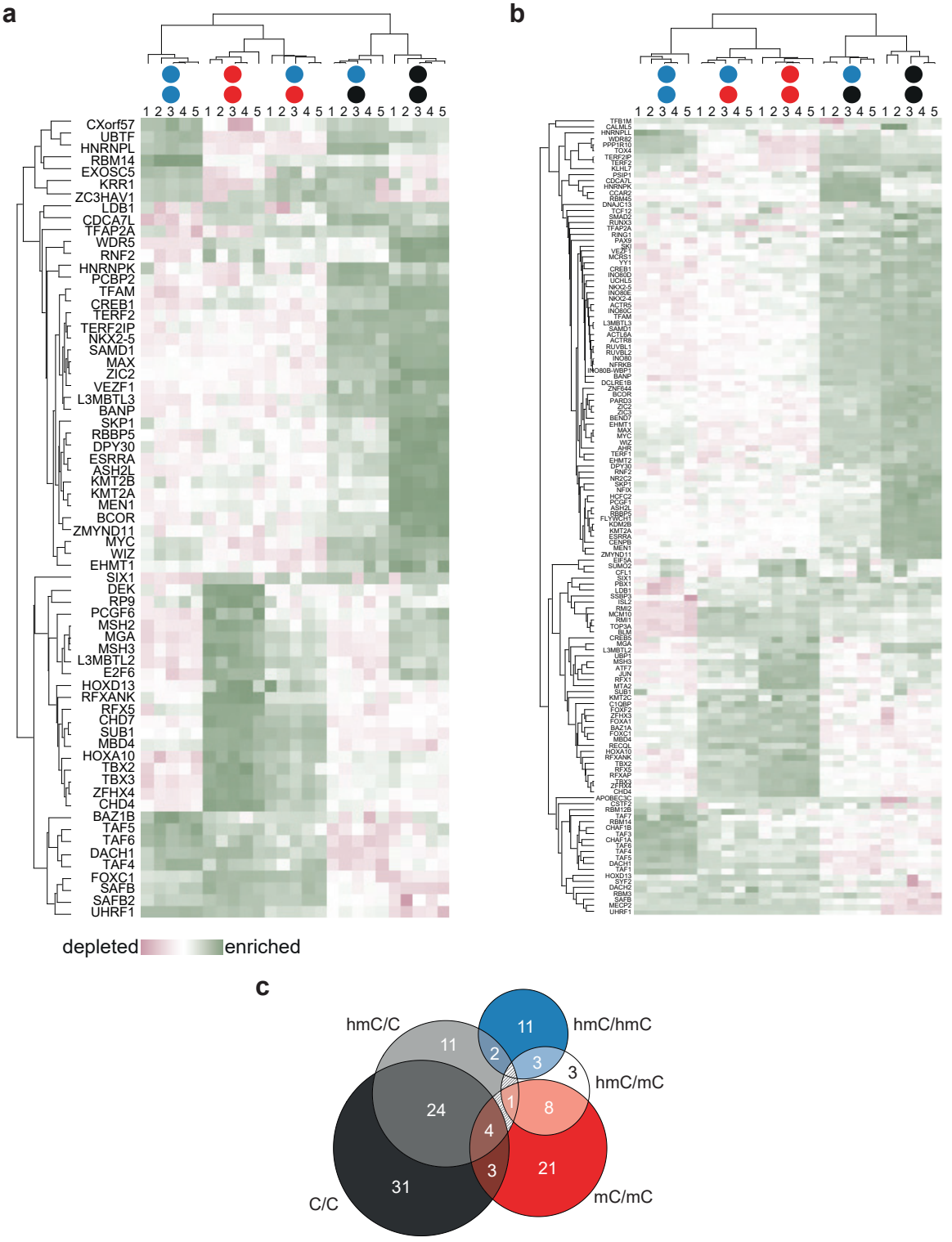


Figure 3.21: Overall comparison of HEK293T reader proteins. (a) Heatmap for experiment 1 of correlation-based clustering of the LFQ-intensities after log2 transformation and normalization by row mean subtraction. Proteins included in the clustering significantly bind to at least one of the probes as determined by an ANOVA test (p-value < 0.025 and s0=1). **(b)** Heatmap for experiment 2 of correlation-based clustering of the LFQ-intensities after log2 transformation and normalization by row mean subtraction. Proteins included in the clustering significantly bind to at least one of the probes as determined by an ANOVA test (p-value < 0.025 and s0=1). **(c)** Venn diagram for significantly enriched proteins for indicated probes, protein names are listed in Table A.31.

3.4.2 Discovery of potential readers from mouse brain nuclear extracts

In addition to the protein pull-down MS studies performed using HEK293T cells, a similar experiment was conducted with proteins extracted from mouse brain tissue, which is biologically relevant given the relatively high abundance of hmC in the brain. Previous studies had already explored mouse brain proteins in the context of symmetric hmC/hmC probes (L. Bai et al., 2021; Spruijt et al., 2013). Actively expressed genes in the brain have previously been found to be enriched in hmC (Hahn et al., 2013; Jin et al., 2011; C.-X. Song et al., 2011), suggesting that certain proteins, which are either uniquely expressed or abundant in brain tissue, could specifically recognize hmC. Identifying hmC-binding proteins directly from brain tissue provides a more physiologically relevant view of potential interactions, giving more insights into the biological role of hmC. To explore interactions between hmC-containing CpG dyads and mouse brain proteins, proteins were extracted from the brain of a 16 weeks old female C57BL6J mouse. The use of tissue rather than cell culture also minimizes the risk of artifacts and biases introduced by artificial culture conditions. In contrast to cultured cells, which often exhibit altered protein expression profiles and may not faithfully represent in vivo physiology, tissue-derived proteins maintain the natural complexity and abundance patterns found in the organism. Therefore, studying brain tissue proteins allows a more accurate and unbiased identification of hmC-binding proteins.

Mouse brain protein pull-down set-up

The mouse brain protein pull-down experiment was performed using five differentially modified Sp1 probes, with five technical replicates per modified probe. Nuclear extracts used for the pull-downs are shown **Figure 3.22a**. While they were highly enriched in nuclear proteins compared to the cytoplasmic extract, there was still a high contamination of cytoplasmic proteins as seen by the β-tubulin signal. The Sp1 probes containing the symmetrically and asymmetrically modified probes were prepared according to the asymmetric probe generation protocol described in **Section 3.1 Figure 3.5** and validated via agarose gel electrophoresis. The corresponding gel is depicted in **Figure 3.22b**. As demonstrated in **Section 3.3**, in addition to the pull-downs conducted for MS, a benchmark pull-down was performed using the MBP-MBD2 protein to validate probe functionality and the general pull-down workflow. The results of this control experiment are shown in the Western blot in **Figure 3.22c**. It should be noted that the benchmark was performed using human MBD2, however, for the purpose of validating probe functionality, this was sufficient, as expected modification preferences could be observed indicating MBD2 successfully outcompeted endogenous mouse proteins to bind to the Sp1 probe. This successful validation of probe functionality and workflow robustness enabled progression to MS analysis.

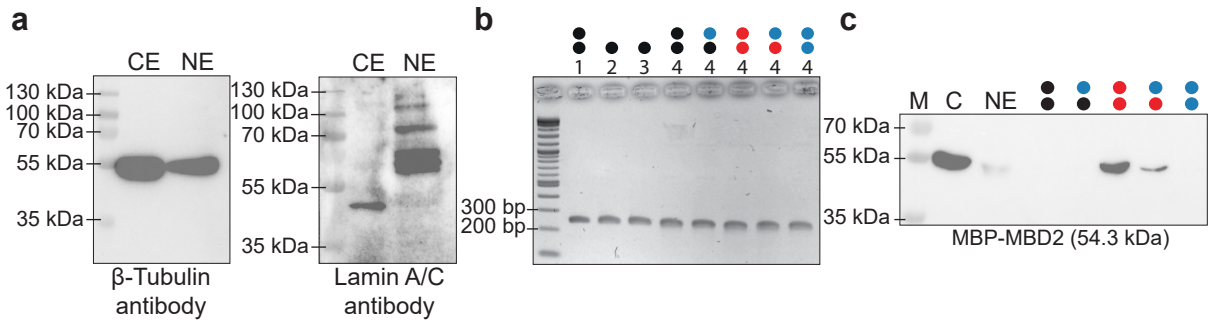


Figure 3.22: Mouse nuclear extracts and Sp1 probes used in mouse nuclear protein pull-downs. (a) Western blot showing the presence of β-Tubulin and Lamin A/C in the cytoplasmic extract (CE) and nuclear extract (NE) Samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M)), expected molecular weights: β-Tubulin 55 kDa, Lamins A and C: 63 and 70 kDa. **(b)** Agarose gel analyses of all employed Sp1 probes used for proteomics pull-downs. Samples were collected during the process of generating (a)symmetric Sp1 probes and run on a 2 % agarose gel with a 1 kb Plus DNA ladder (New England Biolabs). The numbering corresponds to **Section 3.1, Figure 3.5** representing 1: the PCR product, 2: ssDNA, 3: primer hybridized product and 4: (asymmetric) primer extended product. Circles are colored according to the C modifications (C – black, mC – red and hmC – blue) depicting ssDNA (one circle) or dsDNA (two circles). **(c)** Western blot showing the presence of MBP-MBD2 after pull-down with differentially modified Sp1 probes in parallel to the pull-down conducted for MS. Pure MBP-MBD2 was run in lane C as control and mouse nuclear extracts in lane NE. Samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M), expected molecular weight of MBP-MBD2 is 53 kDa.

Evaluation of sample and data quality of mouse brain MS results

As previously demonstrated for HEK293T pull-down experiments (**Subsection 3.4.1**), data quality was carefully assessed for the mouse pull-down samples. First, the number of identified proteins was evaluated. The number of quantified proteins was averaged over all 25 pull-down experiments and is depicted with the corresponding standard deviation in **Figure 3.23a**. The variation within the experiment is extremely low, as indicated by the error bars. On average, 1033 proteins were identified, slightly more than in HEK293T pull-down experiments 2 and 3.

Next, the log2-transformed LFQ-intensities were examined and displayed in a multi-scatter plot to estimate correlations between the individual samples (**Figure 3.23b**). Consistent data should portray similar intensity distributions between replicates and conditions. "Condition" here refers to a single modified probe. Each of the five modifications were analyzed with five technical replicates. The technical replicates within one condition showed strong positive correlation indicating that the measurements were reproducible across technical replicates, suggesting good consistency in the pull-down MS process. However, high correlation was also observed between replicates of different conditions, suggesting that the different modified probes had a minor impact on distinct protein recruitment. This could reflect non-specific binding and potentially a low number of nuclear proteins due to high cytoplasmic contamination in the protein extract.

Finally, a PCA was performed to assess overall variance in the dataset (**Figure 3.23c**). Replicates did not cluster well by modification, especially hmC/mC replicate 1, and hmC/C replicate 2 were identified as clear outliers. All other replicates clustered tightly together independent of the modification they belong to. PC1 and PC2 together explained less than 50 % of the total variance, indicating complex variation without a single dominant trend. To further explore clustering, hierarchical clustering was applied and shown in a heatmap (**Figure 3.23d**). This confirmed that most replicates within a condition did not cluster well, with few exceptions (e.g. replicates 1-3 for C/C, 1,2 and 4 for mC/mC and 2, 3 and 5 for hmC/hmC). The aforementioned outliers were also more distant from other replicates in the clustering tree. This further indicated that differences in protein enrichment across different modifications were less pronounced than in the HEK293T pull-down experiments.

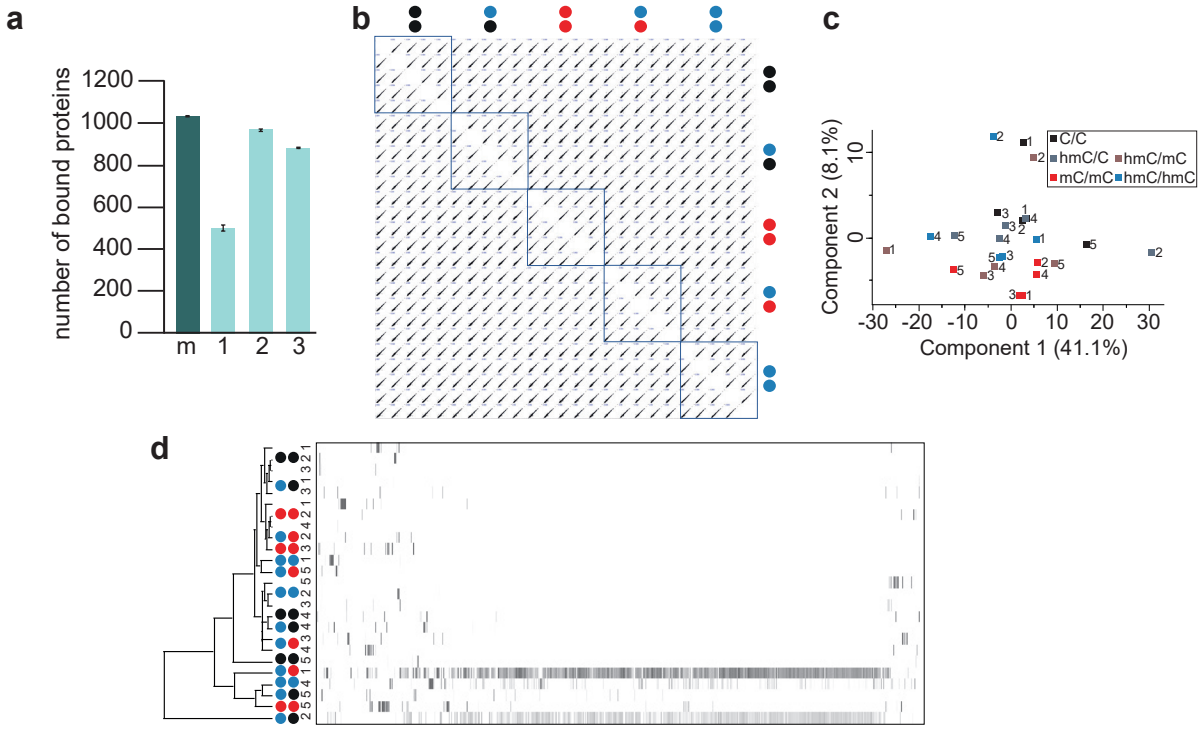


Figure 3.23: Data correlation and quality check of mouse brain MS. (a) Bar diagram representing the number of quantified proteins in the mouse brain pull-down (m) in comparison with HEK293T biological experiments 1-3. Protein number is shown as an average of all technical replicates. Standard deviation is shown as error-bars. **(b)** Multi scatter plot of LFQ intensities showing the correlation of all technical replicates to each other sorted within the modification conditions. Correlations of technical replicates within one group are shown in blue squares. Modifications are depicted as circles, colors correspond to **Figure 3.22**. **(c)** PCA plot of LFQ intensities showing variance across samples. Technical replicates are numbered accordingly and shown in one color per modification as indicated in the legend. **(d)** Hierarchical clustering of samples visualized in a heatmap. Modifications are depicted as circles, colors correspond to **Figure 3.22** and technical replicates are numbered from 1-5.

Identification of potential readers from mouse brain nuclear extracts

In contrast to the HEK293T pull-down experiments (**Subsection 3.4.1**), the overall number of significantly enriched reader proteins was considerably lower in the mouse brain dataset. This reduction may be attributed to either the overall quality of the dataset or smaller differences due to the Sp1 probe or mouse tissue contexts. Among the reader proteins, C/C and mC/mC exhibited the highest number of enriched proteins, each with 16 identified interactors. These were followed by hmC/C with eight, and hmC/mC and hmC/hmC with four identified proteins each (**Figure 3.24a**). While this distribution follows a similar trend to that observed in the HEK293T dataset, a notable difference is the identical number of enriched proteins for C/C and mC/mC.

The relatively lower enrichment differences between C/C and mC/mC are also evident in the corresponding volcano plot, which typically shows substantial differences due to their distinct interaction surfaces. Interestingly, the direct comparison of C/C and mC/mC revealed a greater number of proteins enriched for mC/mC over C/C (**Figure 3.24b**). Among the proteins enriched for mC/mC were several members of the FOX transcription factor family, which were also identified in the study by Iurlaro et al., 2013 in the context of fC/fC. These included Foxk1, a transcriptional regulator implicated in myogenic differentiation (X. Shi et al., 2012), Foxk2, which can bind to T:G mismatches in DNA, despite a lack of reports on enzymatic activity (Fujii and Nakamura, 2010) and Foxp1, a key transcription factor involved in B cell function and lung development, which aids in induction and maintenance of pluripotency through alternative splicing mechanisms (Gabut et al., 2011). Additional enriched pro-

teins included key chromatin regulators and members of the NuRD complex such as Satb1, a chromatin organizer that recruits the NuRD repression complex and physically interacts with its components (Alvarez et al., 2000; Dickinson, 1992; Kloet et al., 2015). NuRD complex components that were enriched with mC/mC over C/C included Mta2, Gatad2b, Rbbp7 and Mbd2. Moreover, Satb2 was enriched, a DNA-binding protein associated with nuclear matrix attachment regions, involved in chromatin remodeling and gene regulation, which plays essential roles the formation of callosal neuron identity and axonal projections (Alcamo et al., 2008; Leone et al., 2015).

Among the proteins preferentially enriched for C/C was Creb1, whose human homologue was enriched for C/C in the HEK293T experiments as well. Leucine repeat adaptor protein 1 (Lurap1), a cytoplasmic protein, was also enriched for C/C. It plays a role in cell polarization and motility by interacting with other proteins, particularly during gastrulation movements, thereby playing a role in developmental processes (X.-N. Cheng et al., 2017). Surprisingly, Trps1 was enriched for C/C, which has been characterized as a repressor within the NuRD and CoREST complexes. It may help maintain histone deacetylation at estrogen receptor (ER)-binding sites facilitating precise spatial and temporal recruitment of ER (Serandour et al., 2018). The core circadian regulator Clock was also enriched for C/C. Clock is a transcriptional activator and chromatin remodeler, which is part of the CLOCK:BMAL1 complex, that binds E-box motifs in core circadian gene promoters, initiating rhythmic transcription through the recruitment of histone-modifying enzymes (Schibler, 2021). Lastly, Baz2b was enriched. Previously, BAZ1B was enriched in the HEK293T experiments for hmC/hmC. These proteins are known to participate in complexes such as NoRC and WICH, which are involved in nucleosome repositioning and transcriptional silencing (Oppikofer et al., 2017).

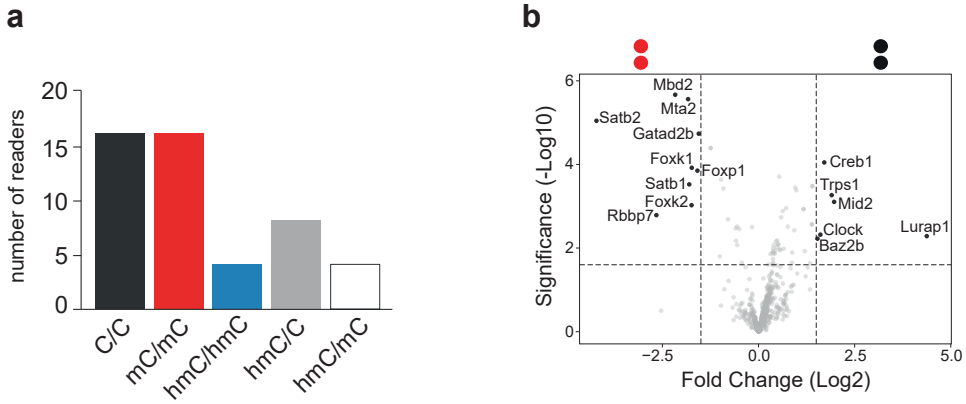


Figure 3.24: General assessment of mouse brain enrichment. (a) Bar diagram showing the distribution of reader numbers for the different modified dyads. **(b)** Volcano plot comparing C/C (black circles) and mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) from the mouse brain pull-down experiment. See Table A.32 for full protein lists.

Readers and antireaders of hmC/hmC from mouse brain. Comparison of hmC/hmC enriched proteins with those identified for the other symmetrically modified probes repeatedly demonstrates a limited number of distinct hmC/hmC readers (**Figure 3.25**). Among them, Uhrf2 was enriched over C/C (**Figure 3.25a**). Interestingly, in the HEK293T pull-down experiments, UHRF1 was consistently enriched instead. Uhrf2 shares a similar domain architecture with Uhrf1 (Bronner et al., 2007) and has been previously reported to be a reader with increased affinity for hmC (Spruijt et al., 2013). Satb2 was found to be enriched over C/C as well, suggesting it is strongly depleted by C/C, as was already observed in comparison to mC/mC. Sfpq was enriched by hmC/hmC, which was enriched in HEK293T experiment 3 for hmC/mC (hmC/hmC was not examined in this set). Notably, other paraspeckle proteins were not co-enriched, possibly indicating that Sfpq bound DNA independently, which is supported by the pres-

ence of a Sfpq DNA-binding domain (X. Song et al., 2005). Sfpq is known to form either homodimers or heterodimers with Nono to bind DNA (Schell et al., 2022). Leucine-rich repeat flightless-interacting protein 2 (Lrrfip2), which was shown to activate the Wnt/β-catenin pathway (J. Liu et al., 2005) was also enriched for hmC/hmC over C/C. It was originally characterized as a novel interactor of the LRR domain of human Flightless I homologue (Fliih), where it negatively regulates NF-κB signaling by interfering with the MyD88–TLR4 interaction in an exposure time-dependent manner (S. Zhang et al., 2013).

Among the proteins that were enriched for C/C and depleted by hmC/hmC was Mafg, a transcriptional repressor that binds to CpG island methylator phenotype genes and recruits a co-repressor complex containing BACH1, CHD8, and DNMT3B, facilitating promoter hypermethylation and transcriptional silencing (Fang et al., 2014). In accordance with the findings from Spruijt et al., 2013, enrichment for C/C was also observed with zinc-finger-containing transcription factors, such as Zbtb2, which was also enriched for C/C in HEK293T experiments, as well as basic leucine zipper-containing proteins such as Jund and Creb1. Creb1 was depleted from mC/mC and hmC/hmC indicating consistent depletion for modified Cs. Parp1 was also enriched for C/C over hmC/hmC, which is a known interactor of Dnmt1 modulated through non-covalent PARylation. This modification has been proposed to inhibit Dnmt1 access to DNA, thereby preserving unmethylated states at regulatory sequences (Caiafa et al., 2009). Other C/C enriched and hmC/hmC depleted proteins included the circadian-rhythm associated protein Arntl, the paraspeckle component Hnrnpk (Naganuma et al., 2012) and Ring1, a key member of PRC1 complexes. Notably, the gene Gm10094, which encodes Sap18b, a subunit of the SIN3A–HDAC1 repression complex, was also enriched. This complex plays a central role in transcriptional repression via histone deacetylation (Y. Zhang et al., 1997).

hmC/hmC over mC/mC enriched proteins comprised Sfpq and Uhrf2, while mC/mC enriched Mbd2, Satb2, Fox family proteins, Bend6, which was shown to be able to accommodate two mCs (X. Zhang et al., 2024) and interestingly, also Creb1 over hmC/hmC indicating a Creb1 selectivity of C/C > mC/mC > hmC/hmC (**Figure 3.25b**).

To conclude on hmC/hmC readers and antireaders in mouse brain, consistent enrichment of Uhrf2 and Sfpq over C/C and mC/mC was observed indicating a specific hmC/hmC preference (blue circle in **Figure 3.25c**). Jund and Creb1 were consistently depleted by hmC/hmC compared with C/C and mC/mC (overlap of black and red circles in **Figure 3.25c**), while Satb2 showed changed preferences as it was enriched for hmC/hmC over C/C but depleted compared to mC/mC indicating a preference of mC/mC > hmC/hmC > C/C (overlap of red and blue circles in **Figure 3.25c**). Notably, more proteins were enriched by mC/mC over hmC/hmC than by C/C over hmC/hmC, which is in contrast to the HEK293T results.

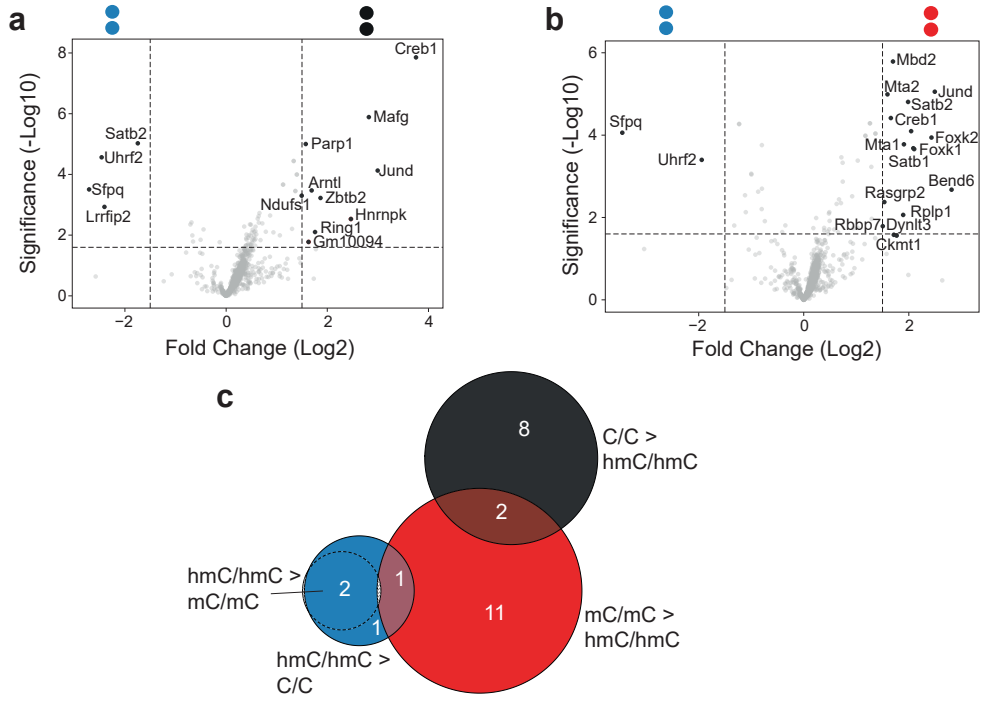


Figure 3.25: Readers and antireaders of hmC/hmC from mouse brain. **(a)** Volcano plot comparing hmC/hmC (blue circles) with C/C (black circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5) **(b)** Volcano plot comparing hmC/hmC (blue circles) with mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5). See Table A.33, Table A.34 for full protein lists. **(c)** Readers and antireaders of hmC/hmC (blue) compared with C/C (black) and mC/mC (red) visualized in a Venn diagram. Corresponding protein names are listed in Table A.35.

Readers and antireaders of hmC/C from mouse brain. Only very few proteins were found to be enriched for hmC/C (**Figure 3.26**). In contrast to the HEK293T pull-down experiments, no protein was significantly enriched for hmC/C over C/C. Also C/C did only enrich few proteins over hmC/C, comprising the aforementioned C/C > hmC/hmC binders Mafg, Creb1 and Jund (**Figure 3.26a**).

hmC/C enriched Creb1 and Hnrnpk over hmC/hmC, which were previously enriched for C/C and depleted for hmC/hmC. Moreover, Ndufs1 was enriched, a core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (Lopez-Fabuel et al., 2016). In addition, Rplp1 was identified, which is part of the pentameric ribosomal stalk complex, that plays an essential role in translation (Artero-Castro et al., 2009) as well as Haus3, which is involved in microtubule nucleation and mitotic spindle organization (Uehara et al., 2009) (**Figure 3.26b**).

hmC/hmC enriched only Sfpq and Satb2 over hmC/C, but not Uhrf2, indicating hmC/C does not fully restrict Uhrf2 binding.

Proteins enriched for hmC/C over mC/mC included Trps1, which was also enriched with C/C over mC/mC, Hnrnpk, which was found to be enriched for C/C over hmC/hmC before, and Mid2, a cytoplasmic E3 ligase (Napolitano et al., 2011) (**Figure 3.26c**). mC/mC over hmC/C enriched proteins comprised again Mbd2, Foxk2 and Foxp1, Satb2, Rbbp7, Mta2, Bend6 and also Dynlt3, a member of the cytoplasmic dynein light chain family, which is essential for proper chromosome alignment and homologous chromosome segregation during meiosis in mouse oocytes (X. Huang et al., 2011).

No protein showed consistent enrichment for hmC/C over all other modified probes attributed to the missing enrichment of proteins for hmC/C over C/C (**Figure 3.26a**). However, Hnrnpk was enriched over hmC/hmC and mC/mC (overlap of gray circles in **Figure 3.26e**). hmC/C was antiread by Satb2 compared to hmC/hmC and mC/mC (overlap between red and blue circles in **Figure 3.26e**). Creb1 showed a decreasing affinity from C/C > hmC/C > hmC/hmC, indicating loss in affinity with

increasing hmC (overlap between black and gray circles in **Figure 3.26e**).

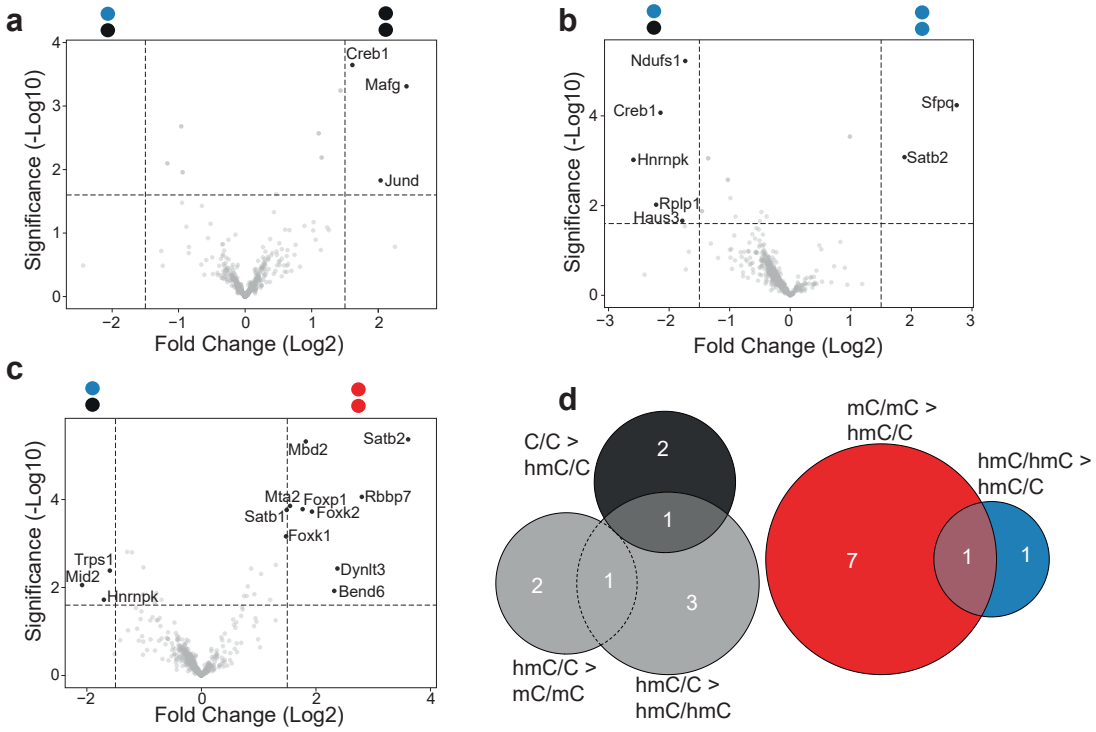


Figure 3.26: Readers and antireaders of hmC/C from mouse brain. (a) Volcano plot comparing hmC/C (blue and black circles) with C/C (black circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5). **(b)** Volcano plot comparing hmC/C (blue and black circles) with hmC/hmC (blue circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5). **(c)** Volcano plot comparing hmC/C (blue and black circles) with mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5). See Table A.36, Table A.37, Table A.38 for full protein lists. **(d)** Readers and antireaders of hmC/C (gray) compared with C/C (black), mC/mC (red) and hmC/hmC (blue) visualized in a Venn diagram. Corresponding protein names are listed in Table A.39.

Readers and antireaders of hmC/mC from mouse brain. Only a little number of proteins was found to be enriched for hmC/mC (**Figure 3.27**). The only proteins enriched over C/C were Satb2 (previous mC/mC and hmC/hmC reader) and Lrrfip2 (previous hmC/hmC reader), indicating a preference of these proteins for methylated and hydroxymethylated probes (**Figure 3.27a**). C/C enriched and hmC/mC depleted proteins included the aforementioned C/C readers Creb1, Zbtb2, Parp1, Trps1, Clock, Jund, Mafg and Arntl, as well as Mapk8ip3, a cytoplasmic brain enriched putative adapter protein believed to link cargo to dynein and kinesin motors (Gowrishankar et al., 2021).

hmC/mC enriched solely Fn3k over mC/mC (**Figure 3.27b**). Fn3k is an enzyme found in the cytoplasm which removes early glycation adducts to mitigate the effects of excessive glycation and restore protein function (Szwergold et al., 2001). mC/mC enriched the three Fox proteins, Creb1, Jund and Satb1, but not Satb2, over hmC/mC.

No protein was significantly enriched by hmC/mC over hmC/hmC, while hmC/hmC enriched Sfpq and Uhrf2 over hmC/mC (**Figure 3.27c**).

hmC/mC enriched Satb2 and Rbbp7 (previous mC/mC reader) over hmC/C (**Figure 3.27d**), while hmC/C enriched the C/C readers Creb1 and Trps1 and also Jup, also known as Plakoglobin, which is involved in cell-cell adhesion and cell signaling by interacting with signaling proteins and transcription factors Zhurinsky et al., 2000 and is further associated with tumor suppressor activity Simcha et al., 1996.

In conclusion, no protein was found to be enriched for hmC/mC over all other modified probes,

however, Satb2 was enriched over C/C and hmC/C (overlap of white circles in **Figure 3.27e**). Creb1 was an antireader of hmC/mC, which was enriched with C/C, hmC/C and mC/mC but not hmC/hmC indicating that its strongly repulsed by hmC and can only bind when hmC occurs together with C in the dyad (overlap of black, red and gray circles in **Figure 3.27e**). Also Jund was an hmC/mC antireader enriched with C/C and mC/mC, which was already found to be depleted by hmC/hmC and one mC in the dyad could not rescue its affinity (overlap of black and red circles in **Figure 3.27e**). Trps1 was enriched with C/C and hmC/C over hmC/mC (overlap black and gray circles in **Figure 3.27e**).

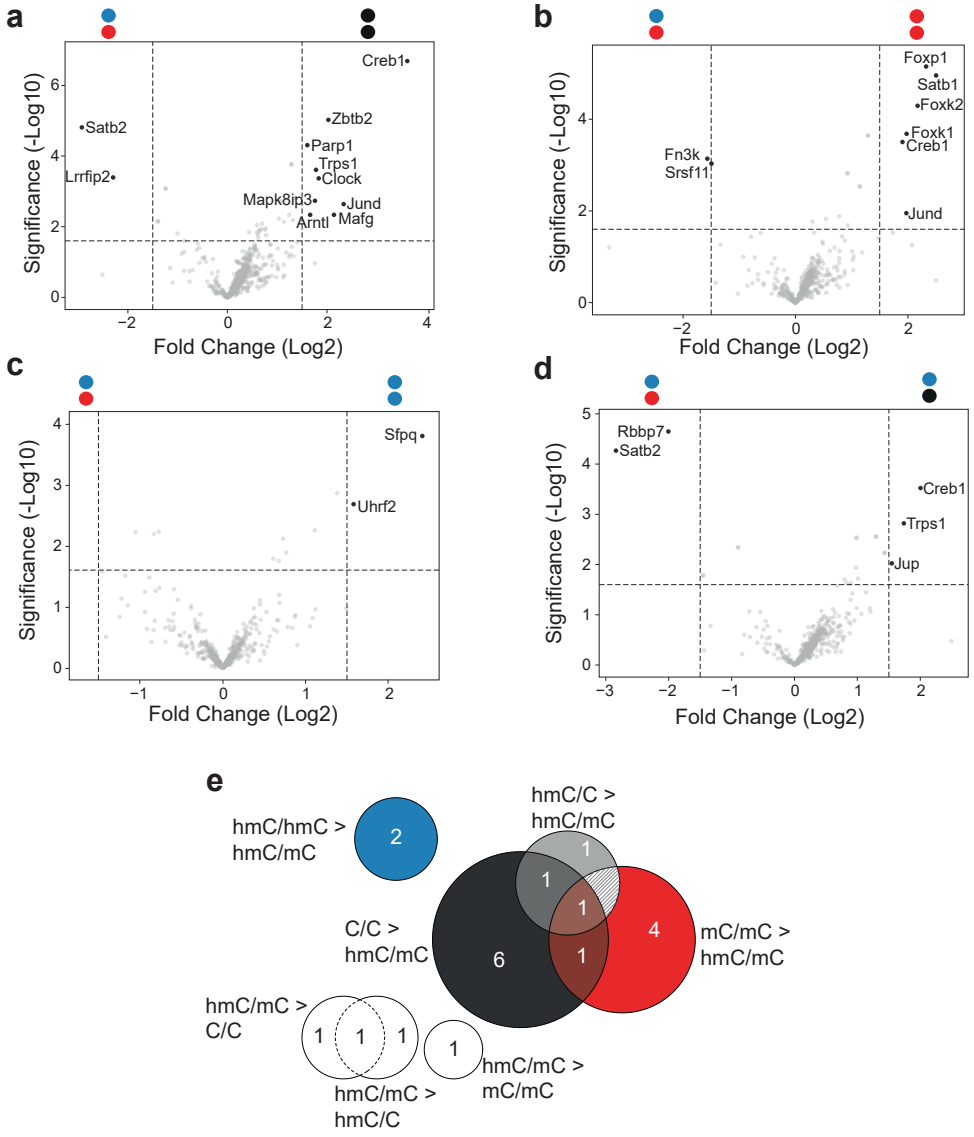


Figure 3.27: Readers and antireaders of hmC/mC from mouse brain. (a) Volcano plots comparing enriched proteins (p-value < 0.025 and log2 fold change >1.5) between hmC/mC (blue and red circles) and C/C (black circles). **(b)** Volcano plots comparing enriched proteins (p-value < 0.025 and log2 fold change >1.5) between hmC/mC (blue and red circles) and mC/mC (red circles). **(c)** Volcano plots comparing enriched proteins (p-value < 0.025 and log2 fold change >1.5) between hmC/mC (blue and red circles) and hmC/hmC (blue circles). **(d)** Volcano plots comparing enriched proteins (p-value < 0.025 and log2 fold change >1.5) between hmC/mC (blue and red circles) and hmC/C (blue and black circles). See Table A.40, Table A.41, Table A.42, Table A.43 for full protein lists. **(e)** Readers and antireaders of hmC/mC (white) compared with C/C (black), mC/mC (red), hmC/hmC (blue) and hmC/C (gray) visualized in a Venn diagram. Corresponding protein names are listed in Table A.44.

Overall comparison and summary of mouse brain reader proteins. The overall number of probe specific binders was quite low in mouse brain experiments, which became already obvious in the data

quality assessment as bound proteins did not fully cluster according to their technical replicates within a modified probe (**Subsection 3.4.2**). Even after determining significantly bound proteins, the clustering was not perfect as can be seen in the hierarchical clustering in **Figure 3.28a** with one outlier for hmC/C and hmC/mC each. Unlike the HEK293T enrichments, where hmC/hmC bound proteins clustered furthest from other dyads and mC/mC and hmC/mC bound proteins clustered closely together, the mouse enrichments showed the opposite pattern with mC/mC bound proteins clustering furthest, while hmC/mC and hmC/hmC bound proteins were closer. Additionally, hmC/C clustered more closely with hmC/mC and hmC/hmC than with C/C, which is another distinction from the HEK293T results suggesting that hmC-dyad bound proteins cluster more tightly in mouse experiments.

The overlaps of readers between the modified probes were small and individual readers for each modified probe emerged (**Figure 3.28b**). The strongest overlap was seen for C/C and hmC/C, however, only three proteins fell into this category, which are Mid2, Hnrnpk and Trps1. Some mC/mC enriched proteins overlapped with readers from other modified probes, such as Creb1, which was enriched with mC/mC, C/C and hmC/C, indicating a overall selectivity of C/C > mC/mC and hmC/C > hmC/hmC and hmC/mC, while Satb2 was enriched with mC/mC, hmC/hmC and hmC/mC with an indicated preference of mC/mC > hmC/hmC > C/C and hmC/C as well as hmC/mC > hmC/C and C/C. Jund was enriched with C/C and mC/mC, Rplp1 with hmC/C and mC/mC and Rbbp7 with hmC/mC and mC/mC. Another overlap was seen for hmC/hmC and hmC/mC with Lrrfip2. All other enriched proteins were specific to one of the modifications, among them were Sfpq and Uhrf2 for hmC/hmC, Jup, Nduf2 and Haus3 for hmC/C and Fn3k for hmC/mC. Notably, Sfpq and Uhrf2 were consistently enriched for hmC/hmC over all other modified probes indicating a strong preference of these proteins for symmetrically modified hmC probes.

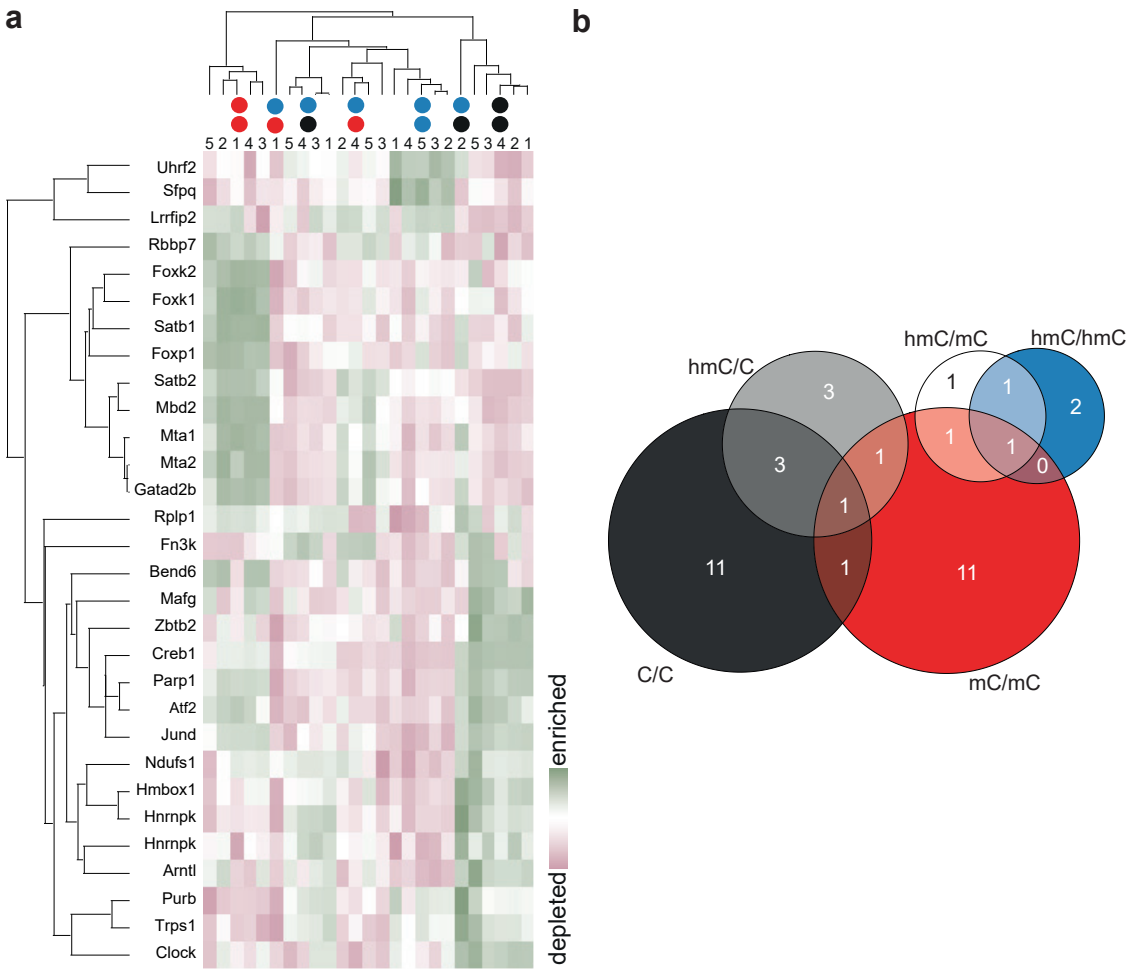


Figure 3.28: Overall comparison of mouse brain reader proteins. (a) Heatmap of correlation-based clustering of the LFQ-intensities after log2 transformation and normalization by row mean subtraction. Proteins included in the clustering significantly bind to at least one of the probes as determined by an ANOVA test (p-value < 0.025 and s0=1). **b** Venn diagram for significantly enriched proteins for indicated probes, proteins are listed in Table A.45.

3.4.3 Essential points

- Pull-down MS studies were conducted using modified promoter DNA probes and nuclear extracts from human HEK293T cells and mouse brain.
- Western blotting confirmed the experiments were functional, while enrichment and clustering in HEK293T experiments was clearer compared to mouse brain. C/C and mC/mC modified probes produced the expected reader profiles.
- hmC-containing dyads generally showed fewer protein interactions than C/C and mC/mC modified probes, though hmC/C in HEK293T experiments was an exception. The enrichment patterns for hmC-containing dyads were not always consistent across all pairwise comparisons. However, while several proteins like RFX5, UHRF1 and MECP2 displayed multiple preferences, affinities could still be predicted. Thus, RFX5 appeared to prefer mC/mC > hmC/mC > C/C and hmC/C > hmC/hmC, UHRF1 showed a preference of hmC/hmC > hmC/mC and hmC/C > C/C as well as a preference of mC/mC > hmC/C > C/C, and MECP2s preferences were mC/mC > hmC/C > C/C as well as hmC/hmC and

hmC/mC and hmC/C > C/C. Satb2 was enriched in the mouse brain pull-down experiments suggesting a preference of mC/mC > hmC/hmC > C/C and hmC/C as well as hmC/mC > hmC/C and C/C.

- Notable hmC/hmC readers included HNRNPLL (in HEK293T experiments) and Uhrf2 and Sfpq (in mouse brain experiments), while antireaders mainly comprised RFX and PRC1.6 complex proteins (in HEK293T experiments) and Jund and Creb1 (in mouse brain experiments).
- hmC/C probes shared many readers with C/C but uniquely enriched RBM45 and CDCA7L, and Hnrnpk in mouse brain experiments, while FOXC1 and DACH1 were identified as antireaders.
- Although no universal hmC/mC readers were identified, several proteins such as NOP16, ZC3HAV1, BAG2, RFX proteins and Satb2 were enriched over multiple modified probes. ATF1, LDB1, and in mouse brain experiments Creb1, Jund, and Trps1 were considered as antireaders, which were particularly sensitive to hmC in general.
- Overall, the study provides valuable framework for further exploration of protein interactions with modified (a)symmetric probes.

The mouse brain used in this chapter was kindly gifted from Dr. Cristina Cadenas from Leibniz-Institut für Arbeitsforschung an der TU Dortmund (IfADo). I gratefully acknowledge their contributions.

3.5 Evaluation of potential reader proteins

The proteomics data provided insights into protein binding preferences towards different modified probes. However, proteins identified in such studies may be components of larger protein complexes or subject to PTMs and thus may not directly interact with DNA. As a result, they may not qualify as *bona fide* C modification reader proteins. To directly assess DNA binding behavior and modification specificity, selected candidate proteins were expressed and purified for use in DNA binding assays. CDSs) were cloned from cDNA libraries and expressed as MBP fusion proteins which allowed for their consistent detection using a MBP-antibody in dot blot and Western blot assays. Initially, the MBP fusion proteins were overexpressed in *Escherichia coli (E. coli)* and the resulting lysates were used in pull-down assays to enable a broad spectrum screening of DNA interactions. Proteins that exhibited promising or unusual binding behavior in these assays were subsequently purified for a more detailed biochemical characterization via EMSA. Among the tested candidates, the three proteins, MYC, MAX and RFX5 showed particularly intriguing interaction patterns. These proteins were re-cloned to further investigate their DNA binding and potential modification activities in greater detail.

3.5.1 Cloning and expression of potential hits

Coding sequences of the candidate proteins were amplified from cDNA libraries and cloned into an expression vector containing an N-terminal MBP tag (Figure A.9a). The following proteins were cloned as MBP fusions and expressed in *E. coli*: MBD2, ATF1, ATF2, CHAF1B, FOXA1, L3MBTL3, MAX, MYC, RFX5, RFXANK, SUB1 and TFAM. Expression was assessed by SDS-PAGE analysis of *E. coli* lysates, which confirmed successful overexpression of the fusion proteins in most cases (**Figure 3.29a**, **Figure A.9b**). Notably, while most proteins showed strong bands at the expected size, L3MBTL3 and RFX5 did not produce a prominent band on the SDS-PAGE gel. However, subsequent Western blot analysis using an anti-MBP antibody confirmed the identity of all expressed proteins, despite lower apparent abundance of L3MBTL3 and RFX5 (**Figure 3.29b**, **Figure A.10a**). Not all lysates displayed a single distinct band in the blot, suggesting the presence of proteolytic cleavage products or prematurely terminated translations, leading to truncated fragments. Since the MBP tag is fused N-terminally, a second Western blot was performed using an anti-His antibody targeting the C-terminal His-tag, which resulted in fewer bands but also showed some smearing (**Figure A.10b, c**). This may be attributed to low quality or affinity of the His-tag antibody. Despite the presence of multiple bands in both blots, the detection of a band at the expected size and confirmation of the presence of the MBP tag by Western blot support the conclusion that the fusion proteins are correctly expressed, allowing to proceed with initial DNA-binding assays.

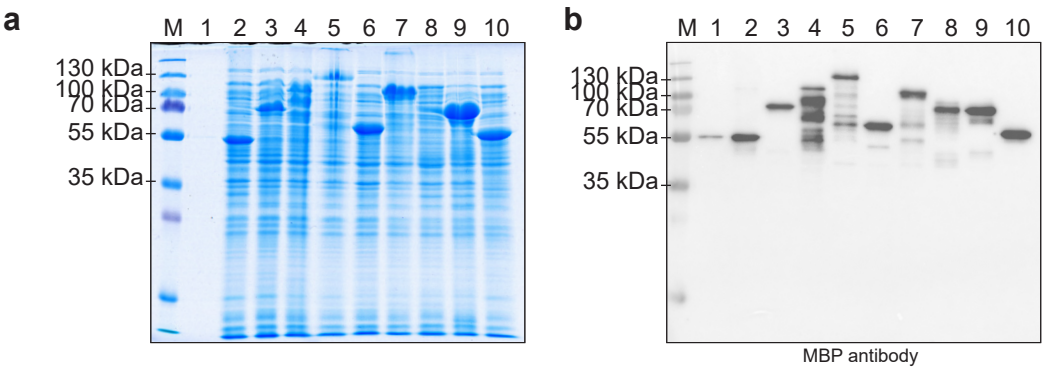


Figure 3.29: SDS-PAGE and Western blot analysis of *E. coli* lysates containing MBP-tagged candidate proteins. **(a)** Whole-cell lysates from *E. coli* cultures induced to overexpress the MBP-tagged candidate proteins were analyzed by SDS-PAGE. Lane 1: purified MBP-MBD2 (54.0 kDa), lane 2: MBP-MBD2 lysate (54.0 kDa), lane 3: MBP-ATF1 lysate (74.4 kDa), lane 4: MBP-CHAF1B lysate (106.7 kDa), lane 5: MBP-L3MBTL3 lysate (133.6 kDa), lane 6: MBP-MAX lysate (63.5 kDa), lane 7: MBP-MYC lysate (94.0 kDa), lane 8: MBP-RFX5 lysate (110.5 kDa), lane 9: MBP-RFXANK lysate (73.3 kDa) and lane 10: MBP-SUB1 lysate (59.6 kDa). Samples were run on a 12 % SDS-PAGE with a Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder (lane M) **(b)** Western Blot using a MBP antibody showing the presence of MBP-tagged candidate proteins. Numbers and molecular weights are corresponding to (a).

3.5.2 DNA-binding assays of candidate proteins

For initial assessment of the binding behavior of candidate proteins, the lysates containing the overexpressed MBP-fusion proteins were employed in pull-down assays to test whether the proteins show the same modification preference trends in this simplified system as observed in the MS studies. Therefore, the standard pull-down assay was performed using the five differentially modified VEGFA probes and protein enrichment was visualized on Dot blots using anti-MBP antibodies. Many proteins did not show modification-specific binding in the Dot blot assay or displayed patterns that were conflicting with MS results (**Figure 3.30a**, Figure A.11). For example, CHAF1B, which was enriched for hmC/mC and depleted for C/C in MS, showed no clear enrichment for hmC/mC and slight depletion for hmC/C and hmC/hmC in Dot blot and Western blot analyses. CHAF1B is regulated by phosphorylation (S. Smith and Stillman, 1991) and it undergoes various protein-protein interactions such as with CHAF1A and PCNA (Martini et al., 1998), which could be factors explaining the lack of modification specificity in Dot blots. Similarly, ATF2, which demonstrated mC/mC affinity and was depleted by C/C and hmC/mC in MS, did not reproduce these preferences in Dot blot analyses. While it showed some enrichment for mC/mC, comparable enrichment was observed for hmC/mC. C/C even enriched ATF2 more than hmC/C and hmC/hmC. ATF2 is another transcription factor dependent on phosphorylation (Bhoumik et al., 2005). L3MBTL3, which showed C/C and hmC/C preference in MS, demonstrated DNA binding without clear modification-specific preferences in the Dot blot. While it can directly bind to DNA via its MBT domains (Nady et al., 2012), the lack of modification preference might be attributed to its involvement in various protein-protein interactions such as with SAMD1, which is suspected to tether L3MBTL3 to its binding sites (Stielow et al., 2021). Consequently, L3MBTL3 may not exhibit distinct modification specificities in isolation. TFAM, which was enriched for C/C and hmC/C in MS studies, did not demonstrate any modification-specific binding in Dot blots. TFAM itself does not have the potential to interact directly with the modified cytosines as it binds in the minor and not the major groove. It was suggested that its binding is however influenced by C modifications due to altered stacking and deformability of the DNA, which subsequently impacts the relative stability of the TFAM-DNA complex (Dostal and Churchill, 2019). Hence, other factors need to be present to observe DNA-binding by TFAM in a modification-specific manner. Likewise, SUB1, which consistently exhibited mC/mC

preference in this and previous MS studies (e.g. L. Bai et al., 2021; Iurlaro et al., 2013), did not confirm this preference under the simplified Dot blot conditions. Given that SUB1 is strongly regulated by phosphorylation (Mustafi et al., 2022), its modification-binding behavior might be affected within *E. coli* lysates, potentially explaining the lack of observed preference in the Dot blot assays. RFXANK only weakly reproduced its MS-detected mC/mC and hmC/mC affinities. RFXANK might not exhibit strong modification preferences when isolated as it is known to bind DNA in complex with RFX5 and RFXAP, with RFX5 binding directly to DNA (Steimle et al., 1995). Nevertheless, a trend consistent with MS findings was still visible, supporting its indirect role in modification-specific binding.

In contrast, some candidate proteins did show modification-specific binding under simplified conditions. Among these were MAX and MYC, which are known C/C binders and also showed preferential hmC/C binding in MS. Strikingly, both proteins additionally showed binding to hmC/hmC in Dot blot assays (**Figure 3.30b**). Furthermore, RFX5 and FOXA1 confirmed their mC/mC and hmC/mC preferences observed in MS across multiple dot blot replicates (**Figure 3.30c**). Subjecting the same samples to Western blots corroborated these preferences and revealed multiple bands under conditions of high affinity binding (**Figure 3.30d**). This suggests that shorter fragments of these proteins may retain the ability to bind to their high affinity targets. The Western blots further showed subtle differences in binding affinities from Dot blot assays, such as a stronger mC/mC over hmC/mC preference for RFX5, prompting more in depth studies described in **Subsection 3.5.3**.

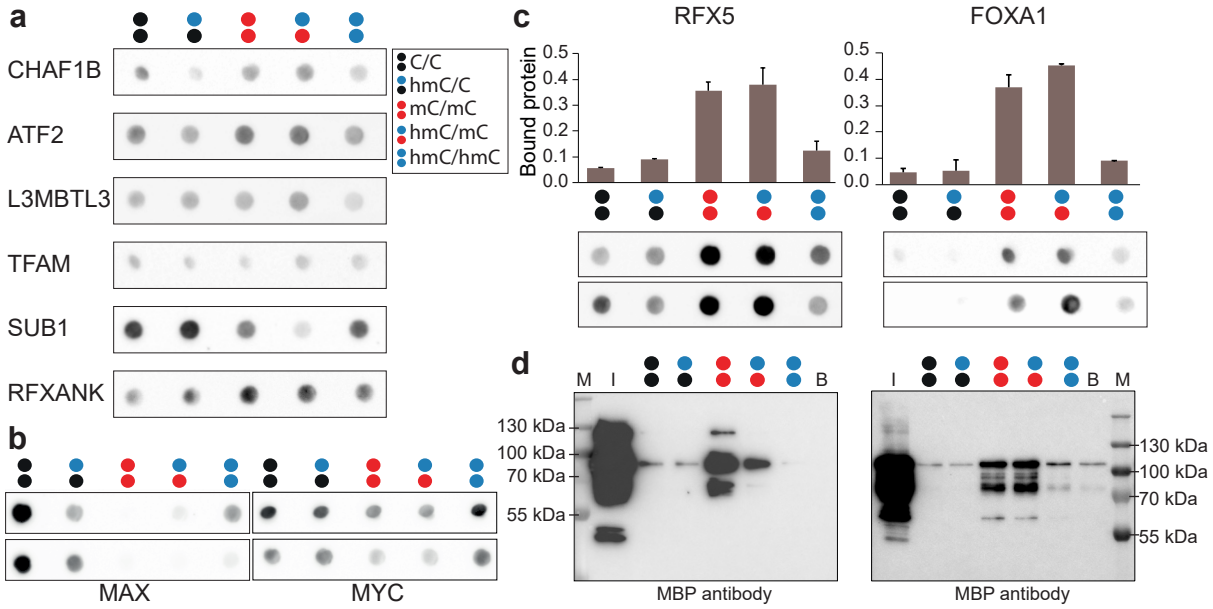


Figure 3.30: Dot blot analyses of modification preferences. (a) Anti-MBP dot blots of enrichments using VEGFA probes and *E. coli* expression extracts of indicated proteins fused to an N-terminal MBP tag. **(b)** Anti-MBP dot blots of enrichments using VEGFA probes and *E. coli* expression extracts of MBP-MAX and MBP-MYC. **(c)** Anti-MBP dot blots of enrichments using VEGFA probes and *E. coli* expression extracts of MBP-RFX5 and MBP-FOXA1 displaying the relative enrichments as bar diagrams. **(d)** Western blot analyses of enrichment samples from (c) using a anti-MBP antibody. Expected sizes are 110.5 kDa for MBP-RFX5 and 94.4 kDa for MBP-FOXA1. Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder was run in lane M, input lysate in lane I and a bead-only control in lane B.

To further investigate DNA binding behavior, selected candidates were purified via nickel affinity chromatography and subsequently examined in EMSA assays, an orthogonal method to pull-down assays allowing direct observation of protein-DNA interactions. The candidate purification process was monitored by SDS-PAGE (Figure A.12), indicating relatively pure elution fractions. For comparison, purified proteins were applied to an SDS-PAGE (**Figure 3.31a**). Only MBP-RFXANK and MBP-SUB1

exhibited clear and distinct bands, while the remaining proteins displayed weak signals or contained impurities. Subsequent EMSA experiments using varying protein concentrations and C/C and mC/mC modified VEGFA probes, revealed that none of the proteins bound to DNA, with an exception of MBP-SUB1, which bound at higher concentrations (**Figure 3.31b**). However, MBP-SUB1 did not distinguish between C/C and mC/mC VEGFA probes. One possible explanation for the lack of binding specificity may be the presence of the MBP tag, which significantly increases the molecular weight of the protein and could interfere with DNA binding or promote protein aggregation, especially when high protein concentrations are used. To circumvent this, the MBP tag was removed post purification using TEV protease, which cleaves between the MBP tag and the protein of interest. The TEV digest results were evaluated via SDS-PAGE (**Figure 3.31c**). RFX5 digest was successful, indicated by the appearance of a 45 kDa MBP band. RFXANK was partially digested with a decrease of the MBP-RFXANK band and a new band appearing around 30 kDa. SUB1 was also successfully suggested. However, TEV digest failed for FOXA1 and ATF2. Notably, the input fractions of the unsuccessful digests were already of low quality with weak protein bands. Using the TEV digested proteins in EMSA assays with C/C and mC/mC modified VEGFA probes indicated that all proteins, including those with apparently failed TEV digests, now bound DNA (**Figure 3.31d**). Despite this, none of the proteins demonstrated specificity for either C/C and mC/mC, including RFX5 and FOXA1 which had previously shown mC/mC > C/C preferences in the dot blot assays. Moreover, the EMSA gels lacked clear protein-DNA complexes suggesting possible aggregation. Repeating the EMSA using all five modified VEGFA probes also failed to reveal specific binding patterns (**Figure 3.31e**). The only exception was SUB1, which however, did not replicate its initial mC/mC preference. These results are consistent with the findings from the pull-down dot blot assays, where RFXANK, SUB1, CHAF1B and ATF2 also did not show modification preferences, likely due to the lack of natural human protein context and possible interference from the MBP tag. Surprisingly, RFX5 and FOXA1, which had shown modification-specific binding in both MS and dot blot assays, also failed to display preferences in EMSA. As FOXA1 purification was not as successful, the focus was put on RFX5 to further investigate its modification preference, as discussed in the next section.

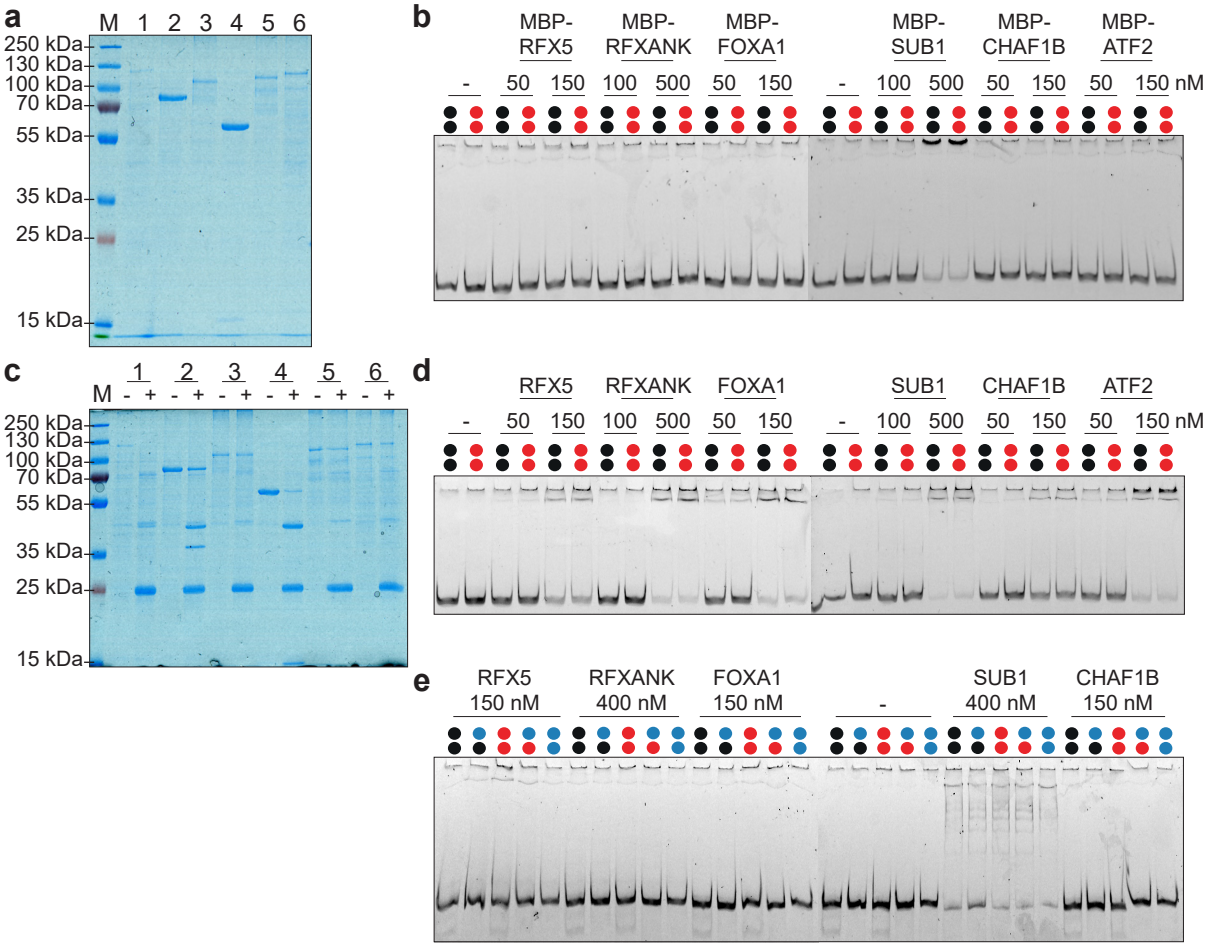


Figure 3.31: EMSA assays of candidate proteins. **(a)** SDS-PAGE displaying the purified candidate proteins: (1) MBP-RFX5 (110.5 kDa), (2) MBP-RFXANK (73.7 kDa), (3) MBP-FOXA1 (94.4 kDa), (4) MBP-SUB1 (59.6 kDa), (5) MBP-CHAF1B (106.7 kDa) and (6) MBP-ATF2 (99.8 kDa). Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder was run in lane M. **(b)** EMSA analyses of candidate proteins in varying concentrations binding to 2 nM C/C (black circles) or mC/mC (red circles) VEGFA probe. **(c)** SDS-PAGE displaying candidate proteins before (-) and after (+) digest with TEV protease. Numbers correspond to (a). TEV protease has a MW of 27 kDa, MBP of 42.5 kDa. **(d)** EMSA analyses of TEV-digested candidate proteins in varying concentrations binding to 2 nM C/C (black circles) or mC/mC (red circles) VEGFA probe. **(e)** EMSA analyses of TEV-digested candidate proteins binding to 2 nM VEGFA probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles).

3.5.3 RFX5 reads symmetrical mC/mC and asymmetrical hmC/mC

The full-length RFX5 (hereafter referred to as RFX5$_{fl}$ did not show modification-specific DNA binding in the previous EMSAs. Earlier studies also indicated that RFX5(1–330) failed to bind to its canonical X-box sequence, despite the fact that the isolated RFX DNA-binding domain (DBD) showed relatively high affinity. This suggested that DBD flanking domains may inhibit DNA binding in vitro (Garvie and Boss, 2008). Interestingly, RFX5$_{fl}$ modification-specific DNA binding was observed in the dot blot assays. However, the presence of bacterial lysate proteins may have facilitated this binding either by direct interaction or competition. For more detailed examination, constructs encoding only the DBD of RFX5 were cloned. RFX5$_{fl}$ comprises five domains, which are an oligomerization domain (OD), a DBD, a dimerization domain, a proline-rich region (P) and a transactivation domain (Garvie et al., 2007) (**Figure 3.32a**). The OD mediates dimerization and is important for interactions with RFXANK and RFXAP. The DBD, containing a helix-turn-helix motif, is responsible for DNA binding. The dimerization domain forms a helical structure and possibly also interacts with RFXANK and RFXAP. The P region

and the transactivation domain contribute to MHCII gene activation and recruit additional proteins (DeSandro et al., 2000) Two constructs encoding the RFX5 DBD (amino acids 75-168, including a few residues from the OD as in prior studies (Chakraborty, Bhattacharya, et al., 2010), **Figure 3.32a**) were cloned. The first one was native $RFX5_{DBD}$ and the second one MBP-$RFX5_{DBD}$. Purification of both constructs was successful (**Figure 3.32b**). The MBP-tagged version was subsequently cleaved using TEV protease, resulting in removal of MBP (referred to as $RFX5_{DBD(TEV)}$ (**Figure 3.32c**). Notably, two protein bands were present before TEV digestion and both of them were cleaved, yielding two bands after TEV digestion. A secondary band was also visible in the native $RFX5_{DBD}$.

To assess whether RFX5's modification preferences are context-dependent, the binding studies were extended beyond the VEGFA probe. Therefore, the X-box sequence, RFX5's physiological target (Mach et al., 1996) was employed. Two X-box sequences were chosen for this purpose. X-box 1 was identical to the sequence used in Spruijt et al., 2013 and X-box 2 was a slightly modified version from the physiological site of the HBV enhancer known to be bound by RFX1 (Gajiwala et al., 2000). EMSAs were performed using either $RFX5_{DBD}$ or $RFX5_{DBD(TEV)}$ incubated with either the VEGFA or the X-box probes that contained modified CpG dyads (**Figure 3.32d**). Interestingly, the VEGFA probe appeared to contain multiple binding sites or prompted RFX5 oligomerization although the oligomerization domain was absent in these constructs. This resulted in multiple protein-DNA bands with $RFX5_{DBD(TEV)}$ showing more complexes than $RFX5_{DBD}$. Notably, $RFX5_{DBD(TEV)}$ exhibited higher binding affinity and showed more bands for both mC/mC and hmC/mC modified probes, indicating enhanced interactions with this construct. Binding to X-box 1 afforded higher protein concentrations but yielded the expected modification preferences with a distinct binding event as indicated by a single protein-DNA band, suggesting a more specific and limited interaction at the canonical site. Lower protein concentrations were sufficient to detect binding to X-box 2. Higher affinity binding yielded two bands, suggesting a possible secondary binding mode. Nevertheless, the modification specificity was maintained. mC/mC binding was consistently favored across all three probes, while hmC/mC was the second most favored modification, although the interaction with $RFX5_{DBD(TEV)}$ was notably stronger than with $RFX5_{DBD}$. To quantify modification preferences, multiple replicates were performed, confirming the preference of mC/mC > hmC/mC > all other modified probes in both X-box sequence contexts (**Figure 3.32e**).

These results and the clear preference of RFX5 for mC/mC and hmC/mC prompted the quantification of $RFX5_{DBD(TEV)}$ binding affinity. EMSAs were performed with increasing protein concentrations and bound DNA fractions were quantified to calculate dissociation constants (K_D) (**Figure 3.32f, Figure A.13a, b**). For X-box 1 the determined K_D values were 207 ±16.4 nM for mC/mC and 572.5 ±115.3 nM for hmC/mC. For X-box 2 binding affinities were stronger with 48.25 ±6.18 nM for mC/mC and 106.49 nM ±15.88 nM for hmC/mC, suggesting that mC/mC is bound with approximately 2-fold higher affinity than hmC/mC in both sequence contexts. Moreover, the marked difference in K_D values between the two X-box probes indicates that $RFX5_{DBD(TEV)}$'s binding affinity is highly context specific. Previous studies have reported a K_D of of 68 ±12 nM for RFX5(88–170) binding to the canonical X1-box, (Garvie and Boss, 2008). However this X-box variant did not contain a CpG site or a modified cytosine. Another study determined dissociation constants of 224 nM and 335 nM for RFX5(76-158) binding to unmodified X-box DNA without CpGs, suggesting that two molecules of RFX5 DBD interact with one molecule of X-box DNA (Chakraborty, Sengupta, et al., 2010). Given the different X-box sequence context and the use of unmodified DNA in earlier studies, direct comparison to the present results are limited. Nonetheless, for the physiologically derived X-box 2 probe, the apparent K_D for mC/mC was lower than previously reported values for unmodified DNA. In this study the K_D for unmodified DNA is estimated to be larger than the one determined for hmC/mC, which could be in the same range with the second mentioned study.

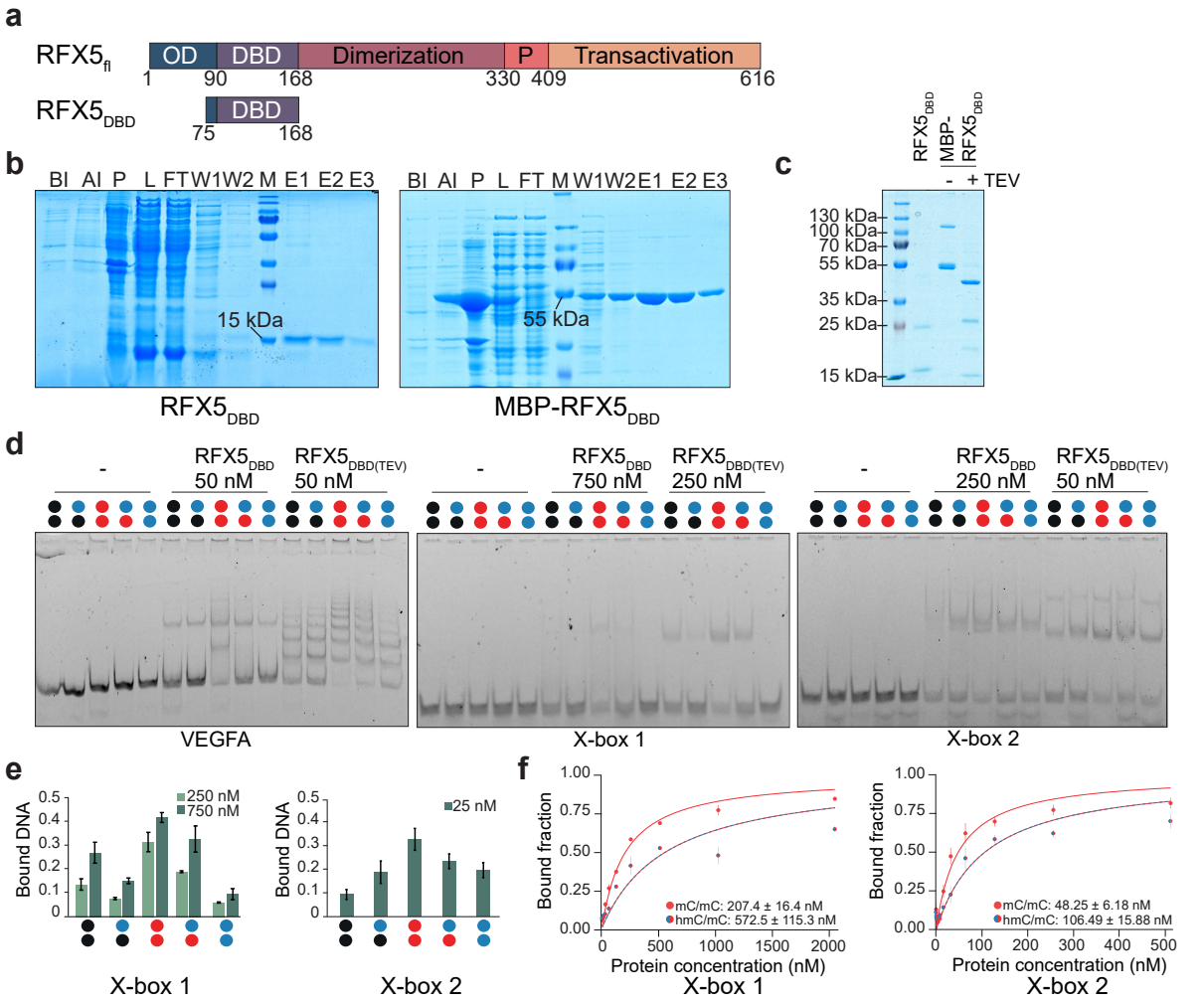


Figure 3.32: Evaluating RFX5 as a mC/mC and hmC/mC reader. (a) Schematic representation of the domain organization of $RFX5_{fl}$ and $RFX5_{DBD}$ constructs. OD: oligomerization domain, DBD: DNA-binding domain, P: proline-rich region. $_{fl}$ and $RFX5_{DBD}$. **(b)** Purification steps of $RFX5_{DBD}$ and MBP-$RFX5_{DBD}$ visualized on a SDS-PAGE gel. Samples were taken from BI (before induction), AI (after induction), P (pellet), L (lysate), FT (flow-through) W (wash), E (elution), Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder was run in lane M. Theoretical MWs are 13.4 kDa for $RFX5_{DBD}$ and 56.2 kDa for MBP-$RFX5_{DBD}$. **(c)** SDS-PAGE gel of MBP-$RFX5_{DBD}$ before and after TEV digest. Theoretical MW of TEV-digested MBP-$RFX5_{DBD}$ is 12.0 kDa. Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder was run in lane M and $RFX5_{DBD}$ was run as reference. **(d)** EMSA analyses of $RFX5_{DBD}$ and $RFX5_{DBD(TEV)}$ binding to 2 nM VEGFA probe, X-box 1 probe or X-box 2 probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles). **(e)** EMSA profiles of $RFX5_{DBD(TEV)}$ binding to modified X-box 1 and X-box 2 probes. Error bars from duplicate experiments. **(f)** EMSA titration experiments and K_D for $RFX5_{DBD(TEV)}$ and mC/mC and hmC/mC modified X-Box 1 and 2 probes.

To conclude, RFX5 plays a critical role in immunity by regulating the transcription of MHC class I and II genes, with mutations in RFX5 leading to disease (Ludigs et al., 2015). It functions as part of the trimeric RFX complex together with RFXANK and RFXAP (Masternak et al., 1998). Earlier studies have indicated that all three components are required for high affinity binding to the X1-box DNA motif (Caretti et al., 2000; DeSandro et al., 2000). Despite this, previous studies reported high affinity DNA binding with RFX5 in isolation (Chakraborty, Sengupta, et al., 2010; Garvie and Boss, 2008), and the present study further demonstrated that the RFX5 DBD alone can bind to symmetrically and asymmetrically modified DNA probes with high affinity. This suggests some level of modular binding activity. Strikingly, recognition of asymmetric dyads by RFX5 has never been reported. Previous reports suggested a mC preference of RFX5, in the context of transcriptional activation of methylated promoter DNA (Niesen et al., 2005) and in pull-down MS studies (Spruijt et al., 2013). The latter attributed this selectivity to a proposed mC binding pocket which can not accommodate hmC, as suggested by homology models based on a DNA-complex of RFX1. Although hmC/hmC almost completely abolished $RFX5_{DBD(TEV)}$ binding, measurable affinity for hmC/mC remained. This supports a model in which mC-selective interaction occurs via a single DNA strand (which would align with the observed 2-fold lower hmC/mC affinity). However, high-affinity binding to the hmC/mC probes contradicts the aforementioned pocket model and these modification preferences were preserved in different probe contexts, indicating that $RFX5_{DBD(TEV)}$ does not simply interact with one mC nucleobase, but indeed reads out the symmetry of hmC CpG marks. The observed selectivity is unlikely to result from oligomerization, since the OD and dimerization domain were absent (Garvie et al., 2007). This data reveals a more nuanced readout of CpG dyads by RFX5 than previously expected by data derived from symmetric modifications. This emphasizes the potential of individual hmC CpG symmetries to cause distinct regulatory outcomes. The biological implication of this read out by RFX5 appears significant, indicating that oxidation of one mC still permits sufficient RFX5 binding to maintain a potential repressive chromatin state.

3.5.4 MYC and MAX read hmC/hmC in probe-specific contexts

MYC and MAX are basic-helix–loop–helix (bHLH) transcription factors that function as dimerization partners (Blackwood and Eisenman, 1991). Together, they bind to unmodified DNA, particularly at E-box motifs, to activate transcription (Kretzner et al., 1992; McMahon et al., 2000). Dot blot assays from section (**Subsection 3.5.2**) using MBP-MAX and MBP-MYC in bacterial lysates revealed expected binding to C/C. Surprisingly, both proteins also showed significant binding to hmC/hmC (**Figure 3.30b** and **Figure A.14**), despite MS data indicated enrichment primarily for C/C > mC/mC, hmC/mC, and hmC/hmC, as well as enrichment for hmC/C > mC/mC and hmC/mC. Only in one experiment MAX was also enriched with hmC/C > hmC/hmC and in experiment 3 MYC and MAX were enriched for hmC/mC > mC/mC (**Subsection 3.4.1**). The expected hmC/C preference was observed in dot blots, further strengthening the indication that MYC and MAX binding is not abolished with hmC on one strand. Interestingly, even fully hmC/hmC modified probes did not abolish their binding, but instead MYC and MAX were recruited by this modification. This was unexpected and suggested that binding affinity is not entirely lost with hmC modifications, which was previously unreported. However, results from the dot blot assays were not fully consistent (**Figure A.11a**). To address potential variability, NE was added to the pull-down dot blot assays to ensure that MBP fused MYC and MAX had access to potential dimerization partners. The previously observed general binding trend emerged, indicating C/C and hmC/C as the preferred binding sites and hmC/hmC ranking next. This effect was more pronounced for MBP-MYC than for MBP-MAX (**Figure 3.33a**). To obtain more robust data, EMSA assays were employed using purified proteins. MAX is capable to bind DNA as as homodimer, although with

lower affinity than the MYC-MAX heterodimer. MYC, on the other hand does not efficiently homodimerize under physiological conditions (Blackwood et al., 1992; Blackwood and Eisenman, 1991). Therefore, MBP-MAX was used in EMSA assays, which confirmed the observed dot blot binding preferences (**Figure 3.33b**). Notably, hmC/hmC binding became more pronounced at higher protein concentrations and hmC/C was bound with almost the same affinity as C/C, showing even stronger binding than observed in the dot blot assays. As the hmC/hmC preference was only observed in the VEGFA probe context, it should be considered that Cs were not only modified in the CpG but also in CpH contexts. To further investigate binding to the VEGFA probe, it was divided into seven fragments. Binding of MBP-MAX to each split fragment was assessed. The canonical E-box sequence (CACGTG) (Kato et al., 1992) is absent from the VEGFA probe but E-box-like motifs are present (**Table 3.4**). Among the seven fragments, splits 5 and 6, which contained E-box-like motifs and elevated number of CpA sites, exhibited the strongest binding affinity (**Figure 3.33c**). Interestingly, split 2 also showed elevated binding, despite lacking an E-box-like motif.

Table 3.4: Properties of VEGFA split probes.

Probe	CpG	CpA	E-box-like motif
Split 1	1	3	NA
Split 2	3	5	NA
Split 3	2	6	NA
Split 4	3	5	NA
Split 5	2	9	GACGTG
Split 6	2	9	TGCGTG
Split 7	2	4	NA

To assess the influence of modified VEGFA sequences on MAX binding, the MBP tag was removed by TEV digestion and MAX was subsequently used in a competition EMSA with split 5 and increasing concentrations of modified full-length VEGFA probes. Unmodified full-length VEGFA competitor completely displaced split 5 binding at 10 nM, while the mC/mC competitor had other effect at this concentration. The hmC/hmC competitor did interfere with binding, albeit to a lesser extent than unmodified VEGFA competitor (**Figure 3.33c**). This reinforced the C/C > hmC/hmC > mC/mC preference observed in the VEGFA probe context.

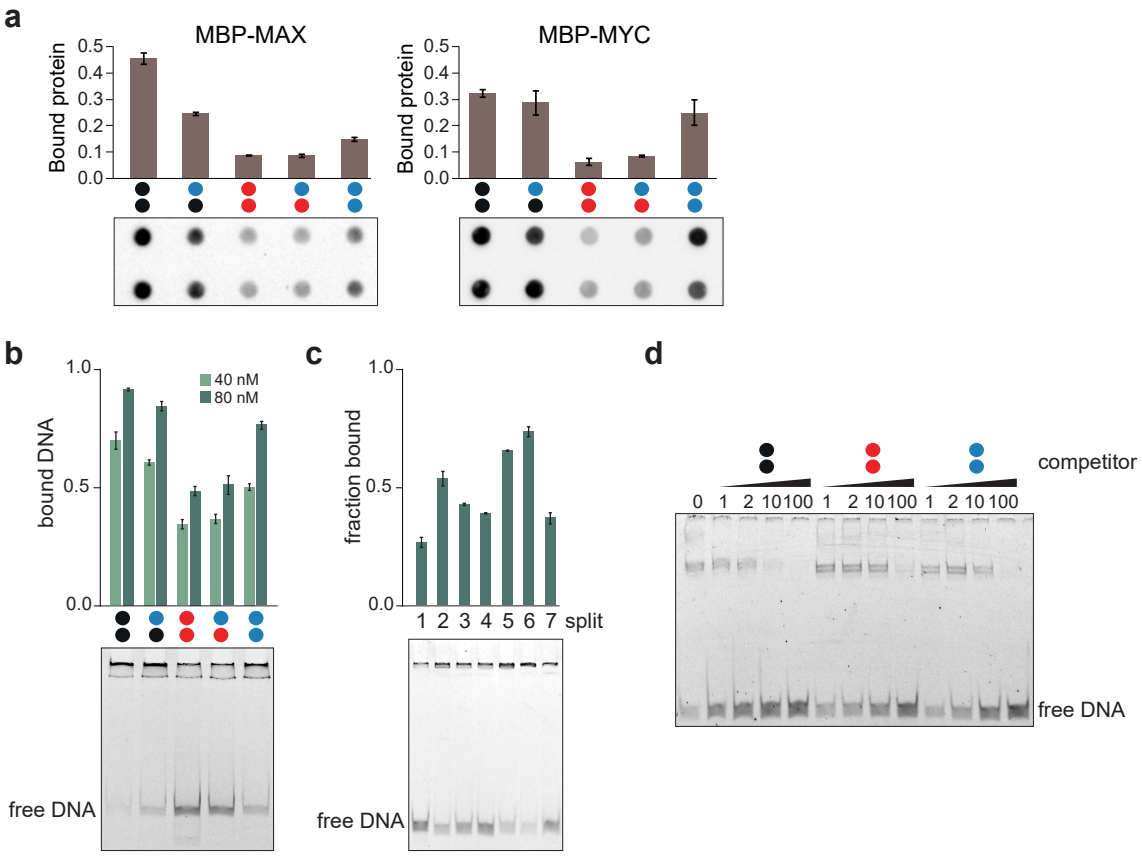


Figure 3.33: MAX and MYC DNA-binding assays. **(a)** Anti-MBP dot blots of enrichments using VEGFA probes and *E. coli* expression extracts of MBP-MAX and MBP-MYC together with nuclear extracts in duplicates displaying the relative enrichments as bar diagrams. **(b)** EMSA profiles of MBP-MAX binding to modified VEGFA probes. Error bars from duplicate experiments. **(c)** EMSA profile of MBP-MAX binding to VEGFA split fragments. **(d)** EMSA analyses of MAX binding to VEGFA split 5 with varying concentrations of differentially modified dark full-length VEGFA (black circles: C/C, red circles: mC/mC and blue circles: hmC/hmC).

Due to the tendency of MBP-fused proteins to aggregated and TEV-digested MAX produced clear bands (compare **Figure 3.33c, d**), the MBP tag was removed by cloning. In addition to the full-length constructs, basic helix-loop-helix leucine zipper (bHLHZ) constructs were generated, containing only the domains responsible for DNA binding and dimerization. This resulted in four different constructs (**Figure 3.34a, b**). The DNA binding ability of these constructs was tested using an unmodified E-box probe, which represents their canonical binding motif. Only MAX_{fl} and MYC_{bHLHZ} showed DNA binding within a reasonable concentration range. MAX_{fl} exhibited very high binding affinity, with partial binding already at 1 nM and complete binding at 20 nM. MYC_{bHLHZ} required higher concentrations, with approximately half of the DNA probe bound at 1000 nM (**Figure 3.34c**). This is in line with expectations of weaker MYC binding affinity.

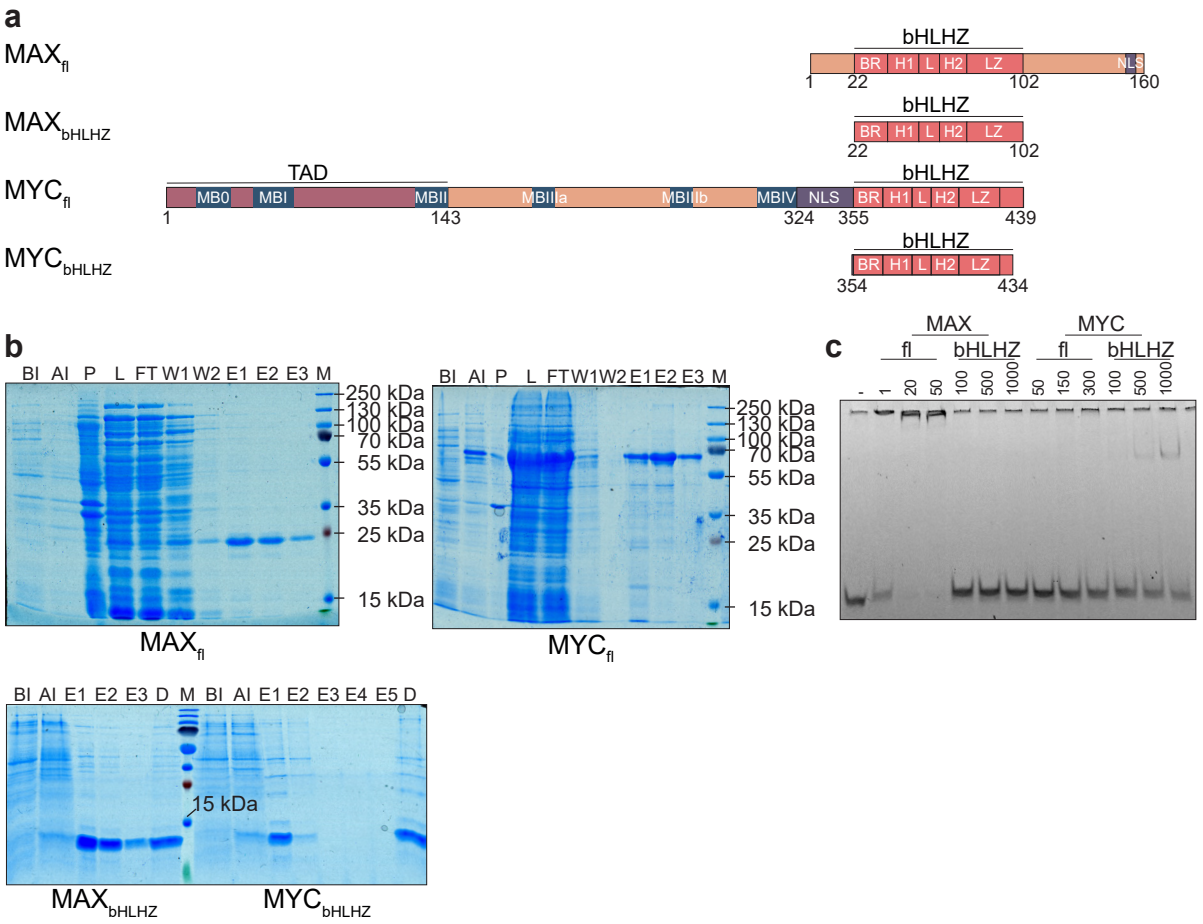


Figure 3.34: MAX and MYC constructs. (a) Schematic representation of the domain organization of MAX and MYC fl and bHLHZ constructs. BR: basic region, H1: helix 1, L: loop, H2: helix 2, LZ: leucine zipper, NLS: nuclear location signal, TAD: transactivation domain, MB: Myc boxes. **(b)** Purification steps of MAX and MYC fl and bHLHZ constructs visualized on a SDS-PAGE gel. Samples were taken from BI (before induction), AI (after induction), P (pellet), L (lysate), FT (flow through), W (wash), E (elution), D (dialysis) and Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder was run in lane M. Theoretical MWs are 22.3 kDa for MAX_{fl}, 52.8 kDa for MYC_{fl} (apparent MW 64 kDa), 126 kDa for MAX_{bHLHZ} and 12.9 kDa for MYC_{bHLHZ}. **(d)** EMSA analyses of MAX and MYC constructs in various concentrations binding to 2 nM unmodified E-box 1 probe.

To further investigate the E-box motif probe context and assess whether the previously observed hmC/hmC also appears within this probe context, a series of E-box like probes were tested. These included the canonical E-box (E-box 1) an several degenerated E-box sequences, representing MYC off-targets frequently identified in ChIP-seq datasets (J. Guo et al., 2014). Among these, the most occurring off-target (termed E-box 2), another highly occurring off-target containing CpA instead of CpG (E-box 3), and two additional E-box like sequences found in the VEGFA split fragments 5 and 6 (E-box 4 and E-box 5) were included. MAX_{fl} exhibited strong C/C binding and bound the canonical E-box 1 probe with high affinity, even at low concentrations. High affinity binding was also observed for the other E-box probes with C/C (**Figure 3.35a**). hmC/C binding was strongest for E-box 2, and moderate binding was observed for E-box 1, while the other hmC/C modified E-box probes showed no detectable interaction. mC/mC, hmC/mC and also hmC/hmC did not yield any binding with MAX_{fl} across all tested E-box contexts. MYC_{bHLHZ} also preferentially bound to unmodified E-box 1 (**Figure 3.35b**). Similar to MAX, it displayed measurable hmC/C binding for E-boxes 2 and 3, and while mC/mC, hmC/mC and hmC/hmC were bound weakly, the extent of interaction was minimal and no differential preference among these modifications was observed. Notably, MAX and MYC showed reduced binding to E-box

probes 4 and 5, containing the E-box like sequences present in the most favored VEGFA splits. These observations align with the notion that E-box 1 is the canonical binding site, while E-boxes 2 and 3 represent commonly observed off-targets. No preferential hmC/hmC binding was observed in any of the E-box contexts for either MAX or MYC. Given that MYC and MAX are known to form heterodimers, it is plausible that modification-specific preferences might emerge upon dimerization. Dimerization among bHLHZ proteins occurs typically through their HLHZ domains, forming a four-helix bundle. This enables sequence-specific DNA binding through an induced-fit mechanism in which the basic regions engage in specific base contacts within the major groove of DNA (Ferré-D'Amaré et al., 1993). To explore this, MAX and MYC constructs were co-incubated at varying ratios and then added to unmodified E-box 1 probe to optimize heterodimer formation. Distinct protein-DNA complexes representing the putative dimer were only observed for the MAX_{fl} and MYC_{bHLHZ} combination (**Figure 3.35c**). Multiple bands of differing mobility indicated potential heterodimer formation, with the middle band corresponding to the heterodimer. Other tested combinations (MAX_{bHLHZ} + MYC_{bHLHZ} and MAX_{fl} + MYC_{fl}) produced only single bands corresponding to homodimers, while MAX_{bHLHZ} and MYC_{fl} showed no binding at all, confirming their previously observed inability to bind DNA (**Figure 3.34c**). As the 1:50 ratio of MAX_{fl} to MYC_{bHLHZ} showed a promising putative dimer band, this condition was used to investigate heterodimer binding towards the differentially modified E-box probes. The emerging binding affinities resembled those of the homodimers, with strong C/C binding, and elevated hmC/C binding especially for E-box 2, while mC/mC, hmC/mC and hmC/hmC showed less pronounced binding (**Figure 3.35d**). The general trend of higher affinity for E-boxes 1-3 was consistent with homodimers. Altogether, these findings for both homo- and heterodimer binding to E-boxes support previous reports showing reduced binding affinity for mC/mC and hmC/hmC compared to C/C (D. Wang et al., 2017). They also indicate that the previously observed hmC/hmC preference is limited to the VEGFA context and does not extend to canonical or degenerate E-box motifs.

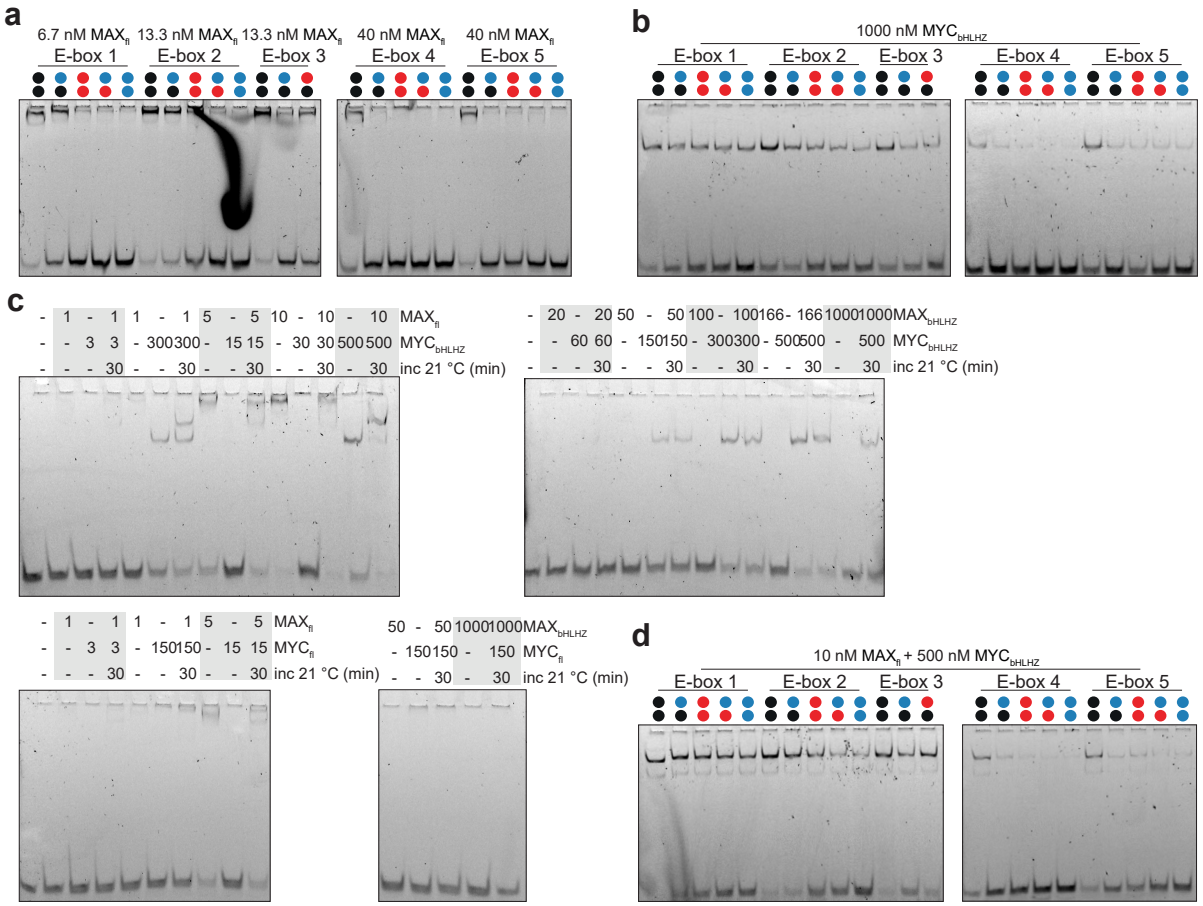


Figure 3.35: Modification preferences and dimerization of MAX and MYC. (a) EMSA analyses of MAX_{fl} binding to 2 nM of indicated E-box probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles). **(b)** EMSA analyses of MYC_{bHLHZ} binding to 2 nM of indicated E-box probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles). **(c)** EMSA analyses of dimerization of MAX and MYC constructs binding to 2 nM of unmodified E-box 1 probe. **(d)** EMSA analyses of putative MAX_{fl} and MYC_{bHLHZ} dimer binding to 2 nM of indicated E-box probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles).

Although the preference for hmC/hmC was not reproduced in a different sequence context and may therefore lack general validity, the binding affinity towards the VEGFA probe was still quantified. Homodimer and heterodimer binding were visualized using an EMSA gel (**Figure 3.36a-c**), and the preference pattern of C/C and hmC/C > hmC/hmC > mC/mC and hmC/mC was consistently reproduced. First, the K_D for the MAX_{fl} homodimer was determined for C/C, mC/mC and hmC/hmC modified VEGFA probes. The expected affinity trend was observed, with a K_D of 4.11 ±0.55 nM for C/C, 14.5 ±2.42 nM for hmC/hmC and 38.5 ±9.3 nM for mC/mC (**Figure 3.36d**). These quantitative measurements confirmed the observed binding preferences. Previous studies were primarily focused on E-box affinity, reporting K_D values of 36 nM for C/C and > 400 nM for mC/mC and hmC/hmC (D. Wang et al., 2017). Next, heterodimer binding affinities to C/C, mC/mC and hmC/hmC modified VEGFA probes were determined using a 1:30 ratio of MAX_{fl} to MYC_{bHLHZ}. While the affinity differences were less pronounced compared to MAX homodimers, the same preference trend was observed (**Figure 3.36e**). The K_D values were 1.704 ±0.254 nM for C/C, 2.868 ±0.550 nM for hmC/hmC and 4.866 nM ±0.864 nM for mC/mC. The observed higher affinity of the heterodimer compared to the MAX homodimer is in agreement with previous studies. While earlier studies have suggested that the MAX homodimer can bind more tightly to DNA, the MYC-MAX heterodimer is considered to be

thermodynamically more stable (Fieber et al., 2001). It should be considered, however, that incomplete heterodimerization in vitro may allow homodimer formation which could influence the apparent K_D values such as those from the presented EMSA assays. The K_D values obtained here are in good agreement with those reported using alternative biophysical methods, although slight differences can be attributed to variations in protein constructs, assay conditions, and measurement techniques. For instance, a SPR study reported K_D values of 15 ±3 nM for the MAX_{fl} homodimer and <4 nM for the MAX_{fl}–MYC_{bHLHZ} heterodimer (Schütz et al., 2022). These values are in strong concordance with the EMSA-derived affinities reported here and support the conclusion that heterodimerization with MYC enhances DNA-binding strength relative to MAX homodimers. Interestingly, the same study found that the MAX_{fl}-MYC_{fl} heterodimer displayed a weaker apparent affinity (<28 nM) than the MAX homodimer, highlighting that full-length protein interactions may vary depending on construct stability and assay format. Similar observations were made in a study using fluorescence polarization (FP), where a K_D of 19.2 ±4.4 nM was reported for the MAX_{bHLHZ} homodimer and of 90.5 ±25.4 for the bHLHZ heterodimer (J. Hu et al., 2005). This suggested that bHLHZ only domains not fully recapitulate the binding behavior of full-length constructs.

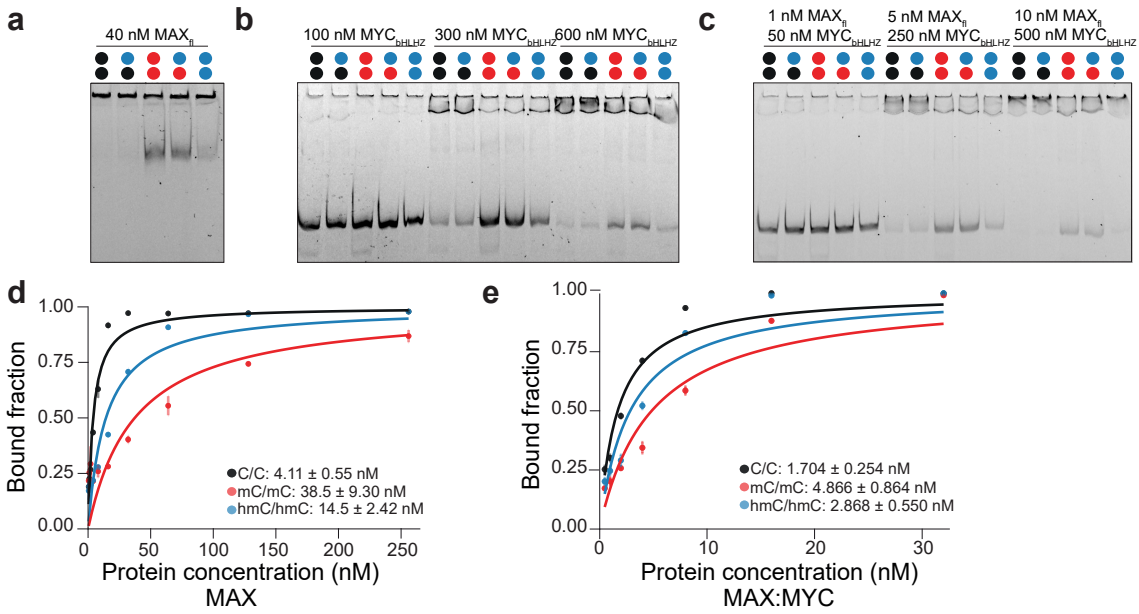


Figure 3.36: MAX_{fl} and MYC_{bHLHZ} binding affinity to the VEGFA probe. **(a)** EMSA analyses of MAX_{fl} binding to 2 nM of VEGFA probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles). **(b)** EMSA analyses of MYC_{bHLHZ} binding to 2 nM of VEGFA probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles). **(c)** EMSA analyses of putative MAX_{fl} and MYC_{bHLHZ} dimers binding to 2 nM of VEGFA probe modified with C/C (black circles), hmC/C (blue and black circles), mC/mC (red circles), hmC/mC (blue and red circles) or hmC/hmC (blue circles). **(d)** EMSA titration experiments and K_D for MAX_{fl} and C/C, mC/mC and hmC/hmC modified VEGFA probes. **(e)** EMSA titration experiments and K_D for MAX_{fl} and MYC_{bHLHZ} putative heterodimer and C/C, mC/mC and hmC/hmC modified VEGFA probes.

In conclusion, the observed binding of MYC and MAX to hmC/hmC was reproducible and robust, but notably limited to a specific sequence context. The VEGFA probe contains modified cytosines not only in CpG but also in CpH contexts. Especially CpA contexts need further investigations on their influence on modification-specific MYC and MAX binding as well as explorations on cooperative or looping-mediated binding mechanisms which can be studied using probes containing multiple E-box motifs. Previous research has suggested that binding to CpH sites may carry distinct selectivity. For example, MECP2 has been shown to selectively recognize hmCpA in certain contexts (Ibrahim et al., 2021), although subsequent studies failed to reproduce these findings (Chhatbar et al., 2022), raising

questions about the broader relevance of such interactions. While MYC and MAX canonically bind to E-box motifs to regulate gene expression via RNA polymerase II (Lüscher and Vervoorts, 2012), the current results suggest that this canonical binding might not fully account for their interaction with hydroxymethylated DNA. Despite the lack of general validity for hmC-dependent binding across E-box and E-box like sequences, the VEGFA promoter context remains an important case. It is well established that MYC is a key regulator of VEGFA expression and plays a central role in vascular development (Y. Shi et al., 2014). In this context, MYC and MAX displayed a clear preference for hmC, supporting the possibility that these transcription factors may exhibit selective binding to modified cytosines in a promoter-specific and sequence-dependent manner. Although the hmC preference may not be universally applicable, sequence dependent high affinity hmC reading of MYC and MAX was observed for the first time and stably reproduced. Future investigations including SELEX and ChIP studies could further elucidate the biological significance.

3.5.5 Essential points

- Candidate hit proteins were cloned, expressed and employed in DNA-binding studies. While many proteins failed to exhibit their expected MS-based binding preferences, likely due to absence of binding partners or lack of PTMs, some demonstrated expected or novel binding behaviors such as FOXA1, RFX5, MYC and MAX.
- RFX5 emerged as a reader of mC/mC and hmC/mC sites, showing a 2-fold reduced affinity for hmC/mC, suggesting that hmC CpG symmetry can influence regulatory outcomes.
- MYC and MAX were observed to have enhanced affinity to hmC/hmC in a sequence-specific manner outside of their canonical E-box motif, revealing a novel binding behavior.
- These findings demonstrate that MS-identified interactions can be recapitulated with expressed proteins when direct DNA binding is involved, independent of cofactors or PTMs. Four proteins fulfilled these criteria, uncovering previously unknown interactions, including those with asymmetric hmC/mC sites.

Some results presented in this chapter, comprising parts of the cloning and expression work, were carried out by Simone Eppmann, and the RFX5 K_D-assay with the X-box 2 probe was performed by Zeyneb Vildan Çakıl. I gratefully acknowledge their contributions.

3.6 Genome wide mapping of hmC/mC sites

To build upon the findings of the proteomics-based pull-down approach, the subsequent project shifted focus from identifying protein interactors of hmC-modified CpG dyads to elucidating the genomic distribution of the suspectedly most prevalent hmC-containing CpG dyad hmC/mC providing complementary insights into the biological context and genomic localization of this modification combination. Although recent studies have shown the prevalence of various dyads (Halliwell et al., 2025; Hardwick et al., 2025), existing techniques remain limited due to low efficiency and high cost and either lack high throughput or accessibility that is required for genome-wide studies.

As a more accessible alternative, I employed an enrichment-based sequencing approach that allows for targeted isolation of DNA fragments that contain the hmC/mC modification. While enrichment-based approaches do not allow for base-resolution mapping, they give insight into the genomic localization of the modified sites. The method is based on a mutant MBD protein derived from MECP2, which was evolved for high selectivity to hmC/mC dyads. Previous work in the Summerer group (Buchmuller et al., 2022; H. Singh et al., 2023) established this mutant MECP2 protein (termed MECP2 HM; mutations K109T, V122A, S134N) that demonstrated strong and selective binding of hmC/mC dyads. MECP2 HM showed a similar degree of selectivity for its target (hmC/mC over mC/mC) as MECP2 wildtype (termed MECP2 WT) does for mC/mC over hmC/mC in EMSA assays. Moreover MECP2 HM showed superior discrimination for hmC/mC against other dyads. I applied this mutant in affinity-based enrichment assays using fragmented genomic DNA (gDNA) from E14 Tg2a mESCs. The DNA was sheared to around 200 bp fragments, adapter ligated and incubated with GST-tagged MECP2 proteins, which were then immobilized on magnetic beads (Diagenode MethylCap kit) (**Figure 3.37**). Both MECP2 WT and MECP2 HM were tested.

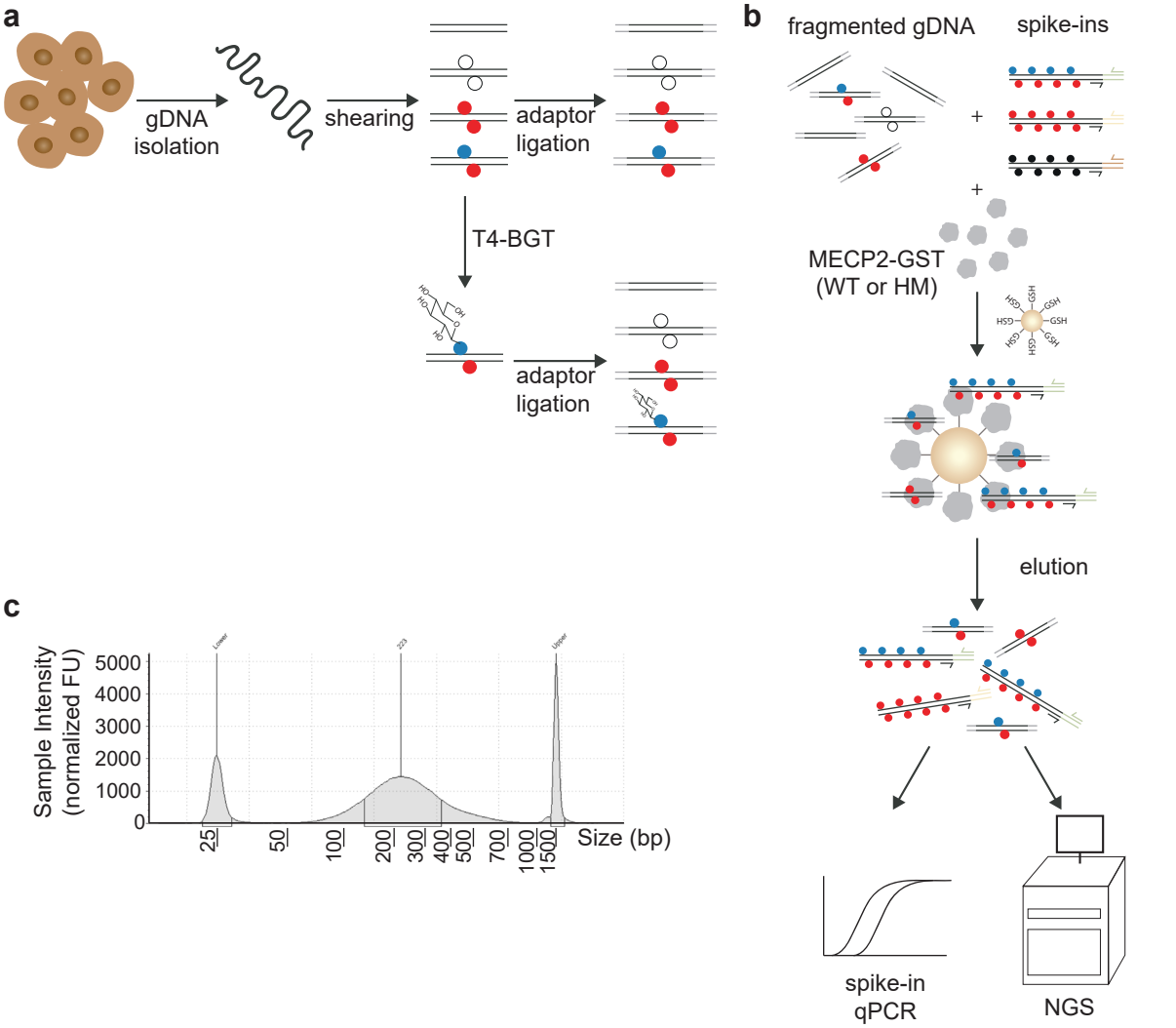


Figure 3.37: Selective enrichment of hmC/mC-modified DNA. (a) gDNA was isolated from E14 Tg2a mESCs and subsequently sheared to a fragment length of 200 bp. DNA was either subjected to glucosylation by T4-BGT or directly adapter ligated before applied in the enrichment assay. **(b)** Workflow of the enrichment process. Fragmented gDNA together with spike-ins bearing either hmC/mC (blue and red circles), mC/mC (red circles) or C/C (black circles) modified CpGs are incubated with MECP2 WT or HM fused to GST. GST-tagged protein together with bound DNA fragments is captured by GSH-coated magnetic beads. Enriched spike-ins were quantified by qPCR using a spike-in specific primer. Enriched gDNA was sequenced by NGS. **(c)** Exemplary fragment distribution profile of sheared gDNA demonstrated successful fragmentation of DNA to an average size of around 200 bp.

3.6.1 Spike-in recovery

To assess the binding specificity, synthetic DNA spike-ins containing defined CpG dyad modifications (C/C, mC/mC, hmC/mC) were included during the enrichment (**Figure 3.37b**). Each spike-in was added in 50,000 copies together with the fragmented gDNA. Three parallel experiments were conducted, using (1) gDNA, MECP2 WT and spike-ins, (2) gDNA, MECP2 HM and spike-ins, and (3) glucosylated gDNA, MECP2 HM and spike-ins (**Figure 3.38a**). Glucosylation of hmC prevents MECP2 HM binding serving as a control for the sequencing experiment as all previously bound hmC/mC sites will be blocked. EMSA confirmed that glucosylation abolished binding of MECP2 HM to hmC-containing probes but not to mC/mC, validating the specificity of the treatment (**Figure 3.38b**).

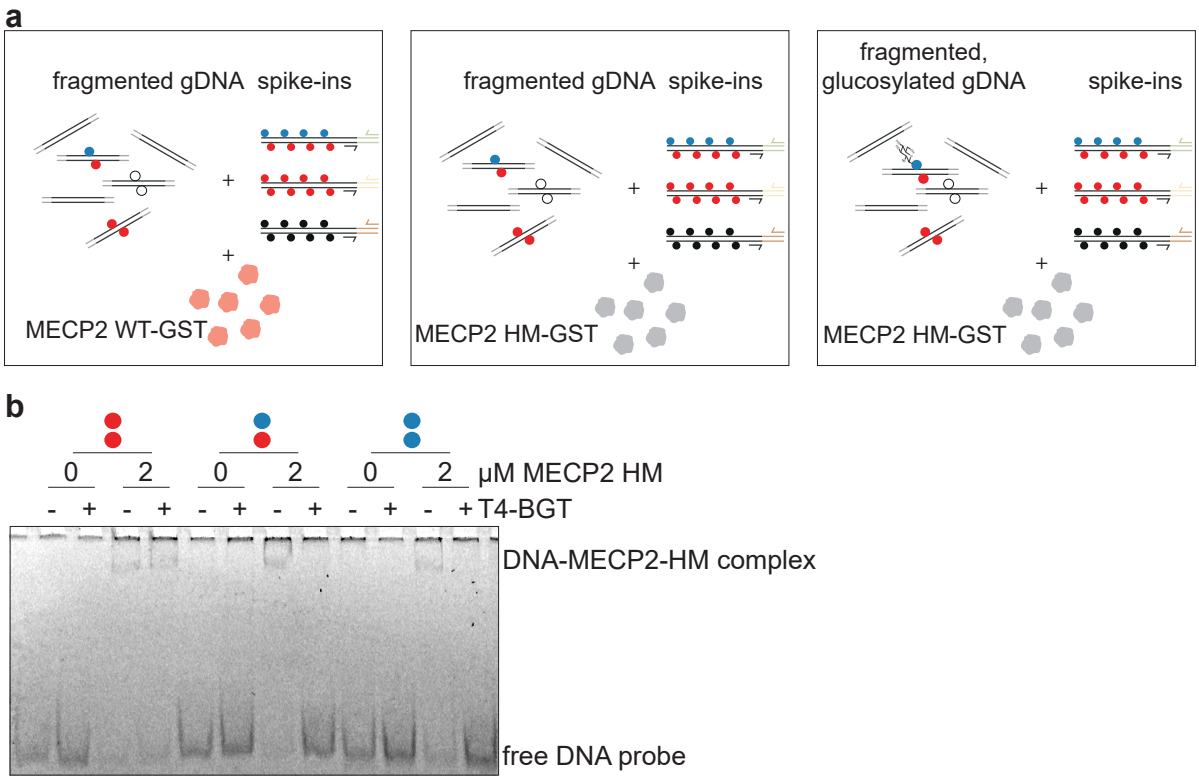


Figure 3.38: Experimental set-up and glucosylation. (a) Three experimental set-ups were conducted. The first one comprised enrichment of gDNA and spike-ins with MECP2 WT, the second one enrichment of gDNA and spike-ins with MECP2 HM, and the third one enrichment of glucosylated gDNA and spike-ins with MECP2 HM. **(b)** Synthetic CG-containing probes in a polyA/T context bearing either mC/mC, hmC/mC, or hmC/hmC-modified CpGs, were incubated with 0 or 2 μm MECP2 HM with or without previous glucosylation by T4-BGT.

Recovered DNA after enrichment assays was analyzed by qPCR using spike-in specific primers (**Figure 3.37b**). MECP2 WT recovered both the mC/mC and hmC/mC spike-ins (50 – 60 %), with slightly higher enrichment of mC/mC (**Figure 3.39a** and **Figure A.15** for further biological replicates). In contrast, MECP2 HM showed the expected reverted selectivity and recovered hmC/mC more than 2-fold compared to mC/mC, indicating selective enrichment. C/C spike-in recovery was low (around 10 %) with both proteins reflecting the weak affinity of MECP2 to unmodified DNA. These results align with previous studies on MECP2 WT selectivity (Buchmuller et al., 2020) and initial spike-in studies with MECP2 HM (Buchmuller et al., 2022). The glucosylated gDNA background had only minimal direct impact on spike-in recovery by MECP2 HM, as the spike-ins themselves were not glucosylated. The recovery profiles differed minimally from the ones without gDNA glucosylation, likely due to reduced competition from glucosylated DNA for MECP2 HM binding (**Figure 3.39a**, **Figure A.15**).

To get more insights into the binding dynamics during the enrichment process, additionally the

intermediate enrichment fractions were analyzed for spike-in presence by qPCR. The C/C spike-in was mostly found in the flow-through and first wash step, indicating weak binding by both proteins (**Figure 3.39b**, **Figure A.16a**, **Figure A.17a**, **Figure A.18a**). No C/C spike-in was found in the second wash step. The elution fraction revealed that MECP2 WT recovered more C/C than MECP2 HM, however, recoveries were between 7-15 %, suggesting an overall low affinity of both proteins towards C/C. The mC/mC spike-in was retained more efficiently, especially with MECP2 WT (**Figure 3.39c**, **Figure A.16b**, **Figure A.17b**, **Figure A.18b**). MECP2 HM showed weaker binding to mC/mC indicated by higher amounts of this spike-in lost in the wash steps (30-40 %). The stronger binding of MECP2 WT to mC/mC was reflected by a minimal amount of the mC/mC spike-in in the wash fractions. The second wash step showed no presence of the spike-ins in all cases, confirming a sufficient wash protocol that only retained strongly bound spike-ins. The elution fraction demonstrated the strong affinity of MECP2 WT to mC/mC with a recovery higher than that of MECP2 HM. For hmC/mC, both proteins retained the spike-in effectively (**Figure 3.39d**, **Figure A.16c**, **Figure A.17c**, **Figure A.18c**). Almost nothing was present in the flow-through, and only a minimal amount was found in the first wash step. The elution fractions contained a lot of hmC/mC spike-in, with slightly more recovery for MECP2 WT. However, only MECP2 HM demonstrated a robust discrimination between mC/mC and hmC/mC, confirming its specificity (**Figure 3.40**, **Figure A.16d-f**, **Figure A.17d-f**, **Figure A.18d-f**).

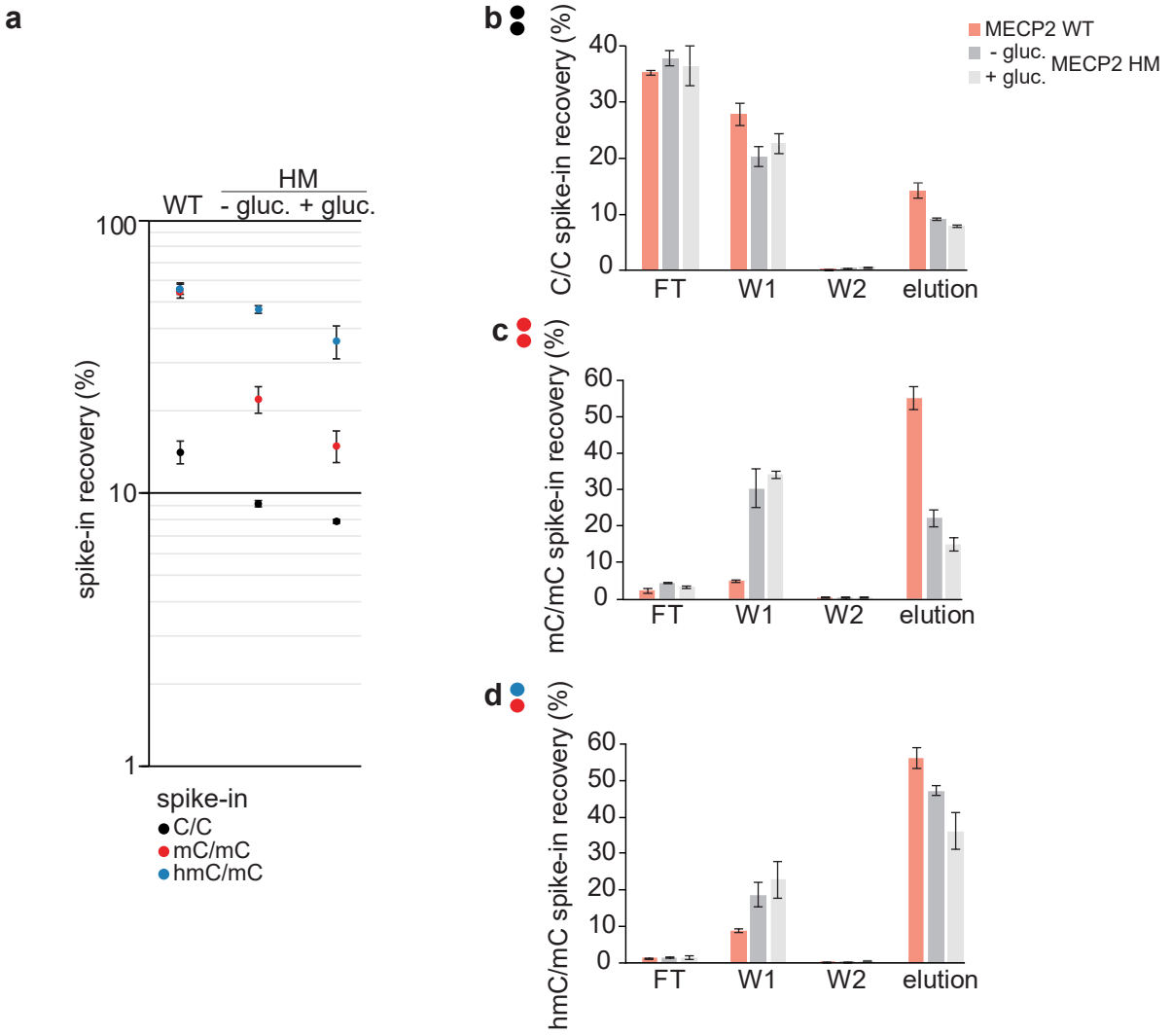


Figure 3.39: Recovery of modified spike-ins during enrichment. **(a)** Spike in recoveries in elution fractions with MECP2 WT and MECP2 HM (with glucosylated and non-glucosylated gDNA background). **(b)** Bar diagram showing the presence of C/C spike-in in the flow-through (FT), first wash (W1), second wash (W2), and elution fraction of enrichment with MECP2 WT and MECP2 HM (with glucosylated and non-glucosylated gDNA background). Bar diagrams of showing the presence of mC/mC spike-in and hmC/mC spike-in are shown in **(c)** and **(d)**, respectively.

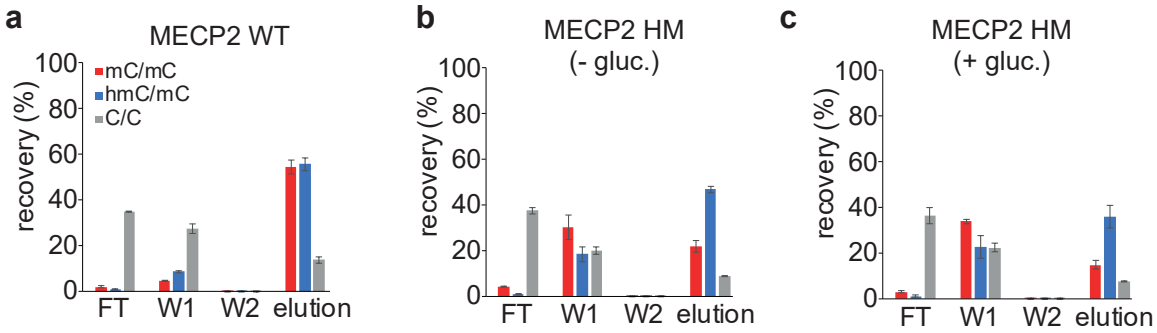


Figure 3.40: Protein-specific recovery of spike-ins. Bar diagrams showing comparative spike-in enrichment in each step of the enrichment process for **(a)** MECP2 WT, **(b)** MECP2 HM, and **(c)** MECP2 HM in presence of glucosylated gDNA.

3.6.2 NGS

After validating the specificity of the enrichments with spike-ins, the eluted gDNA was processed for NGS. Libraries were barcoded, quantified, and assessed for fragment length before they were pooled and sequenced on the Illumina platform. Glucosylated samples provided a background subtracted control for identifying true hmC/mC sites. NGS analysis was performed on all four biological replicates, however, only one replicate yielded sufficient sequencing quality for in-depth analysis, while the other three produced less informative results. Therefore, the following section presents the analysis of the high quality replicate corresponding to **Figure A.18**.

Sequencing yielded between 63 million and 184 million reads per library and over 95% of reads mapped to the *mm10* reference genome. Three technical replicates were conducted and analyzed for each enrichment condition. Spearman correlation analysis for 100 kb genomic bins revealed a high reproducibility (**Figure 3.41a**). MECP2 WT replicates had r>0.97, while MECP2 HM samples with and without glucosylated gDNA reached r>0.99. Indeed, MECP2 HM enrichment using glucosylated DNA formed a distinct group separate from MECP2 WT and MECP2 HM enrichment with untreated gDNA. The input libraries clustered together and correlated less strongly with the enriched libraries (r=0.84−0.89). Heatmaps and average enrichment profiles centered on peak regions (±2 kb) showed strong and reproducible enrichment for MECP2 WT, MECP2 HM and MECP2 HM with glucosylated DNA samples across replicates with sharp signal peaks at the center of identified binding sites relative to the input (**Figure A.19**).

Peak calling identified 91,481 reproducible peaks for MECP2 WT and 149,656 for MECP2 HM (in at least 2 out of 3 technical replicates). A consensus peak set showed that 24.2 % of peaks were unique to MECP2 WT, 54.3 % unique to MECP2 HM, and 21.5 % were shared (**Figure 3.41b**). Both MECP2 WT and MECP2 HM associated peaks were significantly enriched in 3'UTRs, CDS and TTS (±1 kb), while both were depleted (q<0.05) in rRNA and tRNA genes, 5' UTRs and TSS (±1 kb) (**Figure 3.41c, d**). Notably, MECP2 HM peaks were significantly enriched at CpG islands compared to MECP2 WT.

Comparison between MECP2 WT regions and MeDIP-seq revealed a substantial overlap of 64.2 % of peaks (**Figure 3.42a**). However, the majority of MeDIP peaks were unique, while only 7.2 % of MECP2 WT peaks were unique, suggesting potential unspecific enrichment of the antibody-based MeDIP method consistent with earlier reports (Lentini et al., 2018). 54,350 peaks (12 %) were shared. MECP2 WT peaks therefore represent a more restricted but distinct genomic subset relative to MeDIP. Both methods showed enrichment at 3'UTRs, CDS and TTS, and depletion at rRNA and tRNA genes as well as 5'UTRs and TSS (**Figure 3.42b, c**). They diverged clearly at CpG islands where MeDIP peaks were strongly depleted while MECP2 WT peaks were only mildly depleted.

Comparison between MECP2 HM regions and hMeDIP-seq showed 38.99 % overlap (**Figure 3.42d**). While most peaks were unique to hMeDIP (73.4 %), 16.3 % were unique to MECP2 HM. 10.3 % of peaks were shared. This suggests that MECP2 HM enrichment captures a more distinct subset of genomic regions compared to hMeDIP. Feature analysis revealed that CpG islands were most divergent. MECP2 HM peaks were modestly enriched while hMeDIP peaks were strongly depleted (**Figure 3.42e**). At TTS and 3'UTRs, MECP2 HM peaks showed significant enrichment (q<0.05), while non overlapping hMeDIP peaks were slightly depleted (**Figure 3.42e**). This suggests that MECP2 HM binding may preferentially associate with regulatory regions, such as CpG islands, 3'UTRs and TTS, while hMeDIP captures a larger and more unspecific genomic landscape. It should be noted that the targets of both enrichment methods differ, which may further explain the differences in genomic enrichment. While hMeDIP captures hmC on single strands, MECP2 HM recognizes the combinatorial hmC/mC CpG dyad.

As shown before, glucosylation of hmC blocks MECP2 HM binding (**Figure 3.38b**). Comparison

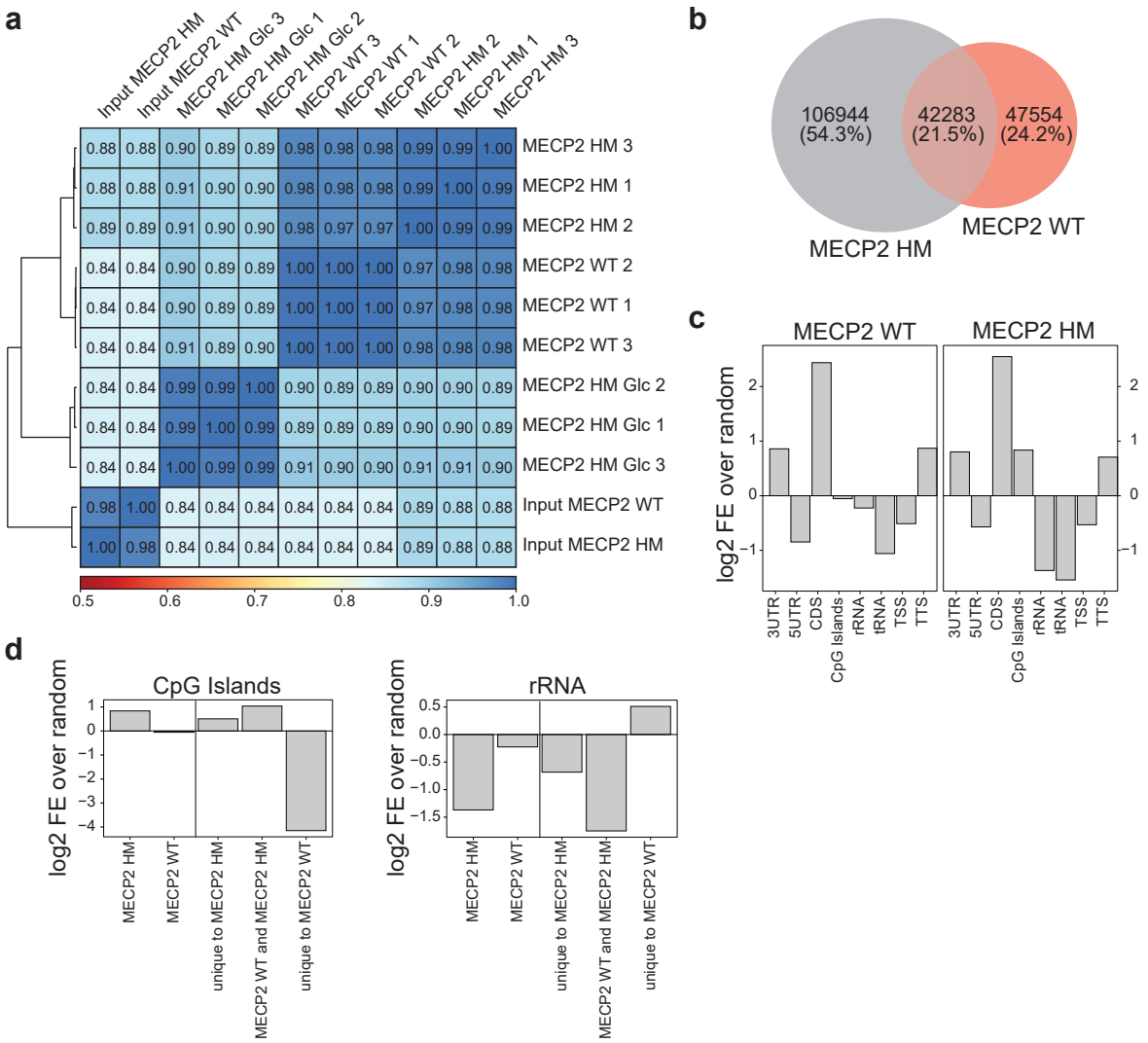


Figure 3.41: NGS analysis of MECP2 WT vs MECP2 HM enrichment. **(a)** Spearman correlation analysis of genome-wide enrichment profiles comparing MECP2 WT and MECP2 HM samples illustrating the overall similarity between the conditions and the technical replicates. **(b)** Venn diagram showing the amount of peaks unique or shared between MECP2 WT and MECP2 HM. **(c)** Genomic feature enrichment analysis showing the preferential localization of MECP2 WT and MECP2 HM binding across the annotated genomic regions. **(d)** Analysis of enrichment or depletion relative to functional genomic elements,demonstrating the differences in MECP2 WT and MECP2 HM binding to CpG islands and rRNA associated loci.

of untreated and glucosylated gDNA enrichments with MECP2 HM showed 119,415 unique peaks (73.9 %) exclusive to untreated DNA enrichment and only 12,008 peaks (7.43 %) unique to glucosylated gDNA enrichment, demonstrating a clear loss in peaks upon DNA glucosylation (**Figure 3.43a**). 30,186 peaks (18.7 %) were shared with both approaches, which may correlate to regions not containing hmC, such as mC/mC sites. IGV browser tracks confirmed consistent enrichment across replicates with regions of overlap and condition specific binding, especially loss in signal at many loci following hmC glucosylation (**Figure 3.43b**).

Normalized read density plots (log2 ratios of normalized coverage between pull-down over input) aligned to protein-coding genes confirmed reproducibility across replicates (**Figure 3.43c**). Both MECP2 WT and MECP2 HM pull-downs showed enrichment across gene bodies with depletion at TSS and TTS, extending into proximal flanking regions. Upon hmC glucosylation, MECP2 HM enrichment was strongly reduced, with the largest signal loss at promoter proximal and termination regions. Together, this indicates that MECP2 WT and MECP2 HM associate with gene bodies and regulatory flanking

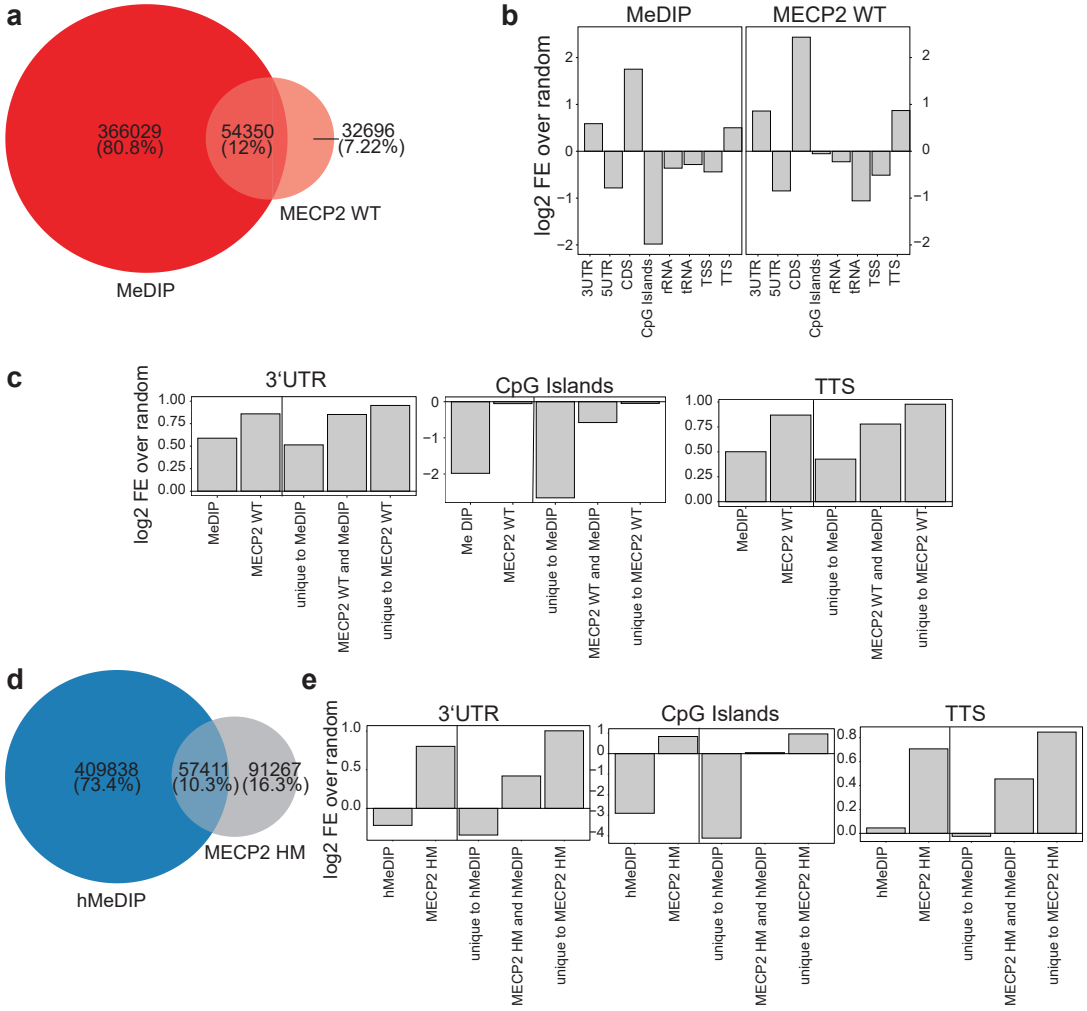


Figure 3.42: Comparison with antibody-based enrichment methods. **(a)** Venn diagram showing the overlap between peaks identified by MeDIP and MECP2 WT, indicating shared and unique enrichment regions. **(b)** Genomic feature enrichment analysis of MeDIP and MECP2 WT peaks across annotated genomic regions. **(c)** Feature-specific analysis highlighting enrichment and depletion at 3'UTR, CpG islands and TTS. **(d)** Venn diagram comparing peaks identified by hMeDIP and MECP2 HM, illustrating overlap and unique binding regions. **(e)** Feature-specific analysis of hMeDIP and MECP2 HM enrichment and depletion at 3'UTRs, CpG islands and TTS.

regions, but MECP2 HM occupancy is significantly diminished upon hmC glucosylation.

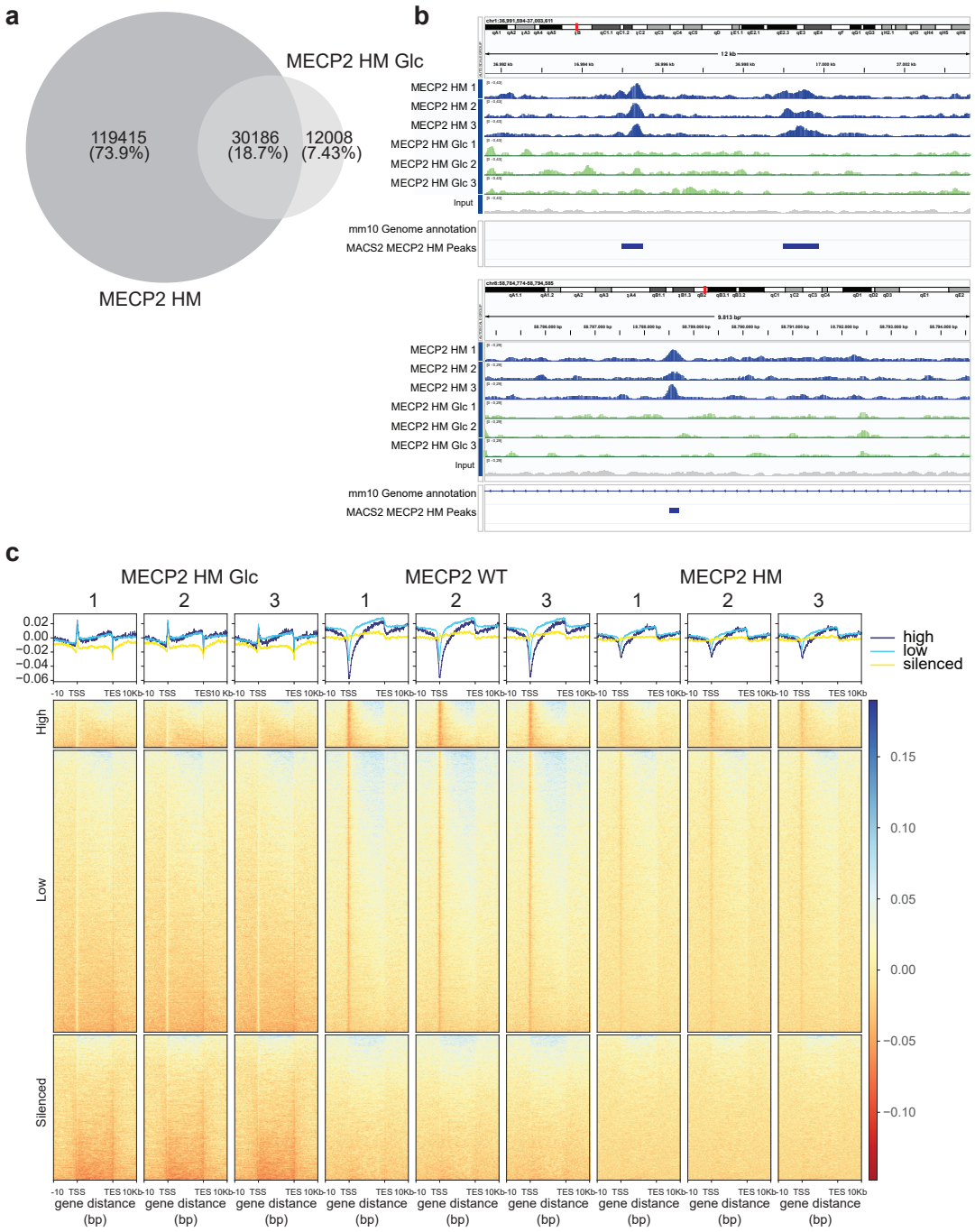


Figure 3.43: Glucosylation control of MECP2 HM enrichment. **(a)** Venn diagrams comparing MECP2 HM enrichment peaks on glucosylated vs non-glucosylated DNA, illustrating overlap and unique binding regions. **(b)** Representative genome browser snapshots (IGV traces) showing peak enrichment profiles with and without glucosylation, highlighting local differences in binding intensity. **(c)** Analysis of relative distribution of MECP2 WT and MECP2 HM binding across genes graded by expression level (dark blue: high, light blue: low, yellow: silenced).

3.6.3 Essential points

- Genomic locations of asymmetric hmC/mC CpG dyads were mapped using a selective enrichment approach employing an evolved MECP2 mutant (MECP2 HM) with higher affinity for hmC/mC than mC/mC.
- Validation of protein specificity was achieved by spike-in quantification using qPCR confirming MECP2 HM selectivity of hmC/mC > mC/mC > C/C.
- Glucosylation of gDNA served as a specificity control by blocking MECP2 HM binding to hmC-containing CpG dyads, allowing for identification of these sites in downstream analysis.
- Fragmented (and optionally glucosylated) gDNA was enriched with MECP2 WT or MECP2 HM and subsequently subjected to Illumina sequencing.
- NGS analysis revealed high reproducibility and specificity with the different enrichment conditions, with a clear separation between MECP2 WT, MECP2 HM, and MECP2 HM with glucosylated gDNA. This confirms the robustness of the enrichment strategy and specificity of the proteins.
- Peak calling and comparative analysis showed that MECP2 WT and MECP2 HM occupy partially overlapping but distinct genomic regions. MECP2 HM showed preferential enrichment at regulatory features such as CpG islands, 3'UTRs and TTS. Glucosylation of hmC substantially reduced the number of peaks, consistent with the blocking of MECP2 HM binding.

The MECP2 HM mutant used in this chapter was created by Dr. Benjamin Buchmüller and expressed and purified by Simone Eppmann. NGS library preparation of one of the four biological replicates was carried out by Dr. Damian Schiller. NGS data analysis was carried out in collaboration with Dr. Sidney Becker and Marlon Zambrano. Figures containing NGS data analysis and their interpretation were prepared by Marlon Zambrano, which I have adapted and modified for this thesis. I gratefully acknowledge their contributions.

4 Materials

4.1 Chemicals and Reagents

4.1.1 Chemicals

Table 4.1: List of chemicals used in this study.

Chemical	CAS number	Supplier	Catalog number
acetic acid	64-19-7	Carl Roth	6755
acetonitrile	75-05-8	Sigma Aldrich	34851-1L
agarose LE	840006	Biozym Scientific	840004
ammonium persulfate (APS)	7727-54-0	Carl Roth	9592
boric acid	10043-35-3	Carl Roth	233-139-2
bovine serum albumine (BSA)	9048-46-8	Cell Signaling Technology	# 9998S
carbenicillin	4800-94-6	Carl Roth	6344
chloroacetamide (CAA)	79-07-2	Acros Organics	10071660
Coomassie Brilliant Blue G 250	6104-58-1	Carl Roth	9598
disodium phosphate	7558-79-4	Roth	4984.1
dithiothreitol (DTT)	3483-12-3	Carl Roth	6908.3
ethanol absolute	64-17-5	Merck	32221
ethidium bromide	1239-45-8	Carl Roth	2218
EDTA	6381-92-6	Alfa Aeser	E5134
EGTA	67-42-5	Carl Roth	200-651-2
formic acid	64-18-6	Carl Roth	200-579-1
glycerol	56-81-5	Carl Roth	6962
HEPES	7365-45-9	Alfa Aeser	J60463.AE
hydrochloric acid, 37 %	7647-01-0	Merck	30721
IGEPAL® CA-630	9036-19-5	Alfa Aesar	J61055
imidazole	288-32-4	Roth/abcr	3899.4/109053
isopropanol	67-63-0	Fisher Scientific	10315720
IPTG	367-93-1	Carl Roth	2316
magnesiumchlorid hexahydrat	7791-18-6	Acros	197530010
2-mercaptoethanol	60-24-2	Merck	8.05740.0250
methanol	67-56-1	Merck	34860
non fat dry milk	68514-61-4	Carl Roth	T145.2
PMSF	329-98-6	Roth	315217174

Chemical	CAS number	Supplier	Catalog number
potassium acetate	127-08-2	Merck	1.04820.1000
potassium chloride	7447-40-7	Carl Roth	A137
potassium dihydrogen phosphate	7778-77-0	Fluka	60220
Rotiphorese® 25 % solution	-	Roth	3043.1
sodium acetate	127-09-3	Sigma Aldrich	S8750
sodium chloride	7647-14-5	Merck	31434
sodium dodecyl sulfate pellets	151-21-3	Roth	CN30.3
sodium hydroxide	1310-73-2	Merck	30620
sucrose	57-50-1	Carl Roth	200-334-9
TEMED	110-18-9	Carl Roth	2367
TFA	76-05-1	Thermo Scientific™	28904
Trizma® Base (Tris)	77-86-1	Carl Roth	AE15
Triton® X-100	9002-93-1	Fluka	93420
Tween® 20	9005-64-5	Fisher Bioreagents	BP337-500
urea	57-13-6	Alfa Aesar	36428
zinc sulfate	7446-20-0	VWR	29253236

4.1.2 Reagents

Table 4.2: List of reagents used in this study.

Reagent	Catalog number	Supplier
BSA, molecular biology grade	B9000S	NEB
cadCTP	040N-2063-5	tebu-bio
Clarity™ Western ECL Substrate	1705060	Bio-Rad
10x rCutSmart™ buffer	B6400S	NEB
DMEM	P04-03600	PAN Biotech
DNA marker, 2-log ladder	N3200S	NEB
dNTP mix	N0447L	NEB
10x DreamTaq™ Buffer	B65	Thermo Scientific™
Dulbecco's Phosphate Buffered Saline (DPBS)	P04-35500	PAN Biotech
Dynabeads™ MyOne™ Streptavidin T1 magnetic beads	65601	Invitrogen™
fetal bovine serum (FBS)	P30-3306	PAN Biotech
L-glutamine	P04-80100	PAN Biotech
glycogen	10901393001	Roche Diagnostics
hmdCTP	NU-932S	Jena Bioscience
5x Phusion™ HF Buffer	B0518S	NEB
HisPur™ Ni-NTA Resin	88221	Thermo Scientific™
human prostate cDNA	10108-A	Biocat
human thyroid cDNA	HD-503	Zyagen
mdCTP	D1035	Zymo Reseach
10x NEBuffer™ 2	B7002S	NEB
NEBNext Sample Purification Beads	E7103	NEB
PageRuler™, prestained	26616	Thermo Scientific™
2x primaQuant SYBRGreen Master Mix with ROX	SL-9902HR-20ML	Steinbrenner

Reagent	Catalog number	Supplier
protease inhibtior, cOmplete tablets	4693124001	Roche
streptomycin	P06-07100	PAN Biotech
10x T4 DNA ligase buffer	B0202S	NEB
10x T4 polynucleotide kinase buffer	B0201S	NEB
10x ThermoPol Buffer	B9004S	NEB
trypsin 0.05 %/EDTA 0.02 %	P10-0231SP	PAN Biotech
UDP-Glucose	S2200S	NEB

Table 4.3: List of enzymes used in this study.

Enzyme	Catalog number	Supplier
DNaseI	M0303L	NEB
DpnI	R0176L	NEB
Dreamtaq DNA polymerase	EP0705	Thermo Scientific™
HphI	R0158S	NEB
Klenow fragment exo-	M0212S	NEB
LysC endoproteinase	P8109S	NEB
Lysozyme	232-620-4	Merck
Mismatch endonuclease I	M0678S	NEB
MspI	R0106S	NEB
NheI	R3131S	NEB
Phusion® Polymerase	M0530S	NEB
rSAP	M0371S	NEB
T4 DNA ligase	M0202S	NEB
T4 polynucleotide kinase	M0201S	NEB
T4-BGT	M0357S	NEB
T5 exonuclease	M0663S	NEB
T7 endonuclease I	M0302S	NEB
Taq DNA ligase	M0208S	NEB
Taq DNA Polymerase	Ep0401	Thermo Scientific™
Trypsin-ultra TM, MS grade	P8101S	NEB
XhoI	R0146S	NEB

Table 4.4: List of antibodies used in this study.

Antibody	Catalog number	Supplier
Anti-5mC antibody (33D3 clone), monoclonal	C15200081-100	Diagenode
Anti-hmC mouse antibody, monoclonal	C15200200	Diagenode
Anti-MBP antibody, monoclonal	E8032S	NEB
Anti-mouse ECL HRP-conjugated antibody	NA931VS	Cytvia
Anti-rabbit ECL HRP-conjugated antibody	NA934VS	Cytvia
β-Tubulin antibody	#2146	Cell Signaling Technology
Lamin A/C antibody	#2032	Cell Signaling Technology
MBD2 antibody, polyclonal	PA5-115551	Invitrogen™

4.1.3 Kits

Table 4.5: List of kits used in this study.

Kit	Catalog number	Supplier
MagMeDIP-seq Package	C02010041	Diagenode
MethylCap Kit	C02020010	Diagenode
Monarch Genomic DNA Purification Kit	T3010	NEB
NEBNext Multiplex Oligos	E7335S	NEB
NEBNext Ultra II DNA Library Preparation Kit	E7103	NEB
NucleoSpin® Plasmid EasyPure kit	740727.50	Macherey-Nagel
NucleoSpin® Gel and PCR Clean-Up kit	740609.50	Macherey-Nagel
Oligo Clean & Concentrator kit	D4060	Zymo Reseacrh
Pierce™ BCA Protein Assay kit red. agent compatible	23250	Thermo Scientific™

4.2 Buffers

Alkylation solution, 1x
50 mM CAA in Denaturing/Reducing buffer

Binding and Washing buffer (B&W), 2x
10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl

Buffer A, 1x
10 mM HEPES (pH 7.9), 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM PMSF

Buffer C, 1x
20 mM HEPES (pH 7.9), 400 mM NaCl, 1 mM each of DTT, EDTA, EGTA and PMSF

Cell lysis buffer (CLB), 1x
10 mM HEPES (pH 7.9), 10 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM DTT and 0.5 mM PMSF

Coomassie brilliant blue staining solution
50 % EtOH, 10 % AcOH, 40 % H_2O, 2.5 % Coomassie® R-250

Denaturation/Reducing buffer, 1x
8 M urea, 50 mM Tris (pH 7.5), 1 mM DTT

Destain solution
10 % AcOH, 20 % EtOH, 70 % H_2O

Dialysis buffer, 1x
20 mM HEPES (pH 7.3), 100 mM NaCl, 10 % glycerol

EMSA buffer, 1x
20 mM HEPES, 30 mM KCl, 1 mM EDTA, 1 mM $(NH_4)_2SO_4$, 0.2 % Tween®20

EMSA loading buffer, 6x
1.5 x TBE, 40 % glycerol

Hybridization buffer, 1x
20 mM HEPES (pH 7.3), 30 mM KCl, 1 mM EDTA, 1 mM $(NH_4)_2SO_4$

IDT buffer, 1x
30 mM HEPES, 100 mM KOAc

LB broth
10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.0

Lysis buffer, 1x
50 mM Tris-HCl (pH 8.0), 20 % sucrose, 1 mM EDTA, 0.5 mM PMSF, 1 mM DTT, 0.1 mg/ml lysozyme, 0.1 % Triton™ X-100

Nuclear extraction buffer (NEB), 1x
20 mM HEPES (pH 7.9), 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF
Phosphate buffered saline (PBS), 1x
137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4
Protein binding buffer (PBB), 1x
50 mM Tris-HCl (pH 8), 150 mM NaCl, 1 mM DTT, 1 mM PMSF, 0.25 % IGEPAL® CA-630
SDS running buffer, 1x
3 g/L Tris base, 14 g/L glycine, 1 % (w/v) SDS, pH 8.3
SDS sample buffer, 4x
250 mM Tris-HCl (pH 6.8), 500 mM DTT and 0.1 % SDS
S.O.C.
20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 20 mM glucose
TBE, 1x
89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3
TBST, 1x
20 mM Tris base, 150 mM NaCl, 0.1 % (w/v) Tween®20
Tris EDTA (TE), 1x
10 mM Tris-HCl, 1 mM EDTA, pH 8.0
Urea buffer, 1x
PBS with 1 % Triton X-100, 1 mM PMSF, 5 mM imidazole, 7 M urea
Urea wash buffer, 1x
10 mM Tris-HCl (pH 7.8), 500 mM NaCl, 1 % Triton™ X-100, 1 mM PMSF, 10 mM imidazole, 7 M urea

4.3 Biological Materials

4.3.1 Cell lines

Table 4.6: Cell lines used in this study.

Cell line	Description	Supplier
HEK293T	Human embryonic kidney cells (transformed with SV40 large T antigen)	MPI Dortmund
HeLa	Human cervical cancer cells (adenocarcinoma, HPV-18 positive)	MPI Dortmund
E14Tg2a	Mouse embryonic stem cells (derived from strain 129/Ola)	Dr. Christian Schröter

4.3.2 Bacterial strains

Table 4.7: List of *E. coli* strains used in this study.

Strain	Genotype	Supplier
GH371	F⁻ *mcrA* Δ*(mrr-hsdRMS-mcrBC)* φ*80lacZ*Δ*M15*Δ*(lac)X74 recA1 endA1 araD139* Δ*(araA–leu)7697 galU galK rpsL(Str^r) nupG fhuA::IS2 upp*$^-$	iGEM

Strain	Genotype	Supplier
BL21-Gold(DE3)	F^- hsdS$_B$ (r_B^- m_B^-) gal dcm$^+$ endA1 ompT λ(DE3) Tet^r Hte	Agilent

4.3.3 Oligonucleotides

Table 4.8: List of oligonucleotides used in this study used for probe generation and PCR assays.

Identifier	Description	Purpose	Sequence 5′ ->3′
o1565	cPCR	Seq. primer	CACTATAGGGGAATTGTGAGCGGATAACAA
o1590	cPCR	Seq. primer	CTAGTTATTGCTCAGCGG
o2904	CG_T probe	EMSA probe	TTTTTTTTTTTCGTTTTTTTTTTT
o2906	CG_A probe	EMSA probe	[6-FAM]AAAAAAAAAAACGAAAAAAAAAAA
o2909	CG_T probe	EMSA probe	TTTTTTTTTTTXGTTTTTTTTTTT,X=mC
o2967	CG_A probe	EMSA probe	[6-FAM]AAAAAAAAAAAXGAAAAAAAAAAA,X=mC
o2968	polyA	competitor	AAAAAAAAAAAAAAAAAAAAAAA
o2969	polyT	competitor	TTTTTTTTTTTTTTTTTTTTTTT
o3112	CG_T probe	EMSA probe	TTTTTTTTTTTXGTTTTTTTTTTT,X=hmC
o3115	CG_A probe	EMSA probe	[6-FAM]AAAAAAAAAAAXGAAAAAAAAAAA,X=hmC
o4123	spike-in oligo	qPCR	[PHOS]TCAGCCTTTCATTGATTGCG
o4124	spike-in quantification	qPCR	CTTCTCCTTTACTAGTGAATTC
o4368	spike-in quantification	qPCR	CTCTTCTGCCTGCTGACCTTTG
o4371	mC spike in fw	adapter	CTTTCATTGATTGCGGATTCCAGAATTCACTAG TAAAGGAGAAGTTGGCTACAGCAA
o4372	mC spike in rv	adapter	ACCACCCTGTTGCTGTAGCCAA
o4373	C spike in fw	adapter	CTTTCATTGATTGCGCACGATAGAATTCACTAG TAAAGGAGAAGGAACCGCTCATTG
o4374	C spike in rv	adapter	CACCATTGGCAATGAGCGGTTC
o4392	hmC spike in fw	adapter	CTTTCATTGATTGCGTCACAGAGAATTCACTAG TAAAGGAGAAGTGTAGCCCTCTGT
o4393	hmC spike in rv	adapter	CTTGAGCACACAGAGGGCTACA
o4450	8NC rv	template	CTTCTCCTTTACTACTCAATTCNNNNCGNNNNG GGAATGAATGAAAGGGTGA
o4627	reference for o4372	qPCR	CTCTTCTGCCTGCTGACCTTTGTCAGCCTTTCA TTGATTGCGGATTCCAGAATTCACTAGTAAAGG AGAAGTTGGCTACAGCAA
o4628	reference for o4374	qPCR	CTCTTCTGCCTGCTGACCTTTGTCAGCCTTTCA TTGATTGCGCACGATAGAATTCACTAGTAAAGG AGAAGGAACCGCTCATTG
o4629	reference for o4393	qPCR	CTCTTCTGCCTGCTGACCTTTGTCAGCCTTTCA TTGATTGCGTCACAGAGAATTCACTAGTAAAGG AGAAGTGTAGCCCTCTGT
o4666	VEGFA rv	PCR primer	AGTGACCCCTGGCCT
o4667	LINE1 fw	PCR primer	TCCCAGCGTGAGCGA
o4668	LINE1 rv	PCR primer	CGCCTTGCAGTTTGATCTCA
o4669	MUC1 fw	PCR primer	GACAGGGAGCGGTTAGAAG
o4670	MUC rv	PCR primer	GGTCCTGTACAGCAAGGT

Identifier	Description	Purpose	Sequence 5′ ->3′
o4675	Carrier hmC fw	Spike-in	TCTTCXGTTTCCTCAGCXGAAGGCTCGAGTCTT CXGTTTCCAAGCTTCAGCXGAAGGCTCTTCTGC CTGCTGACCTTTG,X=hmC
o4677	Carrier mC fw	Spike-in	TCTTCXGTTTCCTCAGCXGAAGGCTCGAGTCTT CXGTTTCCAAGCTTCAGCXGAAGGCTCTTCTGC CTGCTGACCTTTG,X=mC
o4681	VEGFA fw	PCR primer	TTTCCAAAGCCCATTCCCT
o4727	Carrier mC rv	Spike-in	CAAAGGTCAGCAGGCAGAAGAGCCTTXGGCTGA AGCTTGGAAAXGGAAGACTCGAGCCTTXGGCTG AGGAAAXGGAAGA,X=mC
o4728	Carrier C fw	Spike-in	TCTTCCGTTTCCTCAGCCGAAGGCTCGAGTCTT CCGTTTCCAAGCTTCAGCCGAAGGCTCTTCTGC CTGCTGACCTTTG
o4729	Carrier C rv	Spike-in	CAAAGGTCAGCAGGCAGAAGAGCCTTCGGCTGA AGCTTGGAAACGGAAGACTCGAGCCTTCGGCTG AGGAAACGGAAGA
o4736	MUC1 fw	PCR primer	[BtnTg][6-FAM]GACAGGGAGCGGNTAGAAG
o4737	LINE1 fw	PCR primer	[BtnTg][6-FAM]TCCCAGCGNGAGCGA
o4743	MUC1 rv	PCR primer	[Cy5]GGTCCTGTACAGCAAGGT
o4744	LINE1 rv	PCR primer	[Cy5]CGCCTTGCAGTTTGATCTCA
o4745	VEGFA fw	PCR primer	[BtnTg][6-FAM]TTTCCAAAGCCCANTCCCT
o4746	VEGFA rv	PCR primer	[Cy5]AGTGACCCCTGGCCT
o5071	8NmC rv	template	CTTCTCCTTTACTACTCAATTCNNNNXGNNNNG GGAATGAATGAAAGGGTGA,X=mC
o5186	0 mm t7 fw	T7E1 assay	TCACCCTTTCATTCATTCCCGGGGCGGGGGGAA TTGAGTAGTAAAGGAGAAG
o5187	0 mm t7 rv	T7E1 assay	CTTCTCCTTTACTACTCAATTCCCCCCGCCCCG GGAATGAATGAAAGGGTGA
o5195	8 mm t7 fw	T7E1 assay	TCACCCTTTCATTCATTCCCCCCCCGCCCCGAA TTGAGTAGTAAAGGAGAAG
o5201	0 mm endo fw	T7E1 assay	TCACCCTTTCATTCATTCCCAAAACGAAAAGAA TTGAGTAGTAAAGGAGAAG
o5202	0 mm endo rv	T7E1 assay	CTTCTCCTTTACTACTCAATTCTTTTCGTTTTG GGAATGAATGAAAGGGTGA
o5221	1 mm endo fw	T7E1 assay	TCACCCTTTCATTCATTCCCAATACGAAAAGAA TTGAGTAGTAAAGGAGAAG
o5222	c biot fw	T7E1 assay	[BtnTg]TCACCCTTTCATTCATTCCCNNNNCG NNNNGAATTGAGTAGTAAAGGAGAAG
o5224	mc fw biot	T7E1 assay	[BtnTg]TCACCCTTTCATTCATTCCCNNNNXG NNNNGAATTGAGTAGTAAAGGAGAAG,X=mC
o5232	Sp1 fw	PCR primer	CTGGGCGGAAACCAAGTACG
o5233	Sp1 rv	PCR primer	GGAAAAACGCGGACGCTGAC
o5241	8NX probe	template	TCACCCTTTCATTCATTCCCTCCACACAACAAC CATTCTTNNNNCGNNNNAATGTGAGGAGGGTGT TATAGAATTGAGTAGTAAAGGAGAAG
o5242	8NX fw	PCR primer	TCACCCTTTCATTCATTCCCTCCACACAACAAC CATTC

Identifier	Description	Purpose	Sequence 5′ ->3′
o5243	8NX rv	PCR primer	CTTCTCCTTTACTACTCAATTCTATAACACCCT CCTCAC
o5246	12NX probe	template	TCACCCTTTCATTCATTCCCTCCACACAACAAC CATTCTTNNNNCGNNNNCGNNNNAATGTGAGGA GGGTGTTATAGAATTGAGTAGTAAAGGAGAAG
o5247	16NX probe	template	TCACCCTTTCATTCATTCCCTCCACACAACAAC CATTCTTNNNNCGNNNNCGNNNNCGNNNNAATG TGAGGAGGGTGTTATAGAATTGAGTAGTAAAGG AGAAG
o5488	VEGFA probe	template	TTTCCAAAGCCCATTCCCTCTTTAGCCAGAGCC GGGGTGTGCAGACGGCAGTCACTAGGGGGCGCT CGGCCACCACAGGGAAGCTGGGTGAATGGAGCG AGCAGCGTCTTCGAGAGTGAGGACGTGTGTGTC TGTGTGGGTGAGTGAGTGTGTGCGTGTGGGGTT GAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGTC ACT
o5489	MUC1 probe	template	GACAGGGAGCGGTTAGAAGGGTGGGGCTATTCC GGGAAGTGGTGGAAGGAGGGAGCCCAAAACTAG CACCTAGTCCACTCATTATCCAGCCCTCTTATT TCTCGGCCCCGCTCTGCTTCAGTGGACCCGGGG AGGGCGGGGAAGTGGAGTGGGAGACCTAGGGGT GGGCTTCCCGACCTTGCTGTACAGGACC
o5490	LINE1 probe	template	TCCCAGCGTGAGCGACGCAGAAGACGGTGATTT CTGCATTTCCATCTGAGGTACCGGGTTCATCTC ACTAGGGAGTGCCAGACAGTGGGCGCAGGCCAG TGTGTGTGCGCACCGTGCGCGAGCCGAAGCAGG GCGAGGCATTGCCTCACCTGGGAAGCGCAAGGG GTCAGGGAGTTCCCTTTCCGAGTCAAAGAAAGG GGTGACGGACGCACCTGGAAAATCGGGTCACTC CCACCCGAATATTGCGCTTTTCAGACCGGCTTA AGAAACGGCGCACCACGAGACTATATCCCACAC CTGGCTCGGAGGGTCCTACGCCCACGGAATCTC GCTGATTGCTAGCACAGCAGTCTGAGATCAAAC TGCAAGGCG
o5522	VEGFA fw	PCR primer	[BtnTg]TTTCCAAAGCCCATTCCCT[Cy3]
o5688	VEGFA split 1 fw	EMSA probe	[6-FAM]TTTCCAAAGCCCATTCCCTCTTTAGC CAGAGCCGGGGT
o5689	VEGFA split 1 rv	EMSA probe	ACCCCGGCTCTGGCTAAAGAGGGAATGGGCTTT GGAAA
o5690	VEGFA split 2 fw	EMSA probe	[6-FAM]GAGCCGGGGTGTGCAGACGGCAGTCA CTAGGGGGCGCT
o5691	VEGFA split 2 rv	EMSA probe	AGCGCCCCCTAGTGACTGCCGTCTGCACACCCC GGCTC
o5692	VEGFA split 3 fw	EMSA probe	[6-FAM]AGGGGGCGCTCGGCCACCACAGGGAA GCTGGGTGAATG

Identifier	Description	Purpose	Sequence 5′ ->3′
o5693	VEGFA split 3 rv	EMSA probe	CATTCACCCAGCTTCCCTGTGGTGGCCGAGCGCCCCCT
o5694	VEGFA split 4 fw	EMSA probe	[6-FAM]TGGGTGAATGGAGCGAGCAGCGTCTTCGAGAGTGAGGA
o5695	VEGFA split 4 rv	EMSA probe	TCCTCACTCTCGAAGACGCTGCTCGCTCCATTCACCCA
o5696	VEGFA split 5 fw	EMSA probe	[6-FAM]CGAGAGTGAGGACGTGTGTGTCTGTGTGGGTGAGTGAG
o5697	VEGFA split 5 rv	EMSA probe	CTCACTCACCCACACAGACACACACGTCCTCACTCTCG
o5698	VEGFA split 6 fw	EMSA probe	[6-FAM]GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGG
o5699	VEGFA split 6 rv	EMSA probe	CCAACGCCCTCAACCCCACACGCACACACTCACTCACC
o5700	VEGFA split 7 fw	EMSA probe	[6-FAM]TTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGTCACT
o5701	VEGFA split 7 rv	EMSA probe	AGTGACCCCTGGCCTTCTCCCCGCTCCAACGCCCTCAA
o5906	E-box 1 fw	EMSA probe	[6-FAM]TAGGCCAXGTGGGAGG,X=hmC
o5907	E-box 1 rv	EMSA probe	CCTCCCAXGTGGCCTA,X=hmC
o5908	E-box 2 fw	EMSA probe	[6-FAM]TAGGCCAXGCGGGAGG,X=hmC
o5909	E-box 2 rv	EMSA probe	CCTCCCGXGTGGCCTA,X=hmC
o5910	E-box 3 fw	EMSA probe	[6-FAM]TAGGCCAXATGGGAGG,X=hmC
o5911	E-box 4 fw	EMSA probe	[6-FAM]TAGGCGAXGTGGGAGG,X=hmC
o5912	E-box 4 rv	EMSA probe	CCTCCCAXGTCGCCTA,X=hmC
o5913	E-box 5 fw	EMSA probe	[6-FAM]TAGGCTGXGTGGGAGG,X=hmC
o5914	E-box 5 rv	EMSA probe	CCTCCCAXGCAGCCTA,X=hmC
o5917	E-box 1 fw	EMSA probe	[6-FAM]TAGGCCACGTGGGAGG
o5918	E-box 1 rv	EMSA probe	CTCCCACGTGGCCTA
o5919	E-box 2 fw	EMSA probe	[6-FAM]TAGGCCACGCGGGAGG
o5920	E-box 2 rv	EMSA probe	CCTCCCGCGTGGCCTA
o5921	E-box 3 fw	EMSA probe	[6-FAM]TAGGCCACATGGGAGG
o5922	E-box 4 fw	EMSA probe	[6-FAM]TAGGCGACGTGGGAGG
o5923	E-box 4 rv	EMSA probe	CCTCCCACGTCGCCTA
o5924	E-box 5 fw	EMSA probe	[6-FAM]TAGGCTGCGTGGGAGG
o5925	E-box 5 rv	EMSA probe	[6-FAM]CCTCCCACGCAGCCTA
o5926	E-box 1 fw	EMSA probe	[6-FAM]TAGGCCAXGTGGGAGG,X=mC
o5927	E-box 1 rv	EMSA probe	CCTCCCAXGTGGCCTA,X=mC
o5928	E-box 2 fw	EMSA probe	[6-FAM]TAGGCCAXGCGGGAGG,X=mC
o5929	E-box 2 rv	EMSA probe	CCTCCCGXGTGGCCTA,X=mC
o5930	E-box 3 fw	EMSA probe	[6-FAM]TAGGCCAXATGGGAGG,X=mC
o5931	E-box 4 fw	EMSA probe	[6-FAM]TAGGCGAXGTGGGAGG,X=mC
o5932	E-box 4 rv	EMSA probe	CCTCCCAXGTCGCCTA,X=mC
o5933	E-box 5 fw	EMSA probe	[6-FAM]TAGGCTGXGTGGGAGG,X=mC
o5934	E-box 5 rv	EMSA probe	CCTCCCAXGCAGCCTA,X=mC
o5935	E-box 3 rv	EMSA probe	CCTCCCATGTGGCCTA
o6026	X-box 2 fw	EMSA probe	[6-FAM]CCTAGTTGCCXGGCAACCCCC,X=hmC
o6027	X-box 2 rv	EMSA probe	GGGGGTTGCXGGGCAACTAGG,X=hmC

Identifier	Description	Purpose	Sequence 5′ ->3′
o6028	X-box 1 fw	EMSA probe	[6-FAM]CCTGATGAXGACGTACCG,X=hmC
o6029	X-box 1 rv	EMSA probe	CGGTACGTXGTCATCAGG,X=hmC
o6051	X-box 2 fw	EMSA probe	[6-FAM]CCTAGTTGCCCGGCAACCCCC
o6052	X-box 2 rv	EMSA probe	GGGGGTTGCCGGGCAACTAGG
o6053	X-box 1 fw	EMSA probe	[6-FAM]CCTGATGACGACGTACCG
o6054	X-box 1 rv	EMSA probe	CGGTACGTCGTCATCAGG
o6055	X-box 2 fw	EMSA probe	[6-FAM]CCTAGTTGCCXGGCAACCCCC,X=mC
o6056	X-box 2 rv	EMSA probe	GGGGGTTGCXGGGCAACTAGG,X=mC
o6057	X-box 1 fw	EMSA probe	[6-FAM]CCTGATGAXGACGTACCG,X=mC
o6058	X-box 1 rv	EMSA probe	CGGTACGTXGTCATCAGG,X=mC

Table 4.9: Oligonucleotides used for CDS amplification of full-length proteins and primers used for cloning of truncated proteins.

Identifier	Purpose	Sequence 5′ ->3′
o5527	ATF1 fw	GAAAATCTTTATTTTCAGTCTCTCATGGAAGATTCCCACAAGAGTAC
o5528	ATF1 rv	CAGTGGTGGTGGTGGTGGTGCTCAACACTTTTATTGGAATAAAGATCCTTC
o5529	ATF2 fw	GAAAATCTTTATTTTCAGTCTCTCATGAAATTCAAGTTACATGTGAATTCTG
o5530	ATF2 rv	CAGTGGTGGTGGTGGTGGTGCTCACTTCCTGAGGGCTGTG
o5539	CHAF1B fw	GAAAATCTTTATTTTCAGTCTCTCATGAAAGTCATCACTTGTGAAATAGC
o5540	CHAF1B rv	CAGTGGTGGTGGTGGTGGTGCTCAGGGTCCAGACTTTCCG
o5545	FOXA1 fw	GAAAATCTTTATTTTCAGTCTCTCATGTTAGGAACTGTGAAGATGGAAG
o5546	FOXA1 rv	CAGTGGTGGTGGTGGTGGTGCTCGGAAGTGTTTAGGACGGGTC
o5555	L3MBTL3 fw	GAAAATCTTTATTTTCAGTCTCTCATGACTGAATCTGCCTCTAGC
o5556	L3MBTL3 rv	CAGTGGTGGTGGTGGTGGTGCTCAAGTTCATTGTGAGAATTCTTCTCTG
o5557	MAX fw	GAAAATCTTTATTTTCAGTCTCTCATGAGCGATAACGATGACATCG
o5558	MAX rv	CAGTGGTGGTGGTGGTGGTGCTCGCTGGCCTCCATCCG
o5561	MYC fw	GAAAATCTTTATTTTCAGTCTCTCATGCCCCTCAACGTTAGC
o5562	MYC rv	CAGTGGTGGTGGTGGTGGTGCTCCGCACAAGAGTTCCGTAG
o5567	RFX5 fw	GAAAATCTTTATTTTCAGTCTCTCATGGCAGAAGATGAGCCTG
o5568	RFX5 rv	CAGTGGTGGTGGTGGTGGTGCTCTGGGGGTGTTGCTTTTGG
o5571	RFXANK fw	GAAAATCTTTATTTTCAGTCTCTCATGGAGCTTACCCAGCC
o5572	RFXANK rv	CAGTGGTGGTGGTGGTGGTGCTCCTCAGGGTCAGCGGG
o5579	SUB1 fw	GAAAATCTTTATTTTCAGTCTCTCATGCCTAAATCAAAGGAACTTGTTTC
o5580	SUB1 rv	CAGTGGTGGTGGTGGTGGTGCTCCAGTTTTCTTACTGCATCATCAATG
o5591	TFAM fw	GAAAATCTTTATTTTCAGTCTCTCATGGCGTTTCTCCGAAGC
o5592	TFAM rv	CAGTGGTGGTGGTGGTGGTGCTCACACTCCTCAGCACCATATTTTC
o5938	MBP removal fw	ATCGAGGGAAGGCTCGAAAATC
o5939	MBP removal rv	CGCCAGCGCCACTTTATC
o5980	bHLH MYC fw 1	ATGGTCAAGAGGCGAACACACAAC
o5981	bHLH both rv 1	GAGCCTTCCCTCGATCGC
o5982	bHLH MAX fw 1	GCTGACAAACGGGCTCATC
o5983	bHLH both fw 2	GAGCACCACCACCACCACCACT
o5984	bHLH MYC rv 2	TAGCTGTTCAAGTTTGTGTTTCAACTGTTCTCG
o5985	bHLH MAX rv 2	CAGTGCACGGACTTGCTGCTCC
o6035	$RFX5_{DBD}$ fw 1	[PHOS]GAGCACCACCACCACCACCACTG
o6036	$RFX5_{DBD}$ rv 1	GGTCTTCCTCCTTATGCCACTGTAGC
o6037	$RFX5_{DBD}$ fw 2	[PHOS]GGAGACAAAAGCTCAGAGCCAAGT
o6038	$RFX5_{DBD\text{-}MBP}$ rv	GAGAGACTGAAAATAAAGATTTTCGAGCCTTCC

Identifier	Purpose	Sequence 5′ ->3′
o6039	$RFX5_{DBD}$ rv 2	CGCCAGCGCCACTTTATCATC

4.3.4 Plasmids

All plasmids listed here have an encoded ampicillin resistance and a pET21d(+) backbone.

Table 4.10: List of all plasmids used in this study.

Identifier	Construct
p1379	MBP–His_6
p1385	MBP-MBD2-His_6
p2774	GST-MBP-MECP2 WT
p2775	GST-MBP-MECP2 HM
p3597	MBP-ATF1-His_6
p3598	MBP-ATF2-His_6
p3599	MBP-CHAF1B-His_6
p3600	MBP-FOXA1-His_6
p3601	MBP-L3MBTL3-His_6
p3602	MBP-MAX_{FL}-His_6
p3603	MBP-MYC-His_6
p3604	MBP-RFX5-His_6
p3605	MBP-RFXANK-His_6
p3606	MBP-SUB1-His_6
p3607	MBP-TFAM-His_6
p3622	MYC_{FL}-His_6
p3623	MAX_{FL}-His_6
p3636	MAX_{bHLHZ}-His_6
p3637	MYC_{bHLHZ}-His_6
p3675	$RFX5_{DBD}$-His_6
p3676	MBP-$RFX5_{DBD}$-His_6

4.4 Lab equipment, Consumables and Software

4.4.1 Lab equipment

Table 4.11: List of lab equipment used in this study.

Equipment	Model	Supplier
Bunsen Burner	1040/1	Carl Friedrich Usbeck KG
C18 columns	Acclaim™ PepMap™ 100 C18-HPLC-Säulen	Thermo Scientific™
Cabinet Station PCR/MS	PCR Workstation Pro	Peqlab
Camera	PowerShot G10	Canon
Camera Chamber	BDA digital	Biometra
Centrifuge	Mini Centrifuge ROTILABO®	Carl Roth
Centrifuge benchtop	5810 R	Eppendorf
Centrifuge benchtop	5424	Eppendorf

Equipment	Model	Supplier
Centrifuge benchtop with cooling	5424 R	Eppendorf
Chromatography system	ÄKTA Purifier 10 FPLC™ system	GE Healthcare
Concentrator	Concentrator Plus	Eppendorf
Dewar flask	Dewar flask cylindric	KGW Isotherm
Dounce homogenizer	Homogenisator Dounce, 2 ml	Carl Roth
Fluorometer	Quantus™	Promega
Freezer -20 °C	Premium GGU 1500	Liebherr
Freezer -20 °C	ProfiLine GG 4010	Liebherr
Freezer -80 °C	New Brunswick TM HEF® U410	Eppendorf
Gel casting system	Mini-PROTEAN® Tetra Handcast Systems	Bio-Rad
Glass Beads	Glass beads	VWR
Heating Block	ThermoStat™ plus	Eppendorf
HPLC system	UltiMate 3000 RSLCnano System	Thermo Scientific™
Humidified incubator	CellXpert® C170i	Eppendorf
Ice Machine	Scotsman AF20	Fisher Scientific
Imaging system	ChemiDoc	Bio-Rad
Incubator	INE 600	Memmert GmbH
Incubator Shaker	New Brunswick™ I26	Eppendorf
Laser Scanner	Typhoon™ FLA 9500	GE Healthcare
Magnetic rack	MagRack6	Cytiva
Magnetic stirrer	RCT classic	IKA-Werke
Mass spectrometer	Q Exactive™ Plus Hybrid Quadrupol-Orbitrap™ Massenspektrometer	Thermo Scientific™
Microfiltration unit	Bio-Dot® apparatus	Bio-Rad
Microwave	Tecnolux ED 8525 exquisit	Verbeken & Fils
PCR cycler	T-Personal	Biometra
PCR cycler	SimpliAmp™	Thermo Scientific™
pH Meter	FiveEasy™ F20	Mettler-Toledo
Pipette 0.1-2.5 μL	Research plus	Eppendorf
Pipette 0.5-10 μL	Research plus	Eppendorf
Pipette 10-100 μL	Research plus	Eppendorf
Pipette 100-1000 μL	Research plus	Eppendorf
Pipette 20-200 μL	Research plus	Eppendorf
Plate Reader	Infinite® M1000	Tecan
Plate Reader	EnVision®	PerkinElmer
Power supply	PowerPac™ Basic	Bio-Rad
Real time PCR system	CFX384 Touch™	Bio-Rad
Refrigerttor 2-8 °C	ProfiLine FKU 1800	Liebherr
Scale	PM400	Mettler-Toledo
Scale	M-Pact AX224	Sartorius
Scanner	CanonScan 9000F	Canon
Shaker, orbital	Unimax 1010	Heidolph
Sonicator	Bioruptor® Plus	Diagenode

Equipment	Model	Supplier
Sonicator	Digital Sonifier 450 Cell Disruptor	Branson
Spectrophotometer UV-Vis	Nanodrop™ 2000	Thermo Scientific™
TapeStation	4150 TapeStation System	Agilent
Thermomixer	ThermoMixer® F	Eppendorf
Turning wheel	Loopster Digital	IKA-Werke
Vacuum pump	VNC 2	Vacuubrand
Vortex Mixer	Vortex-Genie 2	Scientific Industries
Western blot transfer system	Trans-Blot Turbo Transfer System	Bio-Rad

4.4.2 Consumables

Table 4.12: List of consumables used in this study.

Consumable	Catalog number	Supplier
384-well lightcycler plate	72.1985.202	Sarstedt
96-well plate, clear, for BCA	15045	Thermo Scientific™
Balance tray, 41 x 41 mm	1-1124	Neolab
Balance tray, 89 x 89 mm	1-1125	Neolab
Cuvettes	67.742	Sarstedt
D1000 ScreenTapes	5067-5582	Agilent
Desalting Tips Pierce™ C18 100 μL	87784	Thermo Scientific™
Dialysis tube 12-14 kDa MWCO	2719.1	Roth
Filter tips, low retention, 0.1 - 10 μL	732-3249	VWR
Filter tips, low retention, 1 - 200 μL	732-3253	VWR
Filter tips, low retention, 100 - 1,000 μL	732-3254	VWR
Glass beads	MARI4901005	VWR
Gloves, nitrile	816781633	Semperit
Microcentrifuge tubes 1.5 mL	72.706	Sarstedt
Microcentrifuge tubes 2 mL	72.691	Sarstedt
Microcentrifuge tubes DNA LowBind, 1.5 mL	72.706.700	Sarstedt
Microcentrifuge tubes Protein LoBind, 1.5 mL	022431081	Eppendorf
Microtubes for Bioruptor® Pico, 1.5 mL	C30010016	Diagenode
Nitrocellulose membrane	10600002	Cytvia
Paper boxes	95.64.981	Sarstedt
Paper towels, Kimtech	AA64.2	Carl Roth
Parafilm®	PM-966	Bemis
PCR tubes, 0.2 mL	72.737.005	Sarstedt
PCR tubes, 0.2 mL, Multiply® Pro	72.991.002	Sarstedt
Petri dishes	82.1473.001	Sarstedt
Pipette tips, 10 μL	70.1130	Sarstedt
Pipette tips, 1000 μL	70.3050.020	Sarstedt
Pipette tips, 200 μL	70.760.002	Sarstedt
Plate seals, adhesive	600238	Biozym
PVDF membrane	#12023954	Bio-Rad
Scalpel	5518075	B. Braun
Serological pipette 10 mL	86.1254.001	Sarstedt
Serological pipette 25 mL	86.1685.001	Sarstedt

Consumable	Catalog number	Supplier
Slide-A-Lyzer® dialysis cassettes (3.5 kDa MWCO)	66330	Thermo Scientific™
TC plate	83.3902	Sarstedt
Transfer membrane, nitrocellulose	10600002	Cytvia
Transfer membrane, PVDF, 7.1 x 8.5 cm	10026934	Bio-Rad
Tubes conical 15 mL	62.554.502	Sarstedt
Tubes conical 50 mL	62.547.254	Sarstedt
Waste bags, 200 x 300 mm	E706.1	Carl Roth
Waste bags, 700 x 1120 mm	86.1204	Sarstedt

4.4.3 Software and Tools

Data processing and instrument control

- NanoDrop 2000 Software v1.6 (Thermo Scientific™)
- CFX Manager Software (Bio-Rad)
- Typhoon FLA 9500 Control Software (Cytvia)
- Image Lab Software (Bio-Rad)

Molecular biology and sequence analysis

- SnapGene v4.3 (GSL Biotech LLC)
- NEB Tm Calculator v1.16 (NEB)

Proteomics data analysis

- MaxQuant v2.2.0.0
- Perseus v2.0.7.0 / v2.0.11.0
- VolcaNoseR (web application)
- R (R Core Team)
- RStudio (Posit PBC)

Image analysis

- ImageQuant TL v8.1 1D Gel Analysis (Cytvia)
- ImageJ (Wayne Rasband (NIH))

Visualization

- PyMOL (Schrödinger)

5 Methods

5.1 Molecular biology techniques

5.1.1 Bacterial cells

Bacterial cell cultivation *E. coli* cells were cultivated in LB broth or on LB agar plates at 37 °C supplemented with the appropriate antibiotics. Unless otherwise indicated, liquid cultures were grown aerobically with shaking at 180-220 rpm. For expression cultures, the incubation temperature was adjusted as described in the corresponding section.

Transformation Plasmid DNA was amplified by transformation. Chemically competent *E. coli* GH371 cells were slowly thawed on ice, after which plasmid DNA was added. The mixture was incubated on ice for 20 min, followed by a heat shock at 42 °C for 45 s. Cells were then cooled on ice for 2 min and 500 μL S.O.C. medium was added. The suspension was incubated at 37 °C with shaking at 600 rpm for 45 min. 100 μL of either undiluted transformed bacteria or concentrated bacteria (obtained by centrifugation for 3 min at 500 x g and supernatant discarded) was plated on selective growth media containing 50 μg/mL carbenicillin and incubated overnight at 37 °C .

Plasmid isolation. Plasmids were isolated from *E. coli* using the NucleoSpin® Plasmid EasyPure kit.

5.1.2 Cloning of potential hit proteins

Cloning of MBP-fusion proteins.

To clone protein CDSs into expression vectors, the corresponding DNA sequence was amplified from cDNA libraries, *DpnI* digested, and cloned into the desired expression vector via Gibson assembly. Subsequently, plasmids were transformed into *E. coli* for amplification.

Insert amplification from cDNA libraries. cDNA templates derived from human thyroid or prostate tissue were used to amplify CDSs corresponding to the following proteins: MYC (aa1-439, [NP_AAA36340.1]), MAX (aa1-160, [NP_001394023.1]), RFX5 (aa1-616, [NP_000440.1]), RFXANK (aa1-260, [NP_001357162.1]), L3MBTL3 (aa1-780, [NP_115814.1]), SUB1 (aa1-127, [NP_006704.3]), FOXA1 (aa1-472, [NP_004487.2]), ATF1 (aa1-505, [NP_001243019.1]), TFAM (aa1-246, [NP_003192.1]), and CHAF1b (aa1-559, [NP_005432.1]). Primer sequences including overhangs compatible with subsequent Gibson assembly are listed in Table 4.8. PCR amplifications were performed in a total volume of 50 μL. The reaction mixture contained the following components:

Component	Volume (µl)	Final Concentration
5x Phusion® Buffer	10.0	1x
dNTP mix	1.0	0.2 mM each
Primer forward	2.5	0.5 µM
Primer reverse	2.5	0.5 µM
DMSO	1.5	3 %
Template cDNA	1.0	100 ng
Phusion® Polymerase	0.5	1.0 U
Nuclease-free water	31.0	—

PCR amplification was carried out using the following thermal cycling program:

	Step	Temperature	Time
1	Initial denaturation	98 °C	30 s
2	Denaturation	98 °C	10 s
3	Annealing	62 °C	30 s
4	Extension	72 °C	1 min
5	Final extension	72 °C	10 min
6	Hold	4 °C	∞

Steps 2–4 were repeated for a total of 35 cycles. Following amplification, the cDNA template was removed by adding 1 μL *DpnI* to each 50 μL PCR reaction and incubating at 37 °C for 3 h. PCR products were analyzed on 1 % agarose gels. Reactions yielding a single band of the expected insert size were purified using the NucleoSpin® Gel and PCR Clean-Up kit. For reactions showing multiple bands, the band corresponding to the expected insert size was excised and purified using the NucleoSpin® Gel and PCR Clean-Up kit.

Gibson assembly. Inserts of the correct size were cloned into a *XhoI*-digested pET-21d(+) vector containing a N-terminal MBP tag and a C-terminal His tag by Gibson assembly (see Figure A.9a). The restriction digest was carried out by incubating 1 μg of vector p1379 with 2 μL *XhoI* and 1 μL rSAP in 1x rCutsmart™ buffer in a total volume of 30 μL for 1 h at 37 °C with subsequent heat inactivation of the enzyme for 20 min at 80 °C . The molar insert to vector ratio was 3:1 in a total volume of 5 μL which were mixed with 15 μL of 1.33x Gibson master mix. The Gibson assembly reaction was allowed to proceed at 50 °C for 1 h. 5 μL of each assembly were transformed into *E. coli* GH371 (see Subsection 5.1.1), plated on LB agar containing carbenicillin and incubated at 37 °C overnight.

Colony PCR. The plasmid was tested for its integrity by colony PCR using primers annealing to backbone regions flanking the insert resulting in larger amplicons when the insert is present. For colony PCR, a single colony was resuspended in 10 μL nuclease-free water and the following PCR reaction was prepared in a total volume of 25 μL:

Component	Volume (µl)	Final Concentration
10x Thermopol Buffer	2.5	1x
dNTP mix	0.5	0.2 mM each
Primer o1565	0.5	0.2 µM
Primer o1590	0.5	0.2 µM
Resuspended colony	3.0	-
Taq Polymerase	0.25	1.0 U
Nuclease-free water	17.75	-

PCR amplification was carried out using the following thermal cycling program:

	Step	Temperature	Time
1	Initial denaturation	95 °C	5 min
2	Denaturation	95 °C	30 s
3	Annealing	46 °C	30 s
4	Extension	68 °C	2 min
5	Final extension	68 °C	5 min
6	Hold	4 °C	∞

Steps 2-4 were repeated for a total of 30 cycles. PCR products were analyzed on 1 % agarose gels. Confirmed colonies containing the correct insert were inoculated in LB broth containing carbenicillin and grown anaerobically at 200 rpm at 37 °C overnight. Plasmids were purified using the NucleoSpin® Plasmid EasyPure kit and the DNA sequence was validated by Sanger sequencing.

MYC, MAX and RFX5 cloning.

MBP removal from MAX and MYC. MAX_{FL} and MYC_{FL} were obtained by removing MBP from MAX (aa1-160) and MYC (aa1-439) performing an inverse PCR with a total volume of 50 μL as follows:

Component	Volume (µl)	Final Concentration
10x HF Buffer	10	1x
dNTP mix	1	0.2 mM each
Primer o5938	2.5	0.2 µM
Primer o5939	2.5	0.2 µM
Template (p3602 or p3603)	1	1 ng
Phusion® Polymerase	0.5	1.0 U
Nuclease-free water	32.5	-

PCR amplification was carried out using the following thermal cycling program:

	Step	Temperature	Time
1	Initial denaturation	98 °C	30 s
2	Denaturation	98 °C	10 s
3	Annealing	64 °C	30 s
4	Extension	72 °C	2 min
5	Final extension	72 °C	10 min
6	Hold	4 °C	∞

Steps 2-4 were repeated for a total of 30 cycles. 1 μL *DpnI* was added to the PCR reaction and incubated for 3 h at 37 °C . The PCR product was purified using the NucleoSpin® Gel and PCR Clean-Up kit. 1 μg of purified PCR product was mixed with 2 μL of 10x T4 polynucleotide kinase buffer, 0.8 μL 25 mM ATP and 1 μL T4 polynucleotide kinase in a total volume of 20 μL. The reaction proceeded for 30 min at 37 °C with subsequent heat inactivation for 20 min at 65 °C . Ligation was carried out by mixing 5 μL of phosphorylated product with 1 μL 10x T4 ligase buffer, 1 μL T4 DNA ligase and 3 μL nuclease free water at 16 °C overnight with subsequent heat inactivation at 65 °C for 10 min. 5 μL of he ligated product was transformed into *E. coli* GH371.

MYC_{bHLHZ} and MAX_{bHLHZ} cloning. MYC_{bHLHZ} and MAX_{bHLHZ} were cloned by first removing the N-terminal and then the C-terminal part of p3622 and p3623. PCR reactions were carried out with the same protocol used above for removal of MBP from MYC and MAX. Primers for removing the N-terminal part were o5980 and o5981 for MYC, and o5982 and o5981 for MAX. The PCR products were mixed with 1 μL *DpnI* and incubated for 3 h at 27 °C and 5 μL were subsequently transformed into *E. coli* GH371. Resulting colonies were sequenced and plasmids with that showed successful removal of

the N-terminal part were used as templates for the second PCR. The same PCR protocol was followed using the primers o5983 and o5984 for MYC and o5983 and o5985 for MAX. PCR was followed by *DpnI* digest and transformation into *E. coli* GH371.

$RFX5_{DBD}$ cloning. $RFX5_{DBD}$ was cloned according to Chakraborty, Bhattacharya, et al., 2010. Cloning of $RFX5_{DBD}$ (aa75-168) with or without MBP was achieved in two steps. First, the C-terminal part of RFX5 (aa1-616) (p3604) was removed using the same PCR protocol as used above for MYC and MAX cloning and primers o6035 and o6036. The N-terminal part was subsequently removed using the primers o6037 and o6038. In each case, *DpnI* digest was carried out after PCR, followed by ligation with T4 ligase at 16 °C overnight and transformation of 5 μL of the ligated product into *E. coli* GH371.

5.2 Nucleic acid-based techniques

5.2.1 General procedures

PCR clean-up. PCR products were purified using the NucleoSpin® Gel and PCR Clean-Up kit.

Ethanol precipitation. Up to 300 μL of DNA solution was transferred into a 1.5 mL microcentrifuge tube. 1 μL of glycogen was added as a carrier, followed by 3 M sodium acetate (pH 5.2) at 1/10 of the DNA solution volume. After mixing, three volumes of ice-cold 100 % ethanol were added and the solution was inverted several times to ensure homogeneity, then incubated at –80 °C overnight. Following incubation, DNA was pelleted by centrifugation at 16,000 × g for 15 min at 4 °C . The supernatant was discarded and the pellet was washed with 500 μL of 70 % ice-cold ethanol. After gentle mixing, the sample was centrifuged again at 16,000 × g for 5 min at 4 °C . The supernatant was discarded and the pellet was air-dried at room temperature (RT) for 5 min avoiding overdrying. Finally, the DNA pellet was resuspended in 20 μL of nuclease-free water and incubated at 37 °C for 5 min.

Agarose gel electrophoresis. Agarose gels of 0.8 - 2.0 % were used to analyse DNA probes and PCR products. They were run in 0.5x TBE buffer with 6-12 V/cm. Gels were stained in 0.5 μg/mL EtBr and destained in water. The DNA was visualized by UV fluorescence and documented using a camera.

Gel extraction. To extract DNA from an agarose gel, the corresponding band was cut out and DNA was purified using the NucleoSpin® Gel and PCR Clean-Up kit.

gDNA isolation. Genomic DNA (gDNA) was isolated using the Monarch Genomic DNA Purification Kit according to the manufacturer's instructions.

gDNA fragmentation. gDNA at a concentration of 10 ng/μL in 1x TE was sheared to an average fragment size of 200 bp using a Bioruptor Pico sonication device. Fragmentation was performed for 19 cycles of 30 s on, 30 s off at 4 °C . Sheared DNA was purified by ethanol precipitation overnight at -80 °C . Fragment size distribution was confirmed using a TapeStation using D1000 ScreenTapes.

Concentration determination by Nanodrop and Quantus. DNA concentration was determined using a NanoDrop™ spectrophotometer by recording absorbance at 260 nm with purity assessment from the A_{260}/A_{280} and A_{260}/A_{230} ratios. For quantifying dsDNA only, a Quantus™ Fluorometer was employed, using the QuantiFluor® dsDNA System, following the manufacturer's instructions.

Sanger sequencing. DNA sequences after PCR or cloning were checked by Sanger sequencing by Microsynth Seqlab GmbH.

DNA hybridization. If not indicated otherwise, oligonucleotides were hybridized in equimolar ratios or with 1.5-1.8x molar excess of unlabeled oligonucleotide over labeled oligonucleotide either in hybridization buffer or IDT buffer by heating for 5 min at 95 °C , followed by gradual cooling to RT over night in a Dewar flask containing boiling water.

Glucosylation of hydroxymethylcytosine. 1 μg of sheared gDNA was used for one glucosylation reaction. Reactions were carried out in 1x Cutsmart buffer supplemented with 40 μM UDP-Glucose and 20 U T4-BGT in a total volume of 40 μL and incubated at 37 °C for 16 h. Subsequently, DNA was purified using the Monarch PCR & DNA Purification kit following the manufacturers protocol for fragments < 2 kbp. DNA concentration was measured with Quantus Fluorometer.

5.2.2 DNA probe amplification

DNA probes used in the pull-down assays were amplified by PCR. The protocols corresponding to the different probes are listed in the following paragraphs.

VEGFA probe. PCR reactions to amplify the VEGFA probe were performed in a total volume of 50 μL. The reaction mixture contained the following components:

Component	Volume (µl)	Final Concentration
10x DreamTaq™ Buffer	5.0	1x
dNTP mix	1.0	0.2 mM each
Primer o4666	2.5	0.5 µM
Primer o4681	2.5	0.5 µM
DMSO	2.5	5 %
Template DNA	5.0	500 pg
DreamTaq™ DNA Polymerase	0.25	1.25 U
Nuclease-free water	31.25	—

The dNTP mix contained either unmodified dCTPs, or mdCTP or hmdCTP for modified PCR products. PCR amplification was carried out using the following thermal cycling program:

	Step	Temperature	Time
1	Initial denaturation	95 °C	3 min
2	Denaturation	95 °C	1 min
3	Annealing	59 °C	30 s
4	Extension	72 °C	15 s
5	Final extension	72 °C	5 min
6	Hold	4 °C	∞

Steps 2–4 were repeated for a total of 25 cycles.

MUC1 and LINE1 probes. The conditions for the MUC1 and LINE1 PCRs differed from the VEGFA PCR only in the primers (o4669 and o4670 for MUC1, o4667 and o4668 for LINE1), templates (o5489 for MUC1, o5490 for LINE1) and annealing temperatures (61 °C for MUC1, 63 °C for LINE1).

Sp1 probe. PCR reactions to amplify the Sp1 probe were performed in a total volume of 50 μL. The reaction mixture contained the following components:

Component	Volume (µl)	Final Concentration
10x DreamTaq™ Buffer	5.0	1x
dNTP mix	1.0	0.2 mM each
Primer o5232	2.5	0.5 µM
Primer o5233	2.5	0.5 µM
DMSO	2.5	5 %
Template DNA	1.0	1 ng
DreamTaq™ DNA Polymerase	0.25	1.25 U
Nuclease-free water	35.25	—

The dNTP mix contained either unmodified dCTPs, or mdCTP or hmdCTP for modified PCR products. PCR amplification was carried out using the following thermal cycling program:

	Step	Temperature	Time
1	Initial denaturation	95 °C	3 min
2	Denaturation	95 °C	30 s
3	Annealing	61 °C	30 s
4	Extension	72 °C	30 s
5	Final extension	72 °C	5 min
6	Hold	4 °C	∞

Steps 2–4 were repeated for a total of 25 cycles.

8NX, 12NX, and 16NX probes. The randomized CG core probes contain two flanking sequences adjacent to the random parts, which are complementary to the PCR primers. PCR reactions to amplify the 8NX, 12NX or 16NX probe were performed in a total volume of 50 μL. The reaction mixture contained the following components:

Component	Volume (µl)	Final Concentration
10x DreamTaq™ Buffer	5.0	1x
dNTP mix	1.0	0.2 mM each
Primer o5242	2.5	0.5 µM
Primer o5243	2.5	0.5 µM
Template DNA	1.0	0.4 nM
DreamTaq™ DNA Polymerase	0.25	1.25 U
Nuclease-free water	37.75	—

As templates, o5241, o5246, and o5247 were used for 8NX, 12NX, and 16NX, respectively. The dNTP mix contained either unmodified dCTPs, or mdCTP or hmdCTP for modified PCR products. PCR amplification was carried out using the following thermal cycling program:

	Step	Temperature	Time
1	Initial denaturation	95 °C	3 min
2	Denaturation	95 °C	1 min
3	Annealing	65 °C	30 s
4	Extension	72 °C	15 s
5	Final extension	72 °C	5 min
6	Hold	4 °C	∞

Steps 2–4 were repeated for a total of 25 cycles.

5.2.3 Modified DNA probe preparation

DNA probes were typically prepared as described in Subsection 5.2.2. However, this procedure only allowed to generate symmetrically modified probes. In this part, first the general steps considered for the creation of asymmetric probes are described, followed by the asymmetric probe generation protocol and DNA hybridization and digestion studies.

Techniques for modified asymmetric DNA probes

Modified dCTP incorporation assay. Taq DNA Polymerase, DreamTaq™ DNA Polymerase and Phusion® DNA Polymerase were tested for their ability to incorporate modified dCTPs into the MUC1, LINE1, and VEGFA sequences. PCR reactions and thermo cycling protocols were used each with unmodified dNTP mixes or mixes that contained mdCTP, hmdCTP, or cadCTP instead of dCTP.
The following reaction mixture was prepared for Taq DNA Polymerase in a total volume of 50 μL:

Component	Volume (µl)	Final Concentration
10x ThermoPol Buffer	5.0	1x
dNTP mix	1.0	0.2 mM each
Primer forward	2.5	0.5 µM
Primer reverse	2.5	0.5 µM
DMSO	1.5	3 %
Template DNA	1.0	<1000 ng
Taq DNA Polymerase	0.25	1.25 U
Nuclease-free water	36.25	—

The following thermal cycling program was used. Annealing temperatures differed for different templates, 53 °C was chosen for MUC1, 51 °C for VEGFA, and 53 °C for LINE1.

	Step	Temperature	Time
1	Initial denaturation	95 °C	30 s
2	Denaturation	95 °C	30 s
3	Annealing	indicated above	30 s
4	Extension	68 °C	1 min
5	Final extension	68 °C	5 min
6	Hold	4 °C	∞

Steps 2-4 were repeated for a total of 12 cycles.

The following reaction mixture was prepared for DreamTaq™ DNA Polymerase in a total volume of 50 μL:

Component	Volume (µl)	Final Concentration
10x DreamTaq™ Buffer	5.0	1x
dNTP mix	1.0	0.2 mM each
Primer forward	2.5	0.5 µM
Primer reverse	2.5	0.5 µM
DMSO	1.5	3 %
Template DNA	0.2	2 ng
DreamTaq™ DNA Polymerase	0.25	1.25 U
Nuclease-free water	36.25	—

The following thermal cycling program was used. Annealing temperatures differed for different templates, 56 °C was chosen for MUC1, 57 °C for VEGFA, and 57 °C for LINE1.

	Step	Temperature	Time
1	Initial denaturation	95 °C	3 min
2	Denaturation	95 °C	30 s
3	Annealing	indicated above	30 s
4	Extension	72 °C	1 min
5	Final extension	72 °C	10 min
6	Hold	4 °C	∞

Steps 2-4 were repeated for a total of 12 cycles.

The following reaction mixture was prepared for Phusion® DNA Polymerase in a total volume of 50 μL:

Component	Volume (µl)	Final Concentration
5x HF Buffer	10.0	1x
dNTP mix	1.0	0.2 mM each
Primer forward	2.5	0.5 µM
Primer reverse	2.5	0.5 µM
DMSO	1.5	3 %
Template DNA	0.5	5 ng
Phusion® DNA Polymerase	0.5	1.25 U
Nuclease-free water	36.25	—

The following thermal cycling program was used. Annealing temperatures differed for different templates, 62 °C was chosen for MUC1 and VEGFA, and 63 °C for LINE1.

	Step	Temperature	Time
1	Initial denaturation	98 °C	30 s
2	Denaturation	98 °C	30 s
3	Annealing	indicated above	30 s
4	Extension	72 °C	15 s
5	Final extension	72 °C	10 min
6	Hold	4 °C	∞

Steps 2-4 were repeated for a total of 12 cycles.
All PCR products were purified using the NucleoSpin® Gel and PCR Clean-Up kit and 5 μL of each reaction was applied to a 2 % agarose gel.

Immobilization of DNA probes on streptavidin-coated beads. DNA probes were biotinylated using a biotinylated PCR primer during PCR. Immobilization was performed with Dynabeads™ MyOne™ Streptavidin T1 magnetic beads according to the manufacturer's protocol. The desired bead volume was transferred to a 1.5 mL tube and washed three times with an equal volume of 1x B&W buffer. After 1 min the tube was placed on a magnetic rack to discard the supernatant. The beads were then resuspended in 2x B&W buffer and mixed with an equal volume of biotinylated DNA to yield a final concentration of 1x B&W buffer. DNA immobilization was carried out for 1-3 h at RT on a turning wheel. Binding efficiency was assessed by analyzing the supernatant on an agarose gel or, for fluorescently labeled probes, by fluorescence measurement on a Tecan reader.

NaOH melting. To determine the optimal NaOH concentration for separating the DNA strands, a titration series was carried out using fluorescently labeled probes. Forward strands were FAM-labeled and biotinalyted, while reverse strands were Cy5 labeled (PCR primers: o4737 and o4744 for LINE1, o4745 and o4746 for VEGFA and o4743 and o4736 for MUC1). Fluorescently labeled probes were immobilized on beads and NaOH concentrations ranging from 2.5 mM to 150 mM were applied for 10 min at RT on a turning wheel. Supernatants were collected, adjusted to 1x TE and neutralized with an equal concentration of HCl. Cy5 fluorescence was quantified on a Tecan reader and total strand yield was determined from a standard curve generated with known DNA concentrations.

Asymmetric probe generation

The VEGFA probe (o5488) was amplified via PCR using Dreamtaq™ DNA Polymerase with the biotinylated forward primer o4681 and the reverse primer o4666. 16 pmol of PCR product was immobilized on 20 μL Dynabeads™ MyOne™ Streptavidin T1 in 1x B&W buffer for 3 h at RT on a turning wheel. The beads were washed twice with 1x B&W buffer and then resuspended in 50 μL 20 mM NaOH. After incubation on a turning wheel for 10 min at RT, the supernatant containing ssDNA was separated from the beads with a magnetic rack and then added to a new tube containing a final concentration of 1x TE and 20 mM HCl. The ssDNA was hybridized with 1.5x molar excess of o4681 in 1x hybridization buffer. The annealing mixture was heated to 95 °C and then gradually cooled down to RT in a Dewar filled with boiling water. The primer was extended with Klenow fragment in 1x NEB2 buffer using 0.1 mM dNTPs including the respective modified dCTP for 3 h at 37 °C . The final product was purified with the NucleoSpin® Gel and PCR Clean-Up kit and checked on an 2 % agarose gel.

DNA digestion and hybridization

Restriction enzyme digest. Restriction enzyme (RE) digests to compare probes generated with PCR and with asymmetric primer extension were carried out in a total volume of 10 μL. 20 ng of DNA probe

was incubated with 1 μL rCutsmart^{TM} buffer and 0.5 μL of RE (*MspI* for MUC1, *NheI* for LINE1, and *HphI* for VEGFA). The reactions proceeded for 30 min at 37 °C . Samples were immediately analyzed by agarose gel electrophoresis using 2 % gels.

Mismatch assay of 8NX probes. 8NX probes with and without deliberate mismatches were hybridized by mixing forward and reverse oligonucleotides at 5 μM in hybridization buffer in a total volume of 50 μL. After heating for 5 min at 95 °C , samples were gradually cooled to RT over night in a Dewar flask containing boiling water. Oligonucleotides used for hybridization are listed in Table 4.8. The concentration was determined in a 1:20 dilution by absorption at 260 nm. For T7 endonuclease I digest, 2 μL of enzyme were mixed with 500 ng of DNA probe and 2 μL 10x NEB2 buffer, while for mismatch endonuclease I digest, 2 μL of enzyme were mixed with 500 ng of DNA probe and 2 μL 10x NEB2.1. Samples were adjusted with nuclease-free water to a total volume of 20 μL. Reactions were incubated for 2 h at 37 °C and then immediately analyzed by agarose gel electrophoresis using a 2 % gel.

5.2.4 qPCR

Spike-in probe preparation for qPCR

Preparation of spike-in probes was carried out as previously described in Buchmuller et al., 2022. DNA duplexes carrying modified CpG dyads ("carriers") were ligated to adapter sequences containing general and unique primer binding sites for qPCR quantitation. Carriers were prepared by annealing modified oligonucleotides (C/C: o4728 and o4729, mC/mC: o4677 and o4727, hmC/mC: o4675 and o4727) in hybridization buffer. Annealing was carried out by heating to 95 °C for 5 min followed by gradual cooling in a Dewar flask containing boiling water. Adapters were annealed at a final concentration of 2.5 μM in the same way (o4371 and o4372, o4373 and o4374, and o4392 and o4393) and subsequently extended with 25 mU/μL Klenow fragment in the presence of 0.1 mM dNTPs for 20 min at 37 °C . To introduce 5'-phosphorylation, primer o4123 was annealed to the extended adapters and further extended with Klenow fragment for 30 min at 37 °C . 50 μL of adapters were ligated with 15 pmol of carriers at 16 °C overnight using 200 U T4 DNA ligase. Ligated Spike-in probes were purified using the Macherey-Nagel PCR purification kit, with the NTI buffer diluted to 16 % (v/v) with double distilled water. Probe concentration was determined by qPCR against reference standards (o4628 for o4374, o4627 for o4372 and o4629 for o4393). qPCR primers o4368 and o4124 were premixed as a 2 μM stock. Each reaction contained a 3 μL primer mix, 5 μL 2x primaQuant SYBRGreen Master Mix with ROX and 2 μL of spike-in probe. Cycling conditions were 95 °C for 2 min, followed by 50 cycles of 95 °C for 10 s and 60 °C for 30 s.

Spike-in recovery determination by qPCR.

Following enrichment with MECP2 WT or MECP2 HM, spike-in probe recovery was determined by qPCR against a standard dilution series of spike-in probes with known concentrations ranging from $2*10^{-6}$ to 2 nM. To distinguish individual spike-ins, three different primer pairs were used (C/C: o4374 and o4368, mC/mC: o4372 and o4368, hmC/mC: o4393 and o4368), which were premixed as a 2 μM stock. qPCR was performed on all fractions (flow-through, wash1, wash2, and elution). Each reaction contained 3 μL of corresponding primer mix, 5 μL 2x primaQuant SYBRGreen Master Mix with ROX and 2 μL of fraction to be analyzed. Reactions were carried out by initial denaturation at 95 °C for 2 min, followed by 50 cycles of 95 °C for 10 s and 60 °C for 30 s.

5.2.5 DNA immunoprecipitation

1 μg of sheared gDNA was used as input for one DNA immunoprecipitation (DIP) reaction. DIPs were performed according to the MagMeDIP-seq Package V2 protocol, using either the provided anti-5mC antibody for MeDIP or 0.6 μg per IP of the mouse monoclonal anti-hmC antibody for hMeDIP. Illumina libraries were prepared according to the Diagenode MeDIP-seq library preparation protocol. Final library concentrations were quantified using the Quantus Fluorometer and fragment size distributions were assessed with a TapeStation using a D1000 ScreenTape. Libraries were sequenced on a Illumina NovaSeq X-25B platform using paired-end 2x 150 bp reads with a target depth of 50 million reads per sample.

5.2.6 NGS

End repair and A-tailing. 1 μg of fragmented DNA, either glucosylated or non-glucosylated, was subjected to end repair and adapter ligation using the NEBNext Ultra II DNA Library Preparation Kit. The manufacturer's protocols for end preparation and adapter ligation were followed (version 6.1_5/20). DNA purification and size selection was carried out using NEBNext Sample Purification Beads according to the protocol with bead volumes corresponding to a target fragment size of 200 bp. Elution was carried out as described in the manual and DNA concentration was determined with the Quantus Fluorometer. Fragment size distribution was verified using a TapeStation with D1000 ScreenTapes.

Library preparation. Library preparation was carried out using the NEBNext Ultra II DNA Library Prep Kit for Illumina (Version 6.1_5/20) following the manufacturer's protocol. Enriched DNA fragments were PCR amplified using NEBNext Multiplex Oligos. As amplification was conducted after enrichment and not directly after adapter ligation, DNA input concentrations were treated as threefold higher for determining the appropriate number of PCR cycles, which were determined to be 10 cycles. PCR clean-up was performed according to the "Cleanup of PCR Reaction" protocol (version 6.1_5/20) using NEBNext Sample Purification Beads. Fragment length distribution of the final libraries was assessed with a Tapestation with a D1000 ScreenTape according to the Agilent protocol. dsDNA concentration was measured with the Quantus Fluorometer. Libraries were pooled in equimolar ratios fulfilling the requirements of a library concentration of >2 ng/μL and >70 μL total volume. They were sequenced on a NovaSeq X-25B (Illumina) using paired-end reads, with a target depth of 50 million reads per sample.

Data analysis. Analysis of paired-end sequencing libraries was performed using a standardized, automated Snakemake workflow. The raw read quality was assessed with FastQC with subsequent adapter removal and trimming of low-quality bases using Trim Galore. Cleaned reads were aligned to the *Mus muculus* reference genome (mm10) with Bowtie2. Properly paired reads were kept and PCR duplicates removed with Samtools. Peak calling to identify enriched regions was performed on deduplicated BAM files with MACS2. Technical reproducibility was evaluated by calling peaks independently for each of the three technical replicates together with their respective input controls within each biological replicate. Reproducibility was assessed by generating consensus peak sets across the three independent replicates and retaining only those peaks present in at least two replicates using bedtools. For visualization, normalized BigWig tracks were generated with bamCoverage (deepTools v3.5.0) and final alignments were examined in the Integrative Genomics Viewer (IGV). Statistical significance of overlap between consensus peaks and annotated genomic features was assessed with the Genomic Association Tester (GAT). Additional downstream analyses were conducted with bedtools, deepTools, custom Bash and R scripts.

5.3 Cell biology and protein biochemistry techniques

5.3.1 Mammalian cell culture

HEK293T and Hela cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % (v/v) fetal bovine serum (FBS), 2 mM/L glutamine, 100 U/mL penicillin, and 0.1 mg/mL streptomycin. Cells were cultured in a humidified incubator ($\geq$ 95 % relative humidity) at 37 °C and 5 % CO_2. Cells were passaged every 3-4 days at a dilution ratio of 1:20 or 1:30. Therefore, the medium was removed and cells were washed once with Dulbecco's Phosphate Buffered Saline (DPBS). 1 mL trypsin-EDTA was added and cells were incubated for 5 min at 37 °C until detached. Fresh DMEM was added to a final volume of 10 mL and cells were resuspended by gentle pipetting. An aliquot corresponding to the desired dilution was transferred into a new culture plate containing pre-warmed DMEM.

E14tg2a mouse embryonic stem cells were obtained from Dr. Christian Schröter. The cells were grown in dishes precoated with 0.1 % gelatin (w/v) in GMEM supplemented with 10 % FBS, sodium pyruvate, 50 μM β-mercaptoethanol, glutamax, non-essential amino acids and 10 ng/mL murine leukemia inhibitory factor (LIF, Protein Expression Facility,MPI Dortmund). Cells were passaged every 2-3 days to maintain cultures between 10 % and 90 % confluency. For analysis, near-confluent cultures were released from culture vessels with trypsin, spun down, and snap-frozen in liquid nitrogen before further processing.

5.3.2 Protein extraction from cells and tissues

Nuclear protein extraction from cells

Cells were washed with 1 mL DPBS and detached with 1 mL trypsin-EDTA for 5 min at 37 °C . Cells were harvested by centrifugation at 300 × g for 5 min, washed twice with ice-cold 1× PBS and resuspended in 7 volumes of ice-cold cell lysis buffer (CLB) containing protease inhibitor cocktail. After incubation on ice for 10 min, 0.5 % IGEPAL was added, and the suspension was shortly vortexed and then pipetted 50 times using a 200 μL tip. The lysate was centrifuged at 3000 × g for 20 min at 4 °C and the resulting pellet was washed twice with CLB. The nuclear pellet was resuspended in nuclear extraction buffer (NEB) containing protease inhibitor at half the volume used for CLB. The suspension was incubated at 4 °C and vortexed every 10 min for an overall duration of 50 min. Samples were then centrifuged at 16,000 x g for 20 min and the supernatant was collected, aliquoted, snap-frozen in liquid N_2 and stored at -80 °C . Protein concentrations were determined by BCA assay.

Nuclear protein extraction from tissue

Female C57BL6J mice (16 weeks old) were sacrificed and brains were harvested and provided by Dr. Cristina Cadenas (IfaDo). Brain tissue was kept on dry ice and cut into small pieces using a sterile scalpel. Around 50 mg of brain were used for one round of extraction. Tissue was rinsed with PBS and transferred to a pre-cooled dounce homogenizer with 10 μL/mg of buffer A. Homogenization was performed with 15 strokes with pestle A. Then 0.5 % IGEPAL was added and additional 15 strokes with pestle A were applied. The homogenate was incubated for 10 min on ice to promote cell lysis. The lysate was then transferred to microcentrifuge tubes and centrifuged at 3000 x g for 10 min at 4 °C . The supernatant that contained the cytoplasmic fraction was removed. The nuclear pellet was washed twice with 200 μL of buffer A. The pellet was then resuspended in buffer C at half the volume of buffer A and incubated at 4 °C with vortexing every 10 min for an overall duration of 50 min. Centrifugation was carried out at 13,000 x g for 20 min and the supernatant containing the nuclear protein extract was

aliquoted, snap-frozen in liquid N_2 and stored at -80 °C . Protein concentrations were determined by BCA assay.

5.3.3 Protein detection and quantification

BCA-Assay. Protein concentrations were determined using BCA assay according to the manufacturer's instructions. Briefly, protein samples and bovine serum albumin (BSA) standards were incubated with BCA working reagent in duplicates, and absorbance was measured at 562 nm using a Tecan reader. Protein concentrations were calculated from the standard curve.

SDS-PAGE. Proteins were separated by SDS polyacrylamide gel electrophoresis (PAGE). Protein samples were mixed with 4x SDS sample buffer and heated at 95 °C for 5 min. 12-20 μL of sample were loaded onto polyacrylamide gels (12-15 % resolving gel, 4 % stacking gel). SDS gels were prepared using the Bio-RadMini-PROTEAN® system, pouring first the resolving gel and after polymerization pouring the stacking gel. The resolving gel was prepared with either 12 % (2.15 mL double distilled water, 1.5 mL Rotiphorese Gel 40 per gel) or 15 % (1.78 mL double distilled water, 1.88 mL Rotiphorese Gel 40 per gel) acrylamid concentration, adding 1.25 mL 1.5 M Tris (pH 8.8), 50 μL 10 % SDS solution, 50 μL 10 % APS and 5 μL TEMED. The stacking gel had a final acrylamide concentration of 5 % mixing 1 mL double distilled water with 208 μL Rotiphorese Gel 40, 420 μL 0.5 M Tris (pH 6.8), 16.7 μL 10 % SDS solution, 16.7 μL 10 % APS and 1.67 μL TEMED for one gel. Electrophoresis was performed in SDS running buffer at 220 V for around 45min or until the dye front reached the bottom of the gel. For protein band visualization, gels were stained with Coomassie Brilliant Blue staining solution for 20 min followed by incubation in destaining solution for up to multiple hours and further destaining in double-distilled water. Gels were documented with an image scanner. The staining step was omitted for gels intended for Western blotting.

Western blot. Proteins separated by SDS-PAGE were transferred onto polyvinylidene difluoride (PVDF) membranes using the Trans-Blot®Turbo semi-dry transfer system following the manufacturers "MIXED MW*" protocol for mini gels (1.3 A constant, 7 min). Membranes were blocked in 5 % (w/v) non-fat dry milk in Tris-buffered saline with 0.1 % Tween-20 (TBST) for at least 1 h at RT with gentle agitation. Primary antibodies diluted in 1.25 % (w/v) non-fat dry milk in TBST were incubated with the membranes overnight at 4 °C with gentle agitation. Membranes were washed three times for 5 min with TBST and then incubated with HRP-conjugated secondary antibodies diluted in 2.5 % (w/v) non-fat dry milk in TBST for 1 h at RT with gentle agitation. The membranes were washed three times for 5 min in TBST and protein bands were subsequently visualized using enhanced chemiluminescence (ECL) and imaged with the Bio-Rad ChemiDoc imaging system.

Dot blot. 50 μL of protein sample in PBS buffer were applied to a nitrocellulose membrane using gravity filtration in a Bio-Dot® apparatus). The wells were rinsed twice with TBST using vacuum until dry. The blotted membrane was taken out from the apparatus and blocked for at least 1 h using 5 % (w/v) non-fat dry milk in TBST. Afterwards the membrane was treated in the same way as described for western blots (Subsection 5.3.3).

5.3.4 Recombinant protein expression and purification

Recombinant protein expression

Expression of candidate proteins from proteomics. Expression plasmids encoding the MBP-fusion candidate proteins (MBP-MBD2, MBP-ATF1, MBP-ATF2, MBP-CHAF1B, MBP-FOXA1, MBP-L3MBTL3,

MBP-MAX, MBP-MYC, MBP-RFX5, MBP-RFXANK, MBP-SUB1, and MBP-TFAM, see Table 4.10) were transformed into *E. coli* BL21-Gold(DE3). Fresh overnight cultures of single clones were then diluted to an optical density (OD_{600}) of 0.05 in 30 mL LB broth supplemented with 50 μg/mL carbenicillin, 1 mM $MgCl_2$ and 1 mM $ZnSO_4$ and grown at 37 °C until they reached an OD_{600} of 0.5–0.6. Expression was induced with 1 mM IPTG and cultures were incubated overnight at 25 °C (for MBP-ATF2 and MBP-CHAF1B) or 30 °C (for MBP-FOXA1, MBP-MAX, MBP-MYC, MBP-RFX5, MBP-RFXANK, and MBP-SUB1) shaking at 150 rpm. Cells were harvested and washed once by resuspension in ice-cold 20 mM Tris-HCl (pH 8.0) before they were lysed by sonication (3 min, amplitude 15 %, 4 s on, 2 s off) in 2 mL lysis buffer. After centrifugation at 14,000 x g for 20 min at 4 °C , the cleared lysates were retained, and either snap-frozen in liquid nitrogen and stored at -80 °C or further purified.

Expression of MECP2 WT and MECP2 HM. GST-tagged MBP-MECP2 WT (p2774) and MBP-MECP2 HM (p2775) were recombinantly expressed in *E. coli* BL21-Gold(DE3). 300 mL LB supplemented with 1 mM $MgCl_2$, 1 mM $ZnSO_4$ and carbenicillin was inoculated from a fresh overnight culture and grown at 37 °C with shaking at 220 rpm to an OD_{600} of 0.5-0.6. Cultures were cooled on ice and protein expression was induced by the adding 1 mM IPTG. The expression proceeded overnight at 30 °C with shaking at 150 rpm. The cultures were harvested at 8,000 x g for 20 min at 4 °C , washed twice with 50 mL cold 20 mM Tris-HCl (pH 8.0) and the resulting pellet was frozen until further processing.

Protein purification using affinity chromatography

Purification of candidate proteins. Candidate proteins were expressed as described in Subsection 5.3.4 and 500 μL of lysates were diluted with 1.5 mL of binding buffer and added to 200 μL 50 % Ni-nitriloacetic acid (NTA) HisPur agarose resin that were prior equilibrated in binding buffer. Samples were incubated for 2 min at 16 °C shaking at 700 rpm. The resins were washed twice with 1.5 mL binding buffer containing 90 mM imidazole for 2 min at 16 °C shaking at 700 rpm and the protein was eluted two times 0.2 mL and one time 0.4 mL binding buffer with 500 mM imidazole for 2 min at 16 °C shaking at 700 rpm. Fractions with sufficient purity judged by SDS-PAGE were combined and dialyzed three times against 1 L dialysis buffer in a tube with 12-14 kDa MWCO. The protein concentration was determined with BCA assay and the proteins were snap-frozen in liquid nitrogen and stored at -80 °C .

Purification of MAX_{FL}, MAX_{bHLHZ}, $RFX5_{FL}$, and $RFX5_{DBD}$. Expression was carried out as described in Subsection 5.3.4 using the expression plasmids for MAX_{FL} (p3602), MAX_{bHLHZ} (p3636), $RFX5_{DBD}$ (p3675), and MBP-$RFX5_{DBD}$ (p3676). For purification, harvested cells were resuspended in 2 mL binding buffer before 1 mM PMSF was added. Resuspended cells were sonicated (3 min, amplitude 15 %, 4 s on, 2 s off) and subsequently incubated with 1 mg/mL lysozyme and 10 U DNAseI overnight at 4 °C . After centrifugation at 14,000 x g for 20 min at 4 °C , the cleared supernatants were retained and incubated at 16 °C for 20 min shaking at 700 rpm with 450 μL 50 % Ni-nitriloacetic acid (NTA) HisPur agarose resin that were prior equilibrated in binding buffer. The resins were washed twice with 1 mL binding buffer containing 90 mM imidazole for 5 min at 16 °C shaking at 700 rpm and the fusion protein was eluted in two times 0.2 mL and one time 0.4 mL binding buffer with 500 mM imidazole for 5 min at 16 °C shaking at 700 rpm. Fractions with sufficient purity judged by SDS-PAGE were combined and dialyzed three times against 1 L dialysis buffer in a tube with 12-14 kDa MWCO. The protein concentration was determined with BCA assay and the proteins were snap-frozen in liquid nitrogen and stored at -80 °C .

Purification of MYC_{FL} and MYC_{bHLHZ}. Expression was carried out as described in Subsection 5.3.4. For purification of MYC_{FL} (p3603) and MYC_{bHLHZ} (p3637), the cell pellet was resuspended in 2 mL urea buffer similarly as described in Jung et al., 2017. Cell lysis was achieved by sonication (3 min,

amplitude 15 %, 4 s on, 2 s off) prior to centrifugation for 20 min at 16,000 x g at 4 °C . The supernatant was adjusted to a final concentration of 750 mM NaCl. The adjusted supernatant was incubated at 4 °C for 1 h shaking at 700 rpm with 450 μL 50 % Ni-nitriloacetic acid (NTA) HisPur agarose resin that were prior equilibrated in urea buffer. The resins were washed with 5 mL urea wash buffer. A second wash step was carried out using the same buffer with 100 mM NaCl. The protein was eluted by stepwise elution with urea wash buffer (five times 1 mL) with 100 mM NaCl and 250 mM imidazole. Fractions with sufficient purity judged by SDS-PAGE were combined and dialyzed against three times 500 mL of 10 mM Tris-HCl (pH8 7.8), 100 mM NaCl, 1 % Triton X-100, 0.1 mM PMSF and 7 M urea in a tube with 12-14 kDa MWCO. The protein concentration was determined with BCA assay and the proteins were snap-frozen in liquid nitrogen and stored at -80 °C .

Purification of MECP2 WT and MECP2 HM. To lyse the cells, pellets were resuspended in 20 mL binding buffer supplemented with 1 mM PMSF. The suspension was treated with 0.1 mg/ml lysozyme and 1 U/mL DNase I and incubated overnight at 4 °C . Cells were disrupted by two rounds of sonication on ice (3 min, amplitude 20 %, 4 s on, 30 s off). Centrifugation at 14,000 x g for 20 min at 4 °C removed cellular debris and the cleared supernatant was retained and purified over a Ni-NTA column on an Äkta Purifier 10 FPLC system running an imidazole gradient of 10 mM to 500 mM in binding buffer. Fractions containing pure protein judged by SDS-PAGE were combined and dialyzed three times against dialysis buffer using Slide-A-Lyzer dialysis cassettes (3.5 kDa MWCO). The protein concentration was determined with BCA assay and the proteins were snap-frozen in liquid nitrogen and stored at -80 °C at a concentration of 15 μM.

5.4 Protein-DNA interaction techniques

5.4.1 Pull-down assay implementation

Pull-down assays were initially optimized with the purpose to use them for MS experiments. Optimization was done in multiple steps including the use of a MBD2 benchmark protein, testing different DNA competitor sequences and titrating the amount of nuclear extracts. Pull-down assays were also used for initial reader protein validation by using MBP-tagged proteins in the pull-down which were detected in Dot blots and Western blots.

Pull-down optimizations

MBP-MBD2 benchmark. 5 μL of Dynabeads^{TM} MyOne^{TM} Streptavidin T1 magnetic beads were incubated with 4 pmol biotinylated VEGFA probe in 1x B&W buffer for 1 h at RT. Afterwards, the beads were washed once with 1x B&W buffer followed by two washes with Protein binding buffer (PBB) containing 1x protease inhibitor. The beads were then incubated with 0.1, 1, or 10 nM of MBP-MBD2(aa145-225, p1385) for 4 h at 4 °C . Following incubation, the beads were washed twice with 1x PBS and proteins were eluted by boiling the beads in 15 μL PBS and 5 μL 4x SDS sample buffer for 5 min at 95 °C . Beads were separated by using a magnetic rack and 18 μL of the sample was applied to a 15 % SDS-PAGE with subsequent Western blotting using a MBP antibody at a 1:10,000 dilution.

Nuclear extract titration. The pull-down was conducted as described in Subsection 5.4.1 using 1 nM MBP-MBD2 and 36, 360, or 3600 nM nuclear extracts.

DNA competitor titration. The pull-down was conducted as described in Subsection 5.4.1 using 1 nM MBP-MBD2 and 3600 nM nuclear extracts. For the dA:dT competitor DNA, o2968 and o2969 were

hybridized at 30 μM in hybridization buffer by boiling for 5 min at 95 °C and subsequently gradually cooling in a dewar flask filled with boiling water. dA:dT competitor DNA was added to the pull-down reaction in amounts ranging from 500 ng to 20 μg.

Pull-down assays performed with randomized core CG probes. The pull-down was conducted as described in Subsection 5.4.1 using 1 nM MBP-MBD2, 3600 nM nuclear extracts and varying concentrations of randomized core CG probes. The probes were prepared using the asymmetric probe generation protocol (Subsection 5.2.3) and applied in ranged of 4 to 96 pmol.

Pull-down of nuclear proteins with modified DNA probes for MS experiments

5 μL of Dynabeads^{TM} MyOne^{TM} Streptavidin T1 magnetic beads were incubated with 4 pmol biotinylated DNA probe in 1x B&W buffer for 3 h at RT. Afterwards, the beads were washed once with 1x B&W buffer followed by two washes with PBB containing 1x protease inhibitor. The beads were then incubated with 50 μg (3600 nM) nuclear extracts and 5 μg (1.1 μM) dA:dT competitor DNA in PBB in a total volume of 300 μL on a turning wheel overnight at 4 °C . Following incubation, the beads were washed twice with 1x PBS and stored in 20 μL PBS until further processing. In parallel to the pull-down assays carried out for MS analysis, one pull-down assay with MBP-MBD2 spike-in was performed intended for Western blotting as a quality control as described in Subsection 5.4.1 using 1 nM MBP-MBD2, 3600 nM nuclear extracts and 1.1 μM dA:dT competitor.

5.4.2 Pull-down of candidate proteins

E. coli lysates containing the MBP-tagged candidate protein or purified MBP-tagged candidate protein was used. Pull-downs were performed as described in Subsection 5.4.1 using 1.67 μL of Dynabeads^{TM} MyOne^{TM} Streptavidin T1 magnetic beads to immobilize 1.33 pmol of VEGFA DNA probe which was subsequently incubated with 0.01 or 0.1 μL of *E. coli* lysates or 1 nM purified protein with 16.7 μg of nuclear extracts in the presence of 1.67 μg dA:dT competitor DNA. Pull-down samples were boiled in 15 μL PBS and 5 μL 4x SDS sample buffer at 95 °C for 5 min. 10 μL of these samples were mixed with 40 μL PBS and then analyzed using Dot blot.

5.4.3 EMSA

General EMSA procedure

FAM labeled DNA probes were obtained either by asymmetric probe generation as described in Subsection 5.2.3 using a FAM labeled forward primer and unlabeled reverse primer or by hybridization of 1.5 μM labeled forward strand and 1.8 μM unlabeled reverse strand oligonucleotides (oligonucleotides are listed in Table 4.8) in IDT buffer by heating to 95 °C and then gradually cooling to RT in a Dewar flask filled with boiling water. MBP-fusion proteins were either directly used or the MBP tag was cleaved by incubation with 0.25 μM TEV-protease overnight at 4 °C . Varying concentrations of protein were incubated with 2 nM labeled DNA probe in EMSA buffer in presence of 5 μM dA:dT competitor DNA and 1 mM DTT in a total volume of 15 μL for 20 min at 21 °C . Samples were mixed with 3 μL 6x EMSA loading buffer and loaded to a non-denaturing polyacrylamide gel. EMSA gels were prepared using the Bio-Rad Mini-PROTEAN® system. For a 12 % gel 6.4 mL double distilled water was mixed with 0.5 mL 10x TBE (pH 7.5), 3.0 mL Rotiphorese Gel 40, 10 μL TEMED and 100 μL 10 % APS. 12 % gels were used for EMSA experiments involving VEGFA probes, while 15 % gels were used E-box and X-box probes. Gels were stored at 4 °C after polymerization and pre-run at 200 V for at least 2 h in 0.25x TBE prior to loading. 10 μL of sample was loaded per well and gels were run at 240 V at 4 °C for

60-75 min, depending on DNA probe length. Visualization was performed using the 510 LP filter of the Typhoon FLA-9500 laser scanner at 473 nm with 800-1000 V amplification.

Specific EMSA experiments

Selectivity and affinity determination. EMSAs to determine selectivity of a protein to different C modifications, were carried out with at least two different protein concentrations in duplicate. The EMSA was carried out as described in Subsection 5.4.3. Bound and unbound DNA probes were quantified using ImageQuant TL v8.1 1D Gel Analysis and data were plotted in bar diagrams with error bars.

For affinity determination a protein concentration range was applied ranging from 1 nM up to 2048 nM and experiments were performed at least in duplicates. The EMSA was carried out as described in Subsection 5.4.3. Bound and unbound DNA probes were quantified using ImageQuant TL v8.1 1D Gel Analysis and the R script provided in appendix A.3 was used for determining the putative K_D values.

MYC/MAX dimerization optimization. Heterodimerization of MAX and MYC proteins was optimized by pre-incubating the proteins for 30 min at 21 °C in ratios ranging from 1:3 to 1:50 (MAX:MYC) in varying protein concentrations, testing MAX_{FL} and MAX_{bHLHZ} each with MYC_{FL} and MYC_{bHLHZ}. The putative dimer was subsequently added to 2 nM of E-box 1 probe in EMSA buffer in presence of 5 μM dA:dT competitor DNA and 1 mM DTT in a total volume of 15 μL for 20 min at 21 °C . The EMSA was carried out as described in Subsection 5.4.3.

MYC/MAX competitor assay. Binding of 500 nM MAX_{FL} to 2 nM VEGFA split 5 probe (Table 4.8) was observed in presence of varying concentrations of unlabeled competitor VEGFA DNA in EMSA buffer in presence of 5 μM dA:dT competitor DNA and 1 mM DTT in a total volume of 15 μL. The mixture was incubated for 20 min at 21 °C and the EMSA was carried out as described in Subsection 5.4.3.

hmC glucosylation EMSA. Oligonucleotides containing polyA and polyT with central CGs (Table 4.8) were hybridized by mixing 1.8 μM of the polyA strand with 2.2 μM of the polyT strand in IDT buffer, heating at 95 °C for 5 min and cooling down to RT overnight in a Dewar vessel containing boiling water. A final volume of 227 nM of probe was subjected to glucosylation following the protocol provided in Subsection 5.2.1. Probes were purified using the Oligo Clean & Concentrator kit. Probe concentration was determined by NanoDrop^{TM} spectrophotometer and validated on EMSA gels to ensure uniform probe concentration. Subsequently, 2 nM of glucosylated or non glucosylated probe was mixed with 9 μL double distilled water, 1.5 μL EMSA buffer, 1.5 μL DTT and 1 μL dA:dT competitor and 2 μL of either PBS or of 2 μM MECP2 HM were added. The reaction proceeded at 21 °C for 20 min, then 3 μL of 6x EMSA loading buffer were added and samples were applied to a pre-run 15 % EMSA gel and run at 240 V for 1 h. Imaging was carried out as described in Subsection 5.4.3.

5.4.4 Mass spectrometry

On-bead digestion

Pull-down samples stored in 20 μL PBS were denatured and reduced on the beads by adding 50 μL Denaturation/Reducing buffer and incubating for 30 min at 20 °C and 350 rpm shaking before 5.55 μL alkylation solution was added for 30 min at 20 °C and 350 rpm shaking. Samples were on-bead digested by first adding 1 μg Lys-C dissolved in nuclease free water for 1 h at 37 °C and 350 rpm shaking. The supernatant was transferred into a new tube and 1 μg Trypsin dissolved in 165 μL 50 mM Tris-HCl was

added to the beads and incubated for 1 h at 37 °C and 350 rpm shaking. The supernatants from both digestions were combined and 2 μg Trypsin was added. Incubation was carried out overnight at 37 °C and 350 rpm shaking. The reaction was stopped by adding 20 μL 10 % TFA.

Peptide desalting

The digested peptides were desalted using Pierce^{TM} C18 Tips following the manufacturer's instructions. In brief, the tips were activated by aspirating and dispensing 100 μL 50 % ACN in water twice and then equilibrated with 100 μL 0.1 % TFA. The sample containing the digested peptides was aspirated and dispensed 15 times with a volume of 100 μL. The tip was rinsed by aspirating 0.1 % TFA/5 % ACN twice. The samples were eluted by aspirating 20 μL of 0.1 % formic acid twice and dispensed into a new protein low-bind tube. The eluted peptides were dried in a vacuum concentrator for 1.5 h at 30 °C . Until further processing the dried samples were stored at -20 °C .

MS analysis

The dried peptides were dissolved in 10-20 μL of 0.1 % TFA and analyzed by nanoHPLC-MS/MS using an Ultimate 3000 RSLC nano-HPLC system coupled to a Hybrid-Orbitrap mass spectrometer Q Exactive Plus. 3-5 μL of the peptide solution were injected and enriched on a C18 PepMap 100 column (5 mm, 100 Å, 300 mm ID * 5 mm) using 0.1 % TFA, at a flow rate of 30 μL/min for 5 min. Peptides were then separated on a C18 PepMap 100 column (3 mm, 100 Å, 75 mm ID * 50 cm) with a linear gradient of 5-30 % ACN in H_2O containing 0.1 % formic acid over 90 min, at a flow rate of 300 nL/min. The nano-HPLC system was coupled online using a nano emitter with 10 μm tip diameter. Mass spectra were acquired in the range of m/z 300-1650 at a resolution of 70,000 for full scan, followed by up to ten high-energy collision-dissociation (HCD) MS/MS scans of the most intense, doubly or higher charged ions at a resolution of 17,500. Protein quantification was performed with MaxQuant (v.2.2.0.0) (Cox and Mann, 2008) using the Andromeda search algorithm and searching in parallel the Homo sapiens reference proteome of the UniProt database and a contaminants database implemented in MaxQuant. Searches were carried out for tryptic peptides allowing up to two missed cleavages. Carbamidomethylation was set as a fixed protein modification, and oxidation of methionine and acetylation of the N-terminus were set as variable modifications. The mass accuracy thresholds were 20 ppm for the first search and to 4.5 ppm for the second search. The false discovery rates for peptide and protein identification were set to 0.01. Only proteins for which at least two peptides were quantified were chosen for further validation. Relative quantification of proteins was performed by using the label-free quantification algorithm implemented in MaxQuant. Statistical data analysis of pull-down samples was performed using Perseus (v.2.0.7.0 and 2.0.11.0) (Tyanova and Cox, 2018). Proteins were included if identified in at least four out of five technical replicates in at least one of the two compared conditions. Label-free quantification (LFQ) intensities were log2-transformed, replicates were grouped, missing values were imputed using small normally distributed values, and a two-sided t-test was performed. Volcano plots were generated using the VolcaNoseR web application (Goedhart and Luijsterburg, 2020). Proteins with log2 fold changes below 1.5 and -log p-values below 0.025 were considered significant enriched. Venn diagrams prepared for data visualization were generated using the R script provided in appendix A.3.

5.4.5 Targeted gDNA enrichment.

For enrichment, spike-in probes were premixed equimolarly at 6.04 fmol each and added to 250 ng of sheared gDNA in Buffer B (from MethylCap Kit) in a total volume of 35.45 μL. GST-tagged MBP-MECP2 WT and MBP-MECP2 HM were TEV-digested for 30 min at RT to remove MBP. TEV-digested proteins were then added to the DNA mixture at a final concentration of 1.7 μM. Enrichment was performed

according to the manufacturer's instructions (MethylCap Kit version 6) following the high-salt elution protocol. MECP2 WT and MECP2 HM enrichments were carried out separately, each performed in three technical replicates per biological experiment. Eluted DNA was purified using the Macherey Nagel PCR purification kit. Spike-in probe recovery was determined by qPCR against a standard dilution series of spike-in probes with known concentrations. To distinguish individual spike-ins, three different primer pairs were used (C/C: o4374 and o4368, mC/mC: o4372 and o4368, hmC/mC: o4393 and o4368). qPCR was performed on all fractions (flow-through, wash 1, wash 2 and elution). Reactions were carried out by initial denaturation at 95 °C for 2 min, followed by 50 cycles of 95 °C for 10 s and 60 °C for 30 s. The remaining elution samples were further processed for NGS analysis.

6 Conclusion

This thesis addresses a previously underexplored area in epigenetic regulation by systematically investigating protein interactions with asymmetrically modified probes, with a particular focus on the biologically abundant hmC/mC modification. While previous proteomics studies had focused only on symmetrically modified CpG dyads, this work extends the existing knowledge by directly comparing these with the prevalent asymmetric hmC-containing dyads hmC/C and hmC/mC. For this purpose, I developed a protocol for generating asymmetrically modified DNA probes, identified candidate reader proteins by pull-down MS studies and validated selected candidates through in vitro DNA-binding assays. In a complementary approach I employed an evolved MECP2 mutant to map the prevalent hmC/mC modification in mESCs. These contributions offer a rich basis for understanding how DNA modification asymmetry influences protein binding and, ultimately, gene regulation.

6.1 Asymmetric probe generation allows for identifying asymmetric DNA probe readers

I established a robust and reproducible protocol for generating DNA probes bearing distinct cytosine modifications on each strand. The method is based on PCR generated symmetric DNA probes, followed by strand separation, primer annealing and Klenow-mediated primer extension, incorporating the cytosine modification of choice. This approach allowed for the generation of asymmetrically modified DNA probes containing dyads such as hmC/mC, which had not been previously explored in proteome profiling studies. Earlier proteome profiling studies by L. Bai et al., 2021; Iurlaro et al., 2013; Spruijt et al., 2013; J. Xiong et al., 2016 were focused on fully symmetric and hemi-modified CpG dyads, as well as TF-centric studies by G. Song et al., 2021 that already investigated hemi-modified hmC/C CpGs. However, the most prevalent hmC-containing dyad, hmC/mC, had not been studied yet. The reliability of the asymmetric probe generation protocol was confirmed by extensive probe characterization studies including fluorescent labeling of DNA strands that allowed to monitor each step of the protocol and Sanger sequencing to confirm probe integrity. Benchmarking was performed by pull-down assays using symmetric and asymmetric probes together with MBP-MBD2 with subsequent Western blotting. The results aligned with previous EMSA data (Buchmuller et al., 2020), showing the expected MBD2 preferences for symmetric and asymmetric probes, which further validated the probe design. The choice for promoter-region based probes was consistent with previous studies (Iurlaro et al., 2013), providing a biologically relevant sequence context. This ensured that CpGs were embedded in native sequences allowing to study protein-DNA interactions in a more unbiased manner compared to artificially designed probes. However, promoter-based probes have the inherent drawback of sequence dependence. Moreover, since the here established protocol is PCR-based, all cytosines, in CpG as well as CpH contexts, are subject to modification, which may lead to off-target effects if certain proteins preferentially bind to CpH modified DNA. The asymmetric probe generation protocol also works for

other probe contexts as was demonstrated for the 8NX CG core probe. Here, the probe was designed with long primer-overlapping flanking sequences without guanines on the 5'-strand to prevent cytosine modification incorporation. However, in the random nucleotide region modified cytosines could be still integrated. Overall these probes offer more control over cytosine modification incorporation and eliminate sequence dependence. However, they turned out to be not sufficiently discriminative in later MS experiments, highlighting the need for a higher number of CpG sites. Overall, the here established protocol offers a scalable and cost-effective method, applicable to all PCR-generated probes. While some limitations persist, particularly concerning off-target modifications, these can be addressed in downstream validations. Proteins that are identified to bind to these probes in initial screens need to be further characterized to ensure the binding is neither sequence-specific to the chosen promoter region nor CpH modification biased.

6.2 A pull-down MS approach revealed numerous putative hmC-containing CpG dyad binders

To systematically profile protein interactions with hmC-containing CpG dyads, I employed a pull-down approach coupled with quantitative mass spectrometry. This enabled the identification of proteins that exhibit differential binding to modified DNA, allowing for comparative assessment of binding preferences of different modified probes. Three hmC-containing dyads were investigated in the scope of this theses, namely hmC/C, hmC/mC, and hmC/hmC. In parallel, C/C and mC/mC probes were included as controls to directly compare how proteome profiles change with changes in C modifications. Proteins were isolated from HEK293T cells as a standard and well-characterized model. While hmC is not enriched in kidney, it but present at detectable levels. To complement this, proteins were also extracted from mouse brain tissue, offering a highly relevant biological context, as hmC is enriched in the brain. Each probe modification was tested in five technical replicates per biological replicate. For HEK293T-derived proteins, three independent biological replicates were performed and mouse brain pull-down assays were performed in a single biological replicate. This ensured a strong statistical basis enabling the identification of significantly enriched proteins. As expected based on previous studies (L. Bai et al., 2021; Iurlaro et al., 2013; G. Song et al., 2021; Spruijt et al., 2013), hmC-containing DNA probes showed fewer protein interactions overall than C/C and mC/mC modified probes. However, distinct profiles were observed for each CpG dyad modification. hmC/C probes shared many readers with C/C probes but uniquely enriched RBM45, CDCA7L and Hnrnpk (capital letter names indicate HEK293T-identified proteins, lowercase names indicate mouse brain-identified proteins), while FOXC1 and DACH1 turned out to be antireaders. hmC/hmC readers included HNRNPLL and Uhrf2 and Sfpq, with antireaders comprising RFX proteins and PRC1.6 complex proteins, as well as Jund and Creb1. For the hmC/mC probe no universal readers were found, however, some proteins were enriched over multiple other modified probes such as NOP16, ZC3HAV1, BAG2, RFX and Satb2. Antireaders included Creb1, Jund, and Trps1. Notably, many readers for hmC/mC overlapped with mC/mC and hmC/hmC, suggesting subtle shifts of reader profiles dependent on the probe modification. These probably reflect biological distinctions in how epigenetic information is interpreted. Generally, proteome profiling using MS provides a powerful tool to identify potential DNA-interacting proteins, however, it also has limitations. As the experimental pull-down conditions are relatively mild, not only direct DNA-binding proteins are enriched but also proteins associated with larger complexes or indirect interactors. Moreover, PTMs may significantly influence a proteins binding behavior. Consequently, proteins identified in these studies must be regarded as putative readers. Follow-up validation is necessary to determine if these proteins are direct, modification-specific binders. The comprehensive lists of enriched proteins

for each modified probe are a valuable resource, however, proteins of interest need to be further characterized before these candidates can be classified as *bona fide* readers of hmC/C, hmC/mC or hmC/hmC dyads.

6.3 Validation of candidate readers

To validate the candidate proteins identified by MS, a selection of these proteins was cloned and recombinantly expressed to be used in in vitro DNA-binding assays, with the goal of characterizing their ability to discriminate between different CpG dyad modifications. In initial Dot blot screening assays, proteins were tested for their binding affinity to the five different dyads C/C, mC/mC, hmC/C, hmC/mC, and hmC/hmC, using the same DNA probe as in the MS experiments. While some proteins exhibited modification preferences, others bound with the same preference to all probes or did not bind at all. This suggested that some proteins may have been pulled-down in complex with other proteins or that critical PTMs were missing that are necessary for binding. For example, SUB1 did not replicate its mC/mC and hmC/mC preference from the MS data in the in vitro pull-down assay. However, it was already shown in other studies to have a mC/mC preference (L. Bai et al., 2021) suggesting that its activity may be dependent on PTMs. SUB1 is known to be heavily regulated by phosphorylation (Mustafi et al., 2022), which could explain the absence of modification specific binding with SUB1 expressed in *E. coli*. Similarly, RFXANK did not reproduce its MS binding preferences when recombinantly expressed. However, RFX5, which together with RFXANK and RFXAP constitutes the RFX complex, recapitulated its MS observed mC/mC and hmC/mC preference in vitro. As RFX5 is responsible for DNA-binding, its validation explains why the full RFX complex was enriched in the MS data, even though only RFX5 binds to DNA directly. As RFX5's modification preference was confirmed in the Dot blot assays, its binding was further studied in EMSA assays, using canonical X-box probes, its known DNA-binding motif. RFX5 exhibited mC/mC preference and additionally bound significantly stronger to hmC/mC than to other dyad modifications. Quantitative analysis of binding affinities resulted in K_D values of 207 ±16.4 nM for mC/mC and 572.5 ±115.3 nM for hmC/mC for one X-box probe, and 48.25 ±6.18 nM for mC/mC and 106.49 nM ±15.88 nM for hmC/mC for another X-box probe, indicating a strong influence of sequence context on affinity. This study is the first to demonstrate that RFX5 can bind both mC/mC and hmC/mC dyads in its native context, although affinity decreases with hmC presence in one strand. Given that hmC/mC is much more prevalent in vivo than hmC/hmC, and RFX5 binding is abolished with hmC/hmC, this finding suggests a biologically relevant mechanism for retaining transcriptional instructions even when mC is oxidized to hmC in one strand. Further intriguing results were observed for MYC and MAX, two bHLH transcription factors known to prefer unmodified CpGs. Additionally to their C/C preference, they exhibited increased binding towards hmC/C and hmC/hmC probes in Dot blot assays. Their hmC-containing CpG dyad binding in MS was present but less pronounced. Therefore, their binding behavior was further studied in EMSA assays. Using the VEGFA promoter probe, the binding behavior observed in the Dot blots could be robustly reproduced. Quantitative binding assays showed that the MAX homodimer bound with highest affinity to C/C (K_D = 4.11 ±0.55 nM), followed by hmC/hmC (K_D = 14.5 ±2.42 nM) and weakest to mC/mC (K_D = 38.5 ±9.3 nM). A putative MYC^{bHLHZ}/MAX^{FL} heterodimer showed even stronger binding, with K_D values of 1.704 ±0.254 nM for C/C, 2.868 ±0.550 nM for hmC/hmC and 4.866 nM ±0.864 nM for mC/mC. The VEGFA promoter probe does not contain a canonical E-box, the known DNA binding motif of MYC and MAX, but it has E-box like sequences. To assess the sequence dependence, DNA binding studies of MYC and MAX with the E-box motif were performed, which yielded high affinities towards C/C and less affinity towards hmC/C, but hmC/hmC affinity was not observed, contrasting their VEGFA probe binding behavior.

This suggested that binding affinity is not only determined by cytosine modification state but also dependent on the sequence context. As MYC is a key regulator of VEGFA expression (Y. Shi et al., 2014), this result may indicate a context-specific role for hmC in influencing transcription factor-DNA interactions at target promoters. Together, the validation studies indicated that asymmetric CpG modifications can alter protein-DNA interactions. This study revealed novel reader candidates and highlighted the importance of sequence context and PTMs in epigenetic regulation.

6.4 Selective enrichment to map genomic locations of asymmetric hmC/mC dyads

Building on the pull-down MS identification of putative hmC-containing CpG dyad readers and the validation of selected candidates, I next sought to determine where the asymmetric hmC/mC dyads occur in the genome. To this end, I used a selective enrichment strategy employing an evolved MECP2 variant (MECP2 HM). MECP2s enhanced affinity for hmC/mC was validated by spike-in qPCR assays confirming its preference of hmC/mC > mC/mC > C/C. As a control genomic DNA was glucosylated to block MECP2 HM binding to hmC-containing dyads, which enabled the unambiguous identification of hmC/mC sites in subsequent analysis. Fragmented genomic DNA either untreated or glucosylated was subjected to enrichment by MECP2 HM or wildtype MECP2 (MECP2 WT) and captured DNA fragments were subsequently sequenced by Illumina sequencing. NGS analysis revealed high reproducibility and specificity of the different enrichment conditions, with clear separation between MECP2 WT, MECP2 HM and MECP2 HM with glucosylated DNA. This confirmed the robustness of the enrichment strategy and binding specificity as well as blocking by glucosylation of the proteins. Peak calling and comparative analyses showed further that MECP2 WT and MECP2 HM occupy partially overlapping but distinct genomic regions. Notably, MECP2 HM showed preferential enrichment at regulatory features such as CpG islands, 3'UTRs and TTS. In contrast, glucosylation of hmC substantially reduced the number of peaks as expected due to the blocking of MECP2 HM binding. Comparison of MECP2 mediated enrichment with MeDIP and hMeDIP sequencing data underlined the specificity of MECP2 HM for hmC/mC dyads and the added resolution of this approach relative to the more biased antibody-based methods. Together, this provided a genome-wide map of asymmetric hmC/mC dyads, establishing a foundation for exploring their potential functional roles in gene regulation.

7 Future developments

The present work provides a comprehensive list of putative hmC-containing CpG dyad readers together with initial validation of selected candidates and hmC/mC mapping studies using a selective MeCP2 mutant. While this already establishes a strong foundation for studying the role of epigenetic cytosine modifications, future efforts should extend to the subsequent oxidation products fC and caC, which remain underexplored, especially in asymmetric CpG dyad contexts. Although fC and caC are less abundant in somatic tissues, their enhanced levels in ESCs suggest important roles during embryogenesis and development. Identification and characterization of readers that recognize these modifications will greatly enhance our understanding of their regulatory impact on gene expression. Notably, previous profiling studies (L. Bai et al., 2021; G. Song et al., 2021; Spruijt et al., 2013) have identified a higher number of fC-binding proteins compared to hmC, indicating a distinct and possibly broader reader landscape for fC. Additionally, the chemical reactivity of the formyl group of fC allows for reversible covalent formation of Schiff base linkages with lysine residues in proteins as has been demonstrated for histones (Ji et al., 2017; Raiber et al., 2018). This unique feature can be used to covalently trap and identify fC readers in pull-down experiments and may help to reveal in vivo protein interactions mediated by fC. Extending the proteome profiling analysis to all possible 15 modified CpG dyads will yield a comprehensive list of interacting proteins building the foundation for connecting CpG modifications states with biological outcomes. Additionally, proteome profiles at specific time points or stages during development and cell differentiation will advance the current reader profiling studies to uncover temporal and spatial dynamics of CpG modification interactions.

A major future task will be the functional characterization of candidate readers identified by proteome profiling methods. It will be essential to determine whether these proteins are direct DNA binders, or if they are recruited via protein-protein interactions or as part of larger protein complexes. Moreover, it will be necessary to understand how these proteins are regulated, e.g. by post-translational modifications, and which DNA sequence or chromatin contexts they prefer. Final integration of these efforts into existing databases of known transcription factor binding motifs, and expanding them by including motifs bearing modified cytosines, will be an important contribution to bridge the fields of TFs and epigenetic regulation. Determining their affinity to different epigenetic DNA modifications will be a central question that can be either addressed by EMSAs, as was shown in this study, or extended by higher-throughput methods like Fluorescence anisotropy, SwitchSense (Knezevic et al., 2012) or SPRi. Emerging techniques such as chromatin fractionation and enrichment followed by proteomics (van Mierlo and Vermeulen, 2021) will be essential to study protein complexes that assemble on modified CpGs and can help to elucidate cooperative binding or indirect recruiting mechanisms. One such approach is ChIP–selective isolation of chromatin-associated proteins (SICAP) that is able to identify proteins that are physically associated with chromatin-bound transcription factors (Rafiee et al., 2016).

Another important area comprises the mapping of CpG dyad modification combinations at high resolution and in a locus-specific manner. Technologies such as ChIP-seq will be crucial to determine the genomic binding site of candidate readers. Mapping efforts can be even more enhanced by using

engineered selective reader proteins, through directed evolution (as was done for the hmC/mC selective MECP2 mutant TAYN) or mutagenesis of newly discovered readers like RFX5. Ultimately, this will yield insights into the distribution and function of modified CpG dyads in chromatin organization and gene regulation. More detailed examinations of selected genomic loci can be achieved by targeted approaches such as CRISPR-dCas9-based enrichment (X. Liu et al., 2017; Schmidtmann et al., 2016). By recruiting affinity tags to the loci of interest, such as disease-associated promoters and developmentally regulated genes, associated proteins and complexes can be affinity-enriched and then identified by mass spectrometry. This can further be combined with dyad specific sequencing methods, that localize the CpG dyad modification of choice. Subsequently, targeted identification of proteins at these modified loci can be carried out. Such approaches will shine light on the in vivo interplay between modified CpG dyads and their reader proteins. Looking ahead, single cell epigenomic technologies such as scChIP-seq (Grosselin et al., 2019) or single-cell CUT&Tag (Janssens et al., 2024) will provide insights into cell-to-cell variability in modification patterns and reader binding further advancing our understanding in context dependent regulations. Finally, given the role of aberrant DNA methylation in diseases such as cancer and neurodevelopmental disorders, the systematic identification and functional study of CpG dyad readers may reveal diagnostic markers or therapeutic targets.

Integration of CpG modification data with transcription factor binding maps and histone modifications will be essential to unravel the combinatorial effect of the epigenetic regulatory landscape and its implications for development and disease.

A Supplementary Information

A.1 Supplementary Figures

A.1.1 Supporting figures for Section 3.1

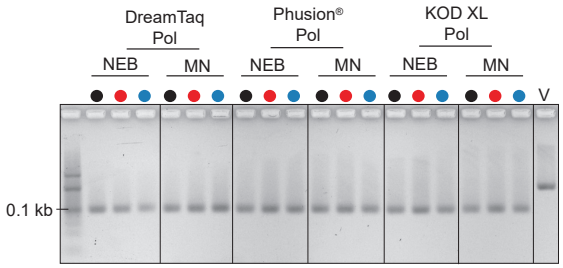


Figure A.1: Incorporation of modified dCTPs into the 8NX probe. Incorporation of C (black circle), mC (red circle) or hmC (blue circle) containing dNTP mixes into the 8NX probe using different polymerases.

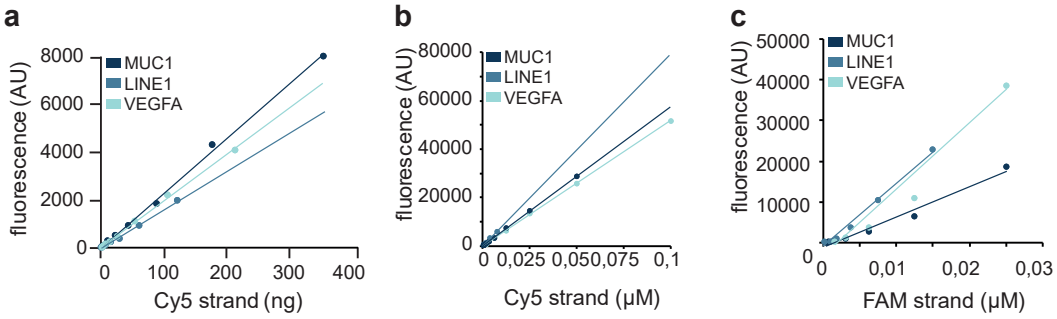


Figure A.2: Standard curves of fluorescently labeled DNA probes. Standard curves using a known concentration range of labeled DNA to determine DNA yield in the asymmetric probe generation process.**a** Titration of Cy5 labeled DNA mass, **b** Concentration series of Cy5 labeled DNA, **c** Concentration series of FAM labeled DNA.

Figure A.3: Sequence verification of final probes. Both the forward and reverse strand of a regular PCR-generated VEGFA probe without the use of modified dCTPs (PCR C) was sequenced as reference. In addition, all five asymmetric VEGFA probes generated by the primer-extension-based probe generation protocol were directly subjected to the same forward and reverse Sanger sequencing as the PCR reference (C/C: CC, hmC/C: HC, hmC/hmC: HH, hmC/mC: HM; mC/mC: MM).

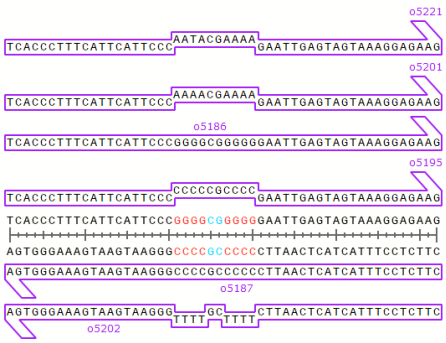


Figure A.4: Oligonucleotides used for 8NX mismatch assay. Oilgonucleotides were hybridized in a way that the 8NX probe contained 0 CC mismatches (o5186 + o5187), 0 TT mismatches (o5201 + o5202), 1 TT mismatch (o5221 + o5202) or 8 CC mismatches (o5201 + o5187).

A.1.2 Supporting figures for Section 3.2

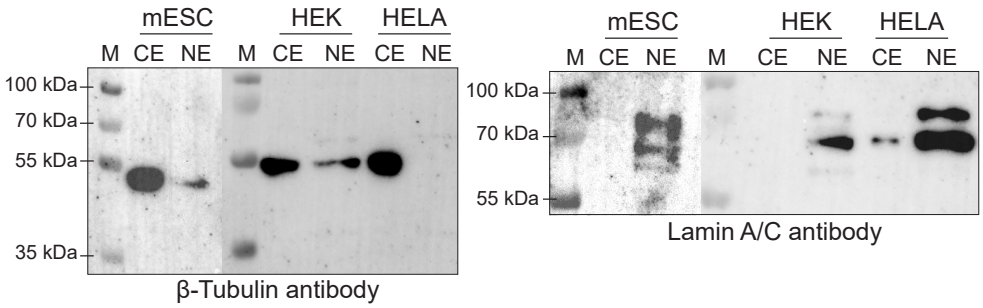


Figure A.5: Nuclear protein extraction from different cell types. The nuclear extraction protocol was applied to mESC, HEK293T (HEK) and HELA cells. Western blots showing the presence of β-Tubulin in cytoplasmic extracts (CE) and nuclear extracts (NE) and Lamins A and C in cytoplasmic extracts (CE) and nuclear extracts (NE). Samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M)), expected molecular weights: β-Tubulin 55 kDa, Lamins A and C: 63 and 70 kDa.

A.1.3 Supporting figures for Section 3.3

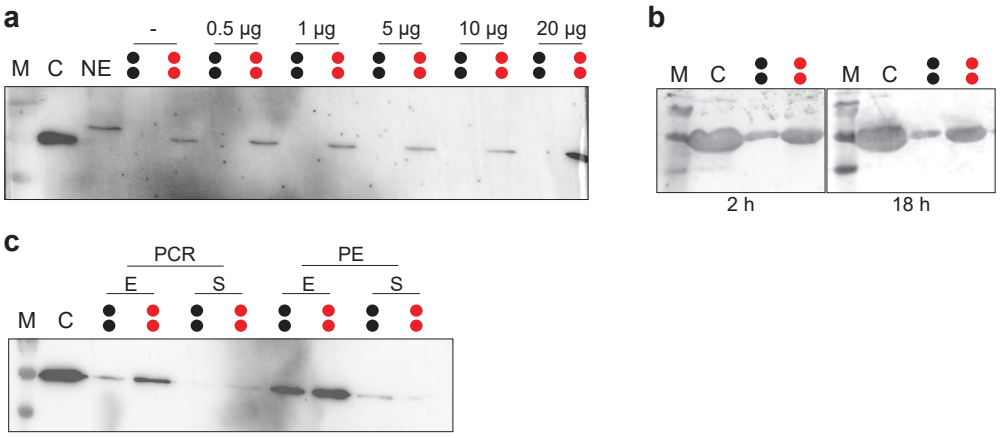


Figure A.6: Optimization of pull-down conditions.(a) Western blot showing binding of 1 nM MBP-MBD2 in the presence of 3.6 μm nuclear extracts to unmodified (black circles) or methylated (red circles) VEGFA probes with varying concentrations of dI:dC competitor DNA. Pure MBP-MBD2 was run in lane C as control. **(b)** Western blot of pull-down samples with protein incubation time during pull-down for 2 h or 18 h. **(c)** Performance of PCR vs PE probes, elution (E) and unbound proteins in supernatant (S) fractions were applied to a Western blot.

A.1.4 Supporting figures for Section 3.4

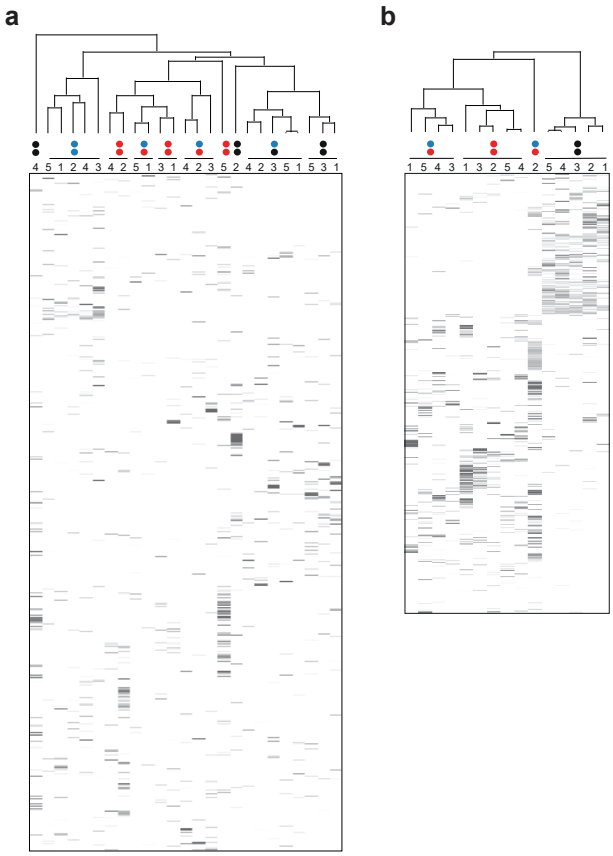


Figure A.7: Heatmap analyses for correlation-based clustering of experiments 2 and 3. The five technical replicates are shown as numbers and the dyad modifications as colored circles: C/C black circles, mC/mC red circles, hmC/hmC blue circles, hmC/C blue and black circles, hmC/mC blue and red circles.

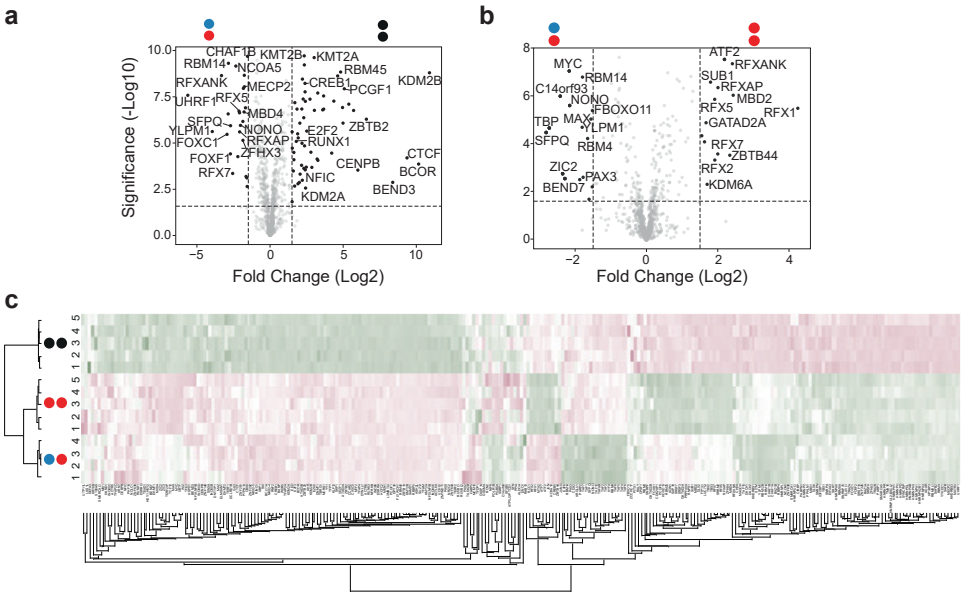


Figure A.8: HEK293T reader proteins from experiment 3. (a) Volcano plot comparing hmC/mC (blue and red circles) with C/C (black circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5). **(b)** Volcano plot comparing hmC/mC (blue and red circles) with mC/mC (red circles) enriched proteins (p-value < 0.025 and log2 fold change >1.5). Corresponding protein lists are in Table A.46, Table A.47. **(c)** Heatmap of correlation-based clustering of the LFQ-intensities after log2 transformation and normalization by row mean subtraction. Proteins included in the clustering significantly bind to at least one of the probes as determined by an ANOVA test (p-value < 0.025 and s0=1).

A.1.5 Supporting figures for Section 3.5

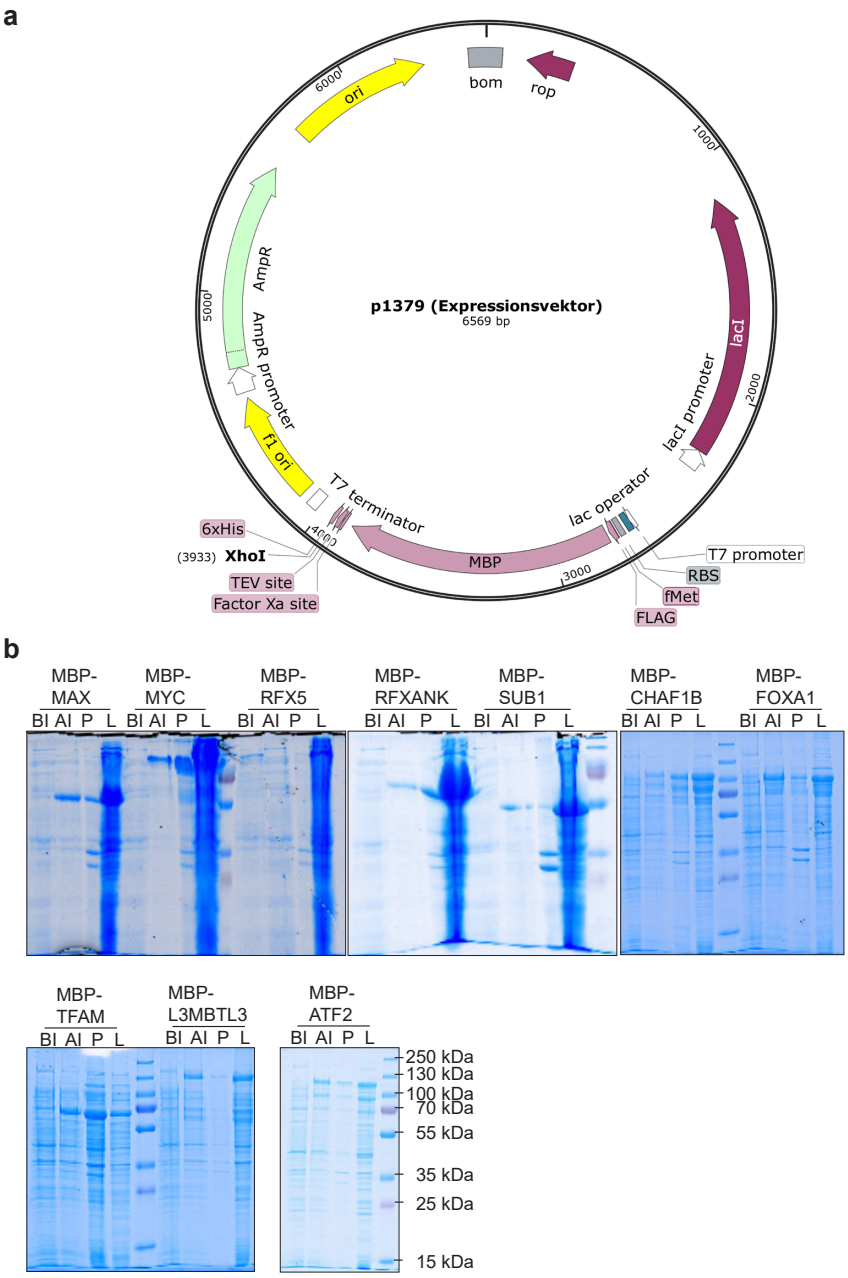


Figure A.9: Expression of candidate proteins in *E. coli* lysates. (a) Plasmid map of the expression vector. The *XhoI* restriction site was used to insert the full-length coding sequences of candidate proteins. **(b)** Coomassie stained 12 % SDS PAGES of expression of candidate proteins in *E. coli* BL21 DE(3) Gold showing the following fractions: before induction (BI), after induction with 1 mM IPTG (AI), pellet fraction (P) after sonication and soluble fraction (L) after sonication referred to as lysate.

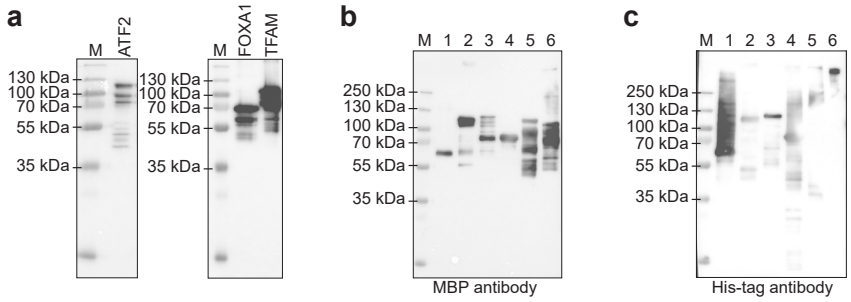


Figure A.10: Western blot analysis of *E. coli* lysates containing MBP-tagged candidate proteins. (a) Western Blot using a MBP antibody showing the presence of MBP-tagged candidate proteins. Expected MWs are 99.8 kDa for MBP-ATF2, 94.4 kDa for MBP-FOXA1, 74.3 kDa for MBP-TFAM. Samples were run on a 12 % SDS-PAGE with a Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder (lane M). **(b, c)** Western Blot using a MBP antibody or a His antibody showing the presence of the following proteins: MBP-MAX (1, 63.5 kDa), MBP-MYC (2, 94.0 kDa), MBP-RFX5 (3, 110.5 kDa), MBP-RFXANK (4, 73.3 kDa), MBP-CHAF1B (5, 106.7 kDa), MBP-FOXA1 (6, 94.4 kDa).

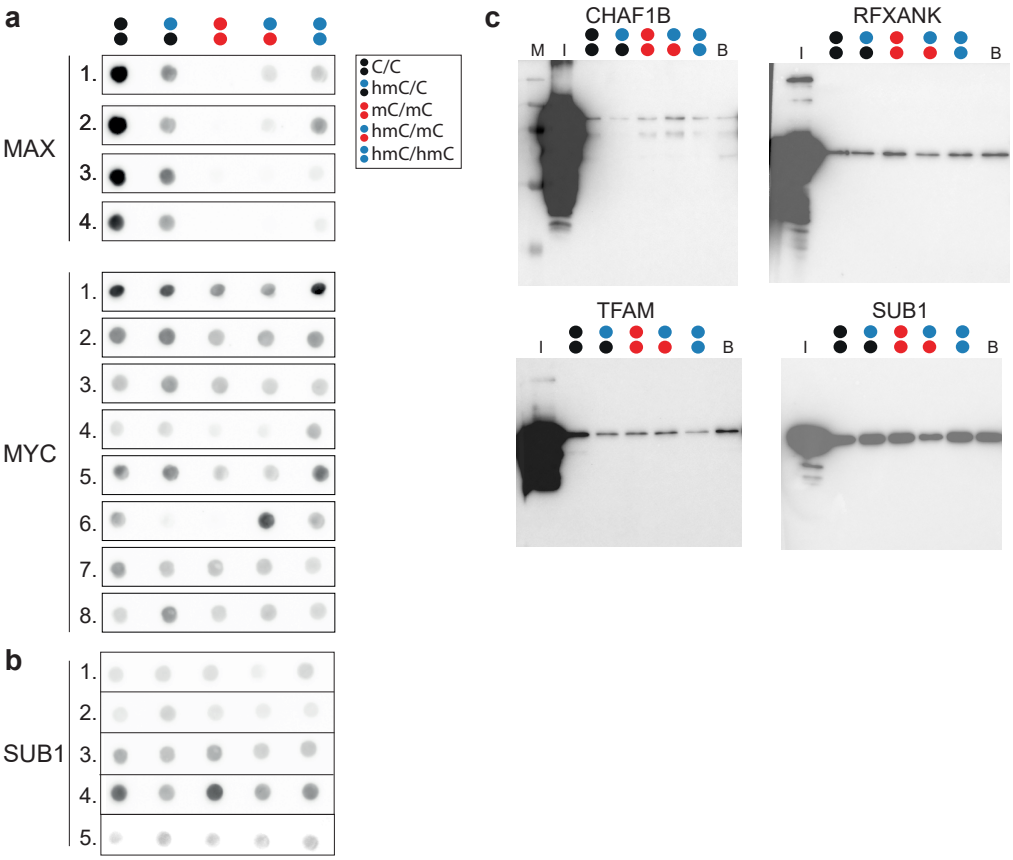


Figure A.11: Dot blot and Western blot analyses of *E. coli* lysate pull-down experiments. (a) Anti-MBP dot blot analyses of eluates from enrichments employing VEGFA probes and *E. coli* expression lysate of MAX and MYC fused to an N-terminal MBP tag. Multiple replicates are shown. **(b)** Anti-MBP dot blot analyses of eluates from enrichments employing VEGFA probes and *E. coli* expression lysate of SUB1 fused to an N-terminal MBP tag. Multiple replicates are shown. **(c)** Anti-MBP Western blot analyses of eluates from enrichments employing VEGFA probes and *E. coli* expression lysate of CHAF1B, RFXANK, TFAM, SUB1 fused to an N-terminal MBP tag.

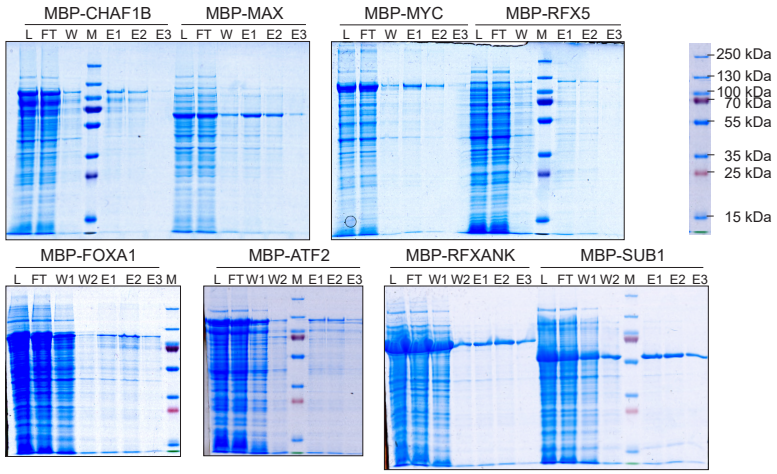


Figure A.12: Expression and purification of candidate proteins. Coomassie stained 12 % SDS PAGES of purification of candidate proteins in *E. coli* BL21 DE(3) Gold showing the following fractions: soluble fraction (L) after sonication referred to as lysate, flow-through (FT) collected after binding to the Ni-NTA resin, wash (W) and elution (E) fractions. Labeled marker (Thermo Scientific^{TM} PageRuler^{TM} Plus Prestained Protein Ladder) is shown for reference. Expected MWs are 106.7 kDa for MBP-CHAF1B, 63.5 kDa for MBP-MAX, 94.0 kDa for MBP-MYC, 110.5 kDa for MBP-RFX5, 94.4 kDa for MBP-FOXA1, 99.8 kDa for MBP-ATF2, 73.3 kDa for MBP-RFXANK lysate, 59.6 kDa for MBP-SUB1.

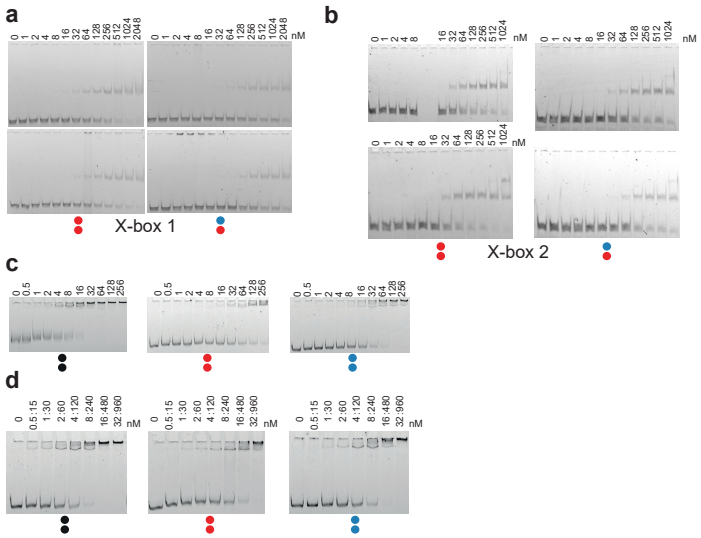


Figure A.13: EMSA gels used for K_D determination. **(a)** EMSA titration of $RFX5_{DBD(TEV)}$ and mC/mC and hmC/mC modified X-Box 1 probes. **(b)** EMSA titration of $RFX5_{DBD(TEV)}$ and mC/mC and hmC/mC modified X-Box 2 probes. **(c)** EMSA titration of MAX_{fl} and C/C, mC/mC and hmC/hmC modified VEGFA probes. **(d)** EMSA titration of MAX_{fl} and MYC_{bHLHZ} and C/C, mC/mC and hmC/hmC modified VEGFA probes.

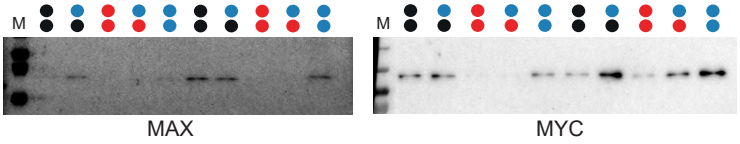


Figure A.14: Western blot analyses of MAX and MAY pull-down experiments with NE. Anti-MBP Western blot analyses of eluates from enrichments employing VEGFA probes and *E. coli* expression lysates of MAX and MYC fused to an N-terminal MBP tag supplemented with NE.

A.1.6 Supporting figures for Section 3.6

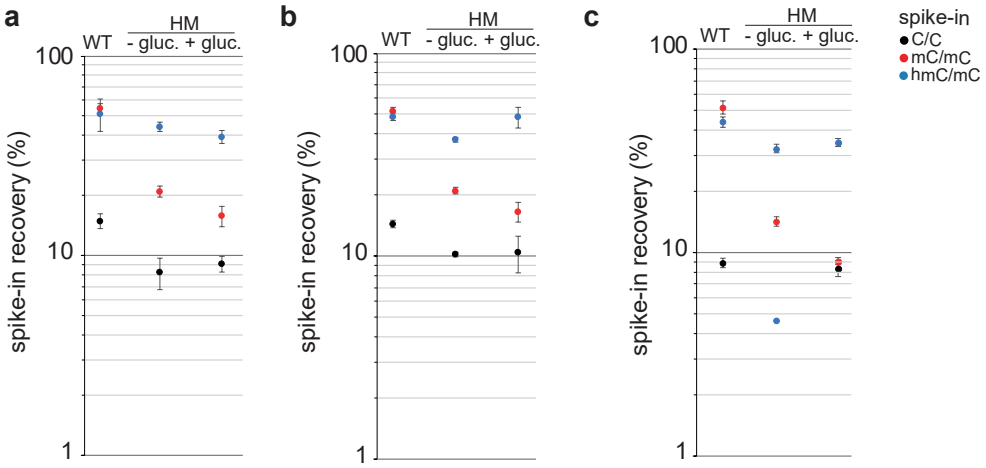


Figure A.15: Spike-in recoveries with MECP2 WT and MECP2 HM. Recovery of C/C, mC/mC, and hmC/mC spike-ins with MECP2 WT and MECP2 HM during the enrichent assays. Three biological replicates are shown corresponding to **(a)** replicate 2, **(b)** replicate 3, **(c)** replicate 4. Replicate 1 is shown in **Figure 3.39**.

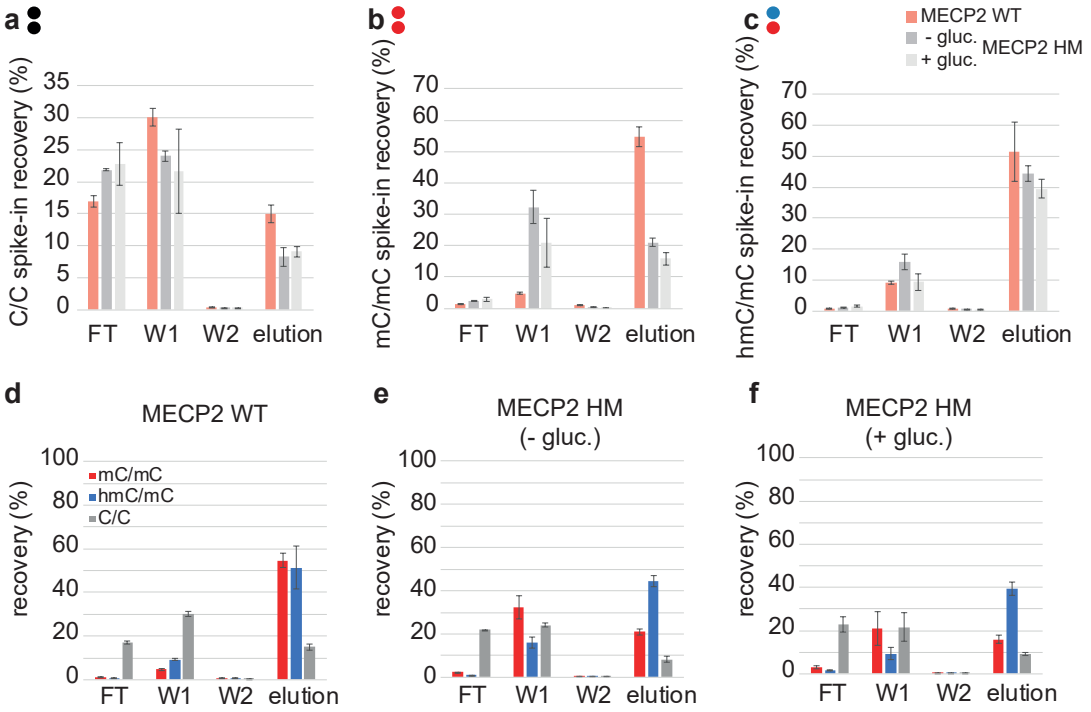


Figure A.16: Spike-in and protein-specific recoveries in replicate 2. Bar diagrams showing the recoveries of **(a)** C/C spike-in, **(b)** mC/mC spike-in, and **(c)** hmC/mC spike-in with MECP2 WT and MECP2 HM. Comparative display of spike-in recoveries is povided in the bar diagrams showing **(d)** MECP2 WT enrichment, **(e)** MECP2 HM enrichment, and **(f)** MECP2 HM enrichment in presence of glucosylated gDNA.

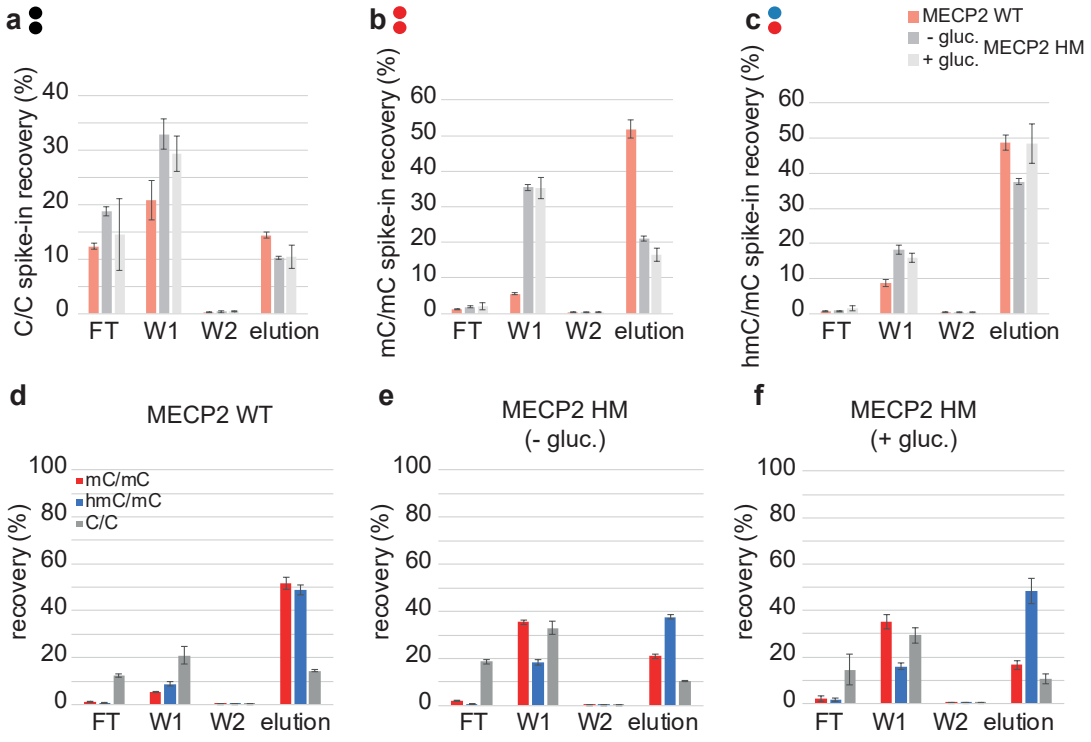


Figure A.17: Spike-in and protein-specific recoveries in replicate 3. Bar diagrams showing the recoveries of **(a)** C/C spike-in, **(b)** mC/mC spike-in, and **(c)** hmC/mC spike-in with MECP2 WT and MECP2 HM. Comparative display of spike-in recoveries is povided in the bar diagrams showing **(d)** MECP2 WT enrichment, **(e)** MECP2 HM enrichment, and **(f)** MECP2 HM enrichment in presence of glucosylated gDNA.

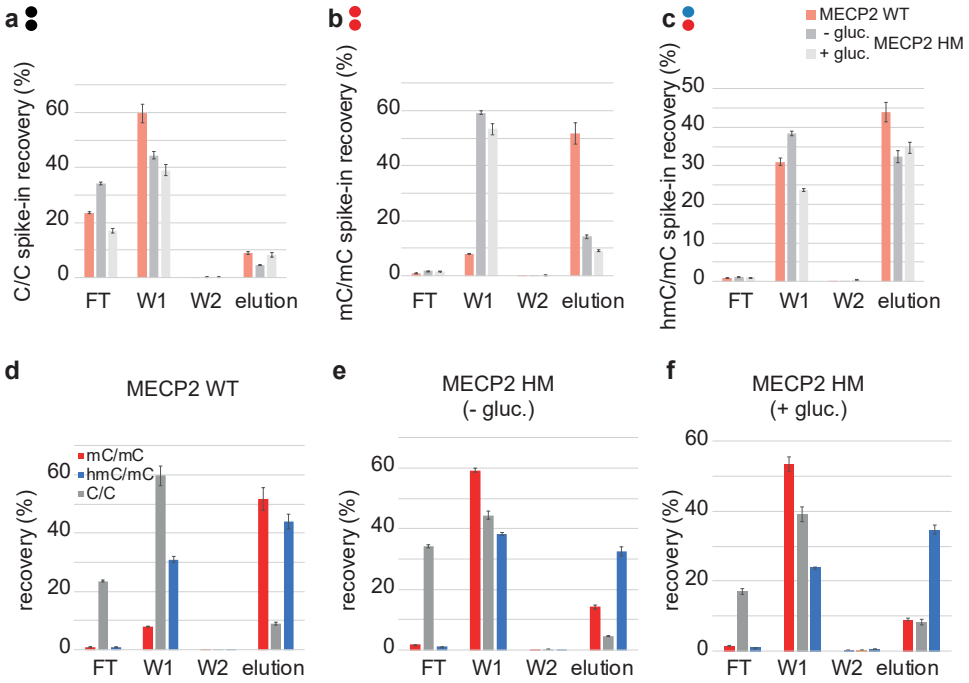


Figure A.18: Spike-in and protein-specific recoveries in replicate 4. Bar diagrams showing the recoveries of **(a)** C/C spike-in, **(b)** mC/mC spike-in, and **(c)** hmC/mC spike-in with MECP2 WT and MECP2 HM. Comparative display of spike-in recoveries is povided in the bar diagrams showing **(d)** MECP2 WT enrichment, **(e)** MECP2 HM enrichment, and **(f)** MECP2 HM enrichment in presence of glucosylated gDNA.

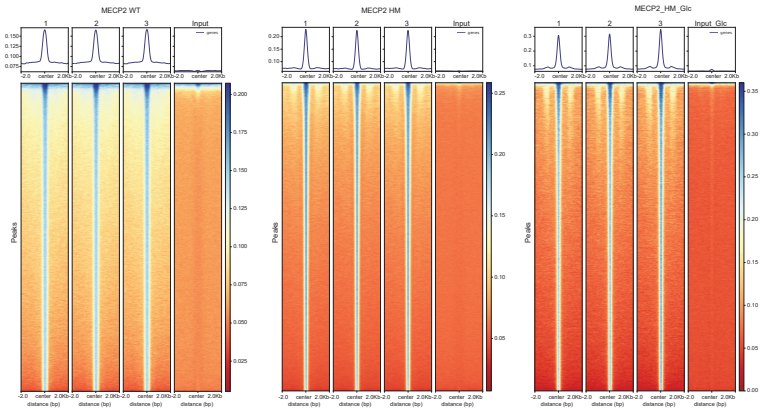


Figure A.19: Heatmaps and average enrichment profiles for MECP2 binding. Heatmaps and average enrichment profiles centered on peak regions (±2 kb) show strong and reproducible enrichment for MECP2 WT, MECP2 HM and MECP2 HM with glucosylated DNA for all replicates. All profiles display sharp signal peaks at the center of the identified binding sites relative to the input.

A.2 Supplementary Tables

Table A.1: VEGFA C/C vs. mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RBM25	Increased	3.346.565.628	1.008.938.487	13.435.950.498
MBD4	Increased	3.127.036.285	9.633.107.855	1.276.014.414
ZNF638	Increased	1.713.297.272	1.032.308.019	12.036.377.462
ZFHX4	Increased	162.724.762	9.870.434.433	11.497.682.053
SAFB	Increased	2.320.854.187	9.029.795.755	11.350.649.942
ACIN1	Increased	2.667.882.538	7.549.617.049	10.217.499.587
CHD3	Increased	1.251.714.706	8.956.700.031	10.208.414.737
SRSF10	Increased	1.344.765.091	876.427.908	10.109.044.171
PRPF40A	Increased	3.714.262.772	6.369.098.797	10.083.361.569
SRSF3	Increased	2.141.437.149	7.852.513.331	9.993.950.480
SRSF7	Increased	1.411.441.422	8.106.643.435	9.518.084.857
ELAVL1	Increased	1.873.759.079	7.643.732.767	9.517.491.846
RBM14	Increased	1.332.397.079	8.038.135.318	9.370.532.397
RBMX	Increased	1.467.813.492	7.775.391.113	9.243.204.605
HNRNPDL	Increased	1.114.562.607	7.947.826.685	9.062.389.292
RFX5	Increased	2.694.632.339	6.290.750.712	8.985.383.051
PNN	Increased	1.781.362.534	7.149.145.947	8.930.508.481
SRSF9	Increased	1.388.673.019	7.402.456.744	8.791.129.763
TBX2	Increased	2.233.210.373	6.412.731.378	8.645.941.751
HNRNPC	Increased	1.139.613.724	7.442.319.481	8.581.933.205
TBX3	Increased	1.238.788.986	7.310.710.827	8.549.499.813
TRA2A	Increased	1.588.219.452	6.896.424.431	8.484.643.883
TRA2B	Increased	1.367.462.921	710.042.518	8.467.888.101
HNRNPA2B1	Increased	1.220.360.184	6.607.687.809	7.828.047.993
HNRNPUL1	Increased	1.407.041.168	6.374.414.626	7.781.455.794
SCAF11	Increased	1.682.430.649	5.971.466.704	7.653.897.353
ATF2	Increased	1.444.956.207	5.966.109.329	7.411.065.536
SAFB2	Increased	222.342.186	5.096.352.048	7.319.773.908
HNRNPA1	Increased	1.007.351.303	6.232.161.397	7.239.512.700
ZC3H18	Increased	1.103.581.619	6.056.003.572	7.159.585.191
SRRM2	Increased	1.433.824.921	5.698.252.808	7.132.077.729
LUC7L3	Increased	1.330.046.844	5.627.630.944	6.957.677.788
SAP18	Increased	1.946.664.047	4.924.150.988	6.870.815.035
UHRF1	Increased	1.280.550.003	5.526.891.026	6.807.441.029
HNRNPA3	Increased	114.202.919	552.821.448	6.670.243.670
BCL2L2	Increased	1.214.004.135	5.454.036.119	6.668.040.254
RNPS1	Increased	1.618.496.323	4.903.448.367	6.521.944.690
FOXB1	Increased	1.383.224.487	5.004.149.857	6.387.374.343
EIF4A3	Increased	1.950.476.074	4.305.814.669	6.256.290.743
FOXC1	Increased	2.480.623.245	3.741.466.071	6.222.089.316
YLPM1	Increased	119.995.842	4.478.415.994	5.678.374.413
MBD1	Increased	1.160.630.417	4.121.599.506	5.282.229.923
MAGOH	Increased	1.754.863.358	3.127.951.585	4.882.814.943
ADAR	Increased	1.021.443.176	3.800.835.831	4.822.279.007
NCOA5	Increased	2.052.378.464	2.208.838.998	4.261.217.461
RBM8A	Increased	108.211.174	2.992.982.251	4.075.093.991
SUB1	Increased	1.262.494.659	2.704.550.582	3.967.045.241
CPSF6	Increased	1.189.268.875	2.184.083.721	3.373.352.596
HNRNPH3	Increased	1.288.208.771	1.942.959.075	3.231.167.846
NUDT21	Increased	1.325.187.302	1.537.084.654	2.862.271.956
ZC3H13	Increased	1.261.281.204	1.580.037.116	2.841.318.320
YBX1	Increased	1.103.059.387	1.343.454.023	2.446.513.410
SAMD1	Decreased	-9.936.119.843	1.076.722.354	20.703.343.383
BEND3	Decreased	-5.028.996.277	1.033.319.229	15.362.188.567
PARD3	Decreased	-4.037.539.291	1.026.337.816	14.300.917.451
ZIC2	Decreased	-3.410.260.773	1.084.605.005	14.256.310.823
KMT2A	Decreased	-3.202.728.271	1.059.006.099	13.792.789.261
DPY30	Decreased	-2.985.953.522	1.071.170.042	13.697.653.942

Name	Change	Fold change (log2)	Significance	Manhattan distance
CTCF	Decreased	-7.314.043.808	6.150.378.829	13.464.422.637
KMT2B	Decreased	-3.580.495.834	9.415.539.899	12.996.035.733
KDM2A	Decreased	-375.408.783	9.101.941.238	12.856.029.068
NFRKB	Decreased	-2.166.162.872	1.023.805.431	12.404.217.182
ZMYND11	Decreased	-3.910.907.364	8.263.580.698	12.174.488.062
WDR5	Decreased	-1.960.358.429	9.876.318.938	11.836.677.367
ASH2L	Decreased	-3.632.388.687	8.064.981.884	11.697.370.571
ZBTB2	Decreased	-4.249.185.562	713.711.481	11.386.300.372
BCOR	Decreased	-3.556.750.107	7.739.506.829	11.296.256.936
INO80	Decreased	-2.092.806.625	9.134.740.887	11.227.547.512
RBBP5	Decreased	-3.422.402.573	7.697.030.938	11.119.433.511
TFAP2A	Decreased	-1.584.738.541	9.117.184.085	10.701.922.626
NKX2-5	Decreased	-1.501.551.819	9.130.386.556	10.631.938.375
RUVBL1	Decreased	-1.672.311.401	850.362.104	10.175.932.441
RUVBL2	Decreased	-1.693.357.849	803.326.585	9.726.623.699
L3MBTL3	Decreased	-4.635.368.347	5.075.496.138	9.710.864.485
BANP	Decreased	-4.137.408.829	5.541.197.893	9.678.606.722
RBM45	Decreased	-3.571.028.519	5.723.915.161	9.294.943.680
PCGF1	Decreased	-3.551.899.338	530.228.728	8.854.186.618
MTF2	Decreased	-2.792.033.005	6.044.193.105	8.836.226.110
UBTF	Decreased	-1.630.604.553	7.138.993.143	8.769.597.696
MAZ	Decreased	-1.702.025.223	693.730.139	8.639.326.613
AHR	Decreased	-2.354.508.591	5.794.670.255	8.149.178.846
SKI	Decreased	-1.272.432.327	6.830.921.233	8.103.353.560
MYC	Decreased	-2.430.158.615	5.618.103.643	8.048.262.258
CENPB	Decreased	-3.018.408.585	5.008.556.598	8.026.965.183
HCFC2	Decreased	-2.669.990.921	5.334.150.332	8.004.141.253
MEN1	Decreased	-2.974.424.744	4.950.698.958	7.925.123.702
NR2C2	Decreased	-1.356.018.829	6.543.946.857	7.899.965.686
CCAR2	Decreased	-246.999.855	5.394.877.736	7.864.876.285
DYNLL1	Decreased	-2.443.249.893	5.222.863.524	7.666.113.417
ZNF644	Decreased	-1.273.524.475	624.391.849	7.517.442.965
NSD1	Decreased	-1.229.821.396	6.240.210.686	7.470.032.082
HNRNPK	Decreased	-1.341.162.491	6.111.829.472	7.452.991.963
SUZ12	Decreased	-241.485.939	4.981.078.945	7.395.938.335
UCHL5	Decreased	-1.791.971.207	5.580.787.608	7.372.758.814
MAX	Decreased	-3.202.392.578	3.915.505.838	7.117.898.416
ACTL6A	Decreased	-1.190.671.539	5.884.066.205	7.074.737.744
MCRS1	Decreased	-1.609.765.625	5.260.821.275	6.870.586.899
BCL11A	Decreased	-1.973.674.774	4.672.449.706	6.646.124.480
INO80D	Decreased	-1.937.439.728	457.778.855	6.515.228.278
RNF2	Decreased	-1.166.133.499	5.194.313.946	6.360.447.445
PCBP2	Decreased	-1.020.946.503	5.309.080.023	6.330.026.526
TFAM	Decreased	-1.498.121.262	4.761.484.487	6.259.605.748
ZIC1	Decreased	-2.956.703.949	3.240.124.461	6.196.828.410
BEND7	Decreased	-1.629.469.681	426.879.202	5.898.261.701
ZBTB10	Decreased	-172.624.588	4.022.775.578	5.749.021.458
INO80C	Decreased	-1.116.659.546	4.580.195.889	5.696.855.435
FLYWCH1	Decreased	-209.370.842	3.494.777.871	5.588.486.291
ARNT	Decreased	-1.953.351.593	3.245.090.083	5.198.441.676
INO80E	Decreased	-1.497.279.739	3.340.999.769	4.838.279.508
WIZ	Decreased	-1.985.710.526	27.857.463	4.771.456.826
NKX2-1	Decreased	-1.159.508.514	3.485.561.857	4.645.070.371
RUNX2	Decreased	-2.035.131.454	233.028.515	4.365.416.604
ACTR5	Decreased	-1.484.466.171	2.854.110.695	4.338.576.866
EZH2	Decreased	-2.185.593.033	1.816.922.364	4.002.515.397
ATF1	Decreased	-2.101.613.998	1.853.755.196	3.955.369.194
ESRRA	Decreased	-1.895.675.278	1.908.342.822	3.804.018.100
NFIA	Decreased	-1.329.872.131	2.434.268.214	3.764.140.344
TAF6L	Decreased	-1.050.883.102	2.457.533.977	3.508.417.079
TADA3	Decreased	-1.023.237.228	2.173.555.074	3.196.792.302

Table A.2: 8NX C/C vs. mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
UHRF1	Increased	4.403.795.242	6.058.117.485	10.461.912.727
BCOR	Decreased	-2.774.842.262	9.210.776.415	11.985.618.677
PCGF1	Decreased	-3.030.790.329	6.620.024.559	9.650.814.888
SAMD1	Decreased	-1.904.212.952	7.691.628.623	9.595.841.575
KDM2B	Decreased	-3.826.803.684	4.883.324.972	8.710.128.656
L3MBTL3	Decreased	-1.763.720.036	6.051.417.383	7.815.137.419
E2F6	Decreased	-1.386.516.094	5.483.516.129	6.870.032.223
CGGBP1	Decreased	-1.175.010.681	5.191.048.532	6.366.059.213
TFAP2A	Decreased	-1.338.393.688	42.785.584	5.616.952.088

Table A.3: 16NX 4 pmol C/C vs. mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
MECP2	Increased	3.160.332.203	6.579.320.372	9.739.652.575
L3MBTL3	Decreased	-4.445.808.887	7.731.141.514	12.176.950.401
SAMD1	Decreased	-4.405.491.352	5.241.440.007	9.646.931.359
ZBTB2	Decreased	-365.839.529	5.622.497.438	9.280.892.728
KMT2A	Decreased	-2.409.463.882	6.663.057.211	9.072.521.092
KDM2B	Decreased	-2.896.011.829	3.502.148.504	6.398.160.333
FLYWCH1	Decreased	-2.073.293.209	4.233.078.205	6.306.371.414
MGA	Decreased	-1.677.447.796	4.209.778.287	5.887.226.083
BCOR	Decreased	-2.167.695.045	3.643.063.533	5.810.758.578
TFAP2A	Decreased	-2.220.872.402	2.924.450.201	5.145.322.603
RBBP5	Decreased	-1.373.711.586	3.317.546.749	4.691.258.335
ASH2L	Decreased	-1.623.683.929	3.021.888.208	4.645.572.137
ZMYND11	Decreased	-1.257.677.555	3.306.039.667	4.563.717.221
RING1	Decreased	-1.212.579.727	3.246.211.204	4.458.790.931
DPY30	Decreased	-1.778.539.658	2.506.096.662	4.284.636.320
CUX1	Decreased	-1.121.804.237	3.093.401.164	4.215.205.401
TAF1	Decreased	-1.310.683.727	2.622.346.351	3.933.030.078
BEND3	Decreased	-1.793.169.498	2.135.813.575	3.928.983.072
CBX3	Decreased	-1.538.285.255	2.142.648.291	3.680.933.546
SKP1	Decreased	-120.152.998	1.681.204.765	2.882.734.745

Table A.4: 16NX 16 pmol C/C vs. mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
MBD4	Increased	116.508.913	3.682.374.886	4.847.464.016
FYTTD1	Increased	2.300.356.865	1.963.879.193	4.264.236.058
L3MBTL3	Decreased	-285.989.809	7.967.669.339	10.827.567.429
KDM2B	Decreased	-3.646.504.879	6.187.488.027	9.833.992.905
ZBTB2	Decreased	-2.802.522.659	5.955.816.747	8.758.339.406
BEND3	Decreased	-3.080.259.323	4.303.947.294	7.384.206.617
SAMD1	Decreased	-3.539.941.311	3.310.649.424	6.850.590.735
MTF2	Decreased	-1.532.690.048	4.860.293.884	6.392.983.932
TFAP2A	Decreased	-2.375.592.232	3.702.627.442	6.078.219.674
BCOR	Decreased	-2.457.061.291	3.388.629.511	5.845.690.802
SKP1	Decreased	-1.498.000.145	3.530.311.114	5.028.311.259
DPY30	Decreased	-1.173.081.875	3.424.260.931	4.597.342.806
PCGF1	Decreased	-2.651.895.046	166.624.267	4.318.137.716
JRKL	Decreased	-1.576.762.676	1.762.186.905	3.338.949.581
MPG	Decreased	-1.352.737.427	1.923.914.656	3.276.652.083
EXOSC4	Decreased	-1.304.419.518	1.390.440.102	2.694.859.620

Table A.5: Venn results for C/C associated with the Venn diagram in Figure 3.14.

Probe	Number of proteins	Protein names
16NC 16 pmol, 16NC 4 pmol, 8NC, VEGFA C	4	SAMD1, L3MBTL3, TFAP2A, BCOR
16NC 16 pmol, 8NC, VEGFA C	1	PCGF1
16NC 16 pmol, 16NC 4 pmol, 8NC	1	KDM2B
16NC 16 pmol, 16NC 4 pmol, VEGFA C	3	DPY30, BEND3, ZBTB2
16NC 4 pmol, VEGFA C	5	ASH2L, ZMYND11, KMT2A, RBBP5, FLYWCH1
16NC 16 pmol , VEGFA C	1	MTF2
16NC 16 pmol, 16NC 4 pmol	1	SKP1
8NC	2	E2F6, CGGBP1
VEGFA C	53	ATF1, ARNT, CENPB, NKX2-1, CCAR2, WIZ, ACTL6A, INO80E, CTCF, SUZ12, WDR5, UBTF, INO80C, ZIC2, KMT2B, RUVBL1, BEND7, BANP, ZBTB10, MAZ, TAF6L, NKX2-5, SKI, RBM45, DYNLL1, NR2C2, NFRKB, MEN1, ZIC1, UCHL5, NSD1, HNRNPK, INO80, PCBP2, ACTR5, MYC, ESRRA, AHR, RUNX2, PARD3, MAX, ZNF644, TADA3, HCFC2, TFAM, EZH2, NFIA, RNF2, MCRS1, RUVBL2, INO80D, BCL11A, KDM2A
16NC 4 pmol	5	CUX1, CBX3, MGA, RING1, TAF1
16NC 16 pmol	3	MPG, JRKL, EXOSC4

Table A.6: Venn results for mC/mC associated with the Venn diagram in Figure 3.14.

Probe	Number of proteins	Protein names
8NmC, VEGFA mC	1	UHRF1
16NmC 16 pmol, VEGFA mC	1	MBD4
VEGFA mC	50	HNRNPUL1, PRPF40A, HNRNPDL, MAGOH, HNRNPA3, RBM14, YBX1, ELAVL1, SUB1, RFX5, ACIN1, ADAR, FOXC1, BCL2L2, RBM25, YLPM1, SAP18, SRSF3, SRRM2, TRA2B, TBX2, HNRNPH3, ZC3H13, NCOA5, RNPS1, HNRNPA1, CHD3, ZC3H18, HNRNPC, CPSF6, TRA2A, SAFB2, EIF4A3, FOXB1, SAFB, HNRNPA2B1, SRSF10, NUDT21, ZFHX4, PNN, LUC7L3, ATF2, SRSF9, TBX3, SRSF7, MBD1, RBM8A, SCAF11, RBMX, ZNF638
16NmC 4 pmol	1	MECP2
16NmC 16 pmol	1	FYTTD1

A.2.1 Supporting tables for Section 3.4

Table A.7: Experiment 1 C/C vs mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
BEND3	Increased	8,246297836	7,19783107	15,44412891
SAMD1	Increased	7,238728714	6,698595346	13,93732406
ZIC2	Increased	5,841230392	7,27181665	13,11304704
BCOR	Increased	6,627060699	5,795210166	12,42227087
ZBTB2	Increased	4,863626862	7,186742466	12,05036933
L3MBTL3	Increased	7,367689323	4,409551134	11,77724046
MAX	Increased	2,675393295	8,688979768	11,36437306
KMT2A	Increased	3,875550461	7,423525858	11,29907632
CTCF	Increased	4,512261581	6,186453942	10,69871552
MEN1	Increased	2,29356575	8,283117955	10,57668371
ASH2L	Increased	2,824770737	7,005151174	9,829921911

Name	Change	Fold change (log2)	Significance	Manhattan distance
KMT2B	Increased	4,862147141	4,951859864	9,814007005
RBBP5	Increased	3,116770935	6,665911522	9,782682457
ESRRA	Increased	1,659718704	7,964251445	9,623970149
NKX2-5	Increased	2,136923218	7,248844543	9,385767761
FLYWCH1	Increased	3,484248734	5,813613017	9,297861751
SKP1	Increased	2,384778595	6,554364228	8,939142823
AHR	Increased	2,949562454	5,830600087	8,780162541
KDM2B	Increased	3,731732941	4,96793726	8,699670201
BANP	Increased	3,632776642	5,022661531	8,655438173
ATF1	Increased	2,109357071	6,388956968	8,498314039
ZMYND11	Increased	2,708562088	5,344565168	8,053127256
MAZ	Increased	1,932572556	6,112053543	8,044626099
MYC	Increased	2,913643265	5,127950471	8,041593736
DPY30	Increased	2,37853241	4,919957059	7,298489469
CENPB	Increased	2,266657639	4,765291187	7,031948826
MTF2	Increased	2,058824921	4,892284519	6,95110944
PCGF1	Increased	2,706035614	4,199351718	6,905387332
TFAP2A	Increased	3,089399338	3,35288847	6,442287808
RBM45	Increased	2,260877609	3,989689597	6,250567206
NFIX	Increased	1,644786835	4,542412282	6,187199117
HCFC2	Increased	2,020028305	3,962107663	5,982135968
ZIC1	Increased	2,427778244	2,264326158	4,692104402
RFX5	Decreased	-7,326299286	10,78077039	18,10706968
RFXANK	Decreased	-5,445960617	8,055566882	13,5015275
RFXAP	Decreased	-4,98131218	7,845123484	12,82643566
FOXC1	Decreased	-5,873378372	6,073922743	11,94730112
UHRF1	Decreased	-4,81105423	5,218519895	10,02957413
SUB1	Decreased	-2,273377991	7,654422238	9,927800229
ZFHX2	Decreased	-1,62878952	8,127642021	9,756431541
MBD4	Decreased	-3,583946991	6,013750546	9,597697537
DEK	Decreased	-2,037175751	4,270951715	6,308127466
FOXA1	Decreased	-1,838240814	3,395559526	5,23380034
RFX7	Decreased	-2,565734863	1,951614178	4,517349041

Table A.8: Experiment 2 C/C vs mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
KDM2B	Increased	7,722176361	12,56360343	20,28577979
ZBTB2	Increased	5,831394577	10,05073696	15,88213154
BCOR	Increased	6,582937622	7,460459719	14,04339734
CENPB	Increased	4,671602631	9,294364673	13,9659673
BEND3	Increased	7,255882263	6,613841247	13,86972351
KMT2A	Increased	1,616791534	11,14853152	12,76532305
MYC	Increased	3,006248093	9,279241789	12,28548988
L3MBTL3	Increased	2,359803009	9,855079964	12,21488297
SAMD1	Increased	2,616728592	9,425908226	12,04263682
ZIC1	Increased	2,202627182	8,9327623	11,13538948
FLYWCH1	Increased	1,887147141	8,273822278	10,16096942
ASH2L	Increased	1,5221138	8,602406208	10,12452001
CREB1	Increased	1,803582764	8,317557675	10,12114044
ZIC2	Increased	2,398022461	7,708094837	10,1061173
MAX	Increased	2,406166458	7,605869771	10,01203623
PCGF1	Increased	3,848865128	5,537412391	9,386277519
ESRRA	Increased	1,652100372	7,646829937	9,298930309
BANP	Increased	2,871908951	5,721381903	8,593290854
TFAP4	Increased	2,667440414	5,700454101	8,367894515
NFIX	Increased	2,376829147	5,867143627	8,243972774
TFAM	Increased	1,743247604	6,487599623	8,230847227
KDM2A	Increased	2,154576492	5,751004352	7,905580844
HCFC2	Increased	1,922961426	5,832421652	7,755383078
AHR	Increased	2,628530121	4,360957034	6,989487155

Name	Change	Fold change (log2)	Significance	Manhattan distance
INO80C	Increased	1,528804016	5,405779776	6,934583792
ARNT	Increased	2,183887863	4,431257363	6,615145226
MEN1	Increased	1,568022537	4,765565686	6,333588223
TBRG1	Increased	2,485676575	3,386619112	5,872295687
NKX2-3	Increased	1,501268005	4,259127743	5,760395748
TFPT	Increased	1,826120377	3,59554894	5,421669317
CBFB	Increased	1,541508484	2,99607987	4,537588354
ZBTB10	Increased	1,595026016	2,322315565	3,917341581
HEY1	Increased	1,653650665	1,872900525	3,52655119
RFX5	Decreased	-6,414311218	10,9446043	17,35891552
RFXAP	Decreased	-5,093232727	9,167652284	14,26088501
FOXC1	Decreased	-2,883320999	9,586320613	12,46964161
MECP2	Decreased	-5,806713867	5,176689628	10,9834035
RFXANK	Decreased	-5,020617676	5,855086277	10,87570395
UHRF1	Decreased	-2,930190277	7,374561082	10,30475136
MBD2	Decreased	-3,050417328	5,636259285	8,686676613
RFX7	Decreased	-2,372740555	5,283477386	7,656217941
FOXA1	Decreased	-2,332907486	4,442750215	6,775657701
ZFHX3	Decreased	-2,781612778	3,793155346	6,574768124
RFX1	Decreased	-2,912943649	2,905448652	5,818392301
FOXF1	Decreased	-2,188570023	2,789619009	4,978189032
PTBP3	Decreased	-2,331261826	2,222870194	4,55413202
SUB1	Decreased	-2,637187576	1,619616981	4,256804557

Table A.9: Experiment 3 C/C vs mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
KMT2A	Increased	3,158500671	14,08243814	17,24093881
SAMD1	Increased	4,407381058	12,33705624	16,7444373
BCOR	Increased	10,2425642	6,224259599	16,4668238
CTCF	Increased	10,45025864	5,271881795	15,72214044
ZIC2	Increased	6,361213684	8,609268961	14,97048265
PCGF1	Increased	5,642613602	9,278207791	14,92082139
KMT2B	Increased	2,147656631	12,69647508	14,84413171
KDM2B	Increased	9,954428864	4,784497187	14,73892605
BEND3	Increased	10,68668327	3,714138821	14,40082209
ASH2L	Increased	2,353954697	11,95301781	14,30697251
ZIC1	Increased	4,178694153	9,835590063	14,01428422
SUZ12	Increased	3,708779526	10,15486017	13,8636397
NFIX	Increased	3,001924896	10,07866827	13,08059317
ZBTB2	Increased	6,752664566	6,223733057	12,97639762
RBBP5	Increased	2,246655655	10,61043532	12,85709098
MTF2	Increased	5,535847855	7,242745625	12,77859348
MYC	Increased	3,236789322	9,473574155	12,71036348
L3MBTL3	Increased	4,87722435	7,65747578	12,53470013
CENPB	Increased	7,30196991	4,989581361	12,29155127
FLYWCH1	Increased	3,338329697	8,710147381	12,04847708
RBM45	Increased	4,450345993	7,460877683	11,91122368
ESRRA	Increased	2,030160141	9,803704921	11,83386506
HCFC2	Increased	5,296779251	6,487873485	11,78465274
TFAM	Increased	1,859225845	9,732823566	11,59204941
BANP	Increased	2,598760223	8,540256939	11,13901716
RUNX2	Increased	2,456557083	8,279029618	10,7355867
KDM2A	Increased	2,666439056	7,980611683	10,64705074
ZMYND11	Increased	5,169730377	5,373140182	10,54287056
MAX	Increased	2,558623886	7,745458573	10,30408246
CCAR2	Increased	2,437768173	7,820874407	10,25864258
ZBTB10	Increased	1,902853775	8,264068846	10,16692262
DNMT1	Increased	4,639563751	5,485837592	10,12540134
EED	Increased	2,918299103	7,169042443	10,08734155
CREB1	Increased	1,650974655	7,943700828	9,594675483

Name	Change	Fold change (log2)	Significance	Manhattan distance
VEZF1	Increased	1,652908325	7,781228009	9,434136334
DPY30	Increased	2,018853378	7,365844886	9,384698264
BCL11A	Increased	1,904291916	7,283526808	9,187818724
DYNLL1	Increased	1,636780167	7,447541555	9,084321722
MXD4	Increased	2,959039688	5,860867615	8,819907303
SKP1	Increased	2,658983231	6,153634054	8,812617285
AHR	Increased	3,243846893	5,550161362	8,794008255
ARNT	Increased	3,250579834	5,499943854	8,750523688
MAZ	Increased	2,884310532	5,69603378	8,580344312
E2F2	Increased	2,839251709	5,617631352	8,456883061
CXXC5	Increased	3,446099472	4,88001465	8,326114122
C14orf93	Increased	2,181939697	5,958224559	8,140164256
TCF12	Increased	1,798645401	6,338240875	8,136886276
TFAP4	Increased	2,964052963	5,108543808	8,072596771
MXI1	Increased	3,364154434	4,496840575	7,860995009
BEND7	Increased	2,692538834	5,008058778	7,700597612
ZBTB25	Increased	2,359072495	4,949310885	7,30838338
MITF	Increased	1,740633392	5,417770879	7,158404271
CEBPB	Increased	1,692731857	5,459195201	7,151927058
NFIA	Increased	2,485860825	4,605467769	7,091328594
GLI1	Increased	2,644717026	4,380811843	7,025528869
ZNF639	Increased	2,041291809	4,841019385	6,882311194
NFIC	Increased	2,579135513	4,153530708	6,732666221
PATZ1	Increased	2,249742889	4,315259006	6,565001895
EZH2	Increased	2,416700745	3,795546784	6,212247529
HAND1	Increased	2,236857605	3,912446407	6,149304012
ESRRG	Increased	2,207956696	2,772010646	4,979967342
TERF1	Increased	1,518280029	3,234426074	4,752706103
RUNX1	Increased	1,563083649	3,051355816	4,614439465
INO80C	Increased	2,15105629	2,449111826	4,600168116
KLF12	Increased	1,853593445	2,580003738	4,433597183
S100A9	Increased	1,912420273	2,157738031	4,070158304
RFXANK	Decreased	-5,730038452	11,57039794	17,30043639
RFX5	Decreased	-4,034650421	10,9573386	14,99198902
RFX7	Decreased	-4,343798065	10,03043048	14,37422855
UHRF1	Decreased	-5,228425598	8,479390427	13,70781603
RFXAP	Decreased	-4,087901306	9,090452552	13,17835386
SAFB	Decreased	-1,92542305	10,31756412	12,24298717
ACIN1	Decreased	-1,581201553	10,59751528	12,17871683
MBD4	Decreased	-2,638188171	9,304235377	11,94242355
FOXC1	Decreased	-4,187403488	7,538685625	11,72608911
ZFHX2	Decreased	-1,622570801	9,509850992	11,13242179
ZNF326	Decreased	-1,582966232	9,39676116	10,97972739
NCOA5	Decreased	-1,776353836	9,161343245	10,93769708
MECP2	Decreased	-1,809353256	8,628804384	10,43815764
FOXF1	Decreased	-3,777761078	6,642624536	10,42038561
PRPF40A	Decreased	-2,053840256	8,138970719	10,19281098
RBM25	Decreased	-1,62498436	7,802127015	9,427111375
MBD2	Decreased	-3,280000305	6,069256076	9,349256381
GATAD2A	Decreased	-2,077544403	7,139285231	9,216829634
ATF2	Decreased	-1,722369766	7,452846067	9,175215833
SUB1	Decreased	-2,23182373	6,812643845	9,044467575
KHDRBS2	Decreased	-1,875942993	6,95629828	8,832241273
SAFB2	Decreased	-1,642021942	7,07635064	8,718372582
RFX1	Decreased	-2,908335876	5,719320979	8,627656855
ZFHX3	Decreased	-2,400116348	5,416297179	7,816413527
ZBTB44	Decreased	-2,382618713	5,328347849	7,710966562
YLPM1	Decreased	-2,148587799	4,792019033	6,940606832
FOXK1	Decreased	-1,527096176	5,284044232	6,811140408
RBM12B	Decreased	-1,664742661	5,061962465	6,726705126
MBD1	Decreased	-1,558323669	5,009938298	6,568261967
RFX2	Decreased	-1,772886276	4,391599939	6,164486215
RBM3	Decreased	-1,65476265	1,783340826	3,438103476

Table A.10: Experiment 1 C/C vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
SAMD1	Increased	7,753874588	9,220785321	16,97465991
BEND3	Increased	7,997872543	7,315769965	15,31364251
ATF1	Increased	4,132014465	9,571323715	13,70333818
ASH2L	Increased	3,91824379	9,481004637	13,39924843
BCOR	Increased	6,792121506	5,979249066	12,77137057
KMT2A	Increased	3,998106003	8,622500104	12,62060611
L3MBTL3	Increased	7,866799164	4,665600234	12,5323994
ZIC2	Increased	5,068614578	6,900892673	11,96950725
WDR5	Increased	1,527470398	10,32574142	11,85321182
RBBP5	Increased	4,412427521	7,360357283	11,7727848
MEN1	Increased	2,775874329	8,500670126	11,27654446
NKX2-5	Increased	2,507110214	8,697081203	11,20419142
MAX	Increased	2,473703003	8,006911759	10,48061476
CTCF	Increased	4,494226456	5,899834029	10,39406049
KDM2B	Increased	4,040622711	5,94450687	9,985129581
ZBTB2	Increased	4,413482285	5,115939838	9,529422123
ESRRA	Increased	1,781591797	7,462716275	9,244308072
CENPB	Increased	2,842333221	6,378337821	9,220671042
TFAM	Increased	2,243678665	6,232608863	8,476287528
BANP	Increased	3,177341461	5,106873813	8,284215274
MAZ	Increased	1,980831528	6,262749564	8,243581092
DPY30	Increased	2,866939163	5,126778547	7,99371771
FLYWCH1	Increased	3,340688324	4,46947457	7,810162894
KMT2B	Increased	4,179572678	3,408199229	7,587771907
AHR	Increased	2,859115982	4,301975533	7,161091515
SKP1	Increased	2,102856445	4,971784252	7,074640697
ZMYND11	Increased	2,876893616	4,171054972	7,047948588
RNF2	Increased	1,720940781	5,272299856	6,993240637
HCFC2	Increased	1,984590149	4,945183223	6,929773372
NFIX	Increased	2,16244545	4,705937783	6,868383233
MSANTD3	Increased	2,025314713	4,665021791	6,690336504
MTF2	Increased	2,525743484	4,058187345	6,583930829
MYC	Increased	1,533311844	4,705724855	6,239036699
E2F6	Increased	1,822513199	4,022572227	5,845085426
ARNT	Increased	1,679194641	4,001773487	5,680968128
PCGF6	Increased	1,505555725	4,081541762	5,587097487
RBM45	Increased	1,99732399	3,418675194	5,415999184
PCGF1	Increased	2,25060997	2,865715111	5,116325081
TFAP2A	Increased	2,481744003	2,403317058	4,885061061
ZIC1	Increased	2,516430664	1,880634649	4,397065313
SSBP2	Increased	1,840375519	2,139644184	3,980019703
LDB1	Increased	1,657877731	2,052402453	3,710280184
RUVBL1	Increased	1,958137512	1,670187726	3,628325238
UHRF1	Decreased	-5,41462059	5,613776404	11,02839699
RBM14	Decreased	-1,889851379	7,645636264	9,535487643
EXOSC5	Decreased	-2,730695343	6,424328743	9,155024086
FOXC1	Decreased	-3,306598282	4,192894259	7,499492541
BAZ1B	Decreased	-2,016291809	3,15214301	5,168434819
CELSR1	Decreased	-1,880156708	1,627985127	3,508141835
SF3A3	Decreased	-1,616300201	1,821409224	3,437709425

Table A.11: Experiment 2 C/C vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
KDM2B	Increased	7,775701141	11,3684083	19,14410944
BEND3	Increased	7,519405746	8,459675063	15,97908081
ZBTB2	Increased	6,098607635	9,729176264	15,8277839
BCOR	Increased	6,528377151	7,80616497	14,33454212
CREB1	Increased	3,732040787	9,765862451	13,49790324

Name	Change	Fold change (log2)	Significance	Manhattan distance
CENPB	Increased	4,192437744	8,430226208	12,62266395
SAMD1	Increased	2,791795731	9,238850953	12,03064668
BANP	Increased	5,409820557	5,95991488	11,36973544
L3MBTL3	Increased	2,392855072	8,921122955	11,31397803
TFAP4	Increased	3,338018799	7,081384036	10,41940284
MEN1	Increased	1,770697021	8,521430789	10,29212781
ZIC1	Increased	1,902907944	8,370386799	10,27329474
ZIC2	Increased	1,978143311	8,080090681	10,05823399
NKX2-1	Increased	2,008750534	7,801224095	9,809974629
PCGF1	Increased	4,063145065	5,611723552	9,674868617
KMT2A	Increased	1,644488907	7,675589947	9,320078854
ESRRA	Increased	1,639367294	7,588550711	9,227918005
TFAM	Increased	2,030957794	6,866707372	8,897665166
FLYWCH1	Increased	1,829023743	6,901893253	8,730916996
ACTR5	Increased	1,694320679	6,945207575	8,639528254
TFPT	Increased	3,169520187	5,40648474	8,576004927
NFIX	Increased	2,088851929	6,344618357	8,433470286
INO80C	Increased	2,168215179	6,255763741	8,42397892
ASH2L	Increased	1,706384277	6,506870394	8,213254671
HCFC2	Increased	2,077851486	6,135119275	8,212970761
MGA	Increased	1,515094376	6,669580585	8,184674961
RBBP5	Increased	1,664810181	6,377556422	8,042366603
NKX2-3	Increased	2,309610748	5,730954069	8,040564817
MYC	Increased	1,612118912	6,027146777	7,639265689
RFX5	Increased	2,810845184	4,781249078	7,592094262
AHR	Increased	2,208069992	3,951371833	6,159441825
VEZF1	Increased	1,681085205	4,405615242	6,086700447
DPY30	Increased	1,65514679	4,319695481	5,974842271
KDM2A	Increased	1,876754379	3,926023227	5,802777606
PBX1	Increased	1,72682457	4,049768118	5,776592688
ARNT	Increased	1,901484299	3,72013845	5,621622749
NFIC	Increased	2,559074783	2,712077329	5,271152112
TBRG1	Increased	2,388790894	2,878744891	5,267535785
RFX1	Increased	2,230177689	3,034911389	5,265089078
ZBTB10	Increased	1,756707764	3,225380191	4,982087955
SMAD1	Increased	1,79101944	1,939437224	3,730456664
RUNX2	Increased	1,818128204	1,608247092	3,426375296
UHRF1	Decreased	-3,302682495	7,940841721	11,24352422
MECP2	Decreased	-5,119957352	4,923687674	10,04364503
CHAF1B	Decreased	-1,995353699	6,503570333	8,498924032
CHAF1A	Decreased	-2,078175354	4,413490615	6,491665969
HNRNPLL	Decreased	-2,200650406	2,516934648	4,717585054
RPP25L	Decreased	-2,254079819	2,023104155	4,277183974
SYF2	Decreased	-1,620495224	1,622459576	3,2429548

Table A.12: Experiment 1 mC/mC vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFXANK	Increased	5,123757553	7,544813854	12,66857141
RFX5	Increased	7,010343742	4,250925279	11,26126902
RFXAP	Increased	4,665046692	5,532053028	10,19709972
ZFHX2	Increased	1,809608459	7,993035052	9,802643511
MBD4	Increased	2,184214401	7,17639857	9,360612971
FOXC1	Increased	2,56678009	6,759156986	9,325937076
SUB1	Increased	1,937084198	7,179530531	9,116614729
MGA	Increased	1,750964737	6,912305206	8,663269943
TBX21	Increased	1,849641037	6,536990216	8,386631253
L3MBTL2	Increased	2,138063812	6,067042248	8,20510606
ATF1	Increased	2,022657394	5,711212482	7,733869876
E2F6	Increased	2,395269012	5,319857432	7,715126444
TBX2	Increased	2,00021553	5,404094934	7,404310464

Name	Change	Fold change (log2)	Significance	Manhattan distance
DEK	Increased	2,265805817	4,957697424	7,223503241
CDX1	Increased	1,603125381	5,373123522	6,976248903
PCGF6	Increased	2,080963898	4,882567749	6,963531647
RFX7	Increased	2,937112045	2,073961725	5,01107377
SIX1	Increased	2,595845795	2,382879829	4,978725624
FOXA1	Increased	1,751510239	2,822981833	4,574492072
MTA2	Increased	2,098211288	2,298744281	4,396955569
LDB1	Increased	1,793291092	1,864034234	3,657325326
CBX3	Increased	1,711323547	1,766575457	3,477899004
EXOSC5	Decreased	-3,017801285	5,522291125	8,54009241
RBM14	Decreased	-1,672227478	6,624288621	8,296516099
RADX	Decreased	-1,850803375	4,08940568	5,940209055
KRR1	Decreased	-1,941794205	3,082620688	5,024414893
SRFBP1	Decreased	-1,905795288	2,216373644	4,122168932
KLHL7	Decreased	-1,540359879	2,121441804	3,661801683
SF3A3	Decreased	-1,701864243	1,942524935	3,644389178

Table A.13: Experiment 2 mC/mC vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFX5	Increased	9,225156403	8,911843681	18,13700008
ATF2	Increased	2,479412842	8,020372162	10,499785
RFX1	Increased	4,476176834	5,985873491	10,46205033
CREB1	Increased	1,928458023	8,247664402	10,17612243
ZFHX3	Increased	1,881183624	8,280146041	10,16132967
FOXC1	Increased	1,554481125	8,366176749	9,920657874
RFXAP	Increased	5,010009003	4,81526418	9,825273183
RFXANK	Increased	5,170384979	4,411510398	9,581895377
MBD2	Increased	3,355878448	5,71111967	9,066998118
RFX7	Increased	2,727877045	5,229205501	7,957082546
BANP	Increased	2,425976181	2,810102095	5,236078276
ZBTB44	Increased	1,945095062	2,909710135	4,854805197
CREB5	Increased	1,752019501	2,717740629	4,46976013
RBBP7	Increased	2,022769928	1,878515152	3,90128508
SUMO2	Increased	1,856627274	1,870853466	3,72748074
SSBP2	Increased	1,769953537	1,854127852	3,624081389
TOX	Decreased	-1,542904663	6,930098781	8,473003444
HNRNPLL	Decreased	-2,530077362	2,993935515	5,524012877

Table A.14: Venn diagram readers for C/C, mC/mC, and hmC/hmC from HEK293T. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, mC/mC, and hmC/hmC modifications in at least one out of two biological experiments (experiments 1 and 2) associated with the Venn diagram in Figure 3.18.

Enriched for	Number of proteins	Protein names
hmC/hmC>C/C, hmC/hmC>mC/mC	4	HNRNPLL, EXOSC5, RBM14, SF3A3
hmC/hmC>C/C, hmC/hmC<mC/mC	1	FOXC1
hmC/hmC<C/C, hmC/hmC<mC/mC	10	ATF1, E2F6, BANP, PCGF6, SSBP2, RFX5, LDB1, MGA, RFX1, CREB1
hmC/hmC>C/C	8	CHAF1B, MECP2, BAZ1B, CHAF1A, RPP25L, SYF2, UHRF1, CELSR1

Enriched for	Number of proteins	Protein names
hmC/hmC<C/C	51	SMAD1, ARNT, CENPB, NKX2-1, DPY30, SAMD1, TBRG1, MTF2, PCGF1, CTCF, WDR5, ASH2L, MSANTD3, BEND3, INO80C, KMT2B, ZIC2, VEZF1, RUVBL1, ZBTB10, NKX2-3, MAZ, TFPT, NFIX, NKX2-5, RBM45, ZMYND11, MEN1, L3MBTL3, ZIC1, ZBTB2, KMT2A, SKP1, ACTR5, NFIC, TFAP2A, PBX1, KDM2B, MYC, BCOR, ESRRA, AHR, TFAP4, RUNX2, MAX, RBBP5, HCFC2, FLYWCH1, TFAM, RNF2, KDM2A
hmC/hmC>mC/mC	5	KRR1, SRFBP1, KLHL7, RADX, TOX
hmC/hmC<mC/mC	22	DEK, TBX21, MBD2, SUMO2, MBD4, ZFHX2, RFXANK, ZFHX3, MTA2, ATF2, CBX3, SUB1, RFX7, TBX2, ZBTB44, SIX1, CDX1, CREB5, RBBP7, L3MBTL2, FOXA1, RFXAP

Table A.15: Experiment 1 C/C vs hmC/C.

Name	Change	Fold change (log2)	Significance	Manhattan distance
BCOR	Increased	3,991450119	8,49268371	12,48413383
KMT2A	Increased	2,986585236	9,295593004	12,28217824
KMT2B	Increased	4,614106369	7,448573232	12,0626796
ZBTB2	Increased	4,558032227	6,155722494	10,71375472
KDM2B	Increased	3,082039642	6,945701251	10,02774089
MTF2	Increased	2,327380753	6,799128246	9,126508999
CTCF	Increased	3,476277161	5,047641836	8,523918997
MEN1	Increased	2,304178619	5,932099138	8,236277757
RBBP5	Increased	2,77963829	5,216797241	7,996435531
SAMD1	Increased	1,630738831	6,235982996	7,866721827
PCGF1	Increased	2,750312424	4,83009786	7,580410284
DPY30	Increased	1,918320084	5,390688662	7,309008746
ASH2L	Increased	2,128744888	5,130283406	7,259028294
ZIC2	Increased	2,231959915	4,205221792	6,437181707
HCFC2	Increased	1,769738007	4,452009451	6,221747458
SKP1	Increased	1,967718124	4,112345876	6,080064
ZMYND11	Increased	1,534513092	4,313799695	5,848312787
E2F6	Increased	1,531050491	4,082758076	5,613808567
NFIX	Increased	1,655892563	3,798495231	5,454387794
CENPB	Increased	1,843591309	3,358826337	5,202417646
FLYWCH1	Increased	1,788798904	3,368600254	5,157399158
EXOSC5	Decreased	-3,222357178	6,259051133	9,481408311
UHRF1	Decreased	-2,296754837	2,911922716	5,208677553
SRP68	Decreased	-1,761000824	1,722502507	3,483503331

Table A.16: Experiment 2 C/C vs hmC/C.

Name	Change	Fold change (log2)	Significance	Manhattan distance
KDM2B	Increased	5,143182755	8,94604341	14,08922617
ZBTB2	Increased	5,227718353	7,97627156	13,20398991
TFAP4	Increased	3,003242874	9,212235021	12,2154779
NFIX	Increased	2,311558914	7,364117243	9,675676157
BCOR	Increased	2,877800751	5,956124935	8,833925686
CENPB	Increased	3,060686111	5,237911627	8,298597738
PCGF1	Increased	2,828242111	4,7760254	7,604267511
HCFC2	Increased	1,613341141	4,92248597	6,535827111
KDM2A	Increased	2,21642952	3,767081363	5,983510883
NFIC	Increased	1,865048218	3,067673901	4,932722119
RBM45	Decreased	-1,564663696	11,34536117	12,91002487

Name	Change	Fold change (log2)	Significance	Manhattan distance
MECP2	Decreased	-4,362238312	5,05262275	9,414861062
UHRF1	Decreased	-1,579175949	5,296139862	6,875315811
CDCA7L	Decreased	-1,624069977	4,862487258	6,486557235

Table A.17: Experiment 1 hmC/C vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
SAMD1	Increased	6,123135757	8,402495	14,52563076
BEND3	Increased	6,96189537	5,811328085	12,77322346
L3MBTL3	Increased	6,921005249	5,316647037	12,23765229
ATF1	Increased	2,686919022	8,003767693	10,69068672
ZIC2	Increased	2,836654663	7,364193577	10,20084824
NKX2-5	Increased	2,250564575	6,202744815	8,45330939
MAX	Increased	1,596856689	6,71737859	8,314235279
RBM45	Increased	2,599118805	4,991385528	7,590504333
BCOR	Increased	2,901272583	3,923068092	6,824340675
TFAM	Increased	1,603358078	5,06255792	6,665915998
ASH2L	Increased	1,789498901	4,577356303	6,366855204
ZIC1	Increased	2,003360748	3,204928539	5,208289287
BANP	Increased	2,019288635	3,030419046	5,049707681
RBBP5	Increased	1,63278923	2,905646122	4,538435352
SYNCRIP	Increased	2,695048904	1,836610746	4,53165965
SIX1	Increased	2,563137054	1,733219427	4,296356481
PAX1	Increased	1,862688828	2,419487605	4,282176433
LDB1	Increased	1,845508575	1,988162082	3,833670657
UHRF1	Decreased	-3,117865753	7,862430124	10,98029588
DACH1	Decreased	-1,708934402	6,86191049	8,570844892
SRFBP1	Decreased	-1,731120682	6,56872709	8,299847772
DNAJB3	Decreased	-1,895962524	5,01243306	6,908395584
FOXC1	Decreased	-2,793143082	2,920591177	5,713734259
BAZ1B	Decreased	-2,132674408	2,782404386	4,915078794

Table A.18: Experiment 2 hmC/C vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
BEND3	Increased	6,805451965	6,761455662	13,56690763
SAMD1	Increased	2,160818863	9,610196494	11,77101536
RBM45	Increased	2,385865021	8,706724293	11,09258931
CCAR2	Increased	1,575449753	9,148013452	10,72346321
CREB1	Increased	2,267418671	8,237244993	10,50466366
NKX2-1	Increased	2,089812469	8,177778506	10,26759098
L3MBTL3	Increased	2,007976151	8,201425783	10,20940193
BCOR	Increased	3,650576401	5,983005647	9,633582048
BANP	Increased	4,423986053	4,287077081	8,711063134
KDM2B	Increased	2,632518387	6,022413863	8,65493225
NKX2-3	Increased	2,08437233	6,337794526	8,422166856
CDCA7L	Increased	2,966637421	4,892246753	7,858884174
INO80C	Increased	1,793208313	5,947746066	7,740954379
RFX5	Increased	2,78036499	4,768565336	7,548930326
AHR	Increased	1,521578598	5,115212791	6,636791389
ARNT	Increased	1,542472839	3,415744023	4,958216862
TFPT	Increased	1,829528046	2,530492489	4,360020535
RUNX2	Increased	1,570289993	2,472313563	4,042603556
PCGF1	Increased	1,631491852	2,105366434	3,736858286
SSBP2	Increased	1,796916962	1,685407144	3,482324106
UHRF1	Decreased	-1,723506546	7,982878627	9,706385173
DACH1	Decreased	-1,501672745	6,842902902	8,344575647
CHAF1A	Decreased	-1,941672134	6,190917336	8,13258947
CHAF1B	Decreased	-1,809923553	5,773512176	7,583435729

Table A.19: Experiment 1 mC/mC vs hmC/C.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFX5	Increased	8,426774216	7,168584193	15,59535841
RFXAP	Increased	4,362547302	7,272215779	11,63476308
MBD4	Increased	3,142383194	8,034188051	11,17657125
RFXANK	Increased	4,958098221	5,923210103	10,88130832
FOXC1	Increased	5,359923172	4,881423982	10,24134715
UHRF1	Increased	2,514299393	7,044304051	9,558603444
SUB1	Increased	2,210567474	7,264405938	9,474973412
DACH1	Increased	1,852444458	6,912226021	8,764670479
MGA	Increased	1,554565811	7,208619027	8,763184838
DEK	Increased	2,34741478	6,104283098	8,451697878
E2F6	Increased	2,103806305	5,794265086	7,898071391
FOXA1	Increased	2,286525726	5,520746187	7,807271913
L3MBTL2	Increased	2,020093536	5,605683369	7,625776905
TBX2	Increased	1,573542786	5,410284618	6,983827404
PCGF6	Increased	1,569924927	4,028657061	5,598581988
MTA2	Increased	1,614209747	2,649940609	4,264150356
BEND3	Decreased	-7,527812958	7,221188792	14,74900175
SAMD1	Decreased	-5,844467545	8,376215849	14,22068339
ZIC2	Decreased	-3,364835739	8,471198924	11,83603466
L3MBTL3	Decreased	-6,35104084	4,445123118	10,79616396
MAX	Decreased	-1,798546982	7,589881293	9,388428275
EXOSC5	Decreased	-2,798733521	5,105237717	7,903971238
BANP	Decreased	-2,479319382	4,76726181	7,246581192
NKX2-5	Decreased	-1,880377579	5,306789769	7,187167348
MYC	Decreased	-2,210822296	4,942680772	7,153503068
FLYWCH1	Decreased	-1,775429153	4,519296842	6,294725995
RBM45	Decreased	-2,743015289	2,918259717	5,661275006
MAPRE2	Decreased	-3,690244675	1,881949694	5,572194369
CSTA	Decreased	-2,675920105	2,551870585	5,22779069
ZIC1	Decreased	-2,147966766	2,766378343	4,914345109
BCOR	Decreased	-2,378233719	2,218776398	4,597010117
TFAP2A	Decreased	-2,076324844	2,337825475	4,414150319
PHF3	Decreased	-1,898693848	2,463815811	4,362509659
PRPF4B	Decreased	-1,922904205	2,074395606	3,997299811
SF3A3	Decreased	-1,641815186	1,932155733	3,573970919
AIFM1	Decreased	-1,673772049	1,798736897	3,472508946

Table A.20: Experiment 2 mC/mC vs hmC/C.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFX5	Increased	6,444791412	11,14292434	17,58771575
FOXC1	Increased	2,871783066	9,531299565	12,40308263
ATF2	Increased	2,38225975	8,765327266	11,14758702
RFXANK	Increased	4,803175735	5,56239351	10,36556925
RFXAP	Increased	4,666832733	5,359110654	10,02594339
MBD2	Increased	3,040773392	5,441423234	8,482196626
MECP2	Increased	1,578052902	6,58243936	8,160492262
ZFHX3	Increased	1,932154465	6,02167013	7,953824595
MTA2	Increased	1,683507919	5,970827112	7,654335031
RFX7	Increased	2,818942261	4,681597338	7,500539599
FOXA1	Increased	2,253225708	5,184576104	7,437801812
RFX1	Increased	3,515278625	3,63341693	7,148695555
FOXF1	Increased	1,681685257	5,377141685	7,058826942
SUB1	Increased	2,673327255	2,794734532	5,468061787
CREB5	Increased	1,904407501	2,067561686	3,971969187
BUB3	Increased	1,617709732	1,819558259	3,437267991
BEND3	Decreased	-7,108484268	10,08075813	17,1892424
RBM45	Decreased	-2,199285507	10,29454252	12,49382803
SAMD1	Decreased	-1,985751724	10,31497287	12,30072459

Name	Change	Fold change (log2)	Significance	Manhattan distance
L3MBTL3	Decreased	-1,974924088	9,039381553	11,01430564
TERF2IP	Decreased	-1,5157444	9,07966667	10,59541107
MYC	Decreased	-2,507802963	7,797472519	10,30527548
CCAR2	Decreased	-1,584217834	8,13865628	9,722874114
BCOR	Decreased	-3,705136871	5,639648101	9,344784972
CDCA7L	Decreased	-1,95133934	7,359051208	9,310390548
KDM2B	Decreased	-2,578993607	6,317815676	8,896809283
MAX	Decreased	-1,79046936	6,73683842	8,52730778
BANP	Decreased	-2,139658356	4,783837215	6,923495571
AHR	Decreased	-2,333420563	3,786514648	6,119935211
ARNT	Decreased	-1,886992264	3,613893189	5,500885453
CENPB	Decreased	-1,610916519	3,37477151	4,985688029
TFPT	Decreased	-1,636933517	1,986793	3,623726517

Table A.21: Venn diagram readers for C/C, hmC/C, mC/mC, and hmC/hmC from HEK293T. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, hmC/C, mC/mC and hmC/hmC modifications in at least one out of two biological experiments (experiments 1 and 2) associated with the Venn diagram in Figure 3.19.

Enriched for	Number of proteins	Protein names
hmC/C>C/C, hmC/C>hmC/hmC, hmC/C>mC/mC	2	CDCA7L, RBM45
hmC/C>C/C, hmC/C<hmC/hmC, hmC/C<mC/mC	1	UHRF1
hmC/C>hmC/hmC, hmC/C>mC/mC, hmC/C<C/C	4	SAMD1, ZIC2, KDM2B, BCOR
hmC/C>C/C, hmC/C>mC/mC	1	EXOSC5
hmC/C>C/C, hmC/C<mC/mC	1	MECP2
hmC/C>mC/mC, hmC/C<C/C	2	CENPB, FLYWCH1
hmC/C<C/C, hmC/C<mC/mC	1	E2F6
hmC/C>hmC/hmC, hmC/C<C/C	3	PCGF1, ASH2L, RBBP5
hmC/C>hmC/hmC, hmC/C>mC/mC	10	ARNT, CCAR2,, BEND3, BANP, TFPT, NKX2-5, L3MBTL3, ZIC1, AHR, MAX
hmC/C>hmC/hmC, hmC/C<mC/mC	1	RFX5
hmC/C<hmC/hmC, hmC/C<mC/mC	2	FOXC1, DACH1
hmC/C>C/C	1	SRP68
hmC/C<C/C	14	DPY30, MTF2, CTCF, KMT2B, NFIX, ZMYND11, MEN1, ZBTB2, KMT2A, SKP1, NFIC, TFAP4, HCFC2, KDM2A
hmC/C>mC/mC	9	CSTA, PRPF4B, AIFM1, TERF2IP, SF3A3, PHF3, TFAP2A, MYC, MAPRE2
hmC/C<mC/mC	19	FOXF1, DEK, MBD2, BUB3, MBD4, RFXANK, ZFHX3, MTA2, ATF2, PCGF6, SUB1, RFX7, TBX2, MAG, CREB5, RFX1, L3MBTL2, FOXA1, RFXAP
hmC/C>hmC/hmC	12	ATF1, NKX2-1, INO80C, NKX2-3, SSBP2, SYNCRIP, LDB1, SIX1, RUNX2, PAX1, TFAM, CREB1
hmC/C<hmC/hmC	5	CHAF1B, BAZ1B, CHAF1A, SRFBP1, DNAJB3

Table A.22: Experiment 1 C/C vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
KMT2A	Increased	3,682699585	11,99193292	15,67463251
SAMD1	Increased	7,717199326	7,729820266	15,44701959
BEND3	Increased	8,366173172	6,359366161	14,72553933
ASH2L	Increased	3,616127777	9,72395549	13,34008327
RBBP5	Increased	3,701060104	9,058843782	12,75990389

Name	Change	Fold change (log2)	Significance	Manhattan distance
ZIC2	Increased	5,758614349	6,980272187	12,73888654
KMT2B	Increased	5,836121368	6,785948528	12,6220699
CTCF	Increased	4,737511444	7,512214748	12,24972619
ATF1	Increased	3,780293274	8,083647024	11,8639403
ZBTB2	Increased	4,799263763	6,440405748	11,23966951
L3MBTL3	Increased	6,704005051	4,072738363	10,77674341
NKX2-5	Increased	2,308169174	8,402771854	10,71094103
KDM2B	Increased	4,107843018	6,434204327	10,54204735
BCOR	Increased	5,97955513	4,39000437	10,3695595
BANP	Increased	3,750226974	6,494632833	10,24485981
MEN1	Increased	2,487025833	7,638995458	10,12602129
TFAM	Increased	1,741912079	8,057947012	9,799859091
ESRRA	Increased	1,612360382	7,937237321	9,549597703
MAX	Increased	2,980612183	6,548724359	9,529336542
FLYWCH1	Increased	3,906797028	5,445973419	9,352770447
DPY30	Increased	2,871689987	6,218282629	9,089972616
SKP1	Increased	2,413048172	6,234613523	8,647661695
MYC	Increased	3,05333252	4,896726169	7,950058689
MAZ	Increased	1,950139999	5,795124091	7,74526409
PCGF1	Increased	2,566146851	4,906588781	7,472735632
CENPB	Increased	2,279859543	4,638681014	6,918540557
MTF2	Increased	2,217001343	4,684612619	6,901613962
AHR	Increased	3,032427979	3,784851049	6,817279028
NFIX	Increased	2,200575638	3,65486186	5,855437498
RBM45	Increased	1,863879395	3,974011965	5,83789136
ZMYND11	Increased	2,302122116	3,317726142	5,619848258
ZIC1	Increased	2,942087173	2,599981887	5,54206906
EHMT1	Increased	1,592380142	3,822169575	5,414549717
HCFC2	Increased	1,981154251	3,308457369	5,28961162
RUVBL1	Increased	1,983606339	1,786718167	3,770324506
LDB1	Increased	1,818797684	1,696417688	3,515215372
RFX5	Decreased	-4,013798523	8,488304397	12,50210292
FOXC1	Decreased	-5,360597992	5,764829118	11,12542711
SUB1	Decreased	-1,906903839	7,808980328	9,715884167
MBD4	Decreased	-2,809260941	4,98962292	7,798883861
UHRF1	Decreased	-3,416285324	4,083985331	7,500270655
FOXA1	Decreased	-2,114162064	4,195825375	6,309987439
RFXAP	Decreased	-2,065058899	3,994269962	6,059328861
RFXANK	Decreased	-1,568976212	3,941262161	5,510238373
ZC3HAV1	Decreased	-1,847460175	3,532620987	5,380081162
CELSR1	Decreased	-1,850858307	1,835143184	3,686001491

Table A.23: Experiment 2 C/C vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
KDM2B	Increased	7,842909241	7,560002739	15,40291198
BANP	Increased	4,110921478	10,86507228	14,97599376
BCOR	Increased	6,740979004	6,327124199	13,0681032
ZBTB2	Increased	5,890924454	6,42657649	12,31750094
BEND3	Increased	7,685234833	4,553070917	12,23830575
CENPB	Increased	4,590294266	7,025432072	11,61572634
SAMD1	Increased	2,602989197	8,993727409	11,59671661
ZIC2	Increased	2,228489304	9,319908231	11,54839754
KMT2A	Increased	1,514105988	9,333777841	10,84788383
MAX	Increased	2,439654541	8,268806631	10,70846117
L3MBTL3	Increased	2,42359314	8,21984783	10,64344097
MEN1	Increased	1,788851547	8,841331629	10,63018318
MYC	Increased	2,813244629	7,764205708	10,57745034
ZIC1	Increased	2,112292099	8,313974079	10,42626618
TFAM	Increased	1,755049515	7,366026409	9,121075924
NKX2-1	Increased	1,530082321	7,43211757	8,962199891

Name	Change	Fold change (log2)	Significance	Manhattan distance
ESRRA	Increased	1,553895569	7,357461684	8,911357253
FLYWCH1	Increased	2,009353638	6,846795227	8,856148865
CREB1	Increased	2,889710999	5,878707636	8,768418635
NFIX	Increased	2,31493187	6,225514127	8,540445997
ASH2L	Increased	1,556942368	6,786606207	8,343548575
INO80C	Increased	2,102071762	6,195821214	8,297892976
RBBP5	Increased	1,600254822	6,520751647	8,121006469
PCGF1	Increased	3,292501068	4,743595785	8,036096853
VEZF1	Increased	1,745923615	5,918284666	7,664208281
TFAP4	Increased	2,856555939	4,771957476	7,628513415
DPY30	Increased	1,810879517	5,441842862	7,252722379
HCFC2	Increased	1,94655838	5,292443607	7,239001987
CBFB	Increased	1,914196396	4,987028466	6,901224862
AHR	Increased	2,824449921	4,006989834	6,831439755
NFIC	Increased	3,095734024	3,63166817	6,727402194
TBRG1	Increased	2,291692352	4,015358683	6,307051035
NKX2-3	Increased	1,787516403	4,092394653	5,879911056
TFPT	Increased	1,910962677	3,388177605	5,299140282
ARNT	Increased	2,250520325	2,945862141	5,196382466
RING1	Increased	1,979864502	3,140208075	5,120072577
KDM2A	Increased	1,842389297	2,55690328	4,399292577
RUNX2	Increased	1,836637115	1,611779834	3,448416949
TIMM8A	Increased	1,658106232	1,710305969	3,368412201
RFX5	Decreased	-4,69571228	9,752324985	14,44803727
FOXC1	Decreased	-2,758992767	10,31252375	13,07151652
MECP2	Decreased	-5,637297821	5,384437059	11,02173488
RFXAP	Decreased	-3,553304291	7,397725799	10,95103009
UHRF1	Decreased	-2,971342087	7,695501618	10,66684371
FOXA1	Decreased	-1,884148788	5,121686072	7,00583486
RFXANK	Decreased	-2,199887848	4,338125152	6,538013
ZFHX3	Decreased	-1,798093796	3,715570149	5,513663945
FOXF1	Decreased	-1,729928207	3,080952139	4,810880346
SUB1	Decreased	-2,857164764	1,790669424	4,647834188
PTBP3	Decreased	-1,951524353	1,620675868	3,572200221
SYF2	Decreased	-1,742301559	1,665109794	3,407411353

Table A.24: Experiment 1 mC/mC vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFX5	Increased	3,312500763	9,42102946	12,73353022
RFXANK	Increased	3,838320541	8,672145361	12,5104659
DEK	Increased	2,466106415	6,934074206	9,400180621
RFXAP	Increased	2,467959976	6,114883363	8,582843339
ATF1	Increased	1,670936203	4,515431305	6,186367508
RFX7	Increased	2,770951462	2,774036048	5,54498751
PCGF6	Increased	1,518431854	3,519877212	5,038309066
MTA2	Increased	1,520041656	2,155812026	3,675853682
LDB1	Increased	1,697550583	1,63762235	3,335172933
KRR1	Decreased	-2,190394211	4,226414605	6,416808816
NOP16	Decreased	-3,392940521	2,113080306	5,506020827
TFAP2A	Decreased	-2,236557388	2,337189195	4,573746583

Table A.25: Experiment 2 mC/mC vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFX5	Increased	1,718598938	7,468044084	9,186643022
MBD2	Increased	3,176625824	5,904289871	9,080915695
MTA2	Increased	1,679240036	6,968369567	8,647609603
RFX1	Increased	2,770393372	5,635833195	8,406226567

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFXAP	Increased	1,539928436	5,362764026	6,902692462
RFXANK	Increased	2,447613907	3,690035395	6,137649302
CFL1	Increased	2,395165634	3,662955411	6,058121045
RFX7	Increased	2,85177269	3,205187931	6,056960621
SUMO2	Increased	2,47405014	2,954374107	5,428424247
BUB3	Increased	1,606002045	2,730419432	4,336421477
HMGB2	Increased	1,982513428	2,171067094	4,153580522
EIF5A	Increased	2,131801987	1,787725297	3,919527284
MDH2	Increased	1,629403305	2,198339923	3,827743228
PPIA	Increased	1,644240952	1,699710285	3,343951237

Table A.26: Experiment 1 hmC/mC vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
SUB1	Increased	1,570610046	7,258737106	8,829347152
FOXC1	Increased	2,05399971	5,972340008	8,026339718
RFXAP	Increased	2,317954254	3,625198324	5,943152578
NOP16	Increased	3,61055603	2,151167913	5,761723943
RFX5	Increased	3,253153419	1,913996415	5,167149834
SIX1	Increased	2,532445908	2,244378813	4,776824721
FOXA1	Increased	1,549806213	3,16200181	4,711808023
UHRF1	Decreased	-1,998335266	6,719014752	8,717350018

Table A.27: Experiment 2 hmC/mC vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFX5	Increased	7,506557465	8,185001955	15,69155942
RFXAP	Increased	3,476562119	3,500268435	6,976830554
RFXANK	Increased	2,566139221	3,717501443	6,283640664
RFX1	Increased	1,752167892	2,51861973	4,270787622
SSBP2	Increased	1,687703705	1,654004542	3,341708247
HNRNPLL	Decreased	-2,19789238	2,338201923	4,536094303

Table A.28: Experiment 1 hmC/C vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
BEND3	Increased	7,45776062	7,362669495	14,82043012
L3MBTL3	Increased	6,117850494	5,007692027	11,12554252
RBM45	Increased	3,36866188	7,627980803	10,99664268
SAMD1	Increased	5,653784943	4,967538293	10,62132324
MYC	Increased	2,307945251	6,784829001	9,092774252
ATF1	Increased	2,33519783	6,400911818	8,736109648
NKX2-5	Increased	2,051623535	5,890785456	7,942408991
ZIC2	Increased	2,992842484	4,638885383	7,631727867
MAX	Increased	2,103765869	5,442450608	7,546216477
BANP	Increased	2,575525284	4,184233243	6,759758527
FLYWCH1	Increased	1,854016495	4,455471072	6,309487567
ZIC1	Increased	1,609263229	3,730321978	5,339585207
PAX1	Increased	2,288753128	2,32361943	4,612372558
LDB1	Increased	1,667733765	1,768098315	3,43583208
RFX5	Decreased	-5,114273453	5,472305041	10,58657849
FOXC1	Decreased	-4,847142792	4,557955671	9,405098463
SUB1	Decreased	-1,844093323	7,255924956	9,100018279
MBD4	Decreased	-2,367697144	6,250886113	8,618583257
RFXAP	Decreased	-2,021530533	4,098659258	6,120189791
FOXA1	Decreased	-1,558623123	3,85863693	5,417260053
NOP16	Decreased	-3,117580795	1,720548817	4,838129612

Table A.29: Experiment 2 hmC/C vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
BANP	Increased	3,378670883	10,43912959	13,81780047
BEND3	Increased	7,830770493	4,889669416	12,72043991
SAMD1	Increased	1,972012329	9,287154095	11,25916642
RBM45	Increased	1,785395813	8,869944617	10,65534043
CCAR2	Increased	1,581245041	8,491674756	10,0729198
ARNT	Increased	1,980984879	7,625452807	9,606437686
L3MBTL3	Increased	2,038714218	7,551279322	9,58999354
NKX2-1	Increased	1,611144257	7,952440487	9,563584744
MAX	Increased	1,823957443	7,473169295	9,297126738
MYC	Increased	2,3147995	6,645763028	8,960562528
BCOR	Increased	3,863178253	4,55665622	8,419834473
INO80C	Increased	1,727064896	5,875826133	7,602891029
CDCA7L	Increased	2,299365616	4,606338169	6,905703785
KDM2B	Increased	2,699726486	3,809797992	6,509524478
NKX2-3	Increased	1,562277985	4,142832207	5,705110192
AHR	Increased	2,18690834	3,311242556	5,498150896
TFPT	Increased	2,002441788	2,340739858	4,343181646
RFX5	Decreased	-4,726192474	9,944232845	14,67042532
FOXC1	Decreased	-2,747454834	10,23002458	12,97747941
RFXAP	Decreased	-3,500031281	7,577380202	11,07741148
RFXANK	Decreased	-2,40447731	5,294831501	7,699308811
SUB1	Decreased	-2,893304443	3,070508867	5,96381331
FOXA1	Decreased	-1,889966202	3,329251296	5,219217498
BAG2	Decreased	-1,552868271	1,842318696	3,395186967

Table A.30: Venn diagram readers for C/C, hmC/C, mC/mC, hmC/mC and hmC/hmC from HEK293T. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, hmC/C, mC/mC, hmC/mC and hmC/hmC modifications in at least one out of two biological experiments (experiments 1 and 2) associated with the Venn diagram in Figure 3.20.

Enriched for	Number of proteins	Protein names
hmC/mC>C/C, hmC/mC>hmC/C, hmC/mC>hmC/hmC, hmC/mC<mC/mC	3	RFXANK, RFX5, RFXAP
hmC/mC>C/C, hmC/mC>hmC/C, hmC/mC>hmC/hmC	3	FOXC1, SUB1, FOXA1
hmC/mC<C/C, hmC/mC<hmC/C, hmC/mC<mC/mC	2	ATF1, LDB1
hmC/mC>hmC/C, hmC/mC>hmC/hmC, hmC/mC>mC/mC	1	NOP16
hmC/mC>C/C, hmC/mC<hmC/hmC	1	UHRF1
hmC/mC>C/C, hmC/mC>hmC/C	1	MBD4
hmC/mC<C/C, hmC/mC<hmC/C	19	ARNT, NKX2-1, SAMD1, BEND3, INO80C, ZIC2, BANP, NKX2-3, TFPT, NKX2-5, RBM45, L3MBTL3, ZIC1, KDM2B, MYC, BCOR, AHR, MAX, FLYWCH1
hmC/mC>hmC/hmC, hmC/mC<mC/mC	1	RFX1
hmC/mC>C/C	7	FOXF1, ZC3HAV1, MECP2, ZFHX3, SYF2, PTBP3, CELSR1
hmC/mC<C/C	30	CENPB, DPY30, TBRG1, MTF2, PCGF1, EHMT1, CTCF, ASH2L, KMT2B, VEZF1, RUVBL1, MAZ, NFIX, ZMYND11, MEN1, ZBTB2, CBFB, KMT2A, SKP1, NFIC, RING1, ESRRA, TFAP4, RUNX2, RBBP5, HCFC2, TIMM8A, TFAM, KDM2A, CREB1
hmC/mC>mC/mC	2	KRR1, TFAP2A
hmC/mC<mC/mC	12	DEK, CFL1, MBD2, BUB3, SUMO2, PPIA, MTA2, MDH2, PCGF6, HMGB2, EIF5A, RFX7
hmC/mC>hmC/hmC	2	SSBP2, SIX1

Enriched for	Number of proteins	Protein names
hmC/mC<hmC/hmC	1	HNRNPLL
hmC/mC>hmC/C	1	BAG2
hmC/mC<hmC/C	3	CDCA7L, CCAR2, PAX1

Table A.31: Venn diagram readers from HEK293T experiments. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, hmC/C, mC/mC, hmC/mC and hmC/hmC modifications in at least one out of two biological experiments (experiments 1 and 2) associated with the Venn diagram in Figure 3.21.

CpG dyad	Number of proteins	Protein names
C/C, hmC/C, hmC/mC, mC/mC	2	RFX5, SSBP2
hmC/C, hmC/hmC, hmC/mC, mC/mC	2	UHRF1, MECP2
C/C, hmC/C, mC/mC	4	ATF1, BANP, LDB1, CREB1
C/C, hmC/C, hmC/mC	1	TFAP2A
C/C, hmC/mC, mC/mC	1	RFX1
hmC/C, hmC/mC, mC/mC	1	SIX1
hmC/hmC, hmC/mC, mC/mC	1	FOXC1
C/C, hmC/C	24	ARNT, ASH2L, BEND3, TFPT, NKX2-5, ZIC1, MYC, BCOR, MAX, RBBP5, TFAM, CENPB, NKX2-1, SAMD1, PCGF1, ZIC2, INO80C, NKX2-3, RBM45, L3MBTL3, KDM2B, AHR, RUNX2, FLYWCH1
C/C, mC/mC	3	PCGF6, E2F6, MGA
hmC/C, hmC/hmC	2	EXOSC5, SF3A3
hmC/mC, mC/mC	8	SUB1, FOXF1, MBD4, RFXANK, ZFHX3, PTBP3, FOXA1, RFXAP
hmC/hmC, mC/mC	1	DACH1
hmC/hmC, hmC/mC	3	CELSR1, KRR1, SYF2
C/C	31	SMAD1, DPY30, MTF2, WDR5, KMT2B, VEZF1, ZMYND11, MEN1, ZBTB2, CBFB, NFIC, ESRRA, HCFC2, KDM2A, TBRG1, EHMT1, CTCF, MSANTD3, RUVBL1, ZBTB10, MAZ, NFIX, KMT2A, ACTR5, SKP1, HEY1, PBX1, RING1, TFAP4, TIMM8A, RNF2
hmC/C	11	CDCA7L, CCAR2, CSTA, SRP68, PRPF4B, AIFM1, SYNCRIP, PAX1, TERF2IP, PHF3, MAPRE2
mC/mC	21	DEK, TBX21, CFL1, BUB3, PPIA, ZFHX2, MTA2, ATF2, MDH2, CBX3, RFX7, TBX2, RBBP7, L3MBTL2, MBD2, SUMO2, HMGB2, EIF5A, ZBTB44, CDX1, CREB5
hmC/mC	3	ZC3HAV1, BAG2, NOP16
hmC/hmC	11	CHAF1A, SRFBP1, DNAJB3, RADX, CHAF1B, HNRNPLL, BAZ1B, RPP25L, KLHL7, RBM14, TOX

Table A.32: Mouse brain C/C vs mC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Lurap1	Increased	4,374020386	2,284489275	6,658509661
Creb1	Increased	1,70856514	4,047642477	5,756207617
Trps1	Increased	1,901485062	3,265637528	5,16712259
Mid2	Increased	1,965319061	3,104633267	5,069952328
Clock	Increased	1,607664871	2,319883211	3,927548082
Baz2b	Increased	1,531928253	2,22440884	3,756337093
Satb2	Decreased	-4,216363907	5,044034683	9,26039859
Mbd2	Decreased	-2,168481827	5,66720921	7,835691037
Mta2	Decreased	-1,832439804	5,561764179	7,394203983

Name	Change	Fold change (log2)	Significance	Manhattan distance
Gatad2b	Decreased	-1,556266785	4,735345858	6,291612643
Foxk1	Decreased	-1,736593246	3,91973821	5,656331456
Rbbp7	Decreased	-2,658253479	2,786861115	5,445114594
Foxp1	Decreased	-1,590343857	3,851641567	5,441985424
Satb1	Decreased	-1,803749847	3,522807059	5,326556906
Foxk2	Decreased	-1,742658615	3,024890358	4,767548973

Table A.33: Mouse brain C/C vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Creb1	Increased	3,753271866	7,855588236	11,6088601
Mafg	Increased	2,821353912	5,888378791	8,709732703
Jund	Increased	2,995639801	4,127584449	7,12322425
Parp1	Increased	1,574351501	5,001847811	6,576199312
Arntl	Increased	1,690711212	3,473111127	5,163822339
Zbtb2	Increased	1,865829849	3,224577208	5,090407057
Hnrnpk	Increased	2,466888428	2,53186657	4,998754998
Ring1	Increased	1,757934189	2,103164509	3,861098698
Gm10094	Increased	1,633867264	1,776696778	3,410564042
Mtmr10	Increased	1,628496552	1,77699355	3,405490102
Uhrf2	Decreased	-2,459907532	4,564393842	7,024301374
Satb2	Decreased	-1,745071411	5,025577786	6,770649197
Sfpq	Decreased	-2,707499313	3,506723993	6,214223306
Lrrfip2	Decreased	-2,404348755	2,931692604	5,336041359

Table A.34: Mouse brain mC/mC vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Jund	Increased	2,494986725	5,054160349	7,549147074
Mbd2	Increased	1,698999786	5,788097918	7,487097704
Satb2	Increased	1,986613846	4,80652388	6,793137726
Mta2	Increased	1,594087601	4,990438789	6,58452639
Foxk2	Increased	2,431923676	3,94108006	6,373003736
Creb1	Increased	2,044706726	4,096106321	6,140813047
Atp6v1d	Increased	1,65866394	4,414992691	6,073656631
Foxk1	Increased	2,086003876	3,68149298	5,767496856
Satb1	Increased	2,102927017	3,660293533	5,76322055
Foxp1	Increased	1,907490921	3,776698174	5,684189095
Bend6	Increased	2,814349747	2,67493884	5,489288587
Rplp1	Increased	1,893349457	2,062004392	3,955353849
Rasgrp2	Increased	1,534702682	2,374255587	3,908958269
Rbbp7	Increased	1,504867554	1,788415338	3,293282892
Sfpq	Decreased	-3,453072357	4,05840654	7,511478897
Uhrf2	Decreased	-1,9411026	3,399205408	5,340308008

Table A.35: Venn diagram readers for C/C, mC/mC and hmC/hmC from mouse brain. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, mC/mC and hmC/hmC modifications associated with the Venn diagram in Figure 3.25.

Enriched for	Number of proteins	Protein names
C/C>hmC/hmC, mC/mC>hmC/hmC	2	Jund, Creb1
hmC/hmC>C/C, mC/mC>hmC/hmC	1	Satb2
hmC/hmC>C/C, hmC/hmC>mC/mC	2	Sfpq, Uhrf2
C/C>hmC/hmC	8	Mtmr10, Arntl, Zbtb2, Parp1, Hnrnpk, Mafg, Gm10094, Ring1
hmC/hmC>C/C	1	Lrrfip2

Enriched for	Number of proteins	Protein names
mC/mC>hmC/hmC	11	Mbd2, Atp6v1d, Mta2, Rbbp7, Satb1, Foxp1, Foxk1, Bend6, Foxk2, Rasgrp2, Rplp1

Table A.36: Mouse brain C/C vs hmC/C.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Mafg	Increased	2,424416351	3,310904538	5,735320889
Creb1	Increased	1,611184692	3,646301546	5,257486238
Jund	Increased	2,035581589	1,828516874	3,864098463

Table A.37: Mouse brain hmC/C vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Ndufs1	Increased	1,73357811	5,223771022	6,957349132
Creb1	Increased	2,142087173	4,070129002	6,212216175
Hnrnpk	Increased	2,588893509	3,02008122	5,608974729
Rplp1	Increased	2,214067459	2,020281489	4,234348948
Haus3	Increased	1,782195282	1,661791592	3,443986874
Sfpq	Decreased	-2,747096634	4,23804937	6,985146004
Satb2	Decreased	-1,885617828	3,080831315	4,966449143

Table A.38: Mouse brain mC/mC vs hmC/C.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Satb2	Increased	3,610266113	5,368223774	8,978489887
Mbd2	Increased	1,823352814	5,320334638	7,143687452
Rbbp7	Increased	2,801776886	4,061156823	6,862933709
Foxk2	Increased	1,931587601	3,725985553	5,657573154
Foxp1	Increased	1,76789856	3,78148096	5,54937952
Mta2	Increased	1,548067093	3,856151011	5,404218104
Dynlt3	Increased	2,375720596	2,432075	4,807795596
Bend6	Increased	2,320231247	1,923545794	4,243777041
Mid2	Decreased	-2,083306885	2,056072544	4,139379429
Trps1	Decreased	-1,598176575	2,38478801	3,982964585
Hnrnpk	Decreased	-1,701974487	1,72258985	3,424564337

Table A.39: Venn diagram readers for C/C, hmC/C, mC/mC and hmC/hmC from mouse brain. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, hmC/C, mC/mC and hmC/hmC modifications associated with the Venn diagram in Figure 3.26.

Enriched for	Number of proteins	Protein names
C/C>hmC/C, hmC/C>hmC/hmC	1	Creb1
hmC/hmC>hmC/C, mC/mC>hmC/C	1	Satb2
hmC/C>hmC/hmC, hmC/C>mC/mC	1	Hnrnpk
C/C>hmC/C	2	Jund, Mafg
mC/mC>hmC/C	7	Mbd2, Dynlt3, Mta2, Rbbp7, Foxp1, Bend6, Foxk2
hmC/C>mC/mC	2	Mid2, Trps1
hmC/C>hmC/hmC	3	Haus3, Ndufs1, Rplp1
hmC/hmC>hmC/C	1	Sfpq

Table A.40: Mouse brain C/C vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Creb1	Increased	3,613555145	6,690040549	10,30359569
Zbtb2	Increased	2,02804451	5,024639123	7,052683633
Parp1	Increased	1,60564003	4,309490813	5,915130843
Trps1	Increased	1,780360413	3,610377242	5,390737655
Clock	Increased	1,834276962	3,369059419	5,203336381
Jund	Increased	2,335227203	2,64012038	4,975347583
Mapk8ip3	Increased	1,759450912	2,734995537	4,494446449
Mafg	Increased	2,140037155	2,33699006	4,477027215
Arntl	Increased	1,662313461	2,336550357	3,998863818
Satb2	Decreased	-2,924352264	4,812783923	7,737136187
Lrrfip2	Decreased	-2,295571136	3,395830632	5,691401768

Table A.41: Mouse brain mC/mC vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Foxp1	Increased	2,326266861	5,150340265	7,476607126
Satb1	Increased	2,50370903	4,955564786	7,459273816
Foxk2	Increased	2,1742836	4,294179761	6,468463361
Foxk1	Increased	1,978218079	3,682655617	5,660873696
Creb1	Increased	1,904990005	3,503303314	5,408293319
Jund	Increased	1,97420311	1,951156477	3,925359587
Fn3k	Decreased	-1,568958664	3,139423277	4,708381941

Table A.42: Mouse brain hmC/mC vs hmC/hmC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Sfpq	Decreased	-2,409205627	3,797770884	6,206976511
Uhrf2	Decreased	-1,57970314	2,680012776	4,259715916

Table A.43: Mouse brain hmC/C vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
Creb1	Increased	2,002370453	3,522820567	5,52519102
Trps1	Increased	1,737911987	2,820275682	4,558187669
Jup	Increased	1,546299744	2,023864914	3,570164658
Satb2	Decreased	-2,839150238	4,268090459	7,107240697
Rbbp7	Decreased	-2,005496979	4,648634337	6,654131316

Table A.44: Venn diagram readers for C/C, hmC/C, mC/mC, hmC/mC and hmC/hmC from mouse brain. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, hmC/C, mC/mC and hmC/hmC modifications associated with the Venn diagram in Figure 3.27.

Enriched for	Number of proteins	Protein names
C/C>hmC/mC, hmC/C>hmC/mC, mC/mC>hmC/mC	1	Creb1
C/C>hmC/mC, mC/mC>hmC/mC	1	Jund
C/C>hmC/mC, hmC/C>hmC/mC	1	Trps1
hmC/mC>C/C, hmC/mC>hmC/C	1	Satb2
C/C>hmC/mC	6	Arntl, Zbtb2, Parp1, Mapk8ip3, Mafg, Clock
hmC/mC>C/C	1	Lrrfip2
mC/mC>hmC/mC	4	Satb1, Foxp1, Foxk1, Foxk2
hmC/mC>mC/mC	1	Fn3k
hmC/hmC>hmC/mC	2	Sfpq, Uhrf2

Enriched for	Number of proteins	Protein names
hmC/mC>hmC/C	1	Rbbp7
hmC/C>hmC/mC	1	Jup

Table A.45: Venn diagram readers from mouse brain experiments. Significantly enriched proteins (p-value < 0.025, FC > 1.5) for C/C, hmC/C, mC/mC, hmC/mC and hmC/hmC modifications associated with the Venn diagram in Figure 3.28.

CpG dyad	Number of proteins	Protein names
C/C, hmC/C, mC/mC	1	Creb1
hmC/hmC, hmC/mC, mC/mC	1	Satb2
C/C, hmC/C	3	Mid2, Hnrnpk, Trps1
C/C, mC/mC	1	Jund
hmC/C, mC/mC	1	Rplp1
hmC/mC, mC/mC	1	Rbbp7
hmC/hmC, hmC/mC	1	Lrrfip2
C/C	11	Mtmr10, Arntl, Lurap1, Zbtb2, Baz2b, Clock, Mafg, Gm10094, Parp1, Mapk8ip3, Ring1
hmC/C	3	Jup, Ndufs1, Haus3
mC/mC	11	Atp6v1d, Dynlt3, Mta2, Foxp1, Bend6, Rasgrp2, Mbd2, Gatad2b, Satb1, Foxk1, Foxk2
hmC/mC	1	Fn3k
hmC/hmC	2	Sfpq, Uhrf2

Table A.46: Experiment 3 C/C vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
KDM2B	Increased	10,89995995	8,798136399	19,69809635
BCOR	Increased	10,15863147	3,865142004	14,02377347
RBM45	Increased	4,825592232	8,839918687	13,66551092
CTCF	Increased	9,363670111	4,198311507	13,56198162
SAMD1	Increased	4,623544312	8,627925174	13,25146949
PCGF1	Increased	5,076825333	7,934207614	13,01103295
ZBTB2	Increased	6,585848999	6,289071186	12,87492019
KMT2A	Increased	3,015033436	9,619508001	12,63454144
ZMYND11	Increased	5,385412788	7,102155807	12,4875686
HCFC2	Increased	5,711881542	6,76266582	12,47454736
KMT2B	Increased	2,336208534	9,712206841	12,04841538
L3MBTL3	Increased	4,956261063	6,920930233	11,8771913
DNMT1	Increased	4,361935234	7,266879139	11,62881437
RBBP5	Increased	2,345146179	9,223830227	11,56897641
BEND3	Increased	8,387664986	2,876047989	11,26371298
SUZ12	Increased	3,662170219	7,546379138	11,20854936
MTF2	Increased	4,989504528	6,077868636	11,06737316
FLYWCH1	Increased	3,261595058	7,704876367	10,96647143
CCAR2	Increased	2,212961006	8,456328156	10,66928916
CREB1	Increased	2,388147354	8,226227838	10,61437519
CXXC5	Increased	3,665742111	6,83432303	10,50006514
BANP	Increased	3,599063492	6,793223079	10,39228657
ASH2L	Increased	2,393615437	7,768727659	10,1623431
NFIX	Increased	2,73001194	7,372367045	10,10237899
EED	Increased	3,031716156	6,767797381	9,799513537
AHR	Increased	2,412821484	7,25552476	9,668346244
RUNX2	Increased	2,263044834	7,379602198	9,642647032
CENPB	Increased	6,003288364	3,538088587	9,541376951
DPY30	Increased	2,171505928	6,840723951	9,012229879
TFAM	Increased	1,694373131	7,183358375	8,877731506
ZBTB10	Increased	1,842144966	6,83526149	8,677406456
TFAP4	Increased	2,304549885	6,369771267	8,674321152
ARNT	Increased	4,214900494	4,457502169	8,672402663

Name	Change	Fold change (log2)	Significance	Manhattan distance
E2F2	Increased	2,509604645	5,646513191	8,156117836
SKP1	Increased	2,036626816	6,076430097	8,113056913
ESRRA	Increased	1,887659645	5,500204695	7,38786434
ZIC2	Increased	3,21755085	4,137566456	7,355117306
BCL11A	Increased	3,271264076	4,065447641	7,336711717
ZIC1	Increased	2,793578911	4,505368952	7,298947863
KDM2A	Increased	2,358344173	4,848944008	7,207288181
ZBTB25	Increased	1,926455879	5,278653199	7,205109078
MXD4	Increased	2,184588718	4,988641463	7,173230181
DYNLL1	Increased	1,508625126	5,60888081	7,117505936
XBP1	Increased	1,793214417	5,08978218	6,882996597
GLI1	Increased	2,858056545	3,685840877	6,543897422
RUNX1	Increased	1,521837807	4,726339596	6,248177403
ATF3	Increased	2,476882553	3,697773062	6,174655615
NKX2-5	Increased	1,6275177	4,491290543	6,118808243
MAZ	Increased	2,405448914	3,551670622	5,957119536
NFIA	Increased	2,154151535	3,798906024	5,953057559
EZH2	Increased	2,309499836	3,632619614	5,94211945
VEZF1	Increased	1,631739521	4,098659444	5,730398965
KLF12	Increased	2,055392456	3,320433719	5,375826175
CEBPB	Increased	2,206313038	2,990031581	5,196344619
ESRRG	Increased	1,590551186	3,49139297	5,081944156
NFIC	Increased	2,429332352	2,567208088	4,99654044
TFAP2C	Increased	2,002018356	2,87005026	4,872068616
HAND1	Increased	1,907557964	2,80011914	4,707677104
ZNF148	Increased	1,680484104	2,673836468	4,354320572
ZNF740	Increased	1,508843899	1,829741543	3,338585442
UHRF1	Decreased	-5,623083687	7,581299906	13,20438359
RBM14	Decreased	-2,846889114	9,311401267	12,15829038
RFXANK	Decreased	-3,319162178	8,648058467	11,96722065
NCOA5	Decreased	-2,340011883	9,161545448	11,50155733
CHAF1B	Decreased	-1,541273308	9,689384667	11,23065798
SAFB	Decreased	-1,759989452	8,662049306	10,42203876
ZNF638	Decreased	-1,759866142	8,040550499	9,800416641
MECP2	Decreased	-1,822759342	7,954991388	9,77775073
YLPM1	Decreased	-3,950691605	5,622849814	9,573541419
RBM12B	Decreased	-2,865865993	6,574569384	9,440435377
RFX5	Decreased	-2,120335102	6,723009842	8,843344944
RBM4	Decreased	-2,088162899	6,635461809	8,723624708
SFPQ	Decreased	-2,702564907	5,93855366	8,641118567
SAFB2	Decreased	-1,680358696	6,905694409	8,586053105
MBD4	Decreased	-1,777537823	6,681695064	8,459232887
FOXC1	Decreased	-2,940714645	5,470296823	8,411011468
KHDRBS2	Decreased	-1,858752632	6,169534683	8,028287315
NONO	Decreased	-2,024374294	5,954586908	7,978961202
RFXAP	Decreased	-2,081328297	5,603840064	7,685168361
FOXF1	Decreased	-2,710388088	4,414893323	7,125281411
ZFHX3	Decreased	-1,855000973	5,146593701	7,001594674
FBXO11	Decreased	-2,202085209	4,267801052	6,469886261
RFX7	Decreased	-2,556787682	3,357777412	5,914565094
PAX3	Decreased	-1,643804836	3,192725775	4,836530611

Table A.47: Experiment 3 mC/mC vs hmC/mC.

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFXANK	Increased	2,410876274	7,96307314	10,37394941
ATF2	Increased	2,188713074	8,158004099	10,34671717
RFX1	Increased	4,245844269	5,949494621	10,19533889
MBD2	Increased	2,43339653	6,538156973	8,971553503
SUB1	Increased	1,797736168	7,130254162	8,92799033
RFXAP	Increased	2,006573009	6,883234681	8,88980769

Name	Change	Fold change (log2)	Significance	Manhattan distance
RFX5	Increased	1,914315319	6,353886213	8,268201532
GATAD2A	Increased	1,672259712	5,295200698	6,96746041
PRPF40A	Increased	1,552702713	4,707735511	6,260438224
ZBTB44	Increased	2,338720226	3,827212452	6,165932678
MTA1	Increased	1,62947998	4,43290892	6,0623889
RFX7	Increased	2,001213264	3,884570184	5,885783448
RFX2	Increased	1,917981815	3,609589161	5,527570976
KDM6A	Increased	1,702680874	2,512640519	4,215321393
MYC	Decreased	-2,163819504	7,635657688	9,799477192
RBM14	Decreased	-1,792959595	7,357598979	9,150558574
C14orf93	Decreased	-2,41564188	6,498505469	8,914147349
NONO	Decreased	-2,151176357	6,076060504	8,227236861
TBP	Decreased	-2,722780323	5,057762018	7,780542341
SFPQ	Decreased	-2,811793423	4,855157394	7,666950817
FBXO11	Decreased	-1,514787769	5,844775164	7,359562933
MAX	Decreased	-1,555712128	5,460997078	7,016709206
YLPM1	Decreased	-1,802103806	5,089702935	6,891806741
RBM4	Decreased	-1,649297619	4,583026568	6,232324187
ZIC2	Decreased	-2,285247231	2,772611862	5,057859093
PAX3	Decreased	-1,775857449	2,830953133	4,606810582
BEND7	Decreased	-1,8712183	2,724746582	4,595964882
RAI14	Decreased	-1,609893131	1,74191323	3,351806361

A.3 Source Code

R scripts

Venn diagrams for reader protein comparison

The following R script was used for visualization of reader protein overlaps. The circle sizes are proportional to the number of proteins in each group.

Listing A.1: General example of R script used for Venn diagram generation with the eulerr package

```
# Install and load the required package
install.packages("eulerr")
library(eulerr)

############### Example dataset ###############

# Replace these with your own gene/protein sets
## here shown with shortened placeholder lists
Group1 <- c("GeneA", "GeneB", "GeneC")
Group2 <- c("GeneB", "GeneD", "GeneE")
Group3 <- c("GeneC", "GeneE", "GeneF")
Group4 <- c("GeneA", "GeneF")
Group5 <- c("GeneD", "GeneG")

# Create a list with all groups
venn_data <- list(
  Group1 = Group1,
  Group2 = Group2,
  Group3 = Group3,
  Group4 = Group4,
  Group5 = Group5
)

############### Analysis and plotting ###############

# Fit Euler diagram
fit <- euler(venn_data)

# Save to PDF
pdf("venn_example.pdf")
plot(fit,
     fills = c("#333333","#E01B1D","#1A7BB6","#B2B2B2","#FFFFFF"),
     quantities = TRUE)
dev.off()
```

Binding curve fitting

The following R script was used for data preparation, model fitting, and visualization of binding curves.

Listing A.2: Example R script for fitting protein-DNA binding curves

```r
# Load required libraries
if (!requireNamespace("ggplot2", quietly = TRUE)) install.packages("ggplot2")
library(ggplot2)

# Step1: Prepare replicate data (mean + SEM)
## Input: protein concentrations, replicate 1, replicate 2
## Output: mean bound fraction + standard error
prepare_data <- function(protein_conc, rep1, rep2) {
  mean_bound <- rowMeans(data.frame(rep1, rep2), na.rm = TRUE)
  se_bound   <- apply(data.frame(rep1, rep2), 1, function(x) sd(x) / sqrt(
      length(x)))
  list(mean = mean_bound, se = se_bound)
}

# Step2: Define binding model (1-site binding)
binding_model <- function(protein_conc, Kd) {
  protein_conc / (Kd + protein_conc)
}

# Step 3: Fit model and extract Kd
## Use non-linear least squares (nls)
## Extract Kd and error estimates
fit_binding <- function(protein_conc, mean_bound) {
  fit <- nls(mean_bound ~ binding_model(protein_conc, Kd), start = list(Kd = 1)
      )
  coef(fit)
}

# Step 4: Input example data (shortened for illustration)
protein_conc <- c(0.5, 1, 2, 4, 8, 16)
rep1 <- c(0.2, 0.25, 0.3, 0.45, 0.6, 0.9)
rep2 <- c(0.18, 0.22, 0.28, 0.42, 0.62, 0.88)

# Step 5: Process
data <- prepare_data(protein_conc, rep1, rep2)
fit_results <- fit_binding(protein_conc, data$mean)

# Step 6: Generate fitted curve for plotting
## Create smooth data points across concentration range
fit_data <- data.frame(protein_conc = seq(min(protein_conc), max(protein_conc),
     length.out = 100))
fit_data$fit <- binding_model(fit_data$protein_conc, fit_results["Kd"])

# Step 7: Plot results
## Scatter plot of mean values
## Error bars (SEM)
```

```
## Fitted curve overlay
ggplot() +
  geom_point(aes(x = protein_conc, y = data$mean), size = 3) +
  geom_errorbar(aes(x = protein_conc,
                    ymin = data$mean - data$se,
                    ymax = data$mean + data$se),
                width = 0.1) +
  geom_line(aes(x = fit_data$protein_conc, y = fit_data$fit), color = "red") +
  labs(x = "Protein concentration", y = "Bound fraction",
       title = "Example binding curve")
```

A.4 License and Credits

License Figure 1.5. Figure 1.5a was adapted and modified from S. Li et al., 2022, licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). A copy of this license is available at: https://creativecommons.org/licenses/by/4.0/

Credits. Figures were prepared using Adobe Illustrator (Adobe Inc.) and the thesis was written and compiled in Overleaf (Digital Science). Parts of this thesis were refined with the assistance of AI language models, specifically ChatGPT (OpenAI), which was used for rewriting certain sections to improve clarity and flow. The abstract was translated using DeepL. These tools were employed solely to enhance language accuracy.

All other contents and figures are original or properly cited.

B References

Abubaker, J., Tiss, A., Abu-Farha, M., Al-Ghimlas, F., Al-Khairi, I., Baturcam, E., Cherian, P., Elkum, N., Hammad, M., John, J., Kavalakatt, S., Khadir, A., Warsame, S., Dermime, S., Behbehani, K., & Dehbi, M. (2013). Dnajb3/hsp-40 cochaperone is downregulated in obese humans and is restored by physical exercise. *PLoS ONE*, *8*, e69217. https://doi.org/10.1371/journal.pone.0069217

Adam, S., Anteneh, H., Hornisch, M., Wagner, V., Lu, J., Radde, N. E., Bashtrykov, P., Song, J., & Jeltsch, A. (2020). Dna sequence-dependent activity and base flipping mechanisms of dnmt1 regulate genome-wide dna methylation. *Nature Communications*, *11*, 3723. https://doi.org/10.1038/s41467-020-17531-8

Adams, C. C., & Workman, J. L. (1995). Binding of disparate transcriptional activators to nucleosomal dna is inherently cooperative. *Molecular and Cellular Biology*, *15*, 1405–1421. https://doi.org/10.1128/MCB.15.3.1405

Adolph, M. B., Mohamed, T. M., Balakrishnan, S., Xue, C., Morati, F., Modesti, M., Greene, E. C., Chazin, W. J., & Cortez, D. (2021). Radx controls rad51 filament dynamics to regulate replication fork stability. *Molecular Cell*, *81*, 1074–1083.e5. https://doi.org/10.1016/j.molcel.2020.12.036

Agarwal, P., Collier, P., Fritz, M. H.-Y., Benes, V., Wiklund, H. J., Westermark, B., & Singh, U. (2015). Cggbp1 mitigates cytosine methylation at repetitive dna sequences. *BMC Genomics*, *16*, 390. https://doi.org/10.1186/s12864-015-1593-2

Akerblom, I. E., Slater, E. P., Beato, M., Baxter, J. D., & Mellon, P. L. (1988). Negative regulation by glucocorticoids through interference with a camp responsive enhancer. *Science*, *241*, 350–353. https://doi.org/10.1126/science.2838908

Alberts, B., Heald, R., Johnson, A., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2022). *Molecular biology of the cell* (7th ed.). W. W. Norton & Company.

Alcamo, E. A., Chirivella, L., Dautzenberg, M., Dobreva, G., Fariñas, I., Grosschedl, R., & McConnell, S. K. (2008). Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. *Neuron*, *57*, 364–377. https://doi.org/10.1016/j.neuron.2007.12.012

Alvarez, J. D., Yasui, D. H., Niida, H., Joh, T., Loh, D. Y., & Kohwi-Shigematsu, T. (2000). The mar-binding protein satb1 orchestrates temporal and spatial expression of multiple genes during t-cell development. *Genes & Development*, *14*, 521–535. https://doi.org/10.1101/gad.14.5.521

Amati, B., & Land, H. (1994). Myc—max—mad: A transcription factor network controlling cell cycle progression, differentiation and death. *Current Opinion in Genetics & Development*, *4*, 102–108. https://doi.org/10.1016/0959-437X(94)90098-1

Amir, R. E., den Veyver, I. B. V., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in x-linked mecp2, encoding methyl-cpg-binding protein 2. *Nature Genetics*, *23*, 185–188. https://doi.org/10.1038/13810

Apuzzo, S., & Gros, P. (2007). Cooperative interactions between the two dna binding domains of pax3: Helix 2 of the paired domain is in the proximity of the amino terminus of the homeodomain. *Biochemistry*, *46*, 2984–2993. https://doi.org/10.1021/bi062107q

Arand, J., Spieler, D., Karius, T., Branco, M. R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., & Walter, J. (2012). In vivo control of cpg and non-cpg dna methylation by dna methyltransferases. *PLoS Genetics*, *8*, e1002750. https://doi.org/10.1371/journal.pgen.1002750

Aravind, L., Anand, S., & Iyer, L. M. (2013). Novel autoproteolytic and dna-damage sensing components in the bacterial sos response and oxidized methylcytosine-induced eukaryotic dna demethylation systems. *Biology Direct*, *8*, 20. https://doi.org/10.1186/1745-6150-8-20

Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M. D., & Wagner, G. P. (2016). The origin and evolution of cell types. *Nature Reviews Genetics*, *17*, 744–757. https://doi.org/10.1038/nrg.2016.127

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., & Shirakawa, M. (2008). Recognition of hemi-methylated dna by the sra protein uhrf1 by a base-flipping mechanism. *Nature*, *455*, 818–821. https://doi.org/10.1038/nature07249

Artero-Castro, A., Kondoh, H., Fernández-Marcos, P., Serrano, M., y Cajal, S. R., & LLeonart, M. (2009). Rplp1 bypasses replicative senescence and contributes to transformation. *Experimental Cell Research*, *315*, 1372–1383. https://doi.org/10.1016/j.yexcr.2009.02.007

Asl, H. F., Talukdar, H. A., Kindt, A. S., Jain, R. K., Ermel, R., Ruusalepp, A., Nguyen, K.-D. H., Dobrin, R., Reilly, D. F., Schunkert, H., Samani, N. J., Braenne, I., Erdmann, J., Melander, O., Qi, J., Ivert, T., Skogsberg, J., Schadt, E. E., Michoel, T., & Björkegren, J. L. (2015). Expression quantitative trait loci acting across multiple tissues are enriched in inherited risk for coronary artery disease. *Circulation: Cardiovascular Genetics*, *8*, 305–315. https://doi.org/10.1161/CIRCGENETICS.114.000640

Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C. H., & Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated dna by the sra domain of human uhrf1. *Nature*, *455*, 822–5. https://doi.org/10.1038/nature07273

Awad, S., & Hassan, A. H. (2008). The swi2/snf2 bromodomain is important for the full binding and remodeling activity of the swi/snf complex on h3- and h4-acetylated nucleosomes. *Annals of the New York Academy of Sciences*, *1138*, 366–375. https://doi.org/10.1196/annals.1414.038

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H. E., Iurlaro, M., Reik, W., Murrell, A., & Balasubramanian, S. (2015). 5-formylcytosine can be a stable dna modification in mammals. *Nature Chemical Biology*, *11*, 555–557. https://doi.org/10.1038/nchembio.1848

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., & Balasubramanian, S. (2014). 5-hydroxymethylcytosine is a predominantly stable dna modification. *Nature Chemistry*, *6*, 1049–1055. https://doi.org/10.1038/nchem.2064

Bai, D., Yang, J., Xue, X., Gao, Y., Wang, Y., Cui, M., He, B., Zeng, H., Xiang, H., Guo, Z., Zhu, L., Gao, J., Zhu, C., Tang, F., & Yi, C. (2025). Single-cell 5-hydroxymethylcytosine landscapes of mouse early embryos at single-base resolution. *Cell Reports*, *44*, 115520. https://doi.org/10.1016/j.celrep.2025.115520

Bai, L., Yang, G., Qin, Z., Lyu, J., Wang, Y., Feng, J., Liu, M., Gong, T., Li, X., Li, Z., Li, J., Qin, J., Yang, W., & Ding, C. (2021). Proteome-wide profiling of readers for dna modification. *Advanced Science*, *8*, 2101426. https://doi.org/10.1002/advs.202101426

Bannai, H., Fukatsu, K., Mizutani, A., Natsume, T., Iemura, S.-i., Ikegami, T., Inoue, T., & Mikoshiba, K. (2004). An rna-interacting protein, syncrip (heterogeneous nuclear ribonuclear protein q1/-nsap1) is a component of mrna granule transported with inositol 1,4,5-trisphosphate receptor

type 1 mrna in neuronal dendrites. *Journal of Biological Chemistry*, *279*, 53427–53434. https://doi.org/10.1074/jbc.M409732200

Bartels, S. J. J., Spruijt, C. G., Brinkman, A. B., Jansen, P. W. T. C., Vermeulen, M., & Stunnenberg, H. G. (2011). A silac-based screen for methyl-cpg binding proteins identifies rbp-j as a dna methylation and sequence-specific binding protein. *PLoS ONE*, *6*, e25884. https://doi.org/10.1371/journal.pone.0025884

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by dna and histone methylation. *Cell*, *143*, 470–484. https://doi.org/10.1016/j.cell.2010.10.012

Battistini, F., Dans, P. D., Terrazas, M., Castellazzi, C. L., Portella, G., Labrador, M., Villegas, N., Brun-Heath, I., González, C., & Orozco, M. (2021). The impact of the hydroxymethylcytosine epigenetic signature on dna structure and function. *PLOS Computational Biology*, *17*, e1009547. https://doi.org/10.1371/journal.pcbi.1009547

Baubec, T., Ivánek, R., Lienert, F., & Schübeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the mbd protein family. *Cell*, *153*, 480–492. https://doi.org/10.1016/j.cell.2013.03.011

Baymaz, H. I., Fournier, A., Laget, S., Ji, Z., Jansen, P. W. T. C., Smits, A. H., Ferry, L., Mensinga, A., Poser, I., Sharrocks, A., Defossez, P.-A., & Vermeulen, M. (2014). Mbd5 and mbd6 interact with the human pr-dub complex through their methyl-cpg-binding domain. *PROTEOMICS*, *14*, 2179–2189. https://doi.org/10.1002/pmic.201400013

Beck, D. B., Petracovici, A., He, C., Moore, H. W., Louie, R. J., Ansar, M., Douzgou, S., Sithambaram, S., Cottrell, T., Santos-Cortez, R. L. P., Prijoles, E. J., Bend, R., Keren, B., Mignot, C., Nougues, M.-C., Õunap, K., Reimand, T., Pajusalu, S., Zahid, M., … Fahrner, J. A. (2020). Delineation of a human mendelian disorder of the dna demethylation machinery: Tet3 deficiency. *The American Journal of Human Genetics*, *106*, 234–245. https://doi.org/10.1016/j.ajhg.2019.12.007

Becker, P. B., & Workman, J. L. (2013). Nucleosome remodeling and epigenetics. *Cold Spring Harbor Perspectives in Biology*, *5*, a017905–a017905. https://doi.org/10.1101/cshperspect.a017905

Bellacosa, A., & Drohat, A. C. (2015). Role of base excision repair in maintaining the genetic and epigenetic integrity of cpg sites. *DNA Repair*, *32*, 33–42. https://doi.org/10.1016/j.dnarep.2015.04.011

Bender, A., Morel, M., Dumas, M., Klopfenstein, M., Osmani, N., Greenberg, M. V. C., Bourc'his, D., Ghyselinck, N. B., & Weber, M. (2025). Uhrf2 mediates resistance to dna methylation reprogramming in primordial germ cells. *Nature Communications*, *16*, 7350. https://doi.org/10.1038/s41467-025-61954-0

Bender, J. (1998). Cytosine methylation of repeated sequences in eukaryotes: The role of dna pairing. *Trends in Biochemical Sciences*, *23*, 252–256. https://doi.org/10.1016/S0968-0004(98)01225-0

Berger, M. F., Philippakis, A. A., Qureshi, A. M., He, F. S., Estep, P. W., & Bulyk, M. L. (2006). Compact, universal dna microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnology*, *24*, 1429–1435. https://doi.org/10.1038/nbt1246

Berry, F. B., Miura, Y., Mihara, K., Kaspar, P., Sakata, N., Hashimoto-Tamaoki, T., & Tamaoki, T. (2001). Positive and negative regulation of myogenic differentiation of c2c12 cells by isoforms of the multiple homeodomain zinc finger transcription factor atbf1. *Journal of Biological Chemistry*, *276*, 25057–25065. https://doi.org/10.1074/jbc.M010378200

Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cdna encoding dna methyltransferase of mouse cells. *Journal of Molecular Biology*, *203*, 971–983. https://doi.org/10.1016/0022-2836(88)90122-2

Bestor, T. H. (1990). Dna methylation: Evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *326*, 179–187. https://doi.org/10.1098/rstb.1990.0002

Bhoumik, A., Takahashi, S., Breitweiser, W., Shiloh, Y., Jones, N., & Ronai, Z. (2005). Atm-dependent phosphorylation of atf2 is required for the dna damage response. *Molecular Cell*, *18*, 577–587. https://doi.org/10.1016/j.molcel.2005.04.015

Bird, A. P. (1980). Dna methylation and the frequency of cpg in animal dna. *Nucleic Acids Research*, *8*, 1499–1504. https://doi.org/10.1093/nar/8.7.1499

Bird, A. P. (1987). Cpg islands as gene markers in the vertebrate nucleus. *Trends in Genetics*, *3*, 342–347. https://doi.org/10.1016/0168-9525(87)90294-0

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J., & Klose, R. J. (2010). Cpg islands recruit a histone h3 lysine 36 demethylase. *Molecular Cell*, *38*, 179–190. https://doi.org/10.1016/j.molcel.2010.04.009

Blackwood, E. M., Lüscher, B., & Eisenman, R. N. (1992). Myc and max associate in vivo. *Genes & Development*, *6*, 71–80. https://doi.org/10.1101/gad.6.1.71

Blackwood, E. M., & Eisenman, R. N. (1991). Max: A helix-loop-helix zipper protein that forms a sequence-specific dna-binding complex with myc. *Science*, *251*, 1211–1217. https://doi.org/10.1126/science.2006410

Bond, C. S., & Fox, A. H. (2009). Paraspeckles: Nuclear bodies built on long noncoding rna. *Journal of Cell Biology*, *186*, 637–644. https://doi.org/10.1083/jcb.200906113

Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W., & Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, *336*, 934–937. https://doi.org/10.1126/science.1220671

Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). Uhrf1 plays a role in maintaining dna methylation in mammalian cells. *Science*, *317*, 1760–1764. https://doi.org/10.1126/science.1147939

Boyer, L. A., Latek, R. R., & Peterson, C. L. (2004). The sant domain: A unique histone-tail-binding module? *Nature Reviews Molecular Cell Biology*, *5*, 158–163. https://doi.org/10.1038/nrm1314

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., & Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, *441*, 349–353. https://doi.org/10.1038/nature04733

Bozhenok, L. (2002). Wstf-iswi chromatin remodeling complex targets heterochromatic replication foci. *The EMBO Journal*, *21*, 2231–2241. https://doi.org/10.1093/emboj/21.9.2231

Bracken, A. P., Brien, G. L., & Verrijzer, C. P. (2019). Dangerous liaisons: Interplay between swi/snf, nurd, and polycomb in chromatin regulation and cancer. *Genes & Development*, *33*, 936–959. https://doi.org/10.1101/gad.326066.119

Brasen, C., Dorosz, J., Wiuf, A., Boesen, T., Mirza, O., & Gajhede, M. (2017). Expression, purification and characterization of the human mta2-rbbp7 complex. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1865*, 531–538. https://doi.org/10.1016/j.bbapap.2017.02.002

Brayer, K. J., & Segal, D. J. (2008). Keep your fingers off my dna: Protein–protein interactions mediated by c2h2 zinc finger domains. *Cell Biochemistry and Biophysics*, *50*, 111–131. https://doi.org/10.1007/s12013-008-9008-5

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Vire, E., Smet, C. D., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., Pelicci, P. G., Amati, B., Kouzarides, T., de Launoit, Y., Croce, L. D., & Fuks, F. (2005). Myc represses transcription through recruitment of dna methyltransferase corepressor. *The EMBO Journal*, *24*, 336–346. https://doi.org/10.1038/sj.emboj.7600509

Brinkman, A. B., Gu, H., Bartels, S. J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., & Stunnenberg, H. G. (2012). Sequential chip-bisulfite sequencing enables direct genome-scale investigation of chromatin and dna methylation cross-talk. *Genome Research*, *22*, 1128–1138. https://doi.org/10.1101/gr.133728.111

Bronner, C., Achour, M., Arima, Y., Chataigneau, T., Saya, H., & Schini-Kerth, V. B. (2007). The uhrf family: Oncogenes that are drugable targets for cancer therapy in the near future? *Pharmacology & Therapeutics*, *115*, 419–434. https://doi.org/10.1016/j.pharmthera.2007.06.003

Buchmuller, B. C., Dröden, J., Singh, H., Palei, S., Drescher, M., Linser, R., & Summerer, D. (2022). Evolved dna duplex readers for strand-asymmetrically modified 5-hydroxymethylcytosine/5-methylcytosine cpg dyads. *Journal of the American Chemical Society*, *144*, 2987–2993. https://doi.org/10.1021/jacs.1c10678

Buchmuller, B. C., Kosel, B., & Summerer, D. (2020). Complete profiling of methyl-cpg-binding domains for combinations of cytosine modifications at cpg dinucleotides reveals differential read-out in normal and rett-associated states. *Scientific Reports*, *10*, 4053. https://doi.org/10.1038/s41598-020-61030-1

Buck-Koehntop, B. A., & Defossez, P.-A. (2013). On how mammalian transcription factors recognize methylated dna. *Epigenetics*, *8*, 131–137. https://doi.org/10.4161/epi.23632

Bürglin, T. R. (2011). Homeodomain subtypes and functional diversity. https://doi.org/10.1007/978-90-481-9069-0_5

Cai, Q., Lin, T., Kamarajugadda, S., & Lu, J. (2013). Regulation of glycolysis and the warburg effect by estrogen-related receptors. *Oncogene*, *32*, 2079–2086. https://doi.org/10.1038/onc.2012.221

Caiafa, P., Guastafierro, T., & Zampieri, M. (2009). Epigenetics: Poly(adp-ribosyl)ation of parp-1 regulates genomic methylation patterns. *The FASEB Journal*, *23*, 672–678. https://doi.org/10.1096/fj.08-123265

Campillos, M. (2003). Transcriptional activation by ap-2alpha is modulated by the oncogene dek. *Nucleic Acids Research*, *31*, 1571–1575. https://doi.org/10.1093/nar/gkg247

Cardozo, T., & Pagano, M. (2004). The scf ubiquitin ligase: Insights into a molecular machine. *Nature Reviews Molecular Cell Biology*, *5*, 739–751. https://doi.org/10.1038/nrm1471

Caretti, G., Cocchiarella, F., Sidoli, C., Villard, J., Peretti, M., Reith, W., & Mantovani, R. (2000). Dissection of functional nf-y-rfx cooperative interactions on the mhc class ii ea promoter. *Journal of Molecular Biology*, *302*, 539–552. https://doi.org/10.1006/jmbi.2000.4028

Carlson, C. D., Warren, C. L., Hauschild, K. E., Ozers, M. S., Qadir, N., Bhimsaria, D., Lee, Y., Cerrina, F., & Ansari, A. Z. (2010). Specificity landscapes of dna binding molecules elucidate biological function. *Proceedings of the National Academy of Sciences*, *107*, 4544–4549. https://doi.org/10.1073/pnas.0914023107

Carlsson, P., & Mahlapuu, M. (2002). Forkhead transcription factors: Key players in development and metabolism. *Developmental Biology*, *250*, 1–23. https://doi.org/10.1006/dbio.2002.0780

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell*, *134*, 25–36. https://doi.org/10.1016/j.cell.2008.06.030

Catania, S., Dumesic, P. A., Pimentel, H., Nasif, A., Stoddard, C. I., Burke, J. E., Diedrich, J. K., Cooke, S., Shea, T., Gienger, E., Lintner, R., Yates, J. R., Hajkova, P., Narlikar, G. J., Cuomo, C. A., Pritchard, J. K., & Madhani, H. D. (2020). Evolutionary persistence of dna methylation for millions of years after ancient loss of a de novo methyltransferase. *Cell*, *180*, 816. https://doi.org/10.1016/j.cell.2020.02.006

Chakraborty, M., Bhattacharya, D., Mukhopadhyay, C., & Chakrabarti, A. (2010). Structure and conformational studies on dityrosine formation in the dna binding domain of rfx5. *Biophysical Chemistry*, *149*, 92–101. https://doi.org/10.1016/j.bpc.2010.04.005

Chakraborty, M., Sengupta, A., Bhattacharya, D., Banerjee, S., & Chakrabarti, A. (2010). Dna binding domain of rfx5: Interactions with x-box dna and rfxank. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1804*, 2016–2024. https://doi.org/10.1016/j.bbapap.2010.07.009

Chanarat, S., & Sträßer, K. (2013). Splicing and beyond: The many faces of the prp19 complex. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1833*, 2126–2134. https://doi.org/10.1016/j.bbamcr.2013.05.023

Charlton, J., Downing, T. L., Smith, Z. D., Gu, H., Clement, K., Pop, R., Akopian, V., Klages, S., Santos, D. P., Tsankov, A. M., Timmermann, B., Ziller, M. J., Kiskinis, E., Gnirke, A., & Meissner, A. (2018). Global delay in nascent strand dna methylation. *Nature Structural & Molecular Biology*, *25*, 327–332. https://doi.org/10.1038/s41594-018-0046-4

Cheblal, A., Challa, K., Seeber, A., Shimada, K., Yoshida, H., Ferreira, H. C., Amitai, A., & Gasser, S. M. (2020). Dna damage-induced nucleosome depletion enhances homology search independently of local break movement. *Molecular Cell*, *80*, 311–326.e4. https://doi.org/10.1016/j.molcel.2020.09.002

Chen, L., Chen, D., Manome, Y., Dong, Y., Fine, H. A., & Kufe, D. W. (1995). Breast cancer selective gene expression and therapy mediated by recombinant adenoviruses containing the df3/muc1 promoter. *Journal of Clinical Investigation*, *96*, 2775–2782. https://doi.org/10.1172/JCI118347

Chen, L., Wang, X., Xie, N., Zhang, Z., Xu, X., Xue, M., Yang, Y., Liu, L., Su, L., Bjaanæs, M., Karlsson, A., Planck, M., Staaf, J., Helland, Å., Esteller, M., Christiani, D. C., Chen, F., & Zhang, R. (2025). A two-phase epigenome-wide four-way gene–smoking interaction study of overall survival for early-stage non-small cell lung cancer. *Molecular Oncology*, *19*, 173–187. https://doi.org/10.1002/1878-0261.13766

Chen, X., Yang, Z., Wang, W., Qian, K., Liu, M., Wang, J., & Wang, M. (2021). Structural basis for rna recognition by the n-terminal tandem rrm domains of human rbm45. *Nucleic Acids Research*, *49*, 2946–2958. https://doi.org/10.1093/nar/gkab075

Cheng, X., Kumar, S., Klimasauskas, S., & Roberts, R. (1993). Crystal structure of the hhai dna methyltransferase. *Cold Spring Harbor Symposia on Quantitative Biology*, *58*, 331–338. https://doi.org/10.1101/SQB.1993.058.01.039

Cheng, X.-N., Shao, M., Li, J.-T., Wang, Y.-F., Qi, J., Xu, Z.-G., & Shi, D.-L. (2017). Leucine repeat adaptor protein 1 interacts with dishevelled to regulate gastrulation cell movements in zebrafish. *Nature Communications*, *8*, 1353. https://doi.org/10.1038/s41467-017-01552-x

Chhatbar, K., Connelly, J., Webb, S., Kriaucionis, S., & Bird, A. (2022). A critique of the hypothesis that ca repeats are primary targets of neuronal mecp2. *Life Science Alliance*, *5*, e202201522. https://doi.org/10.26508/lsa.202201522

Chialastri, A., Sarkar, S., Schauer, E. E., Lamba, S., & Dey, S. S. (2024). Combinatorial quantification of 5mc and 5hmc at individual cpg dyads and the transcriptome in single cells reveals modulators of dna methylation maintenance fidelity. *Nature Structural & Molecular Biology*, *31*, 1296–1308. https://doi.org/10.1038/s41594-024-01291-w

Chiu, T.-P., Rao, S., & Rohs, R. (2023). Physicochemical models of protein–dna binding with standard and modified base pairs. *Proceedings of the National Academy of Sciences*, *120*. https://doi.org/10.1073/pnas.2205796120

Choi, S. H., Flamand, M. N., Liu, B., Zhu, H., Hu, M., Wang, M., Sewell, J., Holley, C. L., Al-Hashimi, H. M., & Meyer, K. D. (2022). Rbm45 is an m6a-binding protein that affects neuronal differentiation and the splicing of a subset of mrnas. *Cell Reports*, *40*, 111293. https://doi.org/10.1016/j.celrep.2022.111293

Choksi, S. P., Lauter, G., Swoboda, P., & Roy, S. (2014). Switching on cilia: Transcriptional networks regulating ciliogenesis. *Development*, *141*, 1427–1441. https://doi.org/10.1242/dev.074666

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors hnf3 (foxa) and gata-4. *Molecular Cell*, *9*, 279–289. https://doi.org/10.1016/S1097-2765(02)00459-8

Clapier, C. R. (2002). A critical epitope for substrate recognition by the nucleosome remodeling atpase iswi. *Nucleic Acids Research*, *30*, 649–655. https://doi.org/10.1093/nar/30.3.649

Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of atp-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, *18*, 407–422. https://doi.org/10.1038/nrm.2017.26

Claude, A. (1946). Fractionation of mammalian liver cells by differential centrifugation: I. problems, methods, and preparation of extract. *Journal of Experimental Medicine*, *84*, 51–59. https://doi.org/10.1084/jem.84.1.51

Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliaha, P. V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D. N., Li, Z., Xu, M., Reik, W., Silva, J. C. R., & Wang, J. (2013). Nanog-dependent function of tet1 and tet2 in establishment of pluripotency. *Nature*, *495*, 370–374. https://doi.org/10.1038/nature11925

Cox, J., & Mann, M. (2008). Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*, 1367–1372. https://doi.org/10.1038/nbt.1511

Crawford, D. J., Liu, M. Y., Nabel, C. S., Cao, X.-J., Garcia, B. A., & Kohli, R. M. (2016). Tet2 catalyzes stepwise 5-methylcytosine oxidation by an iterative and de novo mechanism. *Journal of the American Chemical Society*, *138*, 730–733. https://doi.org/10.1021/jacs.5b10554

Dai, Q., Sanstead, P. J., Peng, C. S., Han, D., He, C., & Tokmakoff, A. (2016). Weakened n3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability. *ACS Chemical Biology*, *11*, 470–477. https://doi.org/10.1021/acschembio.5b00762

de Martin, X., Sodaei, R., & Santpere, G. (2021). Mechanisms of binding specificity among bhlh transcription factors. *International Journal of Molecular Sciences*, *22*, 9150. https://doi.org/10.3390/ijms22179150

de la Rica, L., Rodríguez-Ubreva, J., García, M., Islam, A. B., Urquiza, J. M., Hernando, H., Christensen, J., Helin, K., Gómez-Vaquero, C., & Ballestar, E. (2013). Pu.1 target genes undergo tet2-coupled demethylation and dnmt3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biology*, *14*, R99. https://doi.org/10.1186/gb-2013-14-9-r99

Dean, S. T., Ishikawa, C., Zhu, X., Walulik, S., Nixon, T., Jordan, J. K., Henderson, S., Wyder, M., Salomonis, N., Wunderlich, M., Greis, K. D., Starczynowski, D. T., & Volk, A. G. (2023). Repression of trim13 by chromatin assembly factor chaf1b is critical for aml development. *Blood Advances*, *7*, 4822–4837. https://doi.org/10.1182/bloodadvances.2022009438

Déjardin, J., & Kingston, R. E. (2009). Purification of proteins associated with specific genomic loci. *Cell*, *136*, 175–186. https://doi.org/10.1016/j.cell.2008.11.045

Denissov, S., van Driel, M., Voit, R., Hekkelman, M., Hulsen, T., Hernandez, N., Grummt, I., Wehrens, R., & Stunnenberg, H. (2007). Identification of novel functional tbp-binding sites and general factor repertoires. *The EMBO Journal*, *26*, 944–954. https://doi.org/10.1038/sj.emboj.7601550

DeSandro, A. M., Nagarajan, U. M., & Boss, J. M. (2000). Associations and interactions between bare lymphocyte syndrome factors. *Molecular and Cellular Biology*, *20*, 6587–6599. https://doi.org/10.1128/MCB.20.17.6587-6599.2000

Dickinson, L. (1992). A tissue-specific mar/sar dna-binding protein with unusual binding site recognition. *Cell*, *70*, 631–645. https://doi.org/10.1016/0092-8674(92)90432-C

Dignam, J. D., Lebovitz, R. M., & Roeder, R. G. (1983). Accurate transcription initiation by rna polymerase ii in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research*, *11*, 1475–1489. https://doi.org/10.1093/nar/11.5.1475

Dostal, V., & Churchill, M. E. (2019). Cytosine methylation of mitochondrial dna at cpg sequences impacts transcription factor a dna binding and transcription. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1862*, 598–607. https://doi.org/10.1016/j.bbagrm.2019.01.006

Du, Q., Luu, P.-L., Stirzaker, C., & Clark, S. J. (2015). Methyl-cpg-binding domain proteins: Readers of the epigenome. *Epigenomics*, *7*, 1051–1073. https://doi.org/10.2217/epi.15.39

Dubois-Chevalier, J., Oger, F., Dehondt, H., Firmin, F. F., Gheeraert, C., Staels, B., Lefebvre, P., & Eeckhoute, J. (2014). A dynamic ctcf chromatin binding landscape promotes dna hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Research*, *42*, 10943–10959. https://doi.org/10.1093/nar/gku780

DuBridge, R. B., Tang, P., Hsia, H. C., Leong, P.-M., Miller, J. H., & Calos, M. P. (1987). Analysis of mutation in human cells by using an epstein-barr virus shuttle system. *Molecular and Cellular Biology*, *7*, 379–387. https://doi.org/10.1128/mcb.7.1.379-387.1987

Duina, A. A. (2011). Histone chaperones spt6 and fact: Similarities and differences in modes of action at transcribed genes. *Genetics Research International*, *2011*, 1–12. https://doi.org/10.4061/2011/625210

Dumesic, P. A., Stoddard, C. I., Catania, S., Narlikar, G. J., & Madhani, H. D. (2020). Atp hydrolysis by the snf2 domain of dnmt5 is coupled to both specific recognition and modification of hemimethylated dna. *Molecular Cell*, *79*, 127–139.e4. https://doi.org/10.1016/j.molcel.2020.04.029

Dvorak, H. F. (2002). Vascular permeability factor/vascular endothelial growth factor: A critical cytokine in tumor angiogenesis and a potential target for diagnosis and therapy. *Journal of Clinical Oncology*, *20*, 4368–4380. https://doi.org/10.1200/JCO.2002.10.088

Fang, M., Ou, J., Hutchinson, L., & Green, M. R. (2014). The braf oncoprotein functions through the transcriptional repressor mafg to mediate the cpg island methylator phenotype. *Molecular Cell*, *55*, 904–915. https://doi.org/10.1016/j.molcel.2014.08.010

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W., Koseki, H., Brockdorff, N., Ponting, C. P., Kessler, B. M., & Klose, R. J. (2012). Kdm2b links the polycomb repressive complex 1 (prc1) to recognition of cpg islands. *eLife*, *1*. https://doi.org/10.7554/eLife.00205

Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., & Burley, S. K. (1993). Recognition by max of its cognate dna through a dimeric b/hlh/z domain. *Nature*, *363*, 38–45. https://doi.org/10.1038/363038a0

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S., & Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse es cells and during differentiation. *Nature*, *473*, 398–402. https://doi.org/10.1038/nature10008

Fieber, W., Schneider, M. L., Matt, T., Kräutler, B., Konrat, R., & Bister, K. (2001). Structure, function, and dynamics of the dimerization and dna-binding domain of oncogenic transcription factor v-myc11edited by p. e. wright. *Journal of Molecular Biology*, *307*, 1395–1410. https://doi.org/10.1006/jmbi.2001.4537

Fischer, B. A., Chelbi, S. T., & Guarda, G. (2020). Regulatory factor x 7 and its potential link to lymphoid cancers. *Trends in Cancer*, *6*, 6–9. https://doi.org/10.1016/j.trecan.2019.11.001

Flaus, A. (2006). Identification of multiple distinct snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research*, *34*, 2887–2905. https://doi.org/10.1093/nar/gkl295

Florl, A. R., Löwer, R., Schmitz-Dräger, B. J., & Schulz, W. A. (1999). Dna methylation and expression of line-1 and herv-k provirus sequences in urothelial and renal cell carcinomas. *British Journal of Cancer*, *80*, 1312–1321. https://doi.org/10.1038/sj.bjc.6690524

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., & Turner, S. W. (2010). Direct detection of dna methylation during single-molecule, real-time sequencing. *Nature Methods*, *7*, 461–465. https://doi.org/10.1038/nmeth.1459

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., & Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the uhrf1 sra domain. *PLoS ONE*, *6*, e21306. https://doi.org/10.1371/journal.pone.0021306

Frietze, S., & Farnham, P. J. (2011). Transcription factor effector domains. https://doi.org/10.1007/978-90-481-9069-0_12

Fu, T., Liu, L., Yang, Q.-L., Wang, Y., Xu, P., Zhang, L., Liu, S., Dai, Q., Ji, Q., Xu, G.-L., He, C., Luo, C., & Zhang, L. (2019). Thymine dna glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chemical Science*, *10*, 7407–7417. https://doi.org/10.1039/C9SC02807B

Fujii, Y., & Nakamura, M. (2010). Foxk2 transcription factor is a novel g/t-mismatch dna binding protein. *Journal of Biochemistry*, *147*, 705–709. https://doi.org/10.1093/jb/mvq004

Fujiki, K., Shinoda, A., Kano, F., Sato, R., Shirahige, K., & Murata, M. (2013). Pparγ-induced parylation promotes local dna demethylation by production of 5-hydroxymethylcytosine. *Nature Communications*, *4*, 2262. https://doi.org/10.1038/ncomms3262

Fulton, D. L., Sundararajan, S., Badis, G., Hughes, T. R., Wasserman, W. W., Roach, J. C., & Sladek, R. (2009). Tfcat: The curated catalog of mouse and human transcription factors. *Genome Biology*, *10*, R29. https://doi.org/10.1186/gb-2009-10-3-r29

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E. O., Nedelec, S., Wichterle, H., Woltjen, K., Hughes, T. R., Zandstra, P. W., Nagy, A., Wrana, J. L., & Blencowe, B. J. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*, *147*, 132–146. https://doi.org/10.1016/j.cell.2011.08.023

Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B., & Burley, S. K. (2000). Structure of the winged-helix protein hrfx1 reveals a new mode of dna binding. *Nature*, *403*, 916–921. https://doi.org/10.1038/35002634

Gallina, I., Colding, C., Henriksen, P., Beli, P., Nakamura, K., Offman, J., Mathiasen, D. P., Silva, S., Hoffmann, E., Groth, A., Choudhary, C., & Lisby, M. (2015). Cmr1/wdr76 defines a nuclear genotoxic stress body linking genome integrity and protein quality control. *Nature Communications*, *6*, 6533. https://doi.org/10.1038/ncomms7533

Gama-Sosa, M. A., Midgett, R. M., Slagel, V. A., Githens, S., Kuo, K. C., Gehrke, C. W., & Ehrlich, M. (1983). Tissue-specific differences in dna methylation in various mammals. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *740*, 212–219. https://doi.org/10.1016/0167-4781(83)90079-9

Gao, L., Emperle, M., Guo, Y., Grimm, S. A., Ren, W., Adam, S., Uryu, H., Zhang, Z.-M., Chen, D., Yin, J., Dukatz, M., Anteneh, H., Jurkowska, R. Z., Lu, J., Wang, Y., Bashtrykov, P., Wade, P. A., Wang, G. G., Jeltsch, A., & Song, J. (2020). Comprehensive structure-function characterization of dnmt3b and dnmt3a reveals distinctive de novo dna methylation mechanisms. *Nature Communications*, *11*, 3355. https://doi.org/10.1038/s41467-020-17109-4

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., & Reinberg, D. (2012). Pcgf homologs, cbx proteins, and rybp define functionally distinct prc1 family complexes. *Molecular Cell*, *45*, 344–356. https://doi.org/10.1016/j.molcel.2012.01.002

Garvie, C. W., & Boss, J. M. (2008). Assembly of the rfx complex on the mhcii promoter: Role of rfxap and rfxb in relieving autoinhibition of rfx5. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1779*, 797–804. https://doi.org/10.1016/j.bbagrm.2008.07.012

Garvie, C. W., Stagno, J. R., Reid, S., Singh, A., Harrington, E., & Boss, J. M. (2007). Characterization of the rfx complex and the rfx5(l66a) mutant: Implications for the regulation of mhc class ii gene expression. *Biochemistry*, *46*, 1597–1611. https://doi.org/10.1021/bi6023868

Garvie, C. W., & Wolberger, C. (2001). Recognition of specific dna sequences. *Molecular Cell*, *8*, 937–946. https://doi.org/10.1016/S1097-2765(01)00392-6

Garza, A. S., Ahmad, N., & Kumar, R. (2009). Role of intrinsically disordered protein regions/domains in transcriptional regulation. *Life Sciences*, *84*, 189–193. https://doi.org/10.1016/j.lfs.2008.12.002

Gaubatz, S., Wood, J. G., & Livingston, D. M. (1998). Unusual proliferation arrest and transcriptional control properties of a newly discovered e2f family member, e2f-6. *Proceedings of the National Academy of Sciences*, *95*, 9190–9195. https://doi.org/10.1073/pnas.95.16.9190

Gauchier, M., Kan, S., Barral, A., Sauzet, S., Agirre, E., Bonnell, E., Saksouk, N., Barth, T. K., Ide, S., Urbach, S., Wellinger, R. J., Luco, R. F., Imhof, A., & Déjardin, J. (2019). Setdb1-dependent heterochromatin stimulates alternative lengthening of telomeres. *Science Advances*, *5*. https://doi.org/10.1126/sciadv.aav3673

Gertz, J., Reddy, T. E., Varley, K. E., Garabedian, M. J., & Myers, R. M. (2012). Genistein and bisphenol a exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. *Genome Research*, *22*, 2153–2162. https://doi.org/10.1101/gr.135681.111

Giehr, P., Kyriakopoulos, C., Lepikhov, K., Wallner, S., Wolf, V., & Walter, J. (2018). Two are better than one: Hpoxbs - hairpin oxidative bisulfite sequencing. *Nucleic Acids Research*, *46*, e88–e88. https://doi.org/10.1093/nar/gky422

Gill, G. (2005). Something about sumo inhibits transcription. *Current Opinion in Genetics & Development*, *15*, 536–541. https://doi.org/10.1016/j.gde.2005.07.004

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., & Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*, *5*, e15367. https://doi.org/10.1371/journal.pone.0015367

Goedhart, J., & Luijsterburg, M. S. (2020). Volcanoser is a web app for creating, exploring, labeling and sharing volcano plots. *Scientific Reports*, *10*, 20560. https://doi.org/10.1038/s41598-020-76603-3

Golla, J. P., Zhao, J., Mann, I. K., Sayeed, S. K., Mandal, A., Rose, R. B., & Vinson, C. (2014). Carboxylation of cytosine (5cac) in the cg dinucleotide in the e-box motif (cgcag|gtg) increases binding of the tcf3|ascl1 helix-loop-helix heterodimer 10-fold. *Biochemical and Biophysical Research Communications*, *449*, 248–255. https://doi.org/10.1016/j.bbrc.2014.05.018

Gong, J., Huang, M., Wang, F., Ma, X., Liu, H., Tu, Y., Xing, L., Zhu, X., Zheng, H., Fang, J., Li, X., Wang, Q., Wang, J., Sun, Z., Wang, X., Wang, Y., Guo, C., & Tang, T.-S. (2017). Rbm45 competes with hdac1 for binding to fus in response to dna damage. *Nucleic Acids Research*, *45*, 12862–12876. https://doi.org/10.1093/nar/gkx1102

Gorenjak, V., Vance, D. R., Dade, S., Stathopoulou, M. G., Doherty, L., Xie, T., Murray, H., Masson, C., Lamont, J., Fitzgerald, P., & Visvikis-Siest, S. (2020). Epigenome-wide association study in healthy individuals identifies significant associations with dna methylation and pbmc extract vegf-a concentration. *Clinical epigenetics*, *12*, 79. https://doi.org/10.1186/s13148-020-00874-w

Gowrishankar, S., Lyons, L., Rafiq, N. M., Roczniak-Ferguson, A., Camilli, P. D., & Ferguson, S. M. (2021). Overlapping roles of jip3 and jip4 in promoting axonal transport of lysosomes in human ipsc-derived neurons. *Molecular Biology of the Cell*, *32*, 1094–1103. https://doi.org/10.1091/mbc.E20-06-0382

Gräff, J., & Mansuy, I. M. (2009). Epigenetic dysregulation in cognitive disorders. *European Journal of Neuroscience*, *30*, 1–8. https://doi.org/10.1111/j.1460-9568.2009.06787.x

Grand, R. S., Burger, L., Gräwe, C., Michael, A. K., Isbel, L., Hess, D., Hoerner, L., Iesmantavicius, V., Durdu, S., Pregnolato, M., Krebs, A. R., Smallwood, S. A., Thomä, N., Vermeulen, M., & Schübeler, D. (2021). Banp opens chromatin and activates cpg-island-regulated genes. *Nature*, *596*, 133–137. https://doi.org/10.1038/s41586-021-03689-8

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of dna methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*, 590–607. https://doi.org/10.1038/s41580-019-0159-6

Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyal, F., Frenoy, O., Pousse, Y., Reichen, M., Woolfe, A., Brenan, C., Griffiths, A. D., Vallot, C., & Gérard, A. (2019). High-throughput single-cell chip-seq identifies heterogeneity of chromatin states in breast cancer. *Nature Genetics*, *51*, 1060–1066. https://doi.org/10.1038/s41588-019-0424-9

Gu, J., Chen, Z., Chen, X., & Wang, Z. (2020). Heterogeneous nuclear ribonucleoprotein (hnrnpl) in cancer. *Clinica Chimica Acta*, *507*, 286–294. https://doi.org/10.1016/j.cca.2020.04.040

Guillen-Ahlers, H., Rao, P. K., Levenstein, M. E., Kennedy-Darling, J., Perumalla, D. S., Jadhav, A. Y., Glenn, J. P., Ludwig-Kubinski, A., Drigalenko, E., Montoya, M. J., Göring, H. H., Anderson, C. D., Scalf, M., Gildersleeve, H. I., Cole, R., Greene, A. M., Oduro, A. K., Lazarova, K., Cesnik, A. J., … Olivier, M. (2016). Hyccapp as a tool to characterize promoter dna-protein interactions in saccharomyces cerevisiae. *Genomics*, *107*, 267–273. https://doi.org/10.1016/j.ygeno.2016.05.002

Guo, J., Li, T., Schipper, J., Nilson, K. A., Fordjour, F. K., Cooper, J. J., Gordân, R., & Price, D. H. (2014). Sequence specificity incompletely defines the genome-wide occupancy of myc. *Genome Biology*, *15*, 482. https://doi.org/10.1186/s13059-014-0482-3

Guo, X., Ma, J., Sun, J., & Gao, G. (2007). The zinc-finger antiviral protein recruits the rna processing exosome to degrade the target mrna. *Proceedings of the National Academy of Sciences*, *104*, 151–156. https://doi.org/10.1073/pnas.0607063104

Hahn, M. A., Jin, S.-G., Li, A. X., Liu, J., Huang, Z., Wu, X., Kim, B.-W., Johnson, J., Bilbao, A.-D. V., Tao, S., Yim, J. A., Fong, Y., Goebbels, S., Schwab, M. H., Lu, Q., & Pfeifer, G. P. (2019). Reprogramming of dna methylation at neurod2-bound sequences during cortical neuron differentiation. *Science Advances*, *5*. https://doi.org/10.1126/sciadv.aax0080

Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., Jui, J., Jin, S.-G., Jiang, Y., Pfeifer, G. P., & Lu, Q. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Reports*, *3*, 291–300. https://doi.org/10.1016/j.celrep.2013.01.011

Hainer, S. J., Gu, W., Carone, B. R., Landry, B. D., Rando, O. J., Mello, C. C., & Fazzio, T. G. (2015). Suppression of pervasive noncoding transcription in embryonic stem cells by esbaf. *Genes & Development*, *29*, 362–378. https://doi.org/10.1101/gad.253534.114

Halliwell, D. O., Honig, F., Bagby, S., Roy, S., & Murrell, A. (2025). Double and single stranded detection of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore sequencing. *Communications Biology*, *8*, 243. https://doi.org/10.1038/s42003-025-07681-0

Hardwick, J. S., Dhir, S., Kirchner, A., Simeone, A., Flynn, S. M., Edgerton, J. M., de Cesaris Araujo Tavares, R., Esain-Garcia, I., Tannahill, D., Golder, P., Monahan, J. M., Gosal, W. S., & Balasubramanian, S. (2025). Scotch-seq reveals that 5-hydroxymethylcytosine encodes regulatory information across dna strands. *Proceedings of the National Academy of Sciences of the United States of America*, *122*, e2512204122. https://doi.org/10.1073/pnas.2512204122

Hardwick, J. S., Ptchelkine, D., El-Sagheer, A. H., Tear, I., Singleton, D., Phillips, S. E. V., Lane, A. N., & Brown, T. (2017). 5-formylcytosine does not change the global structure of dna. *Nature Structural & Molecular Biology*, *24*, 544–552. https://doi.org/10.1038/nsmb.3411

Harrison, S. C., & Aggarwal, A. K. (1990). Dna recognition by proteins with the helix-turn-helix motif. *Annual Review of Biochemistry*, *59*, 933–969. https://doi.org/10.1146/annurev.bi.59.070190.004441

Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X., & Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*, *40*, 4841–4849. https://doi.org/10.1093/nar/gks155

Hashimoto, H., Olanrewaju, Y. O., Zheng, Y., Wilson, G. G., Zhang, X., & Cheng, X. (2014). Wilms tumor protein recognizes 5-carboxylcytosine within a specific dna sequence. *Genes & Development*, *28*, 2304–2313. https://doi.org/10.1101/gad.250746.114

Hashimoto, H., Zhang, X., & Cheng, X. (2013). Activity and crystal structure of human thymine dna glycosylase mutant n140a with 5-carboxylcytosine dna at low ph. *DNA Repair*, *12*, 535–540. https://doi.org/10.1016/j.dnarep.2013.04.003

Hata, K., & Sakaki, Y. (1997). Identification of critical cpg sites for repression of l1 transcription by dna methylation. *Gene*, *189*, 227–234. https://doi.org/10.1016/S0378-1119(96)00856-6

Hautbergue, G. M., Hung, M.-L., Walsh, M. J., Snijders, A. P., Chang, C.-T., Jones, R., Ponting, C. P., Dickman, M. J., & Wilson, S. A. (2009). Uif, a new mrna export adaptor that works together with ref/aly, requires fact for recruitment to mrna. *Current Biology*, *19*, 1918–1924. https://doi.org/10.1016/j.cub.2009.09.041

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., & Xu, G.-L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by tdg in mammalian dna. *Science*, *333*, 1303–1307. https://doi.org/10.1126/science.1210944

Helabad, M., Kanaan, N., & Imhof, P. (2014). Base flip in dna studied by molecular dynamics simulationsof differently-oxidized forms of methyl-cytosine. *International Journal of Molecular Sciences*, *15*, 11799–11816. https://doi.org/10.3390/ijms150711799

Hendrich, B., & Bird, A. (1998). Identification and characterization of a family of mammalian methyl-cpg binding proteins. *Molecular and Cellular Biology*, *18*, 6538–6547. https://doi.org/10.1128/MCB.18.11.6538

Hermann, A., Goyal, R., & Jeltsch, A. (2004). The dnmt1 dna-(cytosine-c5)-methyltransferase methylates dna processively with high preference for hemimethylated target sites. *Journal of Biological Chemistry*, *279*, 48350–48359. https://doi.org/10.1074/jbc.M403427200

Hervouet, E., Peixoto, P., Delage-Mourroux, R., Boyer-Guittaut, M., & Cartron, P.-F. (2018). Specific or not specific recruitment of dnmts for dna methylation, an epigenetic dilemma. *Clinical Epigenetics*, *10*, 17. https://doi.org/10.1186/s13148-018-0450-y

Hervouet, E., Vallette, F. M., & Cartron, P.-F. (2009). Dnmt3/transcription factor interactions as crucial players in targeted dna methylation. *Epigenetics*, *4*, 487–499. https://doi.org/10.4161/epi.4.7.9883

Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., & Walkinshaw, M. D. (2008). Mecp2 binding to dna depends upon hydration at methyl-cpg. *Molecular Cell*, *29*, 525–531. https://doi.org/10.1016/j.molcel.2007.12.028

Hoguin, A., Yang, F., Groisillier, A., Bowler, C., Genovesio, A., Ait-Mohamed, O., Vieira, F. R. J., & Tirichine, L. (2023). The model diatom phaeodactylum tricornutum provides insights into the diversity and function of microeukaryotic dna methyltransferases. *Communications Biology*, *6*, 253. https://doi.org/10.1038/s42003-023-04629-0

Holland, P. W. H. (2013). Evolution of homeobox genes. *WIREs Developmental Biology*, *2*, 31–45. https://doi.org/10.1002/wdev.78

Hon, G. C., Song, C.-X., Du, T., Jin, F., Selvaraj, S., Lee, A. Y., Yen, C.-a., Ye, Z., Mao, S.-Q., Wang, B.-A., Kuan, S., Edsall, L. E., Zhao, B. S., Xu, G.-L., He, C., & Ren, B. (2014). 5mc oxidation by tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Molecular Cell*, *56*, 286–297. https://doi.org/10.1016/j.molcel.2014.08.026

Howard, M. J., Foley, K. G., Shock, D. D., Batra, V. K., & Wilson, S. H. (2019). Molecular basis for the faithful replication of 5-methylcytosine and its oxidized forms by dna polymerase β. *Journal of Biological Chemistry*, *294*, 7194–7201. https://doi.org/10.1074/jbc.RA118.006809

Hu, H.-g., Scholten, I., Gruss, C., & Knippers, R. (2007). The distribution of the dek protein in mammalian chromatin. *Biochemical and Biophysical Research Communications*, *358*, 1008–1014. https://doi.org/10.1016/j.bbrc.2007.05.019

Hu, J., Banerjee, A., & Goss, D. J. (2005). Assembly of b/hlh/z proteins c-myc, max, and mad1 with cognate dna: Importance of protein-protein and protein-dna interactions. *Biochemistry*, *44*, 11855–11863. https://doi.org/10.1021/bi050206i

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P., & Xu, Y. (2013). Crystal structure of tet2-dna complex: Insight into tet-mediated 5mc oxidation. *Cell*, *155*, 1545–1555. https://doi.org/10.1016/j.cell.2013.11.020

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., Liu, M., Sun, C., Yin, X., Li, J., Tan, X., Wang, P., Wang, Y., Fang, D., Cui, Q., … Xu, Y. (2015). Structural insight into substrate preference for tet-mediated oxidation. *Nature*, *527*, 118–122. https://doi.org/10.1038/nature15713

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H. N., Shin, J., Cox, E., Rho, H. S., Woodard, C., Xia, S., Liu, S., Lyu, H., Ming, G.-L., Wade, H., Song, H., Qian, J., & Zhu, H. (2013). Dna methylation presents distinct binding sites for human transcription factors. *eLife*, *2*. https://doi.org/10.7554/eLife.00726

Huang, A., Ho, C. S., Ponzielli, R., Barsyte-Lovejoy, D., Bouffet, E., Picard, D., Hawkins, C. E., & Penn, L. Z. (2005). Identification of a novel c-myc protein interactor, jpo2, with transforming activity in medulloblastoma cells. *Cancer Research*, *65*, 5607–5619. https://doi.org/10.1158/0008-5472.CAN-05-0500

Huang, X., Wang, H.-L., Qi, S.-T., Wang, Z.-B., Tong, J.-S., Zhang, Q.-H., Ouyang, Y.-C., Hou, Y., Schatten, H., Qi, Z.-Q., & Sun, Q.-Y. (2011). Dynlt3 is required for chromosome alignment during mouse oocyte meiotic maturation. *Reproductive Sciences*, *18*, 983–989. https://doi.org/10.1177/1933719111401664

Hubner, N. C., Nguyen, L. N., Hornig, N. C., & Stunnenberg, H. G. (2015). A quantitative proteomics tool to identify dna–protein interactions in primary cells or blood. *Journal of Proteome Research*, *14*, 1315–1329. https://doi.org/10.1021/pr5009515

Huen, J., Kakihara, Y., Ugwu, F., Cheung, K. L. Y., Ortega, J., & Houry, W. A. (2010). Rvb1–rvb2: Essential atp-dependent helicases for critical complexesthis paper is one of a selection of papers published in this special issue entitled 8th international conference on aaa proteins and has undergone the journal's usual peer review process. *Biochemistry and Cell Biology*, *88*, 29–40. https://doi.org/10.1139/O09-122

Huff, J. T., & Zilberman, D. (2014). Dnmt1-independent cg methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell*, *156*, 1286–1297. https://doi.org/10.1016/j.cell.2014.01.029

Hurlin, P. J. (1999). Mga, a dual-specificity transcription factor that interacts with max and contains a t-domain dna-binding motif. *The EMBO Journal*, *18*, 7019–7028. https://doi.org/10.1093/emboj/18.24.7019

Huss, J. M., Garbacz, W. G., & Xie, W. (2015). Constitutive activities of estrogen-related receptors: Transcriptional regulation of metabolism by the err pathways in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1852*, 1912–1927. https://doi.org/10.1016/j.bbadis.2015.06.016

Hutnick, L. K., Golshani, P., Namihira, M., Xue, Z., Matynia, A., Yang, X. W., Silva, A. J., Schweizer, F. E., & Fan, G. (2009). Dna hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. *Human Molecular Genetics*, *18*, 2875–2888. https://doi.org/10.1093/hmg/ddp222

Hwang, J. H., Arafeh, R., Seo, J.-H., Baca, S. C., Ludwig, M., Arnoff, T. E., Sawyer, L., Richter, C., Tape, S., Bergom, H. E., McSweeney, S., Rennhack, J. P., Klingenberg, S. A., Cheung, A. T., Kwon, J., So, J., Kregel, S., Allen, E. M. V., Drake, J. M., … Hahn, W. C. (2022). Creb5 reprograms foxa1 nuclear interactions to promote resistance to androgen receptor-targeting therapies. *eLife*, *11*. https://doi.org/10.7554/eLife.73223

Ibrahim, A., Papin, C., Mohideen-Abdul, K., Gras, S. L., Stoll, I., Bronner, C., Dimitrov, S., Klaholz, B. P., & Hamiche, A. (2021). Mecp2 is a microsatellite binding protein that protects ca repeats from nucleosome invasion. *Science*, *372*. https://doi.org/10.1126/science.abd5581

Invivogen. (2025). Invivogen webpage. https://www.invivogen.com/pdrive-muc1

Isakova, A., Groux, R., Imbeault, M., Rainer, P., Alpern, D., Dainese, R., Ambrosini, G., Trono, D., Bucher, P., & Deplancke, B. (2017). Smile-seq identifies binding motifs of single and dimeric transcription factors. *Nature Methods*, *14*, 316–322. https://doi.org/10.1038/nmeth.4143

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of tet proteins in 5mc to 5hmc conversion, es-cell self-renewal and inner cell mass specification. *Nature*, *466*, 1129–1133. https://doi.org/10.1038/nature09303

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*, 1300–1303. https://doi.org/10.1126/science.1210597

Ito, Y., Bae, S.-C., & Chuang, L. S. H. (2015). The runx family: Developmental regulators in cancer. *Nature Reviews Cancer*, *15*, 81–95. https://doi.org/10.1038/nrc3877

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., Andrews, S., Balasubramanian, S., & Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, *14*. https://doi.org/10.1186/gb-2013-14-10-r119

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J. A., Mahony, S., Pugh, B. F., Lee, D., Kaestner, K. H., & Zaret, K. S. (2016). The pioneer transcription factor foxa maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Molecular Cell*, *62*, 79–91. https://doi.org/10.1016/j.molcel.2016.03.001

Janssens, D. H., Greene, J. E., Wu, S. J., Codomo, C. A., Minot, S. S., Furlan, S. N., Ahmad, K., & Henikoff, S. (2024). Scalable single-cell profiling of chromatin modifications with scicut&tag. *Nature Protocols*, *19*, 83–112. https://doi.org/10.1038/s41596-023-00905-9

Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T. P., Lyko, F., Reuter, G., Ankri, S., Nellen, W., Schaefer, M., & Helm, M. (2017). Mechanism and biological role of dnmt2 in nucleic acid methylation. *RNA Biology*, *14*, 1108–1123. https://doi.org/10.1080/15476286.2016.1191737

Jeltsch, A., & Jurkowska, R. Z. (2014). New concepts in dna methylation. *Trends in Biochemical Sciences*, *39*, 310–318. https://doi.org/10.1016/j.tibs.2014.05.002

Jeong, W.-J., Park, J.-C., Kim, W.-S., Ro, E. J., Jeon, S. H., Lee, S.-K., Park, Y. N., Min, D. S., & Choi, K.-Y. (2019). Wdr76 is a ras binding protein that functions as a tumor suppressor via ras degradation. *Nature Communications*, *10*, 295. https://doi.org/10.1038/s41467-018-08230-6

Ji, S., Shao, H., Han, Q., Seiler, C. L., & Tretyakova, N. Y. (2017). Reversible dna–protein cross-linking at epigenetic dna marks. *Angewandte Chemie International Edition*, *56*, 14130–14134. https://doi.org/10.1002/anie.201708286

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of dnmt3a bound to dnmt3l suggests a model for de novo dna methylation. *Nature*, *449*, 248–251. https://doi.org/10.1038/nature06146

Jiang, Q., Ang, J. Y., Lee, A. Y., Cao, Q., Li, K. Y., Yip, K. Y., & Leung, D. C. (2020). G9a plays distinct roles in maintaining dna methylation, retrotransposon silencing, and chromatin looping. *Cell Reports*, *33*, 108315. https://doi.org/10.1016/j.celrep.2020.108315

Jiang, X., Norman, M., Roth, L., & Li, X. (2004). Protein-dna array-based identification of transcription factor activities regulated by interaction with the glucocorticoid receptor. *Journal of Biological Chemistry*, *279*, 38480–38485. https://doi.org/10.1074/jbc.M403948200

Jin, S.-G., Kadam, S., & Pfeifer, G. P. (2010). Examination of the specificity of dna methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Research*, *38*, e125–e125. https://doi.org/10.1093/nar/gkq223

Jin, S.-G., Wu, X., Li, A. X., & Pfeifer, G. P. (2011). Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Research*, *39*, 5015–5024. https://doi.org/10.1093/nar/gkr120

Jin, S.-G., Zhang, Z.-M., Dunwell, T. L., Harter, M. R., Wu, X., Johnson, J., Li, Z., Liu, J., Szabó, P. E., Lu, Q., Xu, G.-l., Song, J., & Pfeifer, G. P. (2016). Tet3 reads 5-carboxylcytosine through its cxxc domain and is a potential guardian against neurodegeneration. *Cell Reports*, *14*, 493–505. https://doi.org/10.1016/j.celrep.2015.12.044

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-dna interactions. *Science*, *316*, 1497–1502. https://doi.org/10.1126/science.1141319

Johnson, P. F., & McKnight, S. L. (1989). Eukaryotic transcriptional regulatory proteins. *Annual Review of Biochemistry*, *58*, 799–839. https://doi.org/10.1146/annurev.bi.58.070189.004055

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J. M., Yan, J., Sillanpää, M. J., Bonke, M., Palin, K., Talukder, S., Hughes, T. R., Luscombe, N. M., Ukkonen, E., & Taipale, J. (2010). Multiplexed massively parallel selex for characterization of human transcription factor binding specificities. *Genome Research*, *20*, 861–873. https://doi.org/10.1101/gr.100552.109

Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., & Taipale, J. (2015). Dna-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, *527*, 384–388. https://doi.org/10.1038/nature15518

Jones, P. A. (2012). Functions of dna methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*, 484–492. https://doi.org/10.1038/nrg3230

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics*, *3*, 415–428. https://doi.org/10.1038/nrg816

Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., & Wolffe, A. P. (1998). Methylated dna and mecp2 recruit histone deacetylase to repress transcription. *Nature Genetics*, *19*, 187–191. https://doi.org/10.1038/561

Jung, L. A., Gebhardt, A., Koelmel, W., Ade, C. P., Walz, S., Kuper, J., von Eyss, B., Letschert, S., Redel, C., d'Artista, L., Biankin, A., Zender, L., Sauer, M., Wolf, E., Evan, G., Kisker, C., & Eilers, M. (2017). Omomyc blunts promoter invasion by oncogenic myc to inhibit gene expression characteristic of myc-dependent tumors. *Oncogene*, *36*, 1911–1924. https://doi.org/10.1038/onc.2016.354

Kafer, G. R., Li, X., Horii, T., Suetake, I., Tajima, S., Hatada, I., & Carlton, P. M. (2016). 5-hydroxymethylcytosine marks sites of dna damage and promotes genome stability. *Cell Reports*, *14*, 1283–1292. https://doi.org/10.1016/j.celrep.2016.01.035

Kang, Y. K., Schiff, R., Ko, L., Wang, T., Tsai, S. Y., Tsai, M.-J., & O'Malley, B. W. (2008). Dual roles for coactivator activator and its counterbalancing isoform coactivator modulator in human kidney cell tumorigenesis. *Cancer Research*, *68*, 7887–7896. https://doi.org/10.1158/0008-5472.CAN-08-1734

Kappes, F., Waldmann, T., Mathew, V., Yu, J., Zhang, L., Khodadoust, M. S., Chinnaiyan, A. M., Luger, K., Erhardt, S., Schneider, R., & Markovitz, D. M. (2011). The dek oncoprotein is a su(var) that is essential to heterochromatin integrity. *Genes & Development*, *25*, 673–678. https://doi.org/10.1101/gad.2036411

Karemaker, I. D., & Vermeulen, M. (2018). Zbtb2 reads unmethylated cpg island promoters and regulates embryonic stem cell differentiation. *EMBO reports*, *19*. https://doi.org/10.15252/embr.201744993

Kato, G. J., Lee, W. M., Chen, L. L., & Dang, C. V. (1992). Max: Functional domains and interaction with c-myc. *Genes & Development*, *6*, 81–92. https://doi.org/10.1101/gad.6.1.81

Kaufman, P. D., Kobayashi, R., Kessler, N., & Stillman, B. (1995). The p150 and p60 subunits of chromatin assemblyfactor i: A molecular link between newly synthesized histories and dna replication. *Cell*, *81*, 1105–1114. https://doi.org/10.1016/S0092-8674(05)80015-7

Kawasaki, H., Chang, H.-W., Tseng, H.-C., Hsu, S.-C., Yang, S.-J., Hung, C.-H., Zhou, Y., & Huang, S.-K. (2014). A tryptophan metabolite, kynurenine, promotes mast cell activation through aryl hydrocarbon receptor. *Allergy*, *69*, 445–452. https://doi.org/10.1111/all.12346

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). Cut&tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*, *10*, 1930. https://doi.org/10.1038/s41467-019-09982-5

Kehle, J. (1998). Dmi-2, a hunchback-interacting protein that functions in polycomb repression. *Science*, *282*, 1897–1900. https://doi.org/10.1126/science.282.5395.1897

Khan, O., Giles, J. R., McDonald, S., Manne, S., Ngiow, S. F., Patel, K. P., Werner, M. T., Huang, A. C., Alexander, K. A., Wu, J. E., Attanasio, J., Yan, P., George, S. M., Bengsch, B., Staupe, R. P., Donahue, G., Xu, W., Amaravadi, R. K., Xu, X., … Wherry, E. J. (2019). Tox transcriptionally and epigenetically programs cd8+ t cell exhaustion. *Nature*, *571*, 211–218. https://doi.org/10.1038/s41586-019-1325-x

Khund-Sayeed, S., He, X., Holzberg, T., Wang, J., Rajagopal, D., Upadhyay, S., Durell, S. R., Mukherjee, S., Weirauch, M. T., Rose, R., & Vinson, C. (2016). 5-hydroxymethylcytosine in e-box motifs acat|gtg and acac|gtg increases dna-binding of the b-hlh transcription factor tcf4. *Integrative Biology*, *8*, 936–945. https://doi.org/10.1039/C6IB00079G

Kigoshi, Y., Tsuruta, F., & Chiba, T. (2011). Ubiquitin ligase activity of cul3-klhl7 protein is attenuated by autosomal dominant retinitis pigmentosa causative mutation. *Journal of Biological Chemistry*, *286*, 33613–33621. https://doi.org/10.1074/jbc.M111.245126

Kitsera, N., Allgayer, J., Parsa, E., Geier, N., Rossa, M., Carell, T., & Khobta, A. (2017). Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemimodified cpg dinucleotide in a gene promoter. *Nucleic Acids Research*, *45*, 11033–11042. https://doi.org/10.1093/nar/gkx718

Kloet, S. L., Baymaz, H. I., Makowski, M., Groenewold, V., Jansen, P. W. T. C., Berendsen, M., Niazi, H., Kops, G. J., & Vermeulen, M. (2015). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *The FEBS Journal*, *282*, 1774–1785. https://doi.org/10.1111/febs.12972

Knezevic, J., Langer, A., Hampel, P. A., Kaiser, W., Strasser, R., & Rant, U. (2012). Quantitation of affinity, avidity, and binding kinetics of protein analytes with a dynamically switchable biosurface. *Journal of the American Chemical Society*, *134*, 15225–15228. https://doi.org/10.1021/ja3061276

Knott, G. J., Bond, C. S., & Fox, A. H. (2016). The dbhs proteins sfpq, nono and pspc1: A multipurpose molecular scaffold. *Nucleic Acids Research*, *44*, 3989–4004. https://doi.org/10.1093/nar/gkw271

Ko, M., An, J., Bandukwala, H. S., Chavez, L., Äijö, T., Pastor, W. A., Segal, M. F., Li, H., Koh, K. P., Lähdesmäki, H., Hogan, P. G., Aravind, L., & Rao, A. (2013). Modulation of tet2 expression and 5-methylcytosine oxidation by the cxxc domain protein idax. *Nature*, *497*, 122–126. https://doi.org/10.1038/nature12052

Köferle, A., Schlattl, A., Hörmann, A., Thatikonda, V., Popa, A., Spreitzer, F., Ravichandran, M. C., Supper, V., Oberndorfer, S., Puchner, T., Wieshofer, C., Corcokovic, M., Reiser, C., Wöhrle, S., Popow, J., Pearson, M., Martinez, J., Weitzer, S., Mair, B., & Neumüller, R. A. (2022). Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. *Cell Reports*, *39*, 110636. https://doi.org/10.1016/j.celrep.2022.110636

Kojima, Y., Tam, O. H., & Tam, P. P. (2014). Timing of developmental events in the early mouse embryo. *Seminars in Cell & Developmental Biology*, *34*, 65–75. https://doi.org/10.1016/j.semcdb.2014.06.010

Kretzner, L., Blackwood, E. M., & Eisenman, R. N. (1992). Myc and max proteins possess distinct transcriptional activities. *Nature*, *359*, 426–429. https://doi.org/10.1038/359426a0

Kriaucionis, S., & Heintz, N. (2009). The nuclear dna base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science*, *324*, 929–930. https://doi.org/10.1126/science.1169786

Kribelbauer, J. F., Laptenko, O., Chen, S., Martini, G. D., Freed-Pastor, W. A., Prives, C., Mann, R. S., & Bussemaker, H. J. (2017). Quantitative analysis of the dna methylation sensitivity of transcription factor complexes. *Cell Reports*, *19*, 2383–2395. https://doi.org/10.1016/j.celrep.2017.05.069

Kumar, S., Cheng, X., Klimasauskas, S., Sha, M., Posfai, J., Roberts, R. J., & Wilson, G. G. (1994). The dna (cytosine-5) methyltransferases. *Nucleic Acids Research*, *22*, 1–10. https://doi.org/10.1093/nar/22.1.1

Kunert, F., Metzner, F. J., Jung, J., Höpfler, M., Woike, S., Schall, K., Kostrewa, D., Moldt, M., Chen, J.-X., Bantele, S., Pfander, B., Eustermann, S., & Hopfner, K.-P. (2022). Structural mechanism of extranucleosomal dna readout by the ino80 complex. *Science Advances*, *8*. https://doi.org/10.1126/sciadv.add3189

Kustatscher, G., Hegarat, N., Wills, K. L. H., Furlan, C., Bukowski-Wills, J.-C., Hochegger, H., & Rappsilber, J. (2014). Proteomics of a fuzzy organelle: Interphase chromatin. *The EMBO Journal*, *33*, 648–664. https://doi.org/10.1002/embj.201387614

Kvietikova, I., Wenger, R. H., Marti, H. H., & Gassmann, M. (1995). The transcription factors atf-1 and creb-1 bind constitutively to the hypoxia-inducible factor-1 (hif-1)dna recognition site. *Nucleic Acids Research*, *23*, 4542–4550. https://doi.org/10.1093/nar/23.22.4542

Kweon, S.-M., Zhu, B., Chen, Y., Aravind, L., Xu, S.-Y., & Feldman, D. E. (2017). Erasure of tet-oxidized 5-methylcytosine by a srap nuclease. *Cell Reports*, *21*, 482–494. https://doi.org/10.1016/j.celrep.2017.09.055

Lafaye, C., Barbier, E., Miscioscia, A., Saint-Pierre, C., Kraut, A., Couté, Y., Plo, I., Gasparutto, D., Ravanat, J.-L., & Breton, J. (2014). Dna binding of the p21 repressor zbtb2 is inhibited by cytosine hydroxymethylation. *Biochemical and Biophysical Research Communications*, *446*, 341–346. https://doi.org/10.1016/j.bbrc.2014.02.122

Lagger, S., Connelly, J. C., Schweikert, G., Webb, S., Selfridge, J., Ramsahoye, B. H., Yu, M., He, C., Sanguinetti, G., Sowers, L. C., Walkinshaw, M. D., & Bird, A. (2017). Mecp2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLOS Genetics*, *13*, e1006793. https://doi.org/10.1371/journal.pgen.1006793

Lahiri, D. K., & Ge, Y.-W. (2000). Electrophoretic mobility shift assay for the detection of specific dna–-protein complex in nuclear extracts from the cultured cells and frozen autopsy human brain tissue. *Brain Research Protocols*, *5*, 257–265. https://doi.org/10.1016/S1385-299X(00)00021-0

Lai, W. K. M., & Pugh, B. F. (2017). Understanding nucleosome dynamics and their links to gene expression and dna replication. *Nature Reviews Molecular Cell Biology*, *18*, 548–562. https://doi.org/10.1038/nrm.2017.47

Laird, C. D., Pleasant, N. D., Clark, A. D., Sneeden, J. L., Hassan, K. M. A., Manley, N. C., Vary, J. C., Morgan, T., Hansen, R. S., & Stöger, R. (2004). Hairpin-bisulfite pcr: Assessing epigenetic methylation patterns on complementary strands of individual dna molecules. *Proceedings of the National Academy of Sciences*, *101*, 204–209. https://doi.org/10.1073/pnas.2536758100

Lambert, J.-P., Mitchell, L., Rudner, A., Baetz, K., & Figeys, D. (2009). A novel proteomics approach for the discovery of chromatin-associated protein networks. *Molecular & Cellular Proteomics*, *8*, 870–882. https://doi.org/10.1074/mcp.M800447-MCP200

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The human transcription factors. *Cell*, *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029

Laszlo, A. H., Derrington, I. M., Brinkerhoff, H., Langford, K. W., Nova, I. C., Samson, J. M., Bartlett, J. J., Pavlenok, M., & Gundlach, J. H. (2013). Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore mspa. *Proceedings of the National Academy of Sciences*, *110*, 18904–18909. https://doi.org/10.1073/pnas.1310240110

Laugesen, A., Højfeldt, J. W., & Helin, K. (2019). Molecular mechanisms directing prc2 recruitment and h3k27 methylation. *Molecular Cell*, *74*, 8–18. https://doi.org/10.1016/j.molcel.2019.03.011

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying dna methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*, 204–220. https://doi.org/10.1038/nrg2719

Lazarevic, V., & Glimcher, L. H. (2011). T-bet in disease. *Nature Immunology*, *12*, 597–606. https://doi.org/10.1038/ni.2059

Lee, D., Kim, J. W., Seo, T., Hwang, S. G., Choi, E.-J., & Choe, J. (2002). Swi/snf complex interacts with tumor suppressor p53 and is necessary for the activation of p53-mediated transcription. *Journal of Biological Chemistry*, *277*, 22330–22337. https://doi.org/10.1074/jbc.M111987200

Lee, J. H., Jomaa, A., Chung, S., Fu, Y.-H. H., Qian, R., Sun, X., Hsieh, H.-H., Chandrasekar, S., Bi, X., Mattei, S., Boehringer, D., Weiss, S., Ban, N., & Shan, S.-o. (2021). Receptor compaction and gtpase rearrangement drive srp-mediated cotranslational protein translocation into the er. *Science Advances*, *7*. https://doi.org/10.1126/sciadv.abg0942

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, *152*, 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Lei, I., Tian, S., Chen, V., Zhao, Y., & Wang, Z. (2020). Swi/snf component baf250a coordinates oct4 and wnt signaling pathway to control cardiac lineage differentiation. *Frontiers in Cell and Developmental Biology*, *7*. https://doi.org/10.3389/fcell.2019.00358

Lentini, A., Lagerwall, C., Vikingsson, S., Mjoseng, H. K., Douvlataniotis, K., Vogt, H., Green, H., Meehan, R. R., Benson, M., & Nestor, C. E. (2018). A reassessment of dna-immunoprecipitation-based genomic profiling. *Nature Methods*, *15*, 499–504. https://doi.org/10.1038/s41592-018-0038-7

Leone, D. P., Heavner, W. E., Ferenczi, E. A., Dobreva, G., Huguenard, J. R., Grosschedl, R., & McConnell, S. K. (2015). Satb2 regulates the differentiation of both callosal and subcerebral projection neurons in the developing cerebral cortex. *Cerebral Cortex*, *25*, 3406–3419. https://doi.org/10.1093/cercor/bhu156

Lercher, L., McDonough, M. A., El-Sagheer, A. H., Thalhammer, A., Kriaucionis, S., Brown, T., & Schofield, C. J. (2014). Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.*, *50*, 1794–1796. https://doi.org/10.1039/C3CC48151D

Li, C., Lu, T., Chen, H., Yu, Z., & Chen, C. (2024). The up-regulation of syncrip promotes the proliferation and tumorigenesis via dnmt3a/p16 in colorectal cancer. *Scientific Reports*, *14*, 21570. https://doi.org/10.1038/s41598-024-59575-6

Li, D., Cao, T., Sun, X., Jin, S., Xie, D., Huang, X., Yang, X., Carmichael, G. G., Taylor, H. S., Diano, S., & Huang, Y. (2020). Hepatic tet3 contributes to type-2 diabetes by inducing the hnf4α fetal isoform. *Nature Communications*, *11*, 342. https://doi.org/10.1038/s41467-019-14185-z

Li, F., Zhang, Y., Bai, J., Greenberg, M. M., Xi, Z., & Zhou, C. (2017). 5-formylcytosine yields dna–protein cross-links in nucleosome core particles. *Journal of the American Chemical Society*, *139*, 10617–10620. https://doi.org/10.1021/jacs.7b05495

Li, H., Liefke, R., Jiang, J., Kurland, J. V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D. J., Bulyk, M. L., Shi, Y., & Wang, Z. (2017). Polycomb-like proteins link the prc2 complex to cpg islands. *Nature*, *549*, 287–291. https://doi.org/10.1038/nature23881

Li, S., Peng, Y., & Panchenko, A. R. (2022). Dna methylation: Precise modulation of chromatin structure and dynamics. *Current Opinion in Structural Biology*, *75*, 102430. https://doi.org/10.1016/j.sbi.2022.102430

Li, T., Wang, L., Du, Y., Xie, S., Yang, X., Lian, F., Zhou, Z., & Qian, C. (2018). Structural and mechanistic insights into uhrf1-mediated dnmt1 activation in the maintenance dna methylation. *Nucleic Acids Research*, *46*, 3218–3231. https://doi.org/10.1093/nar/gky104

Li, W., & Liu, M. (2011). Distribution of 5-hydroxymethylcytosine in different human tissues. *Journal of Nucleic Acids*, *2011*, 1–5. https://doi.org/10.4061/2011/870726

Li, W., Zhang, X., Lu, X., You, L., Song, Y., Luo, Z., Zhang, J., Nie, J., Zheng, W., Xu, D., Wang, Y., Dong, Y., Yu, S., Hong, J., Shi, J., Hao, H., Luo, F., Hua, L., Wang, P., … Liu, J. (2017). 5-hydroxymethylcytosine signatures in circulating cell-free dna as diagnostic biomarkers for human cancers. *Cell Research*, *27*, 1243–1257. https://doi.org/10.1038/cr.2017.121

Lian, C. G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., Lee, C.-w., Hu, D., Lian, B. Q., Kleffel, S., Yang, Y., Neiswender, J., Khorasani, A. J., Fang, R., Lezcano, C., … Shi, Y. G. (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell*, *150*, 1135–1146. https://doi.org/10.1016/j.cell.2012.07.033

LIAO, Y., GU, J., WU, Y., LONG, X., GE, D., XU, J., & DING, J. (2016). Low level of 5-hydroxymethylcytosine predicts poor prognosis in non-small cell lung cancer. *Oncology Letters*, *11*, 3753–3760. https://doi.org/10.3892/ol.2016.4474

Lines, M. A. (2002). Molecular genetics of axenfeld-rieger malformations. *Human Molecular Genetics*, *11*, 1177–1187. https://doi.org/10.1093/hmg/11.10.1177

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghighi, F. G., … Ecker, J. R. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science*, *341*. https://doi.org/10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human dna methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*, 315–322. https://doi.org/10.1038/nature08514

Little, D. R., Lynch, A. M., Yan, Y., Akiyama, H., Kimura, S., & Chen, J. (2021). Differential chromatin binding of the lung lineage transcription factor nkx2-1 resolves opposing murine alveolar cell fates in vivo. *Nature Communications*, *12*, 2509. https://doi.org/10.1038/s41467-021-22817-6

Liu, D., Li, L., Zhang, X.-X., Wan, D.-Y., Xi, B.-X., Hu, Z., Ding, W.-C., Zhu, D., Wang, X.-L., Wang, W., Feng, Z.-H., Wang, H., Ma, D., & Gao, Q.-L. (2014). Six1 promotes tumor lymphangiogenesis by coordinating tgfβ signals that increase expression of vegf-c. *Cancer Research*, *74*, 5597–5607. https://doi.org/10.1158/0008-5472.CAN-13-3598

Liu, J., Bang, A. G., Kintner, C., Orth, A. P., Chanda, S. K., Ding, S., & Schultz, P. G. (2005). Identification of the wnt signaling activator leucine-rich repeat in flightless interaction protein 2 by a genome-wide functional analysis. *Proceedings of the National Academy of Sciences*, *102*, 1927–1932. https://doi.org/10.1073/pnas.0409472102

Liu, K., Xu, C., Lei, M., Yang, A., Loppnau, P., Hughes, T. R., & Min, J. (2018). Structural basis for the ability of mbd domains to bind methyl-cg and tg sites in dna. *Journal of Biological Chemistry*, *293*, 7344–7354. https://doi.org/10.1074/jbc.RA118.001785

Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., Dickerson, K. E., Xie, S., Hon, G. C., Xuan, Z., Zhang, M. Q., Shao, Z., & Xu, J. (2017). In situ capture of chromatin interactions by biotinylated dcas9. *Cell*, *170*, 1028–1043.e19. https://doi.org/10.1016/j.cell.2017.08.003

Liu, Y., Zhang, B., Meng, X., Korn, M. J., Parent, J. M., Lu, L.-Y., & Yu, X. (2017). Uhrf2 regulates local 5-methylcytosine and suppresses spontaneous seizures. *Epigenetics*, *12*, 551–560. https://doi.org/10.1080/15592294.2017.1314423

López, C. M. R., Lloyd, A. J., Leonard, K., & Wilkinson, M. J. (2012). Differential effect of three base modifications on dna thermostability revealed by high resolution melting. *Analytical Chemistry*, *84*, 7336–7342. https://doi.org/10.1021/ac301459x

Lopez-Fabuel, I., Douce, J. L., Logan, A., James, A. M., Bonvento, G., Murphy, M. P., Almeida, A., & Bolaños, J. P. (2016). Complex i assembly into supercomplexes determines differential mitochondrial ros production in neurons and astrocytes. *Proceedings of the National Academy of Sciences*, *113*, 13063–13068. https://doi.org/10.1073/pnas.1613701113

Love, P. E., Warzecha, C., & Li, L. (2014). Ldb1 complexes: The new master regulators of erythroid gene transcription. *Trends in Genetics*, *30*, 1–9. https://doi.org/10.1016/j.tig.2013.10.001

Lu, C., Han, H. D., Mangala, L. S., Ali-Fehmi, R., Newton, C. S., Ozbun, L., Armaiz-Pena, G. N., Hu, W., Stone, R. L., Munkarah, A., Ravoori, M. K., Shahzad, M. M., Lee, J.-W., Mora, E., Langley, R. R., Carroll, A. R., Matsuo, K., Spannuth, W. A., Schmandt, R., … Sood, A. K. (2010). Regulation of tumor angiogenesis by ezh2. *Cancer Cell*, *18*, 185–197. https://doi.org/10.1016/j.ccr.2010.06.016

Lu, X., Han, D., Zhao, B. S., Song, C.-X., Zhang, L.-S., Doré, L. C., & He, C. (2015). Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide dna demethylation dynamics. *Cell Research*, *25*, 386–389. https://doi.org/10.1038/cr.2015.5

Ludigs, K., Seguín-Estévez, Q., Lemeille, S., Ferrero, I., Rota, G., Chelbi, S., Mattmann, C., MacDonald, H. R., Reith, W., & Guarda, G. (2015). Nlrc5 exclusively transactivates mhc class i and related genes through a distinctive sxy module. *PLOS Genetics*, *11*, e1005088. https://doi.org/10.1371/journal.pgen.1005088

Lukauskas, S., Tvardovskiy, A., Nguyen, N. V., Stadler, M., Faull, P., Ravnsborg, T., Aygenli, B. Ö., Dornauer, S., Flynn, H., Lindeboom, R. G. H., Barth, T. K., Brockers, K., Hauck, S. M., Vermeulen, M., Snijders, A. P., Müller, C. L., DiMaggio, P. A., Jensen, O. N., Schneider, R., & Bartke, T. (2024). Decoding chromatin states by proteomic profiling of nucleosome readers. *Nature*, *627*, 671–679. https://doi.org/10.1038/s41586-024-07141-5

Lund, G., Andersson, L., Lauria, M., Lindholm, M., Fraga, M. F., Villar-Garea, A., Ballestar, E., Esteller, M., & Zaina, S. (2004). Dna methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein e. *Journal of Biological Chemistry*, *279*, 29147–29154. https://doi.org/10.1074/jbc.M403618200

Luo, Y., Hara, T., Ishido, Y., Yoshihara, A., Oda, K., Makino, M., Ishii, N., Hiroi, N., & Suzuki, K. (2014). Rapid preparation of high-purity nuclear proteins from a small number of cultured cells for use in electrophoretic mobility shift assays. *BMC Immunology*, *15*, 586. https://doi.org/10.1186/s12865-014-0062-z

Luo, Z., Gao, X., Lin, C., Smith, E. R., Marshall, S. A., Swanson, S. K., Florens, L., Washburn, M. P., & Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Molecular Cell*, *57*, 685–694. https://doi.org/10.1016/j.molcel.2015.01.007

Lüscher, B., & Vervoorts, J. (2012). Regulation of gene transcription by the oncoprotein myc. *Gene*, *494*, 145–160. https://doi.org/10.1016/j.gene.2011.12.027

Lyst, M. J., Ekiert, R., Ebert, D. H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N. R., Robinson, N. D., de Lima Alves, F., Rappsilber, J., Greenberg, M. E., & Bird, A. (2013). Rett syndrome mutations abolish the interaction of mecp2 with the ncor/smrt co-repressor. *Nature Neuroscience*, *16*, 898–902. https://doi.org/10.1038/nn.3434

MacDonald, M. R., Machlin, E. S., Albin, O. R., & Levy, D. E. (2007). The zinc finger antiviral protein acts synergistically with an interferon-induced factor for maximal activity against alphaviruses. *Journal of Virology*, *81*, 13509–13518. https://doi.org/10.1128/JVI.00402-07

Mach, B., Steimle, V., Martinez-Soria, E., & Reith, W. (1996). Regulation of mhc class ii genes: Lessons from a disease. *Annual Review of Immunology*, *14*, 301–331. https://doi.org/10.1146/annurev.immunol.14.1.301

Machado, A. C. D., Zhou, T., Rao, S., Goel, P., Rastogi, C., Lazarovici, A., Bussemaker, H. J., & Rohs, R. (2015). Evolving insights on how cytosine methylation affects protein-dna binding. *Briefings in Functional Genomics*, *14*, 61–73. https://doi.org/10.1093/bfgp/elu040

Machida, S., Takizawa, Y., Ishimaru, M., Sugita, Y., Sekine, S., Nakayama, J.-i., Wolf, M., & Kurumizaka, H. (2018). Structural basis of heterochromatin formation by human hp1. *Molecular Cell*, *69*, 385–397.e8. https://doi.org/10.1016/j.molcel.2017.12.011

Mahlapuu, M., Enerbäck, S., & Carlsson, P. (2001). Haploinsufficiency of the forkhead gene foxf1, a target for sonic hedgehog signaling, causes lung and foregut malformations. *Development (Cambridge, England)*, *128*, 2397–406. https://doi.org/10.1242/dev.128.12.2397

Maiti, A., & Drohat, A. C. (2011). Thymine dna glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine. *Journal of Biological Chemistry*, *286*, 35334–35338. https://doi.org/10.1074/jbc.C111.284620

Malagnac, F., Wendel, B., Goyon, C., Faugeron, G., Zickler, D., Rossignol, J.-L., Noyer-Weidner, M., Vollmayr, P., Trautner, T. A., & Walter, J. (1997). A gene essential for de novo methylation and development in ascobolus reveals a novel type of eukaryotic dna methyltransferase structure. *Cell*, *91*, 281–290. https://doi.org/10.1016/S0092-8674(00)80410-9

Mallona, I., Ilie, I. M., Karemaker, I. D., Butz, S., Manzo, M., Caflisch, A., & Baubec, T. (2021). Flanking sequence preference modulates de novo dna methylation in the mouse genome. *Nucleic Acids Research*, *49*, 145–157. https://doi.org/10.1093/nar/gkaa1168

Mann, I. K., Chatterjee, R., Zhao, J., He, X., Weirauch, M. T., Hughes, T. R., & Vinson, C. (2013). Cg methylated microarrays identify a novel methylated sequence bound by the cebpb|atf4 heterodimer that is active in vivo. *Genome Research*, *23*, 988–997. https://doi.org/10.1101/gr.146654.112

Mann, J. R., Szabo, P. E., Reed, M. R., & Singer-Sam, J. (2000). Methylated dna sequences in genomic imprinting. *Critical Reviews in Eukaryotic Gene Expression*, *10*. https://doi.org/10.1615/CritRevEukarGeneExpr.v10.i3-4.30

Mao, D. Y., Watson, J. D., Yan, P. S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.-L., Farnham, P. J., Huang, T. H.-M., & Penn, L. Z. (2003). Analysis of myc bound loci identified by cpg island arrays shows that max is essential for myc-dependent repression. *Current Biology*, *13*, 882–886. https://doi.org/10.1016/S0960-9822(03)00297-5

Mao, S.-Q., Cuesta, S. M., Tannahill, D., & Balasubramanian, S. (2020). Genome-wide dna methylation signatures are determined by dnmt3a/b sequence preferences. *Biochemistry*, *59*, 2541–2550. https://doi.org/10.1021/acs.biochem.0c00339

Marasco, L. E., & Kornblihtt, A. R. (2023). The physiology of alternative splicing. *Nature Reviews Molecular Cell Biology*, *24*, 242–254. https://doi.org/10.1038/s41580-022-00545-z

Margueron, R., & Reinberg, D. (2011). The polycomb complex prc2 and its mark in life. *Nature*, *469*, 343–349. https://doi.org/10.1038/nature09784

Marsit, C. J., Houseman, E. A., Christensen, B. C., Gagne, L., Wrensch, M. R., Nelson, H. H., Wiemels, J., Zheng, S., Wiencke, J. K., Andrew, A. S., Schned, A. R., Karagas, M. R., & Kelsey, K. T. (2010). Identification of methylated genes associated with aggressive bladder cancer. *PLoS ONE*, *5*, e12334. https://doi.org/10.1371/journal.pone.0012334

Martini, E., Roche, D. M., Marheineke, K., Verreault, A., & Almouzni, G. (1998). Recruitment of phosphorylated chromatin assembly factor 1 to chromatin after uv irradiation of human cells. *The Journal of Cell Biology*, *143*, 563–575. https://doi.org/10.1083/jcb.143.3.563

Masternak, K., Barras, E., Zufferey, M., Conrad, B., Corthals, G., Aebersold, R., Sanchez, J.-C., Hochstrasser, D. F., Mach, B., & Reith, W. (1998). A gene encoding a novel rfx-associated transactivator is mutated in the majority of mhc class ii deficiency patients. *Nature Genetics*, *20*, 273–277. https://doi.org/10.1038/3081

Mateus, A., Kurzawa, N., Becher, I., Sridharan, S., Helm, D., Stein, F., Typas, A., & Savitski, M. M. (2020). Thermal proteome profiling for interrogating protein interactions. *Molecular Systems Biology*, *16*. https://doi.org/10.15252/msb.20199232

McGinty, R. K., Henrici, R. C., & Tan, S. (2014). Crystal structure of the prc1 ubiquitylation module bound to the nucleosome. *Nature*, *514*, 591–596. https://doi.org/10.1038/nature13890

McMahon, S. B., Wood, M. A., & Cole, M. D. (2000). The essential cofactor trrap recruits the histone acetyltransferase hgcn5 to c-myc. *Molecular and Cellular Biology*, *20*, 556–562. https://doi.org/10.1128/MCB.20.2.556-562.2000

Meehan, R., Lewis, J. D., & Bird, A. P. (1992). Characterization of mecp2, a vertebrate dna binding protein with affinity for methylated dna. *Nucleic Acids Research*, *20*, 5085–5092. https://doi.org/10.1093/nar/20.19.5085

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., & Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to dna containing methylated cpgs. *Cell*, *58*, 499–507. https://doi.org/10.1016/0092-8674(89)90430-3

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., & Heintz, N. (2012). Mecp2 binds to 5hmc enriched within active genes and accessible chromatin in the nervous system. *Cell*, *151*, 1417–1430. https://doi.org/10.1016/j.cell.2012.11.022

Mellén, M., Ayata, P., & Heintz, N. (2017). 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proceedings of the National Academy of Sciences*, *114*. https://doi.org/10.1073/pnas.1708044114

Merrifield, R. B. (1963). Solid phase peptide synthesis. i. the synthesis of a tetrapeptide. *Journal of the American Chemical Society*, *85*, 2149–2154. https://doi.org/10.1021/ja00897a025

Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). Dna methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & Development*, *28*, 812–828. https://doi.org/10.1101/gad.234294.113

Minkovsky, A., Sahakyan, A., Rankin-Gee, E., Bonora, G., Patel, S., & Plath, K. (2014). The mbd1-atf7ip-setdb1 pathway contributes to the maintenance of x chromosome inactivation. *Epigenetics & Chromatin*, *7*, 12. https://doi.org/10.1186/1756-8935-7-12

Mohn, F., Weber, M., Schübeler, D., & Roloff, T.-C. (2009). Methylated dna immunoprecipitation (medip). https://doi.org/10.1007/978-1-59745-522-0_5

Mohni, K. N., Wessel, S. R., Zhao, R., Wojciechowski, A. C., Luzwick, J. W., Layden, H., Eichman, B. F., Thompson, P. S., Mehta, K. P., & Cortez, D. (2019). Hmces maintains genome integrity by shielding abasic sites in single-strand dna. *Cell*, *176*, 144–153.e13. https://doi.org/10.1016/j.cell.2018.10.055

Moreno, C. S., Beresford, G. W., Louis-Plence, P., Morris, A. C., & Boss, J. M. (1999). Creb regulates mhc class ii expression in a ciita-dependent manner. *Immunity*, *10*, 143–151. https://doi.org/10.1016/S1074-7613(00)80015-1

Mori, Y., Kataoka, H., Miura, Y., Kawaguchi, M., Kubota, E., Ogasawara, N., Oshima, T., Tanida, S., Sasaki, M., Ohara, H., Mizoshita, T., Tatematsu, M., Asai, K., & Joh, T. (2007). Subcellular localization of atbf1 regulates muc5ac transcription in gastric cancer. *International Journal of Cancer*, *121*, 241–247. https://doi.org/10.1002/ijc.22654

Muhammad, B. A., Almozyan, S., Babaei-Jadidi, R., Onyido, E. K., Saadeddin, A., Kashfi, S. H., Spencer-Dene, B., Ilyas, M., Lourdusamy, A., Behrens, A., & Nateri, A. S. (2018). Flywch1, a novel suppressor of nuclear β-catenin, regulates migration and morphology in colorectal cancer. *Molecular Cancer Research*, *16*, 1977–1990. https://doi.org/10.1158/1541-7786.MCR-18-0262

Müller, U., Bauer, C., Siegl, M., Rottach, A., & Leonhardt, H. (2014). Tet-mediated oxidation of methylcytosine causes tdg or neil glycosylase dependent gene reactivation. *Nucleic Acids Research*, *42*, 8592–8604. https://doi.org/10.1093/nar/gku552

Müller-Hartmann, H., Deissler, H., Naumann, F., Schmitz, B., Schröer, J., & Doerfler, W. (2000). The human 20-kda 5-(cgg) -3-binding protein is targeted to the nucleus and affects the activity of the fmr1promoter. *Journal of Biological Chemistry*, *275*, 6447–6452. https://doi.org/10.1074/jbc.275.9.6447

Mustafi, P., Hu, M., Kumari, S., Das, C., Li, G., & Kundu, T. (2022). Phosphorylation-dependent association of human chromatin protein pc4 to linker histone h1 regulates genome organization and transcription. *Nucleic Acids Research*, *50*, 6116–6136. https://doi.org/10.1093/nar/gkac450

Nady, N., Krichevsky, L., Zhong, N., Duan, S., Tempel, W., Amaya, M. F., Ravichandran, M., & Arrowsmith, C. H. (2012). Histone recognition by human malignant brain tumor domains. *Journal of Molecular Biology*, *423*, 702–718. https://doi.org/10.1016/j.jmb.2012.08.022

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N., & Hirose, T. (2012). Alternative 3'-end processing of long noncoding rna initiates construction of nuclear paraspeckles. *The EMBO Journal*, *31*, 4020–4034. https://doi.org/10.1038/emboj.2012.251

Nakatani, T., Yamagata, K., Kimura, T., Oda, M., Nakashima, H., Hori, M., Sekita, Y., Arakawa, T., Nakamura, T., & Nakano, T. (2015). Stella preserves maternal chromosome integrity by inhibiting 5hmc-induced γh2 ax accumulation. *EMBO reports*, *16*, 582–589. https://doi.org/10.15252/embr.201439427

Napolitano, L. M., Jaffray, E. G., Hay, R. T., & Meroni, G. (2011). Functional interactions between ubiquitin e2 enzymes and trim proteins. *Biochemical Journal*, *434*, 309–319. https://doi.org/10.1042/BJ20101487

Narabayashi, H., Koma, C., Nakata, K., Ikegami, M., Nakanishi, Y., Ogihara, J., Tsuda, M., Hosono, A., Hanazawa, S., & Takahashi, K. (2022). Gut microbiota-dependent adaptor molecule recruits dna methyltransferase to the tlr4 gene in colonic epithelial cells to suppress inflammatory reactions. *Frontiers in Molecular Biosciences*, *9*. https://doi.org/10.3389/fmolb.2022.1005136

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Bello, F. D., & Oliviero, S. (2015). Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter dna methylation dynamics. *Cell Reports*, *10*, 674–683. https://doi.org/10.1016/j.celrep.2015.01.008

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., & Oliviero, S. (2013). Genome-wide analysis identifies a functional association of tet1 and polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biology*, *14*, R91. https://doi.org/10.1186/gb-2013-14-8-r91

Nestor, C. E., & Meehan, R. R. (2014). Hydroxymethylated dna immunoprecipitation (hmedip). https://doi.org/10.1007/978-1-62703-706-8_20

Neufeld, G., Cohen, T., Gengrinovitch, S., & Poltorak, Z. (1999). Vascular endothelial growth factor (vegf) and its receptors. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, *13*, 9–22.

Ng, C., Aichinger, M., Nguyen, T., Au, C., Najar, T., Wu, L., Mesa, K. R., Liao, W., Quivy, J.-P., Hubert, B., Almouzni, G., Zuber, J., & Littman, D. R. (2019). The histone chaperone caf-1 cooperates with the dna methyltransferases to maintain cd4 silencing in cytotoxic t cells. *Genes & Development*, *33*, 669–683. https://doi.org/10.1101/gad.322024.118

Nguyen, L. P., & Bradfield, C. A. (2008). The search for endogenous activators of the aryl hydrocarbon receptor. *Chemical Research in Toxicology*, *21*, 102–116. https://doi.org/10.1021/tx7001965

Niesen, M. I., Osborne, A. R., Yang, H., Rastogi, S., Chellappan, S., Cheng, J. Q., Boss, J. M., & Blanck, G. (2005). Activation of a methylated promoter mediated by a sequence-specific dna-binding protein, rfx. *Journal of Biological Chemistry*, *280*, 38914–38922. https://doi.org/10.1074/jbc.M504633200

Nowock, J., Borgmeyer, U., Pūschel, A. W., Rupp, R. A., & Sippel, A. E. (1985). The tggca protein binds to the mmtv-ltr, the adenovirus origin of replication, and the bk virus enhancer. *Nucleic Acids Research*, *13*, 2045–2061. https://doi.org/10.1093/nar/13.6.2045

O'Flaherty, E., & Kaye, J. (2003). Tox defines a conserved subfamily of hmg-box proteins. *BMC Genomics*, *4*, 13. https://doi.org/10.1186/1471-2164-4-13

Ogawa, H., Ishiguro, K.-i., Gaubatz, S., Livingston, D. M., & Nakatani, Y. (2002). A complex with chromatin modifiers that occupies e2f- and myc-responsive genes in g 0 cells. *Science*, *296*, 1132–1136. https://doi.org/10.1126/science.1069861

Oh, S. S., Plakos, K., Lou, X., Xiao, Y., & Soh, H. T. (2010). In vitro selection of structure-switching, self-reporting aptamers. *Proceedings of the National Academy of Sciences*, *107*, 14053–14058. https://doi.org/10.1073/pnas.1009172107

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). Dna methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*, 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6

Okashita, N., Kumaki, Y., Ebi, K., Nishi, M., Okamoto, Y., Nakayama, M., Hashimoto, S., Nakamura, T., Sugasawa, K., Kojima, N., Takada, T., Okano, M., & Seki, Y. (2014). Prdm14 promotes active dna demethylation through the ten-eleven translocation (tet)-mediated base excision repair pathway in embryonic stem cells. *Development*, *141*, 269–280. https://doi.org/10.1242/dev.099622

Ong, C.-T., & Corces, V. G. (2014). Ctcf: An architectural protein bridging genome topology and function. *Nature Reviews Genetics*, *15*, 234–246. https://doi.org/10.1038/nrg3663

Oppikofer, M., Bai, T., Gan, Y., Haley, B., Liu, P., Sandoval, W., Ciferri, C., & Cochran, A. G. (2017). Expansion of the iswi chromatin remodeler family with new active complexes. *EMBO reports*, *18*, 1697–1706. https://doi.org/10.15252/embr.201744011

Pantier, R., Mullin, N., Hall-Ponsele, E., & Chambers, I. (2020). Tet1 interacts directly with nanog via independent domains containing hydrophobic and aromatic residues. *Journal of Molecular Biology*, *432*, 6075–6091. https://doi.org/10.1016/j.jmb.2020.10.008

Papaioannou, V. E., & Silver, L. M. (1998). The t-box gene family. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *20*, 9–19. https://doi.org/10.1002/(SICI)1521-1878(199801)20:1<9::AID-BIES4>3.0.CO;2-Q

Papamichos-Chronakis, M., Krebs, J. E., & Peterson, C. L. (2006). Interplay between ino80 and swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptationin response to dna damage. *Genes & Development*, *20*, 2437–2449. https://doi.org/10.1101/gad.1440206

Park, Y., Lee, J.-M., Hwang, M.-Y., Son, G.-h., & Geum, D. (2013). Nono binds to the cpg island of oct4 promoter and functions as a transcriptional activator of oct4 gene expression. *Molecules and Cells*, *35*, 61–69. https://doi.org/10.1007/s10059-013-2273-1

Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., McLoughlin, E. M., Brudno, Y., Mahapatra, S., Kapranov, P., Tahiliani, M., Daley, G. Q., Liu, X. S., Ecker, J. R., Milos, P. M., Agarwal, S., & Rao, A. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, *473*, 394–397. https://doi.org/10.1038/nature10102

Patrick, A. N., Cabrera, J. H., Smith, A. L., Chen, X. S., Ford, H. L., & Zhao, R. (2013). Structure-function analyses of the human six1–eya2 complex reveal insights into metastasis and bor syndrome. *Nature Structural & Molecular Biology*, *20*, 447–453. https://doi.org/10.1038/nsmb.2505

Paul, A., Avci-Adali, M., Ziemer, G., & Wendel, H. P. (2009). Streptavidin-coated magnetic beads for dna strand separation implicate a multitude of problems during cell-selex. *Oligonucleotides*, *19*, 243–254. https://doi.org/10.1089/oli.2009.0194

Perera, A., Eisen, D., Wagner, M., Laube, S. K., Künzel, A. F., Koch, S., Steinbacher, J., Schulze, E., Splith, V., Mittermeier, N., Müller, M., Biel, M., Carell, T., & Michalakis, S. (2015). Tet3 is recruited by rest for context-specific hydroxymethylation and induction of gene expression. *Cell Reports*, *11*, 283–294. https://doi.org/10.1016/j.celrep.2015.03.020

Pérez, A., Castellazzi, C. L., Battistini, F., Collinet, K., Flores, O., Deniz, O., Ruiz, M. L., Torrents, D., Eritja, R., Soler-López, M., & Orozco, M. (2012). Impact of methylation on the physical properties of dna. *Biophysical Journal*, *102*, 2140–2148. https://doi.org/10.1016/j.bpj.2012.03.056

Petrif, F., Giles, R. H., Dauwerse, H. G., Saris, J. J., Hennekam, R. C. M., Masuno, M., Tommerup, N., van Ommen, G.-J. B., Goodman, R. H., Peters, D. J. M., & Breuning, M. H. (1995). Rubinstein-taybi syndrome caused by mutations in the transcriptional co-activator cbp. *Nature*, *376*, 348–351. https://doi.org/10.1038/376348a0

Petterson, A., Chung, T. H., Tan, D., Sun, X., & Jia, X.-Y. (2014). Rrhp: A tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. *Genome Biology*, *15*, 456. https://doi.org/10.1186/s13059-014-0456-5

Pfaffeneder, T., Hackner, B., Truß, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeier, C., & Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell dna. *Angewandte Chemie International Edition*, *50*, 7008–7012. https://doi.org/10.1002/anie.201103899

Pfeifer, G. P. (2006). Mutagenesis at methylated cpg sequences. In *Dna methylation: Basic mechanisms* (pp. 259–281). Springer-Verlag. https://doi.org/10.1007/3-540-31390-7_10

Pfeifer, G. P., Szabó, P. E., & Song, J. (2020). Protein interactions at oxidized 5-methylcytosine bases. *Journal of Molecular Biology*, *432*, 1718–1730. https://doi.org/10.1016/j.jmb.2019.07.039

Pierrou, S., Hellqvist, M., Samuelsson, L., Enerbäck, S., & Carlsson, P. (1994). Cloning and characterization of seven human forkhead proteins: Binding site specificity and dna bending. *The EMBO Journal*, *13*, 5002–5012. https://doi.org/10.1002/j.1460-2075.1994.tb06827.x

Ping, S.-Y., Shen, K.-H., & Yu, D.-S. (2013). Epigenetic regulation of vascular endothelial growth factor a dynamic expression in transitional cell carcinoma. *Molecular Carcinogenesis*, *52*, 568–579. https://doi.org/10.1002/mc.21892

Polach, K., & Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *Journal of Molecular Biology*, *258*, 800–812. https://doi.org/10.1006/jmbi.1996.0288

Ponger, L., & Li, W.-H. (2005). Evolutionary diversification of dna methyltransferases in eukaryotic genomes. *Molecular Biology and Evolution*, *22*, 1119–1128. https://doi.org/10.1093/molbev/msi098

Qi, Y., Ranish, J. A., Zhu, X., Krones, A., Zhang, J., Aebersold, R., Rose, D. W., Rosenfeld, M. G., & Carrière, C. (2008). Atbf1 is required for the pit1 gene early activation. *Proceedings of the National Academy of Sciences*, *105*, 2481–2486. https://doi.org/10.1073/pnas.0712196105

Rafiee, M.-R., Girardot, C., Sigismondo, G., & Krijgsveld, J. (2016). Expanding the circuitry of pluripotency by selective isolation of chromatin-associated proteins. *Molecular Cell*, *64*, 624–635. https://doi.org/10.1016/j.molcel.2016.09.019

Raiber, E.-A., Beraldi, D., Ficz, G., Burgess, H. E., Branco, M. R., Murat, P., Oxley, D., Booth, M. J., Reik, W., & Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine dna glycosylase. *Genome Biology*, *13*, R69. https://doi.org/10.1186/gb-2012-13-8-r69

Raiber, E.-A., Portella, G., Cuesta, S. M., Hardisty, R., Murat, P., Li, Z., Iurlaro, M., Dean, W., Spindel, J., Beraldi, D., Liu, Z., Dawson, M. A., Reik, W., & Balasubramanian, S. (2018). 5-formylcytosine organizes nucleosomes and forms schiff base interactions with histones in mouse embryonic stem cells. *Nature Chemistry*, *10*, 1258–1266. https://doi.org/10.1038/s41557-018-0149-x

Rao, R. C., & Dou, Y. (2015). Hijacked in cancer: The kmt2 (mll) family of methyltransferases. *Nature Reviews Cancer*, *15*, 334–346. https://doi.org/10.1038/nrc3929

Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., & Jeltsch, A. (2008). Protein lysine methyltransferase g9a acts on non-histone targets. *Nature Chemical Biology*, *4*, 344–346. https://doi.org/10.1038/nchembio.88

Rauch, J. N., & Gestwicki, J. E. (2014). Binding of human nucleotide exchange factors to heat shock protein 70 (hsp70) generates functionally distinct complexes in vitro. *Journal of Biological Chemistry*, *289*, 1402–1414. https://doi.org/10.1074/jbc.M113.521997

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., & Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone h3 methyltransferases. *Nature*, *406*, 593–9. https://doi.org/10.1038/35020506

Reese, B. E., Bachman, K. E., Baylin, S. B., & Rountree, M. R. (2003). The methyl-cpg binding protein mbd1 interacts with the p150 subunit of chromatin assembly factor 1. *Molecular and Cellular Biology*, *23*, 3226–3236. https://doi.org/10.1128/MCB.23.9.3226-3236.2003

Reiter, F., Wienerroither, S., & Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development*, *43*, 73–81. https://doi.org/10.1016/j.gde.2016.12.007

Ren, X., Ustiyan, V., Pradhan, A., Cai, Y., Havrilak, J. A., Bolte, C. S., Shannon, J. M., Kalin, T. V., & Kalinichenko, V. V. (2014). Foxf1 transcription factor is required for formation of embryonic vasculature by regulating vegf signaling in endothelial cells. *Circulation Research*, *115*, 709–720. https://doi.org/10.1161/CIRCRESAHA.115.304382

Renciuk, D., Blacque, O., Vorlickova, M., & Spingler, B. (2013). Crystal structures of b-dna dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Research*, *41*, 9891–9900. https://doi.org/10.1093/nar/gkt738

Renthal, W., & Nestler, E. J. (2009). Chromatin regulation in drug addiction and depression. *Dialogues in Clinical Neuroscience*, *11*, 257–268. https://doi.org/10.31887/DCNS.2009.11.3/wrenthal

Riggs, A. D., & Pfeifer, G. P. (1992). X-chromosome inactivation and cell memory. *Trends in Genetics*, *8*, 169–174. https://doi.org/10.1016/0168-9525(92)90219-T

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., & Längst, G. (2007). Dna sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proceedings of the National Academy of Sciences*, *104*, 15635–15640. https://doi.org/10.1073/pnas.0702430104

Robertson, A. B., Robertson, J., Fusser, M., & Klungland, A. (2014). Endonuclease g preferentially cleaves 5-hydroxymethylcytosine-modified dna creating a substrate for recombination. *Nucleic Acids Research*, *42*, 13280–13293. https://doi.org/10.1093/nar/gku1032

Robertson, J., Robertson, A. B., & Klungland, A. (2011). The presence of 5-hydroxymethylcytosine at the gene promoter and not in the gene body negatively regulates gene expression. *Biochemical and Biophysical Research Communications*, *411*, 40–43. https://doi.org/10.1016/j.bbrc.2011.06.077

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of specificity in protein-dna recognition. *Annual Review of Biochemistry*, *79*, 233–269. https://doi.org/10.1146/annurev-biochem-060408-091030

Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Cervantes, F., Sanchez, J., Garate, L., Barrios, M., Castillejo, J. A., Navarro, G., Colomer, D., Prosper, F., Heiniger, A., & Torres, A. (2005). Promoter hypomethylation of the line-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia. *Oncogene*, *24*, 7213–7223. https://doi.org/10.1038/sj.onc.1208866

Rosenfeld, M. G., Lunyak, V. V., & Glass, C. K. (2006). Sensors and signals: A coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes & Development*, *20*, 1405–1428. https://doi.org/10.1101/gad.1424806

Rothbart, S. B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A. I., Barsyte-Lovejoy, D., Martinez, J. Y., Bedford, M. T., Fuchs, S. M., Arrowsmith, C. H., & Strahl, B. D. (2012). Association of uhrf1 with methylated h3k9 directs the maintenance of dna methylation. *Nature Structural & Molecular Biology*, *19*, 1155–1160. https://doi.org/10.1038/nsmb.2391

Russell, W. C., Graham, F. L., Smiley, J., & Nairn, R. (1977). Characteristics of a human cell line transformed by dna from human adenovirus type 5. *Journal of General Virology*, *36*, 59–72. https://doi.org/10.1099/0022-1317-36-1-59

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G., Chandran, S., Hay, D. C., Bradley, M., Wilmut, I., & Sousa, P. D. (2011). Lineage-specific distribution of high levels of genomic. *Cell Research*, *21*, 1332–1342. https://doi.org/10.1038/cr.2011.113

Saha, A., Wittmeyer, J., & Cairns, B. R. (2002). Chromatin remodeling by rsc involves atp-dependent dna translocation. *Genes & Development*, *16*, 2120–2134. https://doi.org/10.1101/gad.995002

Saito, M., & Ishikawa, F. (2002). The mcpg-binding domain of human mbd3 does not bind to mcpg but interacts with nurd/mi2 components hdac1 and mta2. *Journal of Biological Chemistry*, *277*, 35434–35439. https://doi.org/10.1074/jbc.M203455200

Sakata, N., Kaneko, S., Ikeno, S., Miura, Y., Nakabayashi, H., Dong, X.-Y., Dong, J.-T., Tamaoki, T., Nakano, N., & Itoh, S. (2014). Tgf- β signaling cooperates with at motif-binding factor-1 for

repression of the α -fetoprotein promoter. *Journal of Signal Transduction*, *2014*, 1–11. https://doi.org/10.1155/2014/970346

Sander, J. D., & Joung, J. K. (2014). Crispr-cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*, 347–355. https://doi.org/10.1038/nbt.2842

Santiago, M., Antunes, C., Guedes, M., Sousa, N., & Marques, C. J. (2014). Tet enzymes and dna hydroxymethylation in neural development and function — how critical are they? *Genomics*, *104*, 334–340. https://doi.org/10.1016/j.ygeno.2014.08.018

Sawatsubashi, S., Murata, T., Lim, J., Fujiki, R., Ito, S., Suzuki, E., Tanabe, M., Zhao, Y., Kimura, S., Fujiyama, S., Ueda, T., Umetsu, D., Ito, T., Takeyama, K.-i., & Kato, S. (2010). A histone chaperone, dek, transcriptionally coactivates a nuclear receptor. *Genes & Development*, *24*, 159–170. https://doi.org/10.1101/gad.1857410

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., & Lyko, F. (2010). Rna methylation by dnmt2 protects transfer rnas against stress-induced cleavage. *Genes & Development*, *24*, 1590–1595. https://doi.org/10.1101/gad.586710

Schell, B., Legrand, P., & Fribourg, S. (2022). Crystal structure of sfpq-nono heterodimer. *Biochimie*, *198*, 1–7. https://doi.org/10.1016/j.biochi.2022.02.011

Schibler, U. (2021). Bmal1 dephosphorylation determines the pace of the circadian clock. *Genes & Development*, *35*, 1076–1078. https://doi.org/10.1101/gad.348801.121

Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F., & Leonhardt, H. (2016). Determination of local chromatin composition by casid. *Nucleus*, *7*, 476–484. https://doi.org/10.1080/19491034.2016.1239000

Schmitges, F. W., Radovani, E., Najafabadi, H. S., Barazandeh, M., Campitelli, L. F., Yin, Y., Jolma, A., Zhong, G., Guo, H., Kanagalingam, T., Dai, W. F., Taipale, J., Emili, A., Greenblatt, J. F., & Hughes, T. R. (2016). Multiparameter functional diversity of human c2h2 zinc finger proteins. *Genome Research*, *26*, 1742–1752. https://doi.org/10.1101/gr.209643.116

Schreiber, E., Matthias, P., Müller, M. M., & Schaffner, W. (1989). Rapid detection of octamer binding proteins with 'mini extracts', prepared from a small number of cells. *Nucleic Acids Research*, *17*, 6419–6419. https://doi.org/10.1093/nar/17.15.6419

Schuettengruber, B., Bourbon, H.-M., Croce, L. D., & Cavalli, G. (2017). Genome regulation by polycomb and trithorax: 70 years and counting. *Cell*, *171*, 34–57. https://doi.org/10.1016/j.cell.2017.08.002

Schütz, S., Bergsdorf, C., Goretzki, B., Lingel, A., Renatus, M., Gossert, A. D., & Jahnke, W. (2022). The disordered max n-terminus modulates dna binding of the transcription factor myc:max. *Journal of Molecular Biology*, *434*, 167833. https://doi.org/10.1016/j.jmb.2022.167833

Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences*, *73*, 804–808. https://doi.org/10.1073/pnas.73.3.804

Sefah, K., Shangguan, D., Xiong, X., O'Donoghue, M. B., & Tan, W. (2010). Development of dna aptamers using cell-selex. *Nature Protocols*, *5*, 1169–1185. https://doi.org/10.1038/nprot.2010.66

Segal, E., & Widom, J. (2009). From dna sequence to transcriptional behaviour: A quantitative approach. *Nature Reviews Genetics*, *10*, 443–456. https://doi.org/10.1038/nrg2591

Seisenberger, S., Peat, J. R., & Reik, W. (2013). Conceptual links between dna methylation reprogramming in the early embryo and primordial germ cells. *Current Opinion in Cell Biology*, *25*, 281–288. https://doi.org/10.1016/j.ceb.2013.02.013

Serandour, A. A., Mohammed, H., Miremadi, A., Mulder, K. W., & Carroll, J. S. (2018). Trps1 regulates oestrogen receptor binding and histone acetylation at enhancers. *Oncogene*, *37*, 5281–5291. https://doi.org/10.1038/s41388-018-0312-2

Sérandour, A. A., Avner, S., Mahé, E. A., Madigou, T., Guibert, S., Weber, M., & Salbert, G. (2016). Single-cpg resolution mapping of 5-hydroxymethylcytosine by chemical labeling and exonuclease digestion identifies evolutionarily unconserved cpgs as tet targets. *Genome Biology*, *17*, 56. https://doi.org/10.1186/s13059-016-0919-y

Sérandour, A. A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Péron, C. L., Madigou, T., Durand, E., Froguel, P., Staels, B., Lefebvre, P., Métivier, R., Eeckhoute, J., & Salbert, G. (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Research*, *40*, 8255–8265. https://doi.org/10.1093/nar/gks595

Shen, H., & Laird, P. W. (2013). Interplay between the cancer genome and epigenome. *Cell*, *153*, 38–55. https://doi.org/10.1016/j.cell.2013.03.008

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H.-L., Zhang, K., & Zhang, Y. (2013). Genome-wide analysis reveals tet- and tdg-dependent 5-methylcytosine oxidation dynamics. *Cell*, *153*, 692–706. https://doi.org/10.1016/j.cell.2013.04.002

Shi, X., Wallis, A. M., Gerard, R. D., Voelker, K. A., Grange, R. W., DePinho, R. A., Garry, M. G., & Garry, D. J. (2012). Foxk1 promotes cell proliferation and represses myogenic differentiation by regulating foxo4 and mef2 factors. *Journal of Cell Science*. https://doi.org/10.1242/jcs.105239

Shi, Y., Xu, X., Zhang, Q., Fu, G., Mo, Z., Wang, G. S., Kishi, S., & Yang, X.-L. (2014). Trna synthetase counteracts c-myc to develop functional vasculature. *eLife*, *3*. https://doi.org/10.7554/eLife.02349

Siddique, A. N., Nunna, S., Rajavelu, A., Zhang, Y., Jurkowska, R. Z., Reinhardt, R., Rots, M. G., Ragozin, S., Jurkowski, T. P., & Jeltsch, A. (2013). Targeted methylation and gene silencing of vegf-a in human cells by using a designed dnmt3a–dnmt3l single-chain fusion protein with increased dna methylation activity. *Journal of Molecular Biology*, *425*, 479–491. https://doi.org/10.1016/j.jmb.2012.11.038

Silberg, D. G., Swain, G. P., Suh, E. R., & Traber, P. G. (2000). Cdx1 and cdx2 expression during intestinal development. *Gastroenterology*, *119*, 961–971. https://doi.org/10.1053/gast.2000.18142

Simcha, I., Geiger, B., Yehuda-Levenberg, S., Salomon, D., & Ben-Ze'ev, A. (1996). Suppression of tumorigenicity by plakoglobin: An augmenting effect of n-cadherin. *The Journal of cell biology*, *133*, 199–209. https://doi.org/10.1083/jcb.133.1.199

Singh, H., Khan, A. A., & Dinner, A. R. (2014). Gene regulatory networks in the immune system. *Trends in Immunology*, *35*, 211–218. https://doi.org/10.1016/j.it.2014.03.006

Singh, H., Das, C. K., Buchmuller, B. C., Schäfer, L. V., Summerer, D., & Linser, R. (2023). Epigenetic cpg duplex marks probed by an evolved dna reader via a well-tempered conformational plasticity. *Nucleic Acids Research*, *51*, 6495–6506. https://doi.org/10.1093/nar/gkad134

Singh, S., Broeck, A. V., Miller, L., Chaker-Margot, M., & Klinge, S. (2021). Nucleolar maturation of the human small subunit processome. *Science*, *373*. https://doi.org/10.1126/science.abj5338

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of dna binding sites. *eLife*, *6*. https://doi.org/10.7554/eLife.21856

Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Andrews, R., & Bird, A. P. (2010). Neuronal mecp2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Molecular Cell*, *37*, 457–468. https://doi.org/10.1016/j.molcel.2010.01.030

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H. J., & Mann, R. S. (2011). Cofactor binding evokes latent differences in dna binding specificity between hox proteins. *Cell*, *147*, 1270–1282. https://doi.org/10.1016/j.cell.2011.10.053

Smith, S., & Stillman, B. (1991). Immunological characterization of chromatin assembly factor i, a human cell factor required for chromatin assembly during dna replication in vitro. *The Journal of biological chemistry*, *266*, 12041–7.

Smith, Z. D., & Meissner, A. (2013). Dna methylation: Roles in mammalian development. *Nature Reviews Genetics*, *14*, 204–220. https://doi.org/10.1038/nrg3354

Song, C.-X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., Gao, J., Liu, P., Li, L., Xu, G.-L., Jin, P., & He, C. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, *153*, 678–691. https://doi.org/10.1016/j.cell.2013.04.001

Song, C.-X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T. J., Zhang, B., Godley, L. A., Hicks, L. M., Lahn, B. T., Jin, P., & He, C. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*, *29*, 68–72. https://doi.org/10.1038/nbt.1732

Song, C.-X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., Liu, B., Xiong, J., Zhang, W., Hu, J., Zhou, Z., Dong, B., Tian, Z., Jeffrey, S. S., Chua, M.-S., So, S., Li, W., Wei, Y., Diao, J., … Quake, S. R. (2017). 5-hydroxymethylcytosine signatures in cell-free dna provide information about tumor types and stages. *Cell Research*, *27*, 1231–1242. https://doi.org/10.1038/cr.2017.106

Song, G., Wang, G., Luo, X., Cheng, Y., Song, Q., Wan, J., Moore, C., Song, H., Jin, P., Qian, J., & Zhu, H. (2021). An all-to-all approach to the identification of sequence-specific readers for epigenetic dna modifications on cytosine. *Nature Communications*, *12*, 795. https://doi.org/10.1038/s41467-021-20950-w

Song, J., Teplova, M., Ishibe-Murakami, S., & Patel, D. J. (2012). Structure-based mechanistic insights into dnmt1-mediated maintenance dna methylation. *Science*, *335*, 709–712. https://doi.org/10.1126/science.1214453

Song, X., Sun, Y., & Garen, A. (2005). Roles of psf protein and vl30 rna in reversible gene regulation. *Proceedings of the National Academy of Sciences*, *102*, 12189–12193. https://doi.org/10.1073/pnas.0505179102

Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial dna motifs on nucleosomes to initiate reprogramming. *Cell*, *161*, 555–568. https://doi.org/10.1016/j.cell.2015.03.017

Sowers, L. C., Shaw, B. R., & Sedwick, W. (1987). Base stacking and molecular polarizability: Effect of a methyl group in the 5-position of pyrimidines. *Biochemical and Biophysical Research Communications*, *148*, 790–794. https://doi.org/10.1016/0006-291X(87)90945-4

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., Ingen, H. V., Leonhardt, H., … Vermeulen, M. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, *152*, 1146–1159. https://doi.org/10.1016/j.cell.2013.02.004

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schübeler, D. (2011). Dna-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, *480*, 490–495. https://doi.org/10.1038/nature10716

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schübeler, D. (2012). Correction: Corrigendum: Dna-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, *484*, 550–550. https://doi.org/10.1038/nature11086

Steimle, V., Durand, B., Barras, E., Zufferey, M., Hadam, M. R., Mach, B., & Reith, W. (1995). A novel dna-binding regulatory factor is mutated in primary mhc class ii deficiency (bare lymphocyte syndrome). *Genes and Development*, *9*, 1021–1032. https://doi.org/10.1101/gad.9.9.1021

Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., & Suske, G. (2018). Mga, l3mbtl2 and e2f6 determine genomic binding of the non-canonical polycomb repressive complex prc1.6. *PLOS Genetics*, *14*, e1007193. https://doi.org/10.1371/journal.pgen.1007193

Stielow, B., Zhou, Y., Cao, Y., Simon, C., Pogoda, H.-M., Jiang, J., Ren, Y., Phanor, S. K., Rohner, I., Nist, A., Stiewe, T., Hammerschmidt, M., Shi, Y., Bulyk, M. L., Wang, Z., & Liefke, R. (2021). The sam domain-containing protein 1 (samd1) acts as a repressive chromatin regulator at unmethylated cpg islands. *Science Advances*, *7*. https://doi.org/10.1126/sciadv.abf2229

Stroud, H., Feng, S., Kinney, S. M., Pradhan, S., & Jacobsen, S. E. (2011). 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biology*, *12*, R54. https://doi.org/10.1186/gb-2011-12-6-r54

Su, M., Kirchner, A., Stazzoni, S., Müller, M., Wagner, M., Schröder, A., & Carell, T. (2016). 5-formylcytosine could be a semipermanent base in specific genome sites. *Angewandte Chemie International Edition*, *55*, 11797–11800. https://doi.org/10.1002/anie.201605994

Su, P.-H., Lai, H.-C., Huang, R.-L., Chen, L.-Y., Wang, Y.-C., Wu, T.-I., Chan, M. W. Y., Liao, C.-C., Chen, C.-W., Lin, W.-Y., & Chang, C.-C. (2019). Paired box-1 (pax1) activates multiple phosphatases and inhibits kinase cascades in cervical cancer. *Scientific Reports*, *9*, 9195. https://doi.org/10.1038/s41598-019-45477-5

Suarez-Alvarez, B., Rodriguez, R. M., Fraga, M. F., & López-Larrea, C. (2012). Dna methylation: A promising landscape for immune system-related diseases. *Trends in Genetics*, *28*, 506–514. https://doi.org/10.1016/j.tig.2012.06.005

Subramanian, S. K., Russ, W. P., & Ranganathan, R. (2018). A set of experimentally validated, mutually orthogonal primers for combinatorially specifying genetic components. *Synthetic Biology*, *3*. https://doi.org/10.1093/synbio/ysx008

Sugishita, H., Kondo, T., Ito, S., Nakayama, M., Yakushiji-Kaminatsui, N., Kawakami, E., Koseki, Y., Ohinata, Y., Sharif, J., Harachi, M., Blackledge, N. P., Klose, R. J., & Koseki, H. (2021). Variant pcgf1-prc1 links prc2 recruitment with differentiation-associated transcriptional inactivation at target genes. *Nature Communications*, *12*, 5341. https://doi.org/10.1038/s41467-021-24894-z

Sun, Z., Dai, N., Borgaro, J. G., Quimby, A., Sun, D., Corrêa, I. R., Zheng, Y., Zhu, Z., & Guan, S. (2015). A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. *Molecular Cell*, *57*, 750–761. https://doi.org/10.1016/j.molcel.2014.12.035

Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D., & Warren, S. T. (1992). Dna methylation represses fmr-1 transcription in fragile x syndrome. *Human Molecular Genetics*, *1*, 397–400. https://doi.org/10.1093/hmg/1.6.397

Swinstead, E. E., Paakinaho, V., Presman, D. M., & Hager, G. L. (2016). Pioneer factors and atp-dependent chromatin remodeling factors interact dynamically: A new perspective. *BioEssays*, *38*, 1150–1157. https://doi.org/10.1002/bies.201600137

Szulik, M. W., Pallan, P. S., Nocek, B., Voehler, M., Banerjee, S., Brooks, S., Joachimiak, A., Egli, M., Eichman, B. F., & Stone, M. P. (2015). Differential stabilities and sequence-dependent base pair opening dynamics of watson–crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. *Biochemistry*, *54*, 1294–1305. https://doi.org/10.1021/bi501534x

Szulwach, K. E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A. K., Gearing, M., Levey, A. I., Vasanthakumar, A., Godley, L. A., Chang, Q., Cheng, X., He, C., & Jin, P. (2011).

5-hmc–mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature Neuroscience*, *14*, 1607–1616. https://doi.org/10.1038/nn.2959

Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F., & Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic dna. *Nucleic Acids Research*, *38*, e181–e181. https://doi.org/10.1093/nar/gkq684

Szwergold, B. S., Howell, S., & Beisswenger, P. J. (2001). Human fructosamine-3-kinase. *Diabetes*, *50*, 2139–2147. https://doi.org/10.2337/diabetes.50.9.2139

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian dna by mll partner tet1. *Science*, *324*, 930–935. https://doi.org/10.1126/science.1170116

Tahirov, T. H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., Sato, K., Kumasaka, T., Yamamoto, M., Ishii, S., & Ogata, K. (2001). Structural analyses of dna recognition by the aml1/runx-1 runt domain and its allosteric control by cbfβ. *Cell*, *104*, 755–767. https://doi.org/10.1016/S0092-8674(01)00271-9

Takashima, K., Lee, D.-J., Trovero, M. F., Rothi, M. H., Mistry, M., Zhang, Y., Li, Z., Davis, C. P., Li, Z., Natale, J., Schmid, E., Haddad, J. A., Hoffmann, G. B., Dietmann, S., Sui, S. H., Oshiumi, H., Lieberman, J., & Greer, E. L. (2023, June). Nop16 is a histone mimetic that regulates histone h3k27 methylation and gene repression. https://doi.org/10.1101/2023.06.13.544862

Takata, M. A., Gonçalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., & Bieniasz, P. D. (2017). Cg dinucleotide suppression enables antiviral defence targeting non-self rna. *Nature*, *550*, 124–127. https://doi.org/10.1038/nature24039

Takayama, S., Xie, Z., & Reed, J. C. (1999). An evolutionarily conserved family of hsp70/hsc70 molecular chaperone regulators. *Journal of Biological Chemistry*, *274*, 781–786. https://doi.org/10.1074/jbc.274.2.781

Tamanaha, E., Guan, S., Marks, K., & Saleh, L. (2016). Distributive processing by the iron(ii)/α-ketoglutarate-dependent catalytic domains of the tet enzymes is consistent with epigenetic roles for oxidized 5-methylcytosine bases. *Journal of the American Chemical Society*, *138*, 9345–9348. https://doi.org/10.1021/jacs.6b03243

Tan, L., Xiong, L., Xu, W., Wu, F., Huang, N., Xu, Y., Kong, L., Zheng, L., Schwartz, L., Shi, Y., & Shi, Y. G. (2013). Genome-wide comparison of dna hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hmedip-seq method. *Nucleic Acids Research*, *41*, e84–e84. https://doi.org/10.1093/nar/gkt091

Tanaka, Y., Kurumizaka, H., & Yokoyama, S. (2005). Cpg methylation of the cenp-b box reduces human cenp-b binding. *The FEBS Journal*, *272*, 282–289. https://doi.org/10.1111/j.1432-1033.2004.04406.x

Tapias, A., Ciudad, C. J., Roninson, I. B., & Noé, V. (2008). Regulation of sp1 by cell cycle related proteins. *Cell Cycle*, *7*, 2856–2867. https://doi.org/10.4161/cc.7.18.6671

Tiranti, V., Rossi, E., Ruiz-Carrillo, A., Rossi, G., Rocchi, M., Didonato, S., Zuffardi, O., & Zeviani, M. (1995). Chromosomal localization of mitochondrial transcription factor a (tcf6), single-stranded dna-binding protein (ssbp), and endonuclease g (endog), three human housekeeping genes involved in mitochondrial biogenesis. *Genomics*, *25*, 559–564. https://doi.org/10.1016/0888-7543(95)80058-T

Tissir, F., & Goffinet, A. M. (2013). Shaping the nervous system: Role of the core planar cell polarity genes. *Nature Reviews Neuroscience*, *14*, 525–535. https://doi.org/10.1038/nrn3525

Torchy, M. P., Hamiche, A., & Klaholz, B. P. (2015). Structure and function insights into the nurd chromatin remodeling complex. *Cellular and Molecular Life Sciences*, *72*, 2491–2507. https://doi.org/10.1007/s00018-015-1880-8

Tyanova, S., & Cox, J. (2018). Perseus: A bioinformatics platform for integrative analysis of proteomics data in cancer research. https://doi.org/10.1007/978-1-4939-7493-1_7

Uehara, R., Nozawa, R.-s., Tomioka, A., Petry, S., Vale, R. D., Obuse, C., & Goshima, G. (2009). The augmin complex plays a critical role in spindle microtubule generation for mitotic progression and cytokinesis in human cells. *Proceedings of the National Academy of Sciences*, *106*, 6998–7003. https://doi.org/10.1073/pnas.0901587106

van Mierlo, G., & Vermeulen, M. (2021). Chromatin proteomics to study epigenetics — challenges and opportunities. *Molecular & Cellular Proteomics*, *20*, 100056. https://doi.org/10.1074/mcp.R120.002208

Vilain, A., Vogt, N., Dutrillaux, B., & Malfoy, B. (1999). Dna methylation and chromosome instability in breast cancer cell lines. *FEBS Letters*, *460*, 231–234. https://doi.org/10.1016/S0014-5793(99)01358-7

Volk, A., & Crispino, J. D. (2015). The role of the chromatin assembly complex (caf-1) and its p60 subunit (chaf1b) in homeostasis and disease. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1849*, 979–986. https://doi.org/10.1016/j.bbagrm.2015.05.009

Waddington, C. H. (1942). The epigenotype. *Endeavour*, *1*, 18–20.

Wang, D., Hashimoto, H., Zhang, X., Barwick, B. G., Lonial, S., Boise, L. H., Vertino, P. M., & Cheng, X. (2017). Max is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Research*, *45*, 2396–2407. https://doi.org/10.1093/nar/gkw1184

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone h2a ubiquitination in polycomb silencing. *Nature*, *431*, 873–878. https://doi.org/10.1038/nature02985

Wang, J., Catania, S., Wang, C., de la Cruz, M. J., Rao, B., Madhani, H. D., & Patel, D. J. (2022). Structural insights into dnmt5-mediated atp-dependent high-fidelity epigenome maintenance. *Molecular Cell*, *82*, 1186–1198.e6. https://doi.org/10.1016/j.molcel.2022.01.028

Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X.-D., & Wang, D. (2015). Molecular basis for 5-carboxycytosine recognition by rna polymerase ii elongation complex. *Nature*, *523*, 621–625. https://doi.org/10.1038/nature14482

Wang, L. J., Jin, H. C., Wang, X., Lam, E. K., Zhang, J. B., Liu, X., Chan, F. K., Si, J. M., & Sung, J. J. (2009). Zic1 is downregulated through promoter hypermethylation in gastric cancer. *Biochemical and Biophysical Research Communications*, *379*, 959–963. https://doi.org/10.1016/j.bbrc.2008.12.180

Wang, X., Zhu, H., Snieder, H., Su, S., Munn, D., Harshfield, G., Maria, B. L., Dong, Y., Treiber, F., Gutin, B., & Shi, H. (2010). Obesity related methylation changes in dna of peripheral blood leukocytes. *BMC Medicine*, *8*, 87. https://doi.org/10.1186/1741-7015-8-87

Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H., Jiao, B., Ren, R., Ye, D., Guan, K.-L., & Xiong, Y. (2015). Wt1 recruits tet2 to regulate its target gene expression and suppress leukemia cell proliferation. *Molecular Cell*, *57*, 662–673. https://doi.org/10.1016/j.molcel.2014.12.023

Wanunu, M., Cohen-Karni, D., Johnson, R. R., Fields, L., Benner, J., Peterman, N., Zheng, Y., Klein, M. L., & Drndic, M. (2011). Discrimination of methylcytosine from hydroxymethylcytosine in dna molecules. *Journal of the American Chemical Society*, *133*, 486–492. https://doi.org/10.1021/ja107836t

Wasserman, W. W., & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics*, *5*, 276–287. https://doi.org/10.1038/nrg1315

Weber, A. R., Krawczyk, C., Robertson, A. B., Kuśnierczyk, A., Vågbø, C. B., Schuermann, D., Klungland, A., & Schär, P. (2016). Biochemical reconstitution of tet1–tdg–ber-dependent active dna demethylation reveals a highly coordinated mechanism. *Nature Communications*, *7*, 10806. https://doi.org/10.1038/ncomms10806

Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., Yu, M., Liu, X., Zhu, P., Li, X., Hou, Y., Guo, H., Wu, X., He, C., Li, R., … Qiao, J. (2014). Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biology*, *15*, R49. https://doi.org/10.1186/gb-2014-15-3-r49

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M. D., & Owen-Hughes, T. (2003). Evidence for dna translocation by the iswi chromatin-remodeling enzyme. *Molecular and Cellular Biology*, *23*, 1935–1945. https://doi.org/10.1128/MCB.23.6.1935-1945.2003

Wierer, M., & Mann, M. (2016). Proteomics to study dna-bound and chromatin-associated gene regulatory complexes. *Human Molecular Genetics*, *25*, R106–R114. https://doi.org/10.1093/hmg/ddw208

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A. C., Rappsilber, J., & Helin, K. (2011). Tet1 and hydroxymethylcytosine in transcription and dna methylation fidelity. *Nature*, *473*, 343–348. https://doi.org/10.1038/nature10066

Wilson, A. S., Power, B. E., & Molloy, P. L. (2007). Dna hypomethylation and human diseases. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1775*, 138–162. https://doi.org/10.1016/j.bbcan.2006.08.007

Wong, S. J., Gearhart, M. D., Taylor, A. B., Nanyes, D. R., Ha, D. J., Robinson, A. K., Artigas, J. A., Lee, O. J., Demeler, B., Hart, P. J., Bardwell, V. J., & Kim, C. A. (2016). Kdm2b recruitment of the polycomb group complex, prc1.1, requires cooperation between pcgf1 and bcorl1. *Structure*, *24*, 1795–1801. https://doi.org/10.1016/j.str.2016.07.011

Wood, M. A., McMahon, S. B., & Cole, M. D. (2000). An atpase/helicase complex is an essential cofactor for oncogenic transformation by c-myc. *Molecular Cell*, *5*, 321–330. https://doi.org/10.1016/S1097-2765(00)80427-X

Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P., & Warren, W. D. (1997). Asymmetric methylation in the hypermethylated cpg promoter region of the human l1 retrotransposon. *Journal of Biological Chemistry*, *272*, 7810–7816. https://doi.org/10.1074/jbc.272.12.7810

Wu, F., Li, X., Looso, M., Liu, H., Ding, D., Günther, S., Kuenne, C., Liu, S., Weissmann, N., Boettger, T., Atzberger, A., Kolahian, S., Renz, H., Offermanns, S., Gärtner, U., Potente, M., Zhou, Y., Yuan, X., & Braun, T. (2023). Spurious transcription causing innate immune responses is prevented by 5-hydroxymethylcytosine. *Nature Genetics*, *55*, 100–111. https://doi.org/10.1038/s41588-022-01252-3

Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., & Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development*, *25*, 679–684. https://doi.org/10.1101/gad.2036011

Wu, H., Wu, X., Shen, L., & Zhang, Y. (2014). Single-base resolution analysis of active dna demethylation using methylase-assisted bisulfite sequencing. *Nature Biotechnology*, *32*, 1231–1240. https://doi.org/10.1038/nbt.3073

Wyatt, G. R. (1950). Occurrence of 5-methyl-cytosine in nucleic acids. *Nature*, *166*, 237–238. https://doi.org/10.1038/166237b0

Xie, C., Long, F., Li, L., Li, X., Ma, M., Lu, Z., Wu, R., Zhang, Y., Huang, L., Chou, J., Gong, N., Hu, G., & Lin, C. (2022). Ptbp3 modulates p53 expression and promotes colorectal cancer cell proliferation

by maintaining ube4a mrna stability. *Cell death & disease*, *13*, 128. https://doi.org/10.1038/s41419-022-04564-8

Xie, H., Wang, M., de Andrade, A., de F. Bonaldo, M., Galat, V., Arndt, K., Rajaram, V., Goldman, S., Tomita, T., & Soares, M. B. (2011). Genome-wide quantitative assessment of variation in dna methylation patterns. *Nucleic Acids Research*, *39*, 4099–4108. https://doi.org/10.1093/nar/gkr017

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu, G.-L., Wang, H., Chen, S., Gao, S., & Zhu, B. (2016). Cooperative action between sall4a and tet proteins in stepwise oxidation of 5-methylcytosine. *Molecular Cell*, *64*, 913–925. https://doi.org/10.1016/j.molcel.2016.10.013

Xiong, X., Chen, H., Zhang, Q., Liu, Y., & Xu, C. (2024). Uncovering the roles of dna hemi-methylation in transcriptional regulation using mspji-assisted hemi-methylation sequencing. *Nucleic Acids Research*, *52*, e24–e24. https://doi.org/10.1093/nar/gkae023

Xu, G.-L., Bestor, T. H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J., & Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a dna methyltransferase gene. *Nature*, *402*, 187–191. https://doi.org/10.1038/46052

Xu, Y., Pan, S., Song, Y., Pan, C., Chen, C., & Zhu, X. (2020). The prognostic value of the chromobox family in human ovarian cancer. *Journal of Cancer*, *11*, 5198–5209. https://doi.org/10.7150/jca.44475

Xu, Y., Sengupta, P. K., Seto, E., & Smith, B. D. (2006). Regulatory factor for x-box family proteins differentially interact with histone deacetylases to repress collagen α2(i) gene (col1a2) expression. *Journal of Biological Chemistry*, *281*, 9260–9270. https://doi.org/10.1074/jbc.M511724200

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A. J., Zheng, L., Zhang, H., Huang, S., Min, J., Nicholson, T., Chen, T., Xu, G., Shi, Y., Zhang, K., & Shi, Y. G. (2011). Genome-wide regulation of 5hmc, 5mc, and gene expression by tet1 hydroxylase in mouse embryonic stem cells. *Molecular Cell*, *42*, 451–464. https://doi.org/10.1016/j.molcel.2011.04.005

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., Diao, J., Wu, F., He, H. H., Cui, Q., Clark, E., Ma, C., Barbara, A., Veenstra, G. J. C., Xu, G., … Shi, Y. G. (2012). Tet3 cxxc domain and dioxygenase activity cooperatively regulate key genes for xenopus eye and neural development. *Cell*, *151*, 1200–1213. https://doi.org/10.1016/j.cell.2012.11.014

Yakhou, L., Azogui, A., Gupta, N., Albert, J. R., Miura, F., Ferry, L., Yamaguchi, K., Battault, S., Therizols, P., Bonhomme, F., Bethuel, E., Sarkar, A., Greenberg, M. V. C., Arimondo, P. B., Cristofari, G., Kirsh, O., Ito, T., & Defossez, P.-A. (2023). A genetic screen identifies bend3 as a regulator of bivalent gene expression and global dna methylation. *Nucleic acids research*, *51*, 10292–10308. https://doi.org/10.1093/nar/gkad719

Yamada, N., Nishida, Y., Tsutsumida, H., Hamada, T., Goto, M., Higashi, M., Nomoto, M., & Yonezawa, S. (2008). Muc1 expression is regulated by dna methylation and histone h3 lysine 9 modification in cancer cells. *Cancer Research*, *68*, 2708–2716. https://doi.org/10.1158/0008-5472.CAN-07-6844

Yamaguchi, K., Chen, X., Rodgers, B., Miura, F., Bashtrykov, P., Bonhomme, F., Salinas-Luypaert, C., Haxholli, D., Gutekunst, N., Aygenli, B. Ö., Ferry, L., Kirsh, O., Laisné, M., Scelfo, A., Ugur, E., Arimondo, P. B., Leonhardt, H., Kanemaki, M. T., Bartke, T., … Defossez, P.-A. (2024). Non-canonical functions of uhrf1 maintain dna methylation homeostasis in cancer cells. *Nature Communications*, *15*, 2960. https://doi.org/10.1038/s41467-024-47314-4

Yang, J., Horton, J. R., Li, J., Huang, Y., Zhang, X., Blumenthal, R. M., & Cheng, X. (2019). Structural basis for preferential binding of human tcf4 to dna containing 5-carboxylcytosine. *Nucleic Acids Research*, *47*, 8375–8387. https://doi.org/10.1093/nar/gkz381

Yang, J., Bashkenova, N., Zang, R., Huang, X., & Wang, J. (2020). The roles of tet family proteins in development and stem cells. *Development*, *147*. https://doi.org/10.1242/dev.183129

Yang, Y., Zeng, C., Lu, X., Song, Y., Nie, J., Ran, R., Zhang, Z., He, C., Zhang, W., & Liu, S.-M. (2019). 5-hydroxymethylcytosines in circulating cell-free dna reveal vascular complications of type 2 diabetes. *Clinical Chemistry*, *65*, 1414–1425. https://doi.org/10.1373/clinchem.2019.305508

Yildirim, O., Li, R., Hung, J.-H., Chen, P. B., Dong, X., Ee, L.-S., Weng, Z., Rando, O. J., & Fazzio, T. G. (2011). Mbd3/nurd complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, *147*, 1498–1510. https://doi.org/10.1016/j.cell.2011.11.054

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on dna binding specificities of human transcription factors. *Science*, *356*. https://doi.org/10.1126/science.aaj2239

Yoder, J. A., Soman, N. S., Verdine, G. L., & Bestor, T. H. (1997). Dna (cytosine-5)-methyltransferases in mouse cells and tissues. studies with a mechanism-based probe. *Journal of Molecular Biology*, *270*, 385–395. https://doi.org/10.1006/jmbi.1997.1125

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, *13*, 335–340. https://doi.org/10.1016/S0168-9525(97)01181-5

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., & Cleary, M. L. (2004). Leukemia proto-oncoprotein mll forms a set1-like histone methyltransferase complex with menin to regulate hox gene expression. *Molecular and Cellular Biology*, *24*, 5639–5649. https://doi.org/10.1128/MCB.24.13.5639-5649.2004

Yu, M., Hon, G. C., Szulwach, K. E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J.-H., Jin, P., Ren, B., & He, C. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, *149*, 1368–1380. https://doi.org/10.1016/j.cell.2012.04.027

Zeng, Y., Yao, B., Shin, J., Lin, L., Kim, N., Song, Q., Liu, S., Su, Y., Guo, J. U., Huang, L., Wan, J., Wu, H., Qian, J., Cheng, X., Zhu, H., Ming, G.-l., Jin, P., & Song, H. (2016). Lin28a binds active promoters and recruits tet1 to regulate gene expression. *Molecular Cell*, *61*, 153–160. https://doi.org/10.1016/j.molcel.2015.11.020

Zhan, X., Yan, C., Zhang, X., Lei, J., & Shi, Y. (2018). Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Research*, *28*, 1129–1140. https://doi.org/10.1038/s41422-018-0094-7

Zhang, J., Gao, Q., Li, P., Liu, X., Jia, Y., Wu, W., Li, J., Dong, S., Koseki, H., & Wong, J. (2011). S phase-dependent interaction with dnmt1 dictates the role of uhrf1 but not uhrf2 in dna methylation maintenance. *Cell Research*, *21*, 1723–1739. https://doi.org/10.1038/cr.2011.176

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.-L., Luo, C., Jiang, H., & He, C. (2012). Thymine dna glycosylase specifically recognizes 5-carboxylcytosine-modified dna. *Nature Chemical Biology*, *8*, 328–330. https://doi.org/10.1038/nchembio.914

Zhang, S., Yan, H., Li, C.-Z., Chen, Y.-H., Yuan, F.-h., Chen, Y.-g., Weng, S.-P., & He, J.-G. (2013). Identification and function of leucine-rich repeat flightless-i-interacting protein 2 (lrrfip2) in litopenaeus vannamei. *PLoS ONE*, *8*, e57456. https://doi.org/10.1371/journal.pone.0057456

Zhang, X., Williams, E. D., Azhar, G., Rogers, S. C., & Wei, J. Y. (2016). Does p49/strap, a srf-binding protein (srfbp1), modulate cardiac mitochondrial function in aging? *Experimental Gerontology*, *82*, 150–159. https://doi.org/10.1016/j.exger.2016.06.008

Zhang, X., Blumenthal, R. M., & Cheng, X. (2024). Dna-binding proteins from mbd through zf to ben: Recognition of cytosine methylation status by one arginine with two conformations. *Nucleic Acids Research*, *52*, 11442–11454. https://doi.org/10.1093/nar/gkae832

Zhang, Y., Wu, K., Shao, Y., Sui, F., Yang, Q., Shi, B., Hou, P., & Ji, M. (2016). Decreased 5-hydroxymethylcytosine (5-hmc) predicts poor prognosis in early-stage laryngeal squamous cell carcinoma. *American journal of cancer research*, *6*, 1089–98.

Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., & Reinberg, D. (1997). Histone deacetylases and sap18, a novel polypeptide, are components of a human sin3 complex. *Cell*, *89*, 357–364. https://doi.org/10.1016/S0092-8674(00)80216-0

Zhang, Z.-M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S. B., Wang, Y., Wang, G. G., & Song, J. (2018). Structural basis for dnmt3a-mediated de novo dna methylation. *Nature*, *554*, 387–391. https://doi.org/10.1038/nature25477

Zhao, J., Gu, T., Gao, C., Miao, G., Palma-Gudiel, H., Yu, L., Yang, J., Wang, Y., Li, Y., Lim, J., Li, R., Yao, B., Wu, H., Schneider, J. A., Seyfried, N., Grodstein, F., Jager, P. L. D., Jin, P., & Bennett, D. A. (2025). Brain 5-hydroxymethylcytosine alterations are associated with alzheimer's disease neuropathology. *Nature Communications*, *16*, 2842. https://doi.org/10.1038/s41467-025-58159-w

Zhao, L., Sun, M.-a., Li, Z., Bai, X., Yu, M., Wang, M., Liang, L., Shao, X., Arnovitz, S., Wang, Q., He, C., Lu, X., Chen, J., & Xie, H. (2014). The dynamics of dna methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Research*, *24*, 1296–1307. https://doi.org/10.1101/gr.163147.113

Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., & Xu, R.-M. (2014). Structural basis for hydroxymethylcytosine recognition by the sra domain of uhrf2. *Molecular Cell*, *54*, 879–886. https://doi.org/10.1016/j.molcel.2014.04.003

Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., & Yi, C. (2017). Single-cell 5-formylcytosine landscapes of mammalian early embryos and escs at single-base resolution. *Cell Stem Cell*, *20*, 720–731.e5. https://doi.org/10.1016/j.stem.2017.02.013

Zhu, H., Wang, G., & Qian, J. (2016). Transcription factors as readers and effectors of dna methylation. *Nature Reviews Genetics*, *17*, 551–565. https://doi.org/10.1038/nrg.2016.83

Zhurinsky, J., Shtutman, M., & Ben-Ze'ev, A. (2000). Plakoglobin and beta-catenin: Protein interactions, regulation and biological roles. *Journal of Cell Science*, *113*, 3127–3139. https://doi.org/10.1242/jcs.113.18.3127

Zilberman, D. (2008). The evolving functions of dna methylation. *Current Opinion in Plant Biology*, *11*, 554–559. https://doi.org/10.1016/j.pbi.2008.07.004

Zuo, Z., Roy, B., Chang, Y. K., Granas, D., & Stormo, G. D. (2017). Measuring quantitative effects of methylation on transcription factor–dna binding affinity. *Science Advances*, *3*. https://doi.org/10.1126/sciadv.aao1799

Eidesstattliche Versicherung (Affidavit)

Engelhard, Lena
Name, Vorname
(Surname, first name)

231201
Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Dortmund,
Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

System-Wide Investigation of Hydroxymethylcytosine Recognition and Distribution in the Mammalian Genome

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Dortmund,
Ort, Datum
(Place, date)

Unterschrift
(Signature)