Mechanisms of transcription factor binding and cofactor activity in chromatin

Inauguraldissertation
Zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Marco Pregnolato

Basel, 2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Erstbetreuer: Prof. Dr. Dirk Schübeler

Zweitbetreuer: Prof. Dr. Marc Bühler

Externer Experte: Prof. Dr. Marc Timmers

Basel, den 18.10.2022

Dekan

Prof. Dr. Marcel Mayor

Acknowledgments

First and foremost, I would like to thank my PhD supervisor Dirk Schübeler for mentoring me in the past five years and for giving me the opportunity to pursue different ideas and projects in his group. I am especially grateful for his constant support both in good and bad times, which strongly helped me staying motivated also when I felt I was not progressing.

I would like to thank Ralph Grand, with whom I shared the footprinting project. He guided me in the first steps in the Schübeler lab and made me grow as a scientist with great discussions and teachings. The entire project would have not been possible without his help and support and working together taught me a lot. He was always there when I needed help, also with other projects and outside the lab, thus I am grateful I had the opportunity to have him as a colleague and friend and I am sure he will be a great mentor for his new students.

A big thank you goes to Leslie Hoerner and Christiane Wirbelauer, for their instrumental help in both projects and for their support. Thank you also Christiane for taking the time to speak easy German with me (and especially for waiting until my brain managed to put a sentence together).
I also want to thank Lukas Burger, who patiently taught me how to analyze bioinformatic data and the critical thinking behind it. Something that was not just key for this thesis, but that I greatly enjoyed and that I hope to do even more in the future.

I would like to thank the members of the FMI facilities, especially Sebastien Smallwood, Sirisha Aluri and Eliza Moreno from the genomics facility, for great help with all the sequencing.
Special thanks go to all the members of the Schübeler lab for creating an amazing and supportive atmosphere, full of ideas and fruitful discussions and also for the fun outside the lab.

I would like to thank the members of my PhD thesis committee, Marc Bühler, Marc Timmers and Piera Cicchetti, for their valuable scientific inputs, discussions and great support throughout my PhD.

Special thanks go to Chiara Azzi, Elida Keller, Carlo Campa and Giorgio Galli. Chiara was my chaperone and became soon a great friend. She is an amazing scientist and colleague and she helped me multiple times with things I always overlook. I really enjoyed our time together.
Elida patiently helped me countless times when I kept forgetting about important documents, her organizational skills still impress me. She makes the PhD life at FMI so much easier.
Carlo was my colleague in Turin and I was lucky to find him again in Basel. Our dinners and discussions were great time and, besides being a friend, he has been an additional mentor who has taught me really a lot.
Giorgio welcomed me in Basel for the first time, giving me the opportunity to work in his lab when I was still unsure of what to do. He helped me and supported me in every way possible, from learning new things to starting the PhD, and I am really grateful for that. PS: although, I am sure he agrees that the most important thing is that, with him, I learnt to keep my samples well labeled and my data well organized.

Finally, I want to thank my friends and family. To my friends, thank you for your support and fun time together, especially during the worst part of the pandemic. To my father, mother and sister, thank you for your unconditional support throughout my studies, you always encouraged me to follow my passion and I will always be grateful for that. To Alessandra, thank you for our time together, for your support despite the distance and for being always there for me during this journey. It meant a lot.

List of abbreviations

5mC	5-methylcytosine
ATAC-seq	Assay for Transposase-Accessible Chromatin followed by sequencing
BANP	BTG3 associated nuclear protein
bp	base pair
CGI	CpG island
ChIP-seq	Chromatin immunoprecipitation sequencing
CpG	Cytosine-phosphate-Guanine
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
DAP-seq	DNA affinity purification sequencing
DBD	DNA binding domain
DNA	Desoxyribonucleic acid
DNMT	DNA methyltransferases
DNMT-TKO	DNA methyltransferases triple knockout
ES	Embryonic stem
FDR	False discovery rate
GpC	Guanine-phosphate-Cytosine
H3K27ac	Histone 3 lysine 27 acetylation
H3KXme	Histone 3 lysine X (e.g. 4, 9, 27, 36) methylation
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HT-SELEX	High-throughput systematic evaluation of ligands by exponential SELEX enrichment
ICR	Imprinted control region ING Inhibitor Of Growth
MBD	Methyl-CpG binding domain
NCOR	Nuclear Receptor Corepressor
NOMe-seq	Nucleosome occupancy and methylation sequencing
NRF1	Nuclear respiratory factor 1
NuRD	Nucleosome Remodeling Deacetylase
PBM	Protein binding microarrays
PHD	Plant Homeodomain

REST	RE1-silencing transcription factor
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
SMF	Single molecule footprinting
TE	Transposable element
TET	Ten-eleven translocation methylcytosine dioxygenase
TF	Transcription factor
TKO	Triple knockout
WT	Wild-type

Table of contents

1 Abstract

Organisms rely on the establishment of specific gene expression patterns to ensure proper development and response to external cues. Transcription factors (TFs) and cofactors are key players in transcriptional regulation and they exert their function in the context of chromatin.

Here, DNA is wrapped around histone octamers forming structures called nucleosomes, which drive its compaction and are thought to regulate access of TFs to DNA. Histones can be post-translationally modified adding yet another layer of regulation. In the context of this thesis, I have studied two aspects of this process.

First, I tested the model that nucleosome position influences TF binding and separates TFs into different classes. Towards this goal, we developed a novel approach to systematically measure in the cellular context the ability of individual motifs to recruit TFs and to evaluate how motif position along nucleosomal DNA affects TF binding and nucleosome remodeling. This revealed that only few motifs allow stable recruitment of their cognate TF and that nucleosome phasing over the motif affects binding to different degrees in a highly TF-specific fashion. In contrast, all tested TFs, including the canonical pioneer factors OCT4-SOX2, displayed strong preference for motifs residing in linker regions and at the entry/exit sites of the nucleosome. This suggests that nucleosomes are negative regulator of TF binding *in vivo* and that chromatin sensitivity is a continuous- rather than binary feature.

Secondly, I have investigated how a cofactor that leads to modification of histones is recruited to chromatin and how it functions. More specifically, I focused on the essential corepressor SIN3A and dissected its activity by exploiting an inducible rapid degradation system. We generated a high-quality genome-wide binding profile of SIN3A and followed the effects of its rapid depletion over time. This enabled us to define its primary targets and revealed that SIN3A, despite being a corepressor, is present at almost all active promoters and only at a small fraction of silent genes. We found that SIN3A can be recruited through two independent mechanisms and its binding is required for constant deacetylation and transcriptional repression. We suggest that SIN3A buffers transcription and acetylation at active genes and is selectively recruited at silent promoters to maintain repression.

2 General introduction

2.1 Transcription factors and chromatin in eukaryotes

The development of both unicellular and multicellular organisms requires the correct spatiotemporal interpretation of genomic information. The information content stored in the DNA allows each organism to respond to different stimuli, such as external cues or different developmental stages. However, the genome experiences very little variation during the lifespan of an individual and yet the information that is read from it can vary greatly to adapt to such stimuli. Key players in the establishment of such gene expression patterns are transcription factors (TFs), DNA binding proteins able to recognize consensus motifs across the genome through their DNA binding domains (DBDs) (Banerji et al., 1981; Maniatis et al., 1987; Vaquerizas et al., 2009). These domains can present different structures and have been used to identify TFs and classify them into families. Some of the major TF families, such as C2H2-zinc finger (ZF), Homeodomain, basic helix-loop-helix (bHLH), basic leucin zipper (bZIP) and nuclear hormone receptor (NHR), were initially identified in the 1980s (Johnson and McKnight, 1989). DNA binding proteins were usually discovered by experimental techniques such as N-terminal peptide sequencing, phage libraries, or one-hybrid screening, while today, identification and classification of new putative TFs are carried out mostly through evaluation of sequence homology with already known TFs (Lambert et al., 2018). Interestingly, almost all the ~100 DBD families that have been characterized to date are thought to derive from a small set of common ancestors that possessed some basic fundamental DNA binding properties. Moreover, it is very likely that we have yet to discover new DBD families, which will further expand the known repertoire of DNA binding proteins. One important limitation of identifying TFs using their homology to other sequences is that such homology does not always translate to the ability of binding DNA. For the majority of DBDs families we have catalogues of their structure in complex with DNA, available in the Protein Data Bank (PDB) (Berman et al., 2000), however, other DBD-containing factors have been shown to not bind DNA or to do so only when a specific DBD subfamily was present (Lambert et al., 2018). Thus, while we improved our ability in identifying new potential TFs, a careful validation of their ability to bind DNA is required to correctly classify each candidate into the right category. Recognition of motifs by the DBD enables TFs to achieve specificity and to precisely regulate transcriptional programs by binding at target

regulatory elements such as promoters and enhancers. The ability of TFs to regulate gene expression and cell fate is well illustrated by different examples, such as the reprogramming of fibroblast into pluripotent stem cells following expression of a specific cocktail of TFs, also called the Yamanaka factors (Takahashi and Yamanaka, 2006), or the important role of the transcriptional repressor REST in inhibiting neurogenesis (Gao et al., 2011).

While we know the importance of TFs for gene regulation, we still lack a comprehensive understanding of how they interpret the genome beyond the recognition of their motif. Indeed, early work in prokaryotes showed that DNA binding proteins known as “regulators” are able to bind upstream of genes and regulate their expression by recognizing specific sequences in a predictable manner (Jacob and Monod, 1961; Ptashne, 1967). In contrast, eukaryotic TFs recognize short and highly degenerate DNA sequences that are predicted to appear by chance every four thousand base pairs in mammalian genomes, thus requiring different features to achieve higher specificity (Wunderlich and Mirny, 2009). In the past ten years, a big effort has been put to develop assays able to expand our knowledge of DNA binding specificity of TFs, in particular by exploiting systematic and high-throughput approaches, such as protein binding microarrays (PBM) (Weirauch et al., 2014), high-throughput systematic evaluation of ligands by exponential enrichment (HT-SELEX) (Jolma et al., 2010) and DAP-seq (Bartlett et al., 2017).
These assays strongly improved our ability to predict binding of TFs *in vitro*, however, *in vivo* binding of the same TFs measured by ChIP-seq revealed that only a minor fraction of the predicted motifs is actually bound in the cell, with only few exceptions. For instance, the pluripotency factors OCT4-SOX2 recognize a composite motif and are thought to open chromatin upon binding. However, only ~20% of their best motifs across the genome are bound in mESCs. Also FOXA1, a chromatin opening TF like OCT4-SOX2, has been shown to bind only ~1% of its top sites. In contrast, the repressor TF REST is able to bind almost all of its top motifs in the same cell line (~95%), suggesting a highly different binding mechanism (Fig. 1) (Arnold et al., 2013; Biggin, 2011; Michael et al., 2020). To date, attempts to predict TF binding based on DNA sequence alone remains a major challenge and the discrepancies observed between *in vitro* and *in vivo* strongly suggest that additional chromatin features likely affect motif recognition. Previous works have proposed different models to explain the

DNA binding preference of specific TFs, such as DNA shape around the motif (Gordân et al., 2013; Levo et al., 2015), TF cooperativity (Jolma et al., 2015) or syntax of the binding sites (e.g spacing and orientation) (Farley et al., 2016). However, these can only partially predict the behavior of those TFs *in vivo*. On the other hand, two other main features of chromatin are thought to impair access to DNA, such as histone proteins, around which DNA is folded, and DNA methylation, a chemical modification known to correlate with gene repression. Understanding how these affect TF binding *in vivo* would allow a better understanding of the principles that govern TF specificity in a chromatinized genome and, in turn, regulate cell fate.

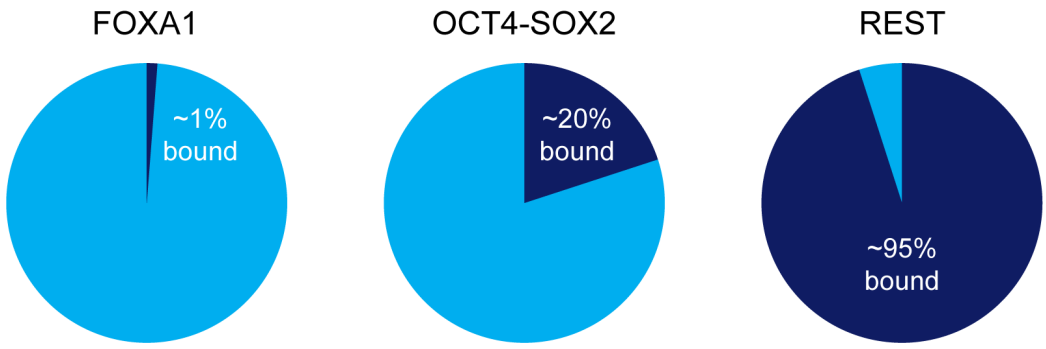


Figure 1: TFs bind only a minority of their predicted motifs in eukaryotes.

Examples of factors that bind only minority of their best motifs across the genome. FOXA1: best motifs in MCF7 cells from (Lupien et al., 2008) selected with 1% FDR (n = ~12,000); OCT4-SOX2: best 1,000 motifs in mESCs from (Michael et al., 2020); REST: best 1,000 motifs in mESCs, own analysis on data from (Arnold et al., 2013). Percentages represent occupancy measured by ChIP-seq.

2.1.1 DNA methylation

Methylation of the fifth carbon on cytosine (5mC) is a common modification across many species (e.g. archaea, bacteria, eukaryotes) and is catalyzed by a highly conserved family of enzymes called DNA methyltransferases (DNMTs) (Goll and Bestor, 2005). Abundance and distribution of 5mC vary greatly in eukaryotic genomes, ranging from nearly absent in the fruit fly *D. melanogaster* (Zemach and Zilberman, 2010) to being the default state in vertebrate genomes such as in humans. Here, this modification occurs at CpG dinucleotides and is predominantly found at transposable elements (TEs) and CpG poor regions, while being depleted at regions with high CpG content, also called GpC islands (CGIs), such as promoters (Fig. 2) (Stadler et al., 2011; Yoder et al., 1997).

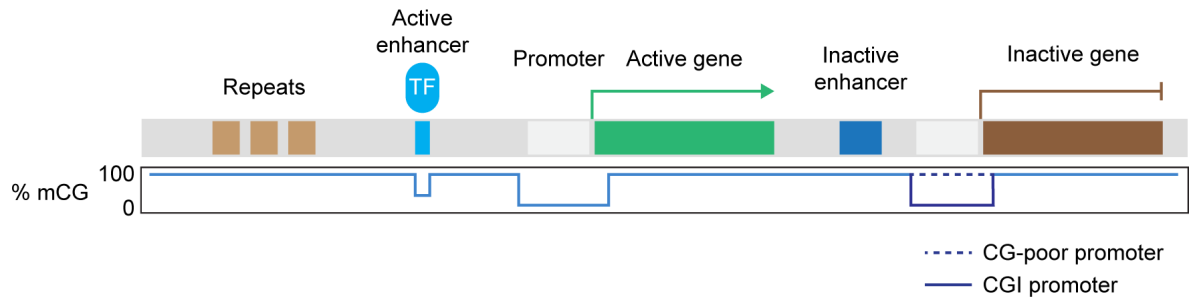


Figure 2: Distribution of DNA methylation at different elements in the mammal genome.

Different genomic regions are illustrated in the top rectangle. The respective levels of CpG methylation are illustrated in the line plot below, expressed as percentage of methylated CpG in a cell population. Adapted from (Schübeler, 2015).

In particular, it has been proposed that DNA methylation arose as a defense mechanism against invasion of viral DNA and TEs, which have the potential to expand, induce rearrangements and overload the transcriptional machinery with their transcripts (Bestor, 1990, 2003). Indeed, cytosine methylation at CpG-rich promoters has been shown to cause strong transcriptional repression (Busslinger et al., 1983; Schübeler et al., 2000) and to drive long-term silencing such as in the case of X chromosome inactivation (Jaenisch and Bird, 2003; Panning and Jaenisch, 1996) and genomic imprinting (Bourc'his et al., 2001; Li et al., 1993). The exact mechanism of transcriptional repression mediated by DNA methylation is currently debated and, to date, two prominent and non-mutually exclusive models have been put forth. One

model suggests that proteins able to recognize methylated CpGs, i.e. methyl-CpG binding domains (MBDs), can do so in a sequence-unspecific manner and translate this modification to repression by interacting with other co-repressors (Tate and Bird, 1993). In another model, presence of 5mC in the motif of a TF is predicted to affect recognition of the motif, leading to different binding efficiency, and thus transcriptional output, depending on the methylation status (Kribelbauer et al., 2020a). While the latter could apply only to TFs binding to CpG-containing motifs, a recent study demonstrated that repression of repetitive element in the mammalian genome is actually explained by a direct inhibition of TF binding, with MBDs being largely dispensable for DNA methylation-mediated transcriptional repression genome wide (Kaluscha et al., 2022). This suggests that DNA methylation is read by TFs and directly translated to transcriptional repression, without involvement of unspecific 5mC readers. Moreover, extending the catalogue of TFs sensitive to DNA methylation would improve our understanding of the regulation of all those genes whose transcription is modulated through methylation of their promoter.

2.1.2 DNA methylation and TF binding

The effect of DNA methylation on TF binding has been tested the first time *in vitro* with gel retardation assays. Here, formation of the TF-DNA complex was prevented when 5mCs were present within the cognate motif, but not outside, suggesting an inhibitory role of motif methylation on TF binding (Watt and Molloy, 1988). Following studies aimed then at a more systematic evaluation of TF sensitivity to DNA methylation, by testing multiple TFs *in vitro* against different methylated versions of their motifs. For instance, a recent study implemented methyl-SELEX and bisulfite-SELEX to assess sensitivity of 529 purified human TFs, full length and DNA binding domain (DBD) only, to methylated and unmethylated DNA sequences (Yin et al., 2017). Here, approximately 40% of TFs were little or not affected by presence of 5mCs, while 23% and 34% showed weaker and stronger binding respectively. Interestingly, not all TFs inhibited by the mark had CpGs in their motif, suggesting the existence of a repulsory mechanism that does not involve the direct interaction between the DBD and the nucleotides composing the motif. Another study in the same year used a similar approach to ask whether the position of the 5mCs can have an effect on TF binding (Kribelbauer et al., 2017). Here, they found that the same TF can be affected in

opposite ways depending on which cytosine is methylated, suggesting a more refined role of DNA methylation in dictating TF binding specificity.

While *in vitro* data have been instrumental to learn how TF are affected by DNA methylation, understanding how those translate *in vivo* and what is the relevance of such relationship requires further study. One of the first attempt to characterize DNA methylation sensitivity *in vivo* focused on the transcription factor CTCF, showing that its binding at the imprinted control region (ICR) *Igf2/H19* occurred primarily at the non-imprinted (thus unmethylated) allele (Bell and Felsenfeld, 2000). Then, subsequent studies showed that abrogation of 5mCs genome-wide had very little effect on localization of CTCF outside of the ICR (Maurano et al., 2015; Stadler et al., 2011), suggesting that sensitivity to DNA-methylation can shape binding in a highly locus-specific fashion. More recent studies have taken advantage of genetic deletion of all three DNA methyltransferases (DNMT TKO) and finally identified multiple TFs that showed high sensitivity to DNA methylation.

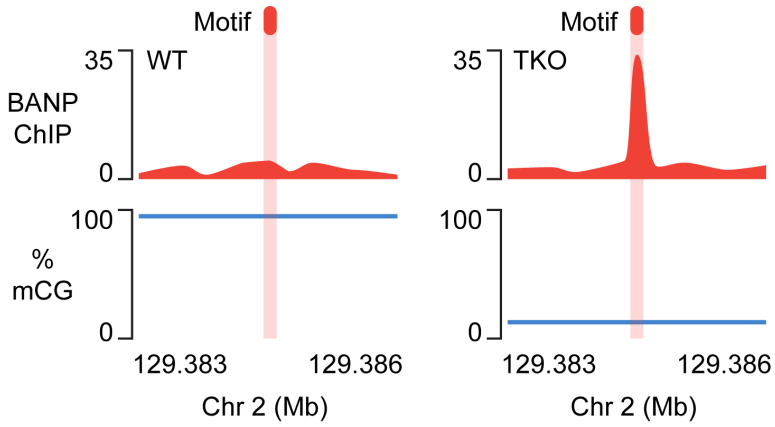


Figure 3: DNA methylation can affect TF binding.

Representative ChIP-seq and CpG methylation tracks in WT and DNMT-TKO cells. Binding of BANP is shown at the same genomic location when methylation is present (WT, left) or absent (TKO, right). BANP motif is highlighted in light orange. Adapted from (Grand et al., 2021).

For instance, comparing the presence of accessible sites with DNaseI digestion between WT and DNMT TKO mouse ES cells revealed that the TF NRF1 was able to bind new motifs and increase chromatin accessibility when these motifs lost DNA methylation (Domcke et al., 2015). Other two recent works found that the activator TFs

BANP and CREB1 are also sensitive to DNA methylation, restricting binding of the first to unmethylated CpG-rich promoters of essential genes and inhibiting binding of the latter at promoters of invasive repetitive elements (Fig. 3) (Grand et al., 2021; Kaluscha et al., 2022). Overall, DNA methylation seems to have a general refractory role for TF binding and has been shown to regulate access to DNA of TFs involved in essential cellular functions.

2.1.3 Nucleosomes and chromatin accessibility

In eukaryotes, genomic DNA is wrapped around histone proteins forming a structure that resembles arrays of “beads on a string”. Here, two copies of the four core histones H2A, H2B, H3 and H4 form octamers and DNA is then wrapped around each octamer for 147 base pairs, generating a structure called nucleosome which is repeated throughout the genome and spaced by short stretches of linker DNA (Fig. 4a) (Kornberg, 1974; Kornberg and Thomas, 1974). Another histone protein named H1 has been shown to bind the nucleosome particle by contacting both the DNA at the core of the nucleosome and part of the linker DNA, helping the compaction and contributing to the formation of the so-called 30nm fiber (Hergeth and Schneider, 2015).

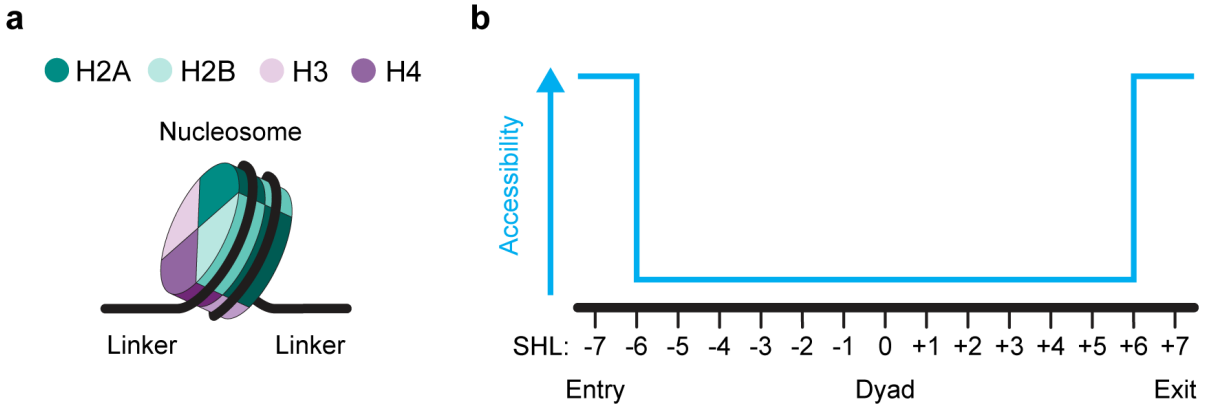


Figure 4: The nucleosome structure.

a) Illustration of a canonical nucleosome (histone H1 not shown). b) Position of each SHL along nucleosomal DNA (147bp) and relative intrinsic accessibility based on DNA contact to histones. Adapted from (Michael et al., 2020).

Within the nucleosome, we can distinguish different discrete locations named superhelical locations (SHL), which range from SHL -7 to SHL +7. These positions represent where the major groove faces towards the histone octamer, with SHL 0 being the dyad, namely the midpoint of the nucleosomal DNA. SHL -7 and SHL +7 represent the entry/exit sites of the nucleosome respectively, where the linker DNA starts and ends its wrapping around the histone octamer (Fig. 4b) (McGinty and Tan, 2015). This configuration generates an inherent accessibility profile along nucleosomal DNA which has been the focus of many studies aimed at understanding the role of nucleosomes in gene regulation. Indeed, DNA wrapping leads to a periodical exposure of the major and minor grooves to the solvent or to the histone

core. This, in turn, can affect how DNA interacts with histones and how proteins can access nucleosomal DNA. Moreover, internal SHLs tend to interact more stably with the histone octamer, while terminal SHLs can experience fluctuations known as nucleosome breathing, a condition where DNA is temporarily detached from the nucleosome and more exposed to external factors (Fierz and Poirier, 2019; Li and Widom, 2004). Therefore, nucleosomes have the potential to regulate how DNA is read out by nuclear factors and understanding how this occurs in the context of the cell will be instrumental to fill the gap between *in vitro* and *in vivo* data.

Initially, chromatin condensation was thought to be the predominant role of the nucleosomes, however, subsequent studies revealed that nucleosomes can have a major influence on gene expression by modulating accessibility of regulatory regions across the genome. Indeed, while they cover the majority of the genome, active regulatory regions typically display low nucleosome occupancy and specific phasing, arguing that nucleosome positioning and gene regulation are tightly linked (Buenrostro et al., 2015; Jiang and Pugh, 2009; Song and Crawford, 2010). For instance, in yeast and metazoans, promoters and termination regions have been shown to be largely nucleosome free, while at transcription start sites (TSSs), nucleosomes are positioned at canonical distances around the site generating a strong nucleosome free region (NFR) (Jiang and Pugh, 2009). Moreover, such strong positioning is restricted to specific sites, such as the already mentioned TSSs and at high-affinity TF binding sites (e.g. CTCF sites), whereas in the rest of the genome nucleosome positioning is much more heterogenous (Fig. 5).

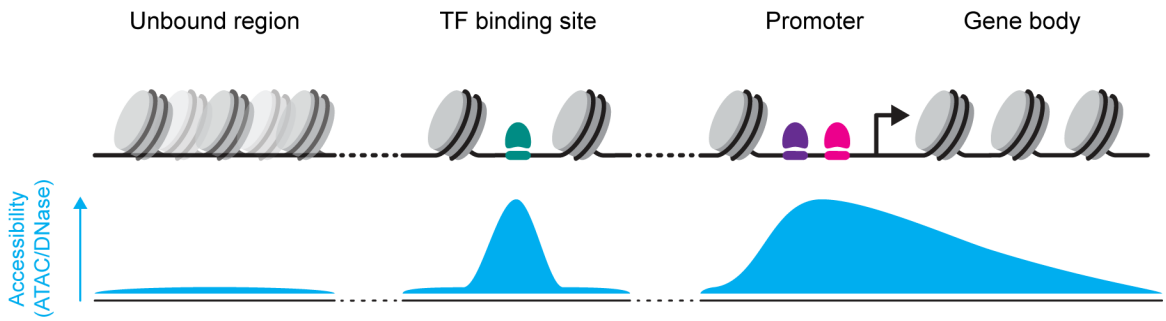


Figure 5: Accessibility differs across the genome.

Examples of different genomic regions with various nucleosome distribution and their relative accessibility profiles according to ATAC-seq or DNAse-I assays.

In fact, despite our ability to assemble nucleosomes *in vitro* through DNA sequence alone using DNA elements like the Widom 601 or the LIN28B (Lowary and Widom, 1998; Soufi et al., 2015), those rules do not apply *in vivo* (Lancrey et al., 2022), arguing that nucleosome positioning is not dictated by DNA sequence and must be enforced by external factors such as TFs and cofactors. This results in specific phasing only at regions where this is likely functional, leaving the rest of the genome covered with nucleosomes in a more unspecific fashion.

2.1.4 Histone variants

Organization of genomic DNA into nucleosomes is not limited to just presence or absence of these structures, but involves other features such as histone variants and histone modifications. These components have been studied for decades and are involved in many different biological processes, yet we lack a mechanistic understanding of how they contribute to chromatin regulation. Each of the core histones possess one or multiple variants whose deposition can be affected by specific conditions and expression can be restricted to specific tissues (Talbert and Henikoff, 2021). For instance, human H2A can be replaced by the variant H2A.X in DNA damage conditions, which is able to recruit the enzymatic machinery required for double-strand break repair. At TSSs, instead, H2A is replaced with H2A.Z, which is thought to enable RNA polymerase II (Pol II) recruitment and thus transcription. Another important variant is CENP-A, a variant of H3 that is found at the centromere of most eukaryotes. It has a fundamental role in building the kinetochore and is one of the most universal histone variants, with only few species that have lost its expression (Talbert and Henikoff, 2021). These and other variants can regulate DNA folding, participate in cell development and differentiation and correlate with differential gene expression, suggesting the existence of a very complex interaction network between histones and the nuclear machinery.

2.1.5 Nucleosomes and TF binding

Most eukaryotic TFs engage with short and degenerate motifs that can occur millions of times just by chance in large genomes. As discussed above, only a minuscule fraction of all these potential binding sites are actually bound in the cell, raising the fundamental question of how specificity is achieved. While methylation has been

shown to effectively restrict or enable access to specific genomic regions, it cannot account for all those TFs that displayed little or no sensitivity to the methylation status of the DNA in *in vitro* assays. In contrast, the genome is pervasively covered by nucleosomes that are known to restrict expression of repeats and likely make the majority of the genome "invisible" to TFs (Isbel et al., 2022). However, regulatory regions need to be accessed at need, thus a balance between repression and activation of those elements must be enforced. A small subset of TFs has been shown to engage with their motif also in the presence of nucleosomes, suggesting that this event would modulate nucleosome positioning and allow establishment of an open regulatory region. This hypothesis led to a binary classification of TFs: on one side, TFs that are chromatin insensitive, also referred to as "pioneer TFs", which are the first to bind closed chromatin; on the other, all those TFs that, in order to access their motif at inaccessible chromatin, require binding of elements of the first group (Fig. 6a-b).

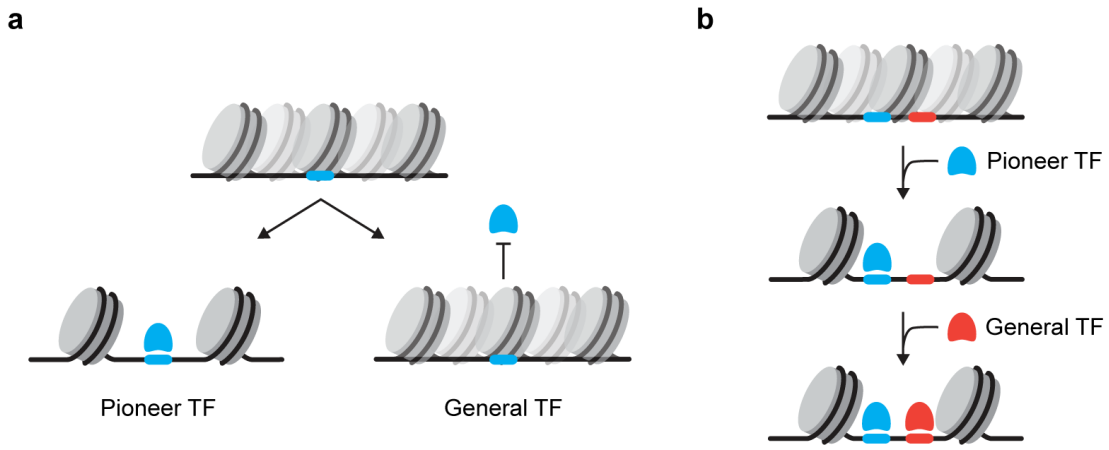


Figure 6: The pioneer factor model.

a) Proposed classification of TFs: pioneer TFs are able to bind their motif when on nucleosomal DNA and general TFs that are inhibited by the presence of nucleosomes. b) Binding events according to the pioneer factor model. A pioneer TF is required to first open chromatin in order to allow a second general TF to bind its motif.

While this model seems intuitive, to which extent this occurs remains a topic of debate. In fact, even pioneer factors bind only a small fraction of their potential binding sites across the genome, arguing that a simple binary classification cannot entirely reflect the complexity observed in reality. Moreover, binding of multiple TFs at the same region suggests that specificity might be achieved by cooperative binding, since their

short motifs have relatively little information and cooperation could improve the ability to bind to specific regions. Accordingly, efforts to generate genome-wide binding maps of TFs such as the Encyclopedia of DNA Elements (ENCODE) (Dunham et al., 2012) or BLUEPRINT (Adams et al., 2012) have revealed that active regulatory regions are often bound by multiple TFs. This also correlates with the presence of open chromatin marks, such as histone acetylation and H3K4 methylation, but whether those are the cause or a consequence of TF binding remains unclear. Understanding the order of events and learning more about potential hierarchies between TFs is therefore critical for decoding the activity of TFs in regulating gene expression.

Binding of TFs at nucleosome occupied regions requires the ability to engage with a motif embedded on nucleosomal DNA. Although eukaryotic motifs are short on average, they are not short enough to reside completely on the solvent-facing portion of the DNA, therefore the nucleosome is likely to hide the motif and prevent its recognition from the TF (Michael and Thomä, 2021). To date, numerous studies have tried to unravel the mechanisms through which TFs manage to engage with their cognate motif on nucleosomes and establish stable binding. Initial work found that nucleosomes can indeed have an inhibitory effect *in vitro* (Li and Wrange, 1993), however, more recent studies suggest that selected TFs are actually able to bind to nucleosomal motifs in a TF-specific fashion. For instance, the reprogramming (Yamanaka) factors OCT4, SOX2 and KLF4 are able to bind a partial motif thanks to a flexible DNA binding domain, which can adapt to the surface of the nucleosome and enable stable binding (Soufi et al., 2015). Moreover, to completely exert its pioneer activity and generate an open chromatin region, OCT4 recruits the chromatin remodeler BRG1, which stabilizes binding of SOX2 at OCT4 occupied sites (King and Klose, 2017). FOXA, instead, can bind nucleosomal DNA because of its similarity in structure to the linker histone H1 and is able to displace nucleosomes without recruiting remodelers (Cirillo, 1998). Alternatively, the previously discussed nucleosome breathing might create chances for opportunistic binding at the entry/exit sites of the nucleosome, potentially allowing more TFs to bind at nucleosome occupied regions. While all these mechanisms are not mutually exclusive, efforts to systematically test TF sensitivity to nucleosomes have indeed revealed that most TFs engage with their motif at the entry/exit sites and that, overall, presence of the nucleosome is largely inhibitory to binding (Fernandez Garcia et al., 2019; Morgunova and Taipale, 2021; Zhu et al., 2018a). Finally, recent structural studies have analyzed

TF-nucleosome interactions at an unprecedented resolution, giving precious insights on how some TFs contact their motif on nucleosomal DNA. Using the Widom's 601 sequence, one work interrogated the ability of the pioneer factors OCT4-SOX2 to bind their combined motif at all possible positions on the nucleosome (Michael et al., 2020). Here, they found that formation of a stable complex was possible when the motif got exposed at the entry/exit sites or when a partial motif faced entirely towards the solvent at these locations. Another work utilized a different nucleosome-containing sequence to test binding of SOX2 and SOX11 and found that these factors could engage with the motif at the more internal SHL+2 and induce local distortion of nucleosomal DNA (Dodonova et al., 2020). Of note, the DNA element used in this latter example contained ten SOX motifs. This, together with the different DNA sequence used to form the nucleosome, might explain the difference observed in SOX binding behavior, thereby highlighting the need for further studies to better disentangle the contribution of motif number and position to TF binding. Moreover, these *in vitro* approaches also present some important limitations. In particular, they rely heavily on reconstituted nucleosomes, as they enable precise targeted perturbation thanks to their highly-predictable and stable positioning. While this strongly benefits the investigation of interactions at the molecular level, it remains unclear how such settings translate to other contexts and *in vivo*, where nucleosomes are less phased and could in turn affect TF binding differently.

Studying TF binding in the context of chromatin *in vivo* is inherently difficult, as cooperativity between factors and bulk assays hinder our ability to isolate the individual contribution of each variable and the relative temporal dynamics. However, the combination of gain- or loss-of-function approaches with techniques to assay nucleosome positioning and TF binding have helped in elucidating binding principles of some TFs as well as dependencies between TFs that likely have different chromatin sensitivity. For instance, a recent work demonstrated that the pioneer factor FOXA1 is not always required for binding of non-pioneer HNF4A at inaccessible chromatin and that it actually requires HNF4A to bind to a subset of its binding sites (Hansen et al., 2022). Interestingly, motif density was critical in driving chromatin opening, with HNF4A requiring slightly higher motif number than FOXA1. This suggests that pioneering activity might be linked to motif density and local concentration of the TF rather than to intrinsic properties. Other works showed that specific TFs (i.e. REST

and CTCF) are able to induce hallmarks of open chromatin, such as DNA demethylation and arrays of phased nucleosomes, revealing that even individual TFs can manipulate chromatin state (Domcke et al., 2015; Sönmezer et al., 2021; Xu et al., 2021). In line with these findings, these TFs have also been shown to bind the majority of their best motifs across the genome (Johnson et al., 2007; Plasschaert et al., 2014), a property that is in stark contrast with that of other pioneer and non-pioneer TFs. Interestingly, comparison of the ratios of bound versus unbound motifs could also be used as a proxy to predict how well a TF can engage with nucleosomal DNA, under the assumption that a TF able to occupy the majority of its best motifs is more likely to do so also at nucleosome occupied regions. In fact, one study revealed that TFs able to generate an accessible site upon binding were also occupying their motifs more frequently (Sherwood et al., 2014). Few years later, another work in yeast identified 29 nucleosome displacing factors and found that tight DNA binding was a key property to compete with nucleosomes, further suggesting that local concentration and residence time might have a role in “pioneering” closed genomic regions (Yan et al., 2018). However, using motif occupancy as a measure for chromatin sensitivity carries also some limitations. In particular, occupancy is not only dictated by affinity of the DNA binding domain for the motif nucleotides, but it can be also affected by the surrounding environment. Factors binding at adjacent motifs can influence the interaction between the TF and its motif through protein-protein interaction, but also the sequence outside the main motif can alter the properties of the DNA strand and affect motif recognition (Heinz et al., 2010; Kribelbauer et al., 2020b). Thus, more refined approaches are required to break down the complexity of the interplay between TFs and nucleosomes *in vivo*.

2.1.6 Histone modifications

On top of histone variants, all histones can undergo post-translational modification. These proteins are mostly globular, but they possess tails that protrude from the main structure that can be modified with a plethora of chemical groups (Strahl and Allis, 2000), such as acetylation, methylation, phosphorylation, ubiquitination and many others (Fig. 7) (Bannister and Kouzarides, 2011).

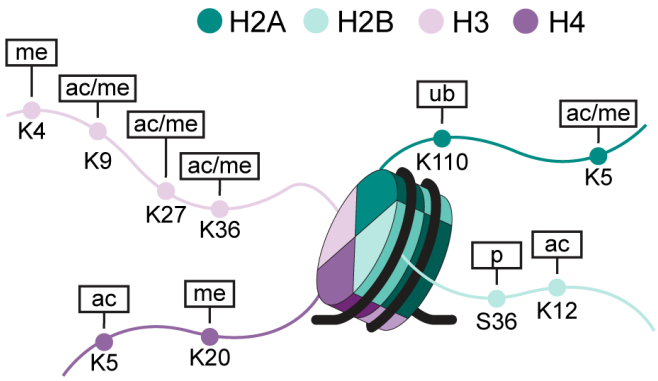


Figure 7: Histone modifications.

Examples of possible tail modifications for each of the four core histones. Length of tails not in scale. Abbreviations are: ac = acetylation; me = methylation; p = phosphorylation; ub = ubiquitination.

The enormous number of known histone modifications makes it hard to describe each of them in detail, nevertheless, there are specific marks that have been correlated with particular chromatin states and are worth to explore more in detail. In particular, transcriptionally inactive and closed chromatin is normally associated with overall low levels of acetylation and methylation of certain residues, such as H3K9, H3K27 and H4K20 (Kouzarides, 2007), while active chromatin displays high levels of acetylation and methylation of different histones, such as H3K4 and H3K36 (Kouzarides, 2007). Furthermore, even within one particular chromatin state, we can identify specific association between genomic elements and histone modifications. For example, both H3K9 and H3K27 methylation associate with repression, but the first is present mostly at silenced repetitive and centromeric elements, whereas the latter is involved in repression of lineage specific genes (Nicetto and Zaret, 2019). Interestingly, marks associated with active and repressed transcription can also coexist, like in the case of H3K4 and H3K27 methylation at the promoters of a subset of silent genes. Initially, these “poised” or “bivalent” state was thought to set genes for rapid activation, for example during development. Recently however, a new study challenged this hypothesis showing that activation of these genes is no more rapid than average and that the role of the H3K4 methylation mark is likely to prevent DNA methylation and subsequent irreversible repression (Kumar et al., 2021). These findings reveal the complexity of chromatin modifications and highlight the need for further studies to help decode their interactions. In fact, little is known about the mechanisms that interpret these marks in the genome and to date different and non-mutually exclusive models have been proposed. One model suggests that addition of groups with different

charges would change the affinity of the histone octamer for the DNA, thereby affecting the compaction of the nucleosome particle (e.g. the negative charge of the acetyl group could repel the negative charges of the DNA and weaken the interaction with the histones) (Bannister and Kouzarides, 2011). On the other hand, these post translational modifications could be read by chromatin factors which, in turn, might act on chromatin modifying its properties and enabling specific processes. Many factors have been identified as able to read, write and erase these modifications and some will be discussed more in detail in a dedicated section as they represent a major focus of this thesis.

Overall, while the association of histone modifications with different chromatin states has been observed and studied for many years, we still lack the mechanistic insights on how they affect gene regulation and on their causal relationship.

2.2 Cofactors and chromatin remodeling

TF binding and establishment of regulatory regions lead to profound reshaping of the chromatin. Transcriptionally active chromatin displays high accessibility, binding of multiple TFs and presence of specific histone marks such as acetylation and H3K4 methylation. In contrast, transcriptionally silent chromatin is characterized by low accessibility, near absence of TF binding and methylation of different residues such as H3K27 and H3K9 (Kouzarides, 2007). This difference in accessibility and histone marks is a consequence of the activity of specific groups of cofactors called chromatin remodelers and modifiers. Cofactors have evolved different ways to manipulate chromatin and their recruitment on chromatin is a key component of gene regulation. They can be classified in different categories: chromatin remodelers, that use ATP to evict or reposition nucleosomes; chromatin writers and erasers, which add or remove chemical groups to and from histone tails and DNA; chromatin readers, that recognize and bind to specific marks (or their absence) on histone tails or TFs. However, while we are increasing our knowledge of their influence of chromatin state, how they achieve specificity across the genome remains unclear. In fact, in contrast to TFs that possess a DBD that recognizes consensus motifs on DNA, cofactors are not able to directly read DNA sequence and must rely on other mechanisms to get access to their targets. While we know little about cofactor recruitment to chromatin, there are now solid evidence that these proteins form large complexes and that numerous subunits possess domains able to interact with TFs and histone modifications (Chen and Dent, 2014). Moreover, the same subunit can participate in formation of different complexes, such as in the case SNF2H and EZH2, the catalytic subunits of the Imitation SWI complex (ISWI) and Polycomb Repressive Complex 2 respectively (Corona and Tamkun, 2004; Margueron and Reinberg, 2011). These features drive the formation of a complex interaction network between TFs, cofactors and histone modifications and might provide a way to achieve specificity independent of a direct recognition of the DNA sequence. Dissecting this network and learning how cofactor binding and activity are regulated will be instrumental to further decode the role of regulatory elements in gene expression. Unfortunately, the promiscuous nature of their interactions makes cofactors inherently difficult to study and major efforts are required to overcome such limitation.

2.2.1 Chromatin remodelers

In order to reposition nucleosomes at regulatory regions and also genome wide, the cell expresses four different families of chromatin remodeling complexes that are able to evict or shift nucleosomes (Fig. 8). All four are ATP-dependent, but their functions do not fully overlap and, within each family, multiple complexes can be observed (Clapier and Cairns, 2009). The four families are SWI/SNF, ISWI, CHD and INO80. All of them share some properties beside the ATP usage, in particular, (a) they have high affinity for nucleosomes, (b) they can recognize histone modifications, (c) they contain subunits able to modulate the ATPase activity and (d) also subunits able to interact with other cofactors and TFs. Of note, remodelers of the CHD and INO80 families can also form stable complexes with histone modifiers, suggesting a close interplay between nucleosome positioning and histone modifications at sites bound by these cofactors.

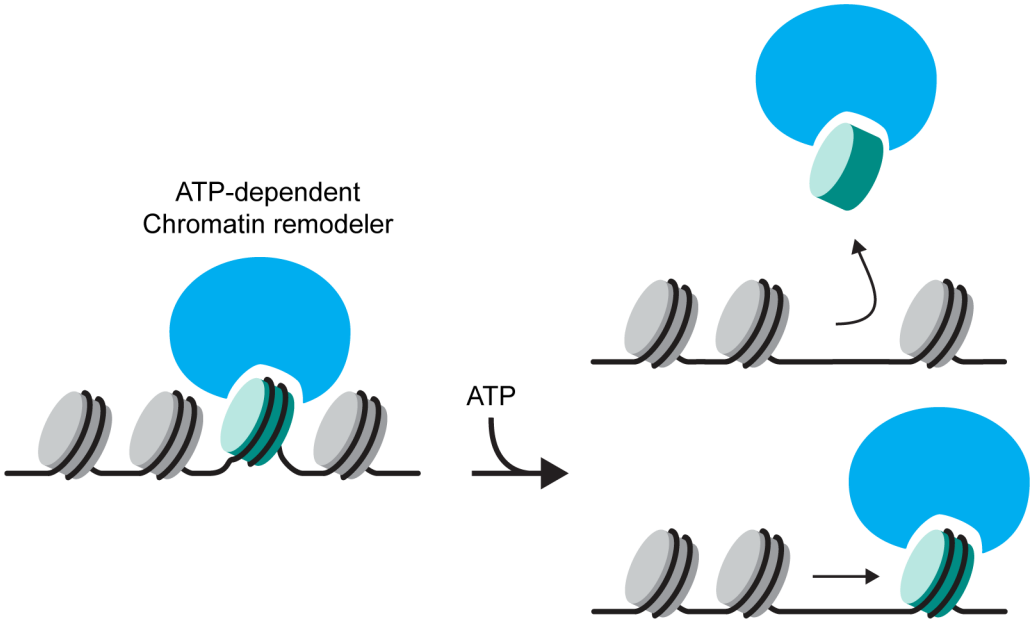


Figure 8: Cromatin remodelers can move nucleosomes with different mechanisms.

Chromatin remodelers exploit ATP hydrolysis to move nucleosomes, resulting in either nucleosome eviction (top right) or nucleosome sliding (bottom right), depending on the complex involved.

Various studies have tried to unravel how remodelers affect chromatin by combining *in vitro* and *in vivo* approaches, finding that different families can have different roles beyond their general ability to move nucleosomes. For instance, a recent work showed that deletion of SNF2H (ISWI) affects nucleosome repeat length (NRL) as well as binding of a specific set of TFs, including CTCF. They also found that loss of BRG1

(SWI/SNF) affects binding of a complete separate group of TFs, while leaving NRL mostly unchanged (Barisic et al., 2019). This argues that remodelers can act on nucleosomes with different mechanisms and can be required in a TF-specific fashion. Yet, whether their binding is restricted to sites occupied by the relative responding TFs, or whether they occupy a bigger fraction of the genome remains to be clarified. Interestingly, another work two years later showed that activity of BRG1 does not reach a steady state, but rather involves a continuous remodeling that enforce local accessibility (Iurlaro et al., 2021). In yeast, other remodelers such as INO80 and its closely-related complex SWR1 have been shown to exert slightly different functions. The first is involved in promoting transcriptional activation and DNA repair, while the latter is able to restructure the nucleosome by exchanging the canonical histone dimer H2A-H2B with the variant H2A.Z-H2B, thereby depositing a histone variant involved in active transcription (Bao and Shen, 2007). Together, these findings have helped unraveling unique features of chromatin remodelers and have instructed us on how their activity can differentially affect chromatin state and TF binding.

2.2.2 Chromatin writers, erasers and readers

Another main group of cofactors is composed by chromatin modifiers. This category is responsible for the deposition and removal of DNA methylation and histone post translational modifications, such as acetylation, methylation, ubiquitination and many more. Like chromatin remodelers, these cofactors display a wide variety of interaction partners and their functional relevance in gene regulation is yet to be fully decoded. In the second part of this thesis, how cofactors regulate acetylation and transcriptional activity at target loci is investigated, thus, the next section will focus on describing what is known about these two features and what remains to be elucidated.

One of the most famous classes of chromatin writers are histone acetyltransferases (HATs), proteins that catalyze the transfer of an acetyl group to histones, TFs and other proteins using acetyl coenzyme A (Ac-CoA) as a donor (Wapenaar and Dekker, 2016). While correlation between histone acetylation and active transcription is clear, we have yet to discover whether this mark is instructive for transcription or whether it is a byproduct needed for further regulation. Initially, histone acetylation by HATs was found to facilitate progression of RNA Pol II through nucleosomes and hence required

for proper gene activation (Barnes et al., 2019). Then, with the advent of CRISPR/Cas9, it has been shown that tethering of the HAT P300 at target loci was sufficient to induce gene activation, suggesting that histone acetylation might directly cause transcriptional activation (Hilton et al., 2015). In contrast, few years later, another study found that acetylation was often occurring as a consequence of transcription and only a minority of sites were actually acetylated in a transcription-independent manner (Martin et al., 2021). Although the study has been debated for its approach (Zencir et al., 2022), it was in line with a previous work that showed that transcription at active enhancers promotes the activity of the HATs CBP/P300, thus describing histone acetylation as an event occurring downstream of transcription (Bose et al., 2017). Together with other works reviewed in (Wapenaar and Dekker, 2016) and (Barnes et al., 2019), these findings argue that HATs might exploit multiple mechanisms to regulate transcription and understanding the general rules of their activity will require further studies.

Another important group of coregulators are histone deacetylases (HDACs). These enzymes are responsible for counteracting the activity of HATs, by removing acetyl groups from different substrates. To date, 18 different HDACs have been identified in mammals and they have been divided in four classes based on their similarity to yeast deacetylases (Park and Kim, 2020). Of particular importance for histone modifications are HDAC1, 2 and 3, which are found in the nucleus and are part, together with other subunits, of bigger corepressor complexes, such as the NuRD complex, Sin3a/HDAC, CoREST and SMRT/N-CoR, with the latter uniquely associated to HDAC3 (Park and Kim, 2020). These complexes have been shown to regulate a plethora of different targets by repressing transcriptional activity, thereby acting as a global tool to regulate gene expression (Ayer, 1999; Kelly and Cowley, 2013; Li et al., 2020). However, while we have a good knowledge of the composition of HDAC-containing complexes and their correlation with transcriptional repression, how activity and recruitment are regulated remains poorly understood. Indeed, despite the strong correlation between deacetylation and transcriptional repression, HDACs binding has been found also at active regulatory regions such as promoters (Kurdistani et al., 2002; Wang et al., 2009), raising the possibility of a positive role of HDACs in gene activation. Although intriguing, this hypothesis is currently based on correlative data and remains largely speculative. Thus, more studies are needed to fully elucidate the molecular mechanisms behind HDAC activity and specificity. Nevertheless, in contrast to HATs,

HDACs have been successfully targeted for anti-cancer therapy. To date, pan inhibitors like SAHA are predominantly used, alone or in combination with other therapies, but efforts to develop more specific and tolerable inhibitors are ongoing (Li et al., 2020).

Specific recruitment of cofactors to chromatin is critical to ensure proper transcriptional output. While they are not able to recognize motifs like TFs do, numerous subunits of coregulator complexes possess domains able to recognize and bind to specific modifications, or their absence, of histone tails and TFs. This mechanism might provide a way to achieve specificity and restrict localization only at regions required for regulation. The first reader of histone acetylation was discovered in 1999, when the bromodomain (BD) of P300 was shown to directly interact with acetylated lysine (Dhalluin et al., 1999). In the following years, many other studies showed that also a chromodomain (CD) and a plant homeodomain (PHD) were able to read methylated lysine (Andrews et al., 2016). The emergence of readers able to interact with different modifications raised the question whether modified histones could provide a sort of "code" that would encode specific transcriptional output. It became clear, however, that the reality was much more complex and that the interplay between different histone modifications required extensive investigation to be deciphered. For instance, H3K4 tri-methylation is a canonical mark of active promoters and can be recognized by the PHD finger of different cofactors.

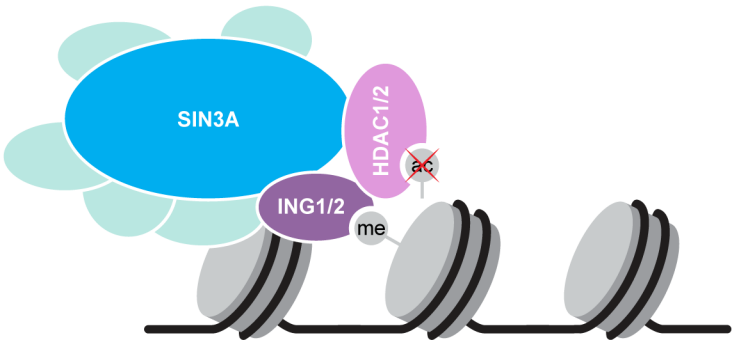


Figure 9: ING proteins link active marks to corepressors: proposed model.

Representation of the proposed model for ING1/2 binding at H3K4me3-positive nucleosome: interaction with other corepressors such as SIN3A and HDAC1/2 could link H3K4me3 to histone deacetylation and gene silencing.

Among those, the ING proteins (ING1 and 2) have been shown to bind to this histone mark both *in vitro* and *in vivo* while simultaneously interacting with corepressors such as SIN3A, HDAC1 and 2 (Fig. 9) (Peña et al., 2008; Shi et al., 2006; Vieyra et al., 2002). Although the exact order of events in their recruitment remains poorly understood, this suggests that active marks can also be instructive for recruitment of corepressors and that the output of specific histone modifications can be strongly context-dependent. Moreover, different subunits have the ability to interact with TFs, thereby linking recruitment of coregulator complexes to TF binding regardless of histone modifications. A known example is CoREST, which has been shown to directly interact with the repressor TF REST as well as with HDACs, and to drive transcriptional repression (Andres et al., 1999; Laugesen and Helin, 2014). Together, these features generate a highly complex crosstalk between chromatin factors and dissecting their molecular interactions will be instrumental to learn the principles that dictate their specificity.

2.3 Methods to study chromatin *in vivo*

2.3.1 Population-based assays

Many techniques have been developed to study chromatin structure and chromatin-bound factors *in vivo* and they can be divided into two main categories: population-based and single-cell or single-molecule assays. Population-based assays capture a snapshot of a specific moment in time and their output represents an average of the entire population. This allows detection and quantification of events occurring at different rates in the cells, however, it also presents strong limitations. For instance, it does not discriminate whether the measured event is occurring in all cells or only at a subset of the population. Also, when two or more events are measured (e.g. binding of two TFs), it does not inform whether they are mutually exclusive or occur simultaneously. Nonetheless, these techniques have allowed extensive analysis of different chromatin features *in vivo*, such as accessibility (DNaseI and ATAC-seq) (Boyle et al., 2008; Buenrostro et al., 2015), nucleosome occupancy (MNase-seq) (Schones et al., 2008), presence of specific DNA binding proteins (ChIP-seq, CUT&RUN) (Skene and Henikoff, 2017; Zhong et al., 2010) and many others. Together, they represent a major toolkit for investigating chromatin biology at the genome-wide scale.

2.3.2 Single-cell and single-molecule assays

Recent advances in single cell and single molecule techniques represent a promising path to go beyond population-based studies and gain better resolution of chromatin dynamics. For instance, single molecule tracking through fluorescent live imaging enables detection of changes after perturbation at the single cell level, potentially also at single molecule (Paakinaho et al., 2017; Swinstead et al., 2016).This technique proves particularly useful in detecting the fast dynamics of TFs on chromatin and the occurrence of transient states that are normally lost in bulk assays. Similarly, genomic footprinting techniques coupled with next generation- and long read sequencing enable readout of TFs and nucleosomes occupancy together with the DNA methylation status at the single molecule level (Sönmezer et al., 2021; Stergachis et al., 2020). Simultaneous detection of different features on the same DNA molecule is critical to go beyond correlation and identify direct relationships between chromatin

proteins. Since genomic footprinting is a major focus of this thesis, it will be described more in detail in the following section.

In conclusion, population-based assays combined with single cell and single molecule techniques could help us to fill the gap between *in vitro* and *in vivo* observations, by allowing a stepwise dissection of the mechanisms that govern binding of TFs and cofactors, nucleosome positioning and other features in the eukaryotic genome. This, in turn, will be fundamental to eventually understand the principles that orchestrate activity of regulatory elements.

2.3.2.1 Genomic footprinting with NOMe-seq

Measuring TF binding and nucleosome occupancy with population-based assays such as ChIP-seq or MNase-seq does not inform whether the observed outputs occur simultaneously or in a mutually exclusive fashion. In contrast, the ability to readout information from individual molecules would allow proper discrimination of the single events. Nucleosome Occupancy and Methylation followed by sequencing (NOMe-seq) is a genomic footprinting technique that takes advantage of a GpC methyltransferase (M. CviPI) to methylate cytosines genome-wide and capture footprints of DNA bound factors (Fig. 10) (Kelly et al., 2012).

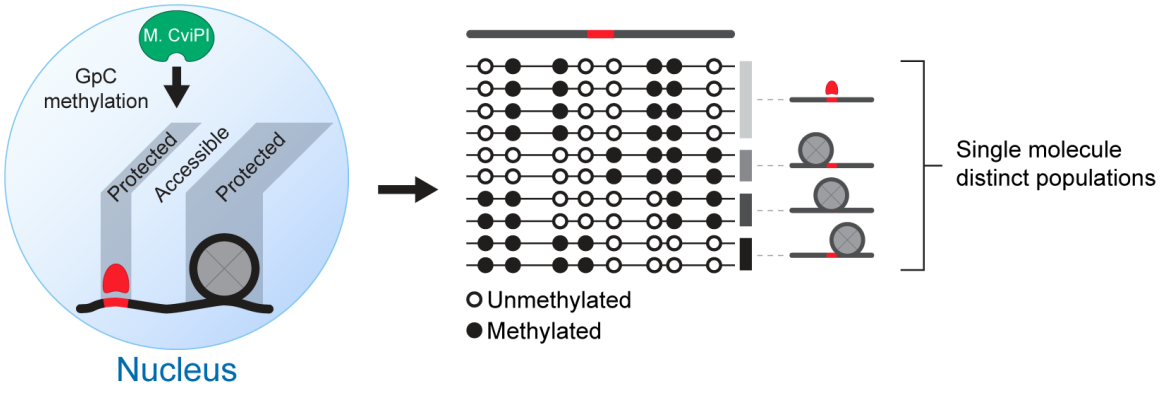


Figure 10: Footprinting the genome with NOMe-seq.

Schematic of NOMe-seq footprinting. Cell nuclei are treated with a GpC methyltransferase without crosslinking. Bound TFs and nucleosomes block methylation from the enzyme and lead to different methylation patterns that can be read at the single molecule level. Reads with similar patterns can be grouped and quantified relative to the others.

First, native nuclear pellets are treated with the enzyme which methylates all unprotected cytosines in the GpC context, thereby generating footprints of unmethylated nucleotides where TFs and nucleosome are present. DNA is then extracted, bisulfite converted and PCR amplified, resulting in the conversion of all the unmethylated Cs to Ts. After next generation sequencing, bioinformatic analysis of the reads allows to discriminate between the endogenous CpG methylation and the ectopic GpC methylation and the latter can be used to evaluate TF- and nucleosome occupancy at the locus of interest. A great advantage of this technique is that information of each single read is retained, thereby enabling detection of footprints at the single molecule (i.e. single cell) level. This allows to ask whether two or more factors bind simultaneously on the same DNA molecule or if their binding is restricted to specific patterns. Furthermore, by combining single read information it is possible to quantify and compare occupancy of different factors at the population level, thus enabling a more precise evaluation of the changes upon perturbation of the system.
While this approach is extremely promising for dissecting molecular interactions, it also comes with some important limitations. For instance, it does not give information on what is bound to DNA. We can infer the presence of a TF or of a nucleosome based on the width of the footprint (nucleosome footprint is normally much larger), but we cannot uniquely assign footprint to specific proteins. In particular, the majority of the motifs in our genome can be bound by multiple TFs that share similar DBDs and NOMe-seq is not able to distinguish which factor is binding at each specific instance. Therefore, orthologous techniques such as ChIP might be required to validate binding of the candidate factor at the locus of interest. Another limitation is that footprints can hide presence of multiple discrete factors, such in the case of a TF that engages with a nucleosome without altering its structure. Here, the larger footprint of the nucleosome would mask the presence of the TF, potentially leading to false conclusions. Again, cross-validation with orthologous techniques can be implemented to exclude such confounders. Similarly, the GpC content of the target locus strongly affect the resolution of the assay. Indeed, NOMe-seq at loci with low GpC content generates poorly-resolved footprints that are difficult to interpret and can be misleading. Finally, we do not know precisely how the GpC methyltransferase competes with DNA binding proteins to access its substrate. Therefore, it is possible that weakly or transiently bound TFs might not be able to protect bound cytosines from methylation, resulting in a lack of footprint that does not reflect the real state.

Overall, combination of NOMe-seq with more standard genomic techniques can be used as an efficient toolkit to dissect the molecular interplay between TFs and nucleosome at the single molecule level.

3 Scope of this thesis

Binding of TFs and cofactors is key to establish proper gene expression patterns. On one hand, how TF binding is regulated by chromatin remains elusive and our ability to predict binding based on DNA sequence is strongly limited. In particular, numerous studies have shown that nucleosomes can act as a barrier to TF binding *in vitro*, but to what extent those findings translates *in vivo* remains unclear. On the other hand, cofactors do not read DNA sequence and how they achieve specificity remains poorly understood. Moreover, their highly-promiscuous interaction network makes it hard to study their recruitment and their effect on chromatin state and transcriptional activity.

In the first part of this thesis, I investigated how nucleosomes affect TF binding *in vivo*. A broadly accepted model has proposed that specific TFs called “pioneers” are able to engage with their motif on nucleosomal DNA, leading to chromatin opening and allowing other TFs to bind. However, evidence supporting this model *in vivo* are mostly correlative. Thus, to go beyond correlation, we developed a novel approach to enforce nucleosome phasing at a specific locus in mouse embryonic stem cells (mESCs) and assay TF binding at different locations along nucleosomal DNA with single molecule footprinting. Firstly, we explored the ability of individual motifs to recruit their cognate TFs in absence of phased nucleosomes. Then, we enforced nucleosome positioning and evaluated the effect of nucleosome position relative to the motif on TF binding, using single molecule analysis to gain quantitative readouts.

In the second part of this thesis, I studied how cofactors achieve binding specificity and how their recruitment affect chromatin state and transcriptional activity. To this end, I focused on SIN3A, a core subunit of the SIN3A/HDAC corepressor complex. SIN3A has been shown to interact with a plethora of TFs that allow its recruitment to drive histone deacetylation and transcriptional repression. However, other studies have found SIN3A also at active promoters where it is thought to act as a coactivator in a context-dependent fashion. Although appealing, this hypothesis is mostly based on late timepoint observations following loss-of-function approaches, which, in the case of essential proteins such as SIN3A, can be misleading due to the onset of secondary effects. To gain a better understanding of SIN3A activity on chromatin and transcription, we took advantage of a degron tag that enabled us to acutely deplete SIN3A in mESCs within one hour and to evaluate its genome-wide localization with

high accuracy. We then exploited the acute depletion to link SIN3A binding profile to changes in chromatin state and transcriptional activity and to identify its primary targets. Lastly, we dissected SIN3A recruitment on chromatin by focusing on histone modifications and manipulating binding of a known SIN3A interaction partner.

4 Results

4.1 Nucleosome mediated motif obstruction impairs occupancy *in vivo* in a highly transcription factor specific fashion (prepared manuscript)

Nucleosome mediated motif obstruction impairs occupancy *in vivo* in a highly transcription factor specific fashion

Ralph S. Grand[1]*, Marco Pregnolato[2,3]*, Leslie Hoerner[2] & Dirk Schübeler[2,3]

1: Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), Im Neuenheimer Feld 282/345, DE 69120 Heidelberg
2: Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH 4058 Basel
3: University of Basel, Faculty of Sciences, Petersplatz 1, CH 4003 Basel

* These authors contributed equally to this work

Correspondence should be addressed to: dirk@fmi.ch

Transcription factors are DNA binding proteins that recognize specific sequence motifs to define proper patterns of gene expression. In eukaryotes, this regulation occurs in a chromatinized genome where nucleosomes are thought to impair motif recognition. Measuring the actual sensitivity of individual TFs to nucleosomes is hindered by the fact that different motifs co-occur within regulatory regions. Here, we apply a reductionist approach to measure in parallel the ability of individual motifs to recruit TFs, revealing that only selected TFs, including previously described pioneer factors, are capable to engage autonomously with a chromatinized template with unstructured nucleosome positioning. To test the ability of individual TFs to bind a defined nucleosomal substrate, we used a novel strategy to generate a highly-phased nucleosome and placed TF motifs at different positions throughout, enabling us to stratify this group further. Here, the pluripotency factors OCT4-SOX2 can only fully engage with motifs residing in the linker and at the entry-exit site, but not within the nucleosome, in contrast to other TFs, such as BANP, which engage and lead to nucleosomal displacement.
These results demonstrate that, even in the context of the cell, nucleosomes can inhibit binding of TFs in a highly TF-specific fashion, with internal sites being intrinsically less accessible, arguing that nucleosome phasing can regulate TF binding.

Introduction

The correct development of multicellular organisms relies on the establishment of specific gene expression patterns. Key players in this process are transcription factors (TFs), proteins that bind short consensus DNA sequence motifs in regulatory regions across the genome (D'Haeseleer, 2006; Inukai et al., 2017). In eukaryotes, this binding occurs in a chromatinized genome, where DNA methylation, nucleosomes and their modifications can affect motif recognition (Li and Wrange, 1993; Padeken et al., 2022; Yin et al., 2017; Zhu et al., 2018). In particular, nucleosomes are constituted by a stretch of 147bp of DNA wrapped around a histone octamer, a structure that restricts access to >95% of the DNA and that has been shown to represent a major barrier for the binding of most TFs *in vitro* (Kornberg, 1974; Michael and Thomä, 2021; Zhu et al., 2018). However, a comprehensive understanding of TF sensitivity to nucleosomes *in vivo* is still lacking and how nucleosomes affect TF binding patterns and in turn transcription remains to be elucidated.

The current perception of gene activation is that specific TFs, known as pioneer factors, are nucleosome insensitive and thereby able to engage with their motifs on nucleosomal DNA to initiate the process of chromatin opening (Cirillo et al., 1998; Hsu et al., 2015; McDaniel et al., 2019; Perlmann and Wrange, 1988; Soufi et al., 2015). This, in turn, can create hierarchies between TFs, since sensitive factors might rely on pioneer factors to bind to their motif and establish a state of open chromatin. Despite its general relevance, this model remains largely speculative due to a limited knowledge of TF sensitivity *in vivo*. In particular, how and to what extent pioneer factors interact with nucleosomes remains unclear. Multiple studies have addressed this question and found that mechanisms of engagement with nucleosomal DNA are TF-specific. For example, the reprogramming (Yamanaka) factors OCT4, SOX2 and KLF4 are able to bind a partial motif thanks to a flexible DNA binding domain, which can adapt to the surface of the nucleosome and enable stable binding (Soufi et al., 2015). Moreover, to completely exert its pioneer activity and generate an open chromatin region, OCT4 recruits the chromatin remodeler BRG1, which stabilizes binding of SOX2 at OCT4 occupied sites (King and Klose, 2017). Notably, these factors have been suggested to bind preferentially to inaccessible chromatin (Soufi et al., 2012, 2015), a feature that is in stark contrast with the general preference of other TFs to bind open chromatin. The paradigm pioneer factor FOXA1, instead, can bind

nucleosomal DNA because of its similarity in structure to the linker histones and is able to displace nucleosomes without recruiting remodelers (Cirillo et al., 2002; Clark et al., 1993). However, *in vitro* studies struggle to recapitulate what happens in the context of the cell, where binding data show that even these TFs occupy just a minor fraction of their top binding sites across the genome (Lupien et al., 2008; Michael et al., 2020). Moreover, despite the large number of *in vivo* studies that investigate TF binding, nucleosome positioning and epigenetic marks, the results remain largely correlative and do not reveal the effect of local chromatin state on the ability of TFs to bind their motifs (King and Klose, 2017; Van Oevelen et al., 2015; Sherwood et al., 2014; Soufi et al., 2012; Wapinski et al., 2013). On the other hand, recent findings showed that FOXA1 is not always required for binding of the non-pioneer factor HNF4A at inaccessible chromatin, with both factors able to pioneer for each other *in vivo* as a function of DNA affinity (Hansen et al., 2022). Another work has shown that the canonical pioneer factors OCT4-SOX2 can only engage with their composite motif at the entry-exit site of a 601 nucleosome, revealing that motif position and nucleosome breathing govern OCT4-SOX2 binding (Michael et al., 2020). Collectively, this argues in favor of a finer stratification of TF groups and highlights the need to study chromatin sensitivity of individual TFs in a more comprehensive manner.

Measuring the actual sensitivity of individual TFs to nucleosomes *in vivo* is inherently difficult due to the fact that different motifs co-occur within regulatory regions (Dunham et al., 2012; Moorman et al., 2006). Thus, to gain a better understanding of how nucleosomes affect TF binding, we developed a reductionist approach that enables us to simultaneously evaluate *in vivo* occupancy of individual TF motifs and nucleosomes in different chromatin contexts using single molecule footprinting (SMF) (Kelly et al., 2012). First, we screened over 100 single motifs occurrences for their ability to be bound in chromatin, revealing that only a minority of TFs, including previously identified pioneer factors such as OCT4-SOX2, are able to engage with their cognate motif in a sequence with unstructured nucleosome positioning. To test the ability of TFs to access their motif on nucleosomal DNA, we implemented a novel strategy to precisely position a nucleosome relative to TF motifs. Strikingly, we found that the pioneer factors OCT4-SOX2 showed a drastic reduction in binding when the motif was nucleosome occupied, similar to non-pioneer factors, while other selected TFs such as BANP, REST and CTCF were able to fully engage with their motif and drive nucleosomal disruption. These results demonstrate that sensitivity to nucleosomes is

strongly TF-dependent *in vivo*, with internal sites being less accessible. This chromatin sensitivity hierarchy likely contributes to shaping specific TF binding patterns across the genome.

Results

A subset of TF motifs are autonomous in factor recruitment and chromatin opening

In the genome, TFs predominantly bind in promoter and enhancer regions that consist of complex motif assemblies making it difficult to distinguish the contribution of individual TFs to chromatin opening. To profile the ability of individual TFs to bind and remodel chromatin *in vivo*, their binding needs to be quantified in a chromatinized template outside of the relative genomic sequence context. Single-molecule footprinting (SMF), which uses methyltransferases and readout by bisulfite sequencing, is able to quantify TF binding and nucleosome positioning at genomic sites as well as engineered genomic loci (Grand et al., 2021; Kelly et al., 2012; Krebs et al., 2017; Pardo et al., 2011; Sönmezer et al., 2021). We set out to define a chromatin sensitivity metric for mouse TFs by introducing single copies of TF motifs into a defined sequence and genomic context and measuring TF binding and chromatin remodeling using SMF **(Fig. 1a)**. More specifically, we used a previously described, *in vitro* derived sequence, that has low nucleosome affinity, high GC content (O/E 1.34), and has not evolved as part of a regulatory region (Yang et al., 2006). We integrated this sequence into a defined locus in mESCs using recombinase-mediated cassette exchange (RMCE) (Lienert et al., 2011), hereafter referred to as RMCE-locus, followed by SMF to detect nucleosome positioning. This resulted in an intermediate accessibility profile, indicative of unstructured nucleosome positioning and low levels of endogenous DNA methylation, as seen at CpG island promoters **(Supplementary Fig. 1a)**.

We then inserted into this sequence over 100 mouse TF consensus motifs from the JASPAR database, ranked based on the highest expression level of the corresponding TF, including 14 scrambled motif controls **(Supplementary Table 1)**. To increase the throughput, we initially compared the integration of a number of constructs individually or as pools of 10 constructs into the RMCE-locus. Following genomic integration, SMF enabled to detect motifs that were bound and if chromatin had been opened relative to control sequences. The SMF obtained for single insertions compared to construct pools where highly reproducible **(Supplementary Fig. 1b)**, thus, we profiled all constructs as pools. A detectable TF footprint was observed for about 15% of the motifs **(Fig. 1b,c)** and only about 3% were also able to open chromatin, including the

TFs CTCF, REST, and BANP **(Fig. 1b,d)**. This enabled us to define a TF chromatin sensitivity hierarchy in absence of structured nucleosome positioning.

We explored a number of TF features to determine if there was a common feature that defined these footprintable TFs. However, the ability to create a TF footprint and open chromatin did not relate to the expression level of the TF, the length of the TF motif, or the DNA binding domain class **(Fig. 1e-g and Supplementary Fig. 1c,d)**. These results indicate that most TFs likely require particular genomic contexts or binding partners to stably bind DNA and open chromatin.

Sequences that position nucleosomes strongly *in vitro* fail to do this in the cellular context in murine embryonic stem cells

The majority of DNA in mammalian genomes is wrapped around nucleosomes, occluding DNA access to proteins that read sequence motifs. However, testing the effect of nucleosome positioning on TF binding *in vivo* is challenging because the majority of nucleosomes in the genome are not precisely phased but instead vary between cells and alleles (Lai et al., 2018). Several sequences have shown to have high affinity for nucleosomes in vitro, most notably the well-established Widom 601 (Lowary and Widom, 1998), for which there is also some evidence that it can position nucleosomes in vivo (Gracey et al., 2010; Perales et al., 2011; Subtil-Rodríguez and Reyes, 2010).

With the goal to generate a locus with precise nucleosome positioning, we inserted the 601 nucleosome positioning sequence into the same genomic locus as for the initial screen and detected resulting nucleosome positioning by SMF. Here, the 601 sequence displayed an intermediate accessibility profile similar to the construct used for the screen, indicative of unstructured nucleosome positioning **(Fig. 2a)**. To better characterized nucleosome occupancy at this locus, we took advantage of the ability to use SMF to look at the position of nucleosomes on individual DNA molecules. This revealed populations of cells that had a nucleosome at relatively defined positions along the sequence **(Fig. 2a)**. However, we found that on average only ~20% of cells had a nucleosome at the same location, thus preventing precise positioning of a nucleosome relative to a TF motif in the majority of cells. This argues that even a sequence that positions well *in vitro* and which has even been challenged to do so in a strength beyond any endogenous sequence, does not position *in vivo*. This is in line with recent results in yeast and suggests that chromatin remodelling enzyme activity

overrides sequence intrinsic features that favor specific nucleosome positioning (Lancrey et al., 2022). It is further compatible with recent reanalysis arguing against a strong contribution of DNA sequence beyond the known effect of AT stretches to disfavor nucleosome binding due to intrinsic stiffness (Zhang et al., 2009).

TF-mediated nucleosome phasing enables precise nucleosome positioning *in vivo*

Previous work, when asking about the location of phased nucleosomes, showed that they reside mostly around transcriptional start sites and in vicinity of TF binding sites (Lai et al., 2018). For example, well positioned nucleosomal arrays are observed upstream and downstream of TF motifs bound by CTCF and REST (Barisic et al., 2019) and individual nucleosomes at these loci can be captured by SMF **(Fig. 2b and Supplementary Fig. 2)**. Thus, as an alternative approach to generate a well-positioned nucleosome, we tested if placing a REST motif at one end of the low nucleosome affinity sequence used for the screen and integrating this construct into the RMCE-locus would result in nucleosome positioning. SMF indeed showed that there was a well-positioned nucleosome next to the footprint for REST in the majority of cells **(Fig. 2c)**. Here, the linker region between the positioning TF and the nucleosome showed small variability, resulting in a nucleosome occupancy over the motif of >80%. Hence, this novel strategy enables to precisely position a nucleosome *in vivo* and provides the opportunity to assess the ability of TFs to bind a nucleosomal substrate directly in the cell.

Nucleosome mediated motif obstruction impairs binding in a TF-specific fashion

To test the ability of TFs to bind and move a nucleosome, we focused on TFs that created a detectable footprint in the screen in the unstructured nucleosome positioning sequence, including also some that did not show significant signal as additional negative controls. For this, we used the REST nucleosome positioning construct and placed the motifs of 17 TFs plus 6 scrambled controls either in the linker next to the REST motif or about 100 bp downstream close to the center of the nucleosome. These constructs were inserted as pools into the RMCE-locus followed by SMF. All the TF motifs that were bound in the unstructured nucleosome positioning sequence were also bound when placed into the linker region **(Supplementary Fig. 3b)**. However, the majority of tested TF motifs did not lead to significant nucleosome disruption when

placed on nucleosomal DNA, including the pioneer TFs OCT4 and SOX2 **(Fig. 3a,b)**. The exceptions were CTCF, REST and BANP, which were able to bind and displace the nucleosome when the motif was placed around the nucleosome dyad **(Fig. 3a,c)**. To gain a higher resolution view of the TF-bound molecules, we extracted and quantified TF binding and nucleosome positioning on individual DNA molecules. This highlighted that only a few TFs are able to bind when their motif is obstructed by a nucleosome **(Fig. 3d)**, suggesting that pioneering activity is limited to a few TFs or is genomic context-dependent.

Motif tiling reveals preference for binding accessible chromatin and nucleosome entry/exit sites

The ability of most TFs to bind their motifs decreases as the motif position approaches the nucleosome dyad *in vitro* (Michael et al., 2020; Zhu et al., 2018). To test if TF binding is similarly impeded *in vivo*, we selected five TFs from different functional classes (repressor, structural, activator) and containing different DNA binding domains (Zinc-fingers, BEN, Basic helix-loop-helix). We tiled their motifs at seven positions across the nucleosome and performed SMF **(Fig 4a)**. This revealed that, while all tested TFs were able to bind in the linker tiles **(Fig. 4a-c)**, only CTCF and REST were able to engage with their motif and move the nucleosome regardless of the tile position **(Fig. 4a)**. This was also true for BANP, although to a lesser extent, making it slightly sensitive to its motif position in the nucleosome **(Fig. 4a,d)**. For BHLHE40 and OCT4-SOX2, the magnitude of the TF footprint decreased as the motif approached the nucleosome dyad, reaching its minimum at the dyad and increasing again moving towards the second linker **(Fig 4a,c,e)**. Notably, at Tile 5, the average profile of the nucleosomal reads fraction differed slightly between the OCT4::SOX2 motif and the scrambled control. It is possible that OCT4-SOX2 recognizes its cognate motif at this position and leads to partial distortion of nucleosomal DNA. However, no changes int the overall nucleosome footprint were detectable, suggesting that this potential distortion is not sufficient to cause nucleosome eviction. Finally, to ensure that positioning the nucleosome with a strong repressor (REST) was not biasing our measurement of chromatin sensitivity, we exchanged the upstream REST motif with a CTCF motif and performed SMF **(Supplementary Fig. 4c)**. All TFs displayed footprints that were highly similar to the ones obtained in the REST-nucleosome context, arguing that their chromatin sensitivity was not affected by the positioning TF

in our setting. Together, these findings suggest that nucleosome positioning *in vivo* generates an intrinsic accessibility profile along nucleosomal DNA and that TFs preferentially engage with motifs residing in the linker regions or entry/exit sites of the nucleosome.

Low accessibility of internal sites results in a strong reduction in binding of the pioneer factors OCT4-SOX2

A limitation of the SMF approach is that it does not give information on simultaneous binding of different factors when they occupy the same portion of DNA. In particular, when a TF binds on a nucleosome, the larger nucleosome footprint will mask the one of the TF. Therefore, absence of TF footprint could not inform whether the TF is bound on the nucleosome without causing nucleosome removal. To test if this might explain the observed differences in TF binding close to the dyad, we focused on the pioneer TFs OCT4-SOX2 and tested their actual binding to motifs at different positions in the nucleosome using ChIP-qPCR. Similarly to SMF, when the motif was placed in the linker region next to REST or CTCF, OCT4 ChIP enrichment was detected and its level was comparable to a highly enriched endogenous locus **(Fig. 4g and Supplementary Fig. 4d)**. In contrast, when the motif was placed more internal to the nucleosome and close to the dyad, the enrichment decreased to almost background **(Fig. 4g and Supplementary Fig. 4d)**. To further validate this finding, we isolated single clones from the initial population of integrants and tested them for directionality of integration before performing ChIP-qPCR. Again, we found stronger OCT4 binding in linker DNA than close to the dyad, where enrichments decreased close to background, strongly resembling the result obtained with the initial cell population **(Supplementary Fig. 4e)**. This result also argues that the slight distortion observed in the average profile of the nucleosomal reads fraction over the OCT4::SOX2 motif **(Fig. 4e)** likely reflects a very weak and unstable engagement of the TFs to the motif. Overall, these findings suggest that even the binding of classical pioneer factors is impeded by the presence of a nucleosome on their motif and that TFs prefer to engage with motifs with higher accessibility such as in linker regions and entry exit/sites.

Discussion

In this study, we survey all known TFs expressed in mESCs for their ability to engage with their chromatinized motifs *in vivo*. To this end, we developed a reductionist approach that enabled us to test the ability of single motifs to recruit their cognate TF in different chromatin contexts and remodel chromatin. We find that only ~10% of motifs are able to recruit TFs within unstructured nucleosome positioning, in agreement with the notion that only a minority of TFs have pioneer activity. Here, only few TFs including REST, CTCF and BANP can create very strong footprints and fully remodel chromatin. Interestingly, ChIP-seq data of these TFs also show that they are able to bind the majority of their motifs *in vivo*, even outside of motif clusters which are typically found in regulatory regions (Grand et al., 2021; Johnson et al., 2007; Plasschaert et al., 2014), in line with the observed low chromatin sensitivity. Surprisingly, the previously identified pioneer TFs OCT4 and SOX2 are individually not able to establish stable binding in this chromatin context and can only do so by binding together to their combined OCT4::SOX2 motif. However, even in this setting, the magnitude of their combined footprint is moderate, challenging their proposed role of chromatin openers. Importantly, absence of detectable binding at specific motifs can be potentially explained by different scenarios, such as weak binding, rapid and transient interactions, or complete absence of engagement. Since our experimental setting is not able to discriminate between these possibilities, more studies are needed to further characterize such motifs. Overall, these findings allow us to stratify TFs based on their ability to engage with individual motifs and reveal that the ability to bind within chromatinized templates is TF specific. Moreover, we find that expression, motif length and DBD class are not sufficient to predict the footprint magnitude, strongly suggesting that these features alone do not dictate chromatin sensitivity and that other factors are at play.

To further define if chromatin engagement is a function of motif position relative to the nucleosome, we enforced precise nucleosome phasing *in vivo*. This became necessary as we find that the Widom 601 *in vitro* positioning sequence does not position nucleosomes *in vivo* **(Fig. 2a)**, likely reflecting the activity of remodeller enzymes and challenging the relevance of *in vitro* experiments with static chromatin. We circumvent this limitation by forcing nucleosome positioning using TFs such as REST and CTCF, in turn generating a physiological chromatin context that can be

precisely manipulated. This enables to contrast existing data *in vitro* with *in vivo*. First, we test the ability of nucleosomes to disrupt binding by positioning a nucleosome at a fixed position relative to the motif and we show that this leads to diminished binding for almost all tested TFs. Indeed, only REST, CTCF and BANP display almost unaltered binding compared to the unstructured nucleosome positioning, while all other tested TFs, including OCT4 and SOX2, display strong nucleosome sensitivity. This approach enabled us to stratify these TFs further and revealed that structured nucleosome positioning has a higher potential to restrict TF binding. Then, we ask what is the contribution of motif position on nucleosomal DNA and select five TFs which encompass different magnitudes of chromatin sensitivity to assay their binding at different positions within the nucleosome. Here, we show that internal locations are inherently less accessible and all TFs display decreased binding when approaching the dyad. Even TFs such as REST, CTCF and BANP, which we show to have low chromatin sensitivity, are somewhat weaker bound at internal sites, arguing that motif position can modulate TF binding also for these factors. Notably, the predicted pioneer factors OCT4 and SOX2 can only bind in linker DNA and at the entry/exit sites of the nucleosome. This is fully in line with previous structural data (Michael et al., 2020) and suggests an important role of nucleosome breathing in governing OCT4-SOX2 binding *in vivo*. Overall, these results reveal that nucleosomes have a general inhibitory role, but each TF is affected differently depending on the motif position relative to the nucleosome.

Previous studies have suggested that pioneer factors are able to initiate chromatin opening by binding to nucleosome occupied regions and proposed that such factors preferentially bind at inaccessible chromatin (Cirillo et al., 2002; McDaniel et al., 2019; Van Oevelen et al., 2015; Soufi et al., 2012; Zaret and Carroll, 2011). However, genome-wide mapping data show that they only occupy a small fraction of their potential binding sites (Lupien et al., 2008; Michael et al., 2020), suggesting that not all inaccessible sites are equally bound. In this study we reveal that nucleosomes inhibit binding also of pioneer factors, thus arguing that their binding at inaccessible regions is not driven by intrinsic preference for nucleosomal substrates. In particular, we find that a single OCT4::SOX2 nucleosomal motif is not sufficient to remodel chromatin, thereby ruling out the possibility of an intrinsic ability of these TFs to bind on nucleosome and drive eviction. Conversely, this raises the possibility that these

TFs might require higher local concentration or cooperation with other TFs to open chromatin, in line with previous data in mouse cell reprogramming (Chronis et al., 2017) and activation of liver genes by FOXA1 in K562 cells (Hansen et al., 2022). Indeed, pioneer factors have been characterized by their binding at sites where other general TFs can access DNA after chromatin has been opened, namely regions where multiple TF motifs are present. Therefore, identifying the individual contribution of each factor is inherently difficult and while pioneer factors seem to have a fundamental role in initiating chromatin remodeling, what is the role of general TFs in this process remains unclear. One stark example of pioneer factors activity is cellular reprogramming, where these TFs target inaccessible chromatin to reshape the gene expression landscape (Soufi et al., 2015; Takahashi and Yamanaka, 2006). However, this process alone has been shown to have very low efficiency (Schlaeger et al., 2014), suggesting that these factors do not act as a simple on/off switch and that more variables govern their activity. The ability of nucleosomes to restrict pioneer factors binding might be at the basis of such heterogeneity and it is possible that general TFs act in concert with pioneer TFs to establish accessible sites by outcompeting nucleosomes at specific positions. In this scenario, pioneer TFs are likely necessary to initiate remodeling, but might not be sufficient. This would fit with the observation that TFs such as OCT4 and SOX2 are primarily found at sites co-occupied by other TFs. To date, general TFs are thought to be fully dependent on pioneer TFs, however, we propose that binding hierarchies among TFs are more complex and likely shaped by cooperativity rather than opportunistic binding.

Preventing binding of TFs by masking motifs with nucleosomes could represent an efficient way to evolve specificity without mutating the DNA. Indeed, short degenerate TF motifs can be found in millions of instances across large genomes and mutating them away to redirect TF binding would require a very high mutation rate which can be deleterious for survival. Moreover, TF concentration might not be sufficient to ensure occupancy at all sites across the genome and unspecific binding could result in spurious transcription. Thus, we speculate that TFs are unlikely to have intrinsic pioneering properties and that chromatin and nucleosome sensitivity is a general regulator of TF binding.

Taken together, our study provides direct *in vivo* evidence of the complex and TF-specific inhibitory effect of nucleosomes on TF binding. It argues for a scenario more complex than a binary distinction of TFs in pioneers and non-pioneers. Instead, the ability to bind in a chromatinized genome can be framed in a broader context of chromatin sensitivity, which we show to be highly TF-specific. New insights into what features shape chromatin sensitivity will be key to understanding how TF binding specificity is achieved across the genome and to improve our ability to predict TF binding starting from DNA sequence and chromatin state.

Figures

Figure 1

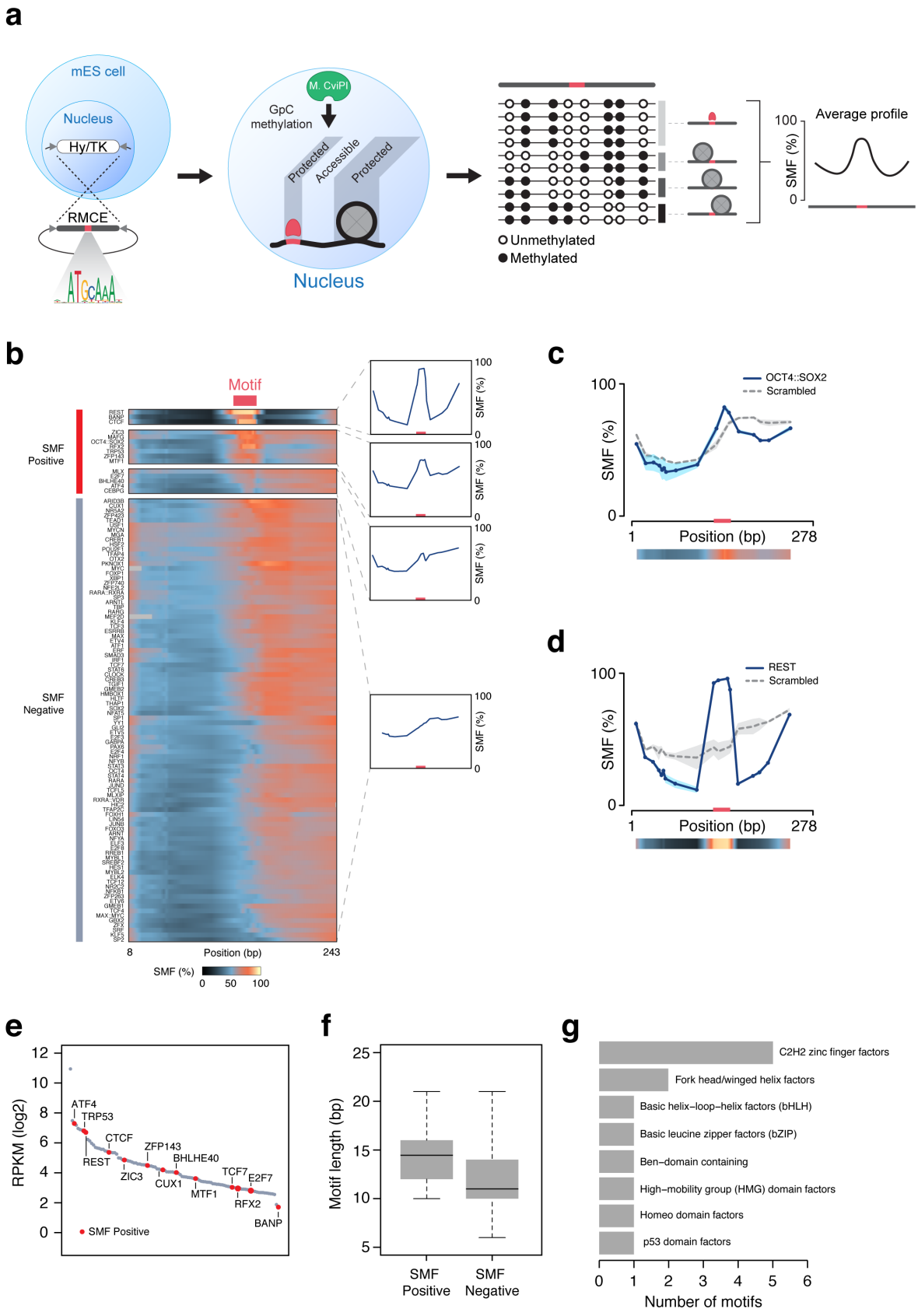

a
mES cell
Nucleus
Hy/TK
RMCE
M. CviPI
GpC methylation
Protected
Accessible
Protected
Nucleus
Unmethylated
Methylated
Average profile
100
SMF (%)
0
b
Motif
SMF Positive
SMF Negative
8
Position (bp)
243
SMF (%)
0 50 100
c
OCT4::SOX2
Scrambled
SMF (%)
Position (bp)
1
278
d
REST
Scrambled
e
RPKM (log2)
ATF4
TRP53
CTCF
REST
ZFP143
ZIC3
BHLHE40
CUX1
MTF1
TCF7
E2F7
RFX2
BANP
SMF Positive
f
Motif length (bp)
SMF Positive
SMF Negative
g
C2H2 zinc finger factors
Fork head/winged helix factors
Basic helix–loop–helix factors (bHLH)
Basic leucine zipper factors (bZIP)
Ben–domain containing
High–mobility group (HMG) domain factors
Homeo domain factors
p53 domain factors
Number of motifs

Figure 1: Screen of motifs able to recruit TFs in mESCs

a) Representation of SMF workflow, from RMCE-mediated integration to SMF output.
b) SMF heatmap of all tested motif inserted into unstructured nucleosome positioning context. Shown is the average of the replicates (n = 2). Line plots on the right-hand side of the heatmap display the average signal of the respective group. Motif position is shown by the red rectangle.
c) SMF average profile over OCT4::SOX2 and OCT4::SOX2 scrambled motif. Lines represent the mean of the replicates (n = 2) and the shades represent the minimum and maximum value at each position. The respective heatmap from b) is shown below the plot.
d) SMF average profile over REST and REST scrambled motif. Lines represent the mean of the replicates (n = 2) and the shades represent the minimum and maximum value at each position. The respective heatmap from b) is shown below the plot.
e) RPKM values of TFs included in the screen in mESCs. TFs that generate a detectable footprint are highlighted in red.
f) Average length of motifs that are SMF Positive or SMF negative. All motifs are included (no outliers were removed).
g) Quantification of DNA binding domains class which bind to motifs that are SMF Positive.

Figure 2

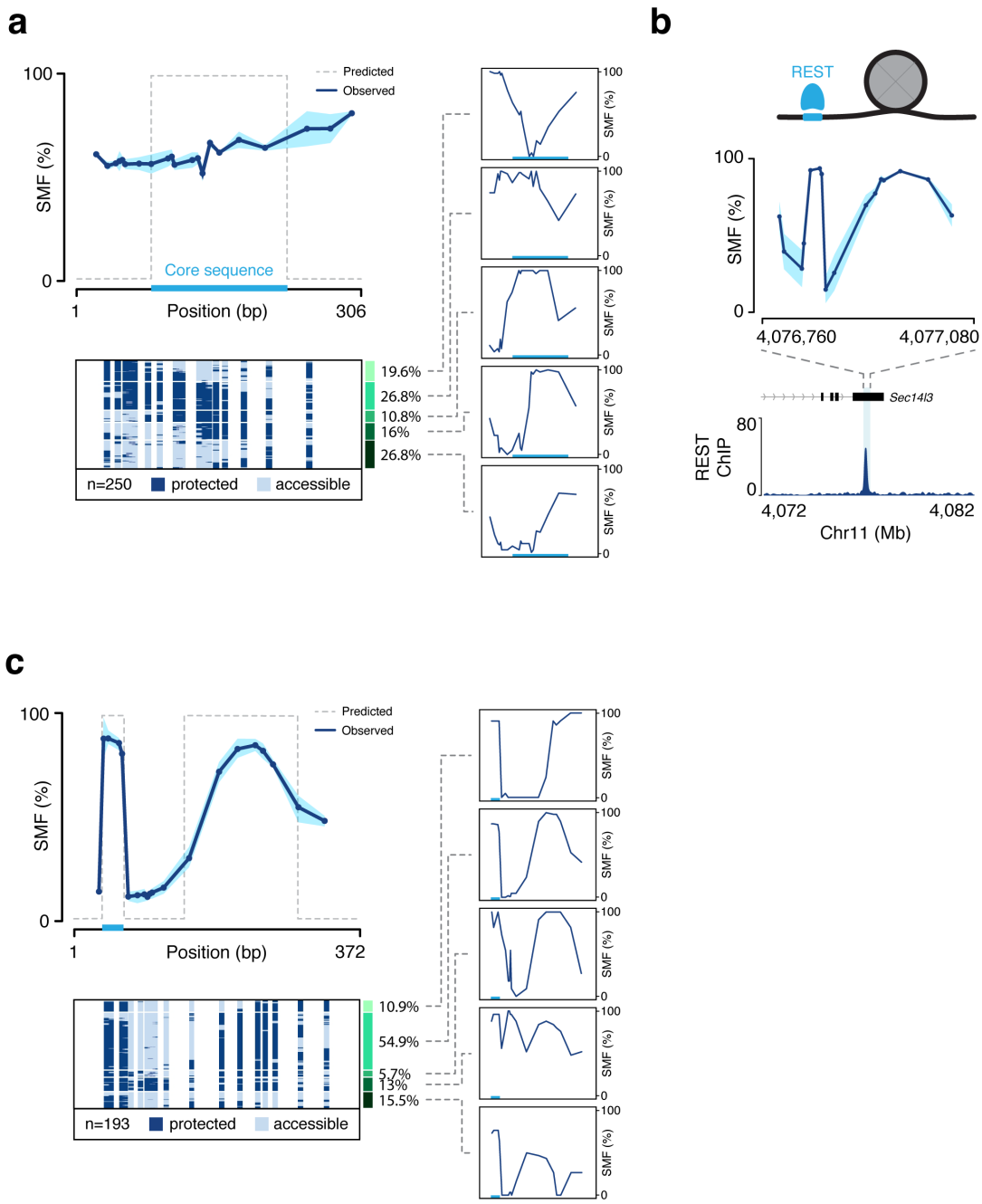

a
100
SMF (%)
0
Predicted
Observed
Core sequence
1
Position (bp)
306
n=250
protected
accessible
19.6%
26.8%
10.8%
16%
26.8%
SMF (%)
b
REST
100
SMF (%)
0
4,076,760
4,077,080
Sec14l3
REST ChIP
80
0
4,072
4,082
Chr11 (Mb)
c
100
SMF (%)
0
Predicted
Observed
1
Position (bp)
372
n=193
protected
accessible
10.9%
54.9%
5.7%
13%
15.5%

Figure 2: Nucleosome positioning *in vivo*

a) Top: SMF average profile of Widom 601. The core sequence is shown by the blue rectangle. Lines represent the mean of the replicates (n = 2) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered by k-means. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.
b) SMF average profile of an endogenous locus bound by REST. The respective ChIP-seq track is shown below, data from (Barisic et al., 2019).
c) Top: SMF average profile of REST-nucleosome construct. REST motif is represented with a blue rectangle. Lines represent the mean of the replicates (n = 3) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered by k-means. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.

Figure 3

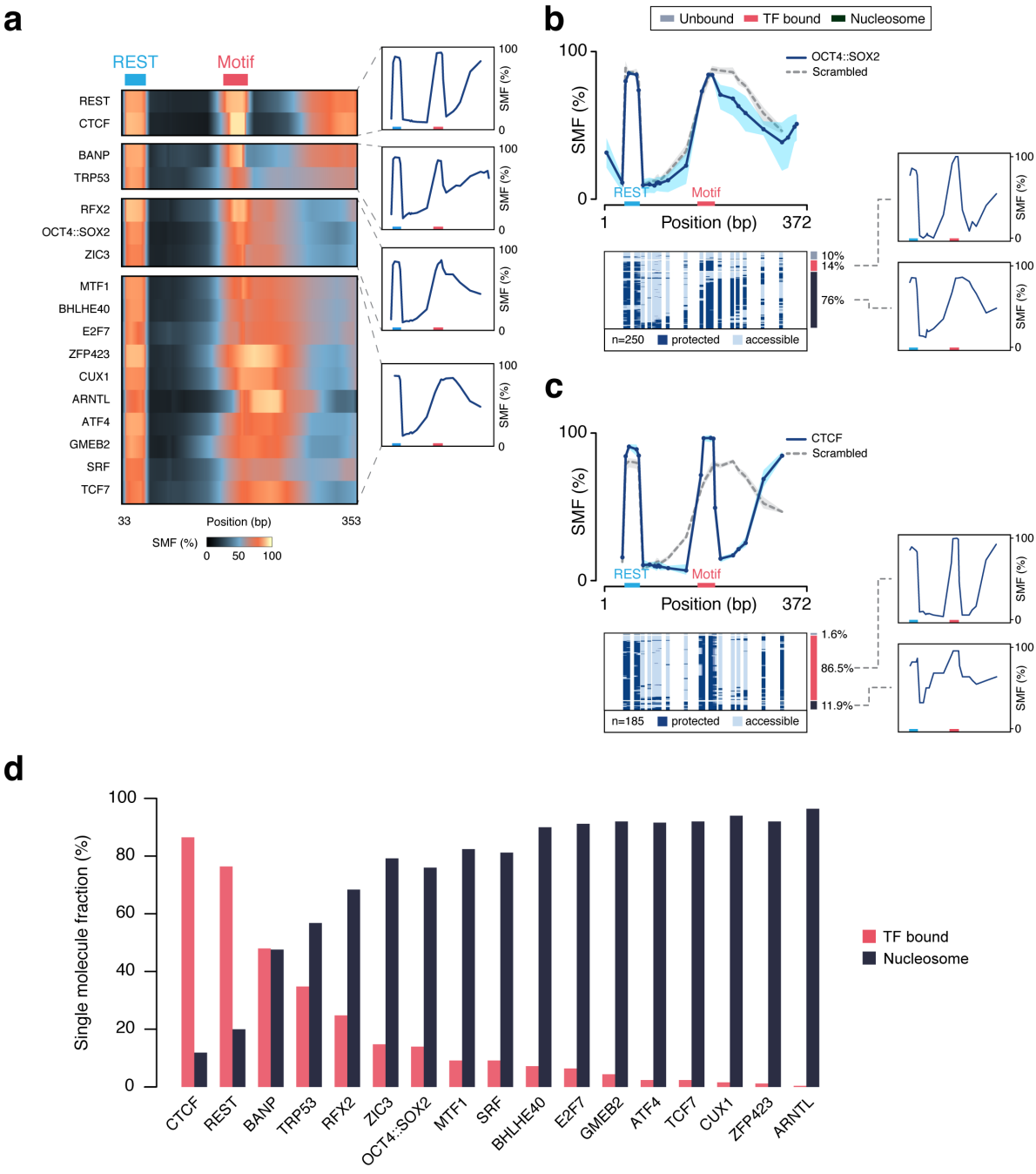

a
REST
Motif
REST
CTCF
BANP
TRP53
RFX2
OCT4::SOX2
ZIC3
MTF1
BHLHE40
E2F7
ZFP423
CUX1
ARNTL
ATF4
GMEB2
SRF
TCF7
33
Position (bp)
353
SMF (%)
0
50
100
b
Unbound
TF bound
Nucleosome
OCT4::SOX2
Scrambled
SMF (%)
REST
Motif
1
Position (bp)
372
10%
14%
76%
n=250
protected
accessible
c
CTCF
Scrambled
1.6%
86.5%
11.9%
n=185
d
Single molecule fraction (%)
TF bound
Nucleosome
CTCF
REST
BANP
TRP53
RFX2
ZIC3
OCT4::SOX2
MTF1
SRF
BHLHE40
E2F7
GMEB2
ATF4
TCF7
CUX1
ZFP423
ARNTL

Figure 3: Nucleosome-mediated obstruction of TF binding

a) SMF heatmap of selected motif inserted into the REST-nucleosome construct. Shown is the average of the replicates (n = 3). Line plots on the right-hand side of the heatmap display the average signal of the respective group. Blue rectangle: REST positioning motif; red rectangle: inserted motif.
b) Top: SMF average profile over OCT4::SOX2 and OCT4::SOX2 scrambled motif in the REST-nucleosome construct. Lines represent the mean of the replicates (n = 3) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered as described in methods. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.
c) Top: SMF average profile over CTCF and CTCF scrambled motif in the REST-nucleosome construct. Lines represent the mean of the replicates (n = 3) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered as described in methods. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.
d) Quantification of the TF bound and nucleosome single molecule fractions for the indicated motifs inserted into the REST-nucleosome construct.

Figure 4

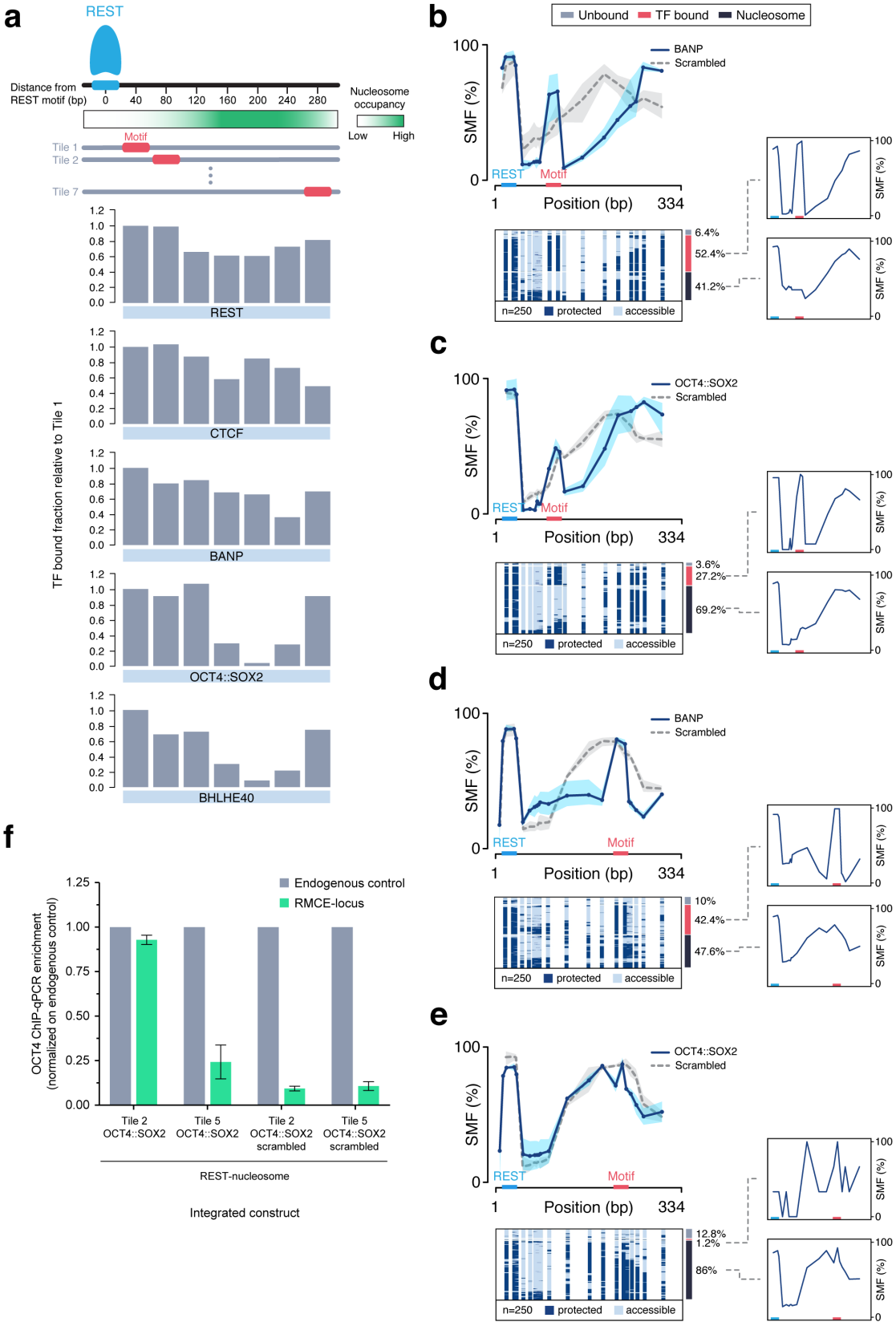
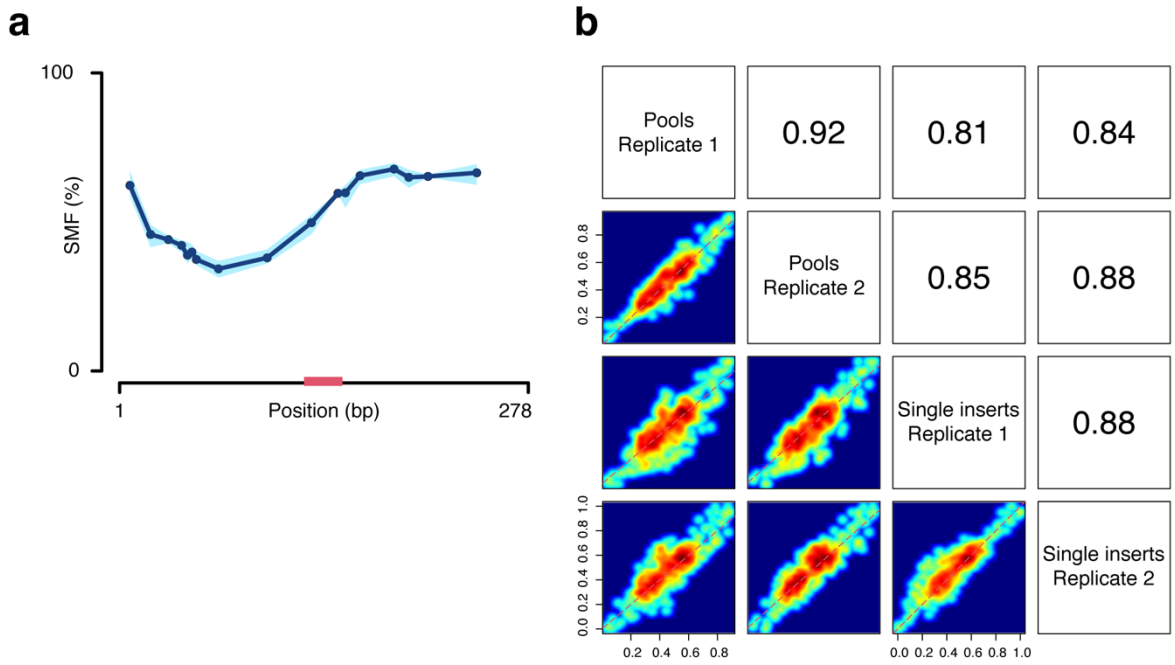

a
REST
Distance from REST motif (bp)
0 40 80 120 160 200 240 280
Nucleosome occupancy
Low High
Motif
Tile 1
Tile 2
Tile 7
TF bound fraction relative to Tile 1
REST
CTCF
BANP
OCT4::SOX2
BHLHE40
b
Unbound
TF bound
Nucleosome
BANP
Scrambled
SMF (%)
REST
Motif
Position (bp)
1
334
6.4%
52.4%
41.2%
n=250
protected
accessible
c
OCT4::SOX2
Scrambled
3.6%
27.2%
69.2%
d
BANP
Scrambled
10%
42.4%
47.6%
e
OCT4::SOX2
Scrambled
12.8%
1.2%
86%
f
Endogenous control
RMCE-locus
OCT4 ChIP-qPCR enrichment (normalized on endogenous control)
Tile 2 OCT4::SOX2
Tile 5 OCT4::SOX2
Tile 2 OCT4::SOX2 scrambled
Tile 5 OCT4::SOX2 scrambled
REST-nucleosome
Integrated construct

Figure 4: Effect of motif position on nucleosomal DNA

a) Top: Schematic of motif tiling relative to the REST positioning motif and of nucleosome occupancy across the construct. Bottom: quantification of the TF bound single molecule fraction for each motif when inserted at the indicated position. Values relative to Tile 1 are shown.
b) Top: SMF average profile over BANP and BANP scrambled motif in the Tile 2 construct. Lines represent the mean of the replicates (n = 3) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered as described in methods. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.
c) Top: SMF average profile over OCT4::SOX2 and OCT4::SOX2 scrambled motif in the Tile 2 construct. Lines represent the mean of the replicates (n = 3) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered as described in methods. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.
d) Top: SMF average profile over BANP and BANP scrambled motif in the Tile 5 construct. Lines represent the mean of the replicates (n = 3) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered as described in methods. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.
e) Top: SMF average profile over POU5F1::SOX2 and POU5F1::SOX2 scrambled motif in the Tile 5 construct. Lines represent the mean of the replicates (n = 3) and the shades represent the minimum and maximum value at each position. Bottom: single reads heatmap clustered as described in methods. Percentage of reads in each cluster and the respective average profile are shown on the right-hand side of the cluster.
f) ChIP-qPCR of OCT4 binding at OCT4::SOX2 and OCT4::SOX2 scrambled motif inserted in tiles 2 and 5 in the REST-nucleosome construct. Signal is shown relative to the enrichment at the endogenous OCT4-bound site. Error bars show the standard deviation between replicates (n = 2).

Supplementary Figure 1

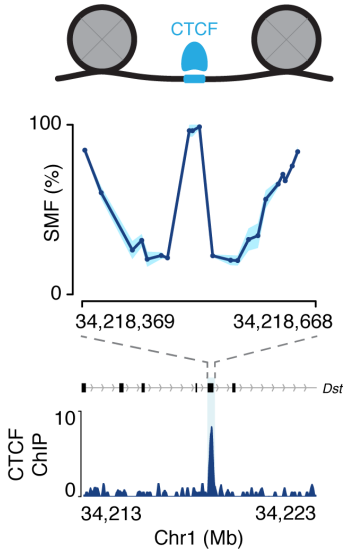


Supplementary figure 1: Generation of SMF for multiple motifs in mESCs

a) SMF average profile of the construct with unstructured nucleosome positioning. Lines represent the mean of the replicates (n = 2) and the shades represent the minimum and maximum value at each position. Red rectangle: position for motif insertion.

b) Pairwise comparison of SMF of selected motifs (n = 41) when integrated as pools or individually. Pearson correlation coefficients are indicated.

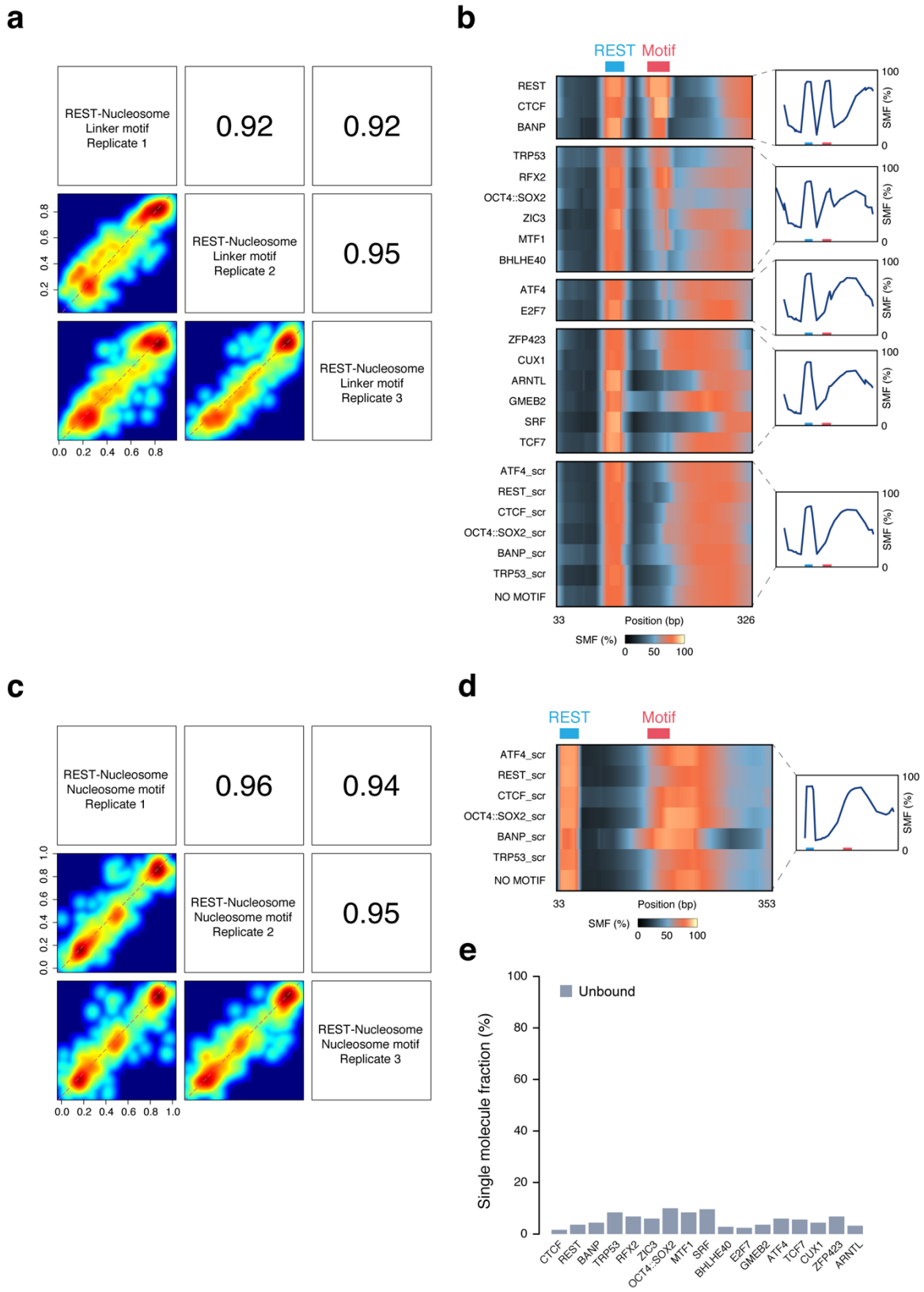
Supplementary Figure 2



Supplementary figure 2: Nucleosome positioning at endogenous TF-bound loci

SMF average profile of an endogenous locus bound by CTCF. The respective ChIP-seq track is shown below, data from (Barisic et al., 2019).

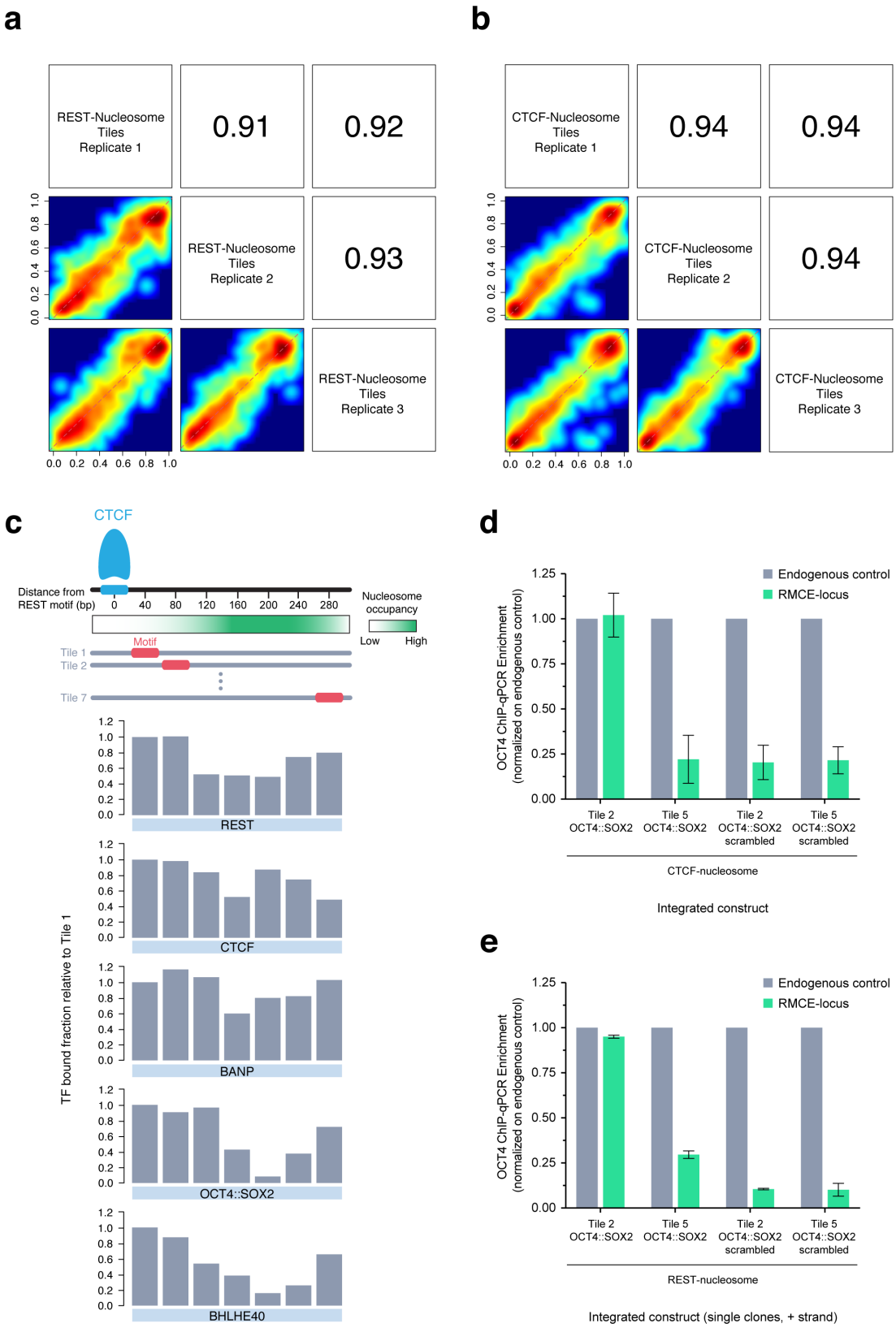
Supplementary Figure 3



Supplementary figure 3: Binding of TFs as a function of nucleosome phasing

a) Pairwise comparison of SMF of selected motifs placed in the linker DNA between REST and the nucleosome in the REST-nucleosome construct. Pearson correlation coefficients are indicated.
b) SMF heatmap of selected motifs placed in the linker in the REST-nucleosome construct. Shown is the average of the replicates (n = 3). Line plots on the right-hand side of the heatmap display the average signal of the respective group. Blue rectangle: REST positioning motif; red rectangle: inserted motif.
c) Pairwise comparison of SMF of selected motifs placed on nucleosomal DNA in the REST-nucleosome construct. Pearson correlation coefficients are indicated.
d) SMF heatmap of scrambled motifs placed on nucleosomal DNA in the REST-nucleosome construct. Shown is the average of the replicates (n = 3). Line plots on the right-hand side of the heatmap display the average signal of the respective group. Blue rectangle: REST positioning motif; red rectangle: inserted motif.
e) Quantification of the unbound single molecule fractions for the indicated motifs inserted into the REST-nucleosome construct.

Supplementary Figure 4



Supplementary figure 4: Effect of motif position along nucleosomal DNA on TF binding

a) Pairwise comparison of SMF of selected motifs tiled across the REST-nucleosome construct. Pearson correlation coefficients are indicated.
b) Pairwise comparison of SMF of selected motifs tiled across the CTCF-nucleosome construct. Pearson correlation coefficients are indicated.
c) Top: Schematic of motif tiling relative to the CTCF positioning motif and of nucleosome occupancy across the construct. Bottom: quantification of the TF bound single molecule fraction for each motif when inserted at the indicated position. Values relative to Tile 1 are shown.
d) ChIP-qPCR of OCT4 binding at OCT4::SOX2 and OCT4::SOX2 scrambled motif inserted in tiles 2 and 5 in the CTCF-nucleosome construct. Signal is shown relative to the enrichment at the endogenous OCT4-bound site. Error bars show the standard deviation between replicates (n = 2).
e) ChIP-qPCR of OCT4 binding at OCT4::SOX2 and OCT4::SOX2 scrambled motif inserted in tiles 2 and 5 in the REST-nucleosome construct in single clones carrying the integrated construct on the plus strand. Signal is shown relative to the enrichment at the endogenous OCT4-bound site. Error bars show the standard deviation between clones (n = 2).

Supplementary Table 1

TF	log2RPKM	Motif	Length	Class	Family
Arid3b	3.6	ATATTAATTAA	11	ARID	ARID-related
Arnt	4.0	CACGTG	6	Basic_helix-loop-helix_factors_(bHLH)	PAS_domain_factors
Arntl	2.8	GGTCACGTGC	10	Basic_helix-loop-helix_factors_(bHLH)	PAS_domain_factors
Atf1	6.2	ATGACGTA	8	Basic_leucine_zipper_factors_(bZIP)	CREB-related_factors
Atf4	7.3	GGATGATGCAATA	13	Basic_leucine_zipper_factors_(bZIP)	ATF-4-related_factors
Banp	1.5	CTCGCGAGAGT	11	Ben-domain_containing	-
Bhlhe40	4.0	ATCACGTGAC	10	Basic_helix-loop-helix_factors_(bHLH)	Hairy-related_factors
Cebpg	3.9	ATTGCGCAAT	10	Basic_leucine_zipper_factors_(bZIP)	CEBP-related
Clock	3.0	AACACGTGTT	10	Basic_helix-loop-helix_factors_(bHLH)	PAS_domain_factors
Creb1	5.4	TGACGTCA	8	Basic_leucine_zipper_factors_(bZIP)	CREB-related_factors
Creb3	3.0	GTGCCACGTCATCA	14	Basic_leucine_zipper_factors_(bZIP)	CREB-related_factors
Ctcf	5.3	TGGCCACCAGGGGGCGCTA	19	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers
Cux1	4.2	TAATCGATAA	10	Homeo_domain_factors	HD-CUT
E2f3	2.9	AAAAATGGCGCCATTTTTT	18	Fork_head/winged_helix_factors	E2F
E2f4	6.1	GGGCGGGAAGG	11	Fork_head/winged_helix_factors	E2F
E2f7	2.7	TTTTCCCGCCAAAA	14	Fork_head/winged_helix_factors	E2F
E2f8	4.0	TTTCCCGCCAAA	12	Fork_head/winged_helix_factors	E2F
Elf3	4.0	AACCCGGAAGTAA	13	Tryptophan_cluster_factors	Ets-related
Elk4	2.6	CCACTTCCGGC	11	Tryptophan_cluster_factors	Ets-related
Erf	4.5	ACCGGAAGTG	10	Tryptophan_cluster_factors	Ets-related
Esrrb	6.8	TCAAGGTCATA	11	Nuclear_receptors_with_C4_zinc_fingers	Steroid_hormone_receptors_(NR3)
Etv4	4.2	ACCGGAAGTA	10	Tryptophan_cluster_factors	Ets-related
Etv5	6.0	ACCGGAAGTG	10	Tryptophan_cluster_factors	Ets-related
Etv6	3.3	AGCGGAAGTG	10	Tryptophan_cluster_factors	Ets-related
Foxh1	2.8	TCCAATCCACA	11	Fork_head/winged_helix_factors	FOX
Foxo3	3.1	GTAAACAA	8	Fork_head/winged_helix_factors	FOX
Foxp1	4.8	CAAAAGTAAACAAAG	15	Fork_head/winged_helix_factors	FOX
Gabpa	5.8	CCGGAAGTGGC	11	Tryptophan_cluster_factors	Ets-related
Gbx2	3.6	ACCAATTAGC	10	Homeo_domain_factors	HOX
Gli2	4.6	GCGACCACACTG	12	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers
Gmeb1	4.0	GAGTGTACGTAAGATGG	17	SAND_domain_factors	GMEB
Gmeb2	3.1	TTACGTAA	8	SAND_domain_factors	GMEB
Hes1	4.1	GGCACGCGTC	10	Basic_helix-loop-helix_factors_(bHLH)	Hairy-related_factors
Hic2	3.7	ATGCCCACC	9	C2H2_zinc_finger_factors	Factors_with_multiple_dispersed_zinc_fingers
Hltf	4.2	AACCTTATAT	10	Tryptophan_cluster_factors	Myb/SANT_domain_factors
Hmbox1	2.7	ACTAGTTAAC	10	Homeo_domain_factors	POU_domain_factors
Hsf2	4.6	TTCTAGAACGTTC	13	Heat_shock_factors	HSF_factors
Irf1	4.0	TTTTACTTTCACTTTCACTTT	21	Tryptophan_cluster_factors	Interferon-regulatory_factors
Junb	2.6	GGATGACTCAT	11	Basic_leucine_zipper_factors_(bZIP)	Jun-related
Jund	4.2	GGTGACTCATC	11	Basic_leucine_zipper_factors_(bZIP)	Jun-related
Klf4	4.8	TGGGTGGGGC	10	C2H2_zinc_finger_factors	Three-zinc_finger_Kruppel-related
Klf5	4.8	GCCCCGCCCC	10	C2H2_zinc_finger_factors	Three-zinc_finger_Kruppel-related
Lin54	3.7	ATTTGAATT	9	CRC_domain	-
Mafg	3.3	AAATTGCTGAGTCAGCATATT	21	Basic_leucine_zipper_factors_(bZIP)	Maf-related
Max	4.5	ACCACGTGCT	10	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Max::Myc	NA	GACCACGTGGT	11	Basic_helix-loop-helix_factors_(bHLH)::Basic_helix-loop-helix_factors_(bHLH)_	bHLH-ZIP_factors::bHLH-ZIP_factors_
Mef2d	2.8	ACTATAAATAGA	12	MADS_box_factors	Regulators_of_differentiation
Mga	5.4	AGGTGTGA	8	T-Box_factors	TBX6-related_factors
Mlx	4.6	ATCACGTGAT	10	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Mlxip	2.4	ATCACGTGAT	10	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Mtf1	3.6	TTTGCACACGGCAC	14	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers
Mybl1	3.2	ACCGTTAACGGT	12	Tryptophan_cluster_factors	Myb/SANT_domain_factors
Mybl2	7.2	AACCGTTAAACGGTC	15	Tryptophan_cluster_factors	Myb/SANT_domain_factors
Myc	2.5	CCATGTGCTT	10	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Mycn	5.6	GCCACGTG	8	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Nfat5	3.8	ATTTTCCATT	10	Rel_homology_region_(RHR)_factors	NFAT-related_factors
Nfe2l2	5.0	CAGCATGACTCAGCA	15	Basic_leucine_zipper_factors_(bZIP)	Jun-related
Nfkb1	3.3	AGGGGAATCCCCT	13	Rel_homology_region_(RHR)_factors	NF-kappaB-related_factors
Nfya	2.7	AGAGTGCTGATTGGTCCA	18	Other_alpha	NFY
Nfyb	6.9	AAATGGACCAATCAG	15	Heteromeric_CCAAT-binding_factors	Heteromeric_CCAAT-binding
Nr2c2	3.1	AGGGGTCAGAGGTCA	15	Nuclear_receptors_with_C4_zinc_fingers	RXR-related_receptors_(NR2)
Nr5a2	4.6	AAGTTCAAGGTCAGC	15	Nuclear_receptors_with_C4_zinc_fingers	FTZF1related(NR5A)
Nrf1	5.6	GCGCCTGCGCA	11	Basic_leucine_zipper_factors_(bZIP)	Jun-related
Oct4/Pou5f1	10.9	TATGCAAAT	9	Homeo_domain_factors	POU_domain_factors
Otx2	3.4	TTAATCCT	8	Homeo_domain_factors	Paired-related_HD_factors
Pax6	3.2	TTCACGCATGAGTT	14	Paired_box_factors	Paired_plus_homeo_domain
Pknox1	2.8	TGACAGGTGTCA	12	Homeo_domain_factors	TALE-type_homeo_domain_factors
Pou2f1	4.7	AATATGCAAATT	12	Homeo_domain_factors	POU_domain_factors
Pou5f1::Sox2	NA	CTTTGTTATGCAAAT	15	Homeo_domain_factors::High-mobility_group_(HMG)_domain_factors	POU_domain_factors::SOX-related_factors
Rara	3.6	GAGGTCAAAAGGTCAATG	18	Nuclear_receptors_with_C4_zinc_fingers	Thyroid_hormone_receptor-related_factors_(NR1)
Rara::Rxra	NA	AGGTCACGGAGAGGTCA	17	Nuclear_receptors_with_C4_zinc_fingers::Nuclear_receptors_with_C4_zinc_fingers	Thyroid_hormone_receptor-related_factors_(NR1)::RXR-related_receptors_(NR2)
Rarg	5.7	AAGGTCAAAAGGTCAA	16	Nuclear_receptors_with_C4_zinc_fingers	Thyroid_hormone_receptor-related_factors_(NR1)
Rest	6.7	TTCAGCACCATGGACAGCGCC	21	C2H2_zinc_finger_factors	Factors_with_multiple_dispersed_zinc_fingers
Rfx2	2.9	CGTTGCCATGGCAACG	16	Fork_head/winged_helix_factors	RFX-related_factors
Rreb1	2.8	CCCCAAACCACCCCCCCCCA	20	C2H2_zinc_finger_factors	Factors_with_multiple_dispersed_zinc_fingers
Rxra	1.7	GGGGTCAAAGGTCA	14	Nuclear_receptors_with_C4_zinc_fingers	RXR-related_receptors_(NR2)
Rxra::Vdr	NA	GGGTCAACGAGTTCA	15	Nuclear_receptors_with_C4_zinc_fingers::Nuclear_receptors_with_C4_zinc_fingers	RXR-related_receptors_(NR2)::Thyroid_hormone_receptor-related_factors_(NR1)
Rxrb	3.3	GGGGTCAAAGGTCA	14	Nuclear_receptors_with_C4_zinc_fingers	RXR-related_receptors_(NR2)
Smad3	2.5	CGTCTAGACA	10	SMAD/NF-1_DNA-binding_domain_factors	SMAD_factors
Sox2	7.5	CCTTTGTT	8	High-mobility_group_(HMG)_domain_factors	SOX-related_factors
Sp1	5.6	GCCCCGCCCCC	11	C2H2_zinc_finger_factors	Three-zinc_finger_Kruppel-related
Sp2	2.5	GCCCCGCCCCCTCCC	15	C2H2_zinc_finger_factors	Three-zinc_finger_Kruppel-related
Sp3	5.3	GCCACGCCCCC	11	C2H2_zinc_finger_factors	Three-zinc_finger_Kruppel-related
Srebf1	2.6	ATCACCCCAC	10	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Srebf2	3.7	ATGGGGTGAT	10	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Srf	2.6	TGACCATATATGGTCA	16	MADS_box_factors	Responders_to_external_signals_(SRF/RLM1)
Stat3	4.4	CTTCTGGGAAA	11	STAT_domain_factors	STAT_factors
Stat4	3.0	TTTCCAGGAAATGG	14	STAT_domain_factors	STAT_factors
Stat6	3.6	CATTTCCTGAGAAAT	15	STAT_domain_factors	STAT_factors
Tbp	5.6	GTATAAAAGGCGGGG	15	TATA-binding_proteins	TBP-related_factors
Tcf12	2.5	AACAGCTGCAG	11	Basic_helix-loop-helix_factors_(bHLH)	E2A
Tcf3	5.0	AACACCTGCT	10	Basic_helix-loop-helix_factors_(bHLH)	E2A
Tcf4	2.6	CGCACCTGCT	10	Basic_helix-loop-helix_factors_(bHLH)	E2A
Tcf7	3.0	AAAGATCAAAGG	12	High-mobility_group_(HMG)_domain_factors	TCF-7-related_factors
Tcfl5	3.5	GGCACGTGCC	10	Basic_helix-loop-helix_factors_(bHLH)	PAS_domain_factors
Tead1	4.7	CACATTCCAT	10	TEA_domain_factors	TEF-1-related_factors
Tfap2c	4.4	TGCCCCAGGGCA	12	Basic_helix-span-helix_factors_(bHSH)	AP-2
Tfap4	2.9	AACAGCTGAT	10	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Tgif1	6.9	TGACAGCTGTCA	12	Homeo_domain_factors	TALE-type_homeo_domain_factors
Thap1	3.4	CTGCCCGCA	9	C2CH_THAP-type_zinc_finger_factors	THAP-related_factors
Trp53	6.8	AACATGCCCGGGCATGTC	18	p53_domain_factors	p53-related_factors
Usf1	5.3	GCCACGTGACC	11	Basic_helix-loop-helix_factors_(bHLH)	bHLH-ZIP
Vdr	-2.7	GAGTTCATTGAGTTCA	16	Nuclear_receptors_with_C4_zinc_fingers	Thyroid_hormone_receptor-related_factors_(NR1)
Xbp1	4.5	AATGCCACGTCATC	14	Basic_leucine_zipper_factors_(bZIP)	XBP-1-related_factors
Yy1	5.2	CAAGATGGCGGC	12	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers
Zfp143	4.5	TACCCACAATGCATTG	16	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers
Zfp263	3.2	GGAGGAGGAGGGGGAGGAGGA	21	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers
Zfp423	3.2	GGCACCCAGGGGTGC	15	C2H2_zinc_finger_factors	Factors_with_multiple_dispersed_zinc_fingers
Zfp740	5.3	CCCCCCCCAC	10	C2H2_zinc_finger_factors	Other_factors_with_up_to_three_adjacent_zinc_fingers
Zfx	3.4	GGGGCCGAGGCCTG	14	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers
Zic3	4.8	GACCCCCCGCTGCGC	15	C2H2_zinc_finger_factors	More_than_3_adjacent_zinc_fingers

Methods

Cell culture and RMCE insertion

Mouse ES cells, TC-1 line, carrying an RMCE cassette in the β-globin locus (Lienert et al., 2011) were used for all experiments and cultured as described in (Mohn et al., 2008). Briefly, plates were coated with 0.2% gelatin (Sigma) and cells were maintained in Dulbecco's Modified Eagle Medium (Invitrogen), supplemented with 15% Fetal Bovine Serum (Invitrogen), Glutamax (Gibco) and Non-essential amino acids (Gibco), betamercaptoethanol (Sigma) and leukemia inhibitory factor (LIF; produced in-house). No synchronization was performed before any of the experiments. RMCE insertion was performed as previously described (Feng et al., 1999; Lienert et al., 2011) In brief, for individual construct insertion, 250'000 cells were transfected with 1ug of L1-insert-1L vector and 500ng of pIC-Cre using Lipofectamine 3000 Reagent (Thermo Fisher Scientific) following manufacturer protocol. For pooled insertions, 4 million cells were electroporated (Amaxa Nucleofection, Lonza) with 25ug of L1-insert-1L vector pool (up to 8 different constructs) and 15ug of pIC-Cre. In both cases, negative selection with Ganciclovir (Roche) at a final concentration of 3uM was carried out 2 days after transfection and for a total of 10 days.

Selection of transcription factors motifs and generation of constructs

Transcription factors expressed in mESCs were ranked based on their mean RPKM value according to (Domcke et al., 2015). Motifs were taken from JASPAR database (Mathelier et al., 2016) by selecting the maximum PWM, followed by the addition of GCT and TGCA sequences at their 5' and 3' respectively, to maximize the possibility to detect a footprint even for factors that do not have a GpC in their motif. For BANP, the motif was not in JASPAR and the motif used in (Grand et al., 2021) was selected. For the constructs where REST and CTCF are used to position nucleosomes, two slightly different top motifs were chosen to avoid repeats in the constructs with two REST or CTCF motifs. Finally, all motifs were expanded to 28bp in length by adding random bases on either end. Scrambled motifs were obtained by randomly shuffling the motif bases while maintaining any CpG dinucleotides. The final constructs were generated by placing the 28mers at the desired position inside a previously described, in-vitro derived sequence, with low nucleosome affinity (LNA), high CpG content (O/E 1.34), which is not part of any known regulatory region (Yang et al., 2006). For RMCE,

the constructs were finally cloned into a plasmid using a multiple cloning site flanked by two inverted L1 Lox sites.

SMF

Single molecule footprinting was performed as previously described (Grand et al., 2021; Kelly et al., 2012) with some modifications. After Ganciclovir selection, 250'000 cells were collected, washed once with PBS and incubated 10min on ice in 1ml of ice-cold lysis buffer (10mM Tris pH 7.5, 10mM NaCl, 3mM $MgCl_2$, 0.1mM EDTA and 0.5% NP-40) to extract nuclei. After centrifugation (800g, 5min, 4ºC), nuclei were washed once with 250ul of ice-cold wash buffer (10mM Tris pH 7.5, 10mM NaCl, 3mM $MgCl_2$ and 0.1mM EDTA) and centrifuged again (800g, 5min, 4ºC). Samples were then resuspended in 100ul of 1x M.CviPI buffer, mixed with 200ul of GpC methyltransferase reaction mix on ice (1x M.CviPI buffer, 1mM SAM, 300nM sucrose, 200U M.CviPI: M0227L, NEB) and then incubated at 37ºC for 15min. Reaction was stopped by adding 300ul of Stop Solution (20mM Tris-HCl pH 8.0, 600mM NaCl, 1% SDS, 10mM EDTA) pre-warmed at 37ºC. To remove proteins and RNA, samples were treated first with proteinase K (200ug/ml) at 65ºC for 16h and then with 10ul of RNAse A (10mg/ml) for 30min at 37ºC. Finally, genomic DNA was extracted with phenol-chloroform purification and isopropanol precipitation as follow. Samples were mixed with 600ul of a 1:1 phenol-chloroform solution, vortexed for 1min at maximum speed and centrifuged at 10'000g for 10min. The supernatant was then transferred to new tubes and mixed with 600ul of chloroform, vortexed and centrifuged again as before. The supernatant was transferred to new tubes and mixed with 600ul of isopropanol and 1ul glycogen and incubated 10min at RT before centrifugation (max speed, 1h, 4ºC). The pellet was then washed with 500ul of ice-cold 70% ethanol and centrifuged (max speed, 15min, 4ºC). The supernatant was discarded and the pellet was dried for 15min at 37ºC and resuspended in 22ul ddH_2O overnight at RT. NanoDrop was used for quantification and 2ug of DNA were converted using the EZ DNA methylation-gold kit (Zymo).

DNA amplification and bisulfite sequencing

Target amplicons were amplified from bis-converted DNA using KAPA HiFi Uracil+ (Roche) with the following program: 95ºC, 4min; [98ºC, 20s; 60ºC, 15s, 72ºC, 20s]x35; 75ºC, 5min; 4ºC, hold. DNA was then purified with AMPure XP Beads (0.6x, Beckman

Coulter) and used as input for library preparation. Sequence libraries were prepared with the NEBNext ChIP-seq Library Prep Master Mix Set from Illumina (E6240), doing 8 cycles of amplification, and loaded on an Illumina MiSeq (500bp paired-end).

Generation of average profiles and single molecule analysis

Read mapping to the relative constructs was carried out using the Bioconductor package QuasR with default settings for bisulfite-converted templates (Gaidatzis et al., 2015) and after adaptor trimming with Trimmomatic. Methylation of Cs in the CpG and GpC contexts was then quantified using qMeth, excluding GCG CCG contexts, where endogenous and exogenous methylation cannot be distinguished, and Cs with coverage <10. Data is shown as 1-methylation on the Y axis and bp position on the X axis. TF motifs are shown in blue (nucleosome-positioning TF) or red (general TF). Single molecule analysis was performed using QuasR in combination with custom functions to cluster reads according to their GCH methylation pattern, with 0 and 1 reflecting unmethylated and methylated status respectively. For each construct, reads from all replicates were pooled and filtered by GCH coverage. Then, 250 reads were randomly sampled and used for downstream analyses. If less than 250 reads were available after filtering, all remaining reads were taken. For the Widom 601 and the REST-nucleosome constructs, single molecules were clustered using *kmeans* and setting k = 5. For all other constructs, three clusters were defined as follow. **Unbound**: >80% of GpC must be methylated; **TF bound**: average of GpC methylation at the motif must be equal to zero after rounding. This ensures that the motif is occupied. Then, at least one of the first two GpCs upstream the motif and at least one of the first two GpCs downstream the motif must be methylated. This ensures presence of linker DNA around the bound TF and accounts for different footprint sizes. When resolution or coverage were limiting (Tiles 1, 6, 7), a subset of linker GpCs was selected for each condition and maintained across constructs; **Nucleosome:** all reads that do not fall in the previous two clusters.

ChIP-qPCR

ChIP samples were prepared as previously described (Barisic et al., 2019) with following modifications: (1) Diagenode Bioruptor Pico was used to sonicate chromatin with 20 cycles, 20sec ON and 40sec OFF, (2) protein A magnetic Dynabeads Magnetic beads (Thermo Fisher Scientific) were used. For immunoprecipitation of OCT4, 70ug

of chromatin were combined with 5ug of Oct-4A (C40A3C1) antibody (Cell Signaling Technologies).

Input and immunoprecipitated materials were resuspended in 40ul and 1ul was used as input for qPCR. Amplification was carried out in 20ul reaction with the PowerUp SYBR Green Master Mix (Applied Biosystems) using the following program: 95ºC, 10min; [98ºC, 15s; 60ºC, 1min]x40, ramping 1.6 ºC/s. To measure enrichment at the RMCE locus, a primer pair encompassing the region from tile 2 to tile 5 was used and compared against a primer pair targeting a single genomic OCT4-bound Oct4::Sox2 motif (Mm10 genome build: chr8:92,740,685-92,746,111).

References

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.

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.

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., and Zaret, K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244–254.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Clark, K.L., Halay, E.D., Lai, E., and Burley, S.K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature *364*, 412–420.

D'Haeseleer, P. (2006). What are DNA sequence motifs? Nat. Biotechnol. *24*, 423–425.

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. Nat. 2015 5287583 *528*, 575–579.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nat. 2012 4897414 *489*, 57–74.

Feng, Y.Q., Seibler, J., Alami, R., Eisen, A., Westerman, K.A., Leboulch, P., Fiering, S., and Bouhassira, E.E. (1999). Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. J. Mol. Biol. *292*, 779–785.

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

Gracey, L.E., Chen, Z.Y., Maniar, J.M., Valouev, A., Sidow, A., Kay, M.A., and Fire, A.Z. (2010). An in vitro-identified high-affinity nucleosome-positioning signal is capable of transiently positioning a nucleosome in vivo. Epigenetics Chromatin *3*.

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.

Hansen, J.L., Loell, K.J., and Cohen, B.A. (2022). A test of the pioneer factor hypothesis using ectopic liver gene activation. Elife *11*.

Hsu, H.T., Chen, H.M., Yang, Z., Wang, J., Lee, N.K., Burger, A., Zaret, K., Liu, T., Levine, E., and Mango, S.E. (2015). TRANSCRIPTION. Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4. Science *348*, 1372–1376.

Inukai, S., Kock, K.H., and Bulyk, M.L. (2017). Transcription factor–DNA binding:

beyond binding site motifs. Curr. Opin. Genet. Dev. *43*, 110–119.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

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.

King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. Elife *6*.

Kornberg, R.D. (1974). Chromatin structure: A repeating unit of histones and DNA. Science (80-.). *184*, 868–871.

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.

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 MNase-seq. Nature *562*, 281.

Lancrey, A., Joubert, A., Duvernois-Berthet, E., Routhier, E., Raj, S., Thierry, A., Sigarteu, M., Ponger, L., Croquette, V., Mozziconacci, J., et al. (2022). Nucleosome Positioning on Large Tandem DNA Repeats of the '601' Sequence Engineered in Saccharomyces cerevisiae. J. Mol. Biol. *434*, 167497.

Li, Q., and Wrange, O. (1993). Translational positioning of a nucleosomal glucocorticoid response element modulates glucocorticoid receptor affinity. Genes Dev. *7*, 2471–2482.

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.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958–970.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*, D110–D115.

McDaniel, S.L., Gibson, T.J., Schulz, K.N., Fernandez Garcia, M., Nevil, M., Jain, S.U., Lewis, P.W., Zaret, K.S., and Harrison, M.M. (2019). Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. Mol. Cell *74*, 185-195.e4.

Michael, A.K., and Thomä, N.H. (2021). Reading the chromatinized genome. Cell *184*, 3599–3611.

Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker,

R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R., et al. (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science (80-.). *368*, 1460–1465.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol. Cell *30*, 755–766.

Moorman, C., Sun, L. V., Wang, J., De Wit, E., Talhout, W., Ward, L.D., Greil, F., Lu, X.J., White, K.P., Bussemaker, H.J., et al. (2006). Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. *103*, 12027–12032.

Van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J.L., Thieffry, D., Beato, M., et al. (2015). C/EBPα Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. Stem Cell Reports *5*, 232–247.

Padeken, J., Methot, S.P., and Gasser, S.M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. Nat. Rev. Mol. Cell Biol. 2022 239 *23*, 623–640.

Pardo, C.E., Darst, R.P., Nabilsi, N.H., Delmas, A.L., and Kladde, M.P. (2011). Simultaneous single-molecule mapping of protein-DNA interactions and DNA methylation by MAPit. Curr. Protoc. Mol. Biol. *Chapter 21*.

Perales, R., Zhang, L., and Bentley, D. (2011). Histone occupancy in vivo at the 601 nucleosome binding element is determined by transcriptional history. Mol. Cell. Biol. *31*, 3485–3496.

Perlmann, T., and Wrange, O. (1988). Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. EMBO J. *7*, 3073–3079.

Plasschaert, R.N., Vigneau, S., Tempera, I., Gupta, R., Maksimoska, J., Everett, L., Davuluri, R., Mamorstein, R., Lieberman, P.M., Schultz, D., et al. (2014). CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. Nucleic Acids Res. *42*, 774–789.

Schlaeger, T.M., Daheron, L., Brickler, T.R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., et al. (2014). A comparison of non-integrating reprogramming methods. Nat. Biotechnol. 2014 331 *33*, 58–63.

Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., Van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat. Biotechnol. 2013 322 *32*, 171–178.

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.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. Cell *151*, 994–1004.

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.

Subtil-Rodríguez, A., and Reyes, J.C. (2010). BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. EMBO Rep. *11*, 751–757.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. Cell *155*, 621–635.

Yang, J.G., Madrid, T.S., Sevastopoulos, E., and Narlikar, G.J. (2006). The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing. Nat. Struct. Mol. Biol. *13*, 1078–1083.

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*.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat. Struct. Mol. Biol. *16*, 847.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81.

4.2 SIN3A binds chromatin through distinct mechanisms and constantly represses transcription

4.2.1 Introduction

TFs and cofactors cooperate to shape gene expression and binding specificity is key to ensure proper localization and activity. TFs achieve specificity by recognizing their cognate motif on DNA and get further restricted by specific chromatin features such as nucleosomes as we have shown in the previous work. In contrast, cofactors form large complexes which do not engage with consensus sequences, relying instead on much more promiscuous interactions to access chromatin. For example, the same cofactor might interact with many different TFs, but also multiple cofactors might interact with the same single TF, generating complex networks that eventually result in different degrees of specificity. These characteristics make cofactors inherently hard to study and a clear understanding of how they gain specificity on chromatin is still lacking.

In this part of the thesis, we focus on the corepressor SIN3A, to investigate its functions as well as its specific recruitment to chromatin. SIN3A is a scaffold protein with no known catalytic function that contains four paired amphipathic helix domains, a HDAC interacting domain (HID) and a highly conserved region (HCR), which allow interaction with numerous cofactors and TFs (Banks et al., 2020; Chandru et al., 2018; Pang et al., 2003; Silverstein and Ekwall, 2005; Viiri et al., 2006). In particular, it is part, together with other subunits, of a bigger complex containing HDAC1 and HDAC2 named SIN3A/HDAC complex (Adams et al., 2018). In mammals, another isoform named SIN3B is also expressed, which shares more than 50% homology with SIN3A across the full protein length with the highest conservation observed at its PAH domains (Ayer et al., 1995). However, its presence is not able to compensate the loss of SIN3A, suggesting a different role of SIN3B in the complex (van Oevelen et al., 2010). The SIN3A/HDAC complex has been shown to be important for deacetylating chromatin and, in turn, for silencing gene expression, with studies showing presence of SIN3A at silent genes or recruitment at active genes and impaired silencing upon its removal (Ayer et al., 1995; Chrysanthou et al., 2022; Huang et al., 1999; Murphy et al., 1999; Nagy et al., 1997). However, other works have also shown impaired gene activation after SIN3A downregulation, thereby proposing a context-dependent activity (Feng et al., 2022; Saha et al., 2016; Zhu et al., 2018b). Although such hypothesis is

appealing, we lack a comprehensive understanding of SIN3A binding specificity and activity, which prevents us to rule out possible confounding factors. Indeed, some major limitations prevent an in-depth analysis of the role of SIN3A on chromatin and gene regulation. Firstly, SIN3A is a broad essential protein, therefore, the only loss of function approacwhes used to date are RNA interference (RNAi) and conditional knockout (cKO). While RNAi and cKO are convenient tools for studying essential genes, they only provide little time resolution and do not enable detection of early events. Here, the onset of secondary effects at later time points might generate confounders, potentially explaining the contrasting results between studies. Secondly, SIN3A occupancy has been often measured via ChIP-qPCR assays, which only allows analysis of few candidate loci at a time. Moreover, commercially available antibodies do not yield high and reproducible SIN3A enrichment in ChIP-seq assays, leading to limited reproducibility across datasets (Williams et al., 2011). Thus, a full map of SIN3A binding across the genome is still lacking. Finally, many TFs have been proposed as candidate recruiters of the SIN3A/HDAC complex, but due to the above-mentioned limitations, a comprehensive understanding of their genome-wide interaction with SIN3A is missing and the effect of such interactions on chromatin state and transcriptional activity remains unclear.

In this study, we took advantage of an inducible degron tag (Nabet et al., 2018) to deplete SIN3A in mouse embryonic stem cells (mESCs) in a temporal-controlled fashion and we characterized its localization across the genome, the impact on transcription and histone acetylation, as well as the mechanisms driving its recruitment. We found that SIN3A was preferentially localized at the TSS of active promoters, where it colocalized with H3K4me3 positive nucleosomes, and only to a minor extent at silent genes and distal regions. SIN3A was also present at REST bound motifs and we demonstrated that its binding was fully REST-dependent, suggesting a different recruitment mechanism compared to TSSs. At bound sites, SIN3A removal led to an increase of H3K27 acetylation that scaled with SIN3A enrichment, suggesting that SIN3A is required to maintain an adequate level of histone acetylation at its targets. At promoters, depletion of SIN3A caused mostly transcriptional upregulation, while downregulation occurred primarily at genes with low- or no SIN3A binding.

Taken together, these results support a model in which SIN3A is selectively recruited on H3K4me3-positive nucleosomes and at TF binding sites to buffer histone acetylation and gene expression predominantly via transcriptional repression.

4.2.2 Results

4.2.2.1 SIN3A endogenous tagging allows acute depletion and reproducible detection by ChIP-seq

Previous attempts to deplete SIN3A in cultured cells and in murine models proved the essential role of this protein for cell survival (Cowley et al., 2005; McDonel et al., 2012), ruling out the possibility to study its functions by genetic deletion. To overcome this limitation, we took advantage of a targeted protein degradation system and inserted an inducible degron tag at the C-terminus of the endogenous *Sin3a* gene in mouse embryonic stem cells (mESCs) using CRISPR/Cas9 **(Fig. 1a)**. Cells carrying homozygous knock-in of the tag were viable and showed no significant changes in cell cycle **(Fig. 1b-c)**. Addition of the degrading compound (dTAG-13) resulted in significant SIN3A depletion already after 30 minutes, which reached its maximum after 1 hour as detected by western blot **(Fig. 1d)**. To test whether such degradation was sufficient to recapitulate a SIN3A knockout phenotype, we treated cells for 3 days and followed their growth and survival over time. While no changes were detected until 6 hours post treatment, cell death became visible starting at 16-hour time point and by day 3 all cells were dead **(Fig. 1e)**. Finally, analysis of the gene expression profile measured by RNA-seq revealed that the SIN3A knockin line was very similar to the parental line, while treatment with dTAG for 6h decreased such correlation **(Fig. 1f)**. Together, these data show that this cell line can be used as a wild type SIN3A background and that, after treatment, it phenocopies a SIN3A full knockout.

The endogenous tagging of SIN3A allowed us to introduce also two HA tags, small viral peptides that can be recognized with high specificity by targeted antibodies and that can improve pulldown efficiency in assays such as ChIP-seq. To test this, we performed SIN3A ChIP-seq and looked at its enrichment and reproducibility across the genome. First, we called peaks in each individual replicate (n = 3), obtaining comparable numbers **(Fig. 2a)** and a reproducible distribution of peaks width **(Fig. 2b)**. Interestingly, we also saw a consistent difference in width between peaks residing within 1kb from a TSS and peaks farther away **(Fig. 2c)**, potentially reflecting context-dependent SIN3A binding mechanisms. Then, we merged the peaks sets and filtered out those with no enrichment in one or more replicates (ChIP enrichment threshold ≥ 2-fold), obtaining a final set consisting of 25,967 peaks. Analysis of enrichment at these sites showed high reproducibility among replicates ($r \geq 0.85$) as well as good

dynamic range **(Fig. 2d)**, with binding detected also at previously identified genomic loci (e.g. *Pou5f1/Oct4*). Moreover, the binding was lost upon treatment with dTAG-13 **(Fig. 2e)**, confirming the high specificity of our SIN3A ChIP-seq dataset. Taken together, this data show that endogenous tagging of SIN3A enables highly-efficient pulldown and detection of SIN3A binding at the genomic scale.

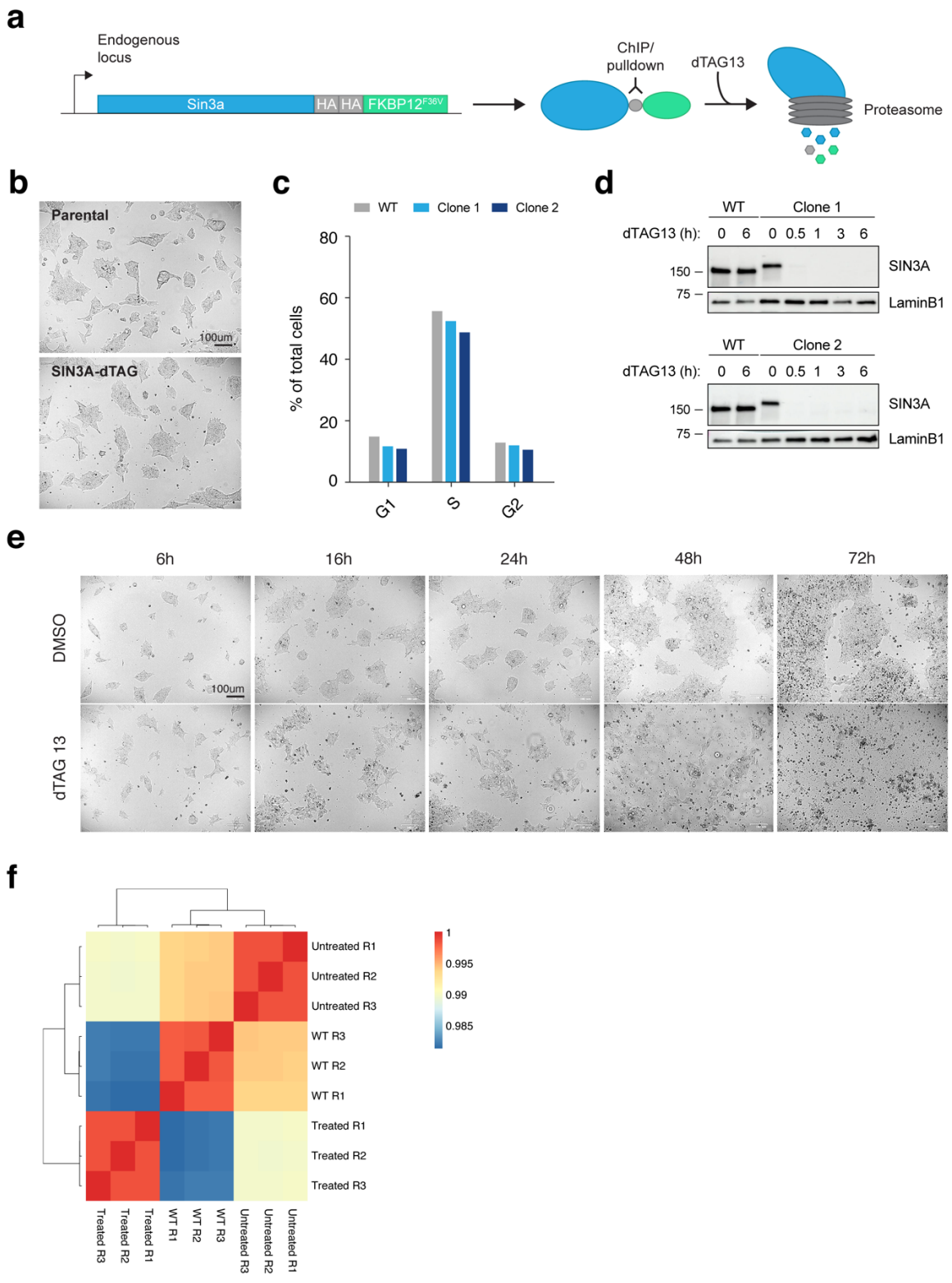


Figure 1: Acute depletion of Sin3a by dTAG inducible system
a) Schematic of the endogenous knockin strategy and of the dTAG degron system.
b) Light microscopy comparison of the Sin3a-dTAG cells with the parental line during standard culture condition. Scale bar=100um
c) FACS analysis of cells pulsed with BrdU. The Y-axis represents the percentage of all analyzed cells.
d) Western blot analysis of Sin3a levels after dTAG treatment at different time points. Unspecific band was used as loading control.
e) Light microscopy time-course of Sin3a-dTAG cells treated with either DMSO or dTAG13. Scale bar=100um
f) Correlation heatmap of expression of all genes measured by RNA-seq. Values represent the Pearson's correlation coefficient.

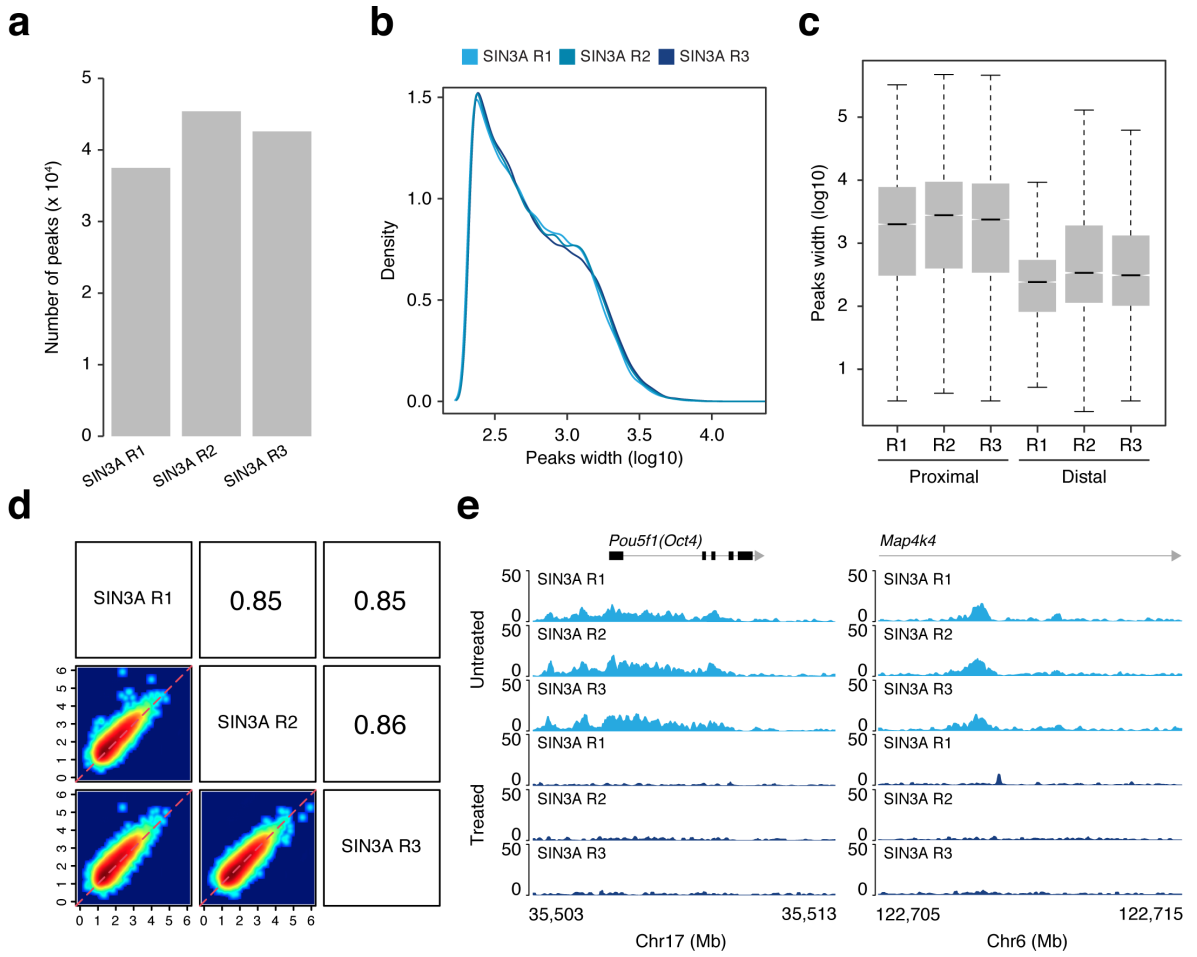


Figure 2: Endogenous tagging enables highly-efficient and reproducible ChIP enrichment genome wide
a) Number of peaks called with MACS2 in the Sin3a ChIP for each individual replicate.
b) Distribution of Sin3a peaks width in each individual replicate.
c) Width of Sin3a proximal (within 1kb from nearest TSS) and distal peaks (farther than 1kb from any TSS) for each individual replicate. Black lines represent the median; boxes represent first and third quartile; whiskers represent maximum and minimum values of distribution after removal of outliers (see methods).
d) Correlation between replicates of Sin3a ChIP-seq enrichment in the unified set of peaks.
e) Examples of individual loci bound by Sin3a with the indicated ChIP-seq tracks. Untreated = DMSO, Treated = dTAG-13 for 6h.

4.2.2.2 SIN3A binds the vast majority of active genes and drives transcriptional repression

Given the numerous studies reporting SIN3A binding at specific promoters and interaction with different TFs, we wanted to check whether SIN3A had specific binding preferences across the genome. First, we compared the distribution of SIN3A ChIP-seq peaks to the distribution of an equal number of random genomic regions, revealing that SIN3A peaks were specifically localized in regions closer than 1kb to TSSs **(Fig. 3a)**. However, peak distribution alone might be misleading, as it does not account for differences in protein enrichment within the peaks. Therefore, to test the relevance of the observed distribution, we looked at SIN3A enrichment as a function of the distance from the nearest TSS. We found that SIN3A binding was >2-fold higher in regions within 1kb from a TSS compared to regions farther away, indicative of specific binding of the complex at proximal sites across the genome **(Fig. 3b)**. Despite its known corepressor nature, the SIN3A/HDAC complex has been also found at promoters of active genes, where it has been proposed to act either as a repressor, or as a coactivator in a context-dependent fashion. While this hypothesis is definitely fascinating, data on SIN3A-mediated transcriptional changes derive mostly from knockdown (RNA interference) and conditional knockout approaches. These are inherently limited by late observational time points, thereby allowing potential secondary effects to kick in, and sometimes by incomplete depletion of the target protein. To gain more insights into the role of SIN3A binding at promoters, we combined ChIP-seq and RNA-seq and characterized the status of such promoters in presence or absence of SIN3A using our rapid degradation system. Strikingly, we found that SIN3A was bound to more than 90% of active promoters, while only 30% of silent genes showed binding at their promoters **(Fig. 3c)**. This suggests that, despite being a corepressor, SIN3A might be broadly required at actively transcribed genes and specifically recruited at a subset of silent genes. To better understand the functional relevance of SIN3A at these promoters, we rapidly degraded SIN3A and took advantage of the exon-intron split analysis (EISA) to detect SIN3A primary targets. This method uses RNA-seq data as input and allows separate analysis of exonic and intronic reads, which reflect mature and nascent RNA respectively. Indeed, mature RNA needs time to show changes occurred at the transcriptional level, whereas nascent RNA is more sensitive to such perturbations and can be used to detect early events. Before the analysis, we filtered genes (23,757, see methods)

based on their intron coverage (minimum of 10 reads) obtaining a final set of 8,319 genes which were then used as input for EISA. First, by looking at the top 1000 SIN3A bound genes, we confirmed that EISA allowed detection of early transcriptional changes, with intronic signal showing upregulation as early as 0.5h post treatment **(Fig. 3d,e)**. In contrast, exon signal reached similar levels only at 6h. Thus, we decided to perform all the subsequent analyses using intronic reads. To better evaluate the impact of SIN3A depletion over time, we monitored transcriptional changes at different time points, this time including all of 8,319 genes. Surprisingly, we noticed that changes occurred very early post treatment (i.e. 0.5 and 1h) were gradually lost or reverted over time, and at 24h the differentially expressed genes showed a profile that differed profoundly from those at 0.5h and 1h **(Fig. 3f)**. This finding reveals that gene expression changes occurring at late time points do not always reflect the true activity of SIN3A, highlighting the need to measure early events in order to avoid confounders. Moreover, at 0.5h, genes were more likely to be upregulated (n = 1,093) rather than downregulated (n = 387) when bound by SIN3A, arguing for a predominant role of SIN3A in transcriptional repression.

Given the presence of SIN3A both at active and silent promoters, we asked which of these two groups was more sensitive to SIN3A depletion. Interestingly, we found that active genes represented roughly 89% and 76% of upregulated and downregulated genes respectively **(Fig. 3g)**, suggesting that SIN3A is required to ensure proper transcriptional activity primarily at active genes. Finally, we noticed that the top bound genes behaved more uniformly after SIN3A depletion compared to all bound genes **(Fig. 3e,f)**, showing an even stronger trend towards upregulation. Thus, we hypothesized that SIN3A binding might correlate with the directionality of response upon its degradation. To gain better understanding of the link between SIN3A and transcriptional activity, we grouped genes based on the SIN3A occupancy at their promoter and looked at transcriptional changes upon SIN3A removal. Here, we found that, at any time point, upregulation frequency was directly linked to the initial SIN3A occupancy, while downregulation mostly occurred at genes with low or no SIN3A binding **(Fig. 3h)**. Although we cannot formally exclude a potential activator function at specific loci, these data strongly argue that the dominant role of SIN3A on chromatin is buffering gene expression via transcriptional repression and that downregulation of genes after its depletion is more likely due to indirect mechanisms.

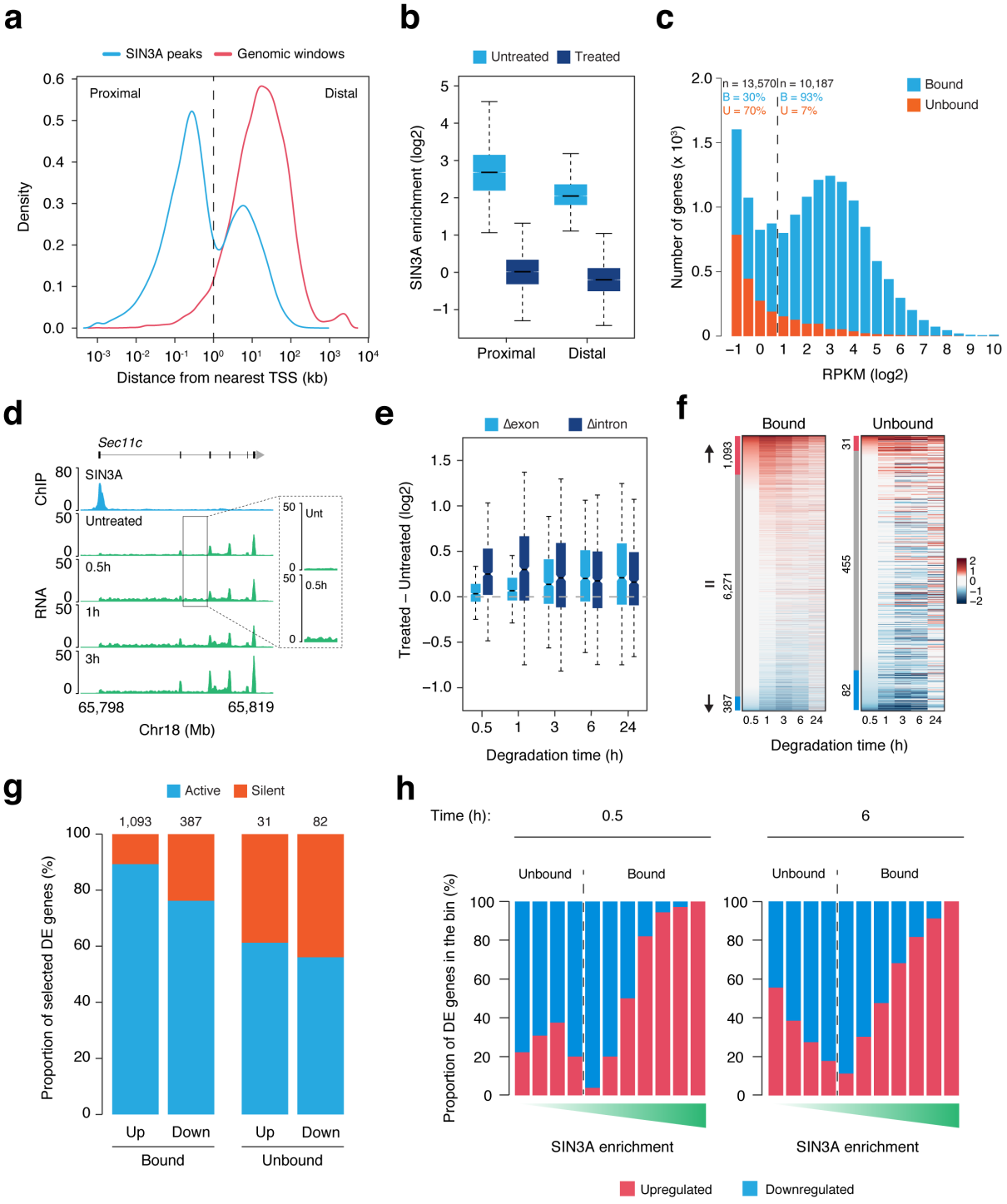


Figure 3: Sin3a binds the vast majority of active genes and drives transcriptional repression
a) Frequency distribution of Sin3a ChIP-seq peaks compared to an equal number of random genomic windows of the same size as a function of distance from nearest TSS.
b) Sin3a ChIP-seq enrichment in peaks proximal to TSSs compared to enrichment in distal peaks. Untreated = DMSO; Treated = dTAG-13 for 6h. Black lines represent the median; boxes represent first and third quartile; whiskers represent maximum and minimum values of distribution after removal of outliers (see methods).
c) Distribution of genes RPKM measured by RNA-seq and colored by occupancy of Sin3a at each respective promoter. Dashed line divides silent genes (log2 RPKM < 1) from active genes (log2 RPKM ≥ 1).

d) Example of a gene bound by Sin3a by ChIP-seq that shows increased intronic signal measured by RNA-seq after Sin3a degradation (inset).
e) Changes in exonic and intronic signal upon Sin3a degradation for the indicated amount of time of the top 1000 Sin3a bound genes (ranked by Sin3a ChIP-seq enrichment at their promoters). Black lines represent the median; boxes represent first and third quartile; whiskers represent maximum and minimum values of distribution after removal of outliers (see methods).
f) Changes in intronic signal of all Sin3a bound and unbound genes upon degradation of Sin3a for the indicated amount of time. Genes are ranked by their change in expression at 0.5h. Annotations are: red, upregulated (log2 FC ≥ 0.5); grey, no change (-0.5 < log2 FC < 0.5); blue, downregulated (log2 FC ≤ –0.5). Values on the left represent the number of genes in each category.
g) Proportion of DE genes from f) that are active or silent in the DMSO condition.
h) Proportion of genes that are differentially expressed at the intronic level (selected with FDR < 10^{-2} to reduce noise) in each bin of Sin3a ChIP-seq enrichment (increasing left to right). Each bin represents a 0.5 (log2) increase in enrichment (see methods).

4.2.2.3 SIN3A colocalizes with H3K4me3 nucleosomes across the genome

The distribution and enrichment of SIN3A peaks across the genome show a clear preference for promoters, however, the exact localization of the complex at these regions remains poorly understood. In particular, we found SIN3A at both silent and active genes, and we wondered if this could be explained by a promoter-specific mechanism or by a shared feature between the two groups. Previous reports showed that SIN3A is in complex with ING1 and ING2, a result that we recapitulated in our cell line by performing SIN3A immunoprecipitation followed by mass spectrometry (data not shown). Both ING proteins possess a PHD domain able to bind to methylated H3K4 and this interaction has been shown to be required for histone deacetylation, gene repression and DNA repair following DNA damage (Peña et al., 2008; Shi et al., 2006). Therefore, it has been proposed that ING1/2 can recruit the SIN3A/HDAC complex to H3K4me3 positive promoters. Of note, this modification is a canonical mark of active genes, but it is also present at poised promoters together with H3K27me3 where transcription is repressed. Thus, we asked whether this feature could link SIN3A binding to both silent and active genes. To answer this question, we split SIN3A bound promoters in two groups, namely transcriptionally repressed and transcriptionally active, and analyzed presence of H3K27ac, H3K4me3 and H3K27me3 in each group. Here, we found that while inactive promoters were on average less acetylated, they showed similar levels of H3K4me3 and higher level of H3K27me3 compared to active genes, suggesting that these might be indeed poised genes and that H3K4me3 is shared feature among SIN3A bound genes **(Fig. 4a)**. In contrast, SIN3A unbound genes showed low levels of all marks **(Fig. 4b)**, with a small number of active genes being H3K4me3 and H3K27ac negative due to wrongly annotated TSSs (data not shown), thus explaining why these active promoters do not have the canonical marks

of active chromatin in our dataset. Together with previous work, this result suggests a direct link between SIN3A binding and presence of H3K4me3, however, whether such link is the result of SIN3A binding directly on nucleosomes or of its stabilization at H3K4 methylated regions remains unclear. To test this, we looked at SIN3A ChIP-seq reads distribution around TSSs, where H3K4me3 is abundant and the strong phasing allows proper discrimination of the single nucleosome particles. We reasoned that binding on nucleosomes would lead to a pattern similar to that of histone ChIP, while a different localization would generate a more distinct pattern. Selecting promoters bound by SIN3A, we found that its enrichment strongly correlates with that of MNase-seq and H3K4me3 ChIP-seq, suggesting a direct localization of SIN3A on nucleosomes **(Fig. 4c)**. To rule out the possibility of this being unspecific enrichment, we performed the same analysis after degrading SIN3A for 6h with dTAG-13. This resulted in a complete loss of signal, confirming that SIN3A specifically colocalizes with nucleosomes at promoters **(Fig. 4c)**. Of note, we tested whether ING2 is responsible for such localization by repeating this analysis in ING2 KO cells and we found that ING2 is dispensable for SIN3A binding at H3K4me3 nucleosomes (data not shown). It remains possible that in our setting ING1 is able to compensate for the loss of ING2 and we are currently investigating this hypothesis by performing ING1 KO and ING1/2 double KO. Together, these findings show that SIN3A colocalizes with H3K4me3 positive nucleosomes genome wide and that its distribution around TSSs mirrors that of H3K4me3.

4.2.2.4 SIN3A buffers acetylation levels by locally driving histone deacetylation

We have shown that binding of SIN3A at promoters is required to establish a constant transcriptional repression and ensure proper levels of expression. However, SIN3A does not possess any catalytic activity and it is thought to act by recruiting other catalytic cofactors that can remodel chromatin. Unexpectedly, we found that SIN3A binding correlated strongly with H3K27 acetylation at promoters **(Fig. 4d)**, a mark that the SIN3A/HDAC complex is known to remove. To gain a better understanding of the relationship between SIN3A binding and histone acetylation, we performed H3K27ac ChIP-seq in cells depleted of SIN3A for 6h and looked at its levels at SIN3A bound promoters. In agreement with the activity of HDAC1/2, we found that degradation of

SIN3A led to a unidirectional hyperacetylation which was directly related to SIN3A occupancy **(Fig. 4e,f)**.

Despite the clear involvement of SIN3A in histone acetylation, many promoters showed limited or no hyperacetylation even at the top end of SIN3A binding, suggesting that some other factors contribute to the regulation of H3K27ac levels. Lack of hyperacetylation might be driven by three main factors: absence of histone acetyltransferases (HATs), absence of modifiable histones and limited sensitivity of the assay. The latter is an intrinsic property that would prove too challenging to overcome, thus, we decided to focus on the first two features. In particular, we reasoned that absence of HATs would translate to an overall low level of H3K27ac at those promoters, while absence of modifiable histones would be a consequence of very high level of acetylation, which leaves little room for further gain. Therefore, we used the basal H3K27ac enrichment as a proxy for these two conditions and looked at its changes upon SIN3A depletion. This analysis revealed that hyperacetylation occurred more strongly at promoters with intermediate levels of H3K27ac **(Fig. 4g)**, arguing that acetylation turnover and substrate availability both contribute to dependency on SIN3A to regulate histone acetylation. Notably, many other residues other than H3K27 underwent hyperacetylation upon SIN3A depletion as we detected by western blot **(Fig. 4h,i)**, which is in agreement with the previously reported lack of specificity of HDAC1/2. Taken together, these data show that SIN3A is a key player in establishing specific levels of chromatin acetylation by locally driving histone deacetylation at promoters across the genome.

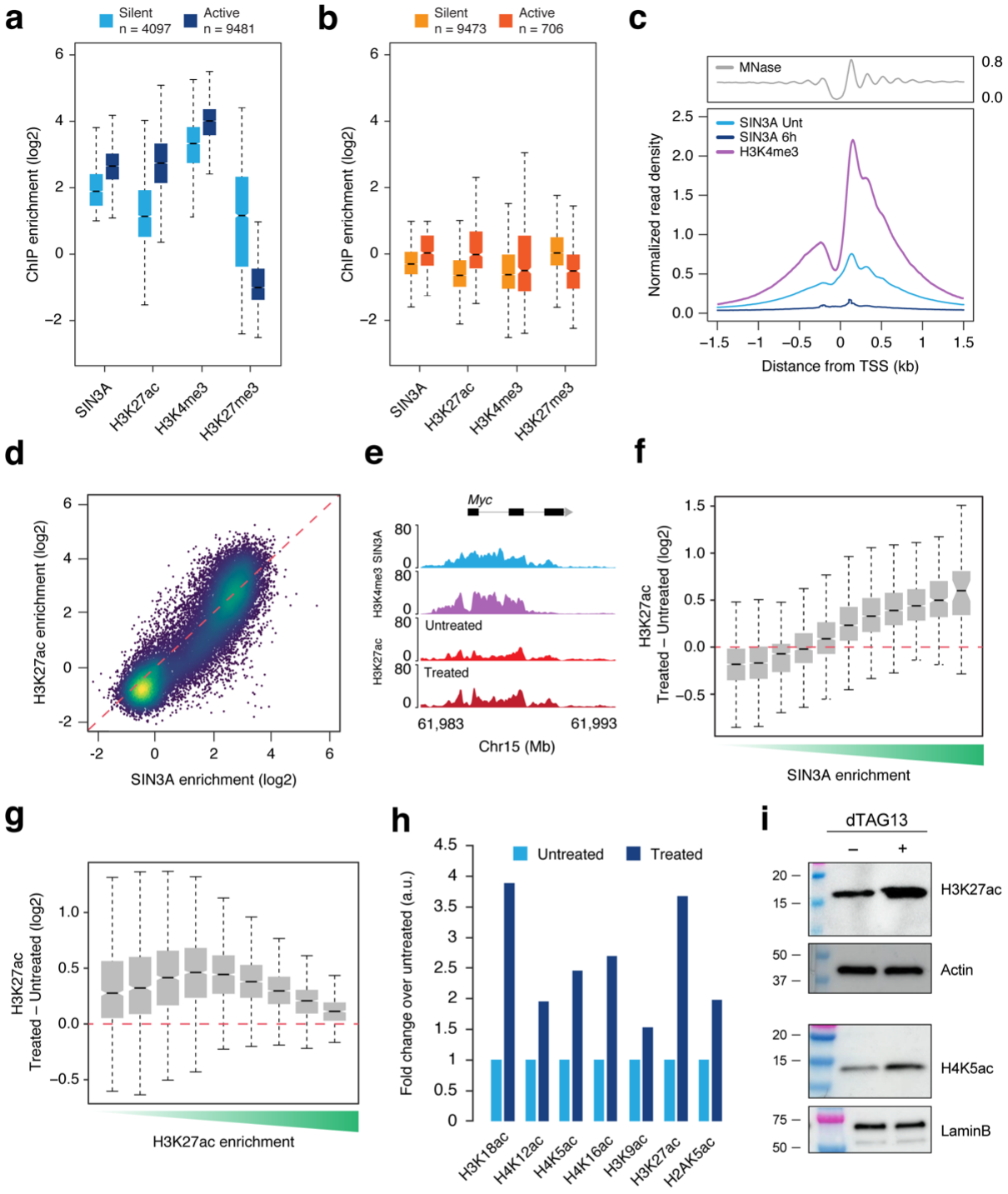


Figure 4: Sin3a colocalizes with H3K4me3 nucleosomes across the genome and buffers acetylation levels by locally driving histone deacetylation

a) ChIP-seq enrichment of Sin3a and different histone marks at Sin3a bound genes and broken down between silent and active genes. Black lines represent the median; boxes represent first and third quartile; whiskers represent maximum and minimum values of distribution after removal of outliers (see methods).

b) ChIP-seq enrichment of Sin3a and different histone marks at Sin3a unbound genes and broken down between silent and active genes. Black lines represent the median; boxes represent first and third quartile; whiskers represent maximum and minimum values of distribution after removal of outliers (see methods).

c) Average ChIP-seq and MNAse-seq read distribution of different datasets around TSSs of Sin3a bound promoters. Unt = Untreated.

d) Pairwise comparison of Sin3a and H3K27ac ChIP-seq enrichment at all promoters.
e) Example of a gene with increased H3K27ac upon Sin3a degradation with the indicated ChIP-seq tracks. Untreated = DMSO, Treated = dTAG-13 for 6h.
f) Changes in H3K27ac ChIP-seq enrichment upon Sin3a degradation at all promoters broken down in bins of increasing Sin3a ChIP-seq enrichment (left to right). Each bin represents a 0.5 (log2) increase in enrichment (see methods). Untreated = DMSO, Treated = dTAG-13 for 6h. Black lines represent the median; boxes represent first and third quartile; whiskers represent maximum and minimum values of distribution after removal of outliers (see methods).
g) Changes in H3K27ac ChIP-seq enrichment upon Sin3a degradation at all promoters broken down in bins of increasing H3K27ac ChIP-seq enrichment in Untreated condition (left to right). Each bin represents a 0.5 (log2) increase in enrichment (see methods). Untreated = DMSO, Treated = dTAG-13 for 6h. Black lines represent the median; boxes represent first and third quartile; whiskers represent maximum and minimum values of distribution after removal of outliers (see methods).
h) Western blot quantification of acetylation of different histone residues upon treatment with DMSO or dTAG-13 for 3h.
i) Examples of western blot gels from h).

4.2.2.5 SIN3A can be recruited by selected TFs in a H3K4me3-independent manner

Looking at SIN3A localization across the genome, we noticed that many of its binding sites lacked H3K4me3 **(Fig. 5a)** and, given the ability of SIN3A to interact with specific TFs, we asked whether TF recruitment could explain this behavior. Previous studies have proposed a plethora of TFs as SIN3A interactors, thereby ruling out the possibility of investigating their colocalization with SIN3A one by one. We reasoned that, if at those loci SIN3A was really recruited by binding of specific TFs, we should expect a higher enrichment of their motifs compared to the rest of the genome. Therefore, we performed motif enrichment analysis on regions bound by SIN3A but devoid of H3K4me3. Interestingly, we found multiple motifs that were significantly enriched and belonged to TFs that have been shown to interact with SIN3A at certain loci **(Fig. 5b)**. Among those, we decided to focus on the RE1-silencing factor (REST), a repressor TF that regulates neuronal genes by interacting with different cofactors (CoREST, SIN3A/HDAC) (Andres et al., 1999; Naruse et al., 1999). REST binds its motif with high affinity both at promoters and distal regions, and drives the formation of a strong DHS thanks to its ability to phase nucleosomes around the binding site. Moreover, REST binding often occurs alone, thereby reducing the probability of confounding effects mediated by the presence of nearby TFs. To test to which extent REST and SIN3A are linked, we used available REST ChIP-seq data in mESCs and compared its enrichment with that of SIN3A at each REST motif across the genome. Despite the difference in dynamic range, we could detect a clear correlation between the two enrichments, suggesting that binding of SIN3A at REST sites might be regulated by the presence of REST **(Fig. 5c)**. Of note, some putative REST motifs that were not bound by the TF were still bound by SIN3A. We found that these sites were strongly

enriched in H3K4me3, a mark that we have shown to correlate with SIN3A binding and that might explain the different behavior of these sites compared to the other REST unbound motifs. Given the direct interaction between REST and SIN3A, we asked whether this could lead to a different binding pattern at REST bound motifs compared to TSSs. To answer this question, we first looked at the average distribution of SIN3A ChIP-seq reads around REST motifs. Here, we saw that SIN3A was highly enriched over REST bound motifs, while no strong enrichment was present on the nearby nucleosomes, suggesting a REST-dependent binding mechanism that led to a narrower localization of the complex **(Fig 5d,e)**. Here, in line with our previous results at H3K4 methylated promoters, degradation of SIN3A led to a slight increase in H3K27ac **(Fig. 5f)**. Indeed, deposition of acetyl groups at inactive chromatin, such as REST bound sites, is generally low or absent, making those sites less responsive to SIN3A removal. Then, to move beyond correlation, we genetically deleted REST in our SIN3A-dTAG cell line and performed ChIP-seq on SIN3A. While no major changes were detectable in SIN3A genome wide localization (data not shown), its enrichment over REST bound motifs was completely abolished, proving the essential role of REST in recruiting SIN3A to chromatin at REST bound regions **(Fig. 5g,h)**. Finally, to test whether SIN3A localization at REST sites was important for REST silencing activity, we compared expression of REST target genes in absence of REST (REST KO) or SIN3A. Here, we found that roughly 50% of genes undergoing upregulation in the REST KO condition were also upregulated in absence of SIN3A **(Fig. 5i)**, arguing that SIN3A is recruited by REST at these genes and drives histone deacetylation and transcriptional repression. Together, these data show that SIN3A can be recruited by specific TFs at their binding sites and that binding occurs over the motif itself rather than on the neighboring nucleosomes. Finally, this mechanism can lead to gene silencing and does not require the presence of H3K4me3-positive nucleosomes, indicating that multiple mechanisms can dictate SIN3A binding specificity.

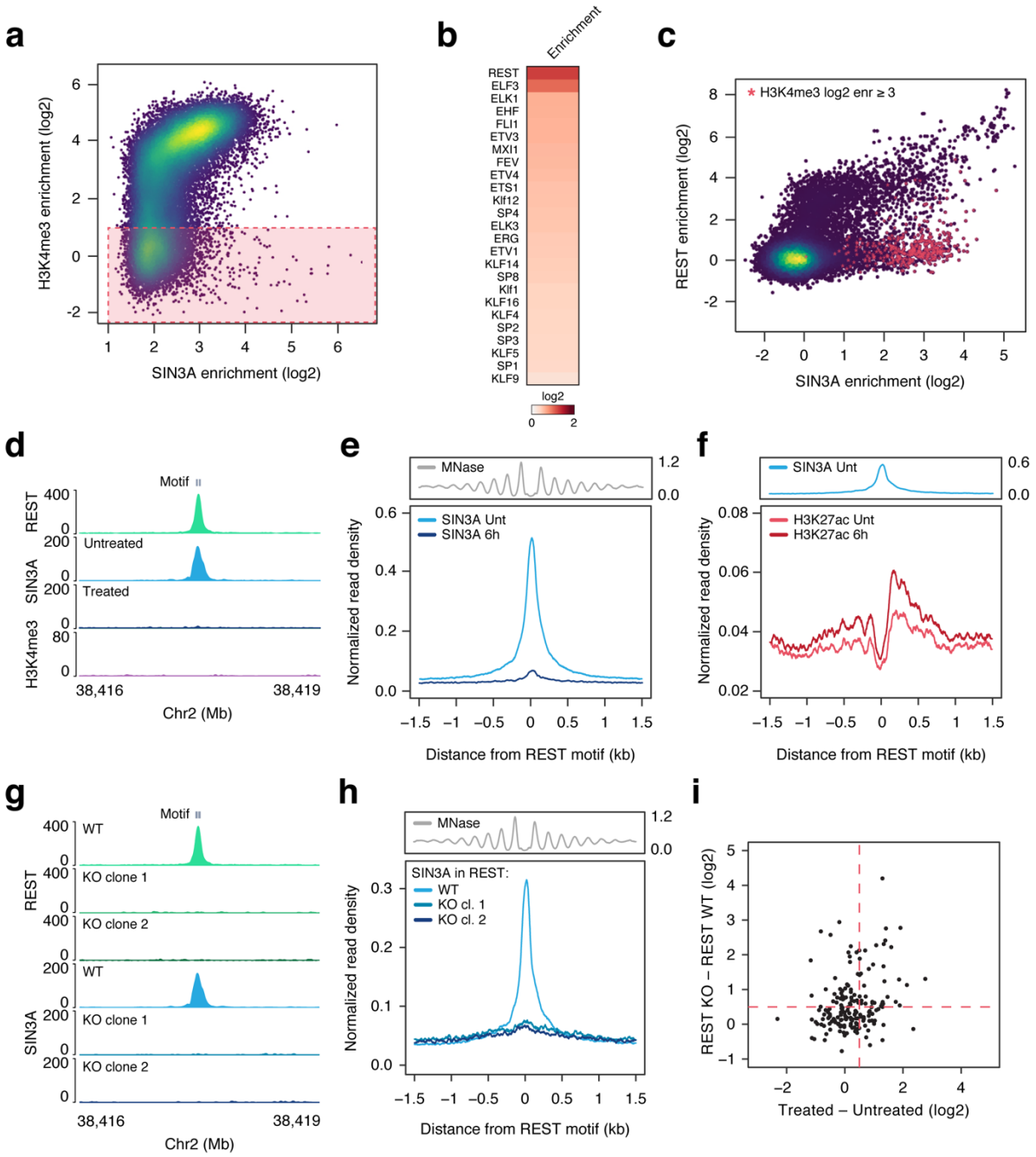


Figure 5: Sin3a can be recruited by selected TFs in a H3K4me3-independent manner
a) Pairwise comparison of Sin3a and H3K4me3 ChIP-seq enrichment at Sin3a peaks. Peaks positive for Sin3a and negative for H3K4me3 are highlighted in red.
b) Motif enrichment analysis performed on peaks highlighted in red in a) (see methods). The top 25 hits are displayed (adj. p-value < 10^{-38}).
c) Pairwise comparison Sin3a and REST ChIP-seq enrichment at REST predicted motifs (see methods). Red star: motifs with H3K4me3 ChIP-seq log2 enrichment ≥ 3.
d) Example of a REST bound site with the indicated ChIP-seq tracks. Motif = REST motif; Untreated = DMSO, Treated = dTAG-13 for 6h.
e) Average Sin3a ChIP-seq and MNase-seq read distribution centered around REST bound motifs. Unt = Untreated.
f) Average H3K27ac and Sin3a ChIP-seq read distribution centered around REST bound motifs. Unt = Untreated.
g) Example of a REST bound site with the indicated ChIP-seq tracks in REST WT and REST KO cells. Motif = REST motif.

h) Average Sin3a ChIP-seq and MNase-seq read distribution around REST bound motifs in REST WT and REST KO cells.
i) Pairwise comparison of changes in expression (intronic signal) in Sin3a degraded cells (3h against DMSO) with changes in REST KO cells (against REST WT). Dashed lines indicate a log2 fold change of 0.5.

4.2.3 Discussion

The complex interaction network of SIN3A and its essential role for cell survival makes it inherently hard to study. In particular, promiscuous binding and impossibility to generate genetic deletions limited our efforts in comprehensively understanding its specificity and activity across the genome. Early studies revealed that SIN3A can interact with specific TFs, such as the MAD-MAX dimer, and drive transcriptional repression, likely exploiting its association with HDAC2 (Ayer et al., 1995, 1996; Laherty et al., 1997). Then, more recent studies implemented loss-of-function and mutagenesis approaches to gain more insights into SIN3A recruitment mechanisms and activity and found indications that SIN3A can also have a positive effect on transcription (Baltus et al., 2009; Solaimani et al., 2014; Zhu et al., 2018b). To date, such context-dependent activity remains debated and no clear mechanism has been proposed. In this study, we investigated on how SIN3A is recruited on chromatin and how its binding affects chromatin state and transcriptional activity genome-wide.

Using CRISPR/Cas9, we inserted a degron tag in the endogenous *Sin3a* locus in mESCs and generated a cell line that, upon delivery of the degrading compound resembled a full SIN3A genetic KO. This tag enabled rapid degradation of the protein, which became undetectable after 1h treatment, and strong ChIP-seq enrichment. Moreover, we took advantage of SIN3A depletion to control for signal specificity, further enhancing the quality of our dataset. Thus, we profiled SIN3A binding and linked it to histone acetylation and transcription, following changes over time upon SIN3A acute degradation. We found that SIN3A binds virtually all active promoters, where it buffers gene expression by driving histone deacetylation and transcriptional repression. It also binds a subset of silent genes, contributing to their silencing. At active and poised promoters, SIN3A colocalizes with H3K4me3 at their TSSs, suggesting a role for this histone modification in SIN3A recruitment regardless of promoter activity. Interestingly, ING2 deletion did not affect SIN3A localization at these sites, arguing that other PHD-domain containing subunits might compensate for its absence. It will be interesting to further test whether and how deletion of ING1 and of

both ING1/2 influence SIN3A binding at H3K4me3-positive nucleosomes. On the other hand, we found that at REST bound sites, SIN3A binding is fully dependent on the presence of REST and does not colocalize with nucleosomes, arguing that SIN3A can be recruited through at least two different mechanisms. Moreover, at a subset of REST target genes, SIN3A-mediated repression is required to maintain silencing rather than buffering their transcription. These results suggests that SIN3A does not work as an on/off switch, but rather as a modulator that contributes to the transcriptional state together with other transcriptional regulators.
Finally, we noticed that transcriptional changes occurring later than 6h post degradation differed significantly from the ones observed at earlier time points. This finding strongly supports the use of a degron tag to study the primary targets of SIN3A and avoid confounders arising from the onset of secondary effects.

Taken together, these results support a model in which SIN3A is recruited with different mechanisms at TSSs and TF binding sites to drive histone deacetylation and transcriptional repression. Nonetheless, the molecular mechanisms behind SIN3A-mediated repression remain to be elucidated. Notably, it has been shown that TF acetylation can be modulated by the presence of the SIN3A/HDAC complex, and this can affect their activity (Icardi et al., 2012; Silverstein and Ekwall, 2005). Therefore, it is possible that SIN3A-mediated regulation of specific TFs may result in transcriptional upregulation. In our work, SIN3A seems to have a role in active transcription only at a small subset of lowly bound genes, but the molecular mechanism remains unclear. Given the observed correlation between SIN3A enrichment and transcriptional repression, it is possible that at sites with low occupancy other indirect mechanisms might overcome the weaker repression established at the TSS. However, more work is needed to fully validate such hypothesis.
Finally, previous works have shown that SIN3A is required for correct development (Halder et al., 2017; Luo et al., 2021; van Oevelen et al., 2010)and it will be intriguing to leverage our mESC line to explore how SIN3A affects the transcriptional program during cell differentiation, for example by simultaneously inducing neuronal differentiation and degrading SIN3A. In conclusion, our work expanded our knowledge on how cofactors can achieve specificity in the context of chromatin and how their binding can modulate the gene expression pattern of the cell.

4.2.4 Methods

Cell culture

Wildtype mouse ES cells, TC-1 line background, were used as WT and for generation of SIN3A-dTAG cell lines. We cultured mouse ES cells as described in (Mohn et al., 2008). Briefly, cells were grown on plates coated with 0.2% gelatin (Sigma) and maintained in Dulbecco's Modified Eagle Medium (Invitrogen), supplemented with 15% Fetal Bovine Serum (Invitrogen), Glutamax (Gibco) and Non-essential amino acids (Gibco), betamercaptoethanol (Sigma) and leukemia inhibitory factor (LIF; produced in-house). Cells used in this study were grown for several passages and not synchronized before any of the experiments.

Generation of SIN3A-dTAG cell line

The SIN3A-dTAG cell line was generated using the CRISPR/Cas9 protocols previously described (Cong et al., 2013), with modifications. Briefly, mouse ES cells were co-transfected (Lipofectamine 3000, Thermo Fisher Scientific) with vectors expressing Cas9 and puromycin resistance and sgRNAs targeting exon 22 of SIN3A. Puromycin (2 µg/ml) was added one day after transfection and kept for 24h. Following a 5-day recovery, individual colonies were expanded and genotyped by western blot and DNA Sanger sequencing. Heterozygous clones were obtained used the following sgRNAs in combination (5' to 3'): sg1: CTCATGGGTGAGGGGCTCGA; sg2: GCTCTGGCTCTGCAGTTAAG. Homozygous clones were obtained by re-targeting the non-tagged allele with the following sgRNA (5' to 3'): sg: CCTGCTCTGGCTCTGCTAAG.

Induced degradation of SIN3A

The induced degradation of SIN3A -dTAG was achieved by adding the compound dTAG13 (Nabet et al., 2018) to the culture medium at a final concentration of 500nM for the amount of time described in each experiment.

Antibodies

Antibody	Type	Vendor	Cat. Number
Actin	wb	Abcam	ab8226
HA	ChIP, wb	Bio Legend	901502

H2AK5ac	wb	Abcam	ab45152
H3K9ac	wb	Abcam	ab32129
H3K18ac	wb	Abcam	ab1191
H3K27ac	ChIP	Abcam	ab4729
H3 acetyl K9+K14+K18+K23+K27	ChIP, wb	Abcam	ab47915
H4K5ac	wb	Upstate	07-327
H4K12ac	wb	Upstate	07-595
H4K16ac	wb	Abcam	ab109463
LaminB1	wb	Abcam	ab16048
REST	ChIP	Santa Cruz	H-290
SIN3A	wb	Abcam	ab3479

ChIP-seq and RNA-seq

ChIP was performed as described in (Barisic et al., 2019) with following modifications: (1) chromatin was sonicated for ~20 cycles of 20 sec. using a Diagenode Bioruptor Pico, with 40 sec. breaks in between cycles, (2) protein A magnetic Dynabeads Magnetic beads (Thermo Fisher Scientific) were used. For SIN3A, 70ug of chromatin were combined with 5ug of anti-HA antibody, for REST, 200ug of chromatin were combined with 5ug of anti-REST antibody and for H3K27ac, 20ug of chromatin were combined with 2ug of anti-H3K27ac antibody. Input and immunoprecipitated DNAs were submitted to library preparation (NEBNext Ultra DNA Library Prep Kit, Illumina). Both input sample and IP sample were amplified during library preparation using 13 PCR cycles. ChIP of SIN3A and H3K27ac in SIN3A -dTAG clone 1 was performed with replicates n = 3 and ChIP of SIN3A and REST in REST KO clones was performed with replicates n = 1 per each clone.

For RNA-seq, RNA was purified using Single Cell RNA Purification Kit (Norgen Biotek). Sequencing libraries were prepared from three biological replicates using TruSeq stranded total RNA Library Prep (Illumina) or TruSeq stranded mRNA Library Prep (Illumina) and sequenced on the Illumina HiSeq 2500.

Reference genome and annotation

All sequence reads were aligned to the mouse GRCm38/mm10 genome assembly (GCA_000001635.2, Dec 2011), obtained from ftp://hgdownload.soe.ucsc.edu/goldenPath/mm10/. Gene annotation was obtained from the Bioconductor package TxDb.Mmusculus.UCSC.mm10.knownGene (version 3.10.0). Promoter regions were defined as a 2kb window centered on known transcript start sites (TSSs), and any position outside of promoters was defined as distal. For data visualization, when multiple TSSs were assigned to a single gene, only the TSS with higher POL-II ChIP-seq enrichment was selected (enrichment calculated as described below in ChIP-seq data analysis, GEO access code: GSM747548). If POL-II enrichment was the same among all TSSs, one single TSS was picked randomly.

RNA-seq data analysis

RNA-seq reads were mapped to the genome using STAR version 2.5.2 with the following parameters: --outFilterMultimapNmax 20 –outMultimapperOrder Random –alignSJoverhangMin 8 –alignSJDBoverhangMin 1 –outSAMmultNmax 1 –outSAMunmapped Within. Alignments overlapping the opposite strand of any exon of a gene were counted using the qCount function from the QuasR Bioconductor package version 1.32.0 with default parameters. Samples were then normalized to the sum of reads falling into exons (instead of library size). Quasi-likelihood method (glmQLFit and glmQLFTest functions) from the edgeR Bioconductor package version 3.34.0 was used with default parameters to identify differentially expressed genes. This method fits a single model to all RNA-seq samples with a design that contains the condition (DMSO or treatment) as single experimental factor.

ChIP-seq data analysis

ChIP-seq reads were aligned to the genome using qAlign from the QuasR Bioconductor package version 1.32.0 with default parameters, only reporting alignments for reads with a unique match to the genome. ChIP and input replicates were pooled for all analyses after initial quality control on single replicates (as shown in Fig. 2), except for REST KO replicates, which have been treated as independent biological replicates. In order to increase resolution of ChIP alignment pileups over TSSs and TF binding sites, reads were shifted towards the 3'-end of the read until maximum overlap between reads on the plus and minus strands was achieved.

Samples were normalized by library size. The scaling factor for each sample was obtained by dividing the respective library size by the smallest. ChIP meta profiles (e.g. around TSS or TF binding sites) were normalized for library size and other technical differences by multiplying each profile by the respective scaling factor and then dividing it by the total number of regions under consideration. Signal is then smoothed by averaging counts in a 51-bp running window.

ChIP enrichment between immuno-precipitated and input samples were calculated first by counting reads in the regions of interest for each sample and adding a pseudocount of 8. Values were then log2 transformed and enrichment was obtained by subtracting the log2 reads count of the input sample from the log2 reads count of the immuno-precipitated sample. Changes of ChIP enrichment between two immuno-precipitated samples were calculated using the same formula. Regions with ChIP enrichment ≥2-fold (log2 ≥ 1) were defined as bound. SIN3A ChIP peaks were identified for each replicate (n = 3) using macs2 version 2.1.3.3 on ChIP samples and input samples as controls with parameters -f BAM -g 1.87e9. Identified peaks were then filtered based on their enrichment over input. Briefly, reads were counted in each peak for each replicate and SIN3A enrichment over input was calculated as described above. All peaks showing enrichment ≥2-fold (log2 ≥ 1) in all replicates were selected and then merged in a final set consisting of 25,967 peaks.

For analyses in enrichment bins, bins were defined as follow. First, log2 ChIP enrichment over input at the target regions was calculated. Based on the obtained enrichment, for SIN3A and H3K27ac, the ranges (-1,4) and (0,4) were used respectively. The bins were defined as 0.5 (log2) increasing steps with the interval [x,x+0.5) starting from the range minimum. To avoid bins with too few elements, the first and last bins were then defined as (-Inf,$range_{min}$) and ($range_{max}$,+Inf) respectively.

Motif enrichment analysis

Motif enrichment analysis in SIN3A peaks devoid of H3K4me3 was performed using the single set motif enrichment analysis the monaLisa package from (Machlab et al., 2022). Briefly, the function calcBinnedMotifEnrR was used with default parameters and background = “genome” to calculate enrichment of motifs in selected SIN3A peaks compared to a set of random genomic regions matched by GC content.

Boxplots

In all boxplots except in Fig. 3e, whiskers extend to ±1.5 x interquartile range (IQR). In Fig. 3e, whiskers extend to ±1 x IQR to allow better visualization. All notches extend to ±1.58 x (IQR/sqrt(n)).

5 General discussion

Binding of TFs and cofactors is instrumental to ensure proper gene expression patterns. TFs recognize and bind to consensus motifs on DNA, but the vast majority of mammalian TFs only bind a small fraction of their potential binding sites in the genome. Different chromatin features have been shown to affect TF binding *in vitro*, such as DNA methylation and nucleosomes. DNA methylation primarily affects TFs binding to CpG-containing motifs, therefore accounting only in part for the regulation of TF binding. In contrast, nucleosomes are thought to act as a general barrier, with only selected TFs called “pioneers” being able to engage with their motif on nucleosomal DNA and open chromatin. This, in turn, would allow general TFs to access DNA in closed chromatin, establishing precise hierarchies between TFs. However, even pioneer TFs do not bind all instances of their motifs, suggesting a certain degree of chromatin sensitivity. In the first part of this study, we aimed at unravelling the impact of nucleosomes on TF binding *in vivo*, in order to bridge the gap between *in vitro* and *in vivo* data. To achieve this, we developed a novel system to position a nucleosome *in vivo* at specific locus in the mESCs genome and tested the ability of individual motifs to recruit TFs at different locations along nucleosomal DNA.

Cofactors do not directly read DNA sequence and thus rely on complex interaction networks to achieve specificity on chromatin. These interactions are often weak and promiscuous, making it hard to identify cofactors partners and activity on chromatin. Moreover, many cofactors are essential and loss-of-function approaches can be hard to establish. In the second part of this study, we aimed at dissecting the role of the corepressor SIN3A in histone acetylation and transcriptional regulation, as well as its recruitment to chromatin. SIN3A is of particular interest because it is an essential protein thought to have a context-dependent function on gene expression. On one hand, it has been shown to be required for targeted gene silencing during different biological processes. This, together with its association with HDAC1/2, classifies SIN3A as a canonical corepressor. On the other hand, different studies found SIN3A binding at active promoters and to be required for activation of specific genes. This led to the conclusion that SIN3A can also act as a context-dependent activator. However, this model is largely speculative and the molecular mechanisms behind SIN3A-mediated regulation remain elusive. Thus, we endogenously tagged SIN3A with a

degron tag to generate the first high quality map of SIN3A binding in mESCs and took advantage of acute protein degradation to identify SIN3A primary targets. This enabled us to avoid potential confounders occurring with the onset of secondary effects.

5.1 Positioning nucleosomes *in vivo*

Moving beyond correlation in the study of the interactions between TFs and nucleosomes *in vivo* requires the ability to simultaneously control nucleosome positioning and TF binding. However, while TFs can be recruited by specific motifs, no sequence has been identified to recruit nucleosomes thus far. Different DNA elements have been successfully used *in vitro* to reconstitute nucleosomes (Lowary and Widom, 1998; Matsumoto et al., 2019; Soufi et al., 2015), but whether their properties hold *in vivo* remains unclear. We demonstrate that even the Widom's 601 element, which has very strong affinity for the histone octamer *in vitro*, is not able to precisely position a nucleosome in mESCs, arguing that nucleosome positioning is not solely dictated by the DNA sequence. This is in line with previous findings finding no evidence for a nucleosome code regulating nucleosome positioning (Stein et al., 2010; Zhang et al., 2009, 2010). It is possible that chromatin remodelers prevent strong positioning by constantly shuffling nucleosomes across the genome even at sites where they are not specifically recruited by TFs. This could result from a "basal" activity driven by the intrinsic affinity of the ATPase subunits of the complexes for the nucleosome particles. In contrast, TFs such as REST and CTCF are able to create arrays of phased nucleosomes upon binding and we show that this ability is maintained also at ectopic sites, allowing to enforce nucleosome positioning at the desired location. Importantly, both REST and CTCF are known to interact with a plethora of factors such as chromatin remodelers, corepressors and other cofactors such as cohesin (Andres et al., 1999; Barisic et al., 2019; Parelho et al., 2008). Thus, observations require careful evaluation and validation to avoid confounding effects driven by non-nucleosome factors. Nevertheless, with proper controls this setting opens up different possibilities to study the effect of nucleosomes on TF binding and transcriptional regulation, and how nucleosome phasing is regulated. For example, changing nucleosome positioning at regulatory elements could inform on how accessibility pattern affects the activity of the element. Alternatively, distorting phasing at regions with positioned nucleosomes such as at TSSs or TF binding sites could be used to learn how positioning is achieved

and what are the principles that govern it. It would be also fascinating to discover new factors able to phase nucleosomes and dissect their role in shaping gene expression and chromatin accessibility.

5.2 TF binding to single instances of their motif

The possibility to assay multiple factors in parallel with NOMe-seq enabled us to demonstrate that TFs differ greatly in their ability to engage with a single instance of their motif even in the absence of nucleosomes. Indeed, we show that the majority of the known TFs expressed in mESCs are not able to generate a footprint despite the absence of a nucleosome on the motif. Notably, this does not mean that these factors never bind to their motif, but rather that this interaction is likely very weak or has a very fast turnover. Such dynamics would create a chance for the GpC methyltransferase to methylate DNA by outcompeting the TF or by opportunistic binding respectively. In line with previous studies *in vitro* (Li and Wrange, 1993; Michael and Thomä, 2021), positioning a nucleosome over the motifs further reduces the fraction of TFs generating a footprint, arguing that a single instance of a motif is even less efficient in recruiting a TF when it resides on nucleosomal DNA. It will be intriguing to explore how combinations of different variables affect the recruitment of specific TFs. For instance, placing more instances of the same motif, or combinations of motifs with different scores could give new insights into TF binding directly *in vivo*. Also, changing the CpG density around the motif might reveal sensitivity of TFs to CG content and exploiting this setting in cells devoid of DNA methylation (e.g. DNMT-TKO) would further help in discriminating sensitivity to CG content from sensitivity to DNA methylation.

5.3 TF binding as a function of nucleosome sensitivity

At regulatory regions, TFs usually bind in clusters, making it hard to study their individual contribution in establishing an accessible site. With our reductionist approach, we generate a simpler scenario where individual TFs and nucleosomes compete to occupy the same DNA sequence. This enables to directly test nucleosome sensitivity of different TFs and we show that only few are able to engage with their motif on nucleosomal DNA. Moreover, we demonstrate that this ability is better represented by a continuous range of binding efficiency rather than by a binary

classification (i.e. pioneers vs non-pioneers). Indeed, even the same TF displays differences in binding depending on motif position along nucleosomal DNA. This suggests that binding in closed chromatin is not an intrinsic feature of selected TFs and could explain why the vast majority of motifs in the mammalian genome are not bound. In particular, we show that even a top motif can be hard to access when placed close to the nucleosome dyad, therefore it is safe to assume that nucleosomes could easily restrict binding of weaker motifs in the genome, which represent the biggest fraction.

While individual TFs struggle to access DNA at nucleosome occupied regions, a recent work has shown that cooperativity between TFs is instrumental to establish stable binding and open chromatin at certain cis-regulatory regions (Sönmezer et al., 2021). Moreover, another study revealed that the TFs FOXA1 and HNF4A can facilitate binding of each other and that FOXA1 only requires slightly less motif instances to open chromatin compared to HNF4A (Hansen et al., 2022). Together, these findings further argue that binding efficiency can be modulated by different factors and that even pioneer TFs can be dependent on other TFs. This, in turn, would generate hierarchies among TFs which are far more complex than the ones predicted by the pioneer factor model. However, how cooperativity between TFs affects binding at specific locations on nucleosomal DNA *in vivo* remains to be determined. Our reductionist approach can be leveraged to explore the role of multiple TF motifs at regions with phased nucleosomes, combining pairs of different TFs and different motif scores and placing them at different positions along nucleosomal DNA. This could inform for instance on the role of motif orientation, number or distance in binding and remodeling closed chromatin, potentially contributing to our ability to predict TF binding from DNA sequence.

5.4 Nucleosomes as a tool to control TF binding

The high occurrence of motifs in large genomes makes it necessary to redirect TFs to the correct functional targets. Indeed, constant binding of all motifs would require very high concentration of each TF and would lead to unspecific gene regulation, conditions that are likely to worsen the fitness of an organism. Masking motifs with nucleosomes represents an efficient way to avoid spurious binding without intervening on the DNA sequence, which instead would require several mutations to occur and to be

maintained. However, nucleosomes act as a roadblock regardless of their position in the genome, thus affecting binding also at regions that need to be bound. Here, presence of multiple motifs could be a feature evolved to outcompete nucleosomes by cooperative binding and specific TFs could have evolved to be necessary for opening chromatin. Nevertheless, these TFs might be necessary but not sufficient, as we have shown in the case of OCT4-SOX2, where these reprogramming factors are not able to evict a nucleosome on their own. Instead, co-occurrence of other motifs might lead to greater distortion of the DNA around the histone octamer, eventually leading to nucleosome eviction. This feature might have shaped regulatory regions by selecting those bearing clusters of motifs and could at least partially explain why only a minority of sites are bound across the genome.

5.5 Linking cofactors binding to activity

Knowing the genomic localization of cofactors is key to understand their activity and infer causality. However, their interactions with chromatin and other DNA binding proteins are usually weak and temporary, thus, mapping their binding in the genome is inherently difficult. Indeed, SIN3A binding has been long studied, yet a high-quality map of its binding sites genome wide has never been obtained. In our study, we take advantage of CRISPR/Cas9 to endogenously tag SIN3A and improve pulldown efficiency in ChIP-seq assays. Moreover, using a degron tag, we coupled the high ChIP-seq enrichment with acute depletion of the protein, enabling specific detection of SIN3A binding in the genome of mESCs. Another important advantage of this approach is that standard genetic deletion or RNAi-mediated knockdown require days before analysis, while treatment with dTAG leads to full SIN3A depletion in just one hour. This allows observation of early events that likely represent the true direct response to SIN3A loss, avoiding confounders driven by secondary effects occurring at later time points. Overall, our strategy can be applied to many other cofactors, especially to those that are essential, to simultaneously obtain a clear map of their genomic localization and a high-resolution overview of their activity in the cell.

5.6 Histone marks, TFs and SIN3A recruitment

Recruitment of cofactors at multiple loci in the genome requires either the presence of a feature that is shared across all loci, or the ability of the complex to recognize

different marks on chromatin. Importantly, these two mechanisms are not mutually exclusive and in our study we show that SIN3A can exploit both to bind its target sites. In particular, it can colocalize with H3K4me3 at promoters where it follows nucleosomes distribution, strongly suggesting a role of this histone modification in directing SIN3A binding to nucleosomal particles **(Fig. 1a)**. Notably, SIN3A do not always colocalize with other histone marks present at promoters and at distal regions, such as H3K27ac or H3K27me3, arguing that its colocalization with H3K4me3 is not driven by an unspecific affinity to nucleosomes. On the other hand, SIN3A also colocalizes with REST at REST bound motifs, where it enriches over the motif similarly to a TF **(Fig. 1b)**. Here, REST is necessary for SIN3A binding, showing that this cofactor can also depend on a TF to bind its targets. While we do not have evidence of a direct interaction between SIN3A and H3K4me3, the highly different binding profiles at TSSs and REST bound sites strongly suggest that SIN3A exploits different binding partners to access chromatin, revealing that individual cofactors can evolve multiple mechanisms to bind their targets. Perturbation of H3K4me3 and of SIN3A domains and interaction partners will be key to determine whether and how this histone mark affects SIN3A binding and to distinguish between direct and indirect recruitment. For instance, addition of H3K4me3 at ectopic loci using CRISPR/dCas9 (Cano-Rodriguez et al., 2016) can be used to test whether its presence is sufficient to recruit SIN3A or if other pre-existing chromatin states are necessary. Alternatively, deleting or mutating SIN3A domains could reveal which parts of the protein are important for its localization, possibly revealing required interactions with other cofactors.

5.7 SIN3A as a broad modulator of transcription

The presence of repressive cofactors such as HDACs at active promoters has been reported more than 10 years ago (Kurdistani et al., 2002; Wang et al., 2009), however, their recruitment and role beyond histone acetylation remain poorly understood. Likewise, SIN3A binding at active genes has been reported in some studies and led to the hypothesis of a context-dependent activity (Baltus et al., 2009; Solaimani et al., 2014; Zhu et al., 2018b). Nonetheless, the putative mechanism behind this context-dependency remains unexplored. We find that SIN3A is actually bound to the vast majority of active promoters, suggesting a broad role in transcriptional regulation.

Moreover, we show that SIN3A enrichment is strongly correlated with repression and histone deacetylation at its primary targets and that genes losing activity upon SIN3A loss are a minority characterized by very low SIN3A enrichment. This suggests that its role at bound TSSs is to tune down transcription and that more indirect mechanisms, such as deacetylation of specific TFs, might drive activation at a small subset of genes by overcoming the weak repression established at the TSS.

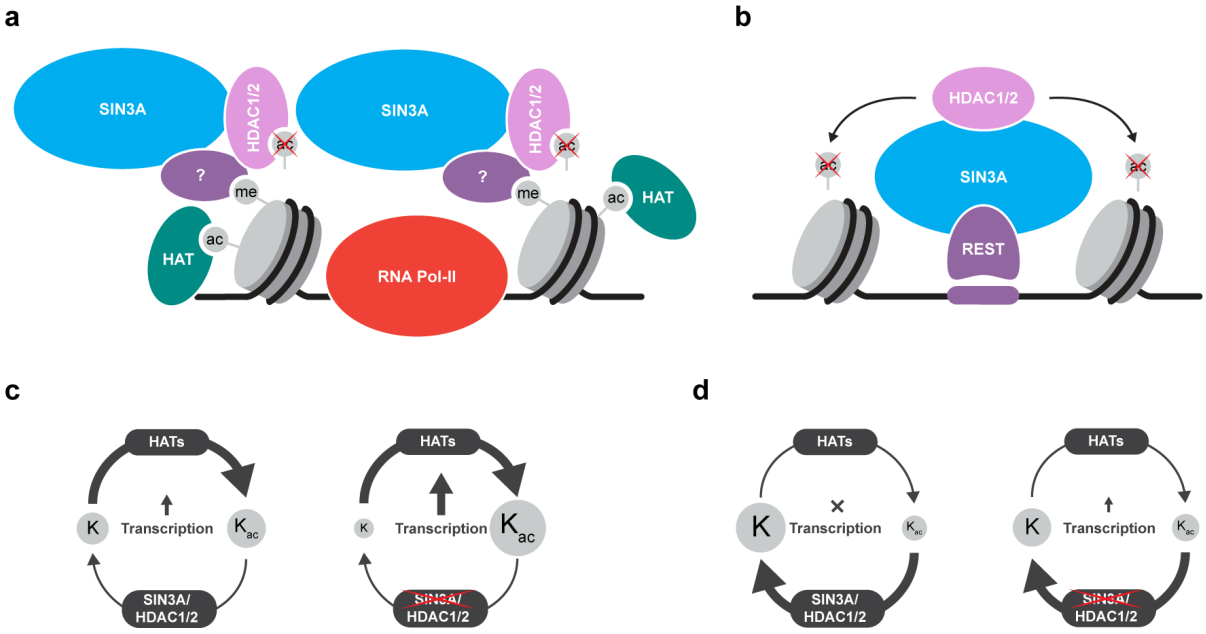


Figure 11: Mechanisms of SIN3A activity in mESCs.

a) Representation of SIN3A binding at active promoters. RNA Pol-II is at the TSS and the -1 and +1 nucleosomes are shown. b) Representation of SIN3A binding at REST bound sites. c) Acetylation turnover at active promoters in the presence (left) or absence (right) of SIN3A. d) Acetylation turnover at REST bound sites in the presence (left) or absence (right) of SIN3A. In a) the "?" stands for an unknown subunit or SIN3A domain that tethers SIN3A to H3K4me3. In a) and b) me = H3K4me3; ac = histone lysine acetylation. In c) and d) rates of acetylation and deacetylation are shown by the arrows size. Total levels of lysine acetylation are represented by the grey circle radius. K = deacetylated histone lysine; K_{ac} = deacetylated histone lysine.

It is tempting to speculate that constantly repressing transcription at active genes might allow the cell to maintain a proper level of expression by reaching an equilibrium between activators and repressors **(Fig. 1c)**. Moreover, shifting this equilibrium would also enable a faster change in transcriptional activity compared to a simple de novo recruitment of activators or repressors. Importantly, while we expanded our knowledge of the effects of SIN3A on transcription and acetylation, whether they are connected remains unclear and what mechanisms dictate SIN3A-mediated repression remains

to be elucidated. It will be intriguing to thoroughly study the interactome of SIN3A and dissect the role of its interaction partners in transcriptional repression by acute depletion with dTAG.

Overall, our study shed light on how cofactors can achieve specificity, how they affect chromatin state and transcriptional activity, and showed that identification of primary targets through acute protein depletion is a powerful strategy to dissect cofactor activity and avoid confounders.

6 Bibliography

Adams, D., Altucci, L., Antonarakis, S.E., Ballesteros, J., Beck, S., Bird, A., Bock, C., Boehm, B., Campo, E., Caricasole, A., et al. (2012). BLUEPRINT to decode the epigenetic signature written in blood. Nat. Biotechnol. *30*, 224–226.

Adams, G.E., Chandru, A., and Cowley, S.M. (2018). Co-repressor, co-activator and general transcription factor: the many faces of the Sin3 histone deacetylase (HDAC) complex. Biochem. J. *475*, 3921–3932.

Andres, M.E., Burger, C., Peral-Rubio, M.J., Battaglioli, E., Anderson, M.E., Grimes, J., Dallman, J., Ballas, N., and Mandel, G. (1999). CoREST: A functional corepressor required for regulation of neural-specific gene expression. Proc. Natl. Acad. Sci. U. S. A. *96*, 9873.

Andrews, F.H., Strahl, B.D., and Kutateladze, T.G. (2016). Insights into newly discovered marks and readers of epigenetic information. Nat. Chem. Biol. 2016 129 *12*, 662–668.

Arnold, P., Schöler, A., Pachkov, M., Balwierz, P.J., Jørgensen, H., Stadler, M.B., Van Nimwegen, E., and Schu̇beler, D. (2013). Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. Genome Res. *23*, 60–73.

Ayer, D.E. (1999). Histone deacetylases: transcriptional repression with SINers and NuRDs. Trends Cell Biol. *9*, 193–198.

Ayer, D.E., Lawrence, Q.A., and Eisenman, R.N. (1995). Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. Cell *80*, 767–776.

Ayer, D.E., Laherty, C.D., Lawrence, Q.A., Armstrong, A.P., and Eisenman, R.N. (1996). Mad proteins contain a dominant transcription repression domain. Mol. Cell. Biol. *16*, 5772–5781.

Baltus, G.A., Kowalski, M.P., Tutter, A. V., and Kadam, S. (2009). A positive regulatory role for the mSin3A-HDAC complex in pluripotency through Nanog and Sox2. J. Biol. Chem. *284*, 6998–7006.

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.

Banks, C.A.S., Zhang, Y., Miah, S., Hao, Y., Adams, M.K., Wen, Z., Thornton, J.L., Florens, L., and Washburn, M.P. (2020). Integrative Modeling of a Sin3/HDAC Complex Sub-structure. Cell Rep. *31*, 107516.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. 2011 213 *21*, 381–395.

Bao, Y., and Shen, X. (2007). INO80 subfamily of chromatin remodeling complexes. Mutat. Res. Mol. Mech. Mutagen. *618*, 18–29.

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.

Barnes, C.E., English, D.M., and Cowley, S.M. (2019). Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription.

Essays Biochem. *63*, 97–107.

Bartlett, A., O'Malley, R.C., Huang, S.S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. *12*, 1659–1672.

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

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Bestor, T.H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *326*, 179–187.

Bestor, T.H. (2003). Cytosine methylation mediates sexual conflict. Trends Genet. *19*, 185–190.

Biggin, M.D. (2011). Animal Transcription Networks as Highly Connected, Quantitative Continua. Dev. Cell *21*, 611–626.

Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell *168*, 135-149.e22.

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.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell *132*, 311–322.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1-21.29.9.

Busslinger, M., Hurst, J., and Flavell, R.A. (1983). DNA methylation and the regulation of globin gene expression. Cell *34*, 197–206.

Cano-Rodriguez, D., Gjaltema, R.A.F., Jilderda, L.J., Jellema, P., Dokter-Fokkens, J., Ruiters, M.H.J., and Rots, M.G. (2016). Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat. Commun. *7*.

Chandru, A., Bate, N., Vuister, G.W., and Cowley, S.M. (2018). Sin3A recruits Tet1 to the PAH1 domain via a highly conserved Sin3-Interaction Domain. Sci. Reports 2018 81 *8*, 1–10.

Chen, T., and Dent, S.Y.R. (2014). Chromatin modifiers: regulators of cellular differentiation. Nat. Rev. Genet. *15*, 93.

Chrysanthou, S., Tang, Q., Lee, J., Taylor, S.J., Zhao, Y., Steidl, U., Zheng, D., and Dawlaty, M.M. (2022). The DNA dioxygenase Tet1 regulates H3K27 modification and embryonic stem cell biology independent of its catalytic activity. Nucleic Acids Res. *50*, 3169–3189.

Cirillo, L.A. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244–254.

Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. Http://Dx.Doi.Org/10.1146/Annurev.Biochem.77.062706.153223 *78*, 273–304.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (80-.). *339*, 819–823.

Corona, D.F.V., and Tamkun, J.W. (2004). Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochim. Biophys. Acta - Gene Struct. Expr. *1677*, 113–119.

Cowley, S.M., Iritani, B.M., Mendrysa, S.M., Xu, T., Cheng, P.F., Yada, J., Liggitt, H.D., and Eisenman, R.N. (2005). The mSin3A Chromatin-Modifying Complex Is Essential for Embryogenesis and T-Cell Development. Mol. Cell. Biol. *25*, 6990.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491–496.

Dodonova, S.O., Zhu, F., Dienemann, C., Taipale, J., and Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nat. 2020 5807805 *580*, 669–672.

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.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., and Levine, M.S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. U. S. A. *113*, 6508–6513.

Feng, J., Zhu, F., Ye, D., Zhang, Q., Guo, X., Du, C., and Kang, J. (2022). Sin3a drives mesenchymal-to-epithelial transition through cooperating with Tet1 in somatic cell reprogramming. Stem Cell Res. Ther. *13*, 1–16.

Fernandez Garcia, M., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H., and Zaret, K.S. (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. Mol. Cell *75*, 921-932.e6.

Fierz, B., and Poirier, M.G. (2019). Biophysics of Chromatin Dynamics. Annu. Rev. Biophys. *48*, 321–345.

Gao, Z., Ure, K., Ding, P., Nashaat, M., Yuan, L., Ma, J., Hammer, R.E., and Hsieh, J. (2011). The Master Negative Regulator REST/NRSF Controls Adult Neurogenesis by Restraining the Neurogenic Program in Quiescent Stem Cells. J. Neurosci. *31*, 9772.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M.L. (2013). Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. Cell Rep. *3*, 1093–1104.

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. Nat. 2021 5967870 *596*, 133–137.

Halder, D., Lee, C.H., Hyun, J.Y., Chang, G.E., Cheong, E., and Shin, I. (2017). Suppression of Sin3A activity promotes differentiation of pluripotent cells into functional neurons. Sci. Reports 2017 71 *7*, 1–11.

Hansen, J.L., Loell, K.J., and Cohen, B.A. (2022). The Pioneer Factor Hypothesis is not necessary to explain ectopic liver gene activation. Elife *11*.

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.

Hergeth, S.P., and Schneider, R. (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep. *16*, 1439.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR/Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510.

Huang, Y., Myers, S.J., and Dingledine, R. (1999). Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. Nat. Neurosci. 1999 210 *2*, 867–872.

Icardi, L., Mori, R., Gesellchen, V., Eyckerman, S., De Cauwer, L., Verhelst, J., Vercauteren, K., Saelens, X., Meuleman, P., Leroux-Roels, G., et al. (2012). The Sin3a repressor complex is a master regulator of STAT transcriptional activity. Proc. Natl. Acad. Sci. U. S. A. *109*, 12058–12063.

Isbel, L., Grand, R.S., and Schübeler, D. (2022). Generating specificity in genome regulation through transcription factor sensitivity to chromatin. Nat. Rev. Genet. 2022 1–13.

Iurlaro, M., Stadler, M.B., Masoni, F., Jagani, Z., Galli, G.G., and Schübeler, D. (2021). Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat. Genet. 2021 533 *53*, 279–287.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33 Suppl*, 245–254.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. 2009 103 *10*, 161–172.

Johnson, P.F., and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. *58*, 799–839.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., et al. (2010). Multiplexed massively

parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. *20*, 861–873.

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.

Kaluscha, S., Domcke, S., Wirbelauer, C., Burger, L., and Schübeler, D. (2022). Direct inhibition of transcription factor binding is the dominant mode of gene and repeat repression by DNA methylation. Nat. Genet.

Kelly, R.D.W., and Cowley, S.M. (2013). The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. Biochem. Soc. Trans. *41*, 741–749.

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.

King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. Elife *6*.

Kornberg, R.D. (1974). Chromatin structure: A repeating unit of histones and DNA. Science (80-.). *184*, 868–871.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure: Oligomers of the histones. Science (80-.). *184*, 865–868.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

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.

Kribelbauer, J.F., Lu, X.J., Rohs, R., Mann, R.S., and Bussemaker, H.J. (2020a). Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. J. Mol. Biol. *432*, 1801–1815.

Kribelbauer, J.F., Loker, R.E., Feng, S., Rastogi, C., Abe, N., Rube, H.T., Bussemaker, H.J., and Mann, R.S. (2020b). Context-Dependent Gene Regulation by Homeodomain Transcription Factor Complexes Revealed by Shape-Readout Deficient Proteins. Mol. Cell *78*, 152-167.e11.

Kumar, D., Cinghu, S., Oldfield, A.J., Yang, P., and Jothi, R. (2021). Decoding the function of bivalent chromatin in development and cancer. Genome Res. *31*, 2170–2184.

Kurdistani, S.K., Robyr, D., Tavazoie, S., and Grunstein, M. (2002). Genome-wide binding map of the histone deacetylase Rpd3 in yeast. Nat. Genet. *31*, 248–254.

Laherty, C.D., Yang, W.M., Jian-Min, S., Davie, J.R., Seto, E., and Eisenman, R.N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell *89*, 349–356.

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.

Lancrey, A., Joubert, A., Duvernois-Berthet, E., Routhier, E., Raj, S., Thierry, A., Sigarteu, M., Ponger, L., Croquette, V., Mozziconacci, J., et al. (2022). Nucleosome Positioning on Large Tandem DNA Repeats of the “601” Sequence Engineered in Saccharomyces cerevisiae. J. Mol. Biol. *434*.

Laugesen, A., and Helin, K. (2014). Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. Cell Stem Cell *14*, 735–751.

Levo, M., Zalckvar, E., Sharon, E., Dantas Machado, A.C., Kalma, Y., Lotam-Pompan, M., Weinberger, A., Yakhini, Z., Rohs, R., and Segal, E. (2015). Unraveling determinants of transcription factor binding outside the core binding site. Genome Res. *25*, 1018–1029.

Li, G., and Widom, J. (2004). Nucleosomes facilitate their own invasion. Nat. Struct. Mol. Biol. *11*, 763–769.

Li, Q., and Wrange, O. (1993). Translational positioning of a nucleosomal glucocorticoid response element modulates glucocorticoid receptor affinity. Genes Dev. *7*, 2471–2482.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nat. 1993 3666453 *366*, 362–365.

Li, G., Tian, Y., and Zhu, W.G. (2020). The Roles of Histone Deacetylases and Their Inhibitors in Cancer Therapy. Front. Cell Dev. Biol. *8*, 1004.

Lowary, P.., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

Luo, L., Dang, Y., Shi, Y., Zhao, P., Zhang, Y., and Zhang, K. (2021). SIN3A Regulates Porcine Early Embryonic Development by Modulating CCNB1 Expression. Front. Cell Dev. Biol. *9*, 191.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. Cell *132*, 958–970.

Machlab, D., Burger, L., Soneson, C., Rijli, F.M., Schübeler, D., and Stadler, M.B. (2022). monaLisa: an R/Bioconductor package for identifying regulatory motifs. Bioinformatics *38*, 2624–2625.

Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987). Regulation of inducible and tissue-specific gene expression. Science *236*, 1237–1245.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nat. 2011 4697330 *469*, 343–349.

Martin, B.J.E., Brind’Amour, J., Kuzmin, A., Jensen, K.N., Liu, Z.C., Lorincz, M., and Howe, L.A.J. (2021). Transcription shapes genome-wide histone acetylation patterns. Nat. Commun. 2021 121 *12*, 1–9.

Matsumoto, S., Cavadini, S., Bunker, R.D., Grand, R.S., Potenza, A., Rabl, J., Yamamoto, J., Schenk, A.D., Schübeler, D., Iwai, S., et al. (2019). DNA damage detection in nucleosomes involves DNA register shifting. Nat. 2019 5717763 *571*, 79–84.

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.

McDonel, P., Demmers, J., Tan, D.W.M., Watt, F., and Hendrich, B.D. (2012). Sin3a is essential for the genome integrity and viability of pluripotent cells. Dev. Biol. *363–318*, 62.

McGinty, R.K., and Tan, S. (2015). NucleosomeStructure and Function. Chem. Rev. *115*, 2255.

Michael, A.K., and Thomä, N.H. (2021). Reading the chromatinized genome. Cell *184*, 3599–3611.

Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker, R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R., et al. (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science (80-.). *368*, 1460–1465.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol. Cell *30*, 755–766.

Morgunova, E., and Taipale, J. (2021). Structural insights into the interaction between transcription factors and the nucleosome. Curr. Opin. Struct. Biol. *71*, 171–179.

Murphy, M., Ahn, J., Walker, K.K., Hoffman, W.H., Evans, R.M., Levine, A.J., and George, D.L. (1999). Transcriptional repression by wild-type p53 utilizes histone deacetylases, mediated by interaction with mSin3a. Genes Dev. *13*, 2490.

Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol. *14*, 431–441.

Nagy, L., Kao, H.Y., Chakravarti, D., Lin, R.J., Hassig, C.A., Ayer, D.E., Schreiber, S.L., and Evans, R.M. (1997). Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. Cell *89*, 373–380.

Naruse, Y., Aoki, T., Kojima, T., and Mori, N. (1999). Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. Proc. Natl. Acad. Sci. U. S. A. *96*, 13691–13696.

Nicetto, D., and Zaret, K.S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr. Opin. Genet. Dev. *55*, 1–10.

van Oevelen, C., Bowman, C., Pellegrino, J., Asp, P., Cheng, J., Parisi, F., Micsinai, M., Kluger, Y., Chu, A., Blais, A., et al. (2010). The mammalian Sin3 proteins are required for muscle development and sarcomere specification. Mol. Cell. Biol. *30*, 5686–5697.

Paakinaho, V., Presman, D.M., Ball, D.A., Johnson, T.A., Schiltz, R.L., Levitt, P., Mazza, D., Morisaki, T., Karpova, T.S., and Hager, G.L. (2017). Single-molecule analysis of steroid receptor and cofactor action in living cells. Nat. Commun. *8*.

Pang, Y.P., Kumar, G.A., Zhang, J.S., and Urrutia, R. (2003). Differential binding of Sin3 interacting repressor domains to the PAH2 domain of Sin3A. FEBS Lett. *548*, 108–112.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev. *10*, 1991–2002.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. Cell *132*, 422–433.

Park, S.Y., and Kim, J.S. (2020). A short guide to histone deacetylases including recent progress on class II enzymes. Exp. Mol. Med. 2020 522 *52*, 204–212.

Peña, P. V., Hom, R.A., Hung, T., Lin, H., Kuo, A.J., Wong, R.P.C., Subach, O.M., Champagne, K.S., Zhao, R., Verkhusha, V. V., et al. (2008). Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. J. Mol. Biol. *380*, 303.

Plasschaert, R.N., Vigneau, S., Tempera, I., Gupta, R., Maksimoska, J., Everett, L., Davuluri, R., Mamorstein, R., Lieberman, P.M., Schultz, D., et al. (2014). CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. Nucleic Acids Res. *42*, 774–789.

Ptashne, M. (1967). Specific Binding of the λ Phage Repressor to λ DNA. Nat. 1967 2145085 *214*, 232–234.

Saha, N., Liu, M., Gajan, A., and Pile, L.A. (2016). Genome-wide studies reveal novel and distinct biological pathways regulated by SIN3 isoforms. BMC Genomics *17*, 1–16.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell *132*, 887–898.

Schübeler, D. (2015). Function and information content of DNA methylation. Nat. 2015 5177534 *517*, 321–326.

Schübeler, D., Lorincz, M.C., Cimbora, D.M., Telling, A., Feng, Y.-Q., Bouhassira, E.E., and Groudine, M. (2000). Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. Mol. Cell. Biol. *20*, 9103–9112.

Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat. Biotechnol. *32*, 171–178.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nat. 2006 4427098 *442*, 96–99.

Silverstein, R.A., and Ekwall, K. (2005). Sin3: A flexible regulator of global gene expression and genome stability. Curr. Genet. *47*, 1–17.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*.

Solaimani, P., Wang, F., and Hankinson, O. (2014). SIN3A, Generally Regarded as a Transcriptional Repressor, Is Required for Induction of Gene Transcription by the Aryl Hydrocarbon Receptor. J. Biol. Chem. *289*, 33655–33662.

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.

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.

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.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van 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.

Stein, A., Takasuka, T.E., and Collings, C.K. (2010). Are nucleosome positions in vivo primarily determined by histone–DNA sequence preferences? Nucleic Acids Res. *38*, 709.

Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. Science *368*.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Swinstead, E.E., Miranda, T.B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T.S., Ball, D., Mazza, D., Lavis, L.D., et al. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell *165*, 593–605.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Talbert, P.B., and Henikoff, S. (2021). Histone variants at a glance. J. Cell Sci. *134*.

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.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 2009 104 *10*, 252–263.

Vieyra, D., Loewith, R., Scott, M., Bonnefin, P., Boisvert, F.M., Cheema, P., Pastyryeva, S., Meijer, M., Johnston, R.N., Bazett-Jones, D.P., et al. (2002). Human ING1 proteins differentially regulate histone acetylation. J. Biol. Chem. *277*, 29832–29839.

Viiri, K.M., Korkeamäki, H., Kukkonen, M.K., Nieminen, L.K., Lindfors, K., Peterson, P., Mäki, M., Kainulainen, H., and Lohi, O. (2006). SAP30L interacts with members of the Sin3A corepressor complex and targets Sin3A to the nucleolus. Nucleic Acids Res. *34*, 3288–3298.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. Cell *138*, 1019–1031.

Wapenaar, H., and Dekker, F.J. (2016). Histone acetyltransferases: challenges in targeting bi-substrate enzymes. Clin. Epigenetics *8*, 1–11.

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.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.

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*, 343–349.

Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet. *25*, 434–440.

Xu, B., Wang, H., Wright, S., Hyle, J., Zhang, Y., Shao, Y., Niu, M., Fan, Y., Rosikiewicz, W., Djekidel, M.N., et al. (2021). Acute depletion of CTCF rewires genome-wide chromatin accessibility. Genome Biol. *22*.

Yan, C., Chen, H., and Bai, L. (2018). Systematic Study of Nucleosome-Displacing Factors in Budding Yeast. Mol. Cell *71*, 294-305.e4.

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 (80-.). *356*, eaaj2239.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. *13*, 335–340.

Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Curr. Biol. *20*.

Zencir, S., Dilg, D., Shore, D., and Albert, B. (2022). Pitfalls in using phenanthroline to study the causal relationship between promoter nucleosome acetylation and transcription. Nat. Commun. 2022 131 *13*, 1–4.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat. Struct. Mol. Biol. *16*, 847.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2010). Evidence against a genomic code for nucleosome positioning. Nat. Struct. Mol. Biol. *17*, 920.

Zhong, M., Niu, W., Lu, Z.J., Sarov, M., Murray, J.I., Janette, J., Raha, D., Sheaffer, K.L., Lam, H.Y.K., Preston, E., et al. (2010). Genome-Wide Identification of Binding Sites Defines Distinct Functions for Caenorhabditis elegans PHA-4/FOXA in Development and Environmental Response. PLOS Genet. *6*, e1000848.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., et al. (2018a). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81.

Zhu, F., Zhu, Q., Ye, D., Zhang, Q., Yang, Y., Guo, X., Liu, Z., Jiapaer, Z., Wan, X., Wang, G., et al. (2018b). Sin3a–Tet1 interaction activates gene transcription and is required for embryonic stem cell pluripotency. Nucleic Acids Res. *46*, 6026–6040.