Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty Of Mathematics, Engineering and Natural Sciences

of the

Ruprecht – Karls – University

Heidelberg

Presented by:
M.Sc. Elisa Kreibich

born in: Berlin, Germany
Oral examination: 23rd May 2023

Epigenetic regulation of enhancer activity in the mammalian genome

Referees:
Dr. James Hackett
Prof. Dr. Benedikt Brors

For the 10-year-old girl who dreamed
of becoming a scientist

Summary

Cell types are defined by their spatiotemporal gene expression patterns and their differential activity of promoters and enhancers. Enhancers are *cis*-regulatory elements in the DNA critical for the acquisition and maintenance of cellular identities by regulating the expression of key genes. Enhancers serve as landing pads for transcription factors (TFs) which are DNA-binding proteins that interpret the genomic code and enhance gene expression upon their binding. However, the underlying DNA sequence does not solely convey binding specificity, and therefore it is still largely elusive what additional factors regulate TF binding.

An important regulatory layer in gene expression are dynamic and reversible epigenetic modifications of chromatin including DNA and histone proteins. To date, dozens of histone modifications have been identified that are associated with different genomic contexts and transcriptional states. For instance, histone H3 lysine acetylation has been generally associated with active chromatin as active enhancers and promoters, while histone H3 tri-methylation at lysine 23 (H3K27me3) is coupled to transcription repression. Yet, the causal contribution of such histone modifications to the regulation of enhancer activity and TF binding is still large unknown.

To address this question, I developed a technical approach to analyse TF binding at DNA molecules where a certain histone modification of interest is present. For this, I combined a genomic enrichment technique with a single molecule footprinting (SMF) approach that allows to detect TF binding at single DNA molecule resolution. However, this experimental set-up paired with different optimization approaches did not produce high enough enrichments of DNA molecules harboring certain histone modifications to suffice the required statistical power. Therefore, the focus was laid on investigating the causal role of DNA methylation.

DNA methylation in CpG context is the most common epigenetic modification in the mammalian genome that covers 70-80% of all CpG dinucleotides. Despite its prevalence, DNA methylation can be highly dynamic, especially at enhancer elements that exhibit reduced methylation levels during their activation. Previous studies have identified that the binding of TFs to enhancers is correlated with the partial loss in DNA methylation and it has been suggested that DNA methylation regulates enhancer activity. This hypothesis has remained elusive up to date, which has multiple reasons. First, the relationship between TFs and DNA methylation is bidirectional. Previous studies have identified many methyl-sensitive TFs *in vitro* whose binding is reduced upon methylation of their DNA binding motif. Some of those have been confirmed by *in vivo* studies, which showed that DNA methylation prevents the spurious binding of those TFs in the genome. Opposingly, TFs have also been identified to be directly responsible for the demethylation of enhancers. In consequence, the bidirectional regulation between DNA methylation and TF binding has prevented the establishment of a causal relationship between them. Second, the cell-to-cell epigenetic variability observed as intermediate methylation at enhancers elements makes common bulk-cell genomics approaches ineffective to identify a direct correlation between DNA methylation and TF binding and to determine whether DNA methylation generally contributes to the regulation of enhancer activity.

In the here presented PhD project, I overcame these issues and limitation by advancing the single molecule footprinting (SMF) approach to resolve chromatin accessibility, TF binding, and simultaneously quantify the presence of DNA methylation on the same DNA molecules. By applying this technology across the murine genome, I demonstrate that TFs can bind most (>90%) enhancers irrespective of the underlying DNA methylation, suggesting that presence of DNA methylation does not generally impede enhancer activity. Yet, for stem cells and three somatic cell types, I identified active enhancers where TF occupancy is directly repressed by DNA methylation, including enhancers involved in the control of key cell identity genes. Using global perturbation assays and orthogonal enhancer activity measurements, I was able to show that at these active sites, DNA methylation directly controls the occupancy levels of TFs such as Max-Myc, that play a key role in the control of stem cell identity and proliferation. In the end, my data suggest a model where the function of DNA methylation extends beyond protecting the genome from spurious TF binding, by directly regulating the activation of cell-type specific enhancers.

This detailed analysis is an important addition to our general knowledge on gene regulation and suggest that while epigenetic factors may have largely redundant functions, their individual contributions can play important and instructive roles in tuning the quantitative expression of key cell-specific genes. Understanding the regulation of such genes involved in cell identity will have important implications in the comprehension of development and disease.

Zusammenfassung

Zelltypen werden durch ihre raum-zeitlichen Genexpressionsmuster und ihre unterschiedliche Aktivität von Promotoren und Enhancern definiert. Enhancer sind *cis*-regulierende DNA Elemente, die für den Erwerb und die Aufrechterhaltung der zellulären Identität entscheidend sind, indem sie die Expression von Schlüsselgenen regulieren. Enhancer dienen als Bindungsplattformen für Transkriptionsfaktoren (TFs). Bei diesen TFs handelt es sich um DNA-bindende Proteine, die den genomischen Code interpretieren und die Genexpression nach ihrer Bindung an regulatorische Elemente verstärken. Die zugrundeliegende DNA-Sequenz ist jedoch nicht allein für die Spezifität der Bindung verantwortlich. Daher ist es noch weitgehend unklar, welche zusätzlichen Faktoren die TF-Bindung regulieren.

Eine wichtige Regulierungsebene bei der Genexpression sind dynamische und reversible epigenetische Modifikationen des Chromatins, einschließlich der DNA und der Histonproteine.

Bislang wurden Dutzende von Histonmodifikationen identifiziert, die mit verschiedenen genomischen Kontexten und Transkriptionszuständen in Verbindung gebracht werden. So wird beispielsweise die Histon-H3-Lysinacetylierung allgemein mit aktivem Chromatin wie aktiven Enhancern und Promotoren in Verbindung gebracht, während die Histon-H3-Tri-Methylierung an Lysin 23 (H3K27me3) mit der Transkriptionsunterdrückung verbunden ist. Der kausale Beitrag dieser Histonmodifikationen zur Regulierung der Enhanceraktivität und TF-Bindung ist jedoch noch weitgehend unbekannt.

Um diese Kausalitätsfrage zu klären, habe ich einen technischen Ansatz entwickelt um die TF-Bindung an DNA-Molekülen zu analysieren an denen bestimmte Histonmodifikationen vorhanden sind. Dazu habe ich eine genomische Anreicherungsmethode mit einer Methode der Einzelmolekül-Genomik (Single Molecule Footprinting, SMF) kombiniert, mit dem die TF-Bindung auf einzelnen DNA-Molekülen nachgewiesen werden kann. Diese Versuchsanordnung in Verbindung mit verschiedenen Optimierungsansätzen führte jedoch nicht zu einer ausreichenden Anreicherung von DNA-Molekülen mit bestimmten Histonmodifikationen, um die erforderliche statistische Aussagekraft zu erreichen. Daher wurde der Schwerpunkt der Dissertation auf die Untersuchung der kausalen Rolle der DNA-Methylierung gelegt.

DNA-Methylierung im CpG-Kontext ist die häufigste epigenetische Modifikation im Genom von Säugetieren und findet sich auf 70-80% aller CpG-Dinukleotide. Trotz ihrer weiten Verbreitung kann die DNA-Methylierung sehr dynamisch sein, insbesondere an Enhancer-Elementen, die während ihrer Aktivierung reduzierte Methylierungswerte aufweisen. Frühere Studien haben gezeigt, dass die Bindung von TFs an Enhancer mit dem teilweisen Verlust der DNA-Methylierung korreliert, und es wurde vermutet, dass die DNA-Methylierung die Enhancer-Aktivität reguliert. Diese Hypothese ist bis heute nicht eindeutig geklärt, was mehrere Gründe hat. Erstens ist die Beziehung zwischen TFs und DNA-Methylierung bidirektional. In früheren Studien wurden viele methylierungsempfindliche TFs *in vitro* identifiziert, deren Bindung bei Methylierung ihres DNA-Bindungsmotivs reduziert wird. Einige von ihnen wurden durch *In-vivo*-Studien bestätigt, die zeigten, dass die DNA-Methylierung die

unerwünschte Bindung dieser TFs im Genom verhindert. Umgekehrt wurden auch TFs identifiziert, die direkt für die Demethylierung von Enhancern verantwortlich sind. Folglich hat die bidirektionale Regulierung zwischen DNA-Methylierung und TF-Bindung bisher verhindert, dass eine kausale Beziehung zwischen ihnen hergestellt werden konnte. Zweitens macht die epigenetische Variabilität von Zelle zu Zelle, die als intermediäre Methylierung an Enhancer-Elementen beobachtet wird, herkömmliche Genomikansätze für Zell-Populationen unwirksam, um eine direkte Korrelation zwischen DNA-Methylierung und TF-Bindung zu identifizieren und zu bestimmen, ob DNA-Methylierung generell zur Regulierung der Enhancer-Aktivität beiträgt.

In dem hier vorgestellten Dissertationsprojekt habe ich diese Probleme und Einschränkungen überwunden, indem ich den Single Molecule Footprinting (SMF)-Ansatz weiterentwickelt habe, um die Chromatin-Zugänglichkeit und die TF-Bindung zu bestimmen und gleichzeitig das Vorhandensein von DNA-Methylierung auf denselben DNA-Molekülen zu quantifizieren. Durch Anwendung dieser Technologie auf das Mausgenom konnte ich zeigen, dass TFs die meisten (>90%) Enhancer unabhängig von der zugrunde liegenden DNA-Methylierung binden können, was darauf hindeutet, dass das Vorhandensein von DNA-Methylierung die Aktivität von Enhancern nicht generell behindert. Für Stammzellen und drei somatische Zelltypen habe ich jedoch aktive Enhancer identifiziert, bei denen die TF-Bindung direkt durch DNA-Methylierung unterdrückt wird, darunter auch Enhancer, die an der Kontrolle wichtiger Zellidentitätsgene beteiligt sind. Mithilfe genomweiter Veränderungen der DNA-Methylierung und orthogonaler Messungen der Enhancer-Aktivität konnte ich zeigen, dass die DNA-Methylierung an diesen aktiven Stellen direkt die Binding von TFs wie Max-Myc kontrolliert, die eine Schlüsselrolle bei der Steuerung der Stammzellidentität und -vermehrung spielen. Letztlich legen meine Daten ein Modell nahe, bei dem die Funktion der DNA-Methylierung über den Schutz des Genoms vor ungewollter TF-Bindung hinausgeht, indem sie die Aktivierung zelltypspezifischer Enhancer direkt reguliert.

Diese detaillierte Analyse ist eine wichtige Ergänzung unseres allgemeinen Wissens über die Genregulierung und lässt vermuten, dass epigenetische Faktoren zwar weitgehend redundante Funktionen haben, ihre einzelnen Beiträge aber wichtige und instruktive Rollen bei der Abstimmung der quantitativen Expression wichtiger zellspezifischer Gene spielen können. Das Verständnis der Regulierung solcher Gene, die an der Zellidentität beteiligt sind, wird wichtige Auswirkungen auf das Verständnis von Entwicklung und Krankheit haben.

Acknowledgements

Doing a PhD is a journey. A wonderful, crazy, annoying, stressful and joyful journey filled with a million tiny moments of self-doubt, learning, laughter, tears, and gratitude. And it cannot be accomplished alone. Here, I want to thank all those people that accompanied me on this journey and that helped me to have the best time of my life.

My biggest acknowledgement goes – how could it be different – to Arnaud. You have been an incredible supervisor and mentor to me in the past years. You have created this amazing lab full of trust and honesty and joy and pretty awesome science, despite all the challenges of a global health crisis and all the other craziness in the world. I was very lucky when you accepted me to join your – back then still tiny – group and I think there would not have been a more perfect lab for me – and a more perfect mentor. You trusted in me and my abilities, when I did not. You were there in the high and the low moments. You pushed me, when I needed to overcome obstacles. And you stopped me and illuminated the way, when I got lost in the forest of analyses. You made sure that I grow, learn and think, and that I come out on the other side as a scientist. It was a pleasure working and writing all the reviews, the spotlight article and the paper with you. I will really miss this and our gossiping and joking over slack. And I will deeply miss the science. Having this unique viewpoint in the (epi)genomics field felt truly special. SMF will always occupy a corner of my heart. So, thank you for everything, Arnaud!

The second biggest acknowledgement goes to my lab mates – Roos, Guido, Valentina, Mathias, Laura, Michela, Kasit and Carlo – as well as the people that have been part of the lab in the past – Can, Alex, Christine and Blanka. I am sitting here on my couch with a drink in one hand and I realize that no words that I could use would truly grasp what I feel. You are just the most amazing bunch of people. And you have always been. No matter if we were four, five or nine people like now, I always felt that the lab is my safe space…the place where I belong and where I can be my true self. And I will miss you all so much. I will miss the interesting, deep and sometimes very weird and dark topics at lunch. I will miss all our beautiful moments in the bubble – the deep and philosophical discussions about science and life, the snack flashes, the endless amount of moments of pure joy and laughter, yet another discussion about strand collapsing, the satisfying sound of the bell of success and the following clapping, pink fluffy unicorns dancing on rainbows, the violent keyboard sounds of Guido coding, the French and Italian curses, the “Ohh god” of Mathias when he comes out of a meeting with Arnaud, Valentina’s “hm” sounds when she is on to something, Roos’ smiling face appearing in the doorframe in the mornings and evenings for the “Helloooo” and the “Have a good eveniiiing”, Laura’s belly tapping at lunch time, all our language comparisons, and a million other joyful moments. If Arnaud made me to a better scientist, you made me to a better person. I learned so much from all of you – personally and scientifically – and you truly had my back at any moment of this 4.5 year long journey. Without you, it would not have been the same. And if I could, I would do a second PhD in the lab, just to spend another four years with you. Thank you all for your honesty, trust, knowledge, laughter, joy, love and support. I am very grateful to have found in you my second family.

I also want to thank all the other great people at EMBL: The incredibly helpful people from all the facilities and especially GeneCore, GBCS and Sarah from the CSDA for their help and support. All members of the Genome Biology Unit and the Epigenetics Club for their input, help and their exciting science presentations. The people from the canteen and cafeteria that provided me with food and coffee throughout my PhD. My TAC members – Jamie Hackett, Jan Korbel and Frank Lyko – for their input and advice. My defense members – Benedikt Brors, Jamie Hackett, Eileen Furlong and Karsten Rippe – for reading my thesis and discussion my PhD project. And of course, my PhD batch! Knowing that 60 other people are going through the same journey with me, the knowing smiles and nods in the hallway and in the cafeteria, the parties, the coffee breaks, the bad movie nights, the Mima cakes…it all made this journey so much better. Special thanks go to: Aline – for all the waaaay to long and therefore to sparse coffee breaks, the walks and talks, and just for being my trust person since our brunch date at the end of the recruitment week. Agata – for all the lovely dinner evenings with awesome food, lots of wine and stories. The SOLA crew. The Brewing Club members. Lucia, Matteo, Max, Maya, Wolfi and Michelle – for their GB internal support and gossip. Clement and the other foodies – for all the awesome cooking and drinks nights. Amanda – for the pastries. And thanks to all the other awesome people that have crossed my path during those years.

Next to all the internal support, I am deeply grateful to have had wonderful friends and family outside of EMBL – near and far – that made these years joyful and happy, that supported me mentally, gave me perspective and believed in me, that made me forget my to-do-list and made me enjoy the moment. So, with lots of love a heartfelt thank you to: Janina and Lukas – for being the "Gäng" for many years now with long nights, amazing food, German pop-cultural references and splendid holidays. Iris and Saskia (plus Ioannis and Vincent) – for being my master lab girls, for all the couples and girls nights with deep conversation, board games and food. My "Mädelz" Dajana, Maschi and Finni – for being there for me for more than 20 years now and believing in me and even trying to understand what I am actually doing here. My family and especially my brothers, Enrico and Mirco, and Katrin, Viktoria, Manuela and my mom – thank you for supporting and believing in me, for accepting my stress-related low contact frequency, and for your faith that I can achieve my goals although I am falling a bit out of the family line with this scientific career choice.

The last but definitely not least huge, heartfelt and loving thank you goes to my partner and best friend Kai. Thank you for all the listening, helping, supporting, discussing, reading of my dissertation, hugging and giving perspective in the last years. Thank you for being there in the bad and the joyful moments, for having my back, believing in me, falling asleep alone while I am still writing and coding on the couch, for waking me up when I yet again fell asleep reading the journal club paper for the next day and for getting me out of the zone when necessary. And most importantly, thank you for making me laugh every single day and filling our home with love and joy.

Yes, doing a PhD is a journey. And the people that you are spending this journey with are maybe the most important part of it. I am very grateful to have had all of you with me, no matter if close or far, so thank you for being a part of my life!

Table of contents

List of publications

Kreibich, Elisa, Kleinendorst, Rozemarijn, Barzaghi, Guido, Kaspar, Sarah, and Krebs, Arnaud R. (2023). Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. Mol. Cell *83*, 787-802.e9. https://doi.org/10.1016/j.molcel.2023.01.017.

Kreibich, Elisa, and Krebs, Arnaud R. (2022). Cofactors: a new layer of specificity to enhancer regulation. Trends Biochem. Sci. *47*, 993–995. https://doi.org/10.1016/j.tibs.2022.07.008.

List of abbreviations

5hmC	DNA hydroxymethylation | 5-hydroxymethylcytosine
5mC	DNA methylation in CpG context | 5-methylcytosine
CA	chromatin accessibility
CGIs	CpG island
ChIP-seq	chromatin immunoprecipitation sequencing
CpG	cytosine followed by guanine base
CRE	*cis*-regulatory element
CUT&RUN	cleavage under targets and release using nuclease
DHS	DNase I hypersensitive site
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
H3K27ac	histone 3 lysine 27 acetylation
H3K4me(1/2/3)	histone 3 lysine 4 (mono-/di-/tri-) methylation
ICR	imprinting control region
mESC	mouse embryonic stem cell
MT	methyltransferase
RNA	ribonucleic acid
SMF	Single Molecule Footprinting
TET	ten-eleven translocation [enzyme]
TF	transcription factor
TFBS	transcription factor binding site
TKO	triple-knockout
TSS	transcription start site
WT	wild-type

1 | Introduction

Epigenetics – a term coined by Waddington in 1942 has, in its original sense, described ”the processes by which the genotype brings the phenotype into being“ (Waddington, 1942). The meaning and implications of this term have changed multiple times since then (Cavalli and Heard, 2019; Huang, 2022; Nicoglou and Merlin, 2017), resulting in its increased usage for different mechanisms. Its application ranges from (I) systems epigenetics – assuming a genome-wide memory and coordination in order to establish and maintain certain gene expression patterns that define the cellular state –, to (II) molecular epigenetics – assuming local *cis*-regulatory regulations by reversible DNA and histone modifications, higher-order 3D chromatin organization and other factors to establish and maintain local switches of gene expression that are not encoded in the DNA. In addition, epigenetics is sometimes defined as the memory of gene regulatory patterns over multiple cell cycles, and sometimes over multiple generations. Some definitions include the impact of the environmental stimuli on cellular states and expression patterns, others focus on the natural establishment and maintenance of cellular states.

This non-specific usage has led to a general confusion about the contexts in which the term is used (Huang, 2022). Therefore, I want to specify the definition of “epigenetics” that I have followed in this PhD dissertation. Here, epigenetics is meant as the local, reversible and dynamic modifications of histones and DNA, such as DNA methylation, that are not encoded in the DNA and are maintained through cell division. The epigenetic modifications contribute to the regulation of gene expression during maintenance and establishment of cellular states.

1.1 | The regulatory network of transcription

Deoxyribonucleic acid (DNA) contains the genetic instructions for the production of proteins and other molecules essential for living organisms. DNA is located in the nucleus where it is organized into chromatin. The basic unit of chromatin is a nucleosome, that consist of an octameric histone core and around 147 base pairs of DNA wrapped around it (Kornberg and Lorch, 1999). Depending on the density of nucleosomes, chromatin can be classified into highly compacted (repressed, gene-poor) heterochromatin and open (active, gene-rich) euchromatin.

Genes are segments of the DNA that encode proteins or non-coding RNAs. The mouse genome contains around 25,000 protein-coding genes and over 15,000 non-coding RNA genes (Blake et al., 2021), whose spatial and temporal expression is crucial for the establishment and maintenance of cellular identities. The regulation of those gene expression patterns is a complex mechanism involving different regulatory layers and numerous contributing factors.

Genes consist of several distinct regions, including the promoter that controls transcription initiation, the transcriptional start site (TSS) where transcription begins, and the gene body that is transcribed into RNA.

For transcription to happen, general transcription factors and the RNA polymerase II have to be sequentially recruited to the TSS and form the pre-initiation complex (PIC) (Sainsbury et al., 2015). Upon transcription activation, multiple processes are activated to ensure proper transcription elongation along the gene body, transcription re-initiation and mRNA splicing (Cramer, 2019; Haberle and Stark, 2018). Those processes include, among others, nucleosomal rearrangements to open up the chromatin, topoisomerase activity to release elongating RNA Polymerase II-induced DNA torsion (Pommier et al., 2016), and chromatin modifications along the gene body (Kim and Wysocka, 2023; Millán-Zambrano et al., 2022).

However, the basal transcriptional activity of a promoter is rather low, wherefore additional genomic elements are needed to enhance transcription and to determine the spatiotemporal and quantitative output. One important part of this gene regulatory network are *cis*-regulatory elements (CREs) that are either proximal to TSSs such as the aforementioned core promoters and promoter-proximal elements, or distal to TSSs such as enhancers, silencer and insulators. Those CREs are genomic regions with a length of a few hundred base pairs that serve as binding platforms for transcription factors (TFs). Their combinatorial activation is crucial to establish distinct expression patterns in different cell types and to enhance or repress transcription in response to environmental stimuli or during cell differentiation.

Enhancers are frequently located kilobases away from the genes they regulate with an average distance of 20-50 kb in vertebrates (Furlong and Levine, 2018). Although they have been shown to often activate the genes closest to them, they can also act on genes that are much further away and even skip neighboring genes (Lettice et al., 2002). In addition, their regulatory contribution seems to be irrespective of their direction (up- or downstream), orientation and distance to their associated gene (Banerji et al., 1981). Enhancers can also regulate multiple genes and, inversely, one promoter can be regulated by multiple enhancers (Andersson et al., 2014; Schoenfelder et al., 2015). Hence, gene regulation by enhancers is a complex mechanism and the criteria that determine which genes an enhancer activates as well as the interplay between multiple enhancers and other *cis*-regulatory modules to achieve the quantitative gene expression output are still largely elusive (Gasperini et al., 2020; Kim and Wysocka, 2023).

Enhancer elements contain clusters of binding sites, whose binding by cell type-specific TFs activates the them and leads to the recruitment of additional regulators such as nucleosome remodelers, chromatin modifiers, or architectural proteins and the Mediator complex (Reiter et al., 2017; Spitz and Furlong, 2012). TFs represent a prominent group of more than 1,600 proteins that harbor different DNA-binding domains that recognize small 6-12 bp-long DNA sequences as binding motifs (Lambert et al., 2018). TFs often have tissue-specific functions and are usually expressed in a cell-type-specific fashion (Lambert et al., 2018).
Since the motif itself provides low sequence specificity, additional features contribute to the specificity of TF occupancy such as DNA shape (e.g., helix twist, minor groove width) (Mathelier et al., 2016; Rohs et al., 2009), chromatin modifications (Isbel et al., 2022; Yin et al., 2017), and cooperativity between multiple TFs. This cooperativity can either be direct through protein-protein interactions as, for instance, shown for Max-Myc (Amoutzias et al., 2008; Jolma et al., 2015; Lüscher, 2001), or indirect through combinatorial action like simultaneous or competitive binding (Reiter et al., 2017; Spitz and Furlong, 2012). Indirect cooperative binding has been shown to be important for nucleosome competition, where multiple TFs are needed to displace nucleosomes to open and activate a CRE (Sönmezer et al., 2021), but also for cell-type-specific localization of TFs to different enhancers in different cell types as shown for the TF PU.1 in the

hematopoietic system (Barozzi et al., 2014; Jolma et al., 2015; Leddin et al., 2011; Spitz and Furlong, 2012). In other cases, so-called pioneer TFs open up the chromatin at CRE to allow additional binding of other TFs, mainly due to their ability to bind their DNA binding motif despite local nucleosome occupancy. Examples of those pioneer TFs are Sox2, whose binding in embryonic stem cells (ESCs) allows subsequent binding of Oct4 (Chen et al., 2014), and Oct4, Sox2, and Klf4 that function as pioneer factors during somatic to pluripotent reprogramming and enable Myc binding (Soufi et al., 2015). Upon binding, TFs cooperate with and recruit different classes of cofactors: the Mediator complex (Richter et al., 2022), nucleosome remodelers (Clapier et al., 2017), histone modifiers (Morgan and Shilatifard, 2020) and other scaffold or adapter proteins (Kim and Wysocka, 2023; Reiter et al., 2017; Spitz and Furlong, 2012). In the end, binding of TFs to CREs and the subsequent recruitment of cofactors leads to the establishment of enhancer-promoter contacts, the recruitment of the transcriptional machinery and ultimately to the initiation of transcription (Kim and Wysocka, 2023).

With the recruitment of cofactors, another layer of gene regulation is implemented: epigenetics. These include, e.g., the post-translational covalent modification of histone proteins at their globular core or their protruding tails as well as the modification of bases in the DNA. Also 3D chromatin organization and non-coding RNA-mediated regulation are part of epigenetics, but will not be discussed further in this dissertation. These epigenetic modifications have the potential to inhibit or enhance the recruitment and binding of TFs and the transcriptional machinery, the recruitment of other cofactors and even to directly impact chromatin accessibility (Calo and Wysocka, 2013; Kim and Wysocka, 2023; Reiter et al., 2017). In addition, chromatin modifications are important for the memory of the transcriptional state. Yet, although a large number of chromatin modifications has been identified, mapped in the genome and correlated with active or repressive regulatory elements, their direct and causal functions on gene regulation and their interplay with other regulators is still largely unknown.

In the end, multiple layers contribute to the complex regulatory network of transcription (Kim and Wysocka, 2023). First, TFs bind to their DNA sequence motifs and open up chromatin for other TFs to bind. This leads to the recruitment of multiple cofactors which initiate, amongst others, chromatin modifications and chromatin remodeling. Third, with the help of the recruited cofactors, distal CREs interact with each other and their promoters, forming a 3D network. And lastly, diverse CREs with different functions form a complex regulatory landscape. Importantly, those layers are not linear, but have multi-directional dependencies and influence each other. For examples, epigenetic modifications that are part of the second layer can also impact the binding of TFs, which are part of the first layer (Isbel et al., 2022).

In summary, gene transcription regulation is a multifaceted process that involves multiple layers and factors that contribute to a complex machinery to ensure the precise spatiotemporal gene expression patterns of individual cell lines. During decades of research, different layers have been discovered, gradually increasing the complexity of the system, and a lot of knowledge on the individual components has been gained. Understanding how these layers integrate, what their specifications and relationships are, and how they contribute to the final transcriptional output will be the next big endeavor of genome biology.

1.2 | Epigenetics

Epigenetics is an important regulatory layer of gene expression. It describes reversible modifications to the DNA or histones that can affect gene expression without changing the underlying DNA sequence. These epigenetic modifications can either enhance or repress gene expression, depending on their type and location. They can, for example, alter chromatin accessibility to regulate the binding of TFs and the transcription machinery, and are important for the recruitment of additional regulators. Due to their stability, they serve as cellular memory to maintain the functional identity of a cell. In addition, they can persist through cell division, enabling cells to transmit their functional identity from one generation of cells to the next. This helps to maintain the functional identity of a tissue or organ. In some cases, those epigenetic modifications can even be maintained across multiple generations of an organism, a process called transgenerational epigenetic inheritance. Due to their regulatory function, epigenetic modifications play important roles in normal cellular processes such as development, differentiation, and aging, as well as in the development of diseases including cancer and neurodegenerative disorders.

1.2.1 | Histone modifications

One part of epigenetics is the covalent post-translational modification of histones at their globular core or their protruding tails. These modifications can have direct or indirect effects on chromatin accessibility, or serve as recruitment and binding platforms for other regulatory factors (Millán-Zambrano et al., 2022). There are several types of histone modifications, including acetylation (ac), methylation (me), phosphorylation, ubiquitination, and SUMOylation, which can have different effects on the structure and function of chromatin depending on their type, position and quantity. Plenty of proteins have been identified that catalyze the addition or removal of histone modifications or that recognize them through specific domains. Most of these proteins are part of large regulatory complexes that harbor different catalytic and binding properties. The redundancies and specificities of the histone modifying enzymes and the larger protein complexes are an important topic of current research (Kreibich and Krebs, 2022).
For many histone modifications, their presence at specific genomic locations is positively or negatively associated with transcriptional activity (Millán-Zambrano et al., 2022), yet it is still largely elusive whether the modifications exhibit causal roles or whether their presence is rather a consequence of transcriptional activity.

1.2.1.1 | Histone acetylation

A histone modification that is generally correlated with transcriptional activity is histone acetylation. It is enriched at active promoters and enhancers and other accessible chromatin regions and is catalyzed by histone acetyltransferases (HATs) and removed by histone deacetylases (HDACs) (Verdin and Ott, 2015). Many nuclear proteins and transcriptional cofactors contain bromodomains that

recognize acetylated lysine residues. Those proteins include HATs themselves (p300, CREB-binding protein [CBP], general control non-depressible 5 [GCN5]), ATP-dependent remodelers, general transcription factors (e.g. TFIID component TATA-Box Binding Protein Associated Factor 1 [TAF1]), factors regulating transcriptional pause release (e.g. Bromodomain-containing protein 4 [BRD4]) and other chromatin modifiers (e.g. the histone methyltransferase Mixed Lineage Leukemia protein [MLL]) (Filippakopoulos and Knapp, 2012).

Two prominent HATs that play crucial roles in transcription regulation at enhancers are p300 and cyclic AMP response element-binding protein (CBP), which have been shown to be either recruited by TFs or bind to pre-existing acetylated lysine residues with their bromodomain (Millán-Zambrano et al., 2022). The main substrate of those two HATs is the lysine 27 residue of the histone 3 (H3K27), whose acetylation is used to distinguish active CREs from poised/primed or inactive ones (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Zentner et al., 2011). Although H3K27ac is the most prominent and widely-studied acetylation mark, also other histone tail residues have been found to be acetylated. As such, H3K9ac as well as H3K14ac, catalyzed by the HATs general control GCN5 and p300/CBP-associated factor (PCAF), are also found at enhancers and active promoters, and have been shown to be positively correlated with active genes and transcription (Gates et al., 2017; Karmodiya et al., 2012). Notably, despite the enrichment of acetylation at active CREs, studies have shown that the H3K27ac mark, for example, is dispensable for enhancer function (Zhang et al., 2020b). In contrast, a recent study showed that inhibition of HDACs in mouse ESCs (mESCs) resulted in an increased occupancy of certain tested TFs (GABPA, MAX and YY1), either in TSS-distal regions (MAX), TSS-proximal regions (GABPA) or transposable elements (YY1) (Cusack et al., 2020). This provides evidence, that histone modifications may have a direct impact at TF binding and enhancer activity.

1.2.1.2 | Histone methylation

Another prominent histone modification is methylation. In contrast to acetylation, the study of histone methylation is complicated by the fact that multiple methylation groups can be added to one residue, resulting in mono- (me1), di- (me2) or even trimethylation (me3) which can have different genomic localizations and functions. Histone methylation is catalyzed by histone methyltransferases (HMTs) and removed by histone demethylases, both of which are part of large cofactor complexes.

Most active promoters in eukaryotes are marked by H3K4me3, which is enriched at the TSS and strongly correlates with transcriptional activity (Benayoun et al., 2014; Chen et al., 2015; Talbert et al., 2019). Interestingly, a wide collection of studies provide evidence that presence of H3K4me3 is rather a result of transcriptional activity than instructing it (Howe et al., 2017; Wang et al., 2022), while others argue for a dual-feedback relationship where transcription promotes deposition of H3K4me3, which in reverse can also promote transcription (Hu et al., 2017; Policarpi et al., 2022). A very recent study shows that H3K4me3 is primarily important for RNA Polymerase II pausing-release and elongation (Wang et al., 2023). Hence, the specific role of H3K4me3 in gene regulation is not yet fully resolved. However, it has been identified as an important factor for epigenetic memory, since it is maintained through transcriptionally quiescent states (Millán-Zambrano et al., 2022).

In contrast to H3K4me3, its mono-methylated version, H3K4me1, is enriched at active enhancers and is mainly deposited by MLL3 and MLL4. While the aforementioned H3K27ac mark is only present at active enhancers, H3K4me1 is also enriched at some inactive enhancers like developmental and differentiation-associated genes. Hence, it is hypothesized that the premarking by H3K4me1 primes enhancers for potential future activities (Calo and Wysocka, 2013). In line with this, evidences are accumulating also for H3K4me1 that the mark itself has little impact on enhancer activity (Dorighi et al., 2017; Rickels et al., 2017).

While H3K4 methylation is generally associated with CREs, H3K27me3 is associated with silenced heterochromatin and repression of enhancers and promoters (Figure 2A and B). It is deposited by the catalytic component of the Polycomb Repressive Complex 2 (PRC2) (Cao et al., 2002; Müller et al., 2002) and bound by PRC1 leading to a subsequent silencing cascade (Guo et al., 2021). In addition, H3K27me3 plays a role in non-canonical genomic imprinting (Inoue et al., 2017a) and serves as epigenetic memory of silenced regions across cell divisions (Escobar et al., 2019). Also for H3K27me3 it has been shown that its deposition can be up- or downstream of transcriptional regulation, either causing transcriptional silencing or being a consequence of it (Millán-Zambrano et al., 2022).

Another histone modification that is correlated with transcriptional repression is H3K9me3 (Figure 2C-E). It is mostly found at transcriptionally silent genomic region, thus being a hallmark of constitutive heterochromatin (Allshire and Madhani, 2018). The repressive function of H3K9me3 is in part driven by heterochromatin protein 1 (HP1), which binds the mark and leads to compaction and repression by different mechanisms including self-oligomerization and recruitment of other chromatin-modifiers (Allshire and Madhani, 2018; Kumar and Kono, 2020; Schoelz and Riddle, 2022). A recent study in Caenorhabditis elegans provides evidence that H3K9me can block binding of certain TFs and the activity of enhancers, as well as promoters of tissue-specific genes, suggesting a role of H3K9me in distinct expression programs during development or differentiation (Figure 2E) (Methot et al., 2021).

Other histone methylation modifications include H3K9me2, which is a hallmark of X chromosome inactivation (Escamilla-Del-Arenal et al., 2013; Rougeulle Claire et al., 2004). Moreover, H3K36me3 marks the gene body of actively transcribed genes and is deposited by the HMT SET domain-containing 2 (SETD2), which is recruited by the phosphorylated *cis*-terminal domain of RNA Polymerase II (Kizer et al., 2005; Sharda and Humphrey, 2022; Sun et al., 2020).

1.2.1.3 | The oversimplifying histone code

By now, dozens of histone modifications have been identified at the tails and cores of histones (Lawrence et al., 2016; Millán-Zambrano et al., 2022) and the repertoire is likely to be expanded even more. To increase the complexity of this, variable combinations of histone modifications can be found at certain genomic regions. This has led to the histone code hypothesis, proposed by Strahl and Allis almost 25 years ago (Strahl and Allis, 2000), suggesting that histone modifications act in specific combinations to drive certain regulatory functions. This hypothesis, however, is oversimplifying the complexity of the system and does not take the dynamic changes and the functions of histone

modifications outside of gene regulation into account. Importantly, histone modifications stand also in relation to other epigenetic factors like DNA methylation (Figure 2), which is increasing the complexity even more. It has been shown that certain histone modification, e.g. H3K36me3 at gene bodies, recruit DNA methyltransferases resulting in methylation of the genomic region (Figure 2F) (Baubec et al., 2015; Rondelet et al., 2016). Other histone modifications inhibit the binding of DNA methyltransferases or recruit enzymes that remove DNA methylation.

In addition, one major point that is missing in this histone code hypothesis and that has not been unraveled yet in the field, is to what extent certain histone modifications play a causal role in gene regulation. As mentioned before, it has been shown for some individual examples already that the histone mark itself does not impact gene regulation. Hence, it needs to be disentangled whether and under which circumstances histone modifications themselves have regulatory roles or whether they are just consequences of the recruitment of large regulatory complexes with different functions, serve as memory markers or simply increase the redundancy of the system. Another aspect that is still elusive is the directionality of the relationship between the histone modifications and other gene regulatory factors. For instance, it is unclear whether the binding of certain TFs to CREs promotes the modification of histones or whether certain histone modifications regulate the binding of TFs.

1.2.2 | DNA methylation

DNA methylation (5mC) is one of the most studied epigenetic modifications in mammals that is defined by the covalent addition of a methyl-group to the fifth carbon of the cytosine base in CpG context. 5mC is generally associated with gene repression, especially in the context of genomic imprinting, repetitive elements as transposons, and germline-specific genes (Ferguson-Smith and Bourc'his, 2018; Greenberg and Bourc'his, 2019; Iurlaro et al., 2017; Mattei et al., 2022). While the majority of CpGs is methylated in most embryonic and somatic cell types, the levels of DNA methylation can differ and be highly dynamic in certain genomic contexts (Ehrlich et al., 1982; Lister et al., 2009; Meissner et al., 2008; Stadler et al., 2011; Ziller et al., 2013). During development, global DNA methylation levels undergo massive changes, whose disruption has detrimental impacts on embryo viability (Greenberg, 2021; Greenberg and Bourc'his, 2019; Iurlaro et al., 2017). In addition, DNA methylation is associated with a large number of diseases, especially cancer, where it is used as an epigenetic marker of tumor progression (Cavalli and Heard, 2019; Greenberg and Bourc'his, 2019; Nishiyama and Nakanishi, 2021).

Nevertheless, despite the long history of DNA methylation research (Mattei et al., 2022) and the enormous amount of insights we have gained in almost a century, its direct causal role in gene regulation in certain genomic contexts and the mechanistic insights into its different repressive functions are still poorly understood. These outstanding questions are an important consideration for many areas of research, ranging from the fundamental understanding of the interconnections in (epi)genetic gene regulation, embryonic development, aging and disease detection and treatment.

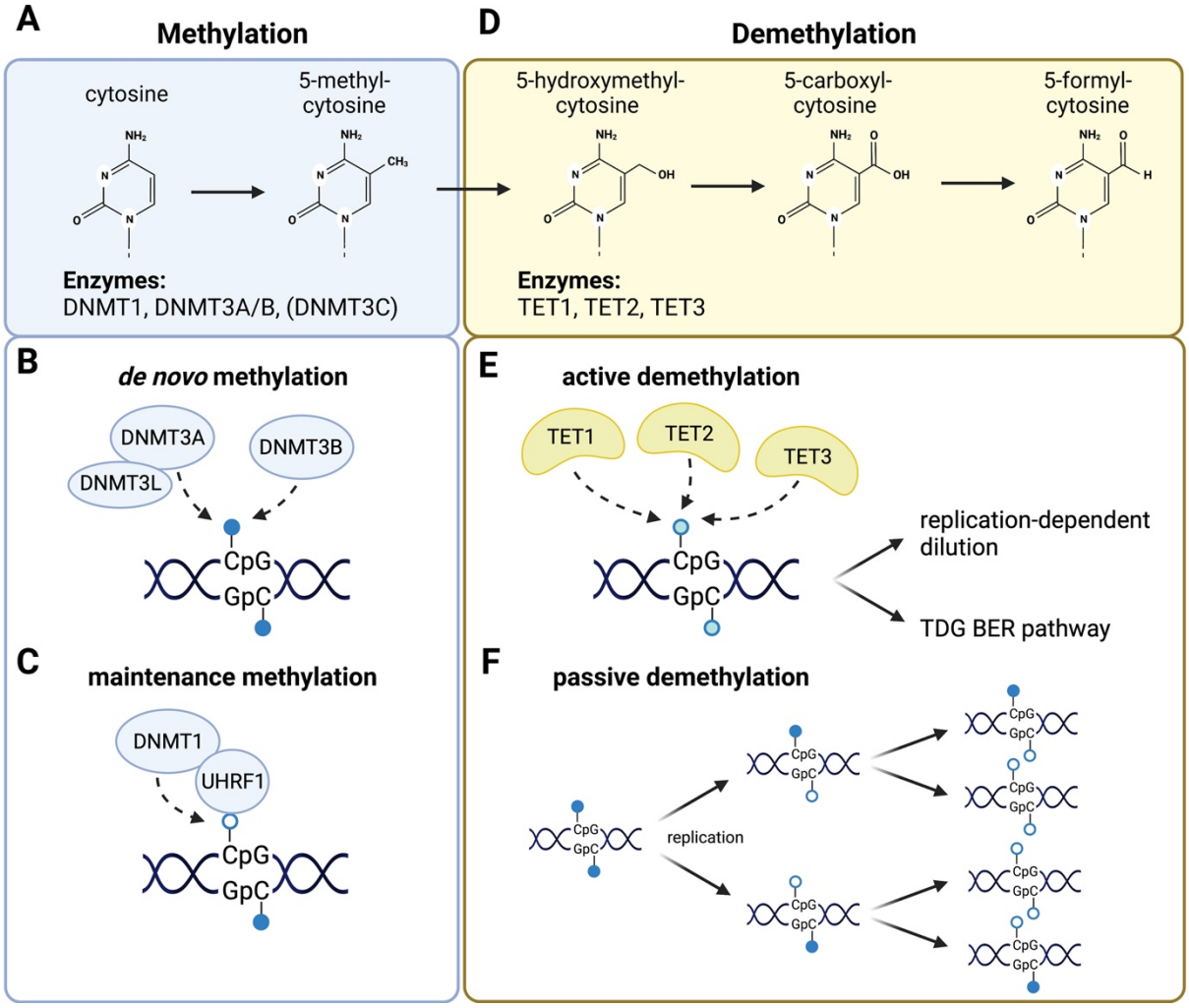

Figure 1: DNA methylation - Its deposition and removal. (A-C) Methylation of the 5th carbon of the cytosine base to 5-methylcytosine is catalyzed by DNA methyltransferases (DNMTs). **(B)** Fully unmethylated sites are methylated by the *de novo* MTs DNMT3A, together with its cofactor DNMT3L, and DNMT3B. **(C)** Hemimethylated sites (e.g. after replication) are recognized and methylated by the maintenance MT DNMT1, together with its cofactor UHRF1. **(D-F)** DNA demethylation can happen through a passive or active mechanism. **(D-E)** In the active mechanism, the methylation mark is successively oxidized in a multi-step process (D) by the enzymes TET1, TET2 and TET3 that are expressed in different cell types. The last step of the demethylation is either achieved by replication-dependent dilution after any of the oxidized cytosine version, or through the thymine DNA glycosylase-dependent base excision repair (TDG BER) pathway (E). **(F)** Passive demethylation happens by dilution of the non-maintained methyl-mark through multiple rounds of replication. During replication, the newly synthesized strand is not methylated. Figure created with BioRender.com

1.2.2.1 | Deposition and maintenance of DNA methylation

5mC is deposited by DNA methyltransferases (DNMTs) that exhibit different functions (Figure 1A). One DNMT family contains the *de novo* MTs DNMT3A and DNMT3B, that target unmethylated regions in the genome in an strongly redundant, ubiquitous and non-selective manner (Okano et al., 1998, 1999) (Figure 1B). In contrast stands DNMT1, which mediates the faithful 5mC maintenance, e.g. after DNA replication (Figure 1C) (Goll and Bestor, 2005; Hermann et al., 2004). For its function, DNMT1 interacts with its cofactor ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1), an E3 ubiquitin-protein ligase, which recognizes hemi-methylated DNA and recruits and activates DNMT1

(Bostick et al., 2007; Ishiyama et al., 2017; Sharif et al., 2007)(reviewed in (Xie and Qian, 2018)). This stimulates the full methylation of hemi-methylated sites where only one DNA strand is methylated (Figure 1C). UHRF1 is has been shown to be an essential cofactor of DNMT1, since its knockout leads to a strong global loss of DNA methylation in vivo (Bostick et al., 2007; Sharif et al., 2007). A recent study suggests that DNMT1 together with UHRF1 may also have *de novo* methylation functionality, at least in the context of transposable elements (Haggerty et al., 2021).
Recently, another family member, DNMT3C, has been found in rodents, where it has been shown to be important for the establishment of 5mC in germ cells and in the repression of retrotransposons (Barau et al., 2016; Jain et al., 2017; Sanchez-Delgado et al., 2016).

The activity and recruitment of DNMTs depends on various cofactors. While UHRF1 is an important cofactor for DNMT1, DNMT3L has been identified as a cofactor for DNMT3A/B. DNMT3L has no catalytic methyltransferase domain itself, but is e.g. important for DNMT3A/B's enzymatic activity in the germline by the formation of a heterotetrameric complex that increases the affinity to DNA (Bourc'his et al., 2001; Jia et al., 2007; Ooi et al., 2007). In addition, chromatin remodeling enzymes play a role in the establishment and maintenance of 5mC patterns (Dennis et al., 2001; Gibbons et al., 2000). The recruitment or blocking of DNMTs is also often mediated by histone modifications or histone modifying enzymes (Cedar and Bergman, 2009; Fuks, 2005; Smith and Meissner, 2013). For example, H3K36me3 recruits DNMT3A/B to transcribed gene bodies (Baubec et al., 2015; Rondelet et al., 2016), while, contrastingly, H3K4 methylation inhibits DNMT3 recruitment and binding (Ooi et al., 2007; Otani et al., 2009). Moreover, also TF and nucleosome occupancy have been shown to block DNMT3A/B's activity at regulatory elements (Ginno et al., 2020), while other TFs have been associated with the specific recruitment of DNMTs (Hervouet et al., 2018).

The loss of DNMTs or of their catalytic activity strongly affects the integrity of a cell and is detrimental during mammalian development. The loss of DNMT1 in mouse embryos leads to a global decrease in DNA methylation, which results in a drastic upregulation of transposable elements (Walsh et al., 1998) and is lethal at early stages of development (Lei et al., 1996; Li et al., 1992). Generally, DNMT1-deficiency is lethal in all dividing somatic cells (Fan et al., 2001; Jackson-Grusby et al., 2001; Li et al., 1992; Sen et al., 2010; Trowbridge et al., 2009). Mice with single knockouts of DNMT3A die a couple of weeks after birth, while DNMT3B-deficiency is lethal during embryogenesis (Okano et al., 1999). Consequently, genetic deletion of both *de novo* MTs causes early embryonic lethality (Okano et al., 1999). Interestingly, mESCs are viable upon loss of DNMTs despite the global DNA demethylation (Lei et al., 1996; Meissner et al., 2005; Okano et al., 1999; Tsumura et al., 2006). However, they are unable to differentiate *in vitro* and die upon induction (Lei et al., 1996; Li et al., 1992; Sakaue et al., 2010). This could partly be due to their inability to silence pluripotency genes such as Oct4 and Nanog (Feldman et al., 2006; Li et al., 2007) and the derepression of transposable elements (Kaluscha et al., 2022).

1.2.2.2 | Erasure of DNA methylation

The removal of DNA methylation can be accomplished through an active or passive mechanism (Figure 1D-F). In the passive mechanism, the 5mC mark is progressively lost by incomplete maintenance after replication (Kagiwada et al., 2013). While after the first cell cycle, the methylation is still present at one DNA strand, the newly synthesized strand is unmethylated. After a second replication cycle, already half of the daughter cells will have no DNA methylation anymore at this genomic location. Hence, without maintenance, the bulk methylation level is gradually reduced over multiple replication cycles (Figure 1F).

Active DNA demethylation involves the enzymatic turnover of the 5mC mark by the ten-eleven translocation (TET) dioxygenase family (Hackett et al., 2013; Ito et al., 2010; Tahiliani et al., 2009; Yamaguchi et al., 2013a). In mice, three TET enzymes (TET1, TET2, TET3) are expressed that recognize the 5mC mark and successively oxidize the methylated cytosine into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Figure 1D) (He et al., 2011; Ito et al., 2010, 2011; Pfaffeneder et al., 2011; Tahiliani et al., 2009). The final step of the demethylation process is either achieved through replication-dependent dilution or through the thymine DNA glycosylase-dependent base excision repair (TDG BER) pathway which leads to the replacement of the modified cytosine with an unmodified version (Figure 1E) (He et al., 2011; Kriaucionis and Heintz, 2009; Maiti and Drohat, 2011; Tahiliani et al., 2009; Weber et al., 2016).

While all TET enzymes exhibit catalytic function, only full-length TET1 and TET3 have a CXXC domain that increases their affinity to CpG-rich regions (Xu et al., 2011, 2012). This may e.g. facilitate TET1's localization to CpG islands, active and bivalent promoters (Williams et al., 2011; Wu et al., 2011; Xu et al., 2011). In contrast, the recruitment and binding of TET2 may be partially explained by its interaction partner the CXXC domain protein IDAX (also known as CXXC4) (Ko et al., 2013). In addition, the CXXC domain also facilities binding to unmethylated CpGs (Ko et al., 2013; Xu et al., 2011, 2012; Zhang et al., 2010a).

Yet, also other factors may play a role in the genomic distribution of TETs. Multiple studies have shown that TET proteins interact with key TFs. For instance, the pluripotency factor Nanog has been shown to interact with and recruit TET proteins (Costa et al., 2013), as well as PR domain zinc finger protein 14 (PRDM14) (Okashita et al., 2014), Polycomb repressive complex 2 (PRC2) (Neri et al., 2013), the RNA-binding protein LIN28A (Zeng et al., 2016), RE1-silencing transcription factor (REST) (Perera et al., 2015) and CTCF (Dubois-Chevalier et al., 2014). Additionally, it has been found that TET2 is recruited to enhancer elements in a highly cell-type-specific manner and colocalized with many TFBS (Rasmussen et al., 2019). These studies suggest that TET recruitment may depend, at least in part, on the interaction with other gene regulatory factors.

Interestingly, an increasing number of studies provide evidence that TET proteins exert additional functions in gene regulation independent of their catalytic activity. TET1, for example, associates with different histone modifying complexes such as the histone deacetylase complex SIN3A (Williams et al., 2011), the MOF histone acetyltransferase complex and even PRC2 (Neri et al., 2013; Wu et al., 2011). The same is true for TET2 which interacts with and recruits the histone deacetylase 2 (HDAC2) (Zhang et al., 2015). This suggests that TET proteins impact the chromatin environment independently of the removal of DNA methylation. In addition, studies using catalytically dead TET mutants showed that

they can confer biological functions and can rescue the phenotype of TET knockouts or knockdowns (Jin et al., 2014; Kaas et al., 2013; Montagner et al., 2016; Tsai et al., 2014; Zhang et al., 2015), strongly arguing for a catalytic-activity-independent function of TET enzymes in the genome.

In contrast to DNMTs, the loss of TET enzymes has more subtle effects, depending on the system tested. In pluripotent stem cells, single, double and triple KOs of the TET family members result in an increase in DNA methylation levels at some genomic loci and a strong reduction of global 5hmC levels (Ficz et al., 2011; Hon et al., 2014; Lu et al., 2014; Wu et al., 2011; Xu et al., 2011). Those DNA methylation increases and 5hmC decreases are, however, not universally distributed within the genome, but show certain patterns and different strengths. Enhancers, for example, show a stronger increase in DNA methylation (Hon et al., 2014), which is in agreement with the recent observation that enhancers exhibit high methylation turn-over that is mostly driven by TET enzymes (Ginno et al., 2020). Notably, these methylation changes have only limited effects on gene transcription (Hon et al., 2014) and maintenance of pluripotency (Dawlaty et al., 2011). In addition, several studies showed that the loss of TET enzymes in ESCs does not inhibit differentiation (Ficz et al., 2011; Hon et al., 2014; Koh et al., 2011) but rather delays the first steps, as it has been shown for *in vitro* methylation into neural progenitor cells (NPs) (Hon et al., 2014).
In germ cells, individual or double TET KOs lead to disrupted germ cell specification and function (Gu et al., 2011; Vincent et al., 2013; Yamaguchi et al., 2012) and disrupted reprogramming of imprinted genes (Hackett et al., 2013; Yamaguchi et al., 2013a).
In multipotent stem cells, loss of TET enzymes has been shown to be associated with disrupted differentiation, e.g. of hematopoietic stem cells (Izzo et al., 2020; Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011; Zhang et al., 2016; Zhao et al., 2015).
TET1/2 deficient cells can even contribute to live animals (Dawlaty et al., 2013). Yet, a recent study has examined the role of the TET machinery in gastrulation by developing chimeras with full or partial TET deficiency (Cheng et al., 2022). They identified that cells deficient for all three TET enzymes (TET triple-KO [TKO]) showed pervasive hypermethylation and mildly perturbed gene expression, disrupted cell signaling and skewed gastrulation. Nevertheless, in chimera context, TET TKO cells retained their potential for differentiation. This argues that neither TET1 nor TET2 are required for cell viability or mammalian development.

1.2.2.3 | Readers of DNA methylation

While DNMTs deposit and TET enzymes remove DNA methylation, other factors read methylated DNA and may contribute to its repressive function. This has led to the hypothesis of two mechanisms - an indirect and a direct one – on how DNA methylation exerts its functions.
In the direct mechanism, DNA methylation directly affects the binding of TFs, important DNA-binding proteins in gene regulation that bind to specific motifs in CREs and recruit the transcriptional machinery and cofactors. Multiple *in vitro* studies have shown that certain TFs are methyl-sensitive, thus are inhibited in binding when their motif is methylated (Héberlé and Bardet, 2019) (see chapter

1.3.1). In contrast, DNA methylation is usually absent (or reduced) at active CREs, arguing that DNA methylation and TF binding are mutually exclusive (Stadler et al., 2011). Yet, despite the large number of *in vitro* identified methyl-sensitive TFs, only a fraction of them has been confirmed *in vivo*. Recently, CREB1 has been shown to be sensitive to DNA methylation *in vivo* where methylation of transposons like intracisternal A-type particle elements inhibits CREB1's binding and, in consequence, the activation of those elements. One of the first methyl-sensitive TFs identified *in vivo* was NRF1 (Domcke et al., 2015). The deletion of all three major DNMTs in mESCs led to an increase in NRF1 binding sites that are usually methylated in wild type (WT) cells. This argues that DNA methylation inhibits the binding of certain TFs to spurious sites in the genome.

The indirect mechanism involves proteins that recognize methylated CpGs and recruit additional silencing factors such as HDACs and DNMTs to maintain methylation states (Bird and Wolffe, 1999; Du et al., 2015a; Nan et al., 1998; Ng et al., 1999). Those methyl-binding proteins can have divers functional domains with which they recognize DNA methylation and can be classified into three structural families: the CpG-methyl-binding domain (MBD) family, the zinc finger family and the SET and RING-finger associated (SRA) family (Buck-Koehntop and Defossez, 2013). Most of those proteins bind to methylated DNA independently of an underlying sequence motif (Buck-Koehntop and Defossez, 2013).

In mammalian cells, four proteins have been identified that harbor a functional MBD enabling them to bind methylated DNA *in vitro* and *in vivo*: MBD1, MBD2, MBD4 and methyl-CpG-binding protein 2 (MeCEP2) (Baubec et al., 2013; Du et al., 2015b; Hendrich and Bird, 1998; Lewis et al., 1992; Meehan et al., 1989). Those proteins interact with repressor complexes such as HDACs, MTs and nucleosome remodelers (Nan et al., 1998; Ng et al., 1999).

Recently, it has been shown that they bind genomic regions in relation to the CpG density, thus mostly binding methylated CpG-dense regions as inactive promoter CpG islands (CGIs) (Baubec et al., 2013). It also had been hypothesized that the binding of MBD proteins represses transcription by inhibiting the binding of TFs to CGIs. Yet, a recent study in which all four MBD proteins have been knocked out revealed only minor impacts on gene expression, suggesting that MBD proteins are not the main drivers of repression by DNA methylation at promoters (Kaluscha et al., 2022).

The second family of DNA methylation binding proteins recognized methylated CpGs with an SRA domain. One member of this family is UHRF1 that specifically recognizes hemi-methylated DNA and recruits DNMT1 to facilitate DNA methylation maintenance (Bostick et al., 2007; Sharif et al., 2007).

The third family contains zinc finger proteins which can also discriminate between methylated and unmethylated DNA. One of those is the TF zinc finger protein 57 (ZFP57) that binds to its methylated motif and is involved in the maintenance of ICRs by recruiting chromatin modifying cofactors that maintain DNA methylation and deposit H3K9me3 at these sites (Quenneville et al., 2011).

1.2.2.4 | Functions of DNA methylation in different genomic contexts

In mammalian cells, 70-80% of all CpG dinucleotides are methylated (Meissner et al., 2008), which has been associated with gene repressive functions, especially for the silencing of transposable elements, repetitive elements and imprinting genes (Ferguson-Smith and Bourc'his, 2018; Greenberg and Bourc'his, 2019; Iurlaro et al., 2017; Mattei et al., 2022). Despite the genome-wide distribution of DNA methylation, certain genomic regions show distinct and dynamic methylation patterns (Lister et al., 2009; Stadler et al., 2011; Ziller et al., 2013).

CpG islands (CGIs)

CpG-dense regions, called CpG islands (CGIs), represent the only genomic context that is predominantly unmethylated (Figure 2A) (Bird et al., 1985). CGIs are present in around two thirds of all promoters, especially of housekeeping and several developmental genes (Gardiner-Garden and Frommer, 1987; Ku et al., 2008; Larsen et al., 1992). The unmethylated state of CGIs is most probably driven by the CpG density itself, as multiple studies have shown (Krebs et al., 2014; Lienert et al., 2011; Long et al., 2016; Wachter et al., 2014). CGIs exhibit transcriptional activity higher than that of non-CGI promoters (Larsen et al., 1992), which could be linked to the lack of DNA methylation. Yet, a recent study provides evidence that this transcriptional activity is rather promoted by the CpG density itself (Hartl et al., 2019). This argues that DNA methylation does not directly regulate CGIs activity.

X chromosome inactivation (XCI)

DNA methylation plays a role in the long-term gene silencing of the inactive X chromosome (Figure 2B). In female mammalian cells, one of the two X chromosomes has to be transcriptionally inactivated to ensure proper gene dosage compensation between male and female cells (Lyon, 1961). This silencing mechanism is primarily driven by the non-coding RNA X-inactive specific transcript (XIST) (Galupa and Heard, 2018). In addition, the CGIs of the inactivated X chromosome appear to be methylated, which is catalyzed by DNMT3B (Gendrel et al., 2012). Yet, this DNA methylation seems to function only as a final lock of the silenced state, since it occurs after initial silencing of the genes (Fang et al., 2019; Galupa and Heard, 2018; Grant et al., 1992; Lock et al., 1987; Singer-Sam et al., 1990).

Transposable and repetitive elements

One of the most important targets of DNA methylation-mediated silencing are transposable elements and in particular retrotransposons (Figure 2C). Transposable elements are mobile genetic elements that represent 37-55% part of the mammalian genome (Alexander et al., 2010; Pourrajab and Hekmatimoghaddam, 2021; Sanchez-Delgado et al., 2016). While all DNMTs have been shown to be involved in their methylation (Bourc'his and Bestor, 2004; Bourc'his et al., 2001; Chen et al., 2003; Dahlet et al., 2020; Walsh et al., 1998), the murine DNMT3C has been found to be important for the methylation of evolutionarily young transposable elements in male germ cells (Barau et al., 2016; Jain et al., 2017). The loss of DNMTs and DNA methylation results in massive reduction of the repression of transposable elements, causing upregulation of neighboring genes and genomic instability (Dahlet

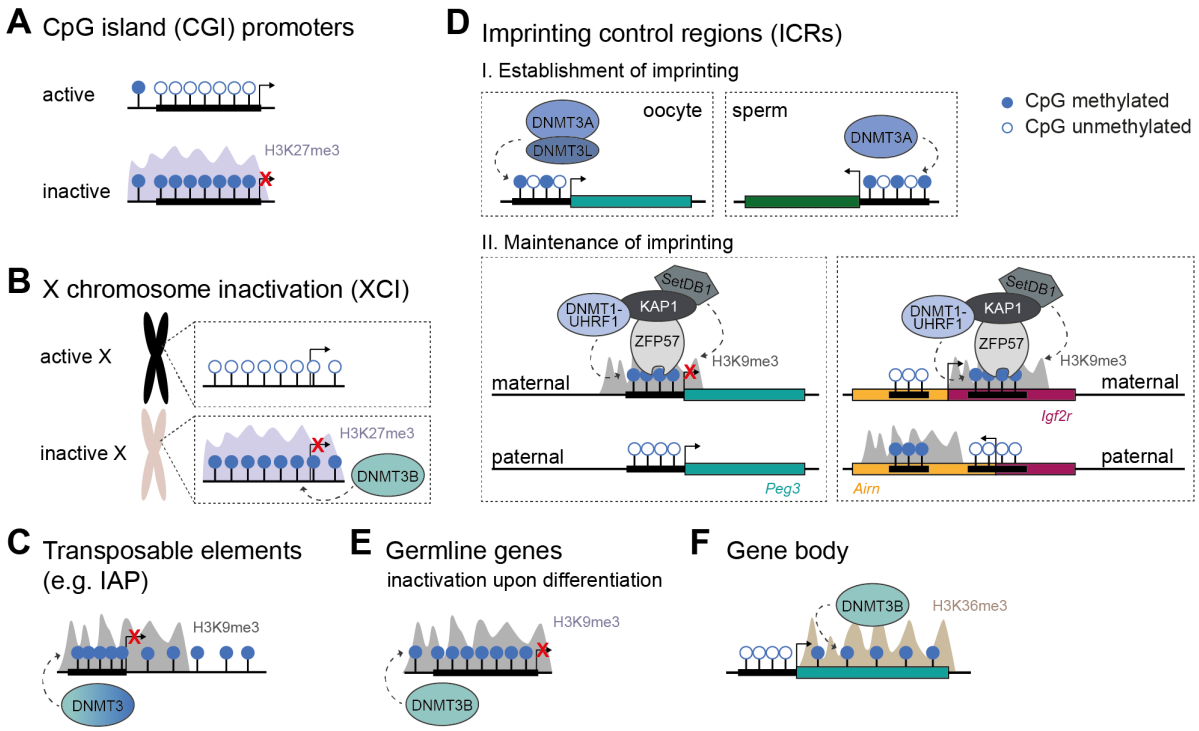

Figure 2: Functions of DNA methylation. (A) DNA methylation at CpG island (CGI) promoters is associated with gene silencing together with H3K27me3. **(B)** DNA methylation, deposited by the *de novo* DNMT3B, is involved in the repression of the inactivated X chromosome together with H3K27me3 and polycomb repression complex 2. **(C)** DNA methylation is involved in the silencing of repetitive elements like transposable elements, which are co-marked by the repressive histone mark H3K9me3. All DNMT enzymes are involved and in the mice germline also DNMT3C, which methylates young retrotransposons like intracisternal A particles (IAP). **(D)** DNA methylation is important for the regulation of imprinting control regions (ICRs) that drive mono-allelic gene expression. I.) The establishment of sex-specific imprinting after germline-specific epigenetic reprogramming is catalyzed by DNMT3A and its co-factor DNMT3L in oocytes, and DNMT3A in sperm. II.) In somatic cells, imprinting is maintained by the transcription factor ZFP57 that binds to its methylated motif in ICRs and recruits the KAP1-mediated silencing machinery, including the histone methyltransferase SetDB1, the deposits H3K9me3, and DNMT1with its cofactor UHRF1 that maintain the DNA methylation. Mono-allelic gene expression can be achieved by inhibiting the promoter (left box) or by alternative gene promoter activation (right box). **(E)** DNA methylation is also involved in germline gene inactivation upon somatic differentiated. This is mediated by DNMT3B, as well as by other silencing mechanisms that also drive H3K9me3. **(F)** At gene bodies, DNA methylation has no repressive function, but is correlated with transcription. For this, DNMT3B is recruited by the H3K36me3 mark downstream of active promoters. Gene body methylation is potentially involved in splicing and inactivation of cryptic intragenic promoters. Figure in parts adapted from (Greenberg and Bourc'his, 2019).

et al., 2020; Inoue et al., 2017b; Karimi et al., 2011). For mESCs, which are viable upon DNMT KOs, compensating mechanisms have been identified to restore long-term transcriptional silencing of transposable elements in hypomethylated genomes (Berrens et al., 2017; Sharif et al., 2007; Walter et al., 2016). Those involve, e.g., the deposition of histone methylation like H3K9me3 and H3K27me3.

In addition, DNA methylation has been found to be involved in the silencing of other repetitive elements like telomeric (Gonzalo et al., 2006), centromeric (Scelfo and Fachinetti, 2019) and pericentromeric (Saksouk et al., 2014) regions together with other silencing mechanisms of constitutive heterochromatin such as H3K9me3 (Nishibuchi and Déjardin, 2017).

Imprinting control regions (ICRs)

DNA methylation-mediated gene silencing plays a fundamental role at imprinting control regions (ICRs) (Figure 2D) (Ferguson-Smith and Bourc'his, 2018; Greenberg and Bourc'his, 2019). ICRs are genomic regions that are differentially methylated between the two parental alleles and drive mono-allelic expression of neighboring imprinted genes (Fitzpatrick et al., 2002; Lin et al., 2003; Proudhon et al., 2012; Thorvaldsen et al., 1998; Williamson et al., 2006; Wutz et al., 1997; Yang et al., 1998). Imprinted genes and their parent-of-origin specific expression are essential for embryonic development as well as for postnatal processes (Cleaton et al., 2014; Ferguson-Smith and Bourc'his, 2018). So far, around 25 ICRs have been identified, 22 of which are maternally methylated (Ferguson-Smith and Bourc'his, 2018).

Maternal ICRs overlap with intragenic CGI promoters. Their activation during oogenesis causes gene body – and ICR – methylation by DNMT3A, assisted by its cofactor DNMT3L (Figure 2D) (Proudhon et al., 2012). In contrast, paternal ICRs are CpG-poor intergenic sequences that are preferentially methylated by DNMT3A in sperm (Bourc'his et al., 2001; Kaneda et al., 2004; Li et al., 1993). During the epigenetic reprogramming post-fertilization, the ICR methylation is maintained through the binding of the ZFP57 to its methylated binding motif, which is enriched in ICRs, and its recruitment of the KRAB-A-interacting protein (KAP1)-centered heterochromatic complex (Quenneville et al., 2011). This causes maintained silencing through the histone methyltransferase SETDB1 and DNA methylation maintenance factors DNMT1 and UHRF1 (Figure 2D) (Messerschmidt et al., 2012; Quenneville et al., 2011; Riso et al., 2016; Zuo et al., 2012). Recently, another zinc finger protein, ZFP445, has been discovered that substantially impacts imprinting regulation in humans and co-regulates genomic imprinting in mice (Takahashi et al., 2019).

Germline gene expression programs

A large set of germline-specific genes are silenced by DNA methylation at the onset of somatic differentiation (Figure 2E). Their CGI-promoters are methylated by the DNMT3 family and loss of DNA methylation leads to reactivation of those genes (Auclair et al., 2014; Borgel et al., 2010; Dahlet et al., 2020; Hackett et al., 2012; Maatouk et al., 2006; Velasco et al., 2010). This establishes that DNA methylation has a causal role for the silencing of germline genes. The underlying mechanisms are still unclear and may involve other silencing co-regulators (Greenberg and Bourc'his, 2019).

Gene bodies

At gene bodies, DNA methylation has no repressive function on the active gene. Quite to the contrary, it is positively correlated with its transcription (Bender et al., 1999; Lister et al., 2009; Varley et al., 2013). Thereby, DNMT3B is recruited by H3K36me3 (Figure 2F) (Baubec et al., 2015), a common histone modification at gene bodies that is also positively correlated with transcriptional activity and transcription elongation (Kizer et al., 2005; Vakoc et al., 2006). Nevertheless, it still needs to be clarified, which exact role DNA methylation plays at gene bodies. On the one hand, it has been shown that it prevents the initiation of transcription from cryptic intragenic promoter sequences (Dahlet et al., 2020; Neri et al., 2017). On the other hand, it may facilitate transcription elongation and/or splicing,

since DNA methylation is enriched at exons (Gelfman et al., 2013; Maunakea et al., 2013; Shayevitch et al., 2018; Zhang et al., 2020a).

Enhancers

DNA methylation has also been found at enhancer. While inactive enhancers show high methylation levels like most of the genome, active enhancer show intermediate methylation levels of 10 to 60%. This has been shown to be a dynamic feature across many different cell types (Hartl et al., 2017; Hodges et al., 2011; Stadler et al., 2011; Ziller et al., 2013), making enhancers the most variable differentially methylated genomic regions. In addition, TET enzymes are enriched at active enhancers (Lu et al., 2014), which drives a high turnover of the local DNA methylation (Ginno et al., 2020). Some studies suggest that the demethylation of enhancers is important for their activation, thus arguing for a causal role of DNA methylation in gene regulation (Luo et al., 2018; Parry et al., 2021; Song et al., 2019), yet others have found its dynamics are rather consequential to the underlying changes in gene expression (Barnett et al., 2020). In conclusion, there is presently only limited evidence that DNA methylation at enhancers is generally instructive.

1.2.2.5 | DNA methylation dynamics in development

DNA methylation plays a crucial role in mammalian development (Greenberg, 2021; Greenberg and Bourc'his, 2019; Hemberger et al., 2009). While in somatic cells, most of the genome is fully methylated (Meissner et al., 2008), there are two dramatic DNA methylation reprogramming events during development in which DNA methylation is largely removed (Figure 3).

The first reprogramming event happens during early embryonic development directly after fertilization (Figure 3). Here, the zygotic genome is rapidly and progressively demethylated until the blastocyst stage where only around 20% of CpGs in the genome remain methylated (Wang et al., 2014). DNA methylation is primarily retained at transposable elements and ICRs (Lane et al., 2003; Smith et al., 2012; Wang et al., 2018, 2014). During this reprogramming, the gamete-specific DNA methylation patterns are removed to reach a pluripotent state. While the overall dynamics of de- and re-methylation and the timing of this methylation reprogramming have been largely identified, some of the mechanistic details are still elusive. Some studies, for instance, argue for a first wave of active demethylation mediated by TET3 (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011) and a subsequent wave of passive genome-wide demethylation during multiple rounds of cell division (Howell et al., 2001), whereas other studies provide evidence for replication- or TET3-independent demethylation (Amouroux et al., 2016; Santos et al., 2013). In addition, it is still under debate whether local *de novo* methylation may co-occur with the global demethylation to ensure methylation retention at desired regions (Amouroux et al., 2016; Zhu et al., 2018). TET enzymes could then be involved in protecting the genome from this methylation activity (Albert et al., 2020; Amouroux et al., 2016; Hill et al., 2018).

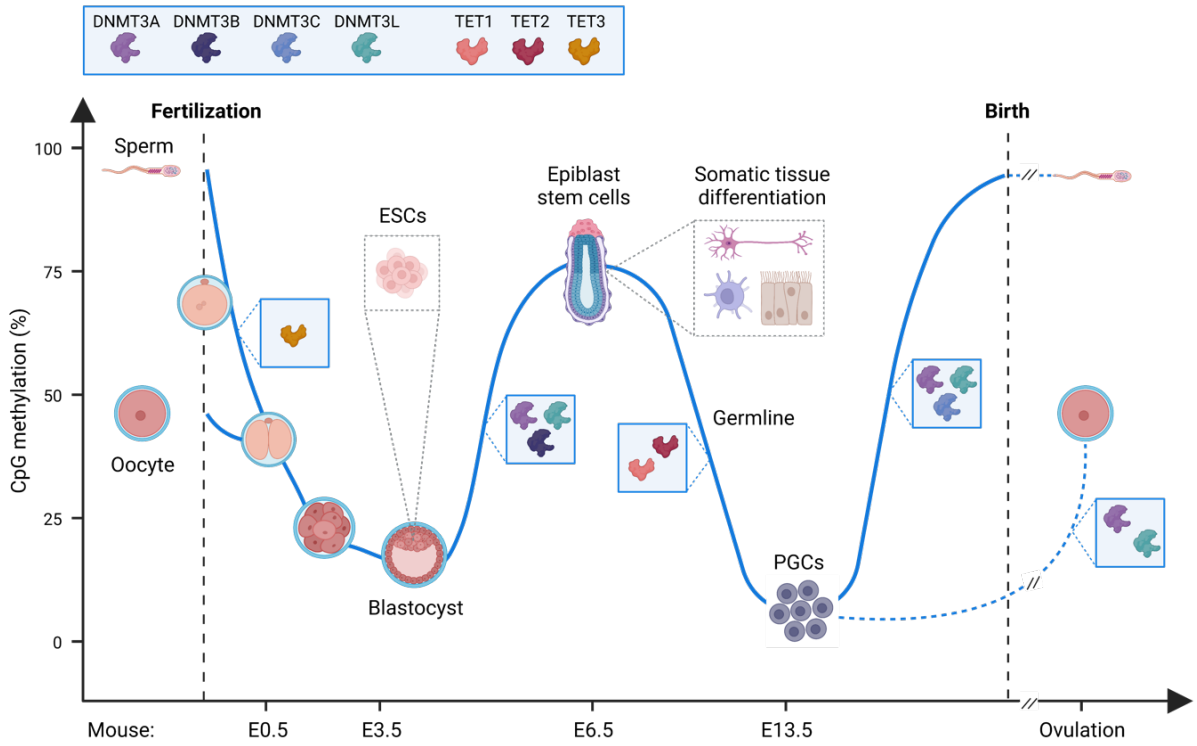

Figure 3: DNA methylation dynamics during embryonic and germline development. Sperms and oocytes have different levels of DNA methylation with ~80% and ~50%, respectively. After fertilization, global DNA methylation levels are massively removed until the blastocyst stage, mainly driven by TET3 and passive demethylation. Upon implantation, the genome is rapidly re-methylated to DNA methylation levels found in somatic tissue of 70-80%. During this process, DNMT3A, DNMT3B and DNMT3L are involved in the *de novo* methylation of the genome. A second wave of reprogramming happens in the germline, where TET1 and TET2 drive the demethylation including at imprinted regions. At the final stage, primordial germ cells (PGC) exhibit only 6-8% global methylation levels, mainly located at transposable elements. Afterwards, the genome is re-methylated by DNMT3A/B and in the mouse male germline also by DNMT3C. ESCs that are derived from the inner cell mass of the blastocyst show somatic methylation levels of 70-80% when cultured in serum/LIF. Blue line shows global DNA methylation levels. Blue boxes show main enzymes involved in the de- or re-methylation process in the different phases. Figure adapted from (Greenberg and Bourc'his, 2019) and created with BioRender.com.

Upon in uterus implantation, the epiblast genome is rapidly re-methylated by the *de novo* MTs DNMT3A/B until it reaches DNA methylation levels of somatic tissues (Figure 3) (Borgel et al., 2010; Seisenberger et al., 2012; Smith et al., 2017; Zhang et al., 2018). During this process, methylation is broadly deposited and potentially guided by H3K36me2/3 markings of the genome that DNMTs bind to (Baubec et al., 2015; Dhayalan et al., 2010; Weinberg et al., 2019). One important genomic context that escapes the DNA methylation are CGI-promoters. Multiple mechanisms have been proposed that explain this methylation protection (Greenberg, 2021). One hypothesis is based on the fact that DNMTs are repelled by H3K4me3, which is found at active promoters (Guo et al., 2015; Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010b). Another hypothesis involves the methylation protection at bivalent promoters. Bivalent genes are developmental genes that are marked by active (H3K4me3) and repressive (PRC2-deposited H3K27me3) histone marks and that are silent during embryogenesis, but become rapidly activated upon developmental cues (Azuara et al., 2006; Bernstein et al., 2006). Those promoters are free of DNA methylation, which could be due to the repelling of DNMTs by H3K4me3 as for active genes, H3K27me3 (Greenberg et al., 2017; Li et al., 2018), the recruitment of TET enzymes

(Boulard et al., 2015; Gu et al., 2018a; Manzo et al., 2017; Verma et al., 2018), and the binding of the TFs Developmental Pluripotency-Associated 2 and 4 (DPPA2/4) which might be involved in the recruitment of both silencing and activating complexes (Eckersley-Maslin et al., 2020; Gretarsson and Hackett, 2020).
Notably, mESCs that are derived from the inner cell mass of the hypomethylated blastocyst at embryonic day (E) 3.5 (Evans and Kaufman, 1981) harbor somatic-tissue methylation levels of 70-80% when cultured under serum/Leukemia inhibitory factor (LIF) conditions (Figure 3). However, when cultured in a medium that contains two inhibitors (2i) of distinct pathways (mitogen-activated protein kinase [MAPK] and glycogen synthase kinase-3 beta [GSK3β] inhibitors), mESCs are in a more "naïve" pluripotent state with low methylation levels that more closely resembles the blastocyst stage (Ficz et al., 2011; Habibi et al., 2013; Leitch et al., 2013).

The second – and even stronger – event of global DNA methylation erasure happens during germline reprogramming for the establishment of primordial germ cells (PGCs) (Figure 3) (Guibert et al., 2012; Hajkova et al., 2002; Lee et al., 2002; Vincent et al., 2013). During this process, a first wave of passive demethylation removes most genomic DNA methylation (Yamaguchi et al., 2013a). Subsequently, TET1 and TET2 enzymes facilitated a wave of active DNA demethylation, which results in the co-occurrence of 5hmC in the PGC genome (Hackett et al., 2013; Vincent et al., 2013; Yamaguchi et al., 2012, 2013b). During this second stage, DNA methylation is primarily removed from ICRs and the promoters of germ line specific genes. After this wave of reprogramming, only 6-8% of CpGs are still methylated in mouse PGCs (Wang et al., 2014), which are mostly enriched in retrotransposons (Guibert et al., 2012; Hajkova et al., 2002). Afterwards, global DNA methylation levels are reestablished in sex-specific manners during male and female germline differentiation up to ~80% and ~50% methylation levels in the sperm and oocyte genome, respectively (Figure 3) (Greenberg and Bourc'his, 2019; Kobayashi et al., 2012; Smallwood et al., 2014; Wang et al., 2014).

The reasons for those two dramatic waves of global DNA demethylation during early embryonic and germline development are still not fully understood. It is speculated that they are necessary to gain epigenetic plasticity at pivotal developmental stages, as well as to limit the inheritance of epimutations acquired by aging, environmental stimuli or diseases (Greenberg and Bourc'his, 2019). Nevertheless, it is indisputable that DNA methylation and its regulation play an important role in mammalian development. Thus, studying DNA methylation and understanding its functions in different genomic contexts has an important impact on our understanding of these developmental mechanisms.

1.2.2.6 | DNA methylation in cancer, ageing and other diseases

DNA methylation is a key epigenetic modification that is frequently dysregulated in diseases. In line with this, mutations in DNMT or TET enzymes have been associated with multiple diseases in humans (Greenberg and Bourc'his, 2019; Paul and Beck, 2014), including cancer (Nishiyama and Nakanishi, 2021), neurological disorders (Younesian et al., 2022), acute myeloid leukemia (AML) (Li et al., 2017)

and the immunodeficiency, centromere instability and facial anomalies (ICF) syndrome (Vukic and Daxinger, 2019).

In cancer, aberrant DNA methylation patterns are thought to be involved in carcinogenesis as well as in the progression of cancer (Baylin and Jones, 2011; McMahon et al., 2017; Nishiyama and Nakanishi, 2021). Thereby, cancer cells show characteristic global hypomethylation and local hypermethylation, leading to dysregulated gene expression and the development of tumorigenic traits. Hypermethylation is e.g. observed at CGI promoters, many of which are associated with tumor suppressor genes (Baylin and Jones, 2011; Jones, 2012). Hypomethylation, in contrast, is associated with the expression of transposable elements, leading to aberrant expression of other genes and genomic instability (Jordà et al., 2017). Due to these common DNA methylation patterns, DNA methylation is often used as a marker for cancer detection, progression and prognosis (McMahon et al., 2017; Nishiyama and Nakanishi, 2021). In addition, cancer treatment strategies targeting epigenetic factors is a growing field of research (Cheng et al., 2019).

Nevertheless, although various cancer types and their progression have been correlated with distinct changes in DNA methylation, the causal effects of these changes are still largely unknown. Interestingly, 90% of the genomic regions showing hypermethylation in cancer cells are already in a silenced state in healthy somatic cells, often associated with Polycomb-mediated silencing (Keshet et al., 2006; Schlesinger et al., 2007). In addition, most cancer types converge to a common methylation landscape over time, which leads to the question, which genomic regions have causal and disease-relevant roles in the initiation and progression of the cancerous traits (Greenberg and Bourc'his, 2019; Mattei et al., 2022).

Changes is DNA methylation over time is also one of the hallmarks of aging (López-Otín et al., 2013). In general, in a similar trend as observed in cancer, global DNA methylation levels decrease with age, while local DNA methylation patterns become more heterogeneous and show local hypermethylation (Horvath and Raj, 2018; Seale et al., 2022). Hypermethylation occurs preferentially at CGI-promoters of developmental and Polycomb-repressed genes and is conserved across different cells and tissues (Lu et al., 2022; Zhou et al., 2018). In comparison, hypomethylation is associated with low CpG-density at intronic and intergenic regions (Pérez et al., 2018) and is less conserved across tissues (Day et al., 2013). These changes in DNA methylation have been linked to age-related diseases and the decline in physiological function (Seale et al., 2022).

It is important to note that although DNA methylation functions as a biomarker to inform on disease onset and progression as well as the "epigenetic age" (Horvath and Raj, 2018; Seale et al., 2022), this does not infer that DNA methylation is causally involved in the disease or aging progress.

1.2.3 | Epigenetics and chromatin features of CREs

The mammalian genome harbors a plethora of epigenetic factors and chromatin features. In addition, these different factors are enriched in certain genomic contexts, resulting in complex combinations that are associated with different transcriptional regulations (Figure 4). In consequence, genomic elements and chromatin states can be identified by analysing those epigenetic combinations (Ernst

and Kellis, 2010). One computational approach to this aim is chromHMM (Ernst and Kellis, 2012), which is based on a multivariate hidden Markov model (HMM) that integrates genome-wide maps of chromatin features such as prevalent histone modifications, DNA accessibility and chromatin-associated proteins as e.g. RNA Polymerase II, CTCF and TFs. The unique combination of chromatin features enables the assignment of chromatin states to each genomic locus in the genome. Those chromatin states include functional elements like promoters, enhancers, insulators, transcribed, and repressed regions.

CREs are important regulators of gene expression patterns to establish and define cellular identities. TSS-proximal CREs are promoters that provide the binding platform for TFs, transcriptional co-factor complexes and the general transcription machinery. When promoters are active, their epigenetic landscape is defined by an enrichment of H3K27ac, H3K4me3 and H3K9ac, increased chromatin accessibility, RNA Polymerase II binding and low levels of DNA methylation at a high CpG density (Figure 4) (see chapter 1.2.2.4). Promoters can also be in a bivalent/poised state, in which they are accessible and bound by TFs and RNA Polymerase II but do not drive expression until further stimuli like differentiation cues. In this state, they are usually co-enriched for H3K4me3 as well as H3K27me3 and the Polycomb-mediated silencing machinery, while their DNA methylation levels are low (Figure 4) (see chapter 1.2.1.2 and 1.2.2.5) (Blanco et al., 2020). Upon activation, the H3K27me3 mark is lost and replaced by H3K27ac. Bivalent promoters are generally associated with developmentally important genes that are already poised for expression in pluripotent cells, but only become activated upon differentiation (see chapter 1.2.2.5) (Bernstein et al., 2006).

Downstream of the TSS, in the gene body, is the region of active transcription which can be defined in chromHMM states as transcription transition and transcription elongation. Upon transcription, gene bodies are generally marked by H3K36me3 and elevated levels of DNA methylation (Figure 4).

Distal CREs like enhancers have slightly different chromatin features. Active enhancers are usually marked by H3K4me1, H3K27ac and H3K9ac (Figure 4). In addition, they show intermediate DNA methylation levels and high chromatin accessibility. Also active enhancers show TF and RNA Polymerase II binding that leads to the transcription of short non-coding RNAs (Sartorelli and Lauberth, 2020). This so-called enhancer RNA can also be used to determine the activity of an enhancer by specific RNA sequencing techniques such as cap analysis gene expression (CAGE) (Andersson et al., 2014; Shiraki et al., 2003), precision run-on sequencing (PRO-seq) (Mahat et al., 2016) as well as capped small RNA sequencing (csRNA-seq) (Duttke et al., 2019).

As promoters, enhancers can be found in a poised state in pluripotent cells, in which they are co-enriched for active and repressive marks such as H3K4me1 and H3K27me3 (Figure 4) (Crispatzu et al., 2021).

Insulators are genomic elements that demarcate domains of independent gene regulation. They serve as boundaries to block permissive promoter activation by enhancers. Insulators are bound by insulator binding proteins, of which CCCTC-binding factor (CTCF) is the most prominent in mammalian cells (Figure 4) (Ali et al., 2016). Upon binding, CTCF mediates long-range interactions and contributes to the topology and 3-dimentional chromatin structure.

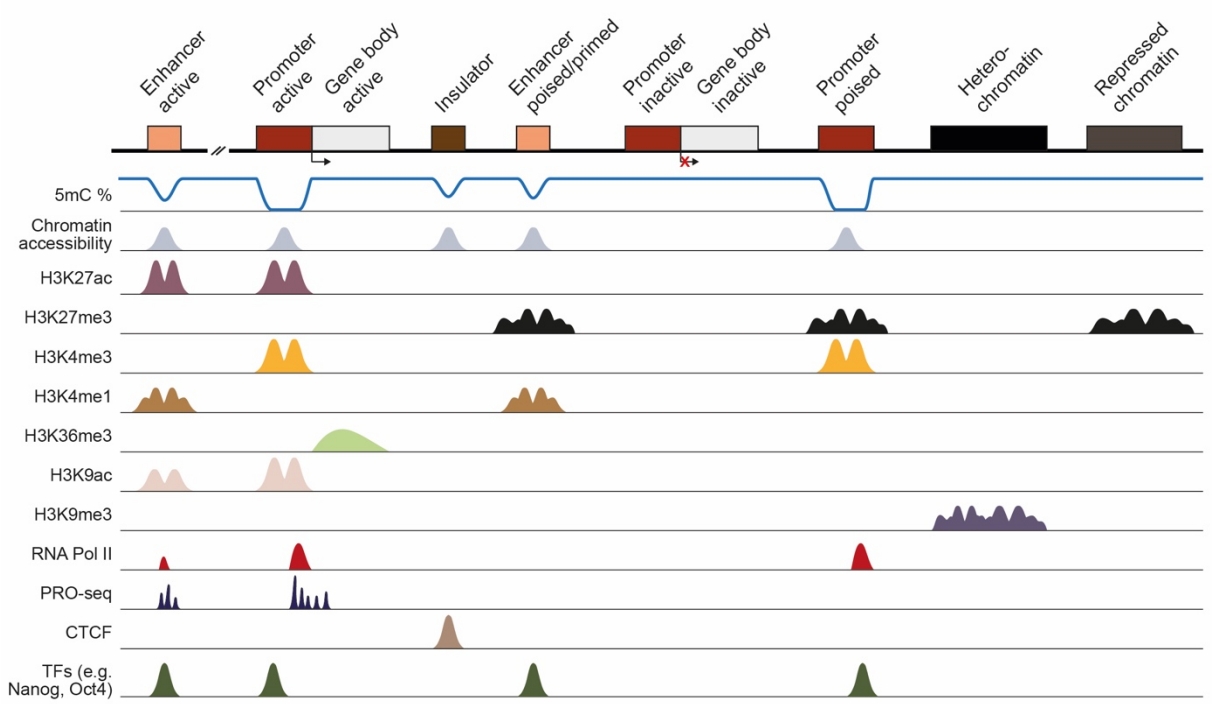

Figure 4: Epigenetic features define chromatin states. The combination of different epigenetic marks and chromatin features can be used to determine chromatin states like *cis*-regulatory elements as enhancers, insulators and promoters and their activation status, as well as hetero- and repressed chromatin. Shown is a schematic genome browser view of different genomic elements and the associated chromatin features as measured by bulk sequencing techniques: DNA methylation, chromatin accessibility, activating and repressing histone modifications, RNA Polymerase II (RNA Pol II), CTCF and other transcription factor (TF) binding, and enhancer activity measured by Precision Run-On sequencing (PRO-seq). Only the most common features are shown. Figure adapted from (Preissl et al., 2023).

In addition to active regulatory elements, also inactive chromatin states can be identified by the analysis of chromatin features. Heterochromatin is, e.g., enriched for the histone mark H3K9me3, while repressed chromatin shows H3K27me3 (Figure 4).

In the last decade, the Encyclopedia of DNA Elements (ENCODE) project has expanded their repertoire of defining genomic regions by incorporating functional and structural data about RNA transcription, chromatin structure and modification, methylome, chromatin looping, and binding of TFs (Abascal et al., 2020). Integrating chromatin states with chromatin looping and gene expression data allows a better understanding of regulatory network that defines cell-type-specific gene expression. Despite those advances in gene element mapping, defining the regulatory network of enhancers, which enhancers contribute – and to what extent – to the regulation of which genes, is still a difficult task and prone to high error rates (Moore et al., 2020).

1.3 | Regulation of enhancers by DNA methylation

As devolved in previous chapters (see chapter 1.2.2), the role of DNA methylation for transcriptional repression is well defined at CGI promoters, transposable elements, imprinted genes and the inactive X chromosome. However, to what extend DNA methylation is instructive for the regulation of enhancer activity is much less understood. Multiple factors have to be considered when evaluating epigenetic enhancer regulation, which will be discussed in the following.

First, enhancers have a generally low CpG density, especially in comparison to CGIs. Thus, methylated CpGs have to be evaluated individually and may have very local effects. Moreover, regulatory impacts of CpG clusters as it has been shown for some CGI promoters cannot be transferred to enhancer regions (Hartl et al., 2019; Krebs et al., 2014).

Second, in contrast to fully unmethylated promoters, enhancers exhibit intermediate methylation levels of 10-60%, which are highly dynamic across different cell types (Hartl et al., 2017; Hodges et al., 2011; Stadler et al., 2011; Ziller et al., 2013). This intermediate state argues for an epigenetic heterogeneity, in which individual cells within a homogeneous cell population have different methylation states of the same enhancer element. This heterogeneity makes it difficult to study methylation-dependent regulation and the correlation between multiple regulatory factors, since common bulk sequencing techniques produce outcomes that average out those heterogeneities.

Third, the direct impact of DNA methylation on TF binding is still largely unknown in different genomic contexts. Although it has been shown that the intermediate methylation is associated with TF binding and elevated TET activity at enhancers (Ginno et al., 2020; Stadler et al., 2011), it is still unclear what the causal relationship between TF binding and DNA methylation is in CpG-poor regions. DNA methylation may impact the binding of some TFs, but TFs can also induce local DNA demethylation, resulting in a dual relationship that is difficult to resolve genome-wide with common detection techniques.

1.3.1 | TFs and their variable relationship with DNA methylation

1.3.1.1 | DNA methylation can impact TF binding

TFs play a crucial part in the regulation of gene expression due to their ability to bind DNA at distinct motifs in a sequence-specific manner and recruit transcriptional co-factors to regulatory elements. Yet, motif sequence alone does not explain the specificity in TF binding, since hundreds of thousands of potential binding sites exist in the genome of which only a fraction is bound *in vivo* at any given moment. Thus, additional features must control in which genomic contexts TFs bind to their motifs.

An important regulatory mechanism of TF binding is cooperativity or competition between multiple TFs, where the binding of one TF to a regulatory element stimulates or inhibits the binding of other TFs (Spitz and Furlong, 2012). Additionally, the shape of the DNA in its 3D conformation has been suggested to potentially affect the binding specificities of TFs (Mathelier et al., 2016; Slattery et al., 2014). Lastly, chromatin features such as nucleosome positioning and epigenetic factors have been

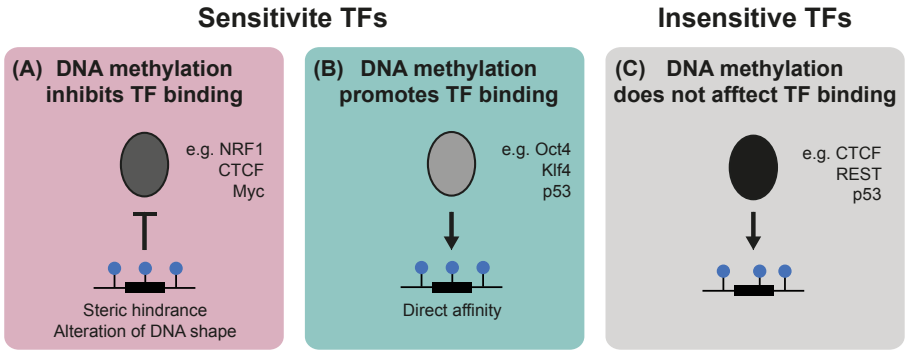

Figure 5: Schematic representation of different DNA methylation sensitivities of transcription factors (TFs). (A) Methylation-sensitive TFs can be inhibited in their binding by steric hindrance or alteration of the DNA shape by DNA methylation of the TF binding motif. **(B)** Methylation-sensitive TFs can be promoted in their binding by direct increase in their affinity by DNA methylation in their motif. **(C)** Insensitive TFs can bind their motif independent of the underlying DNA methylation.

shown to regulate the binding of TFs (Bell et al., 2011). For instance, most TFs cannot access their motif sequence when the region is bound by nucleosomes. Yet, a specific group of TFs, so called "pioneer" factors, have been identified that are able to bind nucleosome-occupied DNA motifs and induce local removal or repositioning of nucleosomes by recruiting chromatin remodeling complexes. This leads to the opening of the region and allows additional TFs to bind (Iwafuchi-Doi and Zaret, 2014).

The observation that DNA methylation is an important regulator of transcription has led to the hypothesis that also CpG methylation may regulate TF binding either through direct interaction between the methylated DNA and the TF or through indirect mechanisms involving other factors such as chromatin remodeling proteins (Baubec and Schübeler, 2014).
Early on, studies using different biochemical assays have supported the notion of the direct mechanism for multiple TFs (Bednarik et al., 1991; Campanero et al., 2000; Iguchi-Ariga and Schaffner, 1989; Prendergast et al., 1991; Watt and Molloy, 1988). In recent years, grand *in vitro* studies have shown that the majority of TFs (60% of the ones tested in (Yin et al., 2017) can be directly affected in their binding when their motif is methylated (Kribelbauer et al., 2017; Yin et al., 2017). They also showed that TFs can be positively or negatively modulated (Figure 5). Yin et al. for example showed that bHLH-, bZIP-, and ETS-family TFs were generally inhibited by DNA methylation, while NFAT factors and members of the homeodomain family preferred binding to motifs containing a methylated CpG. Those comprehensive *in vitro* studies support the hypothesis of a direct inhibitory mechanism of DNA methylation. However, some of the *in vitro* data is rather contradictory and *in vivo* evidence for this direct methylation-sensitivity remains limited (Table 1).
Preferential binding to methylated motifs has been shown *in vivo* for Oct4 (Yin et al., 2017), Klf4 (Hu et al., 2013; Spruijt et al., 2013; Wan et al., 2017) and p53 (Kribelbauer et al., 2017), which is in line with their proposed pioneering activity (Figure 5B). Notably, neither Oct4 nor Klf4 have a CpG in their consensus motif, arguing that this methylation preference has only little effect in the genomic context *in vivo*. ZFP57 and ZFP445 have been shown to bind methylated DNA at ICRs where they fulfill

important roles in the silencing of imprinted genes (Li et al., 2008; Quenneville et al., 2011; Riso et al., 2016; Takahashi et al., 2019).

Negative methylation-sensitivity has been confirmed *in vivo* for NRF1 (Domcke et al., 2015), CTCF (Maurano et al., 2015), Max-Myc (Cusack et al., 2020; Domcke et al., 2015; Yin et al., 2017), BANP (Grand et al., 2021), ONECUT1 (Kaluscha et al., 2022) and CREB1 (Kaluscha et al., 2022) (Figure 5A). Most of these studies used DNMT triple-knockout (TKO) ESCs to identify additional binding events in absence of genomic DNA methylation. This showed for this limited set of TFs that DNA methylation prevents their binding outside of CREs and controls the formation of cryptic binding sites throughout the genome. Yet, whether motif methylation impacts TF binding at active CREs and their regulation remains elusive.

Multiple features have been found that potentially drive the methylation-sensitivity of TFs. Those include the shape of the DNA that is affected by the methylation (Du et al., 2015a; Slattery et al., 2014), the chemical interaction between the methyl-group and the TF (Iguchi-Ariga and Schaffner, 1989; Kribelbauer et al., 2017; Yin et al., 2017) as well as the steric hindrance of the TF (Figure 5) (Iguchi-Ariga and Schaffner, 1989; Kribelbauer et al., 2017; Yin et al., 2017). Hence, the position of the methylated CpG within the motif is a crucial component for its impact as it has been shown for CTCF (Maurano et al., 2015) and ATF4 (Kribelbauer et al., 2017). On a structural basis, methylated cytosine can promote the binding of a TF as it has been shown for HOX TFs, where the methyl-group mimics a thymidine base in the binding pocket (Kribelbauer et al., 2017). Unsurprisingly, also the number and position of the CpG within the TF motifs is impacted by local DNA methylation. Based on the TF motif databank JASPAR, 70% of TF motifs do not contain a CpG (Héberlé and Bardet, 2019), arguing that most TFs might not be affected by local DNA methylation. This stands in contrast to the observation that most TFs are impacted by CpG methylation *in vitro*, even if their canonical motif does not contain a CpG (Yin et al., 2017).

Since the first evidences for a regulatory impact in the late 1980s, many studies have followed with ambivalent results. For example, since CTCF's discovery in 1990 (Lobanenkov et al., 1990), studies have shown that it is in some cases negatively sensitive and in some cases insensitive to DNA methylation (Table 1). Similar contradicting observation have been made for other TFs, arguing that TFs cannot always be classified as per se methyl-sensitive or insensitive.

In conclusion, the functional role of DNA methylation for the regulation of TF binding is still largely unclear and the contradicting results argue for a context specific impact of DNA methylation that is not generalizable for all TFs within a TF class and not for single TFs for all genomic contexts across the genome.

Table 1: Evidence from *in vitro* and *in vivo* studies for TFs with different DNA methylation sensitivities

Sensitive TFs (negative sensitivity)		
TF	***in vitro***	***in vivo***
***ATFs** (activating transcription factors)	- (Mann et al., 2013)[4] - (Yin et al., 2017)[1] - (Kribelbauer et al., 2017)[2]: ATF4 – negative when CpG central in motif - (Zuo et al., 2017)[3]: BATF1 (B cell–activating TF1)	
APs (JUN, FOS) (activator proteins)	- (Yin et al., 2017)[1] - (Mann et al., 2013)[4]	
AP-2 (activating protein 2)	- (Gaston and Fried, 1995; Tate and Bird, 1993)	
BANP (BTG3-associated nuclear protein)		- (Grand et al., 2021)
CEBPG (CCAAT/enhancer binding protein G)	- (Mann et al., 2013)[4]	
CEBPD	- (Mann et al., 2013)[4]	
CREB1 (CAMP responsive element binding protein 1)	- (Iguchi-Ariga and Schaffner, 1989) - (Mancini et al., 1999) - (Tierney et al., 2000) - (Spruijt et al., 2013) - (Yin et al., 2017)[1]	- (Kaluscha et al., 2022): WT vs DNMT TKO neurons at intracisternal A-type particle elements
***CTCF** (CCCTC-binding factor)	- (Bell and Felsenfeld, 2000) - (Filippova et al., 2001) - (Hashimoto et al., 2017): structural insights - (Zuo et al., 2017)[3]	- (Maurano et al., 2015): at specific sites
E2F	- (Gaston and Fried, 1995; Tate and Bird, 1993) - (Yin et al., 2017)[1]	
GABPA (GA binding protein transcription factor subunit alpha)		- (Domcke et al., 2015): potential candidate - (Cusack et al., 2020)
GLI1 (glioma-associated oncogene homolog 1)	- (Zuo et al., 2017)[3]	
MYC	- (Gaston and Fried, 1995; Tate and Bird, 1993) - (Yin et al., 2017)[1]	- (Domcke et al., 2015): potential candidate - (Yin et al., 2017): ChIP-seq in WT vs DNMT TKO mESCs shows minimal effect - (Cusack et al., 2020): Max-Myc
NF-kB (nuclear factor 'kappa-light-chain-enhancer' of activated B-cells)	- (Gaston and Fried, 1995; Tate and Bird, 1993)	
NRF1 (nuclear respiratory factor 1)	- (Kumari and Usdin, 2001)	- (Domcke et al., 2015): first TF shown to be sensitive *in vivo*; binds to more binding sites in DNMT TKO vs WT mESCs (validated by ChIP-seq & reporter construct)
ONECUT1		- (Kaluscha et al., 2022): only at the CpG-containing motif
***SP1** (specificity protein 1)	- (Clark et al., 1997): dependent on position	
USF (upstream TF 1)	- (Yin et al., 2017)[1]	
***YY1** (Yin Yang 1)		- (Cusack et al., 2020)
BHLH family (basic Helix-Loop-Helix)	- (Yin et al., 2017)[1]	
BZIP family (basic leucine zipper)	- (Yin et al., 2017)[1]	
ETS family	- (Yin et al., 2017)[1] - (Gaston and Fried, 1995)	- (Domcke et al., 2015): potential candidate
Sensitive TFs (positive sensitivity)		
TF	***in vitro***	***in vivo***
***ATF**	- (Kribelbauer et al., 2017)[2]: ATF4 – positive	

	when CpG in flanking region)	
CEBPA	- (Rishi et al., 2010): EMSA - (Mann et al., 2013)[4]	
CEBPB	- (Yin et al., 2017)[1]: weakly positive - (Mann et al., 2013)[4] - (Kribelbauer et al., 2017)[2]	- (Mann et al., 2013): ChIP-seq (11% of methylated motifs bound, 54% of unmethylated motifs bound) - (Zhu et al., 2016) ChIP-seq (25% of TFBS methylated)
CEBPE	- (Yin et al., 2017)[1]: weakly positive	
***CEBPG**	- (Yin et al., 2017)[1]: weakly positive	
HOXA9 (and PBX-HOXA9 complex) (Homeobox A9)	- (Hu et al., 2013): EMSA - (Kribelbauer et al., 2017)[2]: structural report	
***HOXB13** (homeobox B13)	- (Yin et al., 2017)[1] - (Zuo et al., 2017)[3]: depending on CpG position - (Yin et al., 2017)[1]: structural report contradicting Zuo et al.	
HOXC11	- (Yin et al., 2017)[1]	
***KAISO/ZBTB33** (Zinc Finger And BTB Domain Containing 33)	- (Yin et al., 2017)[1] - (Prokhortchouk et al., 2001): EMSA - (Daniel et al., 2002): structural report	
KAISO-like ZBTB4 & ZBTB38	- (Hudson et al., 2018): structural report	- (Filion et al., 2006): transient transfection in mice
KLF/ KLF4 (Krüppel-like factor)	- (Yin et al., 2017)[1] - (Spruijt et al., 2013): Klf2/4/5 (proteomics approach & DNA pull-down) - (Hu et al., 2013): different sequence preferences for un-/methylated motifs	- (Spruijt et al., 2013): in mESC 18.5% of Klf4 binding sites methylated - (Hu et al., 2013; Zhu et al., 2016): in hESCs 3% of TFBS have a CpG, of those 48% are methylated - (Wan et al., 2017)
OCT4/POU5F1 (Octamer-Binding Protein 4)	- (Yin et al., 2017)[1]: but no CpG in canonical motif	- (Yin et al., 2017): ChIP-seq in WT vs DNMTs TKO mESCS shows few sites lose binding
PAX (Paired box)	- (Yin et al., 2017)[1]	
p53	- (Kribelbauer et al., 2017)[2]	- (Kribelbauer et al., 2017)
***SP1**	- (Yin et al., 2017)[1]	
ZFP57 (zinc finger protein 57)	- (Zuo et al., 2017)[3]	- (Li et al., 2008; Quenneville et al., 2011; Riso et al., 2016): binds at ICRs in mice
Insensitive TFs		
TF	***in vitro***	***in vivo***
***CTCF**		- (Stadler et al., 2011): WT vs DNMT TKO mESCs show that CTCF binding sites globally unaltered (also validated by stable insertion) - (Maurano et al., 2015): HCT116 cells
***HOXB13**		- (Zuo et al., 2017)[3]: depending on CpG position
***KAISO/ZBTB33**	- (Daniel et al., 2002): structural report	- (Blattler et al., 2013): does not bind methylated DNA but rather highly active promoters marked with high levels of acetylated histones
REST (RE1-Silencing Transcription factor)		- (Stadler et al., 2011): stable insertion
***SP1**	- (Harrington et al., 1988; Höller et al., 1988): EMSA	- (Cusack et al., 2020)
***YY1**	- (Gaston and Fried, 1995)	

* TF for which contradicting evidences exist; [1] *in vitro* system using a methyl-Systematic Evolution of Ligands by EXponential Enrichment (SELEX) and bisulfite-SELEX approach; [2] in vitro EpiSELEX-seq approach; [3] *in vitro* Methyl-Spec-seq approach; [4] *in vitro* approach using methylated binding microarrays

1.3.1.2 | TF binding can impact DNA methylation

While DNA methylation can regulate the binding of some TFs, vice versa, the binding of a TF can also affect the local DNA methylation. Multiple studies investigating chromatin accessibility, TF binding and DNA methylation have shown that the binding of TFs is correlated with reduced levels of DNA methylation at regulatory elements (Detilleux et al., 2022; Feldmann et al., 2013; Ginno et al., 2020; Sérandour et al., 2012; Shen et al., 2013; Song et al., 2013; Stadler et al., 2011). This correlation suggests that TF binding might induce local DNA demethylation (Figure 6). Using an experimental and computational modeling approach, a recent study investigated local methylation dynamics genome-wide in mESCs and showed that the binding of TFs leads to local demethylation in the vicinity of the binding event. This is either achieved by blocking the maintenance and *de novo* DNA methylation by DNMTs (Figure 6A), by increasing the active demethylation by TET enzymes (Figure 6B) or both. Additional studies have found that different TFs drive different mechanisms of DNA demethylation. A recent study has identified multiple TFs like CTCF, REST KLF4, NRF1, Sox2, Myc and others that inhibit *de novo* or maintenance methylation by DNMTs surrounding their bound sites, and TFs that induce local active DNA demethylation by TETs (Vanzan et al., 2021). Some of these TFs have been confirmed also in other studies, as for example Klf4 (Sardina et al., 2018), FOXA1 (Detilleux et al., 2022; Reizel et al., 2021; Yang et al., 2016), REST (Perera et al., 2015), GATA3 (Detilleux et al., 2022) and CTCF (Dubois-Chevalier et al., 2014). Interestingly, it has been shown that TET activity preferentially occurs adjacent to bound sites and extends over multiple hundred base pairs (Ginno et al., 2020; Reizel et al., 2021), while the bound motif itself is protected from TET-mediated demethylation (Ginno et al., 2020). Whether the recruitment of TET enzymes is direct and specific by TFs, or rather indirect by the increased accessibility surrounding the binding site, is still under debate (Figure 6C). On the one hand, specific recruitment has been shown for some individual TFs (Guilhamon et al., 2013; Perera et al., 2015; Rampal et al., 2014; Sardina et al., 2018; Xiong et al., 2016; Zeng et al., 2016). On the other hand, no common features have been found, arguing that TET enzymes may rather be recruited indirectly through other mechanisms (Charlton et al., 2020; Ginno et al., 2020). It is speculated that the loss of DNMT activity at TF bound sites results in unmethylated CpGs in the vicinity that are subsequently

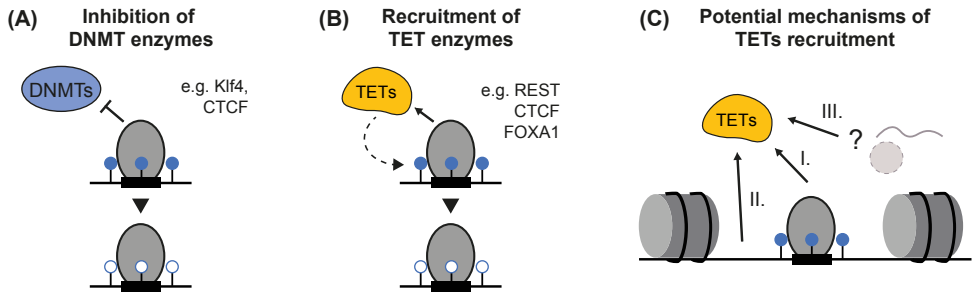

Figure 6: Schematic representation of the local impacts of transcription factor binding on DNA methylation. Binding of transcription factors (TFs) has been associated with reduced DNA methylation levels at distal *cis*-regulatory elements like enhancers that can be induced by different mechanism. **(A)** TF binding can inhibit the activity of DNMTs, leading to locally reduced methylation levels. **(B)** TF binding can recruit TET enzymes that lead to the active demethylation of DNA in the vicinity. **(C)** It is still unclear how TETs are recruited to active *cis*-regulatory elements. Potential mechanisms are (I.) direct recruitment by DNA bound TFs, (II.) attraction by accessible regions adjacent to bound motifs, and (III.) recruitment through other indirect mechanisms involving other factors.

bound by TET enzymes through their CXXC domain or mediated by IDAX (Ginno et al., 2020), while this is contradicted by other studies showing that recruitment of TETs is independent of their CXXC domain (Charlton et al., 2020).
Independent of the recruitment mechanism for TETs, the methylation dynamics at active CREs highly suggest that the binding of a TF leads to increased local accessibility and reduced local DNA methylation, which could enable methyl-sensitive TFs to bind to formerly protected binding sites.

1.3.2 | Epigenetic heterogeneity at active enhancers

Genomic elements that are involved in the regulation of gene expression exhibit different methylation dynamics. While promoters are fully unmethylated in their active state and fully methylated when inactive, enhancers show a much lower dynamic range. Active as well as poised enhancers show intermediate methylation levels of 10-60%. This is a consistent, but dynamic feature across different cell types. To understand DNA methylation dynamics and its potential impact on gene regulation, one has to understand what underlies this intermediate methylation.
Importantly, epigenetic heterogeneity is not limited to DNA methylation (Carter and Zhao, 2021). Single-cell studies of histone modifications have revealed heterogeneity of chromatin states such as H3K4me2 (Rotem et al., 2015) as well as H3K4me3 (Ku et al., 2019) within different cell type populations. And also nucleosome positioning around DHSs has been shown to exhibit cell-to-cell variation, which is positively correlated with variations in accessible regions (Lai et al., 2018).
The consequences of this epigenetic heterogeneity on gene regulation are, however, less clear and limited to individual examples. In addition, the heterogeneity makes it in general challenging to investigate the hierarchical and casual relationships between DNA methylation, TF binding, histone modifications and gene expression (Carter and Zhao, 2021; Jin et al., 2011; King et al., 2016; Zhu et al., 2016).

1.3.2.1 | Bulk- and single-cell approaches for measuring DNA methylation

The most common technique to measure DNA methylation is whole genome bisulfite sequencing (WGBS) (Cokus et al., 2008; Frommer et al., 1992; Mattei et al., 2022). The basic principle of this technique is the conversion of unmethylated cytosines to uracils, which are later detected as thymidines after PCR amplification (Figure 7A). For this, DNA samples are treated with sodium bisulfite, a chemical compound, that reacts with unmethylated cytosines and deaminates them into uracils, while methyl-cytosines are unchanged. After high-throughput sequencing, bisulfite-sensitive alignment strategies allow to determine which cytosines have been methylated. Since DNA methylation is a bi-modal state, an individual cytosine can either be methylated or not (Figure 7A).

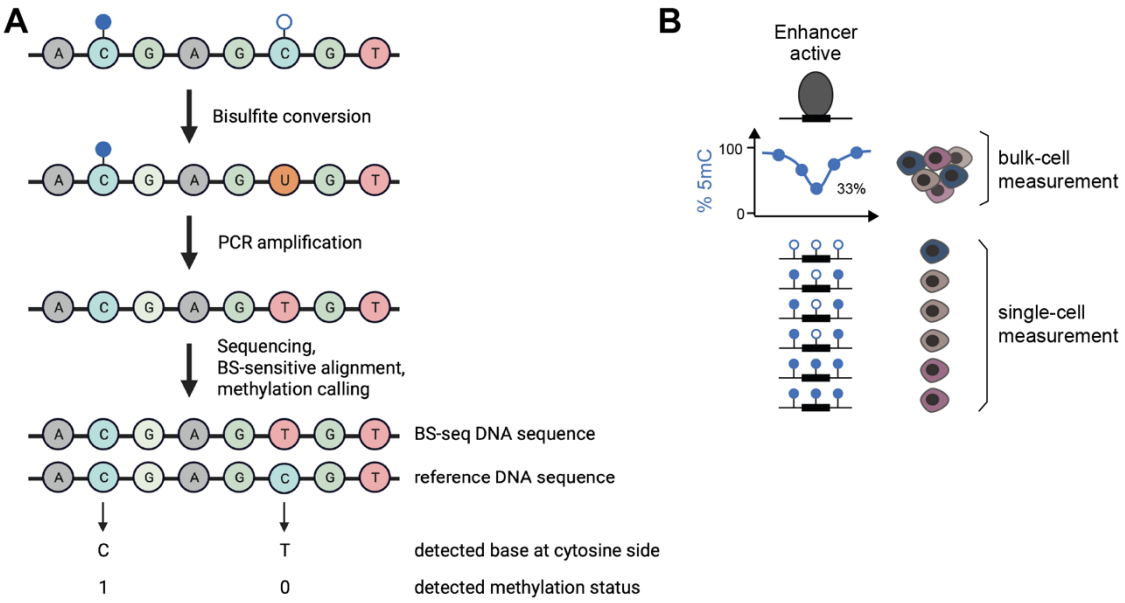

Figure 7: Bisulfite sequencing and epigenetic heterogeneity. (A) Schematic depiction of bisulfite sequencing (BS-seq). For this, DNA is treated with sodium bisulfite that converts all unmethylated cytosines into uracils (and thymidine after PCR amplification), while methyl-cytosines stay cytosines. After high-throughput sequencing, final sequences are aligned BS-sensitively and methylation is called by comparing base calling of cytosine sites against the reference. The final methylation status for each cytosine is either 1 (methylated) or 0 (unmethylated). **(B)** Schematic depiction of epigenetic heterogeneity at active enhancers. Intermediate methylation levels in bulk-measurements represent cell-to-cell variability in the methylation state at a single-cell level.

However, although WGBS can in principle extract methylation states of individual DNA molecules, it is usually used as a bulk measurement technique, in which the final methylation read out for each cytosine is an average over the whole cell population (Figure 7B). This results in potential methylation states of 0% to 100%. In consequence, the observed intermediate methylation at active enhancers is also an average and displays a cell-to-cell variability within the cell population. For example, when a cytosine shows 40% methylation in bulk measurements, that means that 40% of the cells in the population were methylated at this specific cytosine, while 60% were not (Figure 7B). This locus-specific epigenetic heterogeneity within a phenotypically homogeneous cell population has been directly quantified by single-molecule (Landan et al., 2012; Shipony et al., 2014) and single-cell methylation profiling studies (Farlik et al., 2015; Smallwood et al., 2014).

1.3.2.2 | Bulk- and single-cell approaches for measuring chromatin accessibility, histone modifications and TF binding

The most common methods to measure chromatin accessibility genome-wide are DNase I hypersensitive site sequencing (DNase-seq) and assay for transposase-accessible chromatin sequencing (ATAC-Seq). While DNase-seq is based on the sensitivity of accessible regions to be digested by the enzyme DNase I (Figure 8A) (Song and Crawford, 2010), ATAC-seq is based on the sensitivity to Tn5 transposase that cuts and directly ligates sequencing adapters to accessible DNA sequences (Figure 8B) (Buenrostro et al., 2013). Both methods result in the enrichment of reads at

chromatin accessible regions, which are in the case of DNase-seq called DNase hypersensitive sites (DHS). Both techniques have been widely used to detect CREs and changes in their activity. Yet, they share some major disadvantages. First, by digesting or cutting the accessible DNA they lose DNA integrity and are unusable to explore accessibility over longer stretches of DNA. Second, by enriching for accessibility, both methods provide enrichment scores that inform only on the accessible fraction of molecules (Figure 8A and B). Neglecting the inaccessible fraction of molecules, disables them to provide true quantitative measurements of the chromatin accessibility in a cellular population. Third, both methods provide only bulk read-outs by averaging over all cells of a cellular population (Figure 8A and B). This neglects the epigenetic heterogeneity that is integral to many homogeneous cell populations (Carter and Zhao, 2021) and makes is difficult to define true correlations between different gene regulatory layers.
This problem has been partly overcome by the adaptation of both techniques to single-cell read-outs (Buenrostro et al., 2015; Cusanovich et al., 2015; Jin et al., 2015), which resulted in a great wave of evidences for chromatin accessibility heterogeneity between cells in homogeneous populations. One study, for instance, showed that two single cells differ in around 25% of their accessible regions (Jin et al., 2015). Yet, the sparsity of the data of single-cell sequencing methods does not allow for the analysis of cell-to-cell variation at individual CREs (Cusanovich et al., 2015).

The detection of TF binding events has been through a similar advancement. For decades, chromatin immunoprecipitation (ChIP) (O'Neill and Turner, 1995, 1996), later coupled with high-throughput sequencing (ChIP-seq) (Barski et al., 2007; Mikkelsen et al., 2007), has been the most common method to measure DNA-binding as well as histone modification events. It is based on the binding of an antibody to the protein or modification of interest followed by the precipitation and sequencing of the sheered antibody-bound DNA sequences (Figure 8D). ChIP-seq provides bulk information about the localization and relative binding frequency inferred from the peak size of the enrichment. However, as for bulk chromatin accessibility measurement techniques, this information neglects the cell-to-cell heterogeneity of TF binding and histone modifications and does not provide quantitative measurements. The same is true for newer mapping techniques such as Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (Skene and Henikoff, 2017). This method is based on the use of an antibody against the protein of interest to tether DNA cleaving enzyme MNase to the loci where the protein is bound (Figure 8E). The cut DNA fragments are subsequently extracted and sequenced. A similar technique, Cleavage Under Targets and Tagmentation (CUT&Tag) (Kaya-Okur et al., 2019), is based on an analogous principle, but directly labels the cut DNA fragments with sequencing adapters by tethering a Tn5-transposase to the protein of interest. Both techniques neglect the cellular heterogeneity and enrich only for the bound fraction of DNA molecules.

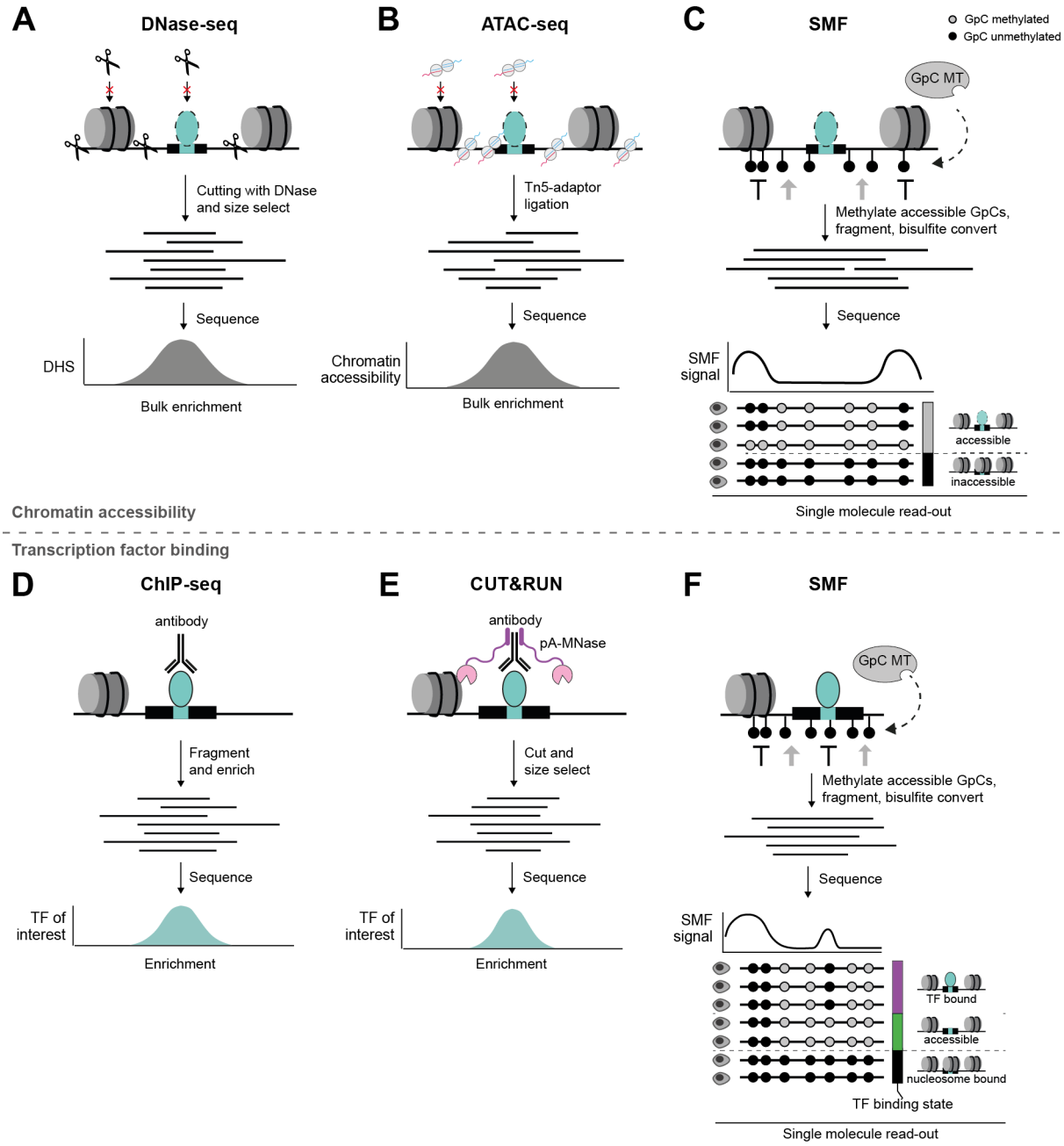

Figure 8: Comparison of bulk-cell genomics techniques measuring chromatin accessibility and transcription factor binding to single molecule footprinting. (A-B) Schematic representation of bulk-cell methods to measure chromatin accessibility. In both cases, only the accessible fraction of molecules is captured. (A) DNase-seq uses a DNase I to cut open chromatin regions. After size selection, DNA fragments are sequenced. The read-out is a bulk enrichment at open chromatin regions, called DNase hypersensitive sites (DHS). (B) ATAC-seq uses Tn5-transposons to cut and directly ligate adapters to open chromatin regions. All adapter-ligated DNA fragments are sequenced and the read-out is a bulk enrichment at open chromatin regions. **(C)** Schematic representation of single molecule footprinting (SMF) that uses a GpC methyltransferase to methylated accessible GpCs. After chromatin fragmentation and bisulfite conversion, all DNA fragments are sequenced. The read-out are individual molecules whose accessibility is depicted by the GpC methylation. All accessibility states (closed/inaccessible and open/accessible chromatin) are captured, enabling quantitative analysis. **(D-E)** Schematic representation of bulk-cell methods to measure transcription factor (TF) binding. In both cases, only the TF bound fraction of molecules is captured. (D) For ChIP-seq, chromatin is fragmented and an antibody against the TF (or histone modification) of interest is used to precipitate the fraction of DNA molecules that are bound by the targeted protein. The read-out is a bulk enrichment at TF bound regions. (E)

(legend continued on next page)

CUT&RUN uses an antibody targeting the TF (or histone modification) of interest and a pA-MNase fusion protein that binds to the antibody. Upon activation, the MNase cuts proximal chromatin, releasing small DNA fragments bound by the TF of interest. All small DNA fragments are sequenced and the read-out is a bulk enrichment at TF bound regions. **(F)** Schematic representation of SMF as described in (C). The read-out are individual molecules whose nucleosome and TF occupancy is depicted by the GpC methylation. All TF bound states (TF bound – purple, accessible -green, nucleosome bound – black) are captured, enabling quantitative analysis of TF binding frequencies. Figure in parts adapted from (Klemm et al., 2019).

1.3.3 | Is DNA methylation instructive at enhancers?

In the end, the fundamental question whether DNA methylation is instructive for the activity of enhancers in mammalian cells is yet to be resolved. The main difficulties in answering this question are the epigenetic heterogeneity and the bidirectional regulation of TF binding and DNA methylation. In addition, most DNA methylation studies rely on disturbing the system either by genetic disruptions or by cell differentiation to drive changes in the DNA methylation whose effects on the cellular system can then be subsequently measured. This however limits the conclusions that can be drawn. For example, the identification of Nrf1 and other proteins as methyl-sensitive TFs was often based on the global removal of DNA methylation which resulted in an increase in sites bound by the respective TF. However, in the wild type cell those sites would anyways never be bound by, e.g., Nrf1. Thus, one can make the conclusions that DNA methylation inhibits the binding of TFs to spurious sites in the genome, and also that the given TF is sensitive to methylation at those sites. Yet, these experiments do not unambiguously reveal whether DNA methylation is regulating the binding of TFs at active enhancers and by this regulating the activity of the enhancer.

Other approaches have used multi-omics techniques to measure the chromatin and epigenetic changes upon differentiation in order to investigate the hierarchies and relationships between different regulatory features. One study, for instance, used ATAC-Me, a method that combines DNA methylation sequencing with ATAC-Seq, to investigate temporal changes in chromatin accessibility and DNA methylation changes during myeloid differentiation (Barnett et al., 2020). This study showed that the opening of CREs upon differentiation is usually preceding the loss of DNA methylation at transitioning enhancers. However, this ATAC-Me approach is, first, based on strong changes in gene expression patterns, and, second, neglects the heterogeneity in DNA methylation and chromatin accessibility, since ATAC-Me cannot resolve this (see chapter 1.3.2.2).

A leap towards understanding the impact of DNA methylation on enhancer activity and transcriptional output has recently been made using a single-cell reporter of genome methylation system (Song et al., 2019). In this, the readout of a fluorescence signal informs about the methylation state of a specific locus and can be used to separate different sub-populations of cells. This system allowed to resolve epigenetic heterogeneity and investigate the dynamics of methylation and its impact on enhancer activity and gene expression of two >10-kb large enhancer clusters, Sox2 and Mir290, that are involved in controlling the expression of essential genes in mESCs. However, this approach only allows to investigate individual loci and is not feasible in a larger genome-wide context.

In conclusion, in recent years, multiple advances have been made to understand the relationship between DNA methylation and enhancer function, but most approaches either do not account for the epigenetic heterogeneity, are not focus on active CREs or give information only on single loci. Therefore, more research and new approaches are needed to fully understand whether DNA methylation is instructive for the activity of enhancers.

1.4 | Aims and design of this study

Epigenetic modification have been studied for multiple decades, continuously growing our understanding of their role in gene regulation (Mattei et al., 2022; Stillman, 2018). Yet, their direct impacts on the regulation of enhancers is still largely elusive. Enhancers are critical for the acquisition of cellular identities during development and their maintenance in healthy tissues. Thus, understanding the regulatory layers contributing to their activity is fundamental. A major challenge in investigating the regulatory contributions at active enhancers is the cell-to-cell variability observed for epigenetic modifications as well as for TF binding. This makes it difficult to determine direct correlations between those features.

In this PhD project, I wanted to address this fundamental question of gene regulation by developing novel genomic approaches to study epigenetic modifications, chromatin accessibility and TF binding at single molecule resolution *in vivo*. The main focus was laid on DNA methylation with the aim to resolve the epigenetic heterogeneity observed at active enhancers and to determine whether DNA methylation generally contributes to the regulation of enhancer activity and TF binding in the mammalian genome (Figure 9A). The second aim of this project was to measure the co-occupancy of histone modifications and TFs at CREs and to investigate their causal relationship (Figure 9B).

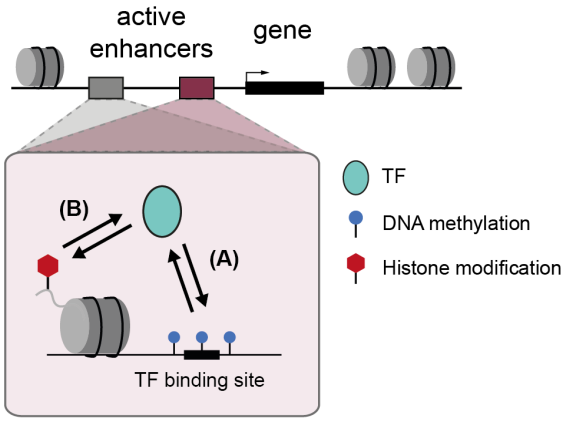

Figure 9: This PhD project aims to understand the causal relationship between transcription factors (TFs) and epigenetic modifications at active enhancers in the mammalian genome. (A) Is DNA methylation instructive for the binding of TFs and the activity of enhancers? **(B)** Do histone modifications have instructive roles for TFs and is this regulation factor specific?

1.4.1 | Design

To address the fundamental question of the causal effect of epigenetic modifications of TF binding and enhancer activity *in vivo*, I explored their co-occupancy with TF binding and chromatin accessibility. To achieve this, I advanced, for the main project, the Single molecule footprinting (SMF) methodology to co-measure TF binding, chromatin accessibility and DNA methylation at the same individual DNA molecules. To investigate DNA methylation in absence of perturbations, I leveraged the natural epigenetic heterogeneity at active enhancers to test the impact of different DNA methylation states on the chromatin accessibility and TF binding. For the second project, I developed a combined approach between SMF and chromatin enrichment technique CUT&RUN to enrich DNA molecules harboring a certain histone modification and to compare TF binding frequencies between the enriched and input DNA molecules. As a model system, I used murine cell lines, especially mESCs.

1.4.1.1 | mESCs as a model system

Mouse ESCs are a well-studied model cell line and widely used for DNA methylation studies. They are pluripotent cells derived from the ICM of E3.5 embryos in the blastocyst stage that have the ability to differentiate into both germline as well as any somatic cell type (Evans and Kaufman, 1981; Martin, 1981). In culture, mESCs are usually maintained under serum/LIF conditions. LIF is a trophic factor for the growth of ESCs that activates the TF STAT3, which inhibits differentiation and promotes self-renewal (Niwa et al., 1998; Smith et al., 1988). Studies have shown that homogeneous mESCs populations exhibit strong heterogeneity in their gene expression patterns (Chambers et al., 2007; Guo et al., 2016; Hayashi et al., 2008). This heterogeneity is less pronounced in mESCs cultured without serum but with two small-molecule kinase inhibitors (2i) (Ying et al., 2008): PD0325901, an inhibitor of mitogen-activated protein kinase (MAPK), and CHIR99021, an inhibitor of glycogen synthase kinase-3 (Gsk3). Mouse ESCs cultured in this 2i/LIF medium show a more naïve pluripotency state (Ying et al., 2008) with low DNA methylation levels that resemble the globally hypomethylated genome of ICM cells (Habibi et al., 2013; Leitch et al., 2013). In contrast, serum/LIF cultured mESCs exhibit methylation levels of 60-80% (Habibi et al., 2013; Leitch et al., 2013; Stadler et al., 2011) due to elevated levels of DNMT3A, DNMT3B and DNMT3L (Habibi et al., 2013), thus presenting methylation levels similar to somatic cell lines.

In this PhD project, I am using mESCs as a model system for multiple reasons. First, mESCs are the only cell line that is viable upon DNMT triple-knockouts (TKO) that leads to full depletion of global DNA methylation (Tsumura et al., 2006). This is associated with additional repressive mechanisms at transposable elements that inhibit their expression despite loss of DNA methylation (Greenberg and Bourc'his, 2019). In somatic cell lines, transposable elements are activated upon DNA methylation removal, potentially contributing to their cell death (Dahlet et al., 2020; Kaluscha et al., 2022; Sharif et al., 2016; Walsh et al., 1998). The second reason is that seemingly homogeneous mESC populations show heterogeneous DNA methylation and gene expression patterns (Singer et al., 2014), making them an interesting object to study the consequences of epigenetic heterogeneity. Lastly, due to their

pluripotent nature, mESCs can be differentiated into other cell lines such as neural progenitor (NP) cells, enabling the investigation of DNA methylation dynamics and their impact on enhancer activity.

1.4.1.2 | Single molecule footprinting as a tool for quantitative measurement of chromatin accessibility and TF binding

A new approach of detecting chromatin accessibility genome-wide has been developed that overcomes some of the limitations of bulk- as well as single-cell techniques: Single molecule footprinting (SMF). It has originally been developed on the small genome of *Drosophila melanogaster* cells (Krebs et al., 2017), but has recently been further advanced to the bigger genome of mouse cells by introducing an enrichment step for regulatory regions (Kleinendorst et al., 2021; Sönmezer et al., 2021). The technique is based on the labeling of accessible regions in the genome with high nucleotide resolution. For this, isolated nuclei are incubated with an exogenous GpC methyltransferase (M.CviPI) that methylates all GpC dinucleotides that are accessible in the moment of treatment (Figure 8C and F). Nucleosomes and proteins like TFs that are bound to the DNA leave unmethylated GpCs as footprints behind (Figure 8F). Reading out the GpC methylation pattern on a single molecule level by bisulfite sequencing, enables to detect chromatin accessibility and TF binding genome-wide.
In contrast to other chromatin accessibility or TF binding detection techniques, SMF does not introduce DNA breakage points, thus allowing to capture continuous accessibility and binding information over long stretches of DNA (Figure 8C and F). And while the high nucleotide resolution of SMF enables to capture short stretches of inaccessibility, such as TF occupancy, the high sequencing depth of SMF allows to detect also rare binding events. Moreover, it provides information about all chromatin accessibility states and binding events, thus resolving the epigenetic heterogeneity and providing quantitative information on chromatin accessibility and TF binding at single regulatory elements.

2 | Results

2.1 | Identifying the causal role of DNA methylation at active mammalian enhancers

DNA methylation in the CpG context (5mC) is a predominant epigenetic mark in the mammalian genome. While 70-80% of CpGs are methylated, two types of genomic regions show decreased levels of methylation (Figure 10A). First, promoters proximal to TSSs show almost no 5mC marks. Second, active enhancers show dynamic intermediate methylation levels of 10-60%, which has been correlated with TF occupancy at those distal CREs (Stadler et al., 2011). However, DNA methylation and TF binding can regulate each other: TFs can be sensitive to DNA methylation and TFs can induce local DNA demethylation (Héberlé and Bardet, 2019). This bidirectional relationship complicates the investigation of the causal impact of DNA methylation at enhancer elements. In addition, bulk-cell measurements fail to resolve the cell-to-cell variability observed at enhancers and indicate only intermediate methylation (Figure 10A). In consequence, this epigenetic heterogeneity has so far prevented the establishment of a causal relationship.

2.1.1 | Genome-wide quantification of DNA methylation and chromatin accessibility on single DNA molecules

To overcome these limitations and to investigate the causal impact of DNA methylation at enhancers, I advanced the SMF technology (Kleinendorst et al., 2021; Krebs et al., 2017; Sönmezer et al., 2021). I aimed to simultaneously measure chromatin accessibility (CA), the binding of TFs, as well as the methylation status at active enhancers on individual DNA molecules *in vivo* (Figure 10B). During SMF, the footprinting procedure with an exogenous GpC methyltransferase (M.CviPI) provides information about CA and TF binding events, by methylating only accessible GpC dinucleotides and leaving GpCs protected by TFs and nucleosomes unmethylated (Figure 10C and F). Thus, the SMF signal can be used as a proxy for the activity of CREs. Importantly, comparison of untreated and M.CviPI-treated samples shows that the endogenous 5mC in CpG context is maintained and unaffected by the footprinting and can be analysed in separation of the GpC context (Figure 10C-D).

Together with Rozemarijn Kleinendorst, I performed SMF in wild-type (WT) mESCs and used this data in addition to available SMF data sets in the Krebs laboratory. To analyse 5mC in relation to CA, I performed single molecule methylation calling separately for all CpGs and GpCs in the genome. Thereby, I considered only those CpGs that are in a NWCGW context (N = any base, W = not C/G), excluding e.g. CpGs in a CCG context, which has been shown to be an off-target of M.CviPI (Kelly et al., 2012) (Figure 10D). This context specificity would in principle allow to analyse 6.1 million CpGs, 30% of all CpGs genome-wide. For GpCs, I only considered the DGCHN context (D = not C, H = not G), accounting for ambiguities between the methylation calls (Figure 10D). To have sufficient coverage to

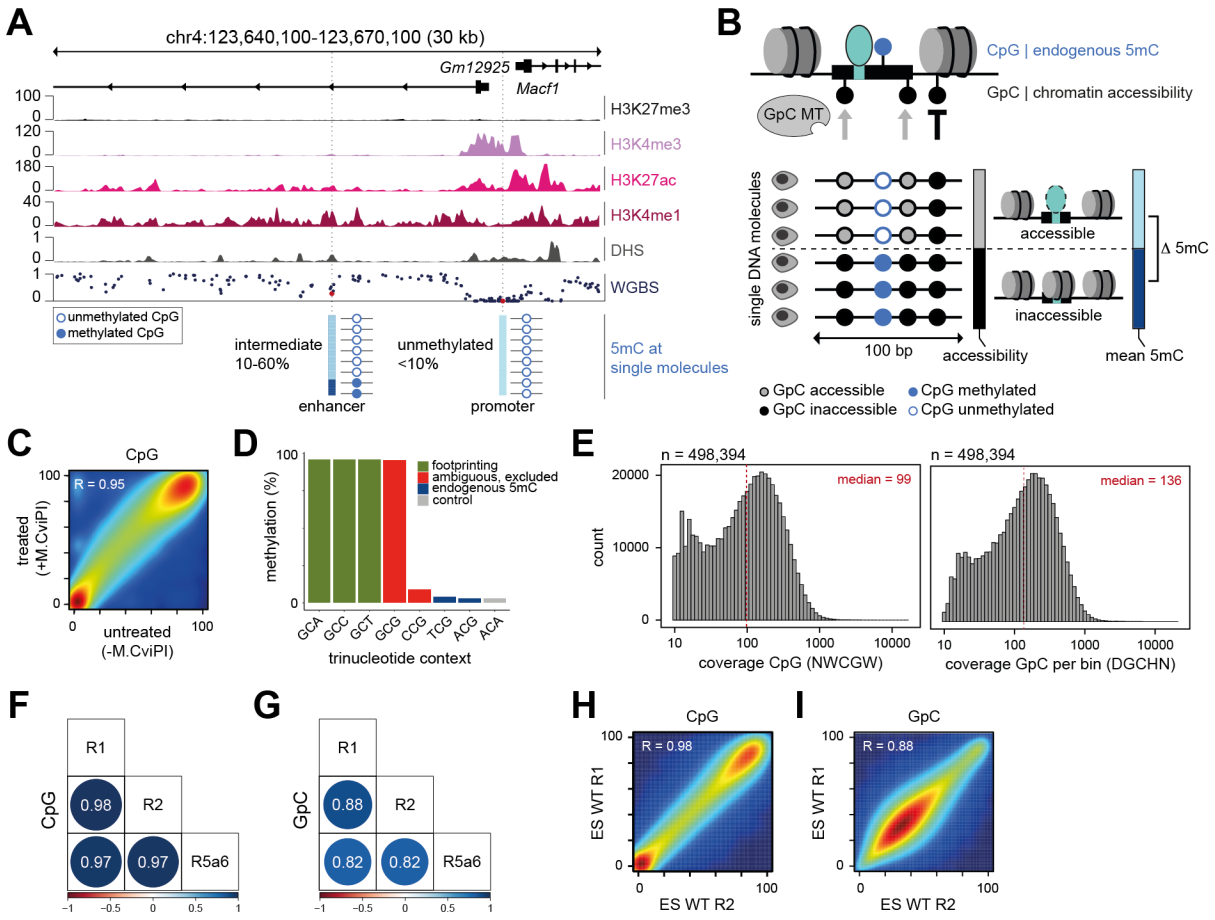

Figure 10: Single Molecule Footprinting approach simultaneously detects 5mC and chromatin accessibility. "**(A)** Single locus example of the heterogeneity in DNA methylation (5mC) at enhancers and promoters when resolved on individual molecules. Upper panel: Genome browser track displaying different chromatin marks and average 5mC as measured by Whole Genome Bisulfite Sequencing (WGBS) around the *Macf1* promoter and its intragenic enhancer. Lower panel: Single molecule representation of 5mC levels (blue - methylated, white - unmethylated) of individual CpGs (red dots in upper panel). 5mC at enhancers is heterogeneous. **(B)** Schematic representation of the experimental strategy used to identify molecular antagonisms between 5mC and chromatin accessibility (CA). Single molecule footprinting (SMF) is performed using the methyltransferase M.CviPI to measure CA in the GpC dinucleotide context, that is distinct from the endogenous 5mC in CpG context. Bisulfite sequencing of the DNA provides continuous information on CA and 5mC over 300 bp long DNA molecules. The accessibility of each molecule is calculated using the methylation of GpCs within the 100 bp surrounding a CpG. The average accessibility is used to classify DNA molecules into accessible and inaccessible fractions. In both fractions, the average 5mC levels are calculated. White/blue lollipops represent the endogenous un-/methylated state, white/black lollipops represent the CA as measured by SMF. **(C)** Treatment by M.CviPI does not interfere with endogenous methylation levels in NWCGW contexts. Smoothed scatter plot comparing the average CpG methylation as measured by WGBS in untreated mESC (data from (Stadler et al., 2011)), with the methylation upon treatment by M.CviPI in SMF data. The Pearson correlation coefficient is indicated. **(D)** Nucleotide contexts methylated by M.CviPI *in vitro.* Barplot representing the average methylation levels of lambda DNA in various trinucleotide contexts upon *in vitro* methylation with M.CviPI in saturating conditions. M.CviPI fully methylates GpCs regardless of the nucleotide context (green bars). It methylates CpGs in GC^{m}G contexts and displays a low level of non-specific activity in CC^{m}G context (<10%) (red bars). Both these contexts were excluded from our analysis, that only considered NWCGW and DGCHN to avoid technical interferences between the quantification of the DNA footprinting signal and the measure of endogenous 5mC levels. The CpGs in other contexts (blue bars) show levels comparable to those observed at non targeted cytosines representing technical noise of our assay (grey bar). Data from (Kleinendorst et al., 2021). **(E)** Distribution of read coverage of individual CpGs and 101 bp GpC bins. Histograms showing the distribution of read counts over CpGs (NWCGW context, left), and GpCs (DGCHN context) in 101 bp bins surrounding CpGs (right) considered for analysis. Red lines indicate the median coverage." (Kreibich et al., 2023) **(F-I)** Measure of (F, H) endogenous DNA methylation and (G, I) DNA footprinting in SMF data is highly reproducible across

(legend continued on next page)

three replicates (R1, R2 and R5a6). (F-G) Correlation matrix of average methylation levels in (F) NWCGW and (G) DGCHN context measured by bisulfite-sequencing between all independent biological replicates showing the Pearson correlation coefficient. (H-I) Smoothed scatter plot examples of (F) and (G) comparing two replicates R1 and R2. The Pearson correlation coefficient is indicated and encoded in the color code.
SMF experiments were performed together with Rozemarijn Kleinendorst. SMF data for R1 and R2 from (Sönmezer et al., 2021).

---

confidently call methylation, we included the bait-capture step in the SMF protocol (see chapter 4.1.2), which enriches for CREs, marking 4% of the mammalian genome. In addition, I applied a coverage threshold of 10x per CpG. In summary, the enrichment and filter steps resulted in 498,394 analyzable CpGs with a median coverage of 100x (Figure 10E). All three replicates showed high reproducibility in the CpG and GpC methylation calls (Figure 10F-I).

To investigate the relationship between 5mC and CA at CREs, I developed a SMF analysis pipeline that enables the co-analysis of the two measurements on the same DNA molecules (Figure 11A). To focus on CREs that are active in the cell line of interest, the first step selects CpGs that have a mean methylation level of 10-60% (Figure 11A – step 1, Figure 11B), which is a distinctive feature of active distal CREs. This methylation cutoff leaves around 100,000 CpGs to analyse. For my mESCs data, those CpGs are enriched at enhancer regions, based on publicly available chromHMM annotations (Figure 11C). In a second step, 101 bp wide bins are created around those intermediately methylated CpGs (Figure 11A – step 2). Using the mean GpC methylation within the bin, each molecule is then sorted based on its CA into an accessible and inaccessible fraction (Figure 11A – step 3). This is followed by the calculation of the mean 5mC of the center CpG in the two CA fractions separately (Figure 11A – step 4). By this, it can be determined whether a certain CA is associated with a certain level of 5mC. Performing the analysis on 102,629 CpG regions shows that on a global scale, both fractions of CA (accessible and inaccessible) are associated with intermediate levels of 5mC (Figure 11D - left), arguing that 5mC is globally not associated with CA. The same pattern can be observed when focusing only on regions annotated as enhancers by chromHMM (Figure 11D - right). Figure 11E shows a single locus example of such a region with no association between 5mC and CA that is localized within a DNase I hypersensitive (DHS) peak, representing high accessibility. The SMF data shows that 33% of the molecules are classified as accessible and that an equal proportion of molecules is methylated at the center CpG within the two CA fraction (28% in the accessible vs 27% in the inaccessible fraction). This data presents first evidence that the majority of CREs are not impacted by 5mC.

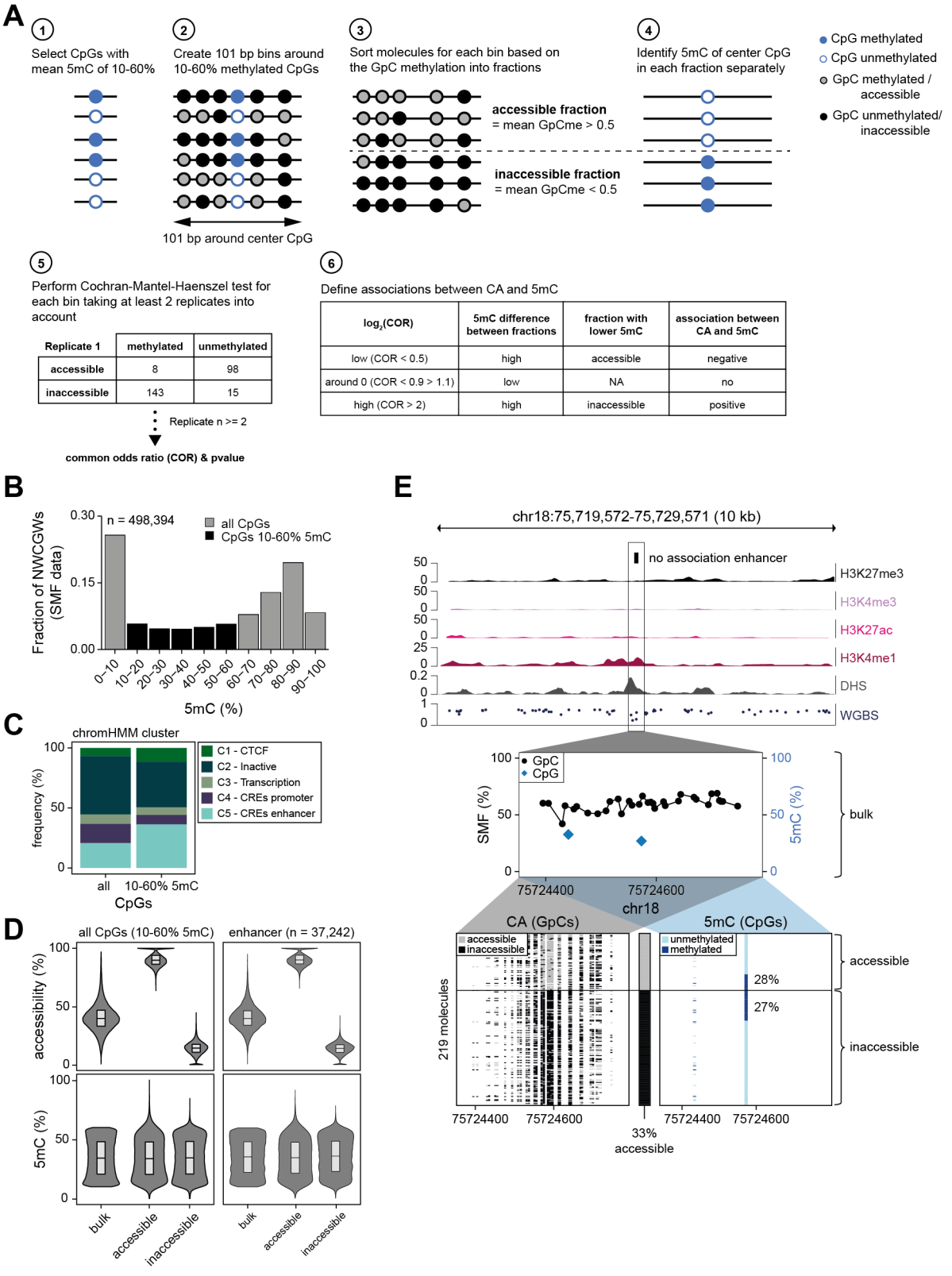

Figure 11: DNA methylation is predominantly neutral to chromatin accessibility at *cis*-regulatory elements. "(A) Schematic representation of the SMF analysis pipeline. **Step 1:** CpGs with mean intermediate (10-60%) DNA methylation (5mC) are selected genome-wide. **Step 2:** 101 bp bins are created around the selected CpGs. **Step 3:** For each bin, DNA molecules are sorted based on their average GpC methylation within the 101 bp bin into an accessible or inaccessible fraction. The mean GpC methylation cut-off for sorting is 0.5 (molecules with a mean GpC methylation of ≥ 50% within the bin are sorted into the accessible fraction). **Step 4:** For each molecule in the two fractions the 5mC

(legend continued on next page)

of the center CpG is identified (methylated/ unmethylated). **Step 5:** 3D contingency tables are created for each bin depicting for every replicate the counts of molecules that are accessible + methylated, inaccessible + methylated, accessible + unmethylated, and inaccessible + unmethylated. Any given bin has to be covered by at least two replicates. The 3D contingency tables are subsequently used as an input for the Cochran-Mantel-Haenszel (CMH) test, a proportionality test probing the association between 5mC and chromatin accessibility (CA) taking replicate information into account. The output of the test are a common odds ratio (COR) and a p-value. **Step 6:** COR and p-value are used to determine the categories of the 5mC-CA association. Bins with a low COR < 0.5 show high negative 5mC-CA association where the molecules in the accessible fraction have lower 5mC levels. Bins with a high COR > 2 show high positive 5mC-CA association where the molecules in the inaccessible fraction have lower 5mC levels. Bins with a COR around 1 show no 5mC-CA association, meaning that molecules in both accessibility fractions have similar levels of 5mC. **(B)** Comparison of CpGs distributions at different methylation states. Bar plots showing the distribution of CpGs (NWCGW context) at binned methylation states in SMF data in mESC. Colored in black are methylation states used for subsequent analyses (10-60%). **(C)** Intermediately methylated CpGs are enriched for CREs as enhancers. Stacked bar plot showing the genomic context annotation for all CpGs (NWCGW context) and intermediately methylated CpGs used for subsequent analyses using publicly available chromHMM annotations clustered into 5 groups (chromHMM data from (Pintacuda et al., 2017)). **(D)** Most intermediately methylated CpGs (left) and active enhancers (right) show no difference in 5mC between the two separated CA fractions. Violin bar plots showing the CA (top) and 5mC (bottom) at CpGs in bulk and in the two separated fractions (all CpGs n = 102,629). Boxplots show median (black middle line), 25th and 75th percentiles (black boundaries). **(E)** Single locus example of a CpG with no 5mC-CA association. Top panel: genome browser track displaying chromatin marks and average methylation as around an intergenic enhancer as shown in Figure 10A. Middle panel: average SMF signal (1 - methylation %) of individual GpCs (black dots) and endogenous 5mC of individual CpGs (blue diamonds). Bottom panels: single DNA molecules stack of genomic locus sorted by CA into accessible (grey) and inaccessible (black) fractions, and molecular 5mC in both fractions. Lower left panel: every column is a single GpC depicting its CA status (gray: accessible, black: inaccessible). Lower right panel: every column is a single CpG depicting its 5mC status (light blue: unmethylated, dark blue: methylated). Percentages represent average 5mC of the center CpG in each fraction. Number of single DNA molecules used is indicated." (Kreibich et al., 2023). Sarah Kaspar (Center for Statistical Data Analysis) advised on the statistical analysis and the CMH test.

2.1.2 | A small set of enhancers shows negative 5mC-CA association

Although globally 5mC and CA do not seem to be associated at active CREs, it could be possible that a fraction of active enhancers is regulated by 5mC. To identify those, I introduced the Cochran-Mantel-Haenszel (CMH) test into the analysis pipeline, which is a proportionality test that probes the statistical association between two given factors and takes replicate and coverage information into account (Figure 11A – step 5-6). Performing the CMH test at 102,400 intermediately methylated CpG regions with high coverage (> 30-fold) shows that CREs can have different degrees of 5mC-CA association (Figure 12A-B). While the largest percentage of sites show no 5mC-CA association, 3.2% of sites with a strong and significant negative 5mC-CA association are identified (Figure 12A-B). At those, the accessible molecules have a higher proportion of unmethylated molecules than the inaccessible molecules. This can be observed at a single locus example of this category where the enhancer is accessible in 45% of the molecules and where the proportion of methylated molecules in the inaccessible fraction is twice as high as in the accessible fraction (31% in the accessible vs 67% in the inaccessible fraction) (Figure 12C).

In addition, 2.4% of sites are identified having a strong positive 5mC-CA association, meaning that the accessible fraction is associated with high 5mC levels (Figure 12A-B). On a functional level this could mean that at those sites, the TFs preferentially bind their motifs in a methylated state. Analysis of chromHMM annotations shows that those positive association sites exhibit a reduced enrichment for enhancers, and are enriched for promoters, bivalent promoters in the inactive cluster, and insulator

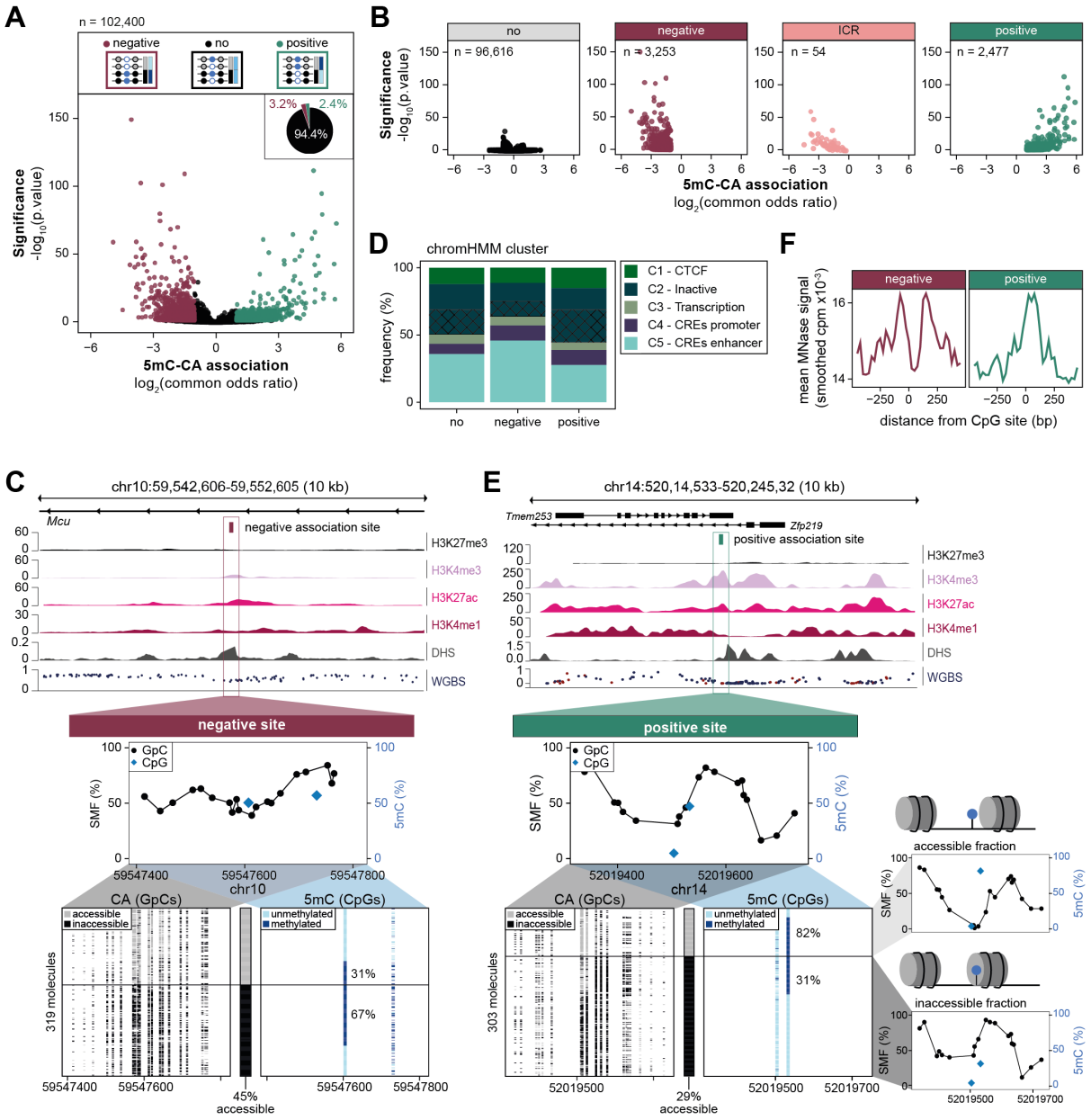

Figure 12: A subset of active enhancers shows negative 5mC-CA association "(A) Negative 5mC-CA association occurs at a small subset of CpGs genome-wide. Volcano plot depicting the common odds ratio (COR) and the p-value of a Cochran-Mantel-Haenszel test that probes the methylation difference between the chromatin accessible and inaccessible fraction of molecules covering 102,400 intermediately methylated CpGs. For most CpGs, 5mC occurs at equivalent levels between accessible and inaccessible molecules (black dots), yet 2,477 CpGs show a positive (cyan dots) and 3,298 CpGs show a negative 5mC-CA association (pink dots). The pie chart represents the relative proportion of CpGs in each category. The legend on top shows a schematic representation of the individual categories. **(B)** Most CpGs within annotated imprinting control regions (ICRs) show strong negative 5mC-CA associations. Volcano plots depicting the results of a Cochran-Mantel-Haenszel test as shown in (A) split by categories. A fraction of CpGs in ICRs show high methylation differences between accessible and inaccessible molecules and, thus, depicts a subclass of negative association sites (n = 45). Those are excluded from the negative association sites, leaving 3,253 non-ICR negative association CpGs for downstream analyses. **(C)** Single locus example of a negative 5mC-CA association site at the *Mcu* enhancer. Same representation as in Figure 11E. **(D)** The majority of CpGs with a negative 5mC-CA association are annotated as enhancers. CpGs with positive 5mC-CA association exhibit a reduced enrichment for enhancers and are enriched for active and bivalent (cross-striped C2 cluster) promoters. Stacked bar plot showing the genomic context annotation for sites with different association categories using publicly available chromHMM annotations clustered into 5 groups. **(E)** Positive 5mC-CA association sites are often located at the boundary of phased nucleosomes. In the two

(legend continued on next page)

positions of the nucleosome, 5mC is high at the linker region but not at the core of the nucleosome. Single locus example of a positive 5mC-CA association site at the *Zfp219* enhancer. Same representation as in Figure 11E. Lower right plots show SMF average plots of the individual fractions (top: accessible fraction, bottom: inaccessible fraction), which illustrate shifting of the nucleosomal footprint and increased 5mC of the CpG within the linker DNA between two nucleosomes. **(F)** Loci with positive 5mC-CA association are enriched with highly phased nucleosomes. Shown are density plots of mean MNase-seq signal (plotted as smoothed counts per million (cpm)) at 1 kb windows surrounding sites with negative (left, pink) and positive (right, cyan) 5mC-CA association." (Kreibich et al., 2023)

---

regions, when compared to all 10-60% methylated regions or those with negative 5mC-CA association (Figure 12D). Single locus analysis of this category indicates that the positive 5mC-CA associated CpGs are located at the border of strongly phased nucleosomes. An example region shows that the center CpG is either located within the nucleosome in the inaccessible fraction, or within the accessible linker region in the accessible fraction when the nucleosome is phased (Figure 12E). It has previously been shown that DNMTs are recruited at nucleosomes and strongly methylate the linker region between two nucleosomes (Xu et al., 2020). This DNMT recruitment could explain the high percentage of accessible molecules methylated at the center CpG (82% of accessible fraction), arguing that this is a secondary effect of the nucleosomal phasing. Analysis of MNase-seq data supports this hypothesis, showing a MNase peak located at positively associated CpGs, identifying strong nucleosomal phasing at those sites (Figure 12F, right). In contrast, negatively associated CpGs are depleted for nucleosomes at the CpG location (Figure 12F, left). In conclusion, these data argue against the hypothesis that positive CpG sites are CREs where TFs preferentially bind to methylated motifs and where 5mC positively regulates enhancer activation. The data rather suggests that the positive 5mC-CA association is a footprint of local chromatin organization around highly phased nucleosomes. Since the aim of this project is to determine a causal role of 5mC at active CREs, I focused all subsequent analyses only on sites with a negative 5mC-CA association.

Differential 5mC levels between accessible and inaccessible molecules at negative 5mC-CA associated sites can have two possible rationales (Figure 13A). First, the sites could be heterogeneously accessible and methylated within the cellular population, arguing for a partial usage of the CRE in association with the underlying 5mC. Second, the sites could be homogeneously accessible and methylated within the cellular population and the heterogeneity is rather driven by allelic differences within the individual cells. This is e.g. observed at ICRs, where 5mC confers monoallelic gene expression at the loci (Ferguson-Smith and Bourc'his, 2018). At those sites, the parental alleles show differential CA and differential 5mC so that one allele is accessible and unmethylated, while the other allele is inaccessible and methylated. Indeed, overlapping the analysed CpG sites with a manually curated list of annotated ICRs (see chapter 4.2.8) identifies 54 CpGs within ICRs (1.4% of all CpGs analysed), of which 83% show a negative 5mC-CA association (Figure 12B). Since CpGs within ICRs should be inherently picked up as negative 5mC-CA sites, this serves as a good positive control for my statistical analysis pipeline. However, it remains unclear whether this differential usage of alleles is explaining the heterogeneity of the remaining 3,253 sites that are located outside of annotated ICRs. To test this, I performed together with Rozemarijn Kleinendorst SMF in F1 hybrid mESCs (129/CAST) (Giorgetti et al., 2016), for which two distant mouse species have been crossed that each contribute one allele (Figure 13B).

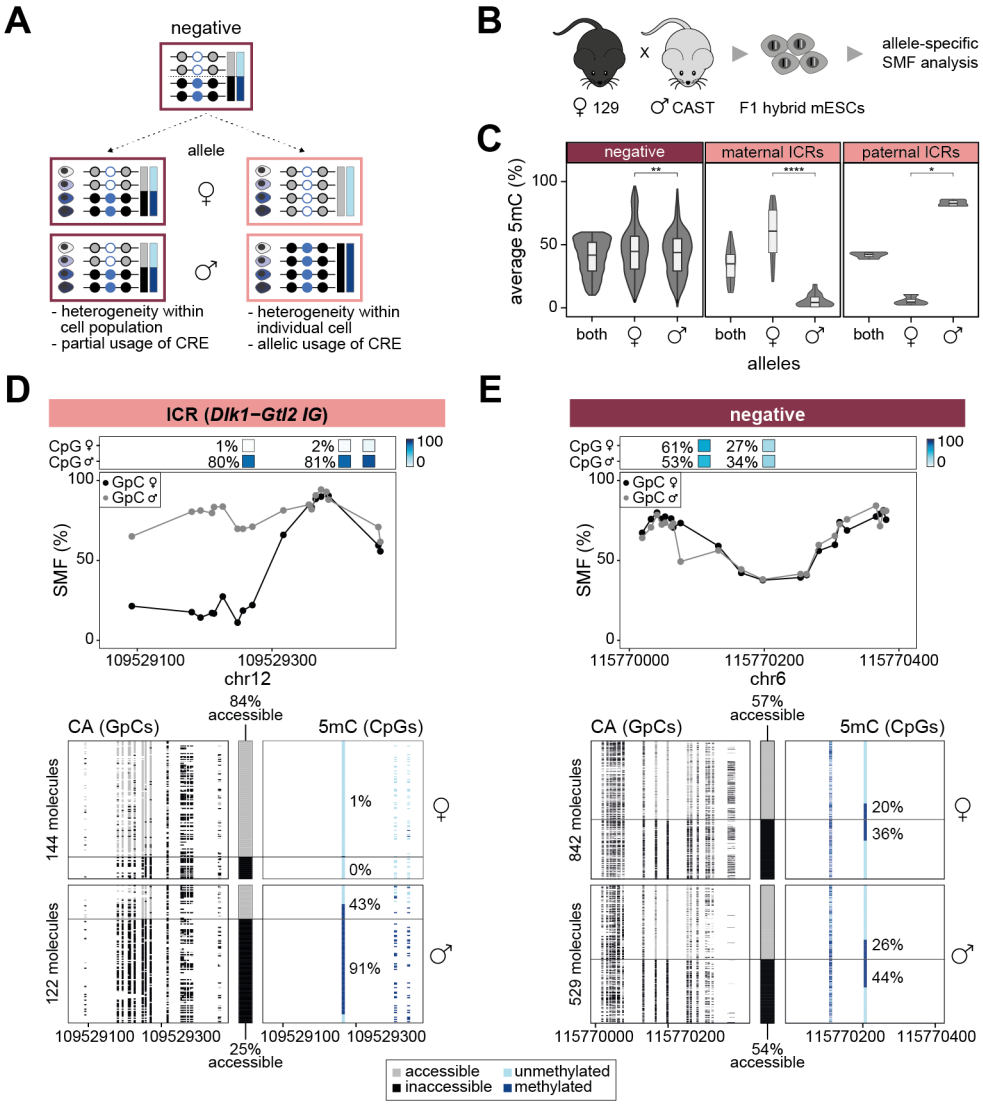

Figure 13: Sites with negative 5mC-CA association are distinct from ICRs.
(A) Schematic representation of the potential interpretations of the negative 5mC-CA association. Left: 5mC and CA levels are differential between cells, while alleles behave similarly. Thus, heterogeneity comes from within the cellular population and is the results of partial usage of CREs. Right: 5mC and CA levels are differential between alleles, while individual cells behave similarly. Thus, heterogeneity comes from within individual cells and is the results of allelic usage of CREs as shown for ICRs. "**(B)** Schematic depiction of F1 hybrid experiment. Mouse from two distant species (129 and CAST) were crossed and embryonic stem cells (ESCs) were derived. SMF was performed on these F1 hybrid mESCs and analysis was performed allele-specific using Single Nucleotide Polymorphisms. **(C)** CpGs with a negative 5mC-CA association are not differentially methylated between alleles. Violin box plots showing the average 5mC methylation distribution at CpGs with a negative 5mC-CA association (left), maternal ICRs (middle) and paternal ICRs (right). Average 5mC levels of individual CpGs were calculated by analysing both alleles together (both), or each allele of a F1 hybrid (129/CAST) mESC line individually. Number of stars illustrates the significance of a Wilcoxon rank test when comparing the two alleles (*p ≤ 0.05; **p ≤ 0.01 ****p ≤ 0.0001). **(D-E)** Single loci exemplifying the difference between an annotated ICR at the *Dlk1-Gtl2* imprinted locus (D) and a newly identified locus with negative 5mC-CA association (E) at allelic resolution. SMF was performed in F1 hybrid (129/CAST) mESCs and CA and 5mC were analysed for both alleles separately. Top panels depict average SMF signals (1 - methylation %) of individual GpCs in the maternal (black dots) or paternal (grey dots) allele. On top, location and methylation status of the CpGs in the two alleles are shown. Lower panels show single DNA molecules for the two alleles separately (top: maternal, bottom: paternal) sorted by CA into an accessible (grey) and inaccessible (black) fraction. Same representation as in Figure 11E. The number of DNA molecules analysed for each allele at both loci is indicated." (Kreibich et al., 2023)
SMF experiments were performed together with Rozemarijn Kleinendorst.

By identifying Single Nucleotide Polymorphisms (SNPs), the two homologous alleles can be computationally separated and 5mC and CA resolved on the individual alleles. Analysing the average 5mC levels of the two alleles at negative 5mC-CA associated sites within or outside of ICRs shows clear differences between the two groups. While the CpGs within maternal or paternal ICRs show allele-specific 5mC levels, the CpGs outside of ICRs show similar 5mC levels at the two alleles (Figure 13C). This is also visible at single locus examples of the two groups. Allele-specific SMF analysis distinctly shows that the CpG region within the ICR *Dlk1-Gtl2 IG* is fully accessible (84% CA) and fully unmethylated (1% 5mC) at the maternal allele, and fully inaccessible (25% CA) and strongly methylated (80% 5mC) at the paternal allele (Figure 13D). This stands in contrast to a CpG site outside of an annotated ICR, where both alleles are similarly accessible (57% and 54% CA) and similarly methylated (27% and 34% 5mC at center CpG) and where both alleles show a negative 5mC-CA association (maternal allele: 20% 5mC in accessible fraction vs 36% 5mC in inaccessible fraction; paternal allele: 26% 5mC in accessible fraction vs 44% 5mC in inaccessible fraction) (Figure 13E). In conclusion, this data argues, that the 3,253 CpG sites outside of ICRs do not represent a new group of ICRs with differential allelic usage of CREs, but that the differential CA and 5mC is related to cellular heterogeneity due to differential usage of CREs within individual cells.

Having identified CpG sites with a negative 5mC-CA association as a distinct group of potentially 5mC-regulated CREs, I set out to further characterize and contrast them to sites with no 5mC-CA association. For this, I excluded CpG sites annotated as inactive or insulators (CTCF) in order to focus on active CREs (see chapter 4.2.14). Furthermore, I used a more stringent set of the no association sites (common odds ratio between 0.9 and 1.1) to account for false negative sites (see chapter 4.2.3). ChromHMM analysis shows that negative 5mC-CA sites are strongly enriched for CREs and especially enhancers (Figure 12D). This enhancer enrichment is supported by the analysis of different chromatin modifications showing that they are enriched for activating marks such as H3K27ac, and H3K4me1, and show high DNase I hypersensitive sites (DHS) signal (Figure 14A). Moreover, there is no enrichment for H3K27me3, arguing that they are not repressed by Polycomb complexes and are therefore different from poised enhancers. In addition, a fraction of sites is bound by CTCF.
Both site groups with either no or negative 5mC-CA association have a low CpG density (median no = 1.75 CpGs/100 bp, median negative = 2 CpGs/100 bp) (Figure 14B), which separates them from CGIs. Since bisulfite sequencing does not differentiate between 5mC and 5-hydroxymethylation (5hmC), it can be questioned whether negative 5mC-CA associated sites are regulated by 5hmC in comparison to no association sites. Data from hydroxymethylated DNA immunoprecipitation (hMeDIP) however, shows that sites in both categories have comparable elevated levels of 5hmC, arguing against any specific regulation by this mark at negative 5mC-CA associated sites (Figure 14C). Yet, two features differentiate the two categories. First, CpG sites with a negative association are on average slightly more accessible (Figure 14D). And second, they are more frequently located at the peak summit of DHS regions, where most of the TF binding is presumed to occur (Figure 14E). This stands in contrast to no association sites that are distributed across the whole DHS regions (Figure 14E). “This data suggests that while CpGs from both categories are found at active CREs, CpGs with a negative 5mC-CA association are mostly located at their center where CA is the highest” (Kreibich et al., 2023).

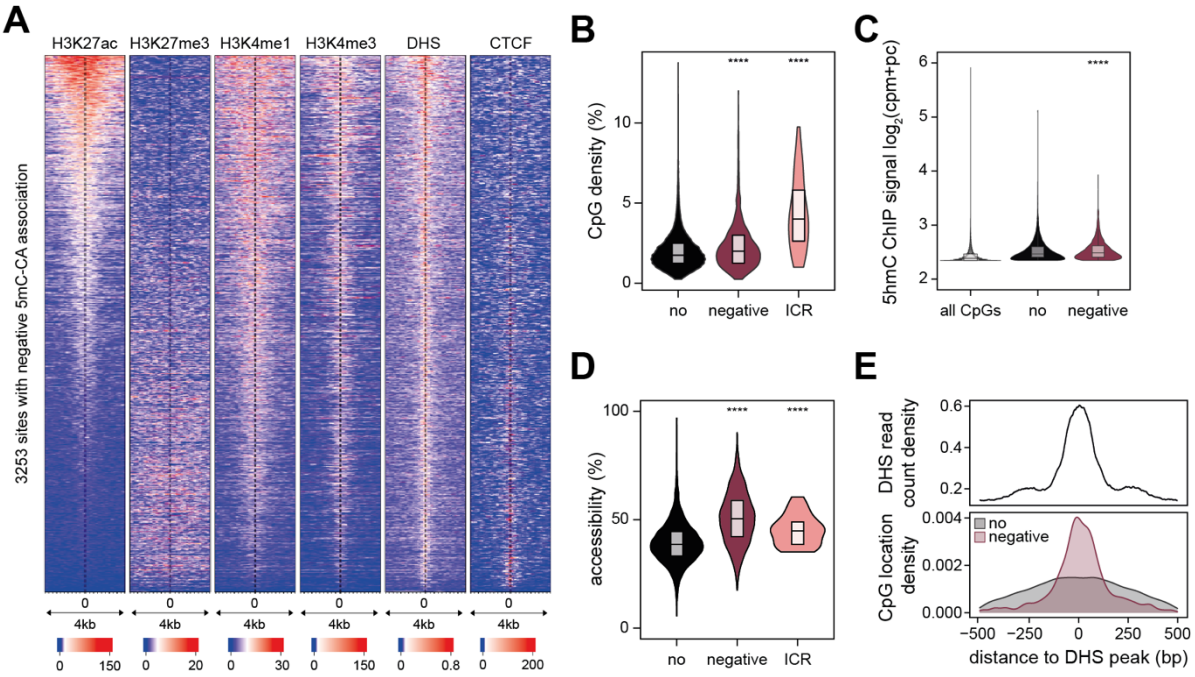

Figure 14: CpGs with negative 5mC-CA association are enriched at peaks of chromatin accessibility. "(A) Sites with negative 5mC-CA association harbor characteristic chromatin marks of enhancers. Heatmap displaying the ChIP-seq counts for H3K27ac, H3K27me3, H3K4me1, H3K4me3, CTCF as well as chromatin accessibility as measured as DNase I hypersensitive sites (DHS) by DNase-seq from publicly available data in a 4 kb window centered on 3,253 CpG sites with negative 5mC-CA association. **(B)** Sites with no and negative 5mC-CA association have low CpG densities. Violin box plots showing the CpG density distribution for each category (with no 5mC-CA association (n = 23,856), negative 5mC-CA association (n = 3,253) and ICR sites (n = 54)), as defined by the number of CpGs per 100 bp in a 400 bp window. **(C)** CpGs with divergent 5mC-CA association show similar levels of 5-hydroxymethylation (5hmC). Violin box plots showing the distribution of 5hmC signal measured by hMeDIP-seq. Plotted is the log2 of hMeDIP signal as counts per millions in a 500 bp window. **(D)** Sites with divergent 5mC-CA association show intermediate chromatin accessibility (CA). Violin box plots showing the distribution of CA as measured by SMF in a 100 bp window. **(E)** CpGs with a negative 5mC-CA association are frequently located at the center of active *cis*-regulatory elements. Lower panel: relative position of CpGs with no (black) or a negative 5mC-CA association (pink) to the peak of accessibility of the regulatory element (summit of DHS). Top panel: average DHS signal for the same regions shown as reference. Stars in this figure illustrate the significance of a Wilcoxon rank test (****$p \leq 0.0001$)." (Kreibich et al., 2023)

In conclusion, the development of my statistical SMF analysis pipeline enabled me to identify a class of active enhancers, where 5mC and CA are negatively associated. This negative association suggests a potential functional role for 5mC at these CREs.

2.1.3 | DNA methylation controls the chromatin accessibility of enhancers with negative 5mC-CA association

The statistical analysis of the SMF data suggests that a small set of enhancers are regulated by 5mC. If this hypothesis is true, the activity of those enhancers should change upon perturbations of the underlying 5mC. To test this, I performed, together with Rozemarijn Kleinendorst, global perturbation experiments followed by SMF and analysed the changes in CA upon removal or increase in 5mC.
In a first step, all three DNMTs (DNMT1, DNMT3A/B) active in mESCs have been knocked-out (DNMT TKO), resulting in cells fully devoid of 5mC. Although the different categories of enhancers sites have

comparable 5mC levels in WT cells (Figure 15A), they respond differently to the DNMT knock-out. In this analysis, I focused on sites with a 5mC of at least 30% in WT cells to be able to detect changes in the CA, and excluded sites annotated as inactive or insulator (see chapter 4.2.14). While sites with no 5mC-CA association show almost no change in their CA, negative 5mC-CA association sites as well as sites within ICRs significantly gain CA (CA = 1-SMF) upon 5mC removal (Figure 15B). This can also be observed for two representative single locus examples. The *Mpp7* enhancer, for which no 5mC-CA association has been detected, shows similar CA upon loss of 5mC (Figure 15C). In contrast, the *Mcu* enhancer, for which a negative 5mC-CA association has been detected, the loss of 5mC results in a strong increase in CA (loss in SMF signal) (Figure 15D).
To investigate the impact of an increase in 5mC instead, all three TET enzymes were knocked-out (TET TKO). This led to a local increase in 5mC in a subset of CpGs, while global 5mC levels barely changed (Figure 15E). In consequence, I focused my analysis on sites with an increase in 5mC of at least 30%. Also in those TET TKO cells, one can observe differences between the enhancer categories. Again, the sites with no 5mC-CA association show only limited changes in their CA (CA = 1-SMF) upon increase in 5mC (Figure 15F). In contrast, sites with a negative association show a significant loss of CA. Looking at the same single locus examples as before, one can observe that at the *Mpp7* enhancer without association, a 5mC increase of 44% (from 27% in WT to 71% in TET TKO) at the center CpG has no impact on the CA of this region (Figure 15C). Yet, the *Mcu* enhancer with a negative association shows a loss in CA upon an 5mC increase of 36% (from 50% in WT to 86% in TET TKO) (Figure 15D).
Interestingly, closer examination of the data reveals that the changes in CA scale with the changes of the underlying 5mC at sites with a negative 5mC-CA association (Figure 15G and H – purple), but not at sites with no or with a positive association (Figure 15G and H – cyan and black). This scaling effect is particularly strong in TET TKOs (R = 0.369), but is also measurable in DNMT TKOs (R = -0.234). This hints towards direct regulation of the negatively associated enhancers' activity by 5mC.

The ICRs serve as a positive control in these perturbation experiments. Due to their known regulation by 5mC, it is expected that they gain CA upon TKO of the DNMTs. In this case, the naturally inactive allele becomes activated and gains CA, while the active allele stays active. This CA gain can be observed in my data as a decrease in SMF signal (Figure 15C). Interestingly, in the TET TKO cells, I observe a strong loss in CA (Figure 15H). However, this loss is rather surprising, since regulation of ICRs by TET enzymes has only been shown in limited cases. Indeed, analysis of the changes in 5mC and CA at individual ICRs reveals that only one ICR, namely *Dlk1-Gtl2 IG*, gains 5mC and loses CA in the TET TKOs (Figure 16A), which can be observed over the whole ICR locus (Figure 16B). This observation is in line with previous studies reporting that TET enzymes regulate the activity of the active allele at the *Dlk1-Gtl2 IG* locus (Aronson et al., 2021).

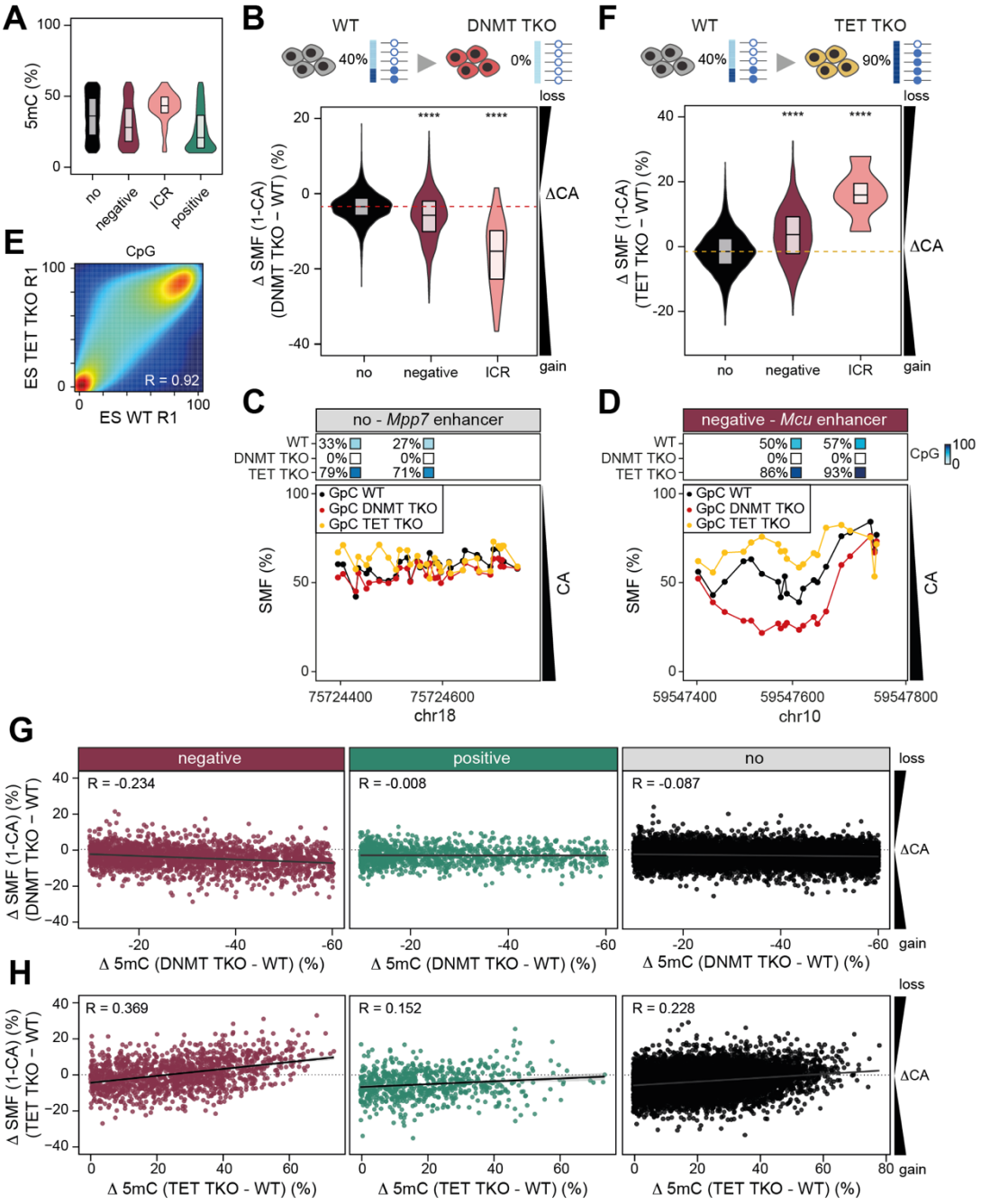

Figure 15: "DNA methylation represses the regulatory activity of enhancers with negative 5mC-CA association. **(A)** Sites without and with negative 5mC-CA association show comparable intermediate DNA methylation (5mC) levels in wild type (WT) mESCs measured by SMF. Violin box plots showing the distribution of average 5mC of sites without (black, stringent no sites, n = 13,050), with negative (pink, n = 2,074) or positive (cyan, n = 1,117) 5mC-CA association or in ICRs (rose, n = 32). **(B)** Chromatin accessibility (CA) is increased upon depletion of 5mC at enhancers with a negative 5mC-CA association. Violin box plots representing the changes in CA as measured by SMF (1-CA) at a 101 bp window upon removal of 5mC in DNMT triple-knockout (DNMT TKO) mESCs. CA is significantly increased at sites with a negative 5mC-CA association (pink, n = 963) and in ICRs (rose, n = 31), but not at sites without 5mC-CA association (black, stringent no sites, n = 7,956). Sites with 5mC changes <30%, or annotated as insulator or inactive, were excluded. Red line depicts the median of "no" sites.**(C-D)** Single locus examples of the CA differences between WT, DNMT TKO and TET TKO at sites with (C) no or (D) negative 5mC-CA association. Top panels: location and methylation status of the CpGs in the three cell lines. Bottom panels: average SMF signal (1-methylation%) of individual GpCs in WT (black dots), DNMT TKO (red dots), and TET TKO (yellow dots) mESCs. **(E)** Global gain of 5mC in TET TKOs for CpGs with intermediate 5mC levels. Smoothed density scatter plot comparing the average 5mC levels in NWCGW context

(legend continued on next page)

measured by bisulfite sequencing in SMF in WT mESCs and an isogenic TET TKO line. The Pearson correlation coefficient is indicated.**(F)** CA is reduced upon increase of 5mC at enhancers with negative 5mC-CA association. Violin box plots representing the changes in CA as measured by SMF upon increase of 5mC in mESC knocked out for the three TET enzymes (TET TKO). CA is significantly reduced at sites with negative 5mC-CA association (pink, n = 539) and in ICRs (rose, n= 6), but not at sites without 5mC-CA association (black, stringent no sites, n = 2,675). Loci where the 5mC increase is less than 30% in TET TKO cells were excluded from the analysis, as well as sites annotated as insulator or inactive by chromHMM. Yellow line depicts the median of "no" sites. Same representation as (B). **(G-H)** CA is a function of 5mC levels at sites with negative 5mC-CA association. Scatterplots comparing the changes in CA (shown as SMF = 1-CA) and 5mC in (G) DNMT TKO (negative sites n = 2,074, positive sites n = 1,117, no association sites n = 13,050) and (H) TET TKO (negative sites n = 1,348, positive sites n = 693, no association sites n = 9,713) for negative 5mC-CA association sites. The decrease or increase in CA is anti-correlated with the increase in 5mC at the locus in TET TKOs and DNMT TKOs, respectively. The regression lines and Pearson correlation coefficients are indicated.
All no association sites shown in this figure are of high confidence with minor methylation difference between fractions (common odds ratio between 0.9 and 1.1). Inactive sites were filtered out for all categories (see methods). Box plots show median (black middle line), 25th and 75th percentiles (black boundaries). The number of stars illustrate the significance of Wilcoxon rank tests when compared to the "no" sites ($p > 0.05$; **$p \leq 0.01$; ****$p \leq 0.0001$)." (Kreibich et al., 2023) CRISPR knock-outs and SMF experiments were performed together with Rozemarijn Kleinendorst.

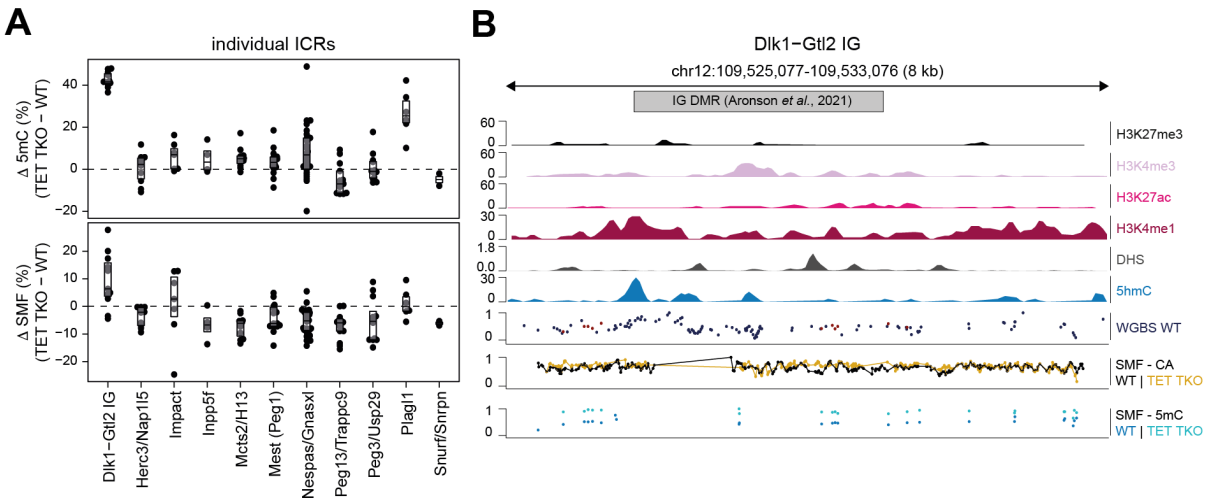

Figure 15: *Dlk1-Gtl2 IG* ICR responds to TET triple-knockout. "(A) An increase in DNA methylation (5mC) and loss of chromatin accessibility (CA) in TET TKOs is specific to the *Dlk1-Gtl2* ICR. Point box plots show changes in 5mC (top panel) and changes in CA (shown as SMF = 1-CA, bottom panel) between WT and TET TKO mESCs for NWCGWs in individual ICRs. Boxplots show median (black middle line), 25th and 75th percentiles (black boundaries). **(B)** The entire *Dlk1-Gtl2 IG* ICR locus shows an increase in 5mC and loss in CA in TET TKOs. Genome browser track of the entire *Dlk1-Gtl2 IG* ICR showing multiple chromatin marks, hydroxy-5mC (5hmC) and 5mC measured by Whole Genome Bisulfite Sequencing (WGBS) in WT mESCs, and CA (SMF = 1-CA) and 5mC measured by SMF in WT and TET TKO mESCs. Grey box shows intergenic differentially methylated region (IG DMR) previously annotated as regulated by TET enzymes (Aronson et al., 2021)." (Kreibich et al., 2023)

2.1.4 | DNA methylation controls the activity of enhancers with negative 5mC-CA association

To investigate the functional consequences of the 5mC perturbation on enhancer activity, I aimed for a more direct read-out of enhancer activity in contrast to the so far used CA as a proxy. Thus, I decided to perform Precision Run-On Sequencing (PRO-seq) that enables the direct measurement of the enhancer activity by measuring the nascent RNA transcribed from genes as well as from enhancers (enhancer RNA). I performed PRO-seq as a 2 biotin-run on experiments in WT, DNMT TKO and TET TKO mESCs and used DESeq2 to analyse differential activities between the mutant cell lines. Comparable density analysis shows that the replicates behave very similarly, while the mutants differ from each other in their signal around TSSs and enhancer sites (Figure 17A-B), arguing for a good quality of the data.
Analysis of the changes in PRO-seq signal upon changes in 5mC shows concordant results with the SMF data. Loss of 5mC in DNMT TKOs leads to a modest, but significant increase in PRO-seq signal at sites with negative 5mC-CA association, an effect that is even stronger at sites within ICRs (Figure 17C). Sites without association, in contrast, show no change in the PRO-seq signal (Figure 17C). In TET TKOs, a gain of 5mC by at least 40% results in a significant decrease in PRO-seq signal at negative association sites in comparison to the no association sites (Figure 17D). This is also apparent at two single locus examples that show changes in PRO-seq signal in the different cell lines at the negative but not at the no association enhancer site (Figure 17E-F). In conclusion, the direct measurement of enhancer activity by PRO-seq further supports the theory that 5mC regulates the activity of enhancers at a small subset of enhancers where a negative 5mC-CA association has been detected.

To answer the question whether the effects at enhancers can be translated into transcriptional changes at genes, I analysed, with the help of Guido Barzaghi, publicly available RNA-seq data of WT, DNMT TKO and TET TKO mESCs. This analysis requires the establishment of gene-enhancer associations to correctly relate the changes in promoter activity to changes in enhancer activity. However, current association tools are still lacking confidence in calling promoter-enhancer contacts, which is even more complicated by the fact that multiple enhancers can impact the activity of a gene to different degrees. For my analysis, I used the Genomic Regions Enrichment of Annotations Tool (GREAT) to assign genes to my list of enhancer regions and analysed the changes in RNA-seq signal at TSSs upon changes in 5mC at the enhancers. This resulted in anecdotal, but non-significant evidence that genes with significant transcriptional changes are coupled to negative association enhancers (Figure 18A). An example for this can be found at chromosome 11, where a negative enhancer is associated with two genes, GLI family zinc finger 2 (*Gli2*) and inhibin subunit beta B (*Inhbb*), at a distance of 114 and 255 kb from their TSSs, respectively. The enhancer shows a strong increase in 5mC (12% in WT to 81% in TET TKO) and a concomitant loss in CA (CA=1-SMF, 70% in WT to 44% in TET TKO) in TET TKOs (Figure 18B). This loss in the enhancer activity is further supported by PRO-seq data (Figure 18B). In concordance, the genes associated with this enhancer based on GREAT, both show a significant loss in RNA-seq signal (Figure 18B).

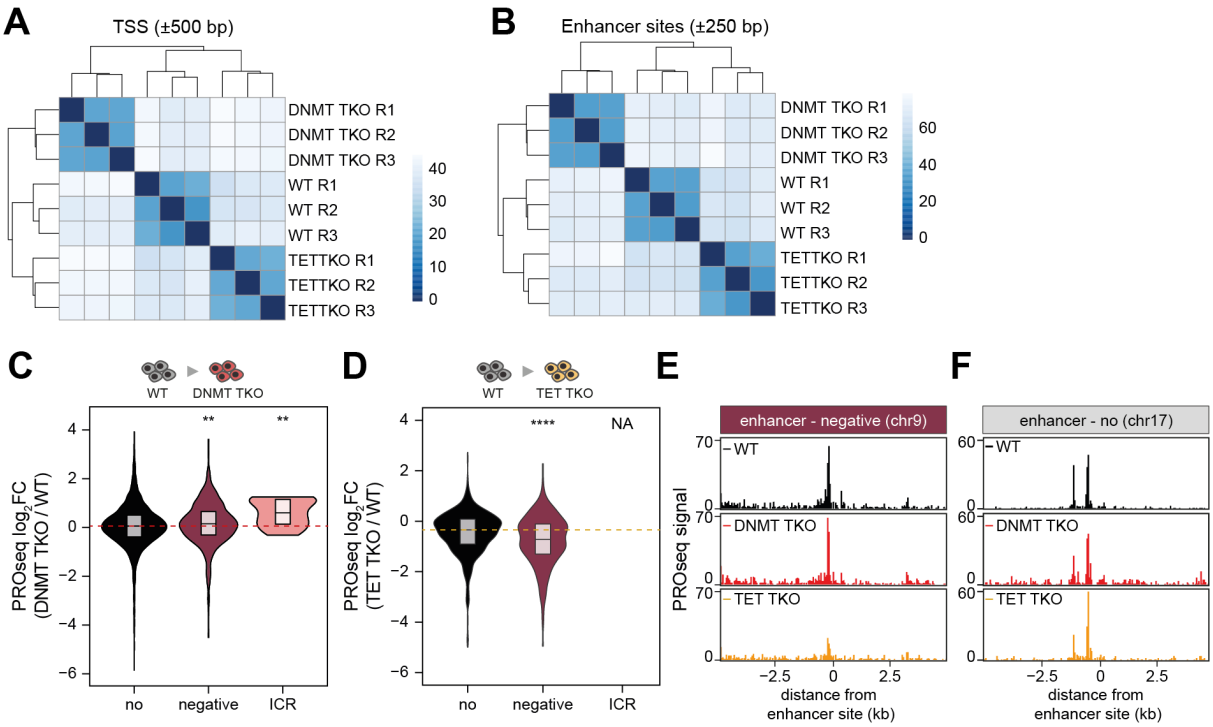

Figure 16: The activity of enhancers with negative 5mC-CA association is impacted by changes in DNA methylation. "(A-B) PRO-seq data shows reproducible changes in RNA Polymerase II activity at transcriptional start sites (TSS) and enhancers. Heatmaps showing the signal distance between all samples at (A) TSS (± 500 bp) and (B) enhancer bins (± 250 bp, using all 10-60% methylated CpG bins excluding inactive regions) as determined by DESeq2. **(C)** Enhancer activity is moderately but significantly increased upon loss of 5mC >40% at sites with negative 5mC-CA association (pink, n = 535) and ICRs (rose, n = 19) but not at sites without association (black, n = 2,675). Violin box plot showing the log2 fold change (FC) between PRO-seq signal at enhancers in the different categories between WT and DNMT TKO mESCs. Horizontal line and statistical test as described in Figure 15B. **(D)** Enhancer activity is significantly reduced upon increase of 5mC >40% at sites with negative 5mC-CA association (pink, n = 251) but not at sites without association (black, n = 964) (ICRs, rose, n = 0). Same representation as in (C). **(E-F)** Example loci of enhancers (E) with negative or (F) without 5mC-CA association showing the PRO-seq signal in WT (black), DNMT TKO (red) and TET TKO (yellow) at a 5 kb window around the enhancer site." (Kreibich et al., 2023)

In addition, a previous study has shown that the enhancer cluster controlling the expression of the pluripotent factor Sox2 in mESCs is regulated by 5mC (Song et al., 2019). Analysing WT SMF data at the *Sox2* enhancer cluster indeed identifies two CpG sites with negative 5mC-CA association (Figure 18C). This supports the former observation and confirms the sensitivity of my single molecule approach to identify enhancer sites that are potentially regulated by 5mC.

In conclusion, my single molecule approach enables the detection of 5mC sensitive enhancers genome-wide in mESCs. Global 5mC perturbation assays reveal that enhancers for which a negative 5mC-CA association has been detected show changes in their CA measured by SMF as well as in their enhancer activity measured by PRO-seq concordant with the changes of the underlying 5mC levels. Analysis of RNA-seq data provides anecdotal evidence that those alterations of the enhancer activity are translated into altered expression levels of associated genes.

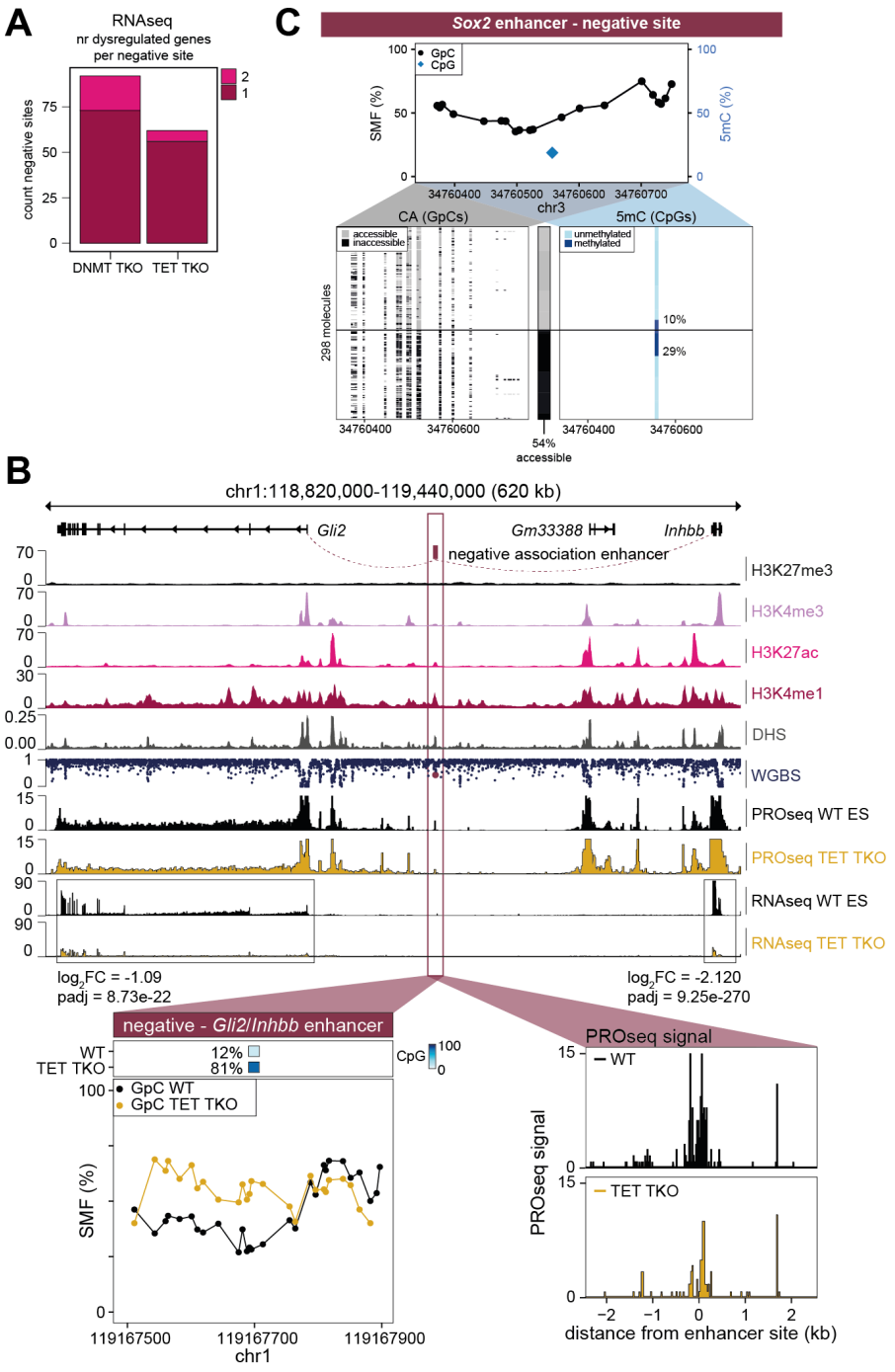

Figure 17: RNA-seq shows anecdotal evidence for impact of 5mC on transcription. "(A) Some enhancers with negative 5mC-CA association are associated with multiple dysregulated genes. Bar plot showing the number of enhancers with negative 5mC-CA association (y-axis) that are annotated to multiple dysregulated genes in RNA-seq (color code) (upregulated in DNMT TKO, downregulated in TET TKO) (RNA-seq data from (Domcke et al., 2015; Huang et al., 2021)). **(B)** Example of a genomic locus where loss of CA at the *Gli2*/*Inhbb* enhancer with negative 5mC-CA association is correlated with a decrease in gene expression at the connected genes upon 5mC increase. Top panel: genome browser track (representation as in Figure 11E) including PRO-seq signal and RNA-seq signal in WT and TET TKO mESCs around an intergenic enhancer with negative 5mC-CA association. The enhancer was linked to the *Gli2* and *Inhbb* genes by GREAT (pink dotted lines). The log2 fold-change and adjusted p-value of the expression changes, as measured by RNA-seq, are displayed. Left bottom panel: CA changes at the *Gli2*/*Inhbb* enhancer in WT versus TET TKO. Same representation as Figure 16D. Right bottom panel: enhancer activity changes as measured by PRO-seq at a 5 kb window around the *Gli2*/*Inhbb* enhancer in WT versus TET TKO.
(C) Identification of two CpGs with negative 5mC-CA association at the *Sox2* enhancer cluster. Single locus example of the 5mC and CA of individual molecules at one of the negatively associated CpGs contained within the *Sox2* enhancer cluster. Representation as in Figure 11E. RNA-seq analysis with performed with the help of Guido Barzaghi.

2.1.5 | Enhancers with negative 5mC-CA association are cell-type specific

Cell types are defined by their spatiotemporal gene expression patterns and their differential activity of promoters and enhancers. Having identified enhancers with negative 5mC-CA association in mESCs raises the question whether those enhancers are also active in other cell types. For this, Rozemarijn Kleinendorst *in vitro* differentiated mESCs into neural progenitor cells (NPs) and generated SMF data from them. Comparison of the 5mC and CA levels at enhancers analysed in mESCs revealed that the ES enhancers with negative 5mC-CA association show a strong increase in 5mC upon differentiation, which is concurrent with a significant loss in CA (Figure 19A-B). Sites with no or with a positive 5mC-CA association also show an increase in 5mC, but only weak changes in CA, while no changes are observed at sites within ICRs (Figure 19A-B). Furthermore, analysis of the changes in 5mC in relation to the changes in CA reveals a strong correlation (R = 0.443) of the two measurements at enhancers with negative 5mC-CA association, arguing that at those sites, the loss in CA is directly linked to the increase in 5mC (Figure 19C). This correlation is not observed at the other sites with no (R = 0.241) or positive (R = -0.046) 5mC-CA association (Figure 19C). These results support the idea that negative association enhancers identified in mESCs are only active in this cell type and become inactivated upon differentiation. This data is further supported by a single locus example of an intragenic enhancer of the *Npepps* gene that is active (CA = 54%), intermediately methylated (mean 5mC = 16%) and with a negative 5mC-CA association in ES cells (5mC in accessible fraction = 3%; 5mC in inaccessible fraction = 32%) (Figure 19E). This enhancer becomes inactive (CA = 23%) and highly methylated (mean 5mC = 73%) in NPs.

In order to generally identify the fate of all ES negative 5mC-CA association sites, I performed my statistical analysis on the NP SMF data and observed that the majority of the sites (63%) gain methylation above 60% and, thus, become inactivated, while others (30%) keep an intermediate methylation state but lose their negative association (Figure 19D). A small percentage of sites either lose 5mC (3%) or keep their negative 5mC-CA association (4%) (Figure 19D).

To collect further evidence for a cell type specific features of negative 5mC-CA association sites, I performed gene ontology (GO) term enrichment analysis using GREAT for this class of ES enhancers and identified that they are related to genes with cell type specific functions such as regulation of stem cell population maintenance (Figure 20F). In summary, comparing the ES and NP SMF data reveals that sites with negative 5mC-CA association in ES cells are a class of *bona fide* ES-specific enhancer that get inactivated upon differentiation into the neural lineage.

During early embryonic development into the blastocyst stage, the cells undergo a massive wave of demethylation that removes all genomic 5mC except at ICRs, retroviral elements and transposable elements. Mouse ES cells are derived from the inner cell mass of the blastocyst during this hypomethylated state. Yet, cultured *in vitro* in serum/LIF conditions, mESCs obtain high methylation patterns similar to post-implantation embryos right after reacquisition of 5mC (Habibi et al., 2013). In consequence, the regulation of a small set of cell-type specific enhancers by 5mC could be a feature

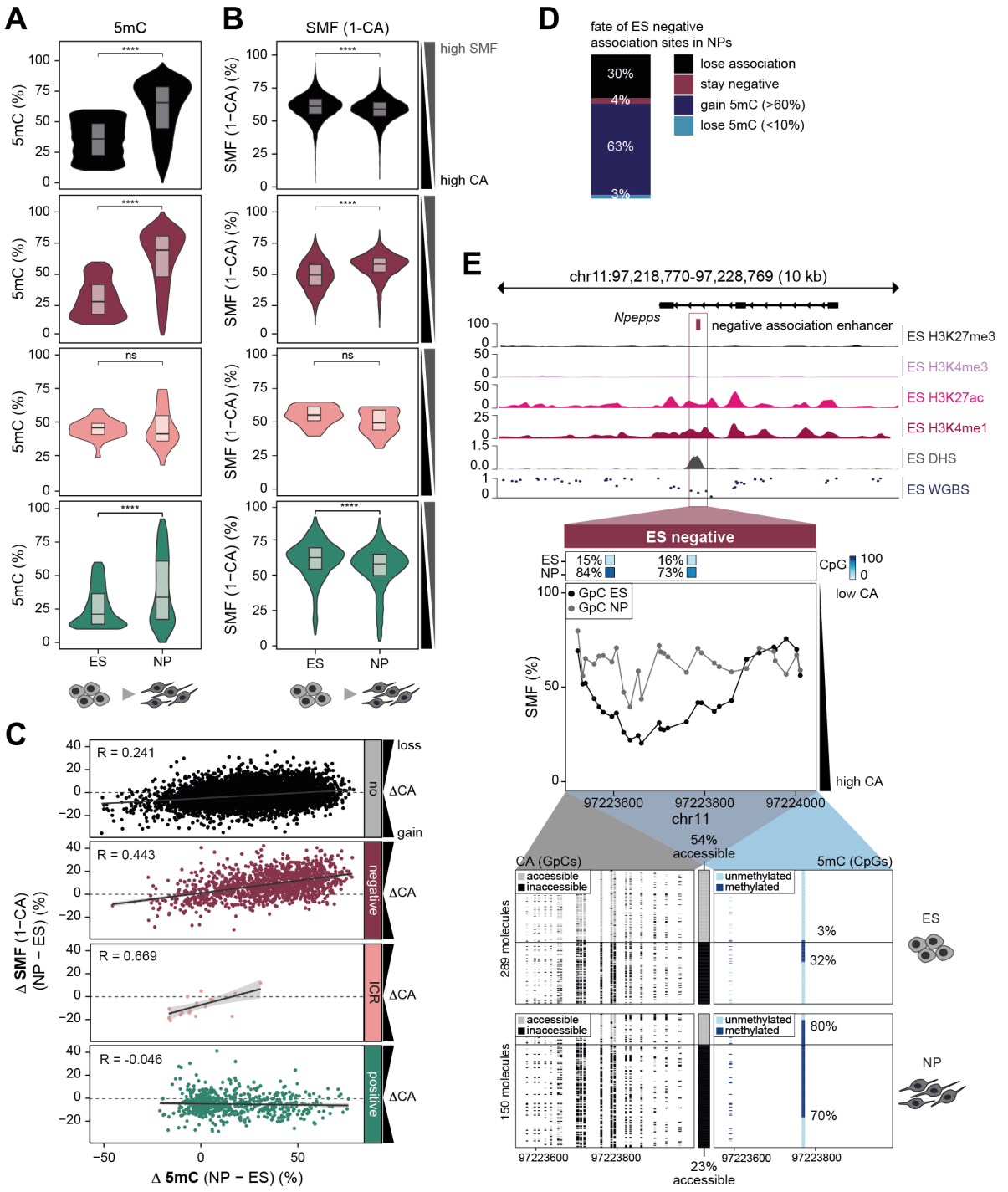

Figure 18: Enhancers with negative 5mC-CA association in ES cells are cell-type specific. (A-B) Sites with negative 5mC-CA association identified in embryonic stem cells (ES) show an increase in DNA methylation (5mC) and a loss of chromatin accessibility (CA) upon differentiation into neural progenitors (NPs). Violin box plots representing the average (A) 5mC and (B) CA measured by SMF (1-CA) at sites without (black, stringent "no" sites, n = 13,050), with negative (pink, n = 2,074) and positive (cyan, n = 1,117) 5mC-CA association, and sites within ICRs (rose, n = 32). Stars illustrate the significance of a Wilcoxon rank test (ns $p > 0.05$, ****$p \leq 0.0001$). **(C)** Loss in CA is anti-correlated with the increase in 5mC in NPs at ES sites with negative 5mC-CA association. Scatterplots comparing the changes in CA (SMF = 1-CA) and 5mC in ES to NP differentiation at sites in the different association categories. The correlation is weaker at sites without (no, black) or positive (cyan) 5mC-CA association. Regression line and Pearson correlation coefficients are displayed. **(D)** Most enhancers with negative 5mC-CA association in ES cells gain high levels of 5mC in NPs. Stacked bar plot depicting the fate of negatively associated enhancers in ES cells upon differentiation into NPs. **(E)** Single locus example of the fate of enhancers with negative 5mC-CA association identified in ES cells during neural differentiation.

(legend continued on next page)

Top panel: genomic context in ES cells 10 kb around the enhancer site (representation as in Figure 10A). Middle panels: average SMF signals (1-methylation%) of individual GpCs in ES (black dots) and NP (gray dots). Bottom panels: single DNA molecule stacks in ES and NP (representation as in Figure 11E).

specific to pluripotent cells. To test this hypothesis, I performed together with Rozemarijn Kleinendorst, SMF for additional murine somatic cell types such as erythrocytes (murine erythroleukemia, MEL) and myoblasts (C2C12). This enabled me to compare the enhancer classes of four different cell lineages: embryonic, neural, erythroid and muscle lineage. Using my statistical analysis pipeline, I was able to identify all three 5mC-CA association classes in all cell lineages (Figure 20A-C). Interestingly, all cell lines showed comparable ratios of the different classes to the ones observed in mESCs, with a predominant number of enhancers without 5mC-CA association and a small percentage (2-4%) of negative 5mC-CA association enhancers (Figure 20A-D). Comparison of the negative 5mC-CA association sites of all four cell lineages reveals that they depict cell-type specific sets of enhancers that show 5mC sensitivity in only one of the cell types and are mostly inactive in the others (Figure 20E). This is consistent with the finding of the GO term analysis showing the associated genes are enriched for cell-type specific GO terms coherent with the given cell type (Figure 20F). The negative 5mC-CA associated enhancers in MEL cells are, for example, enriched for genes related to the regulation of erythroid differentiation, while the C2C12 enhancers are rather enriched for gene functions related to cytoskeleton organization, an important feature in myoblasts.

I wanted to understand whether those enhancers are bound by cell-type specific TFs. Thus, I performed HOMER motif enrichment analysis on the cell-type specific negative 5mC-CA association sites. In all cell lineages, the negative enhancers are enriched for E-box motifs that could be bound by different TFs in the different cell types such as Max-Myc in mESCs or USF1 in erythrocytes (Figure 20G). Those E-box-binding basic-helix-loop-helix (bHLH) TFs have been identified as methyl-sensitive in *in vitro* screens and could, therefore, explain the observed sensitivity of the enhancers. Furthermore, the negative enhancers were also enriched for TF motifs related to the specific cellular identity. Enhancers in MEL cells were, for example, enriched for GATA3/6 motifs, in NPs for Sox10, in mESCs for KLF5 and in C2C12s for HIF2a or MyoD (Figure 20G). Nevertheless, this kind of analysis cannot give definitive answers about the TF binding patterns, which is also complicated by the fact that many TFs have similar motifs.

In conclusion, my SMF analysis pipeline enables the detection of 5mC sensitive sites in multiple somatic cell lines and shows that those sites represent only a small, but cell-type specific, fraction of active enhancers. Consequently, 5mC sensitivity of enhancers is not a stem cell specific feature, but is a regulatory feature across multiple pluripotent and somatic cell lineages.

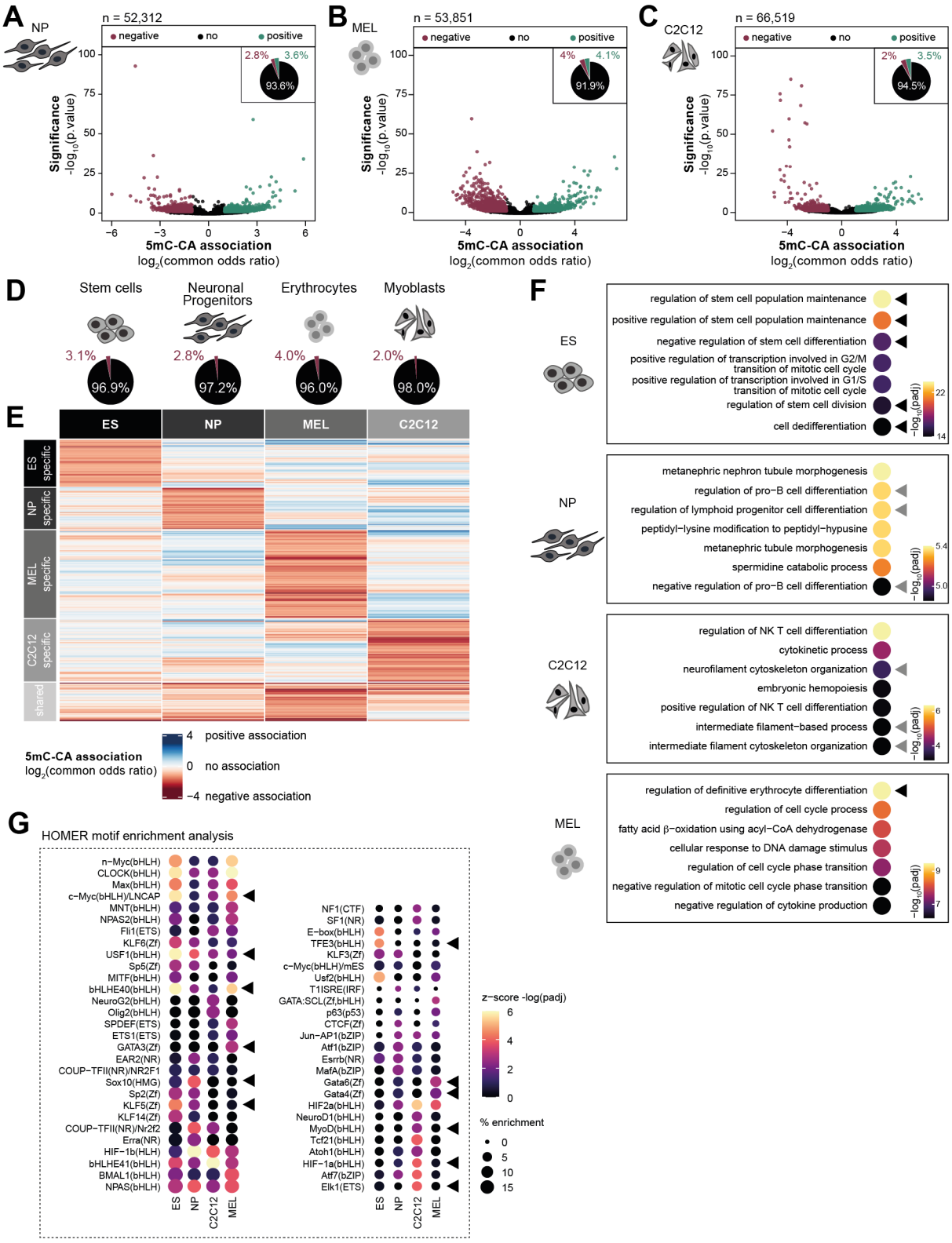

Figure 19: Negative 5mC-CA association enhancers have cell-type specific activity. (A-C) SMF identifies sites with different strengths of 5mC-CA association in three somatic cell lineages. Volcano plots showing the log2 common odds ratio and -log10 p-value of the Cochran-Mantel-Haenszel test for (A) neural progenitors (NP), (B) erythrocytes (murine erythroleukemia cells, MEL) and (C) myoblasts (C2C12). The number of tested 10-60% methylated CpG sites are depicted on top. Representation as in Figure 12A. **(D)** Similar proportions of CpGs with negative 5mC-CA associations are identified in pluripotent and somatic cell lines. Pie charts representing the relative proportion of CpGs without (black) and with negative 5mC-CA association (pink) identified in embryonic stem cells (ES), NP, MEL and C2C12.

(legend continued on next page)

(E) Negative 5mC-CA association is cell-type specific at most CpGs. Heatmap showing the log2 common odds ratio of the association between 5mC and CA (measured by the Cochran-Mantel-Haenszel test) for CpGs with negative 5mC-CA association covered in all four cell lines (n = 1,608). Heatmap was clustered using k-means clustering (k = 5). **(F)** Enhancers with negative 5mC-CA association are connected to gene ontologies (GO) associated with cell-type specific processes. Dot plots showing the adjusted p-value of the 7 top GO term processes (log fold change > 2) identified by GREAT analysis of sites with negative 5mC-CA association in (1) ES, (2) NP, (3) C2C12 and (4) MEL cells. Triangles indicate cell type specific processes for each cell line (black: highly relevant, grey: relevant). **(G)** Cell-type specific sites with negative 5mC-CA association are enriched for E-box motifs as well as for motifs of cell-type specific transcription factors. Dot plot showing the z-score of the -log of the adjusted p-value and the percentage of enrichment of motifs identified with HOMER motif enrichment analysis of cell-type specific negative enhancers in ES, NP, C2C12 and MEL cells. Black triangles indicate cell-type specific motifs.

2.1.6 | Negative 5mC-CA association at enhancers is conveyed by DNA methylation sensitivity of TFs

Having identified active cell type specific enhancers in embryonic as well as somatic cell lines that are regulated by 5mC *in vivo*, provokes the question for the underlying mechanism that establishes this methylation sensitivity. On the one hand, it could be an indirect mechanism in which chromatin marks and methylation binding proteins inhibit the activation of the enhancers by TFs. This scenario is rather unlikely based on the fact that no repressive chromatin marks have been observed at enhancers with a negative 5mC-CA association (Figure 14A). On the other hand, it could be caused by a direct mechanism in which the methylation mark directly inhibits the binding of TFs to their motifs. Large *in vitro* studies and individual *in vivo* studies have shown that some TFs are indeed methyl-sensitive and are unable to bind to their motif in an methylated state (Héberlé and Bardet, 2019). Yet, those studies have often focused on the gain or loss of bound TF binding sites (TFBS) across the genome upon loss or increase in 5mC, respectively, and have neglected the regulation of TF binding by 5mC at the already active CREs. In consequence, using my SMF approach, I wanted to test whether 5mC directly inhibits the binding of TFs *in vivo* at active enhancers.

For this, I leveraged the ability of SMF to quantify TF binding events at individual loci and combined this with the single molecule measurement of 5mC (Figure 21A). I calculated the GpC methylation within three bins: one 30 bp bin centered at the TFBS and two neighboring 15 bp bins up- and downstream of the TFBS. Based on the GpC methylation within those bins, I sorted the individual DNA molecules into a TF bound, accessible and nucleosome bound fraction. In order to identify the impact of 5mC on this binding pattern, I selected for TFBS with a CpG within the 30 bp center bin and tested the association between 5mC and TF binding using the statistical proportionality test. In addition, multiple selection steps have been performed to ensure the binding of a given TF and sufficient statistical power, such as selection for TFBS that are overlaying with ChIP-seq peaks and that show TF binding levels of at least 5% (see chapter 4.2.20). Finally, I analyse 421 TFBS and identified 37% of those with a negative association between 5mC and TF binding. At those TFBS, one can either observe TF binding or 5mC (Figure 21B). Only a marginal percentage of sites (1.5%) showed a positive 5mC-TF association with low significance (Figure 21B).

Among the top negative 5mC-TF association sites are known methyl-sensitive TFs such as CTCF and Max-Myc (Figure 21B). A single locus example of the top negative CTCF site with a binding frequency

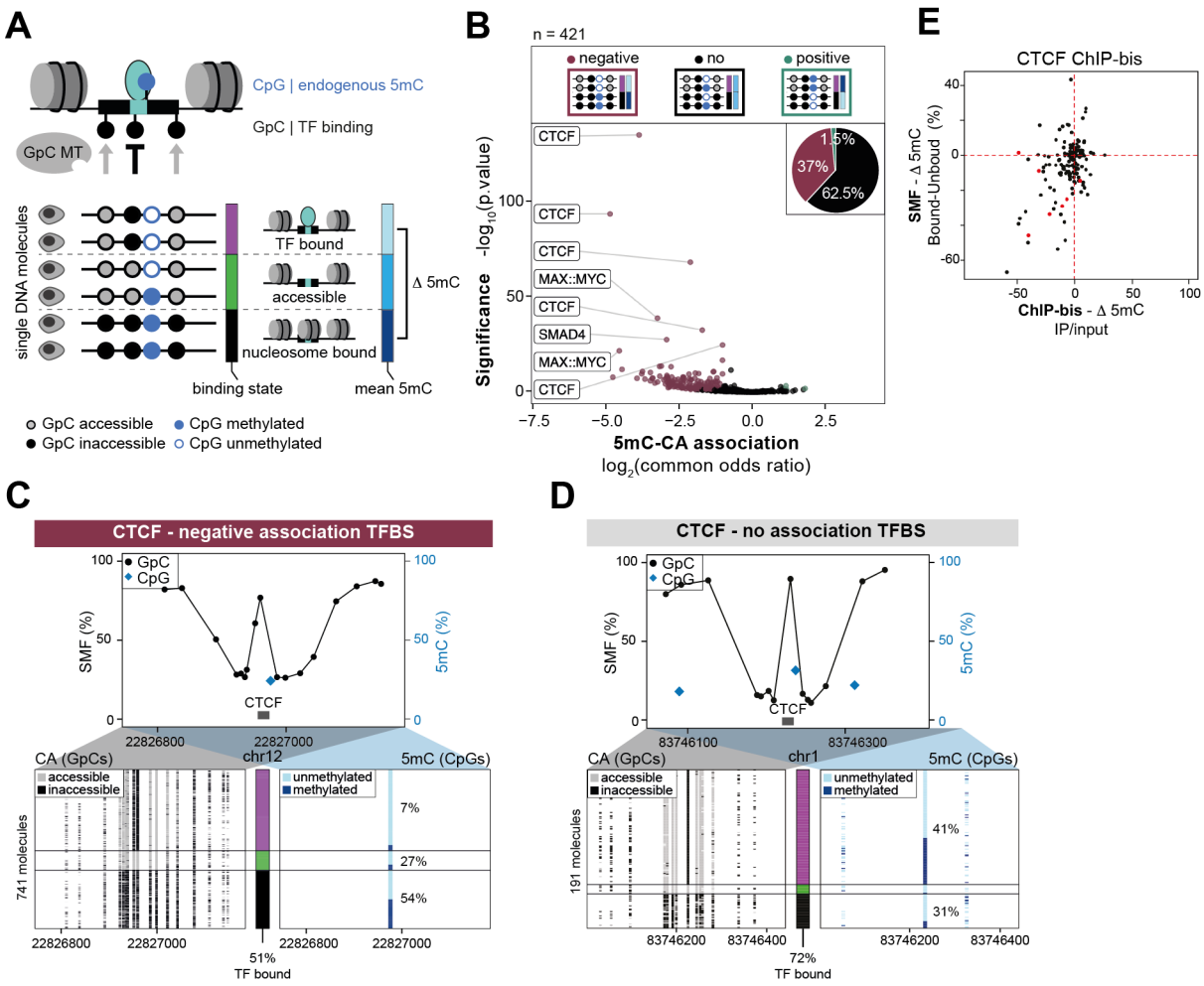

Figure 20: "Negative association between TF occupancy and DNA methylation. (A) Schematic representation of the experimental strategy to quantify the association between DNA methylation (5mC) and transcription factor (TF) binding at molecular resolution using SMF. Short footprints created by TFs were distinguished from large nucleosome footprints by measuring CA over three consecutive bins. DNA molecules were classified into TF-bound (purple bar), accessible (green bar), and nucleosome-bound (black bar) fractions. Identity of the TF was inferred by combining motif information and evidence of binding by ChIP-seq (Kleinendorst et al., 2021). 5mC of the TF bound and nucleosome bound fractions are then tested for association. White/blue lollipops represent 5mC as unmethylated/methylated CpG, black/white lollipops represent CA as inaccessible/accessible GpC. **(B)** Negative association between 5mC and TF binding occurs at a subset of TF-binding sites (TFBS) genome wide. Volcano plot depicting the log2 common odds ratio (COR) and the p-value of a Cochran-Mantel-Haenszel test that probes the association between TF binding and 5mC. For most TFBS, 5mC occurs at equivalent levels between TF bound and nucleosome bound molecules (black dots), yet 156 TFBS show a negative 5mC-TF association (pink dots). Representation as in Figure 12A. Representative motif names are shown for sites with the strongest negative 5mC-TF association. Only TFBS with >5% TF-binding frequency were considered. **(C-D)** Single locus example of the methylation difference between TF bound and nucleosome bound molecules at a CTCF site with (C) negative or (D) no 5mC-TF association. Top panel: average SMF signal (1-methylation%) of individual GpCs (black dots) and average 5mC of individual CpGs (blue diamonds) at a CTCF-binding site (gray box). Lower panels: single DNA molecule stack sorted into a TF-bound (purple), accessible (green), and nucleosome-bound (black) fraction. Representation as in Figure 11E. **(E)** Validation of the negative 5mC-CA association observed at CTCF binding sites. Scatter plot comparing the difference in 5mC between the TF bound and the nucleosome occupied fraction with the 5mC enrichment upon CTCF ChIP-bis (data from (Feldmann et al., 2013)). Each dot is a CTCF binding sites covered in both experiments. The enrichment of CTCF for unmethylated molecules at negatively 5mC-CA associated binding sites (red dots) is observed by both methods." (Kreibich et al., 2023).
Sarah Kaspar (Center for Statistical Data Analysis) advised on the statistical analysis and the CMH test.

of 51% shows a clear negative association, where 7% of the molecules bound by CTCF are methylated in comparison to 54% of the nucleosome bound molecules (Figure 21C). This stands in contrast to a CTCF site with no 5mC-TF association that shows a CTCF binding frequency of 72%, where the TF bound fraction and the nucleosome bound fraction are similarly methylated (41% vs 31%) (Figure 21D). To confirm the methyl-sensitivity of CTCF with an orthogonal approach, I analysed publicly available data obtained by CTCF ChIP-seq coupled with bisulfite sequencing (ChIP-bis). This data supports the observed enriched binding of CTCF to unmethylated DNA molecules (Figure 21E) and validates my SMF approach.

Before going deeper into the analysis of the 5mC-TF association sites, I wondered whether the CA and the TFBS analyses overlap and whether the small footprints in the GpC methylation pattern, caused by the binding of TFs, negatively affect the molecule sorting in the CA analysis.
Comparing the common odds ratios of the two analysis strategies shows an overall good agreement (R = 0.401) with a higher sensitivity of calling a negative association with the TFBS analysis (Figure 22A). The reason for this becomes apparent when comparing the two strategies at single locus examples of TFBS (Figure 22B). In those cases, I first sorted the molecules based on the CA, resulting in an accessible and inaccessible fraction. In a second step, I sorted the molecules in those two fractions separately based on the TFBS into a TF bound, accessible and nucleosome bound fraction. This double sorting analysis illustrates that the inaccessible CA fraction only contains molecules that are nucleosome bound (Figure 22B – bottom right), while the accessible fraction contains the TF bound, accessible and some partly nucleosome bound molecules (Figure 22B – top right). Thus, the accessible fraction is a mixture of multiple binding states. This explains why the CA analysis is less sensitive in calling negative 5mC-TF/CA associations, since the accessible fraction is a mixture of multiple binding states that contaminate the association calling. In contrast, in the TFBS analysis, I use pure states (TF bound vs nucleosome bound), increasing the statistical power. Consequently, the two analysis strategies complement each other. While the TFBS analysis is very robust and sensitive, only a small number of regulatory sites can be tested that fulfill certain criteria (see chapter 3.1.4). Whereas the CA analysis allows the testing of more than 100,000 regulatory sites, but with lower sensitivity. In addition, the double sorting analysis shows that the TF binding events do not interfere with the CA sorting into the accessible fraction (Figure 22B – left). The correct sorting is due to the number of GpCs analysed in the CA analysis, where all GpCs in a 101 bp window are averaged. Since TFs leave only small footprints, affecting one or two GpCs, those are usually averaged out by the remaining accessible GpCs.

To further investigate the role of 5mC at sites with no or negative 5mC-TF association, I focused on the top TF candidates of the statistical test. Those included known methyl-sensitive TFs such as NRF1, CTCF and Max-Myc (Figure 23A). Interestingly, most of their binding sites where fully unmethylated, suggesting that they primarily bind to unmethylated motifs, which is in line with previous reports (Domcke et al., 2015; Maurano et al., 2015; Yin et al., 2017). This becomes especially apparent when

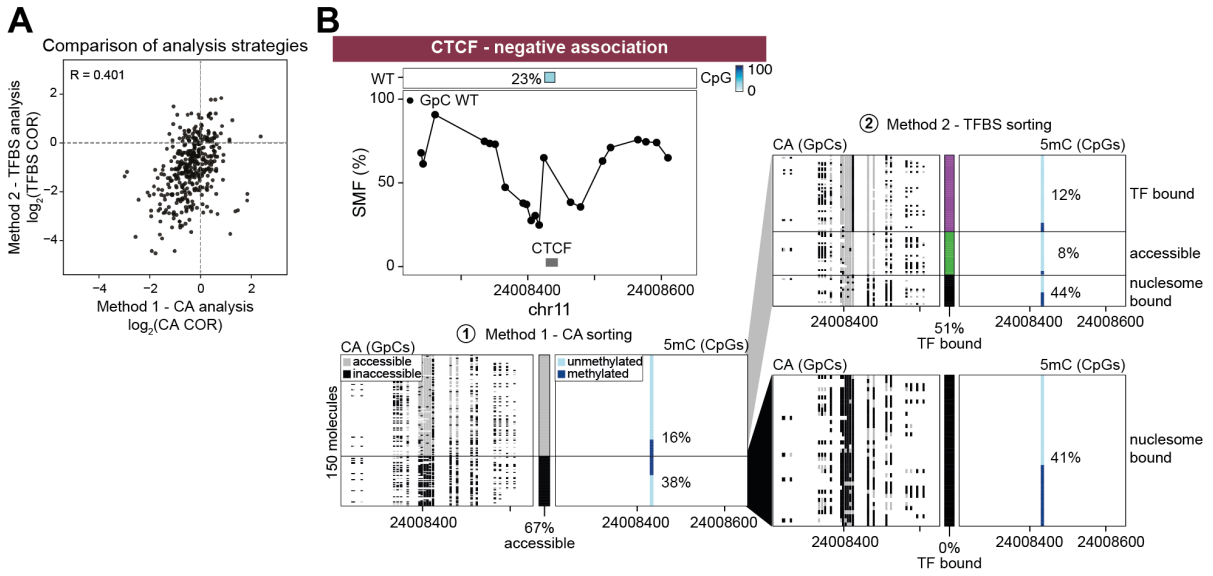

Figure 21: "The two SMF analysis strategies (CA and TFBS) show correlating results. (A) Common odds ratios (COR) calculated with the analysis strategies based on (1) chromatin accessibility (CA) and (2) TF binding site (TFBS) footprints show a good agreement. Scatter plot comparing the log2 COR values derived using the two molecule sorting strategies. The Pearson correlation coefficient is depicted. **(B)** The strategy based on CA accurately sorts most of the TF bound molecules into the accessible ("active") fraction. Single locus examples of a CTCF binding site with negative 5mC-CA association. Top panel: DNA methylation (5mC) and location of CpGs. Middle panel: SMF average plot (representation as in Figure 21C). Lower left panel: SMF single molecule stacks as sorted by analysis method 1 based on CA (representation as in Figure 11E). Lower right panel: single molecule stacks of the molecules from the individual CA sorting fractions (top: accessible, bottom: inaccessible) sorted by analyses method 2 based on TF footprints. TF bound molecules (purple) are only found in the accessible CA fraction. The inaccessible CA fraction only contains nucleosome bound molecules (black)." (Kreibich et al., 2023).

looking at the 5mC distribution of all analysed binding sites for those TFs (Figure 23A-B), where e.g. the vast majority of NRF1 binding sites are to 0% methylated. This observation for NRF1 stands in contrast to CTCF and Max-Myc that are also predominantly bound to unmethylated loci, but that also frequently occupy intermediately methylated binding sites (Figure 23A-B). Looking closer at those loci with intermediate methylation shows a strong negative 5mC-TF association for most Max-Myc binding events (63%) (Figure 23A and C). For CTCF, the fraction of loci with negative 5mC-TF association is in contrast much smaller, while most show no association (Figure 23A and C). This data shows considerable variations in the distribution of the average 5mC and the degree of 5mC-TF association at different loci bound by a given TF, suggesting that the binding of CTCF and Max-Myc are regulated by 5mC in a limited set of loci. As mentioned before, SMF itself is agnostic to the identity of the bound protein. To confirm the binding of CTCF and Max-Myc at the identified footprints, I analysed publicly ChIP-seq data for the two TFs. This data supports their inferred binding and shows that binding sites with negative and without 5mC-TF association are equally bound by the given TF (Figure 23D-E).

Having identified TFs with different degrees of 5mC-TF association, I set out to identify what drives this association at some but not at other TFBS of CTCF and Max-Myc. Previous studies have shown that the methyl-sensitivity of TFs can be impacted by the position of the methylated CpG within the motif

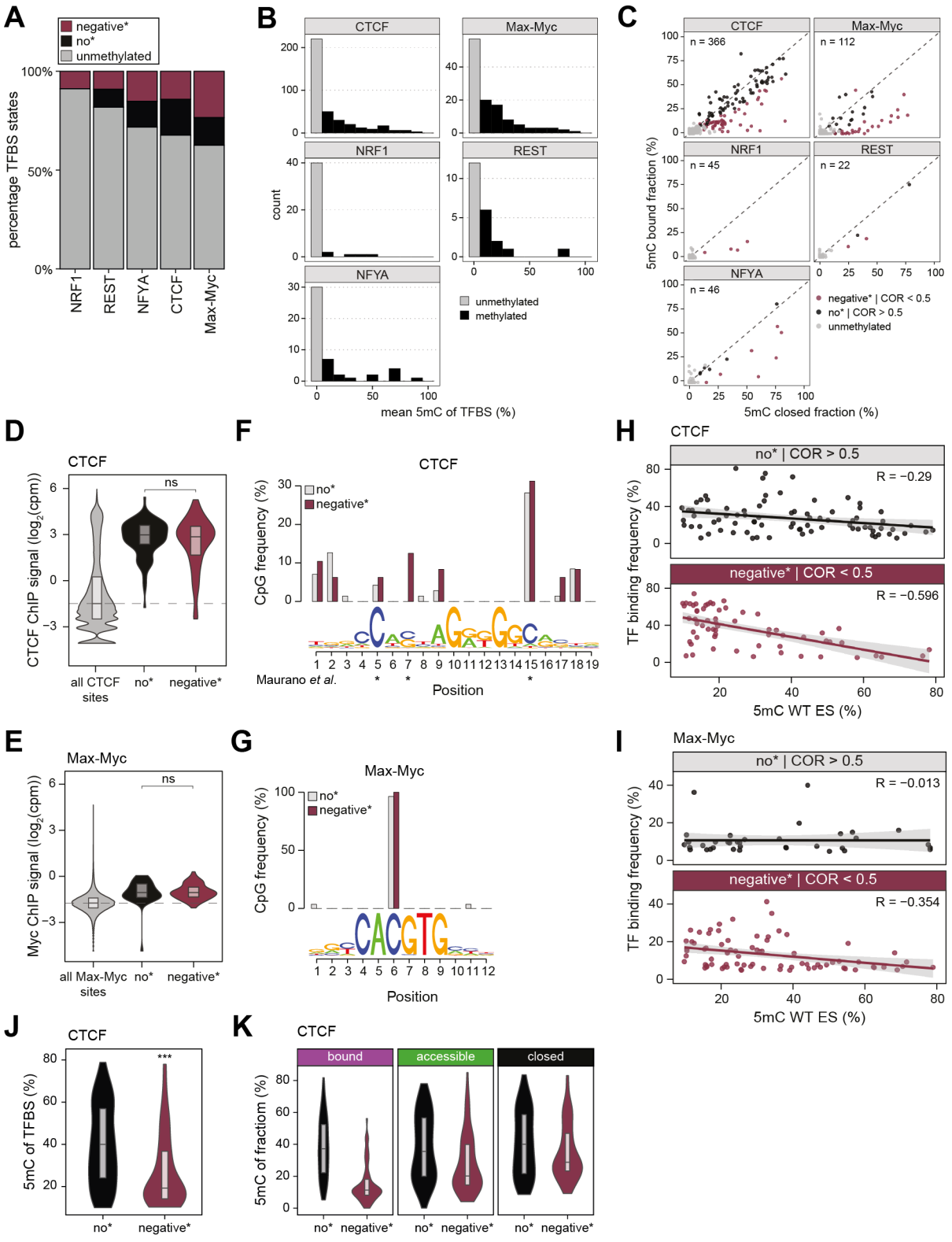

Figure 22: CTCF and Max-Myc are sensitive to DNA methylation TFs with different mechanisms. "(A) Bar chart representing the proportion of TFBS in each 5mC-TF association category for TFs with >10 binding sites covered by SMF. Depicted is the percentage of TF binding sites (TFBS) being unmethylated (<10% 5mC, gray) or intermediately methylated and falling into the no (black) or negative 5mC-TF association (pink) category. **(B)** Distribution of DNA methylation (5mC) at the binding sites of various TFs. Histograms describing the average 5mC at the binding sites of TFs. The binding of all TFs is enriched for unmethylated regions (gray), yet CTCF, Max-Myc, NFYA and REST are also binding at regions with intermediate methylation (black). **(C)** Differential methylation between TF bound molecules

(legend continued on next page)

and nucleosome occupied molecules at a subset of TFBS. Scatter plot showing the average 5mC for the fraction of molecules showing a TF footprint against those showing a nucleosome footprint at the same locus. The analysis is shown for each TF separately. The color code indicates the category of the binding site (pink – negative 5mC-CA association*, black – no 5mC-CA association*, gray – unmethylated). **(D-E)** Validation of the TF binding at sites with no and negative 5mC-TF association by ChIP- seq. Violin box plots showing the distribution of ChIP-seq signal (log2 of the read counts per million [cpm]) at all binding sites, compared to binding at sites without or with negative 5mC- TF association for (D) CTCF and (E) Myc. The vertical line depicts the boundary of the 50th percentile. **(F)** CpG occurrence at specific positions of the intermediately methylated CTCF motifs. Bar plot displaying the frequency of CpG dinucleotides at binding sites without (gray) and with negative 5mC-TF association (pink). CTCF sites with a negative 5mC-TF association frequently have a CpG at position #7. Sites without an association have CpGs at other positions, but not at position #7. The overlap with previously reported CpG positions conferring methyl-dependent binding are indicated by a star (position #5, #7, and #15) (Maurano et al., 2015). **(G)** CpG occurrence at specific positions of the intermediately methylated Max-Myc motifs. Representation as in (F). CpGs mainly occur at position #6, with no significant difference between categories. **(H-I)** TF occupancy is a function of 5mC levels at sites with 5mC-TF association. Scatterplots comparing the changes in TF occupancy and 5mC in WT mESC at sites with no (top) and negative (bottom) 5mC-TF association for (H) CTCF and (I) Max-Myc binding sites. The decrease in TF binding is anti-correlated with the increase in 5mC at the locus at sites with negative 5mC-TF association. Regression line and Pearson correlation coefficients are displayed. 5mC-TF categories are defined using only the COR of the Cochran-Mantel-Haenszel test, as indicated. **(J-K)** CTCF sites with no and negative 5mC-CA association show intermediate levels of 5mC. CTCF sites with negative 5mC-CA association show lower average 5mC levels than sites without association, which is mainly driven by the TF bound fraction (K, purple). Violin box plots showing the average 5mC levels of CTCF binding sites with no* or negative* 5mC-CA association. (K) depicts the 5mC levels of the molecules in the individual fractions from the TFBS sorting." (Kreibich et al., 2023)
Star (*) marks 5mC-TF categories defined with soft criteria using only the common odds ratio of the Cochran-Mantel-Haenszel test as indicated (see chapter 4.2.20).

(Kribelbauer et al., 2017; Yin et al., 2017). This positional effect has, for instance, been demonstrated *in vivo* for CTCF (Hashimoto et al., 2017; Maurano et al., 2015; Wang et al., 2012). To investigate this positional effect on the 5mC-TF association, I analysed the motifs of CTCF and Myc at no and negative association sites in order to detect any differences in the CpG distribution. Indeed, analysing the CTCF TFBS identified a CpG at position #7 that is only present in motifs at CTCF TFBS with negative 5mC-TF association, arguing that the presence and methylation of this CpG dinucleotide physically inhibits the binding of CTCF to those sites (Figure 23F). This position overlaps with those identified by Maurano et al. as drivers of the methylation-sensitivity of CTCF (Maurano et al., 2015). To further support this observation, I investigated whether there are any differences in the 5mC levels of no and negative 5mC-TF associations sites for CTCF that could explain this differential binding behavior. For this, I compared the mean 5mC levels at CTCF sites in both categories. Here, it can be observed that the negative association sites have moderately lower methylation levels (Figure 23J), which is mainly driven by the low 5mC levels in the TF bound fraction (Figure 23K). Therefore, these results support the hypothesis that for CTCF the position of the CpG dinucleotide is the main determinant to differentiate negative from no association sites and not the 5mC levels per se.

Interestingly, I was not able to detect a similar motif-based mechanism for Max-Myc. In both categories, Max-Myc binds to a constrained E-Box motif that contains a CpG site at the center at position #6 (Figure 23G). This suggests that a different mechanism regulates the methyl-sensitive binding of Max-Myc.

Next, I investigated to what extent 5mC is determining the TF binding frequency of CTCF and Max-Myc. The analysis was based on the previous observation that CA and 5mC anti-correlate at enhancers with negative 5mC-CA association (Figure 15G-H). Therefore, I analysed the correlation between the 5mC

levels and the TF binding levels. The advantage of SMF is that the TF binding frequency and the 5mC is measured with a unified scale, allowing the direct comparison of the two measurements. I observed a strong correlation between 5mC and TF binding frequency at CTCF as well as Max-Myc sites with a negative 5mC-TF association (CTCF R = -0.596; Max-Myc R = -0.345), but not at those without association (CTCF R = -0.29; Max-Myc R = -0.013) (Figure 23H-I). At negative association sites, the higher the 5mC levels is, the lower the TF binding frequency is, and vice versa (Figure 23H-I). These results show for the first time a direct correlation between TF binding frequency and DNA methylated *in vivo* and strongly support the hypothesis that 5mC directly regulates the binding of the TFs *in vivo*.

2.1.7 | TET enzymes are crucial for regulating TF binding at DNA methylation sensitive TF binding sites

If 5mC is indeed directly regulating the binding of methyl-sensitive TFs like Max-Myc, perturbations of the underlying 5mC levels would lead to changes in the TF occupancy. To test this hypothesis, I analysed the changes in TF occupancy at TFBS with no or negative 5mC-TF association upon TKO of TET enzymes. This analysis reveals that an increase of 5mC levels by at least 30% leads to a significant reduction of the TF binding frequency at negative associated TFBS that is concomitant with an increase in the nucleosome binding frequency (Figure 24A). Analysing the direct correlation between changes in 5mC and changes in TF occupancy uncovers a strong linear relationship between the increase in 5mC and the loss in TF binding at TFBS with negative 5mC-TF association (R = -0.51) (Figure 24B - right). This stands in contrast to TFBS without 5mC-TF association where only a limited loss in TF occupancy is observed, which is in addition not correlated with the increase in 5mC levels (R = -0.07) (Figure 24B - left).

Focusing on Max-Myc also shows this observed negative correlation between increase in 5mC and loss in TF binding in the SMF data (data not shown). Moreover, analysis of publicly available Myc ChIP-seq data in WT and TET TKO mESCs that shows a negative correlation between the increase in 5mC and the decrease in ChIP-seq signal (R = -0.395) (Figure 24C - right), albeit with a lower dynamic range in comparison to SMF. This anti-correlation is observed only for those Max-Myc BS for which I have identified a negative 5mC-TF association, but not for those with no association (R = 0.01) (Figure 24C - left). This behavior of negative associated Max-Myc binding sites upon increase in 5mC becomes especially apparent in a single locus example (Figure 24D). In WT ES cells, Max-Myc is bound to its motif in 10% of all cells. Strikingly, none of the molecules bound Max-Myc is methylated at the close-by CpG dinucleotide, although it has an average 5mC level of 37%. In the nucleosome bound fraction, 47% of the molecules are methylated. Thus, this Max-Myc BS shows a strong negative association between TF binding and 5mC, which hints in the direction that Max-Myc favors the binding to its motif in an unmethylated state. Upon increase of the 5mC level at the close-by CpG up to 85% in the TET TKOs, the binding of Max-Myc is fully lost, with none of the molecules showing a Max-Myc footprint (Figure 24D – bottom right). This strongly suggests that Max-Myc is unable to bind its motif in a methylated state and that TET enzymes may play a crucial role in the active demethylation of TFBS that are bound by methyl-sensitive TFs.

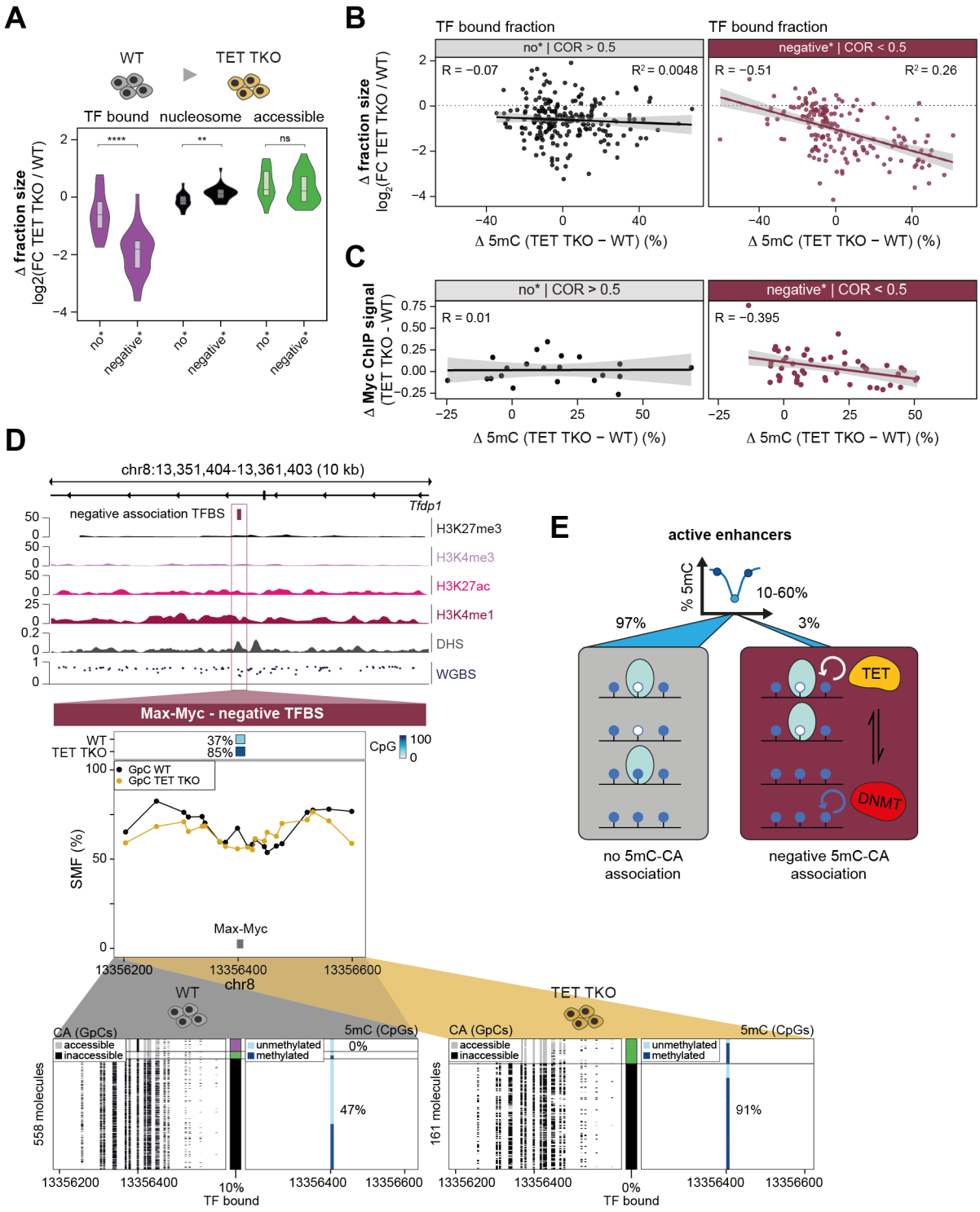

Figure 23: Active 5mC turnover is required for TF binding at loci with negative 5mC-TF associations. (A) Transcription factors (TFs) with negative 5mC-TF association lose binding upon an increase in DNA methylation (5mC) in TET TKOs. Violin box plots represent the changes in states' frequency upon increase of 5mC in TET TKOs. The TF bound fraction is significantly reduced at sites with negative 5mC-TF association (n = 29) in comparison with sites without (n = 17), while the nucleosome bound fraction is increased upon TKO. TFBS with 5mC increase <30% in TET TKOs were excluded. Stars illustrate the significance of a Wilcoxon rank test (ns: $p > 0.05$; **$p \leq 0.01$; ****$p \leq 0.0001$). **(B)** TF occupancy correlates with 5mC changes at sites with negative 5mC-TF association. Scatterplots comparing the changes in TF occupancy and 5mC from WT to TET TKO at sites with no (left) and negative (right) 5mC-TF association. The decrease in TF occupancy is anti-correlated with the increase in 5mC at the locus. Regression line, Pearson correlation coefficients, and coefficient of determination are displayed. Stringent 5mC-TF association categories are

(legend continued on next page)

used. **(C)** Myc ChIP-seq data validates the anti-correlation between Myc binding and 5mC at Myc binding sites with negative 5mC-CA association observed by SMF. Scatter plots comparing the changes in Myc binding measured by ChIP-seq of publicly available data and 5mC in WT vs TET TKO at Myc sites with no* (n = 21) or negative* (n = 52) 5mC-CA association. The regression lines and Pearson correlation coefficient are displayed. **(D)** Single locus example illustrating the loss of TF binding upon 5mC increase at a Max-Myc binding site with negative 5mC-TF association. Top panel: genome browser track (representation as in Figure 10A) of an intragenic *Tfdp1* enhancer with a negative 5mC-TF association Max-Myc binding site. Middle panel: on top: average 5mC of the CpGs in the two cell lines. Below: average SMF signal (1-methylation%) of individual GpCs in WT (black dots) and TET TKO (yellow dots) at a single locus containing a Max-Myc binding site (gray box). Bottom panels: single-molecule stacks for WT (bottom left panels) and TET TKO (bottom right panels). Representation as in Figure 21C. **(E)** Mechanistic model describing the effects of 5mC on enhancers' activity. 5mC does not affect chromatin accessibility nor TF binding at the majority of active enhancers (left), while it controls the activity of a subset of cell-type-specific enhancers (right). Active 5mC turnover by TET enzymes is required for TF binding at methyl-sensitive enhancers. The activity of these enhancers in a cell type is controlled by the equilibrium between demethylation by TETs and *de novo* methylation by DNMTs.
Star (*) marks 5mC-TF categories defined with soft criteria using only the common odds ratio of the Cochran-Mantel-Haenszel test as indicated (see chapter 4.2.20).

2.1.8 | Conclusions on the functional role of DNA methylation in enhancer regulation

In conclusion, my work on the role of 5mC at enhancers has revealed that the vast majority of active enhancers is not regulated by 5mC (Figure 24E). Thus, at those enhancers the observed demethylation is rather a consequence of TF binding and the activation of the enhancer than the cause of it. Nevertheless, careful statistical analysis of the SMF data revealed a small subset of active enhancers that are sensitive to 5mC, where local demethylation is causal for the binding of TFs and the activation of enhancers (Figure 24E). Those enhancers are cell-type specific and are regulating genes that are associated with cell-type specific functions. Perturbation analyses of the underlying 5mC levels uncovered that methyl-sensitive enhancers are regions of high methylation turnover, and that TET enzymes are a crucial component in the regulation of those enhancers since the removal of TET enzymes leads to decreased TF binding and decreased enhancer activity.

2.2 | Investigating the functional role of histone modifications on gene regulation

Histone modifications represent another class of epigenetic modifications that may regulate gene transcription. However, the causal evidence for a direct regulatory role is largely missing. The objective of this second project was to develop a genomics technique to investigate the causal effect of different histone modifications on TF binding and activity of CREs. For this, I aimed to combine the chromatin enrichment technique CUT&RUN with the SMF technology. The combination of those two approaches would enable the read-out of TF binding patterns in direct correlation with the presence of a given histone modification.

2.2.1 | Proof of concept with CTCF provides promising enrichments

CUT&RUN is based on a step-wise protocol in which an antibody is bound to the protein or histone modification of interest, which is subsequently bound by a pA-MNase fusion protein. Upon activation, the MNase cuts the close-by open chromatin, thus releasing small protein-bound DNA fragments from the compacted chromatin. Those small DNA fragments are then extracted and used for library preparation and Illumina sequencing (Figure 25A).
As a proof of principle, I first implemented CUT&RUN in mESCs using an antibody against the CTCF, a TF which has been shown to lead to efficient enrichments over the IgG control in human K562 cells (Skene and Henikoff, 2017). IgG controls serve as a negative control, since the antibody has no target to bind. Thus, IgG controls show the background activity of the non-targeted pA-MNase. The native CTCF CUT&RUN data provided good correlations between individual biological replicates at CTCF sites ($R = 0.79$) (Figure 25B) as well as between native CUT&RUN and published CTCF ChIP-seq data ($R = 0.74$) (Figure 25C). Analysis of single genomic loci and global analysis of the top 20,000 bound CTCF sites revealed a promising enrichment of the native CUT&RUN data at CTCF bound sites. Yet, one could observe slightly lower local enrichment and a wider spreading of the signal in the CUT&RUN data in comparison to the ChIP-seq data (Figure 25D-E). This higher background was also evident when calculating the signal to noise ratio that showed slightly lower values for CUT&RUN samples in comparison to the ChIP-seq data (Figure 25F). Notably, my CUT&RUN data performed better than comparable CTCF CUT&RUN data available at the time (Figure 25D-F), suggesting that my CUT&RUN implementation was successful and promising for the next steps.

Since the final read-out of the CUT&RUN-SMF approach would be bisulfite sequencing, the next step in the methods development was the combination of CUT&RUN with bisulfite sequencing. To have a direct comparison, I split the CUT&RUN sample and performed either native or bisulfite converted sequencing (Figure 25A). The correlation between the native and bisulfite sequenced (bis) sample was reasonably high ($R = 0.80$) (Figure 25G). Moreover, global enrichment comparison at CTCF sites showed a good agreement between the samples, with a slightly elevated background in the bis sample (Figure 25D-F).

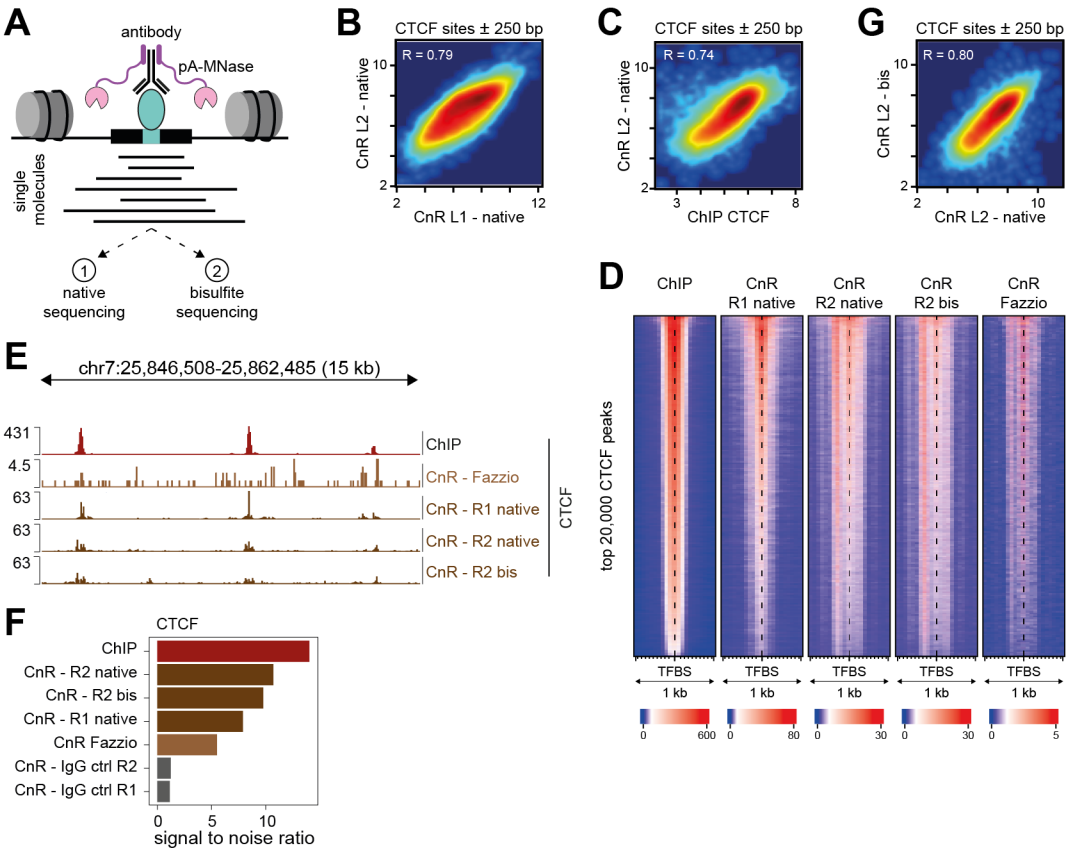

Figure 24: CUT&RUN successfully enriches for CTCF bound sites and can be combined with bisulfite sequencing. **(A)** Schematic representation of the CUT&RUN (CnR) experiment. The protein of interest is bound by a specific antibody, which is subsequently bound by a protein A (pA)-MNase fusion protein. Upon activation with Ca^{2+}, the MNase cuts and digests the DNA in proximity of the bound site. The cut DNA fragments are extracted and sequenced either under (1) native or (2) bisulfite-conversion conditions. **(B-C)** Smoothed scatter plot comparing the enrichment at CTCF binding sites ± 250 bp between (B) two biological CnR replicates or (C) CnR and publicly available CTCF ChIP-seq data in mESCs. The Pearson correlation coefficient is indicated. **(D)** CnR enriches for CTCF bound sites, but with lower local signal and wider spread. Heatmap of the enrichment signal at the top 20,000 bound CTCF binding sites of ChIP-seq data and different CnR samples. Rightmost column shows publicly available CTCF CnR data. Collection window was 1 kb. **(E)** CnR recapitulates CTCF peaks detected by ChIP-seq but with lower amplitude. Genome browser view of a single genomic locus containing three CTCF binding sites. The same data sets as in (D) are shown. **(F)** Signal to noise ratio is lower in CnR samples in comparison to ChIP-seq. Bar plot showing the signal to noise ratio of the different data sets. Signal is measured at the top 20,000 CTCF peaks in a 500 bp window. Background is measured at 20,000 random intergenic regions in a 500 bp window. **(G)** Smoothed scatter plot comparing the enrichment at CTCF binding sites ± 500 bp between the native and bisulfite converted (bis) CnR data. The Pearson correlation coefficient is indicated.

Off note, I optimized the bisulfite sequencing by trying different bisulfite conversion kits that showed, in the end, comparable conversion rates and comparable impacts on the enrichments (data not shown). In conclusion, the initial CUT&RUN experiments targeting CTCF provided promising results that showed a decent enrichment of the protein of interest, albeit lower than with ChIP-seq.

2.2.2 | Combination of CUT&RUN with SMF targeting histone modifications does not provide sufficient enrichment for association studies

Due to the successful implementation of CUT&RUN in the lab and promising proof of concept results using an antibody against CTCF, I moved on to histone modifications and implementing SMF into the protocol (Figure 26A-B). For this, I used antibodies targeting either H3K27me3 or H3K27ac. In a first trial, I performed normal native CUT&RUN for H3K27me3 (Figure 26A). The correlation with publicly available ChIP-seq data at 5 kb windows around active TSS was rather low (R = 0.55) (Figure 26D), but single loci and global of active TSS analyses revealed a promising enrichment of H3K27me3 signal at TSS that was even better than publicly available H3K27me3 CUT&RUN data (Figure 26I, J and L). In a next step, I added SMF to the CUT&RUN procedure (Figure 26B). For this, I first performed the SMF in extracted nuclei before moving on the CUT&RUN enrichment. To evaluate the impact of the SMF procedure itself on the final CUT&RUN signal, I sequenced the data either under native or bisulfite-conversion conditions. For H3K27me3, the SMF had a moderate impact on the enrichment. Comparing H3K27me3 CUT&RUN with H3K27me3 CUT&RUN-SMF samples shows a high correlation between the signals at TSS (Figure 26E). The signal at the top H3K27me3 marked sites is very comparable with a minor loss in the signal to noise ratio (Figure 26J and L). At the single locus example, one can observe a lower enrichment upon addition of SMF (Figure 26I), which is in line with the global signal comparison at active TSS (Figure 26F). While these results still seem promising with acceptable enrichments of the histone modification, implementing the bisulfite sequencing abolishes most of the enrichment and results in very low signal to noise ratio (Figure 26I-L). This can be observed at a single locus (Figure 26I) as well as at the top 10,000 H3K27me3 peaks at TSS (Figure 26J), where the signal is very low and spread out over a wider area.
A similar observation was made with H3K27ac. Here, the native CUT&RUN-SMF approach delivered already low enrichments at H3K27ac marked TSS (Figure 26G, I, K and M), which was even more decreased upon bisulfite sequencing (Figure 26H, I, K and M).

In conclusion, the low histone modification enrichments upon CUT&RUN-SMF and bisulfite sequencing make this approach at the moment inefficient for investigating the direct correlation between histone modifications and TF binding. The approach is based on the comparison of the TF binding frequencies between conventional SMF and enriched CUT&RUN-SMF (Figure 26C) and is dependent on a high enrichment to ensure high statistical power and the ability to detect moderate changes in TF binding. Since the here presented preliminary results of the CUT&RUN-SMF approach did not suffice the enrichment requirements, I stopped the project at this point.

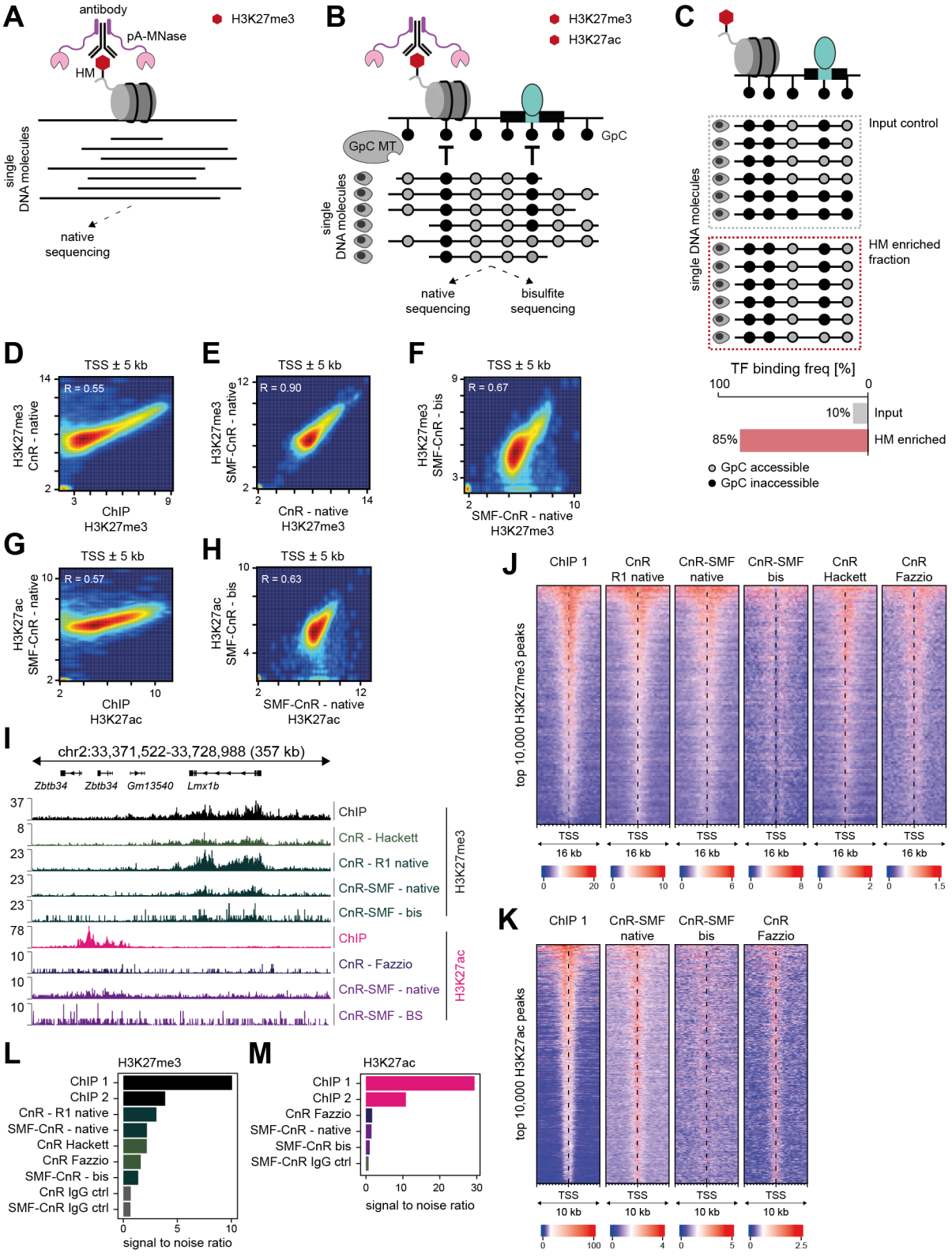

Figure 25: CUT&RUN enrichment is lost during SMF and bisulfite sequencing for histone modifications. (A) Schematic representation of the CUT&RUN (CnR) experiment for histone modifications (HMs). The HM of interest is bound by a specific antibody, which is subsequently bound by a protein A (pA)-MNase fusion protein. Upon activation with Ca^{2+}, the MNase cuts and digests the DNA in proximity of the nucleosome site. The cut DNA fragments are extracted and sequenced. This experiment was performed for H3K27me3. **(B)** Schematic representation of the combined method between CnR and Single Molecule Footprinting (SMF) experiment for HMs. In a first step, extracted nuclei are treated with a GpC methyltransferase (MT) that methylates accessible GpC. Subsequently, the antibody and pA-MNase are

(legend continued on next page)

added to cut and extract the DNA surrounding nucleosomes harvesting the HM of interest. By this, one can read out TF binding patterns in direct correlation with the presence of a HM. For method development purposes, DNA fragments were sequenced either under native or bisulfite-conversion conditions. This experiment was performed for H3K27me3 and H3K27ac. **(C)** Schematic representation of the analysis approach of CUT&RUN-SMF (CnR-SMF). Conventional SMF (grey box) is used as an input control that provides molecules with all TF binding states at a region of interest independent of the presence of local HMs. CnR-SMF (red box) is used to enrich for DNA molecules that harbor a HM of interest. If the HM is directly correlated with TF binding, the TF bound fraction should be increased in comparison to the input SMF data as shown schematically in the bar plot below. Since this approach is built upon the comparison of frequencies in the input vs enriched sample, efficient HM enrichment is crucial to ensure high statistical power and the ability to detect marginal changes in TF binding. **(D-F)** Smoothed scatter plot comparing the H3K27me3 enrichment at TSS ± 2.5 kb between (D) CnR and publicly available H3K27me3 ChIP-seq data in mESCs, (E) native CnR and native CnR-SMF and (F) native and bisulfite converted (bis) CnR-SMF. The Pearson correlation coefficient is indicated. **(G-H)** Smoothed scatter plot comparing the H3K27ac enrichment at TSS ± 2.5 kb between (G) CnR-SMF and publicly available H3K27ac ChIP-seq data in mESCs, (H) native and bis CnR-SMF. The Pearson correlation coefficient is indicated. **(I)** CnR recapitulates H3K27me3 peaks detected by ChIP-seq with high amplitude. Signal is gradually lost with addition of SMF and bisulfite sequencing. CnR for H3K27ac barely recovered ChIP-seq peaks. Genome browser view of a single genomic locus containing regions enriched for H3K27me3 or H3K27ac. The same data sets are shown as in (J) and (K). **(J)** CnR enriches for H3K27me3 marked TSS, but with lower local signal and wider spread. Heatmap of the enrichment signal at the top 10,000 H3K27me3 marked TSS of ChIP-seq data and different CnR and CnR-SMF samples. Two rightmost columns show publicly available H3K27me3 CnR data. Collection window was 16 kb. **(K)** CnR-SMF moderately enriches for H3K27ac marked TSS. Heatmap of the enrichment signal at the top 10,000 H3K27ac marked TSS of ChIP-seq data and different CnR and CnR-SMF samples. Rightmost column shows publicly available H3K27ac CnR data. Collection window was 10 kb. **(L)** Signal to noise ratio is lower in H3K27me3 CnR samples in comparison to ChIP-seq. Bar plot showing the signal to noise ratio of the different H3K27me3 data sets. Signal is measured at the top 5,000 H3K27me3 TSS peaks in a 5 kb window. Background is measured at 5,000 random intergenic regions in a 5 kb window. **(M)** Signal to noise ratio is very low in H3K27ac CnR samples in comparison to ChIP-seq. Bar plot showing the signal to noise ratio of the different data sets. Signal is measured at the top 5,000 H3K27ac TSS peaks in a 5 kb window. Background is measured at 5,000 random intergenic regions in a 5 kb window.

3 | Discussion

Transcription regulation is a complex mechanism integrating multiple regulatory layers and involving dozens of factors. The binding of TFs to CREs plays a pivotal role in the activation of transcription and can have influences in different aspects, such as nucleosome occupancy, the shape of the DNA and the cooperativity with other TFs (Kim and Wysocka, 2023; Spitz and Furlong, 2012). Also epigenetic modifications like histone modifications and DNA methylation may play a role in the regulation of TF binding, but their causal function has remained unclear to date (Héberlé and Bardet, 2019; Isbel et al., 2022; Kim and Wysocka, 2023; Millán-Zambrano et al., 2022).
The aim of my PhD research study was to identify and characterize the functional relationship between epigenetic modifications and TF binding *in vivo* in mammalian cells. This study focuses on enhancers, a class of distal CREs that contains clusters of TF binding sites and whose combinatorial activation is important for the spatiotemporal regulation of gene expression. Therefore, enhancers play a critical role in the acquisition and maintenance of cellular identity during development and in healthy tissue (Spitz and Furlong, 2012). In consequence, answering the central question, whether epigenetic modifications like DNA methylation and histone modifications generally contribute to the regulation of enhancer activity, has important implications in fundamental and translational research.

3.1 | The regulatory role of DNA methylation at enhancers

In the mammalian genome, the majority CpG dinucleotides are methylated, with certain genomic regions that escape this full methylation. One of those are enhancers, that show upon their activation reduced methylation levels of 10-60%, which correlates with the binding of TFs (Stadler et al., 2011). This correlation suggests a regulatory relationship between DNA methylation and TF binding. Yet, despite a decade of research since this discovery, it is still largely elusive whether DNA demethylation is important for the activation of enhancers (Héberlé and Bardet, 2019; Luo et al., 2018; Mattei et al., 2022).
Numerous *in vitro* and some *in vivo* evidences suggested that the affinity of many TFs is reduced upon methylation of their target motifs, preventing their cryptic binding outside of regulatory regions (Domcke et al., 2015; Héberlé and Bardet, 2019). Yet, TF binding is also directly responsible for the demethylation of enhancers by direct or indirect recruitment of TET enzymes and inhibition of DNMTs (Ginno et al., 2020; Héberlé and Bardet, 2019; Stadler et al., 2011). This bidirectional regulation between DNA methylation and TF binding has prevented the establishment of a causal relationship between them. One major challenge in addressing this question is the cell-to-cell variability in DNA methylation that characterizes active enhancers. A bulk-cell measurement of 10-60% methylation indicates that 10-60% of the cells within the cellular population were methylated at this one specific CpG, while the remaining cells were not. This epigenetic heterogeneity makes bulk genomics approaches non-suitable to study the relationship between DNA methylation and TF binding.

To overcome these limitations and solve the paradox of the bidirectional regulation, I advanced the SMF technology (Kleinendorst et al., 2021; Krebs et al., 2017; Sönmezer et al., 2021) to simultaneously measure the comprehensive methylation status of the bound enhancers as well as binding of TFs on individual DNA molecules *in vivo*. By this, I was able to resolve the epigenetic heterogeneity and associate enhancer activity or TF binding with the methylation status at a single molecule resolution. I applied this technology across the murine genome of pluripotent as well as somatic cell lines. I found that the vast majority of enhancers are not instructed by DNA methylation and can be active despite being methylated. Yet, I identified a small subset of cell-type specific enhancers in the different cell lines that are methyl-sensitive and respond to changes of the underlying DNA methylation. Furthermore, I identified known methyl-sensitive TFs such as CTCF and Max-Myc and revealed, that Max-Myc binding sites at enhancer are dependent on TET-mediated methylation turnover to remove DNA methylation in order for Max-Myc to bind.

3.1.1 | Moving epigenomics into the single molecule space

The results of my PhD dissertation provide evidence that overcoming the limitations of bulk-cell measurement techniques allows a deeper understanding of the co-dependencies of gene regulatory factors and is essential to detect subtle changes in regulatory activities.
Transcription regulation is a highly complex mechanism with many layers and factors (Kim and Wysocka, 2023). Thus, it becomes difficult to determine the hierarchies and what is cause and what is consequence. Since the development of genomic techniques to localize proteins and histone modifications along the genome, a large amount of factors have been mapped (Abascal et al., 2020; Luo et al., 2020). Yet, as it turned out, localization does not necessarily mean causation and may depend on the genomic context. In consequence, it is still largely elusive if epigenetic modifications are causal for transcription regulation or are merely consequences of the recruitment of large cofactor complexes (Isbel et al., 2022; Millán-Zambrano et al., 2022; Reiter et al., 2017).

Moving beyond bulk-cell techniques and to co-detect multiple regulatory features (e.g. epigenetic modifications and TF binding) at the same time, allows the evaluation of direct correlations and to understand the causal relationships between them. The SMF technology presented here has multiple advantages over commonly used genomic techniques (Krebs, 2021). In comparison to bulk-cell chromatin accessibility measurements such as ATAC-seq and DNase-seq, SMF detects all accessibility states at a high nucleotide resolution. With SMF, I was able to directly compare the methylation state in the accessible as well as in the inaccessible fraction of DNA molecules, allowing me to make statistical claims about all present states. Having only the accessible fraction – as it is the case for ATAC-seq and DNase-seq – neglects the cellular heterogeneity and removes important information from the analysis. When comparing, e.g., the two given examples of an enhancer with no 5mC-CA association (Figure 10E) and an enhancer with negative 5mC-CA association (Figure 11C), one can observe that both enhancers show comparable methylation states in their accessible fractions (28% vs 31%). Yet, one of them is methyl-sensitive, while the other one is not. This difference between the two sites is only revealed when co-analysing the inaccessible molecules: while the methylation-insensitive

enhancer also shows low methylation in the inaccessible fraction (27%), a large part of the inaccessible molecules at the methyl-sensitive enhancer are methylated (61%). This evidently shows that capturing and resolving cellular heterogeneity is an important feature of future genomics techniques.
Another advantage of SMF and its ability to capture all methylation and accessibility states is that one can use the natural epigenetic heterogeneity as an internal *in vivo* experiment in absence of perturbations. In most methylation studies, cells or individual loci are perturbed in their methylation state and subsequent changes in TF binding and accessibility are used to identify methyl-sensitive sites (Domcke et al., 2015; Grand et al., 2021; Héberlé and Bardet, 2019). Yet, with SMF I can make this comparison already in the unperturbed WT cells, since both methylation states are captured and can directly measure their effects. In consequence, the first read-out and detection step of methyl-sensitive enhancers is free of secondary effects that may come along with perturbation assays and allows to focus on naturally active enhancers. As discussed in the next chapter, analysis of large perturbation assays usually identified methyl-sensitive TFs based on new binding sites outside of their usually bound CREs. Yet, this data gives only limited insights into what is happening *in vivo* at active enhancers. SMF changes the view point from these spurious TFBSs to those within active enhancers.
Along with this comes the advantage of SMF to identify moderate changes in chromatin accessibility and TF binding. In comparison to common genomic techniques, that primarily provide enrichments, SMF provides high coverage quantitative read-outs that allow to detect and quantify even small changes. This is an important point to consider when investigating active CREs. At inactive sites the removal of DNA methylation may lead to a sudden binding of TFs that drives their activation. In contrast, at active sites, one does not expect such an on-off behavior, but rather a modulation of their activity. Thus, the expected changes in enhancer activity are much more moderate than at inactive sites, which are difficult to measure with common enrichment techniques.

A further advantage of SMF is its capacity to investigate enhancers genome-wide. A recent study dissected the impact of DNA methylation on two large enhancer clusters with a single-cell fluorescent reporter system and found that their activity is instructed by DNA methylation dynamics (Song et al., 2019). Although this study presents a new approach to study DNA methylation and makes an important contribution to the field, it has a locus-specific view point, preventing generalizable claims. Opposingly, using SMF I was able to investigate the methylation sensitivity of thousands of active enhancers genome-wide, allowing to make the general claim that a predominant number enhancers are not methyl-sensitive, while a cell-type specific subset is instructed by DNA methylation. Those can be used in further studies for deeper dissections of how the methylation dynamics change their activity.

In addition, the SMF-based analysis of DNA methylation is not limited to a certain cell type. A large fraction of methylation studies has been performed in mESCs, due to their viability upon DNMT KOs, in order to compare unmethylated and methylated states. Since SMF resolves the natural epigenetic heterogeneity, methylation perturbations are not needed for the primary identification of methyl-sensitive enhancers. Hence, one can move beyond mESCs and investigate somatic cell lineages as

shown in this study. This will, moreover, allow to expand the collection of methyl-sensitive TFs to those that are active in different somatic cell lines.

In the last years, single-cell techniques became much more prominent in the genomics field and are used to investigate similar questions as presented here (Preissl et al., 2023). However, although single-cell techniques have obvious advantages over bulk-cell techniques, at the current moment, they have still limited use in answering the same question to the same extent as SMF. One major drawback of single-cell techniques is their low coverage at individual loci. The here presented study is highly dependent on the high coverage provided by SMF (median coverage was around 100x [Figure 10E]) for individual cytosines to capture rare events and moderate changes. This cannot be achieved with current single-cell techniques for which coverages of less than 10x per locus are common at affordable costs. Nevertheless, they provide information about different genomic loci in the same cell, allowing to draw conclusions about global methylation states and co-accessibilities (see chapter 3.1.5).

In conclusion, SMF is a valuable advance in genomics to resolve epigenetic heterogeneity present at CREs and to capture rare TF binding events and moderate changes in TF occupancy and CA. In addition, it enable the investigation of co-occurring events at the same single molecules, allowing to dissect factor dependencies at high resolution genome-wide in pluripotent as well as somatic cell lineages.

3.1.2 | Methylation-sensitive TFs – *in vitro* does not mean *in vivo*

Large *in vitro* studies have supported the notion that the majority of TFs with a CpG in their target motif or vicinity are either positively or negatively affected by DNA methylation (Kribelbauer et al., 2017; Yin et al., 2017), yet the *in vivo* evidences for these methylation sensitivities remain rare (Héberlé and Bardet, 2019).
One difficulty in the analysis of TF sensitivity *in vivo* is that one has to compare the binding capacity of a TF to its motif in a methylated and unmethylated state. To achieve this, most studies have exploited mESCs that are viable upon DNMT TKO. In addition, most studies have used ChIP-seq or comparable protein mapping techniques to investigate the changes in TF occupancy and binding patterns. Yet, as discussed before, those enrichment techniques give only semi-quantitative output and often limit the analysis to spurious binding sites that become bound upon DNA methylation removal.
In this study, SMF is used to quantitatively measure TF binding and investigate the impact of CpG methylation within the motif or its vicinity in its *in vivo* chromatin context). Leveraging the natural epigenetic heterogeneity of DNA methylation at CREs allows to compare TF occupancies between the methylated and unmethylated fraction of molecules at each locus *in vivo* in absence of perturbations. In consequence, SMF enables to investigate the impact of DNA methylation on TF binding directly at active CREs, allowing to draw conclusions about whether DNA methylation contributes to the regulation of enhancers.
The here presented SMF data reveals that most TFs are not methyl-sensitive *in vivo* and that the activity of the vast majority of enhancers is independent of the underlying DNA methylation (Figure 12A, 20A-C and 21B). This stands in contrast to the results of the large *in vitro* studies (Kribelbauer et

al., 2017; Yin et al., 2017) and suggests *in vitro* identified methylation-sensitivity may not affect *in vivo* binding patterns. A simple explanation for this disagreement could be that the TFs do not encounter the *in vitro* tested methyl-sensitive motifs in the *in vivo* context, either because they are inaccessible, or outside of CREs or always in a non-methylated state. In consequence, despite putative methylation-sensitivity *in vitro*, DNA methylation can have little impact on TF binding patterns *in vivo*.

Nevertheless, I identified multiple TFs that have been associated with methylation-sensitivity in the past as e.g. CTCF (Maurano et al., 2015), the general transcription activator Max-Myc (Yin et al., 2017) and NRF1 (Domcke et al., 2015) as well as REST which has been shown to be insensitive to DNA methylation but drives local DNA demethylation upon binding (Stadler et al., 2011), thus supporting the *in vivo* studies. Interestingly, for those methyl-sensitive TFs, I observed different distributions of average DNA methylation (Figure 23A), which suggests that *in vivo*, the chromatin context plays an important role for the consequences of a TF's methylation-sensitivity.

For NRF1, it has clearly been shown that DNA methylation inhibits its binding to spurious sites in the genome and together with my data it becomes clear that NRF1 primarily binds to unmethylated sites. The strong methylation-sensitivity could be related to NRF1's motif sequence that contains multiple CpG sites. In contrast, CTCF can bind independently of the underlying DNA methylation, depending on the motif sequence (Figure 23F) (Maurano et al., 2015) and thus, the genomic context. In addition, the binding frequency of CTCF can be modulated by local DNA methylation at some TFBS, where the methylated CpG is at a specific position in the motif (Figure 23H). In comparison, Max-Myc also shows different context-dependent methylation-sensitivities, which are, however, not related to Max-Myc's motif sequence (Figure 23G). The occupancy of Max-Myc is tuned by the local DNA methylation and loss of TET enzymes abolishes its binding (Figure 23I and 24B-D). This suggests that the recruitment of TET enzymes plays a major role for Max-Myc binding *in vivo*. Similar conclusions have been drawn in a previous study investigating the effect of TET2 loss in hematopoietic stem cells (Rasmussen et al., 2019). Using ATAC-seq, they found decreased CA in TET2 KO vs WT cell lines at E-Box motifs, but were not able to detect direct changes of Myc occupancy. Yet, transcriptome analysis showed a decreased expression in Myc-targeted genes upon TET2 KO, arguing that loss of TET2 impairs activation of CREs by Myc. This is in line with reports showing TET2 recruitment to the promoters of Myc-targeted genes by the transcriptional co-activator SMAD nuclear interacting protein 1 (SNIP1) in cancer cells (Chen et al., 2018). Using SMF, I was now able to directly show *in vivo* that loss of TET enzymes has indeed a direct effect on Myc occupancy.

Based on the observation that the average DNA methylation levels vary at Max-Myc TFBSs, I would speculate that, at sites with intermediate methylation levels, the activity of the CRE is dependent on the binding of Max-Myc, while at the unmethylated TFBSs, Max-Myc is co-binding with other methyl-insensitive TFs that open and demethylate the region prior to Max-Myc binding. This hypothesis needs to be confirmed in further experiment as e.g. perturbations of Myc or other co-binding TFs.

In conclusion, my study shows that some TFs are methyl-sensitive *in vivo* and that the regulation of their binding levels by DNA methylation can be context specific. Using high resolution single molecule data, I was able to show that DNA methylation not only represses cryptic binding sites, but also plays a role in modulating the binding intensity of TFs at enhancers. One of those TFs is Max-Myc that has

been shown to be a transcriptional amplifier that broadly enhances the transcription levels of active genes and promotes cell proliferation, which has important implications in cancer cells (Kress et al., 2015; Lin et al., 2012; Nie et al., 2012). Thus, it can be hypothesized that by controlling the binding levels of Max-Myc, DNA methylation participates in the global control of transcription levels.

Previous *in vitro* as well as individual *in vivo* studies have also shown that certain TFs preferentially bind their motifs in a methylated state (Kribelbauer et al., 2017; Yin et al., 2017). The SMF data from my study does not show evidence for this. Although a positive 5mC-CA association was detected for a small fraction of CREs, further analysis revealed that the CpGs were enriched at the linker DNA at the edge of strongly positioned nucleosomes (Figure 12C-F). This leads to the conclusion that the positive association is not evidence for preferential TF binding to methylated DNA, but rather a consequence of specific chromatin organization. In addition, the TFBS analysis revealed little evidence for positively associated TFs. The reasons for this may lay in the sequence motif constraints for the SMF analysis and other limitation discussed later (see chapter 3.1.4).

3.1.3 | DNA methylation at enhancers – mostly non-instructive

A large fraction of DNA methylation is focused on specific genomic elements like promoters, repetitive elements and ICRs, where the methylation state acts as an important repressive mark (Baubec and Schübeler, 2014). At those sites, DNA methylation levels are quite homogeneous with either high or low levels (Stadler et al., 2011). In contrast, distal CREs like enhancers show intermediate methylation levels that represent an epigenetic heterogeneity within the cellular population (Baubec and Schübeler, 2014). Since the resolution of this heterogeneity is pivotal to study DNA methylation at enhancers, it has been difficult in the past to determine whether enhancer activation is instructed by DNA methylation (Luo et al., 2018; Mattei et al., 2022).
In the here presented study, I resolved the epigenetic heterogeneity using a single-molecule approach and performed simultaneous measurements of DNA methylation, CA and TF binding at enhancer elements. This resolution enabled me to investigate the impact of DNA methylation on enhancer activity *in vivo* in mESCs and multiple somatic cell lines in the absence of perturbations. I discovered that the vast majority of active enhancers are not instructed by DNA methylation. Yet, for ES cells and three somatic cell types, I identified active enhancers where transcription factor occupancy is directly repressed by DNA methylation, including enhancers involved in the control of key cell identity genes. These results are in line with previous studies exploring DNA methylation at enhancers during differentiation (Barnett et al., 2020) and at single enhancer clusters associated with pluripotency (Song et al., 2019). In the end, these results suggest that presence of DNA methylation does not generally impede enhancer activity, but plays a crucial role in regulating the activity of key cell-type specific genes important for the acquisition and maintenance of cellular identities.

However, the question remains what determines the DNA methylation dynamics at enhancers to ensure proper activation. My finding that TET enzymes play an important part in the active demethylation of methyl-sensitive enhancers supports former studies that show high TET occupancy

and high DNA methylation turnover at active enhancers (Ginno et al., 2020). Therefore, two scenarios can be hypothesized on how the methylation dynamics are regulated at methyl-sensitive enhancers.
On the one hand, a preceding event can lead to the recruitment of TET enzymes to methyl-sensitive enhancers to drive local demethylation and let methyl-sensitive TFs bind and activate the enhancer. This would mean that in cells where local methylation is detected, TETs have not been recruited.
On the other hand, DNMTs and TETs could both occupy enhancer elements leading to a constant methylation turnover, whose dynamics are defined by the concentration and activity of the enzymes at each binding site. In this scenario, cells with local methylation are in the methylated-state of the turnover process and importantly, slight changes in enzyme occupancy or activity can push the methylation state in the one or the other direction. At most enhancers, methyl-insensitive TFs can bind at any time in this turnover cycle and activate the enhancer, leading to the recruitment of additional TET and other chromatin modifying enzymes that shift the methylation state further towards the unmethylated state. Subsequently, additional TFs are able to bind to the now unmethylated enhancer, including methyl-sensitive TFs. Those enhancers are suggestively dependent on the initial binding of methyl-insensitive TFs. In contrast, a fraction of enhancers would be dependent on methyl-sensitive TFs. Those are much more dependent on the TET-mediated demethylation during the methylation turnover process. Thus, a slight increase in TET occupancy would also increase the window of opportunity for those TFs to bind and to activate the enhancer.
This second described scenario is, in my opinion, much more likely since it needs only small changes in TET or DNMT occupancy to shift the TF binding outcome for methyl-sensitive enhancers. Which of these two scenarios is taking place at mammalian enhancers needs to be investigated in future studies such as tethering of TET and DNMT enzymes to specific enhancers or by perturbing the nuclear concentration of these enzymes. Yet, in both cases DNA methylation present an additional regulatory layer at a subset of key enhancers to tune the binding of TFs and, thus, the activity of the enhancer.

One interesting observation that has been made in the perturbation assay is that the effects on enhancer activity were much more prominent in TET TKOs than in DNMT TKOs. On one hand, this observation can have technical reasons as active enhancers already show low methylation levels (10-60%). In consequence, a loss of DNA methylation only affects a small to medium proportion of cells and the maximum change in methylation that we can observe is 60%. A loss of DNA methylation at lowly methylated sites will, rationally, have much lower effects on the CA and TF binding. In contrast, in TET TKO cells the change in methylation can be up to 90%, increasing the potential effect it can have.
On the other hand, a more biological reason may account for this differential behavior. Enhancers show high methylation dynamics and turnover. As discussed above, TET-mediated demethylation provides a window of opportunity for methyl-sensitive TFs to bind and open the region for potentially more TFs. The loss in DNMTs and loss in DNA methylation may widen the window of opportunity, but does not necessarily mean that a higher amount of the methyl-sensitive TF will bind. This may be affected by the local TF concentration and competition with other TFs.

3.1.4 | Current limitations of SMF

SMF is a valuable tool for the quantitative analysis of co-occurring events of TF binding and epigenetic modifications. Yet, the SMF analysis approach is generally limited to certain genomic contexts and has, in its current form, certain limitations that restrict the number of CREs as well as the number and identity of TFBSs to analyse.

First of all, the basic concept of this SMF approach is based on the naturally occurring heterogeneity. Thus, only those genomic regions can be analysed that show intermediate methylation levels, excluding for instance promoter regions. The current statistical approach restricts the analysis to sites with average methylation levels between 10% and 90%, when below or above that, the applied proportionality test fails.

Second, SMF is highly dependent on sufficient read coverage at each locus tested in order to call the 5mC-CA association with high confidence. To achieve these high confidence calls in the large mammalian genome, the SMF procedure includes a DNA capture step to enrich the sequencing library for CREs. Although this approach covers a large set of active enhancers, around 25% of enhancer in mESCs are lost upon this complexity reduction. To investigate sites that are currently not part of the enriched regions or show only low coverage, one would need to design additional baits or perform locus-specific amplicon SMF (Kleinendorst et al., 2021).

One major drawback of the current version of SMF is the sequence restriction to GpCs. Hence, only those genomic loci can be analysed that contain at least one GpC within the collection bin (101 bp) (see chapter 4.2.3). For the TFBS analysis, at least one GpC must be present in each of the three collection bins (centered on motif, and up- and downstream) (see chapter 4.2.20). Unfortunately, the mammalian genome is depleted of cytosines in comparison to species without DNA methylation like *Drosophila melanogaster* due to the mutagenic potential of DNA methylation (Holliday and Grigg, 1993), already limiting the number of sites. In addition, due to the strand collapsing procedure in the single molecule analysis and the off-target methylation of the M.CviPI enzyme in CCG context (Kelly et al., 2012) (Figure 10D), additional sequencing contexts have to be removed from the analysis, leaving only the NWCGW (N = any base, W = no C/G) context for CpGs and the DGCHN (N = any base, D = no C, H = no G) context for GpCs. These limitations largely restrict the number of TF binding sites and CREs that can be investigated. In consequence, 18% of the active enhancers in mESCs cannot be analyzed, and from all ChIP-seq-validated TFBS containing a CpG, I can analyse <10%. Especially for the analysis of TFBS, this is an important consideration to make and explains why I do not recapitulate all methyl-sensitive TFs and why my study does not show evidence for methylation-preferred TF binding.

Moreover, the low GpC density could potentially lead to false positive or false negative associations in the CA analysis. The current analysis does not include a minimum GpC threshold. Thus, the CA calling of some CREs is dependent of only one or two GpCs, which may not represent the accessibility of the entire region. Including a threshold would reduce the number of false associations, but would also reduce the number of detectable sites. This limitation could be overcome by advancing SMF to use additional base modifications as accessibility read-out, e.g., by methylating adenines (see chapter 3.1.5).

In relation, the window used for the CA call could be further optimized. In the current analysis, 101 bp represented a good compromise between multiple factors to consider. Those include the occupancy of nucleosomes at the border of accessible regions and the occupancy of TFs within the accessible region, both of which result in low GpC methylations that contaminate the high GpC methylation of an accessible region. To optimize the analysis of accessible CREs, one option would be to first annotate nucleosomes surrounding active CREs and subsequently analyse the accessible area in the center. This nucleosome-driven accessibility determination would allow CA analyses without a stringent size cutoff and could potentially improve the statistical power due to lower contaminating reads. Yet, this analysis would require the unsupervised detection of nucleosome footprints and their boundaries which can be difficult with the given read length. Therefore this approach was not feasible in this study but would be a great implementation in future SMF analyses.

Next to the GpC content, the TFBS analysis is further limited by the dependency of ChIP-seq or CUT&RUN data to infer TF identity. Here, a larger set high quality sequencing data as well as better motif-TF assignments would help to investigate the methylation effect on more TFs. In addition, unsupervised non-data-driven detection of footprints and their correlation with DNA methylation would enable a more global picture of the relationship between DNA methylation and TFs, yet, with limited information about the protein identity. This unsupervised footprint detection is currently being implemented in the lab.

3.1.5 | Future directions

In this study, I identified methyl-sensitive enhancers and TFs in pluripotent as well as somatic cell lines, confirmed them by global perturbation assays and inferred that TET enzymes play a crucial role for the local regulation. In the future, multiple additional experiments could be done to further dissect and refine the list of enhancers and TFs as well as the mechanism that determines the activation of enhancers by DNA demethylation.

Global perturbations are prone to have secondary effects on gene expression and chromatin state, which can be hard to disentangle from primary ones. To limit these secondary effects, local perturbations using, e.g., a dead Cas9 system would be a good approach to tether DNMTs or TETs to individual sites (Carlini et al., 2022; Morita et al., 2016) and measure their local effects with amplicon SMF (Kleinendorst et al., 2021). Yet, even the newest advances of this system achieve only limited increase in DNA methylation (Policarpi et al., 2022), which may not be sufficient to observe relevant changes in chromatin accessibility and TF binding depending on the locus and context. In addition, the tethering complexes are quite large, potentially preventing the binding of TFs or the methylation by SMF enzymes. In contrast to the *trans* approach, one could also perturb the system in *cis* by single-site mutations of TF motifs to either disrupt TF binding or lower TF affinity to investigate their impact on local DNA methylation levels.

One could also explore further the cross-talk between multiple enhancers and promoters and potential co-occurring dynamics. Going in this direction, one could explore single-cell WGBS (Farlik et al., 2015) or even single-cell nucleosome, methylation and transcription sequencing (scNMT-seq) data (Clark et al., 2018) to evaluate whether all active methyl-sensitive enhancers show similar methylation levels

in a cell population and how waves of DNA methylation during development affect the activity of methyl-sensitive enhancers. Another approach in this direction would be the use of long-molecule sequencing such as Oxford Nanopore technology (Liu et al., 2021) together with SMF to investigate the synchronization of methylation patterns of neighboring enhancers on a single molecule basis. By this, one could investigate whether multiple enhancers controlling the same gene are methylated equally when active, whether their methylation levels drop simultaneously and how those changes are coordinated with TF binding and promoter activation. Evidence for those coordinated methylation dynamics have already been shown using Methyl-HiC (Li et al., 2019), hence, it would be interesting to combine long-molecule sequencing of DNA methylation with simultaneous read-outs of accessibility and TF occupancy.

The use of long-molecules sequencing techniques would also allow the advancement of SMF to include other base-modifications such as adenine methylation (6mA), which can be directly read out by those techniques. 6mA-mediated footprinting would achieve a much higher resolution of the TF binding and the CA and would allow a considerably larger number of regions and TFBS to be analysed.

Lastly, the here developed SMF approach allows the identification of methyl-sensitive CREs and TFs in absence of perturbations, thus enabling the analysis of any cell line. Investigation of high coverage SMF data in a wide-range of cell lineages would allow the identification of so-far unknown methyl-sensitive TFs.

3.1.6 | Impact on other fields of research

The results of the here presented study have important implications for a wide range of research fields ranging from developmental research to neuroscience and translational research. Changes in DNA methylation have been associated with aging and a wide-spectrum of diseases including cancer, yet the causal relationship still remains unclear (Greenberg and Bourc'his, 2019; Nishiyama and Nakanishi, 2021; Seale et al., 2022; Younesian et al., 2022). Are changes in DNA methylation dynamics a cause for diseases and aging or are they merely a consequence of other chromatin changes? For example, hyper- and local hypo-methylation are a typical observation in cancer epigenomes (Nishiyama and Nakanishi, 2021). A recent study has found that these changes in DNA methylation can be in part driven by deregulated binding of TFs that guide local DNA demethylation (Detilleux et al., 2022). In this case, DNA methylation serves as a marker for cancer onset or progression, but would only be a consequence of the dysregulation of the TF itself.

My study adds another viewpoint to the cause-consequence relationship. When most enhancers are methylation-insensitive, do changes in DNA methylation in cancer epigenomes actually have an impact on the transcriptional dysregulation? Using the here presented approach and concept enables to determine which methylation changes could have transcription effects and which are just markers without effects on the cancer transcriptome.

The same viewpoint can be applied to the aging field, in which the concept of the "epigenetic clock" suggests that changes in DNA methylation are involved in aging processes and that reversal of the epigenome to its earlier state could lead to cellular rejuvenation (Seale et al., 2022). My data argues

against this concept, since most methylation changes at enhancers have little to no consequences on chromatin state.

In conclusion, my PhD study contributes to a more differentiated view on DNA methylation that suggests that DNA methylation – at least at enhancers – Is most often the consequence of other genetic or epigenetic changes and does not necessarily contribute to the onset, progress and severity of a disease.

3.2 | The regulatory role of histone modifications on gene regulation

Post-translational modifications of the histones by transcriptional cofactors represent an additional layer of gene regulation (Kim and Wysocka, 2023; Stillman, 2018). Many cofactors and protein domains have been identified that can modify histones at their protruding tail or globular core. Different histone modifications have been associated with either active or repressive genomic regions, but it is unclear to which extent they contribute to the regulation of gene expression or whether they are laid down in consequence of transcriptional activity where they serve as mark and memory for chromatin state. In this regard, also the relationship between histone modifications and TF binding is still elusive. Little research has been performed looking at the regulatory consequences of histone modifications on TF binding and vice versa. One major difficulty in the investigation of this questions is the heterogeneity in histone modification and TF binding within cellular populations. So far, most studies have focused on bulk enrichment analyses and used local correlation between the different factors as means of causation. Yet, as for the DNA methylation studies, those experimental approaches do not account for the epigenetic heterogeneity and cannot determine co-occupancy of the different factors within a single cell.

In the here presented histone modification project, I aimed to overcome these limitations by developing a method that enables the simultaneous measurement of histone modifications and TF binding at the same single DNA molecules. My objective was to combine CUT&RUN with SMF in order to enrich for the histone modification of interest and to investigate the TF binding patterns in this fraction of molecules in comparison to the non-enriched bulk molecules. This would enable the association of TF binding patterns with specific histone modifications and determine cause-consequence relationships as well as context-specific interdependencies. This methods approach was however unsuccessful due to the low signal-to-noise ratio and insufficient enrichment by CUT&RUN in comparison to ChIP, especially after the combination of CUT&RUN with SMF and bisulfite sequencing. A recent study has identified a generally low correlation between ChIP-seq and CUT&RUN determined peaks (Hu et al., 2023). In consequence, the project was stopped to focus on the DNA methylation study.

In the future, other approaches are needed to answer these important research questions about the causal effect of histone modifications. One option could be to focus on the CUT&Tag technology, which is working similarly, but has been shown to be more efficient for the enrichment of histone modifications and shows lower signal-to-noise ratios (Kaya-Okur et al., 2019). In addition, CUT&Tag

has recently been successfully coupled with bisulfite sequencing (Li et al., 2021), which would already overcome one of the challenges. Nevertheless, it would need to be determined whether CUT&Tag coupled with bisulfite sequencing reaches sufficient enrichments for the correlation analyses, in contrast to my attempts with CUT&RUN. Another approach would integrate SMF with the recent development of Directed Methylation with Long-read sequencing (DiMeLo-seq) (Altemose et al., 2022). This technology uses an antibody-binding fusion construct that generates genomic 6mA footprints according to the presence and location of a histone modification of interest and enables the single-molecule read-out using Nanopore sequencing (Liu et al., 2021). This double footprinting approach would allow a direct associated between the presence of histone modifications and TF binding as done here with DNA methylation. In addition, the use of Nanopore would enable to simultaneously analyse multiple individual enhancers, enhancer clusters as well as enhancer-promoter associations.

3.3 | Epigenetics and its contributions to gene regulation

In the end, the fundamental question remains: Do epigenetic modifications have a causal role for the regulation of gene expression at enhancers?

The idea of the histone code in which a specific set of histone marks defines the regulatory outcome as proposed by Strahl and Allis (Strahl and Allis, 2000) has long been overcome (Henikoff and Shilatifard, 2011). In addition, a decade of studies in which modifications and cofactors have been mapped throughout the genome, which have revealed large redundancies between factors (Abascal et al., 2020). Now, new approaches are used to carefully dissect the individual layers and their interdependencies. Dependent on the viewpoint, the experiments provide different hypotheses on whether individual epigenetic modifications and cofactors have instructive roles or whether they are merely important for the memory of state.

For cofactors, for example, two recent studies provide evidence that the general regulatory potential of common cofactors is context-dependent. Extensive dissections have shown that some cofactors are globally important for the activation of CREs, while others are fully dispensable or act only in specific CRE contexts (Haberle et al., 2019; Neumayr et al., 2022). Surprisingly, even the well-studied Mediator complex, that has been believed to be generally important for enhancer function, was dispensable in a specific set of enhancers (Neumayr et al., 2022). This shows clearly, that we are overestimating the generality of regulatory factors and that context plays a much more important role (Kreibich and Krebs, 2022; Reiter et al., 2017). In addition, those studies have been performed episomally, hence, outside of the chromatin context, likely even underestimating the complexity of context-specific dependencies.

Similar conclusion can be drawn from studies dissecting the impact of histone modifications. H3K4me3, for instance, has been associated with promoter activation and has been shown to be directly involved in the recruitment of the PIC (Lauberth et al., 2013; Vermeulen et al., 2010) and inhibition of *de novo* methylation (Ooi et al., 2007). Yet, it has also been shown to be a result of transcription (Wang et al., 2022), arguing against an instructive role of H3K4me3 for transcription

activation (Howe et al., 2017), with new evidences suggesting a more important role in regulating the promoter proximal pause-release step of RNA Polymerase II for transcription elongation (Wang et al., 2023). A recent study directly dissected the instructive potential of H3K4me3 in a single-cell reporter assay and found that H3K4me3 has indeed activation capacities, but only in a subset of CRE contexts dependent on the TF motif composition (Policarpi et al., 2022).

In relation, H3K4me1 is usually associated with active and poised enhancers, but has been shown to be largely dispensable for enhancer activation itself, in contrast to the methyltransferases that deposit H3K4me1 (Dorighi et al., 2017). And also for H3K27ac, a mark that has been widely associated with activation of enhancers and promoters, it has been shown that the loss of this acetylation has little effect on the remaining chromatin landscape at active CREs and transcriptional output (Zhang et al., 2020b). Yet, deposition of H3K27ac can drive transcription activation in a context-dependent manner, similar to H3K4me3 (Policarpi et al., 2022).

In addition, interdependencies, synergistic action as well as redundancies between chromatin modification may play a role in driving a quantitative transcriptional output. As, for instance, in the case for H3K27ac, where its loss is potentially backed-up by acetylation marks at other histone tail residues (Zhang et al., 2020b) or in the interplay between DNA methylation and H3K27ac (Cusack et al., 2020; Song et al., 2019) or H3K4 methylation (Millán-Zambrano et al., 2022). In other cases, known interdependencies can also be context-specific as recently shown for H3K4me1 and H3K27ac in mESCs where the loss of H3K4me1 has only limited impact on H3K27ac deposition and transcriptional activation (Boileau et al., 2023).

The context-specific impact is similarly observable for DNA methylation at enhancer elements. On the one hand, genome-wide analysis of enhancer activity during differentiation (Barnett et al., 2020) or loss of the modifying enzymes can show little instructive behavior of DNA methylation. On the other hand, single loci dissections of cell-type specific enhancers like the two core pluripotency enhancers *Sox2* and *Mir290* (Song et al., 2019) show dependencies on DNA methylation dynamics. This context-specific impact of DNA methylation is further supported by the here presented study that shows that in the vast majority, DNA methylation is not instructive, while certain cell-type specific sets of enhancers are regulated by local DNA methylation levels.

In summary, my data in combination with our current knowledge on cofactors, histone modifications as well as DNA methylation support the hypothesis that on a global scale, the epigenetic contribution to gene transcription regulation is largely redundant and based on an interplay between multiple contributors. On a single locus scale, however, the individual chromatin modifications can play important and instructive roles and by this, tune the quantitative expression of important genes. In the case of DNA methylation, it seems that it especially impacts cell-type specific genes and, thus, contributes to the spatiotemporal determination and maintenance of cell lineages. In the end, a similar and resolution of the epigenetic heterogeneity and the deconstruction of which loci are dependent on which histone modifications will be important to understand the local interplay of epigenetic factors and how they contribute in concert to the spatiotemporal expression of genes and the establishment of different cell lineages.

4 | Methods and Materials

4.1 | Experimental Methods

4.1.1 | Cell culture

The study utilized four different types of mouse embryonic stem cells (mESCs): male 159 wild-type, DNMT triple-knockout (Domcke et al., 2015), TET triple-knockout (Ginno et al., 2020), and F1 hybrid (129/CAST) (Giorgetti et al., 2016) cells. These cells were grown on 10 cm culture-plates coated with 0.2% gelatin in ES medium, consisting of Dulbecco's Modified Eagle Medium (DMEM) supplemented with 15% Fetal Bovine Serum (FBS), Leukemia Inhibitory Factor (LIF), 2-Mercaptoethanol, 2 mM L-Glutamine, and 1x non-essential amino acids (NEAA). The cells were maintained at 37°C and 5% CO2, with the medium being changed daily and cells being split every other day.
C2C12 murine myoblasts (a gift from Moritz Mall, DKFZ) were grown on collagen-coated plates in low glucose DMEM medium/F10 nutrient mix supplemented with 20% FBS and 2 mM L-glutamine.
MEL (murine erythroleukemia) cells (DSMZ, #ACC501) were grown in DMEM without Sodium Pyruvate media supplemented with 20% FBS and 2 mM L-glutamine.
The *in vitro* differentiation of ES cells into neural progenitors (NPs) was performed as previously described (Bibel et al., 2007).

4.1.2 | Single Molecule Footprinting

Single Molecule Footprinting (SMF) with Targeted Enrichment was conducted as described in previous studies (Kleinendorst et al., 2021; Sönmezer et al., 2021). In short, for each reaction, 250,000 cultured cells were harvested using trypsin and washed twice with 1x PBS. The cell pellets were resuspended in ice-cold lysis buffer, incubated for 10 min on ice, and centrifuged at 1,000x g at 4°C for 5 min. The nuclei were resuspended in 94.5 µl 1x M.GpC buffer.
The GpC methyltransferase treatment was performed by adding 150 µl of the freshly made GpC methyltransferase mix (consisting of 1x M.GpC buffer, 300 mM sucrose, and 64 µM SAM) and 200 U M.CviPI to the nuclei. The mixture was incubated at 37°C for 7.5 min and replenished with 100 U M.CviPI and 128 pmol of SAM for a second incubation round at 37°C for 7.5 min. 300 µl of the prewarmed stop solution (20 mM Tris-HCl (pH 7.9), 600 mM NaCl, 1% (wt/vol) SDS and 10 mM EDTA) and 6 µl proteinase K (20 mg/ml stock) were added and incubated overnight at 55°C.
The following day, the DNA was extracted using phenol-chloroform and treated with RNase A at 37°C for 30 min. 1 to 2.5 µg of DNA was fragmented into 300 bp fragments via sonication using the Covaris model S2 and bait-captured against *cis*-regulatory elements using the SureSelect XT Mouse Methyl-Seq Kit Enrichment System for Illumina Multiplexed Sequencing Library protocol as previously described (Kleinendorst et al., 2021). The library DNA was bisulfite converted using the ZYMO EZ DNA Methylation-Gold Kit according to the manufacturer's protocol. PCR amplification and indexing was

done using the SureSelect XT Mouse Methyl-Seq Kit. The prepared libraries were run on an Illumina sequencing platform using a NextSeq500 High 150 bp paired-end mode. SMF experiment were performed together with Rozemarijn Kleinendorst.

4.1.3 | PRO-seq

Polymerase run-on sequencing (PRO-seq) was performed as previously described (Mahat et al., 2016) with some minor modifications. In contrast to the original protocol, mouse ES cells were harvested with Trypsin-EDTA on ice and washed twice with ice-cold 1x PBS, followed by centrifugation at 200x g at 4°C. The cells were incubated in cold permeabilization buffer (10 mM Tris-HCl pH 7.4, 300 mM sucrose, 10 mM KCl, 5 mM MgCl2, 1 mM EGTA, 0.05% Tween-20, 0.1% NP40 substitute, 0.5 mM DTT, 1 tablet of PIC per 50 ml, 4 units/ml RNase inhibitor) on ice for 10 min, centrifuged at 1000x g at 4°C and washed once with the same buffer. The permeabilized cells were then resuspended in storage buffer (10 mM Tris-HCl, pH 8.0, 25% (vol/vol) glycerol, 5 mM MgCl2, 0.1 mM EDTA, and 5 mM DTT), counted using DAPI staining, snap-frozen in liquid nitrogen and stored at -80°C.

For the experiment, 10 million cells were used per sample, with a 5% spike-in of permeabilized Drosophila KC cells (0.5 million cells). Three replicates were performed for each ES cell line, one replicate for all cell lines at a time.

The nuclear run-on reaction was performed as a 2 biotin run-on, in which 28 µl of the 2x nuclear-run on master mix (10 mM Tris-Cl pH 8.0, 5 µM MgCl2, 1 mM DTT, 300 mM KCl) was mixed with biotin-11-UTP (1 mM), biotin-11-CTP (1 mM), ATP and GTP (10 mM each), H2O, RNase inhibitor and 2% Sarkosyl. The reaction mix was pre-heated at 37°C, the calculated number of cells added, and incubated at 37°C for 5 min with soft shaking (300 rpm). Divergent from the original protocol, the reaction was stopped by adding 700 µl of the RNA Lysis Buffer from the ZYMO Quick-RNA Miniprep kit and vortexing. RNA extraction was performed according to the manual of the kit. The final RNA was eluted from the columns with two times 50 µl H2O and pooled to a final volume of 100 µl.

To perform the base hydrolysis, the RNA was denatured for 40 s at 65°C and snap-cooled on ice. 25 µl of ice-cold 1N NaOH were added and incubated on ice for 10 min. The RNA was precipitated by adding Tris-HCl (pH 6.8), NaCl, GlycoBlue, and 100% EtOH, followed by centrifugation at 20,0000x g at 4°C for 20 min. The RNA pellet was washed with 70% EtOH, air-dried, and resuspended in 50 µl H2O plus 1 µl RNase Inhibitor.

The first round of biotin RNA enrichment was performed by adding 50 µl pre-washed streptavidin beads to the RNA and incubating it on a rotator (8 rpm) at room temperature (RT) for 20 min. Afterwards, the beads were washed twice with ice-cold high-salt buffer (50 mM Tris-HCl [pH 7.4], 2 M NaCl and 0.5% (vol/vol) Triton X-100), twice with ice-cold binding buffer (10 mM Tris-HCl [pH 7.4], 300 mM NaCl and 0.1% (vol/vol) Triton X-100), and once with low-salt buffer (5 mM Tris-HCl [pH 7.4] and 0.1% (vol/vol) Triton X-100) before cleaning up the RNA using Trizol and chloroform as described above.

For the 3'RNA adaptor ligation, the RNA pellet was dissolved in 4 µl of VRA3 3'RNA adaptor dilution (10 µM), heat-denatured at 65°C for 20 s and placed on ice. To this, 6 µl of the RNA ligation mix containing the T4 RNA ligase (1x T4 RNA ligase buffer [20x], 1mM ATP, 10% PEG, 4 U/µl RNase

inhibitor, 1 U/µl T4 RNA ligase I) were added and incubated at 20°C for 4 hours. The reaction was kept overnight at 4°C.

The next day, the second round of biotin RNA enrichment and RNA clean-up was performed as described above. The final RNA pellet was dissolved in 5 µl H2O, heat-denatured at 65°C for 20 s and snap-cooled on ice. The following 5' cap repair reaction was performed by mixing the RNA with 5 µl of the enzyme mix containing RppH and ThermoPol Reaction Buffer (50% H2O, 1x ThermoPol Reaction Buffer [10x], 2 U/µl RNase inhibitor, 0.5 U/µl RppH) and incubation at 37°C for 1 hour. For the hydroxyl repair, 90 µl of the PNK mix (1x PNK buffer [10x], 1mM ATP, 1 U/µl RNase inhibitor, 0.25 U/µl PNK, H2O to 90 µl) were added and incubated at 37°C for 1 hour. The RNA was cleaned-up using Trizol and chloroform as described above.

For the final 5'RNA adaptor ligation, the RNA pellet was dissolved in 4 µl of VRA3 5'RNA adaptor dilution (10 µM), heat-denatured at 65°C for 20 s and placed on ice. Again, 6 µl of the RNA ligation mix were added and incubated at 20°C for 4 hours. The reaction was kept at 4°C overnight.

The next day, biotin RNA was enriched a third time as described above and the final RNA pellet dissolved in 10 µl H2O. Reverse transcription (RT) of the final RNA was performed with the Superscript IV transcriptase and the RP1 reverse-transcription primer. For this, 2.5 µl of the RT primer mix (0.5 µl RP1 reverse-transcription primer [100 µM], 1 µl 12.5 mM dNTP mix, 1 µl H2O) were added to the RNA, heated up to 70°C for 2 min and chilled on ice for 2 min. After a quick spin-down, 6 µl RT buffer mix (4 µl SSIV Buffer [5x], 1 µl 0.1 M DTT, 1 µl RNase inhibitor) were added and incubated at 37°C for 5 min. Then 1 µl of the SuperScript IV Reverse Transcriptase (200 U/µl) were added and the sample mixed. The mixture was incubated for 10 min at 45°C, for 30 min at 50°C and for 10 min at 80°C. Afterwards, 6 µl H2O were added to the reverse-transcribed cDNA.

The correct number of PCR cycles was determined by test PCR and Bioanalyzer analysis. In the end, the final library was PCR-amplified for 14 cycles and cleaned-up using SPRI beads at a ratio of 1.8x. To remove primer dimers, the library was further size selected with SPRI beads at ratio of 1.3x. Finally, it was sequenced on a NextSeq 2000 P3 with 50 bp single-end conditions.

4.1.4 | CUT&RUN – native and bis

CUT&RUN was performed as previously described with some minor changes (Skene and Henikoff, 2017). The used pA-MNase fusion protein was produced by the PEPCore facility at EMBL and was stored as a 0.5 mg/ml stock in 50% glycerol at -20°C. The experiment was performed with fresh mESC cultures, using 1 million cells per sample. First, cells were harvested, counted and the appropriate number of cells for all samples washed twice with 1x PBS. To account for loss and pipetting errors, 1 million more cells than needed were used, plus 10%. The cells were centrifuged for 3 min at 600x g at RT and resuspended in 1.4 ml of nuclear extraction (NE) buffer (20 mM HEPES-KOH [pH 7.9], 10 mM KCl, 0.5 mM Spermidine, 0.1% Triton X-100, 1x cOmplete EDTA-free mini) by gentle pipetting. After centrifuging for 5 min at 3200 rpm at RT, the cells were resuspended in 1.2 ml of NE buffer. 1 mio nuclei per sample were distributed into 1.5 ml tubes. 250 µl of binding buffer (20 mM HEPES-KOH [pH 7.9], 10 mM KCl, 1 mM CaCl, 1mM MnCl2) were added to each sample (final volume 400 µl).

To prepare the magnetic beads, Bio-Mag Plus Concanavalin A coated beads (Polysciences, Inc. #86057) were gently resuspended, and 100 µl bead slurry (50 µl beads bind up to 10 million cells) were withdrawn and transferred to 900 µl binding buffer in a 2 ml tube. The tube was placed on a strong magnet stand, and the beads were washed twice in 1 ml binding buffer. The beads were resuspended in 1.4 ml binding buffer (200 µl/sample). 200 µl of the bead slurry were slowly added to the nuclei (400 µl), while gently vortexing. Samples were rotated for 10 min at RT on a rotor (set to 25).

To prevent early activation of the MNase, the sample was treated with a blocking solution prior to antibody and pA-MNase treatment. The tube was placed on the magnet stand and the liquid was pulled off. 1 ml of blocking buffer (wash buffer with 2 mM EDTA) was added and mixed by inverting 10x. The samples were incubated for 5-10 min at RT while rotating. Antibodies were prepared in wash buffer (20 mM HEPES [pH 7.5], 150 mM NaCl, 0.5 mM Spermidine, 0.1% BSA, 1x cOmplete EDTA-free) (250 µl per antibody/reaction) to prepare a 2x antibody mix. The primary antibodies used were anti-H3K27me3 (rabbit, monoclonal, Cell signaling #9733), anti-IgG (rabbit, polyclonal, Abcam #ab2046540), and anti-CTCF (rabbit, polyclonal, Millipore #07-729-25U) at a final dilution of 1:100.

Samples were washed once with wash buffer using the magnet stand and 10x inversion. The samples were resuspended in 250 µl of wash buffer with gentle pipetting and while gently vortexing, 250 µl of the prepared 2x primary antibody in wash buffer dilution was added. The samples were incubated on a rotator for ≥2 hours at 4°C, followed by a quick spin and two washes with 500 µl of wash buffer.

To digest the antibody bound chromatin, 2x pA-MNase in wash buffer was prepared at a final 1x concentration of 700 ng/ml, which equals 350 ng of the fusion protein. To prepare this, 5.6 µl of the 0.5 mg/ml stock of pA-MNase were added to 2 ml of wash buffer. Samples were put on a magnet stand and the liquid was pulled. Samples were resuspended in 250 µl of wash buffer, then while gently vortexing, 250 µl of the 2x pA-MNase in wash buffer were added. Samples were incubated for ≥1 hour on a rotator at 4°C, followed by a quick spin and two washes with 1 ml of wash buffer. Before activating the pA-MNase, the nuclei sample were resuspended in 100 µl wash buffer and put onto a 0°C metal block on ice filled with cold water for 5 min. To activate the digestion, 2 µl of 100 mM CaCl2 per 150 µl were added to each sample while vortexing very carefully (final CaCl2 concentration 2 mM). Samples were put back on 0°C and incubated for 30 min while carefully vortexing every few minutes. The digestion was stopped by adding 100 µl 2x STOP+ buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.1% NP40) and careful vortexing.

To release the CUT&RUN fragments from the insoluble nuclear chromatin, samples were incubated for 10-30 min at 37°C. After centrifugation at 16,000x g for 5 min at 4°C and placing on a magnet stand, the supernatant was transferred to a new tube. To each sample, 2 µl of 10% SDS and 2.5 µl of proteinase K (20 mg/ml stock) were added. The mixture was inverted and incubated for 10 min at 70°C.

To purify and size-select the DNA fragments, I used AMPure XP beads. First, AMPure XP beads were added to the samples at a ratio of 0.5X (100 µl) and carefully mixed by pipetting. Samples were incubated at RT for 5 min and placed on a magnet stand. The supernatant was transferred to fresh tubes, and the beads were discarded. This first selection step eliminated fragments >1 kb. Then, 1.3X volume of beads (260 µl) were added to the supernatant, and the mixture was carefully mixed by pipetting. Samples were incubated at RT for 5 min and placed on a magnet stand. The supernatant was discarded, and the beads were kept. This second selection step selected fragments between 100 bp

and 1 kb. While remaining on the magnet stand, beads were washed with 400 µl of 80% freshly prepared ethanol and incubation at RT for 30 s. Ethanol was removed and beads were air-dried for up to 2 min while avoiding over-drying. DNA fragments were eluted by adding 50 µl of H2O to the beads, mixing and incubation at RT for 5 min. Samples were placed on the magnet stand and the eluate carefully transferred to a new tube without disturbing the beads. Final DNA was quantified using Qubit and quality and size distribution evaluated via BioAnalyzer.

The library preparation for Illumina sequencing was either prepared under native or bisulfite (bis) conditions. In both cases, the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB #E7645) was used according to the manufacturer's manual. As a starting amount, 6 to 20 ng of DNA were used. For bisulfite sequencing, the bisulfite conversion was performed after adapter ligation and clean-up without size selection using the EZ DNA Methylation Gold kit (ZYMO, #D5005) according to the manual. PCR amplification of the bisulfite-converted library was performed using KAPA HiFi HotStart Uracil+ Kit (#KR0413) at a total volume of 50 µl according to the manual. The native library was amplified using the NEBNext Ultra II Q5 Master Mix according to the manual. Libraries were cleaned-up using Ampure XP beads and including a size selection step to select for DNA fragments between 120 and 500 bp. Final elution was done in 17 µl elution buffer, of which 15 µl were transferred to a new tube, followed by Qubit measurement and BioAnalyzer quality check. The libraries were sequences at a NextSeq Mid with 75 bp paired-end.

On an additional note, during the CUT&RUN optimization phase, different concentrations of pA-MNase and antibodies were used. For the bisulfite conversion, I also tested the Imprint® DNA Modification Kit (Sigma, Cat#MOD50), which revealed comparable conversion ratios. Library enrichment efficiency was measured by qPCR before Illumina sequencing.

4.1.5 | CUT&RUN combined with SMF

For the CUT&RUN-SMF approach, the CUT&RUN and the SMF protocol were combined. First, cells were harvested (250,000 cells per reaction plus 1 reaction and 10% to account for pipetting errors and loss during washing steps = harvest 3 mio cells for 10 reactions). Cell were washed, nuclei extracted and bound to the Bio-Mag Plus Concanavalin A coated beads as described above in the CUT&RUN protocol (see Methods and Materials 4.1.4). After the bead binding, samples were put on a magnet stand and the liquid was pulled off. Bead-bound nuclei were washed once with 1 ml ice-cold wash buffer SMF (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl2, 0.1 mM EDTA) and then resuspended in 1.1 ml (for 10 reactions) ice-cold wash buffer SMF. 100 µl were distributed each into 1.5 ml low-binding tubes and placed on a magnet stand. The liquid was pulled off and nuclei were resuspended in 94.5 µl 1x M.GpC buffer. While gently vortexing, 150 µl of freshly prepared GpC MT Mix (1.3x M.GpC buffer, 0.6 mM sucrose, 1.28 mM SAM) and 50 µl of M.CviPI (4U/µl) were added to each sample and incubated at 37°C for 7.5 min. The reaction was replenished with 100 U M.CviPI and 128 pmol of SAM for a second incubation round at 37°C for 7.5 min. After the full GpC MT treatment, the samples were immediately put on an ice-cold magnet stand on ice, the liquid was removed and samples washes twice with 1 ml

ice-cold wash buffer CnR (20 mM HEPES [pH 7.5], 150 mM NaCl, 0.5 mM Spermidine, 0.1% BSA, 1x cOmplete EDTA-free) by inverting the tubes around 10x. Afterwards the samples were incubated in 1 ml blocking buffer (wash buffer CnR with 2 mM EDTA). Antibody dilutions were prepared with a 1:400 dilution in 150 µl wash buffer CnR per sample for a final dilution of 1:200. The primary antibodies used were anti-H3K27me3 (rabbit, monoclonal, Cell signaling #9733), anti-IgG (rabbit, polyclonal, Abcam #ab2046540), and anti-H3K27ac (rabbit, polyclonal, Abcam #ab4729). Samples were resuspended in 150 µl cold wash buffer CnR and while gently vortexing, 150 µl of the prepared 2x primary antibody in wash buffer dilution was added. The samples were incubated on a rotator over night at 4°C, followed by a quick spin and two washes with 500 µl of wash buffer.

The remaining protocol followed the conventional CUT&RUN protocol as described above (see Methods and Materials 4.1.4). The final CnR-SMF libraries were sequences at a NextSeq Mid with 75 bp paired-end.

4.2 | Computational analyses

4.2.1 | Pre-processing of sequencing data

Pre-processing of the SMF, WGBS and CTCF ChIP-bis data was done as previously described (Kleinendorst et al., 2021). For this, Illumina adaptors and low quality bases were trimmed and low quality reads were removed using Trimmomatic (0.36) (Bolger et al., 2014). Pre-processed reads were mapped to the bisulfite-indexed *Mus musculus* reference genome (BSgenome.Mmusculus.UCSC.mm10) using the R package QuasR (Gaidatzis et al., 2015), which uses the aligner Bowtie (Langmead et al., 2009), with specific alignment parameters (alignmentParameter = -e 70 -X 1000 -k 2 --best –strata) and keeping only uniquely aligned reads. Duplicated reads were removed using the tool MarkDuplicates from Picard (2.15.0) (Broad Institute, 2019).

Sequencing reads of hybrid F1 SMF data were trimmed for Illumina adaptors using Trim Galore! (v0.6.3) (Krueger et al., 2021) and competitively aligned to two versions of the bisulfite-indexed *Mus musculus* reference genome (BSgenome.Mmusculus.UCSC.mm10) injected with the SNPs(Keane et al., 2011) for the species of interest using QuasR (Gaidatzis et al., 2015). Finally, duplicated reads were removed using the tool MarkDuplicates from Picard (2.15.0) (Broad Institute, 2019).

Data from DNase-seq and PRO-seq were pre-processed with Trimmomatic (0.36) (Bolger et al., 2014) and aligned to the *Mus musculus* genome (BSgenome.Mmusculus.UCSC.mm10) using QuasR (Gaidatzis et al., 2015), keeping only uniquely aligned reads. For DNase-seq, duplicated reads were removed using the tool MarkDuplicates from Picard (2.15.0) (Broad Institute, 2019).

Data from ChIP-seq was pre-processed with the help of GBCS team at EMBL using the Galaxy platform (Afgan et al., 2018). Trim Galore! (Krueger et al., 2021) (Galaxy Tool version 0.4.3.1) was used to trim adaptor sequences and low quality bases. Bowtie(Langmead et al., 2009) (Galaxy Tool version 1.2.0) was used for alignment (for paired-end default parameters were used; for single-end parameters --

best --strata --maxbts 800 -m 1 were used) to the *Mus musculus* reference genome (USCS mm10), keeping only uniquely aligned reads (Galaxy Tool "Filter SAM or BAM, output SAM or BAM" 1.1.2 with filters on MAPQ>10 and bitwise filters to remove unmapped, not primary and supplementary alignments). Duplicated reads were removed using Picard (Broad Institute, 2019) MarkDuplicates (Galaxy Tool version Galaxy Version 2.7.1.1). Finally, normalized (reads per genome coverage) signal coverage bigwig files were obtained with deepTools (Ramírez et al., 2016) bamCoverage (Galaxy tool version 3.0.2.0) and a bin size of 50 bases.

RNA-seq data was processed using the nf-core/rnaseq pipeline (v3.4) (Patel et al., 2022) distributed through nf-core (Ewels et al., 2020). The pipeline was launched with default parameters and each sample was annotated for strandedness.

CUT&RUN and CnR-SMF data were pre-processed with Trimmomatic (0.36) (Bolger et al., 2014), trimming an additional base after the adapter, and aligned to the *Mus musculus* genome (BSgenome.Mmusculus.UCSC.mm10) using QuasR (Gaidatzis et al., 2015), keeping only uniquely aligned reads. Bisulfite-converted CUT&RUN and CnR-SMF data were aligned to the mapped to the bisulfite-indexed *Mus musculus* reference genome. Duplicated reads were removed using the tool MarkDuplicates from Picard (2.15.0) (Broad Institute, 2019).

4.2.2 | SMF - Single molecule methylation call

“Single molecule methylation state calling for all cytosines with at least 10-fold coverage was performed using QuasR (Gaidatzis et al., 2015) as previously described (Kleinendorst et al., 2021; Sönmezer et al., 2021). For this, the strand information are collapsed. Only CpGs in NWCGW (N = any base, W = no C or G) and GpCs in DGCHN (D = no C, H = no G, N = any base) context were considered to avoid interferences between CpG and GpC methylation calling due to strand collapsing or M.CviPI off-targets.” (Kreibich et al., 2023)

4.2.3 | SMF – Single molecule computation of CA and 5mC

“The analysis was focused on CpGs which have heterogeneous methylation across cells by selecting those with intermediate methylation (10-60%) (Figure 11A – step 1). The chromatin accessibility (CA) of each molecule covering those CpGs was calculated using the mean GpC methylation of all DGCHNs within a 101 bp bin surrounding it (Figure 11A – step 2). This mean accessibility value was then used to classify each molecule into an accessible (≥50% methylated) or inaccessible (<50% methylated) fraction (Figure 11A – step 3). Next, the methylation of the CpG in the center of the bin was determined (Figure 11A – step 4).

To test for the significance of the difference in DNA methylation (5mC) between the two fractions of molecules, a Cochran-Mantel-Haenszel test was performed including a pseudo-count of 1 and applying a continuity correction (Figure 11A – step 5). We only considered CpGs covered in at least 2 out of 2 or 3 biological replicates and had a 30-fold coverage. The resulting common odds ratio (COR) and p-value were used to group sites into positive (COR ≥ 2.0 and a p-value < 0.05), negative (COR ≤ 0.5 and

p-value < 0.05) or no 5mC-CA association (Figure 11A – step 6). Consequently, sites without a 5mC-CA association included sites with high methylation difference but no significant p-value due to variability between the replicates or low coverage. Therefore, when comparing negative 5mC-CA association and no 5mC-CA association sites in subsequent analyses, higher stringency sites with no association were defined by further selecting for a COR between 0.9 and 1.1." (Kreibich et al., 2023)

4.2.4 | SMF – CA, 5mC, fraction size and methylation

"For the mean chromatin accessibility (GpC methylation), first the mean methylation of all GpCs within the 101 bp bin surrounding a center CpG was calculated for each biological replicate and then of this the weighted mean of all replicates. For the mean DNA methylation (CpG methylation) the weighted mean of all replicates was calculated for the center CpG. Coverage was used as weighing value. A coverage cut-off of 10 reads per bin was applied to ensure reliable methylation calling." (Kreibich et al., 2023)

4.2.5 | SMF – Single locus plots

"For plotting SMF data for single loci, functions of the R package SingleMoleculeFootprinting (Kleinendorst et al., 2021) were adapted to plot CpGs and GpCs individually and to include the chromatin accessibility state sorting. " (Kreibich et al., 2023)

4.2.6 | Defining chromatin states with chromHMM

"To annotate the chromatin state of the tested CpG sites, previously defined chromHMM states for mouse embryonic stem cells for USCS mm10 (Pintacuda et al., 2017) were used that were in addition manually grouped into five chromHMM clusters:
C1 – CTCF (1_Insulator),
C2 – Inactive (2_Intergenic, 3_Heterochromatin, 5_RepressedChromatin, 6_BivalentChromatin, 12_LowSignal/RepetitiveElements),
C3 – Transcription (9_TranscriptionTransition, 10_TranscriptionElongation),
C4 – CREs promoter (7_ActivePromoter),
C5 – CREs enhancer (4_Enhancer, 8_StrongEnhancer, 11_WeakEnhancer)." (Kreibich et al., 2023)

4.2.7 | Genome browser tracks

"Bigwig files publicly available datasets of chromatin modifications (H3K27me3 (Liu and Kraus, 2017), H3K27ac (Shen et al., 2012), H3K4me1 (Stadler et al., 2011), H3K4me3 (Shen et al., 2012), DNase-seq (Domcke et al., 2015), WGBS (Stadler et al., 2011)), RNA-seq data (Domcke et al., 2015; Huang et al., 2021), and CUT&RUN data (Gretarsson and Hackett, 2020; Hainer et al., 2019), and generated PRO-seq data were loaded and the signal scores at the given genomic locus window plotted as area plots or point plots (for WGBS and SMF) using the R package ggplot2(Wickham, 2016) using different degrees of smoothing depending on the window size. Gene promoter and exons for the USCS mm10 reference

genome were identified using the R package TxDb.Mmusculus.UCSC.mm10.knownGene (Team BC and Maintainer BP, 2019).” (Kreibich et al., 2023)

4.2.8 | Allelic differences

“Imprinting Control Regions (ICRs) were defined by manually curating previous annotations (Xie et al., 2012). For this, only those ICRs were used that are maintained in the mESCs in cultured conditions as identified by F1 hybrid lines analysis. In addition, only those CpGs were kept as a reference that had an average chromatin accessibility of more than 35%, excluding CpGs located at nucleosomes. The average DNA methylation of all sites with negative 5mC-CA association was calculated for both alleles together or separately. C to T SNPs were masked in the analysis. Average DNA methylation of each category was plotted as violin box plots and an unpaired Wilcoxon rank test between the two separated alleles was applied using the stats_compare_means function of the ggpubr package with the maternal allele as reference group. On an additional note, the final manually curated ICRs are shown in Figure 16A.” (Kreibich et al., 2023)

4.2.9 | CpG density

“Number of CpGs was determined in a 400 bp window around all tested CpGs. The number of CpG per 100 bp was plotted as CpG density for each category as violin box plots. ” (Kreibich et al., 2023)

4.2.10 | hMeDIP-seq analysis

“Reads of hMeDIP-seq data (Xu et al., 2011) at a 500 bp window surrounding all NWCGWs were counted using the qCount function from the QuasR (Gaidatzis et al., 2015). Read counts were normalized by the total amounts of reads and a pseudo-count (pc) of 5 was added. The log2 of this counts per million (cpm+pc) was plotted for each category as violin box plots. ” (Kreibich et al., 2023)

4.2.11 | DHS peak analysis

“Published DHS peaks from ENCODE were used to calculate the distance of CpGs of each category to the summit of the closest DHS peak and plotted as a density plot. Only CpGs that are annotated as enhancers by chromHMM (C5 cluster) and that have a DHS peak within 500 bp were considered. Furthermore, the read density of DNase-seq data (Domcke et al., 2015) was calculated at a 1 kb window around the DHS peak summits with the function qProfile from the QuasR (Gaidatzis et al., 2015).” (Kreibich et al., 2023)

4.2.12 | Chromatin modifications

“Bigwig files of chromatin modifications (H3K27me3 (Liu and Kraus, 2017), H3K27ac (Shen et al., 2012), H3K4me1 (Stadler et al., 2011), H3K4me3 (Shen et al., 2012), DNase-seq (Domcke et al., 2015), CTCF (Stadler et al., 2011)) of publicly available datasets were loaded and the signal scores at 4 kb

surrounding tested CpGs plotted using the R package EnrichedHeatmap (Gu et al., 2018b) using 50 bp windows and smoothing. ” (Kreibich et al., 2023)

4.2.13 | MNase-seq analysis

“Read density of MNase-seq data (Barisic et al., 2019) was calculated at a 1 kb window around the CpG bin with negative or positive 5mC-CA association using the function qProfile from the QuasR (Gaidatzis et al., 2015) and plotted as smoothed counts per million (cpm). ” (Kreibich et al., 2023)

4.2.14 | SMF – Change of CA in perturbation assays

“Delta chromatin accessibility (CA defined as 1-SMF signal) between wild type (WT) and triple-knockout (TKO) ES cell lines was calculated by subtraction of the SMF signal (1-CA) in TKO from the SMF signal in WT for CpGs of each category in WT ES cells. To focus on active *cis*-regulatory elements, sites annotated as insulator (chromHMM cluster C1) or inactive (chromHMM cluster C2) were removed. In addition, to focus on sites where a potential change in CA could be correlated with changes in DNA methylation, CpGs were filtered for their DNA methylation change. For DNMT TKOs, only those CpGs were considered that had a starting DNA methylation of at least 30% in WT cells, resulting in a loss of DNA methylation of ≥ 30% in DNMT TKOs. For TET TKOs, only those CpGs were considered that showed an increase in DNA methylation of at least 30% from WT to TET TKOs. Delta SMF (1-CA) was plotted as violin box plots and an unpaired Wilcoxon rank test was applied using the stats_compare_means function of the ggpubr package using sites without 5mC-CA association as reference group.“ (Kreibich et al., 2023)

4.2.15 | PRO-seq analysis

“Read counts from generated PRO-seq data in wild-type (WT) and triple-knockout (TKO) cell lines was performed for each replicate at 500 bp collection windows surrounding all intermediately methylated bins, excluding bins annotated as insulator (chromHMM cluster C1) or inactive (chromHMM cluster C2). Using those count matrices as input, differential Polymerase II activity analysis was performed using DESeq2 (v1.30.1) (Love et al., 2014) after filtering out bins with less than 10 counts across all samples. Changes in PRO-seq signal (log2 fold-change) across replicates between WT and TKOs were plotted as violin box plots after filtering for DNA methylation changes of at least 40% (up in TET TKOs, down in DNMT TKOs). Single locus examples were plotted as IGV snapshots of the Bigwig files.” (Kreibich et al., 2023)

4.2.16 | RNA-seq analysis

“Differential gene expression analysis was performed using DESeq2 (v1.30.1) (Love et al., 2014) after filtering out genes with less than 10 counts across all samples. Genes associated with a log2 fold change greater, or smaller, than 2, or -2, and an adjusted p-value lower than 0.05 were considered differentially expressed for DNMT TKOs and TET TKOs, respectively. Enhancer-gene associations from the GREAT analysis were used to identify genes that are potentially regulated by sites with negative

5mC-CA association. For this I only considered loci that showed DNA methylation and chromatin accessibility changes in the TKO cell lines of at least 10% and 5%, respectively." (Kreibich et al., 2023)

4.2.17 | Enhancer-Gene associations and GO terms

"Genomic Regions Enrichment of Annotations Tool (GREAT, 4.0.4) (McLean et al., 2010) was used through the R package rGREAT (Gu, 2021) to perform enhancer-gene associations and Gene Ontology (GO) term analysis. For this, sites with negative 5mC-CA association in each cell line were used as an input. The rest of the genome was used as background region and USCS mm10 as reference genome. The model "Basal plus extension" was applied with 5 kb upstream and 1 kb downstream for proximal, and 1000 kb for distal. Curated regulatory domains were included. The GO Biological Process enrichment tables were extracted and filtered for a fold enrichment of the binominal test (Binom_Fold_Enrichment) of ≥ 2. The top 7 processes were selected for plotting, based on their Bonferroni corrected p-value of the binominal test (Binom_Adjp_BH). To focus on cell-type-specific sites, I only considered sites with a negative 5mC-CA association in a single cell type (ES cells were used as comparison for all somatic cell lines; NPs were used as comparison for ES cells)." (Kreibich et al., 2023)

4.2.18 | Defining cell-type specificity

"To define whether sites with negative 5mC-CA association are cell-type specific, a k-means clustering of the log2 common odds ratio from the Cochran-Mantel-Haenszel performed in each of the four cell lines (ES, NP, MEL, C2C12) was applied with k = 5. In order to include loci gaining DNA methylation from one cell type to the other, I used an average DNA methylation cut-off of 5-85% for the Cochran-Mantel-Haenszel test. A heatmap of the k-means clustered log2 common odds ratios was plotted using the function Heatmap of the R package ComplexHeatmap (Gu et al., 2016)." (Kreibich et al., 2023)

4.2.19 | Motif enrichment analysis

"Motif enrichment analysis of cell-type-specific enhancers with negative 5mC-CA association was performed using Hypergeometric Optimization of Motif EnRichment (HOMER, v4.11) (Heinz et al., 2010) with a window size of 101 bp around tested CpGs, USCS mm10 as reference genome, and using the repeat-masked sequence. Z-score of the -log p-value of known motifs was calculated for each group and top 20 motifs of each group were selected for plotting. Motifs were hierarchical clustered and the z-score of the -log p-value and the enrichment ("% of Target Sequences with Motif") were plotted. The same definition of cell-type-specific sites with negative 5mC-CA association was used as for the GREAT analysis." (Kreibich et al., 2023)

4.2.20 | SMF – Quantification of 5mC-TF association

"To identify transcription factor (TF) binding states, a 30 bp bin was created centered on a curated list of transcription factor binding sites (TFBS) from the JASPAR database (2018) (Khan et al., 2018) for *Mus Musculus* USCS mm10 reference genome, and two 10 bp bins 10 bp up- and downstream of the

center bin. To ensure to have only single instances of each genomic location of a TFBS, the TFBS were annotated with a published list of TFBS motif clusters (Glaser et al., 2021; Khan et al., 2018) and redundant TFBS were removed. The identity of the TFs creating the footprints were inferred by the presence of a binding motif and ChIP-seq signal for that TF at the locus (Kleinendorst et al., 2021; Sönmezer et al., 2021).

We only analysed TFBS that had a least one CpG within the 30 bp bin centered around the motif. The mean GpC methylation of each molecule was calculated in each of the three bins for every considered TFBS, which was then used to classify each molecule into a TF bound, nucleosome bound or accessible fraction and to define the TF binding frequency as previously described (Kleinendorst et al., 2021; Sönmezer et al., 2021). The analysis was focused on sites that are bound by the TF (>5% of the molecules) and have an average CpG methylation level of 10-80%. To test for a significant difference in the DNA methylation between the TF bound and nucleosome bound fraction, a Cochran-Mantel-Haenszel test was performed including a pseudo-count of 1 and applying a continuity correction. We only considered TFBS that were covered in at least 2 out of 3 biological replicates and had a 30-fold coverage. The common odds ratio (COR) and p-value of the Cochran-Mantel-Haenszel test were used to discriminate sites without association from those with a negative 5mC-TF association (COR $\leq$ 0.5 and p-value < 0.05) or positive (COR $\geq$ 2.0 and p-value < 0.05). For subsequent analyses, only "soft criteria" (only the COR) were used define sites with negative 5mC-TF association (COR < 0.5) from sites without an association (COR > 0.5). These are marked with a (*)star in text and figures." (Kreibich et al., 2023)

4.2.21 | SMF – state frequencies and TFBS methylation

"Binding state frequencies were calculated as means of all replicates. The CpG methylation of each TFBS and of the binding state fractions of each TFBS were calculated as weighted means of all replicates using the coverage as a weighing value." (Kreibich et al., 2023)

4.2.22 | Comparison CTCF ChIP-bis vs SMF

"For all CTCF sites covered by SMF, I separated the nucleosome bound fraction of the reads and the TF bound fraction of the reads. We collected the average DNA methylation in each of the fractions and calculated the difference between the fractions. For the same binding sites, I collected the average DNA methylation in the input fraction (WGBS data (Stadler et al., 2011)) and in the CTCF bound fraction (ChIP-bis (Feldmann et al., 2013)) and calculated the methylation difference between the fractions. We applied a minimal coverage of >9 reads to calculate the average methylation. The delta-delta plot shows the comparison of the methylation difference between the TF bound/nucleosome bound fractions as measured by SMF or ChIP-bis." (Kreibich et al., 2023)

4.2.23 | ChIP-seq analysis – CTCF and Max-Myc

"Myc ChIP-seq datasets (Chronis et al., 2017; Das et al., 2014; Mor et al., 2018) were merged. Reads of ChIP-seq data (CTCF ChIP data (Stadler et al., 2011), and merged Myc ChIP-seq data) at a 500 bp

window surrounding unique CTCF or Max-Myc binding sites were counted using the qCount function from QuasR (Gaidatzis et al., 2015). Read counts were normalized by the total amounts of reads and the log2 of the counts per million (cpm) was plotted for all binding sites and sites categorized by their 5mC-TF association as violin box plots. For the comparison between WT and TET TKO, ChIP signal Myc ChIP data (Yin et al., 2017) from WT and TET TKO ES cells quantified as described above, the delta was quantified and plotted against the delta in DNA methylation (measured by SMF)." (Kreibich et al., 2023)

4.2.24 | Motif analysis CTCF and Max-Myc

"For each unique TFBS covered by SMF with no or a negative 5mC-TF association, I extracted the DNA sequences of the motifs. We then calculated the frequency of having a CpG at each of the motif positions in each category. We compared the CpG frequency for all motifs for which there are at least 10 instances in both categories. To detect enrichments of CpGs at specific motif positions, I plotted the frequency of motifs having a motif at each position." (Kreibich et al., 2023)

4.2.25 | SMF – Change of TF binding upon perturbations

"The difference in binding state fraction size (log2 fold change) between wild type (WT) and TET triple-knockout (TET TKO) ES 5mC-TF association in WT ES cells. To focus on sites where a potential change in TF binding could be correlated with a change in DNA methylation, only those TFBS were considered that showed an increase in DNA methylation of at least 30% from WT to TET TKOs. In Figure 24A, the log2 fold change of all binding state fractions is plotted as violin box plots and an unpaired Wilcoxon rank test was applied using the stats_compare_means function of the ggpubr package with sites without 5mC-TF association as reference group. In Figures R15B-C, the log2 fold change of the TF bound fraction of the sites with negative 5mC-TF association is plotted as a function of the change in DNA methylation of the TFBS from WT to TET TKOs." (Kreibich et al., 2023)

4.2.26 | CUT&RUN – Correlation plots

Using the CUT&RUN data bam files, read count was performed at sites of interest (for CTCF: 500 bp window centered at bound CTCF sites; for histone modifications: 10 kb window at transcriptional start sites [TSS]) using the qCount function from QuasR (Gaidatzis et al., 2015). Log2 of read counts (plus pseudo count of 1) was plotted as a smoothed density scatter plot comparing individual samples and the Pearson correlation was calculated.

4.2.27 | CUT&RUN – Heatmaps

Bigwig files of publicly available ChIP data (CTCF ((Stadler et al., 2011), H3K27me3 (Liu and Kraus, 2017), H3K27ac (Shen et al., 2012)) and CUT&RUN data (CTCF (Hainer et al., 2019), H3K27me3 (Gretarsson and Hackett, 2020; Hainer et al., 2019), H3K27ac (Hainer et al., 2019)) and self-produced CUT&RUN and CnR-SMF data were loaded.

For CTCF, the signal scores in a window of 1 kb surrounding the top 20,000 CTCF bound sites (according to ChIP data) were plotted. For H3K27me3, the signal scores in a window of 16 kb surrounding the top 10,000 H3K27me3 peaks at TSS (according to ChIP data) were plotted. For H3K27ac, the signal scores in a window of 10 kb surrounding the top 10,000 H3K27ac peaks at TSS (according to ChIP data) were plotted. Heatmap plotting was done using the R package EnrichedHeatmap AAA with 50 bp windows and smoothing.

4.2.28 | CUT&RUN – Signal to noise ratio

To calculate the signal to noise ratio, signal was counted at the top sites (according to ChIP-seq data) and at random intergenic regions annotated by chromHMM, averaged and divided. For CTCF, the top 5,000 CTCF bound sites were used with a window of 500 bp. For H3K27me3 and H3K27ac, the top 5,000 TSS sites with the strongest peaks were used with a window of 5 kb.

4.2.29 | Statistical analysis and reproducibility

"For SMF, replicate information of at least two (three for WT ES and DNMT TKO ES) replicates were used for each cell line.
All statistical analyses were performed using R software and are described in the figure legends and methods section.
Violin box plots were plotted by the R package ggplot2 with arguments outlier.alpha = 0 and coef = 0. The upper and lower boundaries of the box plot represent the 25th and the 75th percentiles. The central line represents the median. P-values for in violin box plots were calculated by unpaired Wilcoxon rank test using the R package ggpubr and stars used to depict significance (ns: $p > 0.05$, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$).
Pearson correlation coefficients were calculated using the R package corrr (Kuhn et al., 2022) with the option pairwise.complete.obs, and a linear regression line was plotted with the R package ggplot2 using method = "lm".
The Cochran-Mantel-Haenszel test was performed under the criteria that the region of interest was covered by at least two replicates with a mean coverage of 30 reads. A pseudo-count of 1 was added to all possible states and the test was performed with a continuity correction.
SMF single locus plots were plotted through ggplot2 and using the functions of the SingleMoleculeFootprinting (Kleinendorst et al., 2021) R package as basis adapted to the needs of this study. " (Kreibich et al., 2023)

4.3 | Materials

4.3.1 | Reagents and kits

Table 2: Reagents and commercial kits used in different experiments

Experiment	Reagent/ Commercial Kits	Source	Identifier
Cell culture	Dulbecco's Modified Eagle Medium (DMEM), high glucose	Gibco	Cat#41965039
	FBS Embryomax	Millipore	Cat#ES-009-B
	L-glutamine	Gibco	Cat#A2916801
	Leukemia inhibitory factor (LIF)	prepared in house	N/A
	MEM Non-Essential Amino Acids Solution (100×)	Gibco	Cat#11140050
	2-Mercaptoethanol	Merck	Cat#M6250
	Sodium pyruvate	Gibco	Cat#11360070
	Trypsin-EDTA (0.25%)	Gibco	Cat#25200056
	Gelatin	Sigma	Cat#G-1890
SMF	GpC Methyltransferase (M.CviPI)	NEB	Cat#M0227
	KAPA HiFi HotStart Uracil+ Kit	Roche	Cat#KK2802
	Chloroform	Sigma	Cat# 366919
	IGEPAL CA-630	Sigma	Cat# I8896
	Magnesium chloride ($MgCl_2$)	Sigma	Cat# M8266
	Phenol equilibrated, stabilized:chloroform:isoamyl alcohol 25: 24: 1	PanReac AppliChem	Cat# A0889
	Proteinase K	Sigma	Cat#124568
	RNase A, DNase- and protease-free	Sigma	Cat# R6513
	S-adenosylmethionine (SAM)	NEB	Cat#86867-01-8
	Sodium chloride (NaCl)	Sigma	Cat#S7653
	Sodium dodecyl sulfate solution (10%)	Sigma	Cat#71736
	Sucrose Ultrapure MB grade	Sigma	Cat#1.08421
	Titriplex III (EDTA disodium salt dihydrate)	Sigma	Cat#T1503
	Trizma base	Sigma	Cat#T1503
	2-propanol	Sigma	Cat# I9516
	EZ DNA Methylation-Gold Kit	Zymo Research	Cat#D5005
	SureSelectXT Mouse Methyl-Seq Capture Library	Agilent	Cat#931052
	SureSelectXT Methyl-Seq Reagent Kit	Agilent	Cat#G9651A
	NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB	Cat#E7645L
	NEBNext Multiplex Oligos for Illumina	NEB	Cat#E7335L
PRO-seq	Biotin-11-UTP	PerkinElmer	Cat#NEL544001EA
	Biotin-11-CTP	PerkinElmer	Cat#NEL542001EA
	GTP	Roche	Cat#11277057001
	ATP	Roche	Cat#11277057001
	Streptavidin M280 beads	Invitrogen	Cat#112.06D
	GlycoBlue	Invitrogen	Cat#AM9515
	SUPERase•In™ RNase Inhibitor	Invitrogen	Cat#AM2694

	T4 RNA ligase I	NEB	Cat#M0204
	RNA 5’ pyrophosphohydrolase (RppH)	NEB	Cat#M0356S
	ThermoPol Reaction Buffer	NEB	Cat#B9004S
	T4 polynucleotide kinase (PNK)	NEB	Cat#M0201
	Superscript IV reverse transcriptase	Invitrogen	Cat#18090010
	Phusion polymerase	NEB	Cat#M0530
	Quick-RNA Miniprep Kit	Zymo Research	Cat#R1054
CUT&RUN	Calcium chloride (CaCl2)	Merck	Cat#C1016
	cOmplete™, Mini, EDTA-free Protease-Inhibitor-Cocktail	Merck	Cat#11836170001
	Concanavalin A coated magnetic beads	BioMag Plus	Cat#86057
	EDTA	Carl Roth	Cat#K714
	EGTA	Sigma	Cat#E3889-25g
	Mangan chloride (MnCl2)	Merck	Cat#244580
	Magnesium chloride (MgCl2)	Sigma	Cat# M8266
	Nonidet P 40 substitute (NP-40)	Sigma	Cat#74385-1l
	RNase A, DNase- and protease-free	Sigma	Cat# R6513
	Spermidine		Cat#
	Sodium chloride (NaCl)	Sigma	Cat#S7653
	Sodium dodecyl sulfate solution (10%)	Sigma	Cat#71736
	TritonX-100	Merck	Cat#T8787
	pA-MNase fusion (0.5 mg/ml)	Produced in house	
	EZ DNA Methylation-Gold Kit	Zymo Research	Cat#D5005
	KAPA HiFi HotStart Uracil+ Kit	Roche	Cat#KK2802
	NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB	Cat#E7645L
	NEBNext Multiplex Oligos for Illumina	NEB	Cat#E7335L

4.3.2 | Equipment

- Bioanalyzer 2100 instrument (Agilent, Cat#G2939BA)
- Bioanalyzer DNA 1000 Kit (Agilent, Cat#5067-1504)
- Bioanalyzer High Sensitivity DNA Kit (Agilent, Cat#5067-4627)
- Centrifuge, refrigerated, with fixed-angle rotor (Eppendorf, Model#5427R)
- Centrifuge with fixed-angle rotor (Eppendorf, Model #5425)
- Centrifuge with swinging bucket (Eppendorf, Model #5810R)
- Heater block with wells for 1.5 ml tubes (Thermo)
- Magnetic rack for PCR tubes
- Magnetic rack for 1.5 ml tubes; Dynamag (Thermo, Cat#12321D)
- Microcentrifuge (Roth)
- MicroTUBE holder (Covaris, Cat#500114)
- 1.5 ml microcentrifuge tubes (Eppendorf, Cat#22-282)
- 1.5 ml microcentrifuge safe-lock tubes (Eppendorf, Cat#30120086)
- 1.5 ml microcentrifuge DNA LoBind tubes (Eppendorf, Cat#30108051)
- 0.2 ml PCR tubes (Eppendorf, Cat#30124359)
- Snap-Cap microTUBEs (Covaris, Cat#520045)
- S-series focused ultrasonicator (Covaris, S2 model)
- Thermal cycler (Biorad, C1000 touch, Cat#1851148/1851196)
- Vacuum concentrator (Eppendorf)
- Vortex mixer (Vortex Genie; VWR)
- Water baths (VWR)

4.3.3 | Cell lines

Table 3: Cell lines used

Cell line	Reference	Identifier
Mouse ES159 embryonic stem cells (129, male) - WT	Stadler et al., 2011	https://web.expasy.org/cellosaurus/CVCL_IT51
Mouse ES159 embryonic stem cells (129, male) - DNMT TKO	Domcke et al., 2015	N/A
Mouse ES159 embryonic stem cells (129, male) - TET TKO	Ginno et al., 2020	N/A
Mouse ES F1 hybrid cells (129/CAST)	Giorgetti et al., 2016	N/A
Mouse myoblast cells (C2C12)	Moritz Mall	N/A
Mouse erythroleukemia cells (MEL)	DSMZ	ACC501

4.3.4 | Antibodies

Table 4: Antibodies used for CUT&RUN

Target	Characteristics	Provider	Catalogue number
CTCF	rabbit, polyclonal	Millipore	Cat#07-729-25U
H3K27ac	rabbit, polyclonal	Abcam	Cat#ab4729
H3K27me3	rabbit, monoclonal	Cell signaling	Cat#9733
Mouse IgG	rabbit, polyclonal	Abcam	Cat#ab2046540

4.3.5 | Oligos

Table 5: Oligos used in PRO-seq

Name	Sequence
Oligos for PRO-seq - VRA3	GAUCGUCGGACUGUAGAACUCUGAAC- /inverted dT/
Oligos for PRO-seq - VRA5	CCUUGGCACCCGAGAAUUCCA
Oligos for PRO-seq - RP1	AATGATACGGCGACCACCGAGATCTACACGT TCAGAGTTCTACAGTCCGA
Oligos for PRO-seq - RPI-n	CAAGCAGAAGACGGCATACGAGAT NNNNNN GTGACTGGAGTT CCTTGGCACCCGAGAATTCCA

4.3.6 | Data sets

Table 6: Publicly available data sets used

Data set	Reference	Accession number
ChIP Myc mESC WT	Chronis et al., 2017; Das et al., 2014; Mor et al., 2018	GSM1059010; GSM2417145; GSM2739921
ChIP CTCF mESC WT	Stadler et al., 2011	GSM747534; GSM747535; GSM747536
ChIP H3K4me1 mESC WT	Stadler et al., 2011	GSM747542
ChIP H3K4me3 mESC WT	Shen et al., 2012	GSM723017
ChIP H3K27Ac mESC WT	Shen et al., 2012	GSM851278
ChIP H3K27me3 mESC WT	Liu and Kraus, 2017	GSM1910634
ChIP-bis CTCF mESC WT	Feldmann et al., 2013	GSM978373; GSM1230236
ChIP Myc mESC WT	Yin et al., 2017	ERX1965640; ERX1965641
ChIP Myc mESC TET TKO	Yin et al., 2017	ERX1965638; ERX1965639
CUT&RUN CTCF mESC	Hainer et al., 2019	GSM3416698; GSM3022414
CUT&RUN H3K27ac mESC	Hainer et al., 2019	GSM3416722; GSM3022440
CUT&RUN H3K27me3 mESC	Hainer et al., 2019; Gretarsson and Hackett, 2020	GSM3022438; GSM3416720; GSM4407939
DNase-seq mESC WT	Domcke et al., 2015	GSM1657364
DHS peaks mESC WT	ENCODE	GSM1014154
hMeDIP-seq mESC WT	Xu et al., 2011	GSM706674

MNase-Seq mESC WT	Barisic et al., 2019	GSM3058339; GSM3058340
RNA-seq mESC WT	Domcke et al., 2015	GSM1657368; GSM1657369; GSM1657370
RNA-seq mESC DNMT TKO	Domcke et al., 2015	GSM1657371; GSM1657372; GSM1657373
RNA-seq mESC WT	Huang et al., 2021	GSM3602183; GSM3602184; GSM3602185; GSM3602186; GSM3602187; GSM3602188
RNA-seq mESC TET TKO	Huang et al., 2021	GSM3602201; GSM3602202; GSM3602203
SMF mESC WT	Sönmezer et al., 2021	E-MTAB-9123
SMF mESC DNMT TKO	Sönmezer et al., 2021	E-MTAB-9123
SMF M.CviPI treated lambda DNA	Kleinendorst et al., 2021	E-MTAB-10815
WGBS mESC WT	Stadler et al., 2011	GSM748786
ChromHMM states for mESC WT	Pintacuda et al., 2017	https://github.com/guifengwei/ChromHMM_mESC_mm10
JASPAR database (2018)	Khan et al., 2018	https://jaspar.genereg.net
Mouse Imprinting Control Regions (ICRs) - manually curated	Xie et al., 2012	N/A
TFBS motif clusters	Glaser et al., 2021; Khan et al., 2018	N/A

Table 7: Data sets produced in this study deposited at open data repositories

Data set	Repository	Accession number
SMF mESC WT NO R5 and R6	ArrayExpress	E-MTAB-11700
SMF mESC DNMT TKO DE R5 and R6	ArrayExpress	E-MTAB-11700
SMF mESC TETTKO NO R1 and R2	ArrayExpress	E-MTAB-11700
SMF C2C12 NO R1 and R2	ArrayExpress	E-MTAB-11700
SMF MEL NO R1 and R2	ArrayExpress	E-MTAB-11700
SMF NP NO R1 and R2	ArrayExpress	E-MTAB-11700
SMF F1 (129/CAST) NO R1 and R2	ArrayExpress	E-MTAB-11700
PRO-seq mESC WT R1, R2 and R3	ArrayExpress	E-MTAB-12601
PRO-seq mESC DNMT TKO R1, R2 and R3	ArrayExpress	E-MTAB-12601
PRO-seq mESC TET TKO R1, R2 and R3	ArrayExpress	E-MTAB-12601

4.3.7 | Software and Packages

Table 8: Software and packages used for data analysis

Name	Reference	Deposition
Original R scripts for the analyses of this study	This study	10.5281/zenodo.7534802
R scripts to call methylation of single molecules	Soenmezer et al., 2021	https://github.com/Krebslabrep/Soenmezer_2020_SMF/
R-4.1.0, or higher		https://www.r-project.org
ComplexHeatmap (R package)	Gu et al., 2016	10.1093/bioinformatics/btw313
corrr (R package)	Kuhn et al., 2022	https://corrr.tidymodels.org/index.html
DESeq2 (v1.30.1) (R package)	Love et al., 2014	10.18129/B9.bioc.DESeq2
EnrichedHeatmap (R package)	Gu et al., 2018	10.18129/B9.bioc.EnrichedHeatmap
QuasR (R package)	Gaidatzis et al., 2015	https://www.bioconductor.org/packages/release/bioc/html/QuasR.html
rGREAT (R package)	Gu, 2021	10.18129/B9.bioc.rGREAT
SingleMoleculeFootprinting (R package)	Kleinendorst et al., 2021	10.18129/B9.bioc.SingleMoleculeFootprinting
Tidyverse (R package)	Wickam et al., 2019	https://tidyverse.tidyverse.org/index.html
Bowtie	Langmead et al., 2009	https://bowtie-bio.sourceforge.net/index.shtml
deepTools	Ramírez et al., 2016	https://deeptools.readthedocs.io/en/develop/
Galaxy platform	Afgan et al., 2018	https://usegalaxy.org
Genomic Regions Enrichment of Annotations Tool (GREAT, 4.0.4)	McLean et al., 2010	http://great.stanford.edu/public/html/
Hypergeometric Optimization of Motif EnRichment (HOMER, v4.11)	Heinz et al., 2010	http://homer.ucsd.edu/homer/
nf-core/rnaseq pipeline (v3.4)	Patel et al., 2022	https://nf-co.re/rnaseq
nf-core/chipseq pipeline (v2.0)	Ewels et al., 2020	https://nf-co.re/chipseq
Picard (2.15.0)	Broad Institute, 2019	https://broadinstitute.github.io/picard/
Trim Galore! (v0.6.3)	Krueger et al., 2021	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Trimmomatic (0.36)	Bolger et al., 2014	https://github.com/usadellab/Trimmomatic
TxDb.Mmusculus.UCSC.mm10.knownGene	Team BC and Maintainer BP, 2019	10.18129/B9.bioc.TxDb.Mmusculus.UCSC.mm10.knownGene

Bibliography

Abascal, F., Acosta, R., Addleman, N.J., Adrian, J., Afzal, V., Ai, R., Aken, B., Akiyama, J.A., Jammal, O.A., Amrhein, H., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710. https://doi.org/10.1038/s41586-020-2493-4.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. *46*, W537–W544. https://doi.org/10.1093/nar/gky379.

Albert, J.R., Au Yeung, W.K., Toriyama, K., Kobayashi, H., Hirasawa, R., Brind'Amour, J., Bogutz, A., Sasaki, H., and Lorincz, M. (2020). Maternal DNMT3A-dependent de novo methylation of the paternal genome inhibits gene expression in the early embryo. Nat. Commun. *11*, 5417. https://doi.org/10.1038/s41467-020-19279-7.

Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M., and Gerstein, M.B. (2010). Annotating non-coding regions of the genome. Nat. Rev. Genet. *11*, 559–571. https://doi.org/10.1038/nrg2814.

Ali, T., Renkawitz, R., and Bartkuhn, M. (2016). Insulators and domains of gene expression. Genome Archit. Expr. *37*, 17–26. https://doi.org/10.1016/j.gde.2015.11.009.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol. *19*, 229–244. https://doi.org/10.1038/nrm.2017.119.

Altemose, N., Maslan, A., Smith, O.K., Sundararajan, K., Brown, R.R., Mishra, R., Detweiler, A.M., Neff, N., Miga, K.H., Straight, A.F., et al. (2022). DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. Nat. Methods *19*, 711–723. https://doi.org/10.1038/s41592-022-01475-6.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat. Cell Biol. *18*, 225–233. https://doi.org/10.1038/ncb3296.

Amoutzias, G.D., Robertson, D.L., Van de Peer, Y., and Oliver, S.G. (2008). Choose your partners: dimerization in eukaryotic transcription factors. Trends Biochem. Sci. *33*, 220–229. https://doi.org/10.1016/j.tibs.2008.02.002.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461. https://doi.org/10.1038/nature12787.

Aronson, B.E., Scourzic, L., Shah, V., Swanzey, E., Kloetgen, A., Polyzos, A., Sinha, A., Azziz, A., Caspi, I., Li, J., et al. (2021). A bipartite element with allele-specific functions safeguards DNA methylation imprints at the Dlk1-Dio3 locus. Dev. Cell *56*, 3052-3065.e5. https://doi.org/10.1016/j.devcel.2021.10.004.

Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol. *15*, 545. https://doi.org/10.1186/s13059-014-0545-5.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538. https://doi.org/10.1038/ncb1403.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308. https://doi.org/10.1016/0092-8674(81)90413-x.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909–912. https://doi.org/10.1126/science.aah5143.

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. Nature *569*, 136–140. https://doi.org/10.1038/s41586-019-1115-5.

Barnett, K.R., Decato, B.E., Scott, T.J., Hansen, T.J., Chen, B., Attalla, J., Smith, A.D., and Hodges, E. (2020). ATAC-Me Captures Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions. Mol. Cell *77*, 1350-1364.e6. https://doi.org/10.1016/j.molcel.2020.01.004.

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014). Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers. Mol. Cell *54*, 844–857. https://doi.org/10.1016/j.molcel.2014.04.006.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823–837. https://doi.org/10.1016/j.cell.2007.05.009.

Baubec, T., and Schübeler, D. (2014). Genomic patterns and context specific interpretation of DNA methylation. Curr. Opin. Genet. Dev. *25*, 85–92. https://doi.org/10.1016/j.gde.2013.11.015.

Baubec, T., Ivánek, R., Lienert, F., and 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.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243–247. https://doi.org/10.1038/nature14176.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome — biological and translational implications. Nat. Rev. Cancer *11*, 726–734. https://doi.org/10.1038/nrc3130.

Bednarik, D.P., Duckett, C., Kim, S.U., Perez, V.L., Griffis, K., Guenthner, P.C., and Folks, T.M. (1991). DNA CpG methylation inhibits binding of NF-kappa B proteins to the HIV-1 long terminal repeat cognate DNA motifs. New Biol. *3*, 969–976. .

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485. https://doi.org/10.1038/35013100.

Bell, O., Tiwari, V.K., Thomä, N.H., and Schübeler, D. (2011). Determinants and dynamics of genome accessibility. Nat. Rev. Genet. *12*, 554–564. https://doi.org/10.1038/nrg3017.

Benayoun, B.A., Pollina, E.A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E.D., Devarajan, K., Daugherty, A.C., Kundaje, A.B., Mancini, E., et al. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell *158*, 673–688. https://doi.org/10.1016/j.cell.2014.06.027.

Bender, C.M., Gonzalgo, M.L., Gonzales, F.A., Nguyen, C.T., Robertson, K.D., and Jones, P.A. (1999). Roles of cell division and gene transcription in the methylation of CpG islands. Mol. Cell. Biol. *19*, 6690–6698. https://doi.org/10.1128/MCB.19.10.6690.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326. https://doi.org/10.1016/j.cell.2006.02.041.

Berrens, R.V., Andrews, S., Spensberger, D., Santos, F., Dean, W., Gould, P., Sharif, J., Olova, N., Chandra, T., Koseki, H., et al. (2017). An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. Cell Stem Cell *21*, 694-703.e7. https://doi.org/10.1016/j.stem.2017.10.004.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.-A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat. Protoc. *2*, 1034–1043. https://doi.org/10.1038/nprot.2007.147.

Bird, A.P., and Wolffe, A.P. (1999). Methylation-induced repression--belts, braces, and chromatin. Cell *99*, 451–454. https://doi.org/10.1016/s0092-8674(00)81532-9.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99. https://doi.org/10.1016/0092-8674(85)90312-5.

Blake, J.A., Baldarelli, R., Kadin, J.A., Richardson, J.E., Smith, C.L., and Bult, C.J. (2021). Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. Nucleic Acids Res. *49*, D981–D987. https://doi.org/10.1093/nar/gkaa1083.

Blanco, E., González-Ramírez, M., Alcaine-Colet, A., Aranda, S., and Di Croce, L. (2020). The Bivalent Genome: Characterization, Structure, and Regulation. Trends Genet. *36*, 118–131. https://doi.org/10.1016/j.tig.2019.11.004.

Blattler, A., Yao, L., Wang, Y., Ye, Z., Jin, V.X., and Farnham, P.J. (2013). ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes. Epigenetics Chromatin *6*, 13. https://doi.org/10.1186/1756-8935-6-13.

Boileau, R.M., Chen, K.X., and Blelloch, R. (2023). Loss of MLL3/4 decouples enhancer H3K4 monomethylation, H3K27 acetylation, and gene activation during embryonic stem cell differentiation. Genome Biol. *24*, 41. https://doi.org/10.1186/s13059-023-02883-3.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma. Oxf. Engl. *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. *42*, 1093–1100. https://doi.org/10.1038/ng.708.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and 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.

Boulard, M., Edwards, J.R., and Bestor, T.H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. Nat. Genet. *47*, 479–485. https://doi.org/10.1038/ng.3272.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96–99. https://doi.org/10.1038/nature02886.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536–2539. https://doi.org/10.1126/science.1065848.

Broad Institute (2019). Picard Toolkit. Broad Institute, GitHub repository.

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

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490. https://doi.org/10.1038/nature14590.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837. https://doi.org/10.1016/j.molcel.2013.01.038.

Campanero, M.R., Armstrong, M.I., and Flemington, E.K. (2000). CpG methylation as a mechanism for the regulation of E2F activity. Proc. Natl. Acad. Sci. *97*, 6481–6486. https://doi.org/10.1073/pnas.100340697.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science *298*, 1039–1043. https://doi.org/10.1126/science.1076997.

Carlini, V., Policarpi, C., and Hackett, J.A. (2022). Epigenetic inheritance is gated by naïve pluripotency and Dppa2. EMBO J. *41*, e108677. https://doi.org/10.15252/embj.2021108677.

Carter, B., and Zhao, K. (2021). The epigenetic basis of cellular heterogeneity. Nat. Rev. Genet. *22*, 235–250. https://doi.org/10.1038/s41576-020-00300-0.

Cavalli, G., and Heard, E. (2019). Advances in epigenetics link genetics to the environment and disease. Nature *571*, 489–499. https://doi.org/10.1038/s41586-019-1411-0.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. *10*, 295–304. https://doi.org/10.1038/nrg2540.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234. https://doi.org/10.1038/nature06403.

Charlton, J., Jung, E.J., Mattei, A.L., Bailly, N., Liao, J., Martin, E.J., Giesselmann, P., Brändl, B., Stamenova, E.K., Müller, F.-J., et al. (2020). TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat. Genet. *52*, 819–827. https://doi.org/10.1038/s41588-020-0639-9.

Chen, J., Zhang, Z., Li, L., Chen, B.-C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., et al. (2014). Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. Cell *156*, 1274–1285. https://doi.org/10.1016/j.cell.2014.01.062.

Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., Su, J., Rodriguez, B., Xi, Y., Xia, Z., Chen, X., et al. (2015). Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. Nat. Genet. *47*, 1149–1157. https://doi.org/10.1038/ng.3385.

Chen, L.-L., Lin, H.-P., Zhou, W.-J., He, C.-X., Zhang, Z.-Y., Cheng, Z.-L., Song, J.-B., Liu, P., Chen, X.-Y., Xia, Y.-K., et al. (2018). SNIP1 Recruits TET2 to Regulate c-MYC Target Genes and Cellular DNA Damage Response. Cell Rep. *25*, 1485-1500.e4. https://doi.org/10.1016/j.celrep.2018.10.028.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol. Cell. Biol. *23*, 5594–5605. https://doi.org/10.1128/MCB.23.16.5594-5605.2003.

Cheng, S., Mittnenzweig, M., Mayshar, Y., Lifshitz, A., Dunjić, M., Rais, Y., Ben-Yair, R., Gehrs, S., Chomsky, E., Mukamel, Z., et al. (2022). The intrinsic and extrinsic effects of TET proteins during gastrulation. Cell *185*, 3169-3185.e20. https://doi.org/10.1016/j.cell.2022.06.049.

Cheng, Y., He, C., Wang, M., Ma, X., Mo, F., Yang, S., Han, J., and Wei, X. (2019). Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduct. Target. Ther. *4*, 62. https://doi.org/10.1038/s41392-019-0095-0.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e20. https://doi.org/10.1016/j.cell.2016.12.016.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. *18*, 407–422. https://doi.org/10.1038/nrm.2017.26.

Clark, S.J., Harrison, J., and Molloy, P.L. (1997). Sp1 binding is inhibited by (m)Cp(m)CpG methylation. Gene *195*, 67–71. https://doi.org/10.1016/s0378-1119(97)00164-9.

Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat. Commun. *9*, 781. https://doi.org/10.1038/s41467-018-03149-4.

Cleaton, M.A.M., Edwards, C.A., and Ferguson-Smith, A.C. (2014). Phenotypic Outcomes of Imprinted Gene Models in Mice: Elucidation of Pre- and Postnatal Functions of Imprinted Genes. Annu. Rev. Genomics Hum. Genet. *15*, 93–126. https://doi.org/10.1146/annurev-genom-091212-153441.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215–219. https://doi.org/10.1038/nature06745.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374. https://doi.org/10.1038/nature11925.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54. https://doi.org/10.1038/s41586-019-1517-4.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936. https://doi.org/10.1073/pnas.1016071107.

Crispatzu, G., Rehimi, R., Pachano, T., Bleckwehl, T., Cruz-Molina, S., Xiao, C., Mahabir, E., Bazzi, H., and Rada-Iglesias, A. (2021). The chromatin, topological and regulatory properties of pluripotency-associated poised enhancers are conserved in vivo. Nat. Commun. *12*, 4344. https://doi.org/10.1038/s41467-021-24641-4.

Cusack, M., King, H.W., Spingardi, P., Kessler, B.M., Klose, R.J., and Kriaucionis, S. (2020). Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors. Genome Res. *30*, 1393–1406. https://doi.org/10.1101/gr.257576.119.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914. https://doi.org/10.1126/science.aab1601.

Dahlet, T., Argüeso Lleida, A., Al Adhami, H., Dumas, M., Bender, A., Ngondo, R.P., Tanguy, M., Vallet, J., Auclair, G., Bardet, A.F., et al. (2020). Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. Nat. Commun. *11*, 3153. https://doi.org/10.1038/s41467-020-16919-w.

Daniel, J.M., Spring, C.M., Crawford, H.C., Reynolds, A.B., and Baig, A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. Nucleic Acids Res. *30*, 2911–2919. https://doi.org/10.1093/nar/gkf398.

Das, P.P., Shao, Z., Beyaz, S., Apostolou, E., Pinello, L., De Los Angeles, A., O'Brien, K., Atsma, J.M., Fujiwara, Y., Nguyen, M., et al. (2014). Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity. Mol. Cell *53*, 32–48. https://doi.org/10.1016/j.molcel.2013.11.011.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.-C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.-W., Page, D.C., et al. (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. Cell Stem Cell *9*, 166–175. https://doi.org/10.1016/j.stem.2011.07.010.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev. Cell *24*, 310–323. https://doi.org/10.1016/j.devcel.2012.12.015.

Day, K., Waite, L.L., Thalacker-Mercer, A., West, A., Bamman, M.M., Brooks, J.D., Myers, R.M., and Absher, D. (2013). Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biol. *14*, R102. https://doi.org/10.1186/gb-2013-14-9-r102.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev. *15*, 2940–2944. https://doi.org/10.1101/gad.929101.

Detilleux, D., Spill, Y.G., Balaramane, D., Weber, M., and Bardet, A.F. (2022). Pan-cancer predictions of transcription factors mediating aberrant DNA methylation. Epigenetics Chromatin *15*, 10. https://doi.org/10.1186/s13072-022-00443-w.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation *. J. Biol. Chem. *285*, 26114–26120. https://doi.org/10.1074/jbc.M109.089433.

Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature *528*, 575–579. https://doi.org/10.1038/nature16462.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D. 2nd, Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol. Cell *66*, 568-576.e4. https://doi.org/10.1016/j.molcel.2017.04.018.

Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015a). DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. *16*, 519–532. https://doi.org/10.1038/nrm4043.

Du, Q., Luu, P.-L., Stirzaker, C., and Clark, S.J. (2015b). 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., and Eeckhoute, J. (2014). A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. Nucleic Acids Res. *42*, 10943–10959. https://doi.org/10.1093/nar/gku780.

Duttke, S.H., Chang, M.W., Heinz, S., and Benner, C. (2019). Identification and dynamic quantification of regulatory elements using total RNA. Genome Res. *29*, 1836–1846. https://doi.org/10.1101/gr.253492.119.

Eckersley-Maslin, M.A., Parry, A., Blotenburg, M., Krueger, C., Ito, Y., Franklin, V.N.R., Narita, M., D'Santos, C.S., and Reik, W. (2020). Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment. Nat. Struct. Mol. Biol. *27*, 696–705. https://doi.org/10.1038/s41594-020-0443-3.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. *10*, 2709–2721. https://doi.org/10.1093/nar/10.8.2709.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. *28*, 817–825. https://doi.org/10.1038/nbt.1662.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216. https://doi.org/10.1038/nmeth.1906.

Escamilla-Del-Arenal, M., da Rocha, S.T., Spruijt, C.G., Masui, O., Renaud, O., Smits, A.H., Margueron, R., Vermeulen, M., and Heard, E. (2013). Cdyl, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. Mol. Cell. Biol. *33*, 5005–5020. https://doi.org/10.1128/MCB.00866-13.

Escobar, T.M., Oksuz, O., Saldaña-Meyer, R., Descostes, N., Bonasio, R., and Reinberg, D. (2019). Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. Cell *179*, 953-963.e11. https://doi.org/10.1016/j.cell.2019.10.009.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156. https://doi.org/10.1038/292154a0.

Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat. Biotechnol. *38*, 276–278. https://doi.org/10.1038/s41587-020-0439-x.

Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiewicz, D., Bates, B., Lee, P.P., Kuhn, R., et al. (2001). DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. J. Neurosci. Off. J. Soc. Neurosci. *21*, 788–797. https://doi.org/10.1523/JNEUROSCI.21-03-00788.2001.

Fang, H., Disteche, C.M., and Berletch, J.B. (2019). X Inactivation and Escape: Epigenetic and Structural Features. Front. Cell Dev. Biol. *7*, 219. https://doi.org/10.3389/fcell.2019.00219.

Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., and Bock, C. (2015). Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics. Cell Rep. *10*, 1386–1397. https://doi.org/10.1016/j.celrep.2015.02.001.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat. Cell Biol. *8*, 188–194. https://doi.org/10.1038/ncb1353.

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Schübeler, D. (2013). Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. PLOS Genet. *9*, e1003994. https://doi.org/10.1371/journal.pgen.1003994.

Ferguson-Smith, A.C., and Bourc'his, D. (2018). The discovery and importance of genomic imprinting. ELife *7*, e42368. https://doi.org/10.7554/eLife.42368.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and 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.

Filion, G.J.P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P.-A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol. Cell. Biol. *26*, 169–181. https://doi.org/10.1128/MCB.26.1.169-181.2006.

Filippakopoulos, P., and Knapp, S. (2012). The bromodomain interaction module. FEBS Lett. *586*, 2692–2704. https://doi.org/10.1016/j.febslet.2012.04.045.

Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V., and Tapscott, S.J. (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat. Genet. *28*, 335–343. https://doi.org/10.1038/ng570.

Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat. Genet. *32*, 426–431. https://doi.org/10.1038/ng988.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. *89*, 1827–1831. https://doi.org/10.1073/pnas.89.5.1827.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Differ. Gene Regul. *15*, 490–495. https://doi.org/10.1016/j.gde.2005.08.002.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science *361*, 1341–1345. https://doi.org/10.1126/science.aau0320.

Gaidatzis, D., Lerch, A., Hahne, F., and Stadler, M.B. (2015). QuasR: quantification and annotation of short reads in R. Bioinformatics *31*, 1130–1132. https://doi.org/10.1093/bioinformatics/btu781.

Galupa, R., and Heard, E. (2018). X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. Annu. Rev. Genet. *52*, 535–566. https://doi.org/10.1146/annurev-genet-120116-024611.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282. https://doi.org/10.1016/0022-2836(87)90689-9.

Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310. https://doi.org/10.1038/s41576-019-0209-0.

Gaston, K., and Fried, M. (1995). CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. Nucleic Acids Res. *23*, 901–909. https://doi.org/10.1093/nar/23.6.901.

Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.-J., et al. (2017). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. J. Biol. Chem. *292*, 14456–14472. https://doi.org/10.1074/jbc.M117.802074.

Gelfman, S., Cohen, N., Yearim, A., and Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. Genome Res. *23*, 789–799. https://doi.org/10.1101/gr.143503.112.

Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev. Cell *23*, 265–279. https://doi.org/10.1016/j.devcel.2012.06.011.

Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs, D.R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat. Genet. *24*, 368–371. https://doi.org/10.1038/74191.

Ginno, P.A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., Zilbermann, F., Peters, A.H.F.M., Edenhofer, F., Smallwood, S.A., et al. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. Nat. Commun. *11*, 2680. https://doi.org/10.1038/s41467-020-16354-x.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., et al. (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575–579. https://doi.org/10.1038/nature18589.

Glaser, L.V., Steiger, M., Fuchs, A., van Bömmel, A., Einfeldt, E., Chung, H.-R., Vingron, M., and Meijsing, S.H. (2021). Assessing genome-wide dynamic changes in enhancer activity during early mESC differentiation by FAIRE-STARR-seq. Nucleic Acids Res. *49*, 12178–12195. https://doi.org/10.1093/nar/gkab1100.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721.

Gonzalo, S., Jaco, I., Fraga, M.F., Chen, T., Li, E., Esteller, M., and Blasco, M.A. (2006). DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nat. Cell Biol. *8*, 416–424. https://doi.org/10.1038/ncb1386.

Grand, R.S., Burger, L., Gräwe, C., Michael, A.K., Isbel, L., Hess, D., Hoerner, L., Iesmantavicius, V., Durdu, S., Pregnolato, M., et al. (2021). BANP opens chromatin and activates CpG-island-regulated genes. Nature *596*, 133–137. https://doi.org/10.1038/s41586-021-03689-8.

Grant, M., Zuccotti, M., and Monk, M. (1992). Methylation of CpG sites of two X-linked genes coincides with X-inactivation in the female mouse embryo but not in the germ line. Nat. Genet. *2*, 161–166. https://doi.org/10.1038/ng1092-161.

Greenberg, M.V.C. (2021). Get Out and Stay Out: New Insights Into DNA Methylation Reprogramming in Mammals. Front. Cell Dev. Biol. *8*. https://doi.org/10.3389/fcell.2020.629068.

Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. *20*, 590–607. https://doi.org/10.1038/s41580-019-0159-6.

Greenberg, M.V.C., Glaser, J., Borsos, M., Marjou, F.E., Walter, M., Teissandier, A., and Bourc'his, D. (2017). Transient transcription in the early embryo sets an

epigenetic state that programs postnatal growth. Nat. Genet. *49*, 110–118. https://doi.org/10.1038/ng.3718.

Gretarsson, K.H., and Hackett, J.A. (2020). Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. Nat. Struct. Mol. Biol. *27*, 706–716. https://doi.org/10.1038/s41594-020-0445-1.

Gu, Z. (2021). rGREAT: Client for GREAT Analysis.

Gu, T., Lin, X., Cullen, S.M., Luo, M., Jeong, M., Estecio, M., Shen, J., Hardikar, S., Sun, D., Su, J., et al. (2018a). DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. Genome Biol. *19*, 88. https://doi.org/10.1186/s13059-018-1464-7.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610. https://doi.org/10.1038/nature10443.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinforma. Oxf. Engl. *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

Gu, Z., Eils, R., Schlesner, M., and Ishaque, N. (2018b). EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC Genomics *19*, 234. https://doi.org/10.1186/s12864-018-4625-x.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res. *22*, 633–641. https://doi.org/10.1101/gr.130997.111.

Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G.A., Feber, A., Teschendorff, A.E., Gomez, V., Hergovich, A., Tirabosco, R., Fernanda Amary, M., et al. (2013). Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. Nat. Commun. *4*, 2166. https://doi.org/10.1038/ncomms3166.

Guo, G., Pinello, L., Han, X., Lai, S., Shen, L., Lin, T.-W., Zou, K., Yuan, G.-C., and Orkin, S.H. (2016). Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis. Cell Rep. *14*, 956–965. https://doi.org/10.1016/j.celrep.2015.12.089.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature *517*, 640–644. https://doi.org/10.1038/nature13899.

Guo, Y., Zhao, S., and Wang, G.G. (2021). Polycomb Gene Silencing Mechanisms: PRC2 Chromatin Targeting, H3K27me3 "Readout", and Phase Separation-Based Compaction. Trends Genet. *37*, 547–565. https://doi.org/10.1016/j.tig.2020.12.006.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. *19*, 621–637. https://doi.org/10.1038/s41580-018-0028-8.

Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., and Stark, A. (2019). Transcriptional cofactors display specificity for distinct types of core promoters. Nature *570*, 122–126. https://doi.org/10.1038/s41586-019-1210-7.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369. https://doi.org/10.1016/j.stem.2013.06.002.

Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. Dev. Camb. Engl. *139*, 3623–3632. https://doi.org/10.1242/dev.081661.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448–452. https://doi.org/10.1126/science.1229277.

Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A.S., Mattei, A.L., Bailly, N., Gottfreund, J., Giesselmann, P., Weigert, R., Brändl, B., et al. (2021). Dnmt1 has de novo activity targeted to transposable elements. Nat. Struct. Mol. Biol. *28*, 594–603. https://doi.org/10.1038/s41594-021-00603-8.

Hainer, S.J., Bošković, A., McCannell, K.N., Rando, O.J., and Fazzio, T.G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. Cell *177*, 1319-1329.e11. https://doi.org/10.1016/j.cell.2019.03.014.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech. Dev. *117*, 15–23. https://doi.org/10.1016/s0925-4773(02)00181-8.

Harrington, M.A., Jones, P.A., Imagawa, M., and Karin, M. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. Proc. Natl. Acad. Sci. U. S. A. *85*, 2066–2070. https://doi.org/10.1073/pnas.85.7.2066.

Hartl, D., Krebs, A.R., Jüttner, J., Roska, B., and Schübeler, D. (2017). Cis-regulatory landscapes of four cell types of the retina. Nucleic Acids Res. *45*, 11607–11621. https://doi.org/10.1093/nar/gkx923.

Hartl, D., Krebs, A.R., Grand, R.S., Baubec, T., Isbel, L., Wirbelauer, C., Burger, L., and Schübeler, D. (2019). CG dinucleotides enhance promoter activity independent of DNA methylation. Genome Res. *29*, 554–563. https://doi.org/10.1101/gr.241653.118.

Hashimoto, H., Wang, D., Horton, J.R., Zhang, X., Corces, V.G., and Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Mol. Cell *66*, 711-720.e3. https://doi.org/10.1016/j.molcel.2017.05.004.

Hayashi, K., Lopes, S.M.C. de S., Tang, F., and Surani, M.A. (2008). Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. Cell Stem Cell *3*, 391–401. https://doi.org/10.1016/j.stem.2008.07.027.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (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.

Héberlé, É., and Bardet, A.F. (2019). Sensitivity of transcription factors to DNA methylation. Essays Biochem. *63*, 727–741. https://doi.org/10.1042/EBC20190033.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112. https://doi.org/10.1038/nature07829.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Hemberger, M., Dean, W., and Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat. Rev. Mol. Cell Biol. *10*, 526–537. https://doi.org/10.1038/nrm2727.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. *18*, 6538–6547. https://doi.org/10.1128/MCB.18.11.6538.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends Genet. TIG *27*, 389–396. https://doi.org/10.1016/j.tig.2011.06.006.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J. Biol. Chem. *279*, 48350–48359. https://doi.org/10.1074/jbc.M403427200.

Hervouet, E., Peixoto, P., Delage-Mourroux, R., Boyer-Guittaut, M., and Cartron, P.-F. (2018). Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. Clin. Epigenetics *10*, 17. https://doi.org/10.1186/s13148-018-0450-y.

Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. Nature *555*, 392–396. https://doi.org/10.1038/nature25964.

Hodges, E., Molaro, A., Dos Santos, C.O., Thekkat, P., Song, Q., Uren, P.J., Park, J., Butler, J., Rafii, S., McCombie, W.R., et al. (2011). Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol. Cell *44*, 17–28. https://doi.org/10.1016/j.molcel.2011.08.026.

Höller, M., Westin, G., Jiricny, J., and Schaffner, W. (1988). Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. Genes Dev. *2*, 1127–1135. https://doi.org/10.1101/gad.2.9.1127.

Holliday, R., and Grigg, G.W. (1993). DNA methylation and mutation. Mutat. Res. *285*, 61–67. https://doi.org/10.1016/0027-5107(93)90052-h.

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., et al. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol. Cell *56*, 286–297. https://doi.org/10.1016/j.molcel.2014.08.026.

Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. *19*, 371–384. https://doi.org/10.1038/s41576-018-0004-3.

Howe, F.S., Fischl, H., Murray, S.C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? BioEssays News Rev. Mol. Cell. Dev. Biol. *39*, 1–12. https://doi.org/10.1002/bies.201600095.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., and Chaillet, J.R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell *104*, 829–838. https://doi.org/10.1016/s0092-8674(01)00280-x.

Hu, D., Gao, X., Cao, K., Morgan, M.A., Mas, G., Smith, E.R., Volk, A.G., Bartom, E.T., Crispino, J.D., Di Croce, L., et al. (2017). Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification. Mol. Cell *65*, 460-475.e6. https://doi.org/10.1016/j.molcel.2017.01.013.

Hu, D., Abbasova, L., Schilder, B.M., Nott, A., Skene, N.G., and Marzi, S.J. (2023). CUT&Tag recovers up to half of ENCODE ChIP-seq peaks in modifications of H3K27. BioRxiv 2022.03.30.486382. https://doi.org/10.1101/2022.03.30.486382.

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., et al. (2013). DNA methylation presents distinct binding sites for human transcription factors. ELife *2*, e00726. https://doi.org/10.7554/eLife.00726.

Huang, S. (2022). Towards a unification of the 2 meanings of "epigenetics." PLOS Biol. *20*, e3001944. https://doi.org/10.1371/journal.pbio.3001944.

Huang, Z., Yu, J., Cui, W., Johnson, B.K., Kim, K., and Pfeifer, G.P. (2021). The chromosomal protein SMCHD1 regulates DNA methylation and the 2*c*-like state of embryonic stem cells by antagonizing TET proteins. Sci. Adv. *7*. https://doi.org/10.1126/sciadv.abb9149.

Hudson, N.O., Whitby, F.G., and Buck-Koehntop, B.A. (2018). Structural insights into methylated DNA recognition by the C-terminal zinc fingers of the DNA reader protein ZBTB38. J. Biol. Chem. *293*, 19835–19843. https://doi.org/10.1074/jbc.RA118.005147.

Iguchi-Ariga, S.M., and Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev. *3*, 612–619. https://doi.org/10.1101/gad.3.5.612.

Inoue, A., Jiang, L., Lu, F., Suzuki, T., and Zhang, Y. (2017a). Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature *547*, 419–424. https://doi.org/10.1038/nature23262.

Inoue, K., Ichiyanagi, K., Fukuda, K., Glinka, M., and Sasaki, H. (2017b). Switching of dominant retrotransposon silencing strategies from posttranscriptional to transcriptional mechanisms during male germ-cell development in mice. PLoS Genet. *13*, e1006926. https://doi.org/10.1371/journal.pgen.1006926.

Iqbal, K., Jin, S.-G., Pfeifer, G.P., and Szabó, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc. Natl. Acad. Sci. *108*, 3642–3647. https://doi.org/10.1073/pnas.1014033108.

Isbel, L., Grand, R.S., and Schübeler, D. (2022). Generating specificity in genome regulation through transcription factor sensitivity to chromatin. Nat. Rev. Genet. *23*, 728–740. https://doi.org/10.1038/s41576-022-00512-6.

Ishiyama, S., Nishiyama, A., Saeki, Y., Moritsugu, K., Morimoto, D., Yamaguchi, L., Arai, N., Matsumura, R., Kawakami, T., Mishima, Y., et al. (2017). Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. Mol. Cell *68*, 350-360.e7. https://doi.org/10.1016/j.molcel.2017.09.037.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and 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., and 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.

Iurlaro, M., Meyenn, F., and Reik, W. (2017). DNA methylation homeostasis in human and mouse development. Curr. Opin. Genet. Dev. *43*, 101–109. https://doi.org/10.1016/j.gde.2017.02.003.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. *28*, 2679–2692. https://doi.org/10.1101/gad.253443.114.

Izzo, F., Lee, S.C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat. Genet. *52*, 378–387. https://doi.org/10.1038/s41588-020-0595-4.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., et al. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. *27*, 31–39. https://doi.org/10.1038/83730.

Jain, D., Meydan, C., Lange, J., Claeys Bouuaert, C., Lailler, N., Mason, C.E., Anderson, K.V., and Keeney, S. (2017). rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. PLOS Genet. *13*, e1006964. https://doi.org/10.1371/journal.pgen.1006964.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and 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.

Jin, B., Li, Y., and Robertson, K.D. (2011). DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? Genes Cancer *2*, 607–617. https://doi.org/10.1177/1947601910393957.

Jin, C., Lu, Y., Jelinek, J., Liang, S., Estecio, M.R.H., Barton, M.C., and Issa, J.-P.J. (2014). TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. Nucleic Acids Res. *42*, 6956–6971. https://doi.org/10.1093/nar/gku372.

Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T.M., Childs, R., et al. (2015). Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature *528*, 142–146. https://doi.org/10.1038/nature15740.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and 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. Nat. Rev. Genet. *13*, 484–492. https://doi.org/10.1038/nrg3230.

Jordà, M., Díez-Villanueva, A., Mallona, I., Martín, B., Lois, S., Barrera, V., Esteller, M., Vavouri, T., and Peinado, M.A. (2017). The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. Genome Res. *27*, 118–132. https://doi.org/10.1101/gr.207522.116.

Kaas, G.A., Zhong, C., Eason, D.E., Ross, D.L., Vachhani, R.V., Ming, G.-L., King, J.R., Song, H., and Sweatt, J.D. (2013). TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. Neuron *79*, 1086–1093. https://doi.org/10.1016/j.neuron.2013.08.032.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J. *32*, 340–353. https://doi.org/10.1038/emboj.2012.331.

Kaluscha, S., Domcke, S., Wirbelauer, C., Stadler, M.B., Durdu, S., Burger, L., and Schübeler, D. (2022). Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. Nat. Genet. *54*, 1895–1906. https://doi.org/10.1038/s41588-022-01241-6.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900–903. https://doi.org/10.1038/nature02633.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell *8*, 676–687. https://doi.org/10.1016/j.stem.2011.04.004.

Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics *13*, 424. https://doi.org/10.1186/1471-2164-13-424.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat. Commun. *10*, 1930. https://doi.org/10.1038/s41467-019-09982-5.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289–294. https://doi.org/10.1038/nature10413.

Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. *22*, 2497–2506. https://doi.org/10.1101/gr.143008.112.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H., et al. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat. Genet. *38*, 149–153. https://doi.org/10.1038/ng1719.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46*, D260–D266. https://doi.org/10.1093/nar/gkx1126.

Kim, S., and Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. Reimagining Cent. Dogma *83*, 373–392. https://doi.org/10.1016/j.molcel.2022.12.032.

King, A.D., Huang, K., Rubbi, L., Liu, S., Wang, C.-Y., Wang, Y., Pellegrini, M., and Fan, G. (2016). Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. Cell Rep. *17*, 289–302. https://doi.org/10.1016/j.celrep.2016.08.083.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol. Cell. Biol. *25*, 3305–3316. https://doi.org/10.1128/MCB.25.8.3305-3316.2005.

Kleinendorst, R.W.D., Barzaghi, G., Smith, M.L., Zaugg, J.B., and Krebs, A.R. (2021). Genome-wide quantification of transcription factor binding at single-DNA-molecule resolution using methyl-transferase footprinting. Nat. Protoc. https://doi.org/10.1038/s41596-021-00630-1.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220. https://doi.org/10.1038/s41576-018-0089-8.

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc. Natl. Acad. Sci. *108*, 14566–14571. https://doi.org/10.1073/pnas.1112317108.

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Äijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., et al. (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.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., et al. (2012). Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. PLOS Genet. *8*, e1002440. https://doi.org/10.1371/journal.pgen.1002440.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., et al. (2011). Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells. Cell Stem Cell *8*, 200–213. https://doi.org/10.1016/j.stem.2011.01.008.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. Cell *98*, 285–294. https://doi.org/10.1016/S0092-8674(00)81958-3.

Krebs, A.R. (2021). Studying transcription factor function in the genome at molecular resolution. Trends Genet. *37*, 798–806. https://doi.org/10.1016/j.tig.2021.03.008.

Krebs, A.R., Dessus-Babus, S., Burger, L., and Schübeler, D. (2014). High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. ELife *3*, e04094. https://doi.org/10.7554/eLife.04094.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Mol. Cell *67*, 411-422.e4. https://doi.org/10.1016/j.molcel.2017.06.027.

Kreibich, E., and Krebs, A.R. (2022). Cofactors: a new layer of specificity to enhancer regulation. Trends Biochem. Sci. *47*, 993–995. https://doi.org/10.1016/j.tibs.2022.07.008.

Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S., and Krebs, A.R. (2023). Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. Mol. Cell *83*, 787-802.e9. https://doi.org/10.1016/j.molcel.2023.01.017.

Kress, T.R., Sabò, A., and Amati, B. (2015). MYC: connecting selective transcriptional control to global RNA production. Nat. Rev. Cancer *15*, 593–607. https://doi.org/10.1038/nrc3984.

Kriaucionis, S., and 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., and Bussemaker, H.J. (2017). Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. Cell Rep. *19*, 2383–2395. https://doi.org/10.1016/j.celrep.2017.05.069.

Krueger, F., James, F., Ewels, P., Afyounian, E., and Schuster-Boeckler, B. (2021). FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo. https://doi.org/10.5281/ZENODO.5127899.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242. https://doi.org/10.1371/journal.pgen.1000242.

Ku, W.L., Nakamura, K., Gao, W., Cui, K., Hu, G., Tang, Q., Ni, B., and Zhao, K. (2019). Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. Nat. Methods *16*, 323–325. https://doi.org/10.1038/s41592-019-0361-7.

Kuhn, M., Jackson, S., and Cimentada, J. (2022). corrr: Correlations in R.

Kumar, A., and Kono, H. (2020). Heterochromatin protein 1 (HP1): interactions with itself and chromatin components. Biophys. Rev. *12*, 387–400. https://doi.org/10.1007/s12551-020-00663-y.

Kumari, D., and Usdin, K. (2001). Interaction of the transcription factors USF1, USF2, and alpha -Pal/Nrf-1 with the FMR1 promoter. Implications for Fragile X mental retardation syndrome. J. Biol. Chem. *276*, 4357–4364. https://doi.org/10.1074/jbc.M009629200.

Lai, B., Gao, W., Cui, K., Xie, W., Tang, Q., Jin, W., Hu, G., Ni, B., and Zhao, K. (2018). Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature *562*, 281–285. https://doi.org/10.1038/s41586-018-0567-3.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

Landan, G., Cohen, N.M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S., Zalcenstein, D.A., Goldfinger, N., Zundelevich, A., et al. (2012). Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat. Genet. *44*, 1207–1214. https://doi.org/10.1038/ng.2442.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genes. N. Y. N 2000 *35*, 88–93. https://doi.org/10.1002/gene.10168.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095–1107. https://doi.org/10.1016/0888-7543(92)90024-m.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell *152*, 1021–1036. https://doi.org/10.1016/j.cell.2013.01.052.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet. *32*, 42–56. https://doi.org/10.1016/j.tig.2015.10.007.

Leddin, M., Perrod, C., Hoogenkamp, M., Ghani, S., Assi, S., Heinz, S., Wilson, N.K., Follows, G., Schönheit, J., Vockentanz, L., et al. (2011). Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. Blood *117*, 2827–2838. https://doi.org/10.1182/blood-2010-08-302976.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Dev. Camb. Engl. *129*, 1807–1817. https://doi.org/10.1242/dev.129.8.1807.

Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Dev. Camb. Engl. *122*, 3195–3205. https://doi.org/10.1242/dev.122.10.3195.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316. https://doi.org/10.1038/nsmb.2510.

Lettice, L.A., Horikoshi, T., Heaney, S.J.H., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. *99*, 7548–7553. https://doi.org/10.1073/pnas.112212199.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914. https://doi.org/10.1016/0092-8674(92)90610-o.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926. https://doi.org/10.1016/0092-8674(92)90611-F.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365. https://doi.org/10.1038/366362a0.

Li, G., Liu, Y., Zhang, Y., Kubo, N., Yu, M., Fang, R., Kellis, M., and Ren, B. (2019). Joint profiling of DNA methylation and chromatin architecture in single cells. Nat. Methods *16*, 991–993. https://doi.org/10.1038/s41592-019-0502-z.

Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., Jing, N.-H., Chen, T., Li, E., Sasaki, H., et al. (2007). Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Mol. Cell. Biol. *27*, 8748–8759. https://doi.org/10.1128/MCB.01380-07.

Li, R., Grimm, S.A., and Wade, P.A. (2021). CUT&Tag-BS for simultaneous profiling of histone modification and DNA methylation with high efficiency and low cost. Cell Rep. Methods *1*, 100118. https://doi.org/10.1016/j.crmeth.2021.100118.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev. Cell *15*, 547–557. https://doi.org/10.1016/j.devcel.2008.08.014.

Li, Y., Xu, Q., Lv, N., Wang, L., Zhao, H., Wang, X., Guo, J., Chen, C., Li, Y., and Yu, L. (2017). Clinical implications of genome-wide DNA methylation studies in acute myeloid leukemia. J. Hematol. Oncol.J Hematol Oncol *10*, 41. https://doi.org/10.1186/s13045-017-0409-z.

Li, Y., Zheng, H., Wang, Q., Zhou, C., Wei, L., Liu, X., Zhang, W., Zhang, Y., Du, Z., Wang, X., et al. (2018). Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. Genome Biol. *19*, 18. https://doi.org/10.1186/s13059-018-1390-8.

Li, Z., Cai, X., Cai, C.-L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.-C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509–4518. https://doi.org/10.1182/blood-2010-12-325241.

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. *43*, 1091–1097. https://doi.org/10.1038/ng.946.

Lin, C.Y., Lovén, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional Amplification in Tumor Cells with Elevated c-Myc. Cell *151*, 56–67. https://doi.org/10.1016/j.cell.2012.08.026.

Lin, S.-P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A.C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. Nat. Genet. *35*, 97–102. https://doi.org/10.1038/ng1233.

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., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322. https://doi.org/10.1038/nature08514.

Liu, Z., and Kraus, W.L. (2017). Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. Mol. Cell *65*, 589-603.e9. https://doi.org/10.1016/j.molcel.2017.01.017.

Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foox, J., Mason, C., Carroll, M., Cheng, A., et al. (2021). DNA methylation-calling tools for Oxford

Nanopore sequencing: a survey and human epigenome-wide evaluation. Genome Biol. *22*, 295. https://doi.org/10.1186/s13059-021-02510-z.

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene *5*, 1743–1753. .

Lock, L.F., Takagi, N., and Martin, G.R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell *48*, 39–46. https://doi.org/10.1016/0092-8674(87)90353-9.

Long, H.K., King, H.W., Patient, R.K., Odom, D.T., and Klose, R.J. (2016). Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. Nucleic Acids Res. *44*, 6693–6706. https://doi.org/10.1093/nar/gkw258.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

Lu, A.T., Fei, Z., Haghani, A., Robeck, T.R., Zoller, J.A., Li, C.Z., Lowe, R., Yan, Q., Zhang, J., Vu, H., et al. (2022). Universal DNA methylation age across mammalian tissues. BioRxiv 2021.01.18.426733. https://doi.org/10.1101/2021.01.18.426733.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. *28*, 2103–2119. https://doi.org/10.1101/gad.248005.114.

Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: In the right place at the right time. Science *361*, 1336–1340. https://doi.org/10.1126/science.aat6806.

Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. *48*, D882–D889. https://doi.org/10.1093/nar/gkz1062.

Lüscher, B. (2001). Function and regulation of the transcription factors of the Myc/Max/Mad network. Gene *277*, 1–14. https://doi.org/10.1016/S0378-1119(01)00697-7.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372–373. https://doi.org/10.1038/190372a0.

Maatouk, D.M., Kellam, L.D., Mann, M.R.W., Lei, H., Li, E., Bartolomei, M.S., and Resnick, J.L. (2006). DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. Dev. Camb. Engl. *133*, 3411–3418. https://doi.org/10.1242/dev.02500.

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat. Protoc. *11*, 1455–1476. https://doi.org/10.1038/nprot.2016.086.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chem. *286*, 35334–35338. https://doi.org/10.1074/jbc.C111.284620.

Mancini, D.N., Singh, S.M., Archer, T.K., and Rodenhiser, D.I. (1999). Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. Oncogene *18*, 4108–4119. https://doi.org/10.1038/sj.onc.1202764.

Mann, I.K., Chatterjee, R., Zhao, J., He, X., Weirauch, M.T., Hughes, T.R., and Vinson, C. (2013). CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. Genome Res. *23*, 988–997. https://doi.org/10.1101/gr.146654.112.

Manzo, M., Wirz, J., Ambrosi, C., Villaseñor, R., Roschitzki, B., and Baubec, T. (2017). Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. EMBO J. *36*, 3421–3434. https://doi.org/10.15252/embj.201797038.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78*, 7634–7638. https://doi.org/10.1073/pnas.78.12.7634.

Mathelier, A., Xin, B., Chiu, T.-P., Yang, L., Rohs, R., and Wasserman, W.W. (2016). DNA Shape Features Improve Transcription Factor Binding Site Predictions In Vivo. Cell Syst. *3*, 278-286.e4. https://doi.org/10.1016/j.cels.2016.07.001.

Mattei, A.L., Bailly, N., and Meissner, A. (2022). DNA methylation: a historical perspective. Trends Genet. *38*, 676–707. https://doi.org/10.1016/j.tig.2022.03.010.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. *23*, 1256–1269. https://doi.org/10.1038/cr.2013.110.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. *12*, 1184–1195. https://doi.org/10.1016/j.celrep.2015.07.024.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010).

GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501. https://doi.org/10.1038/nbt.1630.

McMahon, K.W., Karunasena, E., and Ahuja, N. (2017). The Roles of DNA Methylation in the Stages of Cancer. Cancer J. Sudbury Mass *23*, 257–261. https://doi.org/10.1097/PPO.0000000000000279.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and 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.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877. https://doi.org/10.1093/nar/gki901.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770. https://doi.org/10.1038/nature07107.

Messerschmidt, D.M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. Science *335*, 1499–1502. https://doi.org/10.1126/science.1216154.

Methot, S.P., Padeken, J., Brancati, G., Zeller, P., Delaney, C.E., Gaidatzis, D., Kohler, H., van Oudenaarden, A., Großhans, H., and Gasser, S.M. (2021). H3K9me selectively blocks transcription factor activity and ensures differentiated tissue integrity. Nat. Cell Biol. *23*, 1163–1175. https://doi.org/10.1038/s41556-021-00776-w.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560. https://doi.org/10.1038/nature06008.

Millán-Zambrano, G., Burton, A., Bannister, A.J., and Schneider, R. (2022). Histone post-translational modifications — cause and consequence of genome function. Nat. Rev. Genet. *23*, 563–580. https://doi.org/10.1038/s41576-022-00468-7.

Montagner, S., Leoni, C., Emming, S., Della Chiara, G., Balestrieri, C., Barozzi, I., Piccolo, V., Togher, S., Ko, M., Rao, A., et al. (2016). TET2 Regulates Mast Cell Differentiation and Proliferation through Catalytic and Non-catalytic Activities. Cell Rep. *15*, 1566–1579. https://doi.org/10.1016/j.celrep.2016.04.044.

Moore, J.E., Pratt, H.E., Purcaro, M.J., and Weng, Z. (2020). A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. Genome Biol. *21*, 17. https://doi.org/10.1186/s13059-019-1924-8.

Mor, N., Rais, Y., Sheban, D., Peles, S., Aguilera-Castrejon, A., Zviran, A., Elinger, D., Viukov, S., Geula, S., Krupalnik, V., et al. (2018). Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naive Pluripotency. Cell Stem Cell *23*, 412-425.e10. https://doi.org/10.1016/j.stem.2018.07.004.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. Cancer Cell *20*, 11–24. https://doi.org/10.1016/j.ccr.2011.06.001.

Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. Nat. Genet. *52*, 1271–1281. https://doi.org/10.1038/s41588-020-00736-4.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., et al. (2016). Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. Nat. Biotechnol. *34*, 1060–1065. https://doi.org/10.1038/nbt.3658.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. Cell *111*, 197–208. https://doi.org/10.1016/S0092-8674(02)00976-5.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386–389. https://doi.org/10.1038/30764.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. (2013). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. Genome Biol. *14*, R91. https://doi.org/10.1186/gb-2013-14-8-r91.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77. https://doi.org/10.1038/nature21373.

Neumayr, C., Haberle, V., Serebreni, L., Karner, K., Hendy, O., Boija, A., Henninger, J.E., Li, C.H., Stejskal, K., Lin, G., et al. (2022). Differential cofactor dependencies define distinct types of human enhancers. Nature *606*, 406–413. https://doi.org/10.1038/s41586-022-04779-x.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999). MBD2 is a transcriptional repressor

belonging to the MeCP1 histone deacetylase complex. Nat. Genet. *23*, 58–61. https://doi.org/10.1038/12659.

Nicoglou, A., and Merlin, F. (2017). Epigenetics: A way to bridge the gap between biological fields. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. *66*, 73–82. https://doi.org/10.1016/j.shpsc.2017.10.002.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., et al. (2012). c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. Cell *151*, 68–79. https://doi.org/10.1016/j.cell.2012.08.033.

Nishibuchi, G., and Déjardin, J. (2017). The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Res. *25*, 77–87. https://doi.org/10.1007/s10577-016-9547-3.

Nishiyama, A., and Nakanishi, M. (2021). Navigating the DNA methylation landscape of cancer. Trends Genet. TIG *37*, 1012–1027. https://doi.org/10.1016/j.tig.2021.05.002.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. *12*, 2048–2060. https://doi.org/10.1101/gad.12.13.2048.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat. Genet. *19*, 219–220. https://doi.org/10.1038/890.

Okano, M., Bell, D.W., Haber, D.A., and 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., et al. (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.

O'Neill, L.P., and Turner, B.M. (1995). Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. EMBO J. *14*, 3946–3957. https://doi.org/10.1002/j.1460-2075.1995.tb00066.x.

O'Neill, L.P., and Turner, B.M. (1996). Immunoprecipitation of chromatin. Methods Enzymol. *274*, 189–197. https://doi.org/10.1016/s0076-6879(96)74017-x.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717. https://doi.org/10.1038/nature05987.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep. *10*, 1235–1241. https://doi.org/10.1038/embor.2009.218.

Parry, A., Rulands, S., and Reik, W. (2021). Active turnover of DNA methylation during cell fate decisions. Nat. Rev. Genet. *22*, 59–66. https://doi.org/10.1038/s41576-020-00287-8.

Patel, H., Ewels, P., Peltzer, A., Hammarén, R., Botvinnik, O., Sturm, G., Moreno, D., Vemuri, P., silviamorins, Pantano, L., et al. (2022). nf-core/rnaseq: nf-core/rnaseq v3.6 - Platinum Platypus. https://doi.org/10.5281/ZENODO.6327553.

Paul, D.S., and Beck, S. (2014). Advances in epigenome-wide association studies for common diseases. Trends Mol. Med. *20*, 541–543. https://doi.org/10.1016/j.molmed.2014.07.002.

Perera, A., Eisen, D., Wagner, M., Laube, S.K., Künzel, A.F., Koch, S., Steinbacher, J., Schulze, E., Splith, V., Mittermeier, N., et al. (2015). TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. Cell Rep. *11*, 283–294. https://doi.org/10.1016/j.celrep.2015.03.020.

Pérez, R.F., Tejedor, J.R., Bayón, G.F., Fernández, A.F., and Fraga, M.F. (2018). Distinct chromatin signatures of DNA hypomethylation in aging and cancer. Aging Cell *17*, e12744. https://doi.org/10.1111/acel.12744.

Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeier, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed Engl. *50*, 7008–7012. https://doi.org/10.1002/anie.201103899.

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., et al. (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol. Cell *68*, 955-969.e10. https://doi.org/10.1016/j.molcel.2017.11.013.

Policarpi, C., Munafò, M., Tsagkris, S., Carlini, V., and Hackett, J.A. (2022). Systematic Epigenome Editing Captures the Context-dependent Instructive Function of Chromatin Modifications. BioRxiv https://doi.org/10.1101/2022.09.04.506519.

Pommier, Y., Sun, Y., Huang, S.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. Nat. Rev. Mol. Cell Biol. *17*, 703–721. https://doi.org/10.1038/nrm.2016.111.

Pourrajab, F., and Hekmatimoghaddam, S. (2021). Transposable elements, contributors in the evolution of organisms (from an arms race to a source of raw materials). Heliyon *7*, e06029. https://doi.org/10.1016/j.heliyon.2021.e06029.

Preissl, S., Gaulton, K.J., and Ren, B. (2023). Characterizing cis-regulatory elements using single-cell epigenomics. Nat. Rev. Genet. *24*, 21–43. https://doi.org/10.1038/s41576-022-00509-1.

Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell *65*, 395–407. https://doi.org/10.1016/0092-8674(91)90457-a.

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. *15*, 1613–1618. https://doi.org/10.1101/gad.198501.

Proudhon, C., Duffié, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., Saadeh, H., Holland, M.L., Oakey, R.J., Rakyan, V.K., et al. (2012). Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. Mol. Cell *47*, 909–920. https://doi.org/10.1016/j.molcel.2012.07.010.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol. Cell *44*, 361–372. https://doi.org/10.1016/j.molcel.2011.08.032.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283. https://doi.org/10.1038/nature09692.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

Rampal, R., Alkalin, A., Madzo, J., Vasanthakumar, A., Pronier, E., Patel, J., Li, Y., Ahn, J., Abdel-Wahab, O., Shih, A., et al. (2014). DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. Cell Rep. *9*, 1841–1855. https://doi.org/10.1016/j.celrep.2014.11.004.

Rasmussen, K.D., Berest, I., Keβler, S., Nishimura, K., Simón-Carrasco, L., Vassiliou, G.S., Pedersen, M.T., Christensen, J., Zaugg, J.B., and Helin, K. (2019). TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. Genome Res. *29*, 564–575. https://doi.org/10.1101/gr.239277.118.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Genome Archit. Expr. *43*, 73–81. https://doi.org/10.1016/j.gde.2016.12.007.

Reizel, Y., Morgan, A., Gao, L., Schug, J., Mukherjee, S., García, M.F., Donahue, G., Baur, J.A., Zaret, K.S., and Kaestner, K.H. (2021). FoxA-dependent demethylation of DNA initiates epigenetic memory of cellular identity. Dev. Cell *56*, 602-612.e4. https://doi.org/10.1016/j.devcel.2021.02.005.

Richter, W.F., Nayak, S., Iwasa, J., and Taatjes, D.J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *23*, 732–749. https://doi.org/10.1038/s41580-022-00498-3.

Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y.-H., Wang, L., Rendleman, E.J., et al. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat. Genet. *49*, 1647–1653. https://doi.org/10.1038/ng.3965.

Rishi, V., Bhattacharya, P., Chatterjee, R., Rozenberg, J., Zhao, J., Glass, K., Fitzgerald, P., and Vinson, C. (2010). CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. Proc. Natl. Acad. Sci. U. S. A. *107*, 20311–20316. https://doi.org/10.1073/pnas.1008688107.

Riso, V., Cammisa, M., Kukreja, H., Anvar, Z., Verde, G., Sparago, A., Acurzio, B., Lad, S., Lonardo, E., Sankar, A., et al. (2016). ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells. Nucleic Acids Res. *44*, 8165–8178. https://doi.org/10.1093/nar/gkw505.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein–DNA recognition. Nature *461*, 1248–1253. https://doi.org/10.1038/nature08473.

Rondelet, G., Dal Maso, T., Willems, L., and Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J. Struct. Biol. *194*, 357–367. https://doi.org/10.1016/j.jsb.2016.03.013.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat. Biotechnol. *33*, 1165–1172. https://doi.org/10.1038/nbt.3383.

Rougeulle Claire, Chaumeil Julie, Sarma Kavitha, Allis C. David, Reinberg Danny, Avner Philip, and Heard Edith (2004). Differential Histone H3 Lys-9 and Lys-27 Methylation Profiles on the X Chromosome. Mol. Cell. Biol. *24*, 5475–5484. https://doi.org/10.1128/MCB.24.12.5475-5484.2004.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143. https://doi.org/10.1038/nrm3952.

Sakaue, M., Ohta, H., Kumaki, Y., Oda, M., Sakaide, Y., Matsuoka, C., Yamagiwa, A., Niwa, H., Wakayama, T., and Okano, M. (2010). DNA Methylation Is Dispensable for the Growth and Survival of the Extraembryonic Lineages. Curr. Biol. *20*, 1452–1457. https://doi.org/10.1016/j.cub.2010.06.050.

Saksouk, N., Barth, T.K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.-E., et al. (2014). Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol. Cell *56*, 580–594. https://doi.org/10.1016/j.molcel.2014.10.001.

Sanchez-Delgado, M., Court, F., Vidal, E., Medrano, J., Monteagudo-Sánchez, A., Martin-Trujillo, A., Tayama, C., Iglesias-Platas, I., Kondova, I., Bontrop, R., et al. (2016). Human Oocyte-Derived Methylation Differences Persist in the Placenta Revealing Widespread Transient Imprinting. PLoS Genet. *12*, e1006427. https://doi.org/10.1371/journal.pgen.1006427.

Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., and Dean, W. (2013). Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. Epigenetics Chromatin *6*, 39. https://doi.org/10.1186/1756-8935-6-39.

Sardina, J.L., Collombet, S., Tian, T.V., Gómez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., et al. (2018). Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell *23*, 727-741.e9. https://doi.org/10.1016/j.stem.2018.08.016.

Sartorelli, V., and Lauberth, S.M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. Nat. Struct. Mol. Biol. *27*, 521–528. https://doi.org/10.1038/s41594-020-0446-0.

Scelfo, A., and Fachinetti, D. (2019). Keeping the Centromere under Control: A Promising Role for DNA Methylation. Cells *8*. https://doi.org/10.3390/cells8080912.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat. Genet. *39*, 232–236. https://doi.org/10.1038/ng1950.

Schoelz, J.M., and Riddle, N.C. (2022). Functions of HP1 proteins in transcriptional regulation. Epigenetics Chromatin *15*, 14. https://doi.org/10.1186/s13072-022-00453-8.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. *25*, 582–597. https://doi.org/10.1101/gr.185272.114.

Seale, K., Horvath, S., Teschendorff, A., Eynon, N., and Voisin, S. (2022). Making sense of the ageing methylome. Nat. Rev. Genet. *23*, 585–605. https://doi.org/10.1038/s41576-022-00477-6.

Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature *463*, 563–567. https://doi.org/10.1038/nature08683.

Sérandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Péron, C.L., et al. (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic Acids Res. *40*, 8255–8265. https://doi.org/10.1093/nar/gks595.

Sharda, A., and Humphrey, T.C. (2022). The role of histone H3K36me3 writers, readers and erasers in maintaining genome stability. DNA Repair *119*, 103407. https://doi.org/10.1016/j.dnarep.2022.103407.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912. https://doi.org/10.1038/nature06397.

Sharif, J., Endo, T.A., Nakayama, M., Karimi, M.M., Shimada, M., Katsuyama, K., Goyal, P., Brind'Amour, J., Sun, M.-A., Sun, Z., et al. (2016). Activation of Endogenous Retroviruses in Dnmt1(-/-) ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. Cell Stem Cell *19*, 81–94. https://doi.org/10.1016/j.stem.2016.03.013.

Shayevitch, R., Askayo, D., Keydar, I., and Ast, G. (2018). The importance of DNA methylation of exons on alternative splicing. RNA N. Y. N *24*, 1351–1362. https://doi.org/10.1261/rna.064865.117.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.-L., Zhang, K., and 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.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120. https://doi.org/10.1038/nature11243.

Shipony, Z., Mukamel, Z., Cohen, N.M., Landan, G., Chomsky, E., Zeliger, S.R., Fried, Y.C., Ainbinder, E., Friedman, N., and Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature *513*, 115–119. https://doi.org/10.1038/nature13458.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc. Natl.

Acad. Sci. U. S. A. *100*, 15776–15781. https://doi.org/10.1073/pnas.2136655100.

Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells. Mol. Cell *55*, 319–331. https://doi.org/10.1016/j.molcel.2014.06.029.

Singer-Sam, J., Grant, M., LeBon, J.M., Okuyama, K., Chapman, V., Monk, M., and Riggs, A.D. (1990). Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the Pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation. Mol. Cell. Biol. *10*, 4987–4989. https://doi.org/10.1128/mcb.10.9.4987-4989.1990.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. ELife *6*, e21856. https://doi.org/10.7554/eLife.21856.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordân, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends Biochem. Sci. *39*, 381–399. https://doi.org/10.1016/j.tibs.2014.07.002.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods *11*, 817–820. https://doi.org/10.1038/nmeth.3035.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220. https://doi.org/10.1038/nrg3354.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688–690. https://doi.org/10.1038/336688a0.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344. https://doi.org/10.1038/nature10960.

Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb. Protoc. *2010*, pdb.prot5384. https://doi.org/10.1101/pdb.prot5384.

Song, C.-X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (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, Y., Berg, P.R., Markoulaki, S., Soldner, F., Dall'Agnese, A., Henninger, J.E., Drotar, J., Rosenau, N., Cohen, M.A., Young, R.A., et al. (2019). Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs. Mol. Cell *75*, 905-920.e6. https://doi.org/10.1016/j.molcel.2019.06.045.

Sönmezer, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., and Krebs, A.R. (2021). Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. Mol. Cell *81*, 255-267.e6. https://doi.org/10.1016/j.molcel.2020.11.015.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and 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.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626. https://doi.org/10.1038/nrg3207.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (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., Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495. https://doi.org/10.1038/nature10716.

Stillman, B. (2018). Histone Modifications: Insights into Their Influence on Gene Expression. Cell *175*, 6–9. https://doi.org/10.1016/j.cell.2018.08.032.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45. https://doi.org/10.1038/47412.

Sun, Z., Zhang, Y., Jia, J., Fang, Y., Tang, Y., Wu, H., and Fang, D. (2020). H3K36me3, message from chromatin to DNA damage repair. Cell Biosci. *10*, 9. https://doi.org/10.1186/s13578-020-0374-z.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (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.

Takahashi, N., Coluccio, A., Thorball, C.W., Planet, E., Shi, H., Offner, S., Turelli, P., Imbeault, M., Ferguson-Smith, A.C., and Trono, D. (2019). ZNF445 is a primary regulator of genomic imprinting. Genes Dev. *33*, 49–54. https://doi.org/10.1101/gad.320069.118.

Talbert, P.B., Meers, M.P., and Henikoff, S. (2019). Old cogs, new tricks: the evolution of gene expression in a chromatin context. Nat. Rev. Genet. *20*, 283–297. https://doi.org/10.1038/s41576-019-0105-7.

Tate, P.H., and Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Curr. Opin. Genet. Dev. *3*, 226–231. https://doi.org/10.1016/0959-437X(93)90027-M.

Team BC, and Maintainer BP (2019). TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation package for TxDb object(s). R package version 3.4.7 https://doi.org/10.18129/B9.bioc.TxDb.Mmusculus.UCSC.mm10.knownGene.

Thorvaldsen, J.L., Duran, K.L., and Bartolomei, M.S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev. *12*, 3693–3702. https://doi.org/10.1101/gad.12.23.3693.

Tierney, R.J., Kirby, H.E., Nagra, J.K., Desmond, J., Bell, A.I., and Rickinson, A.B. (2000). Methylation of transcription factor binding sites in the Epstein-Barr virus latent cycle promoter Wp coincides with promoter down-regulation during virus-induced B-cell transformation. J. Virol. *74*, 10468–10479. https://doi.org/10.1128/jvi.74.22.10468-10479.2000.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442–449. https://doi.org/10.1016/j.stem.2009.08.016.

Tsai, Y.-P., Chen, H.-F., Chen, S.-Y., Cheng, W.-C., Wang, H.-W., Shen, Z.-J., Song, C., Teng, S.-C., He, C., and Wu, K.-J. (2014). TET1 regulates hypoxia-induced epithelial-mesenchymal transition by acting as a co-activator. Genome Biol. *15*, 513. https://doi.org/10.1186/s13059-014-0513-0.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R., et al. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells Devoted Mol. Cell. Mech. *11*, 805–814. https://doi.org/10.1111/j.1365-2443.2006.00984.x.

Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. Mol. Cell. Biol. *26*, 9185–9195. https://doi.org/10.1128/MCB.01529-06.

Vanzan, L., Soldati, H., Ythier, V., Anand, S., Braun, S.M.G., Francis, N., and Murr, R. (2021). High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its binding sites. Nat. Commun. *12*, 3337. https://doi.org/10.1038/s41467-021-23630-x.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. *23*, 555–567. https://doi.org/10.1101/gr.147942.112.

Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proc. Natl. Acad. Sci. *107*, 9281–9286. https://doi.org/10.1073/pnas.1000473107.

Verdin, E., and Ott, M. (2015). 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. Nat. Rev. Mol. Cell Biol. *16*, 258–264. https://doi.org/10.1038/nrm3931.

Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q.V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.-J., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat. Genet. *50*, 83–95. https://doi.org/10.1038/s41588-017-0002-y.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., et al. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980. https://doi.org/10.1016/j.cell.2010.08.020.

Vincent, J.J., Huang, Y., Chen, P.-Y., Feng, S., Calvopiña, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., et al. (2013). Stage-Specific Roles for Tet1 and Tet2 in DNA Demethylation in Primordial Germ Cells. Cell Stem Cell *12*, 470–478. https://doi.org/10.1016/j.stem.2013.01.016.

Vukic, M., and Daxinger, L. (2019). DNA methylation in disease: Immunodeficiency, Centromeric instability, Facial anomalies syndrome. Essays Biochem. *63*, 773–783. https://doi.org/10.1042/EBC20190035.

Wachter, E., Quante, T., Merusi, C., Arczewska, A., Stewart, F., Webb, S., and Bird, A. (2014). Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. ELife *3*, e03397. https://doi.org/10.7554/eLife.03397.

Waddington, C.H. (1942). The epigenotype. Endeavour *1*, 18–20. .

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. *20*, 116–117. https://doi.org/10.1038/2413.

Walter, M., Teissandier, A., Pérez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. ELife *5*, e11418. https://doi.org/10.7554/eLife.11418.

Wan, J., Su, Y., Song, Q., Tung, B., Oyinlade, O., Liu, S., Ying, M., Ming, G.-L., Song, H., Qian, J., et al. (2017). Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. ELife *6*, e20068. https://doi.org/10.7554/eLife.20068.

Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., Chen, C., Kou, X., Zhao, Y., Chen, J., et al. (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat. Cell Biol. *20*, 620–631. https://doi.org/10.1038/s41556-018-0093-4.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. *22*, 1680–1688. https://doi.org/10.1101/gr.136101.111.

Wang, H., Fan, Z., Shliaha, P.V., Miele, M., Hendrickson, R.C., Jiang, X., and Helin, K. (2023). H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. Nature *615*, 339–348. https://doi.org/10.1038/s41586-023-05780-8.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. Cell *157*, 979–991. https://doi.org/10.1016/j.cell.2014.04.017.

Wang, Z., Chivu, A.G., Choate, L.A., Rice, E.J., Miller, D.C., Chu, T., Chou, S.-P., Kingsley, N.B., Petersen, J.L., Finno, C.J., et al. (2022). Prediction of histone post-translational modification patterns based on nascent transcription data. Nat. Genet. *54*, 295–305. https://doi.org/10.1038/s41588-022-01026-x.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. *2*, 1136–1143. https://doi.org/10.1101/gad.2.9.1136.

Weber, A.R., Krawczyk, C., Robertson, A.B., Kuśnierczyk, A., Vågbø, C.B., Schuermann, D., Klungland, A., and Schär, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nat. Commun. *7*, 10806. https://doi.org/10.1038/ncomms10806.

Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature *573*, 281–286. https://doi.org/10.1038/s41586-019-1534-3.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (ggplot2: Elegant Graphics for Data Analysis).

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348. https://doi.org/10.1038/nature10066.

Williamson, C.M., Turner, M.D., Ball, S.T., Nottingham, W.T., Glenister, P., Fray, M., Tymowska-Lalanne, Z., Plagge, A., Powles-Glover, N., Kelsey, G., et al. (2006). Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster. Nat. Genet. *38*, 350–355. https://doi.org/10.1038/ng1731.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat. Commun. *2*, 241. https://doi.org/10.1038/ncomms1240.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389–393. https://doi.org/10.1038/nature09934.

Wutz, A., Smrzka, O.W., Schweifer, N., Schellander, K., Wagner, E.F., and Barlow, D.P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature *389*, 745–749. https://doi.org/10.1038/39631.

Xie, S., and Qian, C. (2018). The Growing Complexity of UHRF1-Mediated Maintenance DNA Methylation. Genes *9*. https://doi.org/10.3390/genes9120600.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. Cell *148*, 816–831. https://doi.org/10.1016/j.cell.2011.12.035.

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu, G.-L., Wang, H., et al. (2016). Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. Mol. Cell *64*, 913–925. https://doi.org/10.1016/j.molcel.2016.10.013.

Xu, T.-H., Liu, M., Zhou, X.E., Liang, G., Zhao, G., Xu, H.E., Melcher, K., and Jones, P.A. (2020). Structure of nucleosome-bound DNA methyltransferases DNMT3A and DNMT3B. Nature *586*, 151–155. https://doi.org/10.1038/s41586-020-2747-1.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol. 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., et al. (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.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443–447. https://doi.org/10.1038/nature11709.

Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K., and Zhang, Y. (2013a). Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res. *23*, 329–339. https://doi.org/10.1038/cr.2013.22.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013b). Role of Tet1 in erasure of genomic imprinting. Nature *504*, 460–464. https://doi.org/10.1038/nature12805.

Yang, T., Adamson, T.E., Resnick, J.L., Leff, S., Wevrick, R., Francke, U., Jenkins, N.A., Copeland, N.G., and Brannan, C.I. (1998). A mouse model for Prader-Willi syndrome imprinting-centre mutations. Nat. Genet. *19*, 25–31. https://doi.org/10.1038/ng0598-25.

Yang, Y.A., Zhao, J.C., Fong, K.-W., Kim, J., Li, S., Song, C., Song, B., Zheng, B., He, C., and Yu, J. (2016). FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. Nucleic Acids Res. *44*, 8153–8164. https://doi.org/10.1093/nar/gkw498.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*, eaaj2239. https://doi.org/10.1126/science.aaj2239.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523. https://doi.org/10.1038/nature06968.

Younesian, S., Yousefi, A.-M., Momeny, M., Ghaffari, S.H., and Bashash, D. (2022). The DNA Methylation in Neurological Diseases. Cells *11*. https://doi.org/10.3390/cells11213439.

Zeng, Y., Yao, B., Shin, J., Lin, L., Kim, N., Song, Q., Liu, S., Su, Y., Guo, J.U., Huang, L., et al. (2016). Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. Mol. Cell *61*, 153–160. https://doi.org/10.1016/j.molcel.2015.11.020.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. *21*, 1273–1283. https://doi.org/10.1101/gr.122382.111.

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. (2010a). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell Res. *20*, 1390–1393. https://doi.org/10.1038/cr.2010.156.

Zhang, J., Zhang, Y.-Z., Jiang, J., and Duan, C.-G. (2020a). The Crosstalk Between Epigenetic Mechanisms and Alternative RNA Processing Regulation. Front. Genet. *11*. .

Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu, Y., Li, X., Zhao, D., Liu, Y., Wang, C., Zhang, X., et al. (2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature *525*, 389–393. https://doi.org/10.1038/nature15252.

Zhang, T., Zhang, Z., Dong, Q., Xiong, J., and Zhu, B. (2020b). Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biol. *21*, 45. https://doi.org/10.1186/s13059-020-01957-w.

Zhang, X., Su, J., Jeong, M., Ko, M., Huang, Y., Park, H.J., Guzman, A., Lei, Y., Huang, Y.-H., Rao, A., et al. (2016). DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. Nat. Genet. *48*, 1014–1023. https://doi.org/10.1038/ng.3610.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., et al. (2010b). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. *38*, 4246–4253. https://doi.org/10.1093/nar/gkq147.

Zhao, Z., Chen, L., Dawlaty, M.M., Pan, F., Weeks, O., Zhou, Y., Cao, Z., Shi, H., Wang, J., Lin, L., et al. (2015). Combined Loss of Tet1 and Tet2 Promotes B Cell, but Not Myeloid Malignancies, in Mice. Cell Rep. *13*, 1692–1704. https://doi.org/10.1016/j.celrep.2015.10.037.

Zhou, W., Dinh, H.Q., Ramjan, Z., Weisenberger, D.J., Nicolet, C.M., Shen, H., Laird, P.W., and Berman, B.P. (2018). DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat. Genet. *50*, 591–602. https://doi.org/10.1038/s41588-018-0073-4.

Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. Nat. Rev. Genet. *17*, 551–565. https://doi.org/10.1038/nrg.2016.83.

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., et al. (2018). Single-cell DNA methylome sequencing of human preimplantation embryos. Nat. Genet. *50*, 12–19. https://doi.org/10.1038/s41588-017-0007-6.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477–481. https://doi.org/10.1038/nature12433.

Zuo, X., Sheng, J., Lau, H.-T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M., Xu, G., et al. (2012). Zinc Finger Protein ZFP57 Requires Its Co-factor to Recruit DNA Methyltransferases and Maintains DNA Methylation Imprint in Embryonic Stem Cells via Its Transcriptional Repression Domain*. J. Biol. Chem. *287*, 2107–2118. https://doi.org/10.1074/jbc.M111.322644.

Zuo, Z., Roy, B., Chang, Y.K., Granas, D., and Stormo, G.D. (2017). Measuring quantitative effects of methylation on transcription factor–DNA binding affinity. Sci. Adv. *3*, eaao1799. https://doi.org/10.1126/sciadv.aao1799.