Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht – Karls – University

Heidelberg

Presented by

Valentina Baderna, M.Sc.

Born in Genova, Italy

Oral examination: 1st July 2025

Dissecting genetic and epigenetic contribution to chromatin opening through a chromatin-integrated reporter assay

Referees: Prof. Dr Ingrid Lohmann

Dr Eileen Furlong

I. Acknowledgements

First and foremost, I would like to thank Arnaud for the opportunity to join his lab and for his guidance throughout this journey. I've learned a great deal over the past few years, and I'm grateful for your support and mentorship.

I'd also like to thank the members of my thesis advisory committee – Prof. Dr. Ingrid Lohmann, and Dr. Eileen Furlong – for their valuable feedback and insightful suggestions throughout my PhD. A special thanks to Prof. Dr. Duncan Odom and Dr. Justin Crocker for kindly agreeing to be part of my defense committee.

I am deeply thankful to the GeneCore EMBL facility and team, which made all of my experiments possible. In particular, I want to thank Vladimir Benes for always being helpful and for your practical advice.

To all the past and present members of the lab: thank you for making the lab such a supportive, intellectually stimulating, and fun place to work. I've genuinely enjoyed learning from you, laughing with you, and figuring things out together. A special shout-out to Elisa, Roos, Blanka, Lö, Mathias and Michi – it has been a real pleasure to work alongside you throughout this time.

I would like to thank my friends and everyone I've met over the past few years. Whether in the lab, during a coffee break, or outside of work, your presence has made this time more enjoyable and balanced. Of course, a special thank goes to the kicker group – Silvi, Vale, Nico, Simo, Marcolino, Toni, Babbuccio, Cate (Ila per gli amici) and all the others for all the post-lunch breaks (and more!) spent playing, laughing, and venting together. And Marcolino, thanks for being my German translator "di fiducia", all along. The Furlongers, for always making me feel welcome, sharing cakes with me (and adopting me over lunch, from time to time.

Silvi, grazie per tutto; le chicchiarate, lo shopping di piante, gli aperitivi, le cene, i film, il burrachino coi tuoi. Il tuo esserci sempre. Tiago, for all the Italian-Portuguese nights and adventures. Vale, per essere la mia coach, la mia estetista, pasticciera, anche stilista, e soprattutto una cara, cara amica (insomma, da diesc). Nico, il mio compare di procrastinazione, perché alla fine, tu ci sei sempre (e per il cassette pieno di cibo anche per me).

Mamma e Papà, grazie per tutto l'amore e il sostegno costante in tutti questi anni, senza

di voi questo percorso non sarebbe stato possibile. Grazie per tutte le chiamate e per tutte le volte che siete venuti a trovarci, nonostante la distanza. Vi voglio bene.

Ale, amore mio, grazie. Grazie per il caffè ogni mattina, per farmi iniziare la giornata sempre nel modo giusto. Grazie per credere sempre in me, anche quando magari io non ci credo più molto. Grazie per spronarmi quando sono pigra, e farmi pedalare fino a Königstuhl (nonostante i miei lamenti). Grazue per gli abbracci nei momenti di sconforto, e le lunghe chiacchiarate per aiutarmi a uscirne. Grazie per aver ragionato insieme me quando non riuscivo a smettere di pensare al lavoro, che avessi problemi tecnici o domande biologiche. Grazie per essere la mia persona preferita, il mio sostegno più grande, sempre.

II. Summary

Chromatin accessibility at *cis*-regulatory elements (CREs) is influenced by both *cis*-encoded and *trans*-acting mechanisms. This includes short DNA motifs recognized by transcription factors (TFs), as well as chromatin post-translational modifications on histones (PTMs). A challenge in identifying determinants of chromatin accessibility is that TF binding and deposition of activating chromatin marks are often concomitant at endogenous CREs, making it difficult to disentangle their individual contribution to chromatin accessibility and transcription activation. To overcome this limitation, I established a chromatin-integrated reporter assay to test whether regulatory sequences (~250bp) can recruit TFs, open chromatin, and activate transcription at an ectopic site. I created libraries containing hundreds of regulatory sequences and inserted them at the same landing pad located at a chromatin locus devoid of activating or repressing chromatin marks. I profiled whether these short genetic elements could recapitulate their endogenous chromatin accessibility using a parallelized chromatin accessibility assay based on targeted Single Molecule Footprinting. I analyzed ~400 fragments containing promoters and enhancers active in mouse embryonic stem cells (mESCs), and compared the chromatin accessibility at ectopic sites with their endogenous levels. As expected, CTCF-containing sequences (used as positive control) autonomously re-establish chromatin accessibility at identical levels when inserted at the ectopic locus. In contrast, sequences bound by other TFs, including NRF1, NFY and pluripotency factors such as KLF4, OCT4 show only partial recapitulation of their endogenous accessibility when isolated at the ectopic site; and this reduction was not explained by CpG content alone. To investigate whether sequence context beyond the core ~250 bp CRE could explain reduced accessibility, I performed a fragment-length experiment on a single CRE. Although preliminary, the results suggest that accessibility at the ectopic site reached endogenous levels at ≥1 kb. This suggests that, although the regulatory sequences are genetically the same of their endogenous counterparts, the ~250 bp fragments inserted at the ectopic locus may lack additional sequence features required for full activity – such as histone PTMs, whose deposition depends on cooperative interactions between multiple TFs. To further explore this possibility, I next investigated the role of H3K27Ac and found that the extent to which CREs establish chromatin accessibility is anti-correlated with the presence of H3K27Ac at the endogenous locus, suggesting that H3K27Ac may enhance accessibility in its native context. To test this hypothesis, I globally perturbed acetylation by chemically inhibiting the histone acetyltransferase p300, that deposits acetylation at a subset

of enhancers in mESCs. As predicted, chromatin accessibility was reduced at endogenous loci, and the amplitude of the loss scaled with the degree of H3K27Ac reduction, but almost unaffected at the ectopic locus, which is located within a chromatin neutral environment. Finally, I implemented a proof-of-concept reporter assay to simultaneously quantify chromatin accessibility and transcriptional output for multiple ectopically inserted CREs, demonstrating the feasibility of parallel measurements at scale.

III. Zusammenfassung

Die Zugänglichkeit von Chromatin an cis-regulatorischen Elementen (CREs) wird sowohl durch cis-kodierte als auch trans-wirkende Mechanismen beeinflusst. Dazu gehören kurze DNA-Motive, die von Transkriptionsfaktoren (TFs) erkannt werden, sowie posttranslationale Modifikationen (PTMs) des Chromatins. Eine Herausforderung bei der Identifizierung der Bestimmungsfaktoren für Chromatinzugänglichkeit besteht darin, dass TF-Bindung und die Anlagerung aktivierender Chromatinmarkierungen an endogenen CREs häufig gleichzeitig auftreten, was es schwierig macht, ihren jeweiligen Beitrag zur Chromatinzugänglichkeit und Transkriptionsaktivierung auseinanderzuhalten. Um diese Einschränkung zu überwinden, habe ich einen chromatinintegrierten Reporter-Assay entwickelt, um zu testen, ob regulatorische Sequenzen (~250) Transkriptionsfaktoren rekrutieren, Chromatin öffnen und Transkription an einem ektopischen Ort aktivieren können. Ich habe Bibliotheken mit hunderten von regulatorischen Sequenzen erstellt und sie an derselben Zielstelle eingefügt – einem Chromatin-Locus ohne aktivierende oder reprimierende Chromatinmarkierungen. Mit einem parallelisierten Chromatinzugänglichkeits-Assay basierend auf gezieltem Single Molecule Footprinting habe ich analysiert, ob diese kurzen genetischen Elemente ihre endogene Chromatinzugänglichkeit nachbilden können. Ich habe etwa 400 Fragmente untersucht, die Promotoren und Enhancer enthalten, die in embryonalen Stammzellen der Maus (murine ES-Zellen) aktiv sind, und die Chromatinzugänglichkeit an ektopischen Stellen mit ihren endogenen Niveaus verglichen. Wie erwartet, stellten CTCF-haltige Sequenzen (verwendet als positive Kontrolle) die Chromatinzugänglichkeit autonom auf identischem Niveau wieder her, wenn sie an die ektopische Stelle eingefügt wurden. Im Gegensatz dazu zeigten Sequenzen, die von anderen TFs wie NRF1, NFY und Pluripotenzfaktoren wie KLF4, OCT4 gebunden werden, nur eine teilweise Reaktivierung ihrer endogenen Funktion, wenn sie isoliert an der ektopischen Stelle getestet wurden – dieser Rückgang konnte nicht allein durch den CpG-Gehalt erklärt werden. Um zu untersuchen, ob die reduzierte Zugänglichkeit von CREs über den ~300 bp hinausreichenden Sequenzkontext erklärt werden könnte, habe ich ein Fragmentlängen-Experiment mit einem einzelnen CRE durchgeführt. Trotz vorläufiger Ergebnisse, deuten diese darauf hin, dass die Zugänglichkeit an der ektopischen Stelle bei einer Fragmentlänge von ≥1 kb das endogene Niveau erreichte. Dies legt nahe, dass die ~250 bp langen Fragmente, obwohl genetisch identisch mit ihren endogenen Gegenstücken, möglicherweise zusätzliche Sequenzmerkmale für die volle Aktivität fehlen – etwa Histon-

PTMs, deren Anlagerung von kooperativen Interaktionen mehrerer TFs abhängt. Um diese Möglichkeit weiter zu untersuchen, habe ich die Rolle von H3K27Ac analysiert und festgestellt, dass das Ausmaß, in dem CREs Chromatinzugänglichkeit herstellen, negativ mit der Präsenz von H3K27Ac am endogenen Locus korreliert. Dies deutet darauf hin, dass H3K27Ac in seinem natürlichen Kontext die Zugänglichkeit fördern kann. Um diese Hypothese zu testen, habe ich die Acetylierung global gestört, indem ich die Histonacetyltransferase p300 – die an einem Teil der Enhancer in murine ES-Zellen Acetylierung anbringt – chemisch gehemmt habe. Wie vorhergesagt, nahm die Chromatinzugänglichkeit an endogenen Loci ab, und das Ausmaß des Verlusts korrelierte mit der Stärke der H3K27Ac-Reduktion – war jedoch an der ektopischen Stelle, die sich in einer chromatinneutralen Umgebung befindet, kaum betroffen.

Abschließend habe ich einen Proof-of-Concept-Reporter-Assay entwickelt, um gleichzeitig die Chromatinzugänglichkeit und die Transkriptionsaktivität für mehrere ektopisch eingefügte CREs zu quantifizieren, und damit die Machbarkeit paralleler Messungen im großen Maßstab demonstriert.

Table of Contents

1. INTRODUCTION

Deoxyribonucleic acid (DNA) is the fundamental genetic material that encodes the information required for the development, function, and maintenance of all living organisms. In eukaryotic cells, the DNA is stored in the nucleus, where the information contained within genes – DNA sequences that encode for a function – can be read and turned into functional molecules with specific roles.

Gene expression begins with transcription, the process whereby the genetic information encoded in the DNA is decoded and transformed into RNA. Gene transcription initiates at the transcription start site (TSS) – the first transcribed nucleotide of a gene – which is located in specific DNA sequences called promoters, located at the beginning of the gene (Lenhard et al., 2012). The binding of transcription factors (TFs) to the promoter mediates the recruitment of the general transcription factors (GTFs) and eventually of RNA polymerase II (Pol II), forming a transcription-competent machinery (Figure 1). Once transcribed, the RNA can either carry out its function in its original form – hence referred to as non-coding RNA – or be translated into a protein (Cramer, 2019). Both non-coding RNAs and proteins then perform specific functions related to cellular activity or contribute to the regulation of gene expression itself, ultimately influencing which RNAs and proteins will be produced.

Within the nucleus, DNA is packaged into chromatin through its association with histone proteins. The nucleosome, composed of ~147 base pairs of DNA wrapped around an octamer of histones (H2A, H2B, H3, H4), forms the basic unit of chromatin (Kornberg and Lorch, 1999). The degree of chromatin compaction, and consequently its accessibility to regulators, is a critical determinant of gene expression. Chromatin can exist in different states: euchromatin, which is relatively decondensed and transcriptionally active, and heterochromatin, which is highly compacted and generally associated with transcriptional repression.

The advent of next-generation sequencing (NGS) – which allows to determine the nucleotide sequence of full genomes – revealed that only 1–2% of the mammalian genome encodes proteins, while the remaining 98% is composed of non-protein-coding DNA ("The ENCODE (ENCyclopedia Of DNA Elements) Project," 2004; The ENCODE Project Consortium, 2012). Much of this non-coding material contains regulatory information, like *cis*-regulatory elements (CREs), which play critical roles in regulating gene expression. In addition to promoters, CREs include elements such as enhancers,

silencers, and insulators, all of which function by interacting with specific DNA-binding proteins that, in turn, may associate with other regulatory proteins with distinct functions, commonly referred to as co-factors (Cramer, 2019; Kim and Wysocka, 2023; "The ENCODE (ENCyclopedia Of DNA Elements) Project," 2004; The ENCODE Project Consortium, 2012).

Among regulatory proteins are those responsible for local, reversible, and dynamic modifications of DNA and histone proteins that do not alter the underlying DNA sequence but can be stably inherited through cell divisions (Figure 1). This layer of regulation is known as epigenetics. Although the causal relationship between epigenetic modifications and gene expression is still a matter of debate, these modifications contribute to, and indeed represent, an additional layer of gene expression regulation (Cavalli and Heard, 2019).

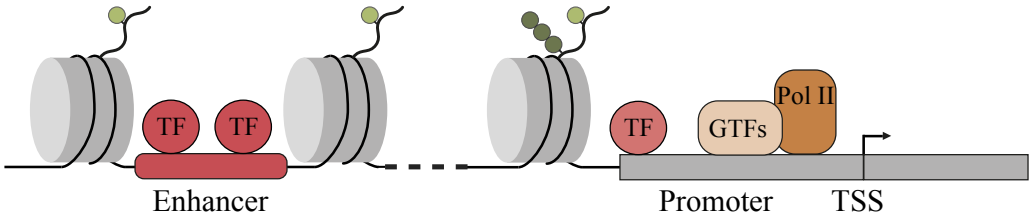


Figure 1. Schematic of gene regulation. Transcription begins at the transcription start site (TSS), located within a promoter region at the 5' end of the gene. Transcription factors (TFs) bind to promoter sequences and recruit general transcription factors (GTFs), which in turn enable the assembly of RNA polymerase II (Pol II) into a transcription-competent complex. Enhancers, bound by other TFs, further modulate the activity of the promoter. Histone post-translational modifications (dark and light green dots) add another layer of regulation. Together, these elements shape complex patterns of gene expression.

Therefore, if we ask the question, "How do cells interpret the DNA sequence to determine when, where, and how much genes should be expressed?" the answer is inherently complex. In eukaryotic cells, a highly interconnected and coordinated regulatory network exists – comprising both *cis*-regulatory elements and *trans*-acting factors – that collectively govern gene expression. This regulatory network ensures that genes are expressed at the correct time, in the appropriate cell types, and at precise quantitative levels. Importantly, this regulation is not restricted to the cellular level: it also orchestrates gene expression across tissues and developmental stages, enabling the coordination of complex physiological functions at the organismal scale (Kim and Wysocka, 2023; Millán-Zambrano et al., 2022).

Because of the scope of this PhD dissertation, I will explore the role of TFs in

regulating gene expression and their interaction with DNA, with a particular focus on *cis*-regulatory elements. I will then examine promoters and enhancers, exploring how they establish specific patterns of gene expression and interact within the three-dimensional nuclear architecture. Finally, I will introduce the concept of epigenetics, with an emphasis on histone post-translational modifications and their importance in gene regulation.

1.1 Interplay between transcription factors and DNA sequence

Transcription factors (TFs) are regulatory proteins that modulate gene expression by binding to defined DNA sequences referred to as transcription factor binding sites (TFBSs) or motifs. These sequences, typically between 6-12 base pairs in length, are recognized through evolutionarily conserved DNA-binding domains (DBDs) within the TF (De Boer and Taipale, 2024). DBDs confer sequence specificity and ensure that TFs bind to the appropriate genomic locations in response to various cues (Lambert et al., 2018). In addition to their DBDs, TFs typically harbor distinct functional domains that modulate the transcriptional output through the recruitment of cofactors. Among these functional domains are activation domains (ADs) and repression domains (RDs), which regulate transcription not by directly interacting with DNA, but by recruiting co-activators (CoAs) and co-repressors (CoRs), respectively (Kim and Wysocka, 2023). Collectively, these functional regions are also referred to as effector domains, as they are responsible for executing the regulatory activity of the TF.

1.1.1 DNA-binding domains

DBDs confer sequence specificity by facilitating the formation of hydrogen bonds between specific amino acid residues in the TF and nucleotide bases in the DNA, resulting in differential binding affinities across potential sites. Common DBD architectures include zinc finger domains, which bind DNA modularly via coordinated zinc ions; helix-turn-helix (HTH) motifs, in which one α-helix inserts into the DNA major groove; leucine zippers, which promote dimerization and transcriptional activation; and helix-loop-helix (HLH) domains, which enable formation of homo- or heterodimers critical for regulatory specificity (Hughes, 2011).

Each TF has an optimal binding sequence associated with the lowest (most favorable)

free energy of binding, alongside a large number of lower-affinity sites that still contribute functionally to regulatory specificity (Crocker et al., 2016). For this reason, TF motifs are not defined by a precise and rigid nucleotide sequence, but rather as position weight matrices (PWMs), where each position in the sequence is associated with the probability of each nucleotide to be present. Although high-throughput *in vitro* assays and comparative genomics have enabled the characterization of motifs for a large fraction of the ~1,600 human TFs, the presence of the motif is alone insufficient to predict binding *in vivo* (Cusanovich et al., 2014; Jolma et al., 2015, 2013; Lambert et al., 2018). Most TFs occupy only a small subset of the motifs present in the genome at a given time, a phenomenon attributable to the chromatin context in which binding occurs. Experimental data confirm that high-affinity sites are relatively rare, whereas many more lower-affinity sites exist across the genome (Badis et al., 2009; Slattery et al., 2011). Importantly, the low-affinity sites have an important physiological relevance. For example, during development it has been shown that they are crucial to encode spatiotemporal gene expression. For instance, in a Pax6-dependent enhancer of the mouse *Neurogenin2* gene, a conserved low-affinity Pax6-binding site is critical for restricting gene expression to regions of high Pax6 concentration. Substitution of this site with a high-affinity variant results in ectopic expression (Scardigli et al., 2003). Similarly, single-nucleotide variants (SNVs) that increase TF binding affinity can disrupt tissue specificity and lead to abnormal phenotypes (i.e., polydactyly) in both mouse and human limbs (Lim et al., 2024). Also, in the marine invertebrate *Ciona*, affinity-enhancing SNVs in a heart enhancer cause ectopic expression of FoxF in non-cardiac cells, resulting in severe heart defects, including the formation of two beating hearts (Jindal et al., 2023). Therefore, TFBSs seem to have evolved to have the appropriate affinity, and what is referred to as 'low-affinity' sites are likely the result of the optimization of their affinity to ensure the proper spatiotemporal pattern of gene expression.

1.1.2 Transcriptional Activation and Repression Domains

In addition to binding their target DNA motifs – which on its own can contribute to gene regulation by increasing DNA accessibility or altering its secondary structure – TFs can also perform additional regulatory functions (Kim and Wysocka, 2023). These functions are mediated through the action of TFs' effector domains and can either activate or repress transcription, by recruiting co-activators (CoAs) or co-repressors (CoRs). Our

knowledge of TF effector domains has been recently extended by a series of high-throughput functional screens that systematically mapped them across thousands of human proteins, resulting in the annotation of over 1,000 domains, of which hundreds were novel (Alerasool et al., 2022; DelRosso et al., 2023; Tycko et al., 2020).

Mechanistically, many of these effector domains contain intrinsically disordered regions (IDRs) – sequences that lack stable secondary or tertiary structure under physiological conditions (Trivedi and Nagarajaram, 2022). Despite their structural flexibility, IDRs can adopt defined conformations upon interaction with binding partners, enabling TFs to engage multiple cofactors. Through these interactions, TFs can recruit both CoAs and CoRs, with the specific combination of recruited factors regulating the transcriptional outcome.

Activation domains promote transcription by recruiting CoAs, such as histone acetyltransferases (e.g., p300/CBP), chromatin remodelers, and the Mediator complex, which collectively enhance chromatin accessibility and facilitate Pol II recruitment and activation at promoters and enhancers (Gillespie et al., 2020; Iurlaro et al., 2021; Narita et al., 2023; Schick et al., 2021). In contrast, repression domains function by engaging CoRs, including histone deacetylases (HDACs), that in turn remove activating marks or add repressive ones, thereby reinforcing a repressive chromatin state and limiting access to the transcriptional machinery (Gillespie et al., 2020; Perissi et al., 2010). Importantly, the regulatory potential of transcriptional cofactors is highly context-dependent. Whereas some cofactors contribute broadly to CRE activation, others are either non-essential or exert their function only in specific regulatory settings (Haberle et al., 2019; Neumayr et al., 2022). Interestingly, some enhancers have been shown to function independently of the Mediator complex, which has traditionally been regarded as a core component of enhancer activity (Neumayr et al., 2022). Overall, the balance between activating and repressive cofactors is central to transcriptional regulation, allowing TFs to fine-tune gene expression in response to various stimuli.

Adding yet another layer of complexity, the biophysical properties of effector domains – particularly IDRs – have been linked to the ability of TFs and cofactors to organize into nuclear condensates or hubs. Many transcriptional regulators – including Pol II, Mediator, p300, and pluripotency factors in mouse embryonic stem cells (mESCs) such as OCT4 – form transient high-concentration foci in the nucleus (Basu et al., 2020; Boehning et al., 2018; Boija et al., 2018; Contreras-Martos et al., 2017; Sabari et al., 2018). These hubs,

enriched at clusters of enhancers (also called super-enhancers) are thought to concentrate TFs, GTFs, and cofactors to promote robust transcription of lineage-defining genes (Cho et al., 2018; Chong et al., 2018; Hnisz et al., 2017). IDRs within TFs and cofactors are essential for this phenomenon, mediating weak, multivalent interactions that facilitate the formation of condensates. These dynamic compartments may enhance transcriptional efficiency by increasing the local concentration of transcriptional machinery at regulatory elements, enabling rapid and sustained gene activation in response to developmental or environmental cues. However, their physiological relevance *in vivo* remains debated and is still under investigation.

1.1.3 Pioneer and cumulative hypothesis

To engage their target binding motifs within chromatin, TFs must deal with the nucleosomal barrier, which imposes steric and energetic constraints that limit access to DNA. Chromatin exists in a continuum of dynamic states – condensed, permissive, and open – each corresponding to varying levels of DNA accessibility to regulatory proteins (Morgunova and Taipale, 2017). Open chromatin represents a configuration in which DNA is exposed and thus permissive to transcriptional activation. In contrast, in the condensed state, nucleosomes restrict access by sterically hindering TF binding and creating physical and energetic barriers to the transcriptional machinery. Permissive chromatin represents an intermediate state in which nucleosome occupancy is sufficiently dynamic to allow initial TF engagement and localized nucleosome remodeling, but full accessibility has not yet been established. In this way, permissive and open chromatin differ in the extent to which DNA is exposed and functionally engaged by the transcriptional machinery. Transitions between these chromatin states are mediated by the coordinated action of sequence-specific TFs and chromatin-associated remodeling complexes (Klemm et al., 2019).

Recent studies have emphasized that chromatin accessibility is not a static feature but a transient state that requires continuous regulatory input to be maintained. Maintenance of open chromatin, particularly at CREs, depends on persistent TF binding, and in many cases, on the activity of ATP-dependent chromatin remodelers. The mammalian SWI/SNF complex has emerged as a key regulator in this context, playing an essential role in restoring local accessibility and counteracting the intrinsic tendency of nucleosomes to reassemble over regulatory DNA. This remodeling activity is critical for

ensuring stable TF occupancy and enabling transcriptional competence (Iurlaro et al., 2021).

In this context, pioneer transcription factors (pioneer TFs) constitute a unique subclass of TFs capable of binding to their recognition motifs within nucleosomal DNA, even in regions of condensed chromatin (Figure 2). These factors can trigger chromatin remodeling and are often required for the initial activation of previously silent regulatory elements. Pioneer activity was first identified *in vivo* through footprinting studies of the *Alb1* enhancer in the developing mouse endoderm, where TFs such as FOXA1 and GATA4 were shown to be able to access their motifs prior to chromatin opening and gene activation (Cirillo et al., 2002; Gualdi et al., 1996). Unlike non-pioneer TFs, which can generally only bind to accessible DNA, pioneer TFs are able to engage with compacted chromatin and facilitate the recruitment of cofactors and additional TFs. Large-scale structural and functional studies have since revealed mechanistic diversity in how these factors interact with nucleosomes, identifying at least five distinct pioneer binding modes (Bulyk et al., 2023). Although high-throughput experiments have shown that a large majority of TFs (209 out of 220 tested) can bind nucleosomal DNA in a cell-free assay, whether this property translates into pioneer activity or leads to chromatin opening *in vivo* remains to be determined (Zhu et al., 2018). Indeed, pioneering activity has so far been confirmed for only a limited number of TFs through orthogonal approaches such as structural or *in vivo* studies and examples include GATA, FOXA1, SOX2, SOX11, SOX19, OCT4 and NANOG (Cirillo et al., 2002; Dodonova et al., 2020; Gualdi et al., 1996; Michael et al., 2023; Pálfy et al., 2020; Veil et al., 2019). Therefore, withing living cells most TFs have only limited chromatin-opening potential when acting alone. Yet, not all CREs are bound by known pioneers factors. In this context, another model proposes that chromatin remodeling at CREs arises from the cooperative and cumulative binding of multiple TFs at closely spaced sites. Such combinatorial binding can produce sufficient chromatin remodeling activity to displace nucleosomes and establish transcriptionally competent chromatin (Barral and Zaret, 2024; Reiter et al., 2017). This model provides an explanation for how enhancer activation can occur in the absence of classical known pioneer factors, pointing to a more flexible and context-dependent logic of chromatin regulation (Figure 2).

Overall, two main models have been proposed to explain how TFs contribute to chromatin opening: the pioneer factor model and the “cumulative” model. These

frameworks offer alternative, yet not mutually exclusive, mechanisms by which TFs can establish chromatin accessibility.

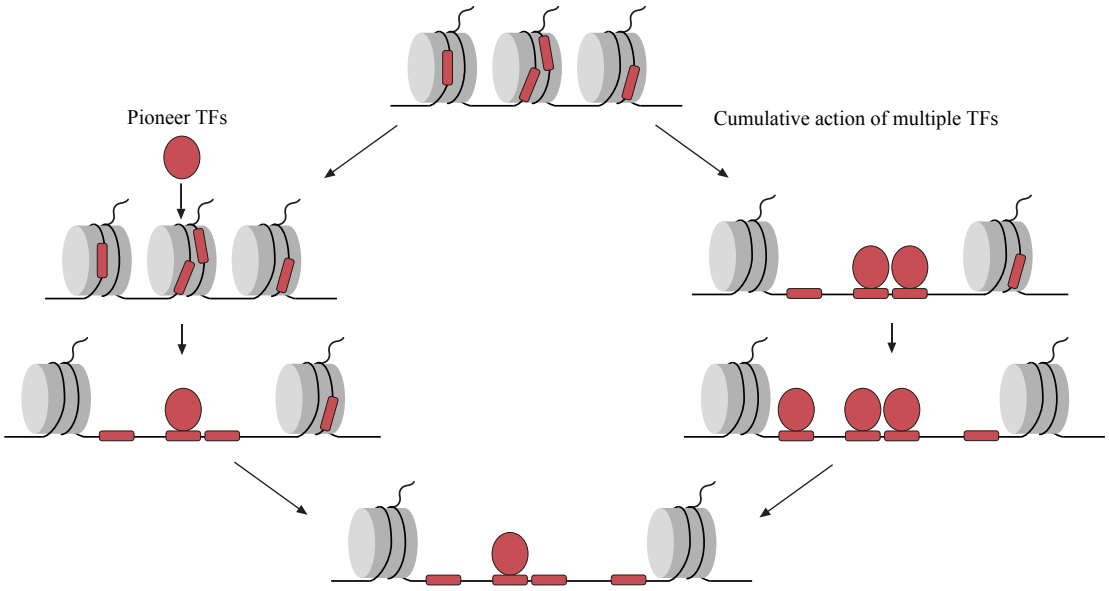


Figure 2. Chromatin opening is mediated by transcription factors (TFs). Nucleosomes act as physical and energetic barriers that limit access to DNA for TFs and other regulatory proteins. Two models explain how this barrier can be overcome by TFs: (left) pioneer TFs are specialized factors capable of directly binding nucleosomal DNA, promoting nucleosome eviction and enabling access to other TFs; (right) most TFs cannot displace nucleosomes alone, but their combined action can collectively overcome the barrier. Red rectangle: TF binding motifs, red circles: TFs.

1.1.4 TF cooperativity and motif syntax

Although important to define which TFs might bind to specific CREs, the mere presence or absence of TF binding motifs it is not sufficient to predict TF binding in the genome or predict the pattern of gene expression (Zeitlinger, 2020). CREs typically contain multiple TF binding sites (TFBSs), and current models postulate that the combinatorial action of many TFs, along with the global arrangement of DNA motifs within a given CRE (also referred to as motif syntax) that recruit TFs to CREs, defines their unique properties and governs how TFs bind and cooperate. Key features of the syntax include motif composition (which motifs are present), orientation (the directionality of each motif), affinity (how strongly each TF binds), order (the sequence in which the motifs appear), and spacing (the distance between motifs) (Jindal and Farley, 2021). These parameters influence how TFs interact with DNA and with each other, ultimately

affecting the regulatory output of an enhancer (Figure 3).

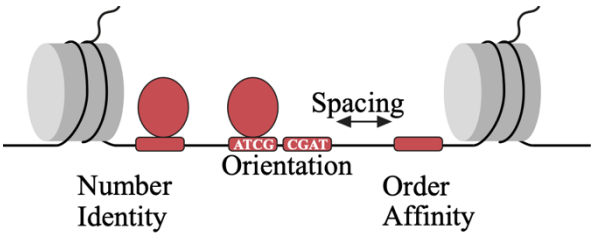


Figure 3. Motif arrangement at CREs. CREs typically contain 4-6 TF binding sites. TF binding and cooperativity can be influenced by the global motif arrangement at the *cis*-regulatory elements (i.e., motif composition, orientation, affinity, order and distance). Red rectangle: TF binding motifs, red circles: TFs.

To explain how TFs cooperate at enhancers, three main models have been proposed (Spitz and Furlong, 2012). The enhanceosome model suggests a rigid organization where TFBSs must follow precise rules of spacing, orientation, and order to function properly, as seen in the *Interferon-β* enhancer (Maniatis, 1995). In contrast, the billboard model proposes a more flexible system, where the presence of the correct TFBSs is sufficient for enhancer function, regardless of their arrangement. The TF-collective model instead presents an intermediate view, where enhancers are activated by a combination of DNA binding and TF-TF interactions, allowing for functional output even without strict motif organization (Spitz and Furlong, 2012). While each of them has been observed, there is still not enough evidence to define a single, generalizable model of CRE grammar, and it is likely that these models are not mutually exclusive. Growing evidence points toward the importance of motif syntax in shaping enhancer function across different contexts. In *Drosophila*, for example, altering the spacing and order of interacting TF motifs in synthetic enhancers can change tissue-specific activity, demonstrating the influence of motif arrangement and distance (Erceg et al., 2014). Also, developmental enhancers often rely on suboptimal motif configurations, suggesting that binding affinity and motif context are more critical than perfect matches to consensus sequences (Crocker et al., 2015; Farley et al., 2015).

More recently, machine learning and deep neural network approaches have been applied to investigate the role of motif syntax in enhancer function. For example, machine learning approaches applied to ChIP-nexus data in mESCs have revealed that TF motif pairs often function optimally with a ~10 bp periodic spacing and directional cooperativity – a principle referred to as "soft syntax" – as observed for NANOG and its binding partners (Avsec et al., 2021). In human cells, massively parallel reporter assays combined with machine learning models revealed that at most enhancers TFs generally

work additively, with limited evidence for strict motif grammar or specific TF–TF interactions, but rather weak TF spacing and orientation preferences (Sahu et al., 2022). Finally, models like DeepSTARR have been developed to learn *cis*-regulatory rules directly from DNA sequence (De Almeida et al., 2022). These models identify both known TF motifs and syntax features such as motif spacing and flanking sequences, which influence enhancer activity. They can accurately predict enhancer strength and guide the design of synthetic enhancers. A recent study extended this approach by combining deep learning with transfer learning to design tissue-specific enhancers in the *Drosophila* embryo. Models trained on single-cell ATAC-seq data and refined with *in vivo* assays predicted 40 synthetic enhancers, 78% of which were active, with 68% showing tissue-specific function, including 100% accuracy for central nervous system and muscle enhancers. These results confirm that motif syntax and TF cooperativity can be used to design functional enhancers with defined regulatory outputs (De Almeida et al., 2024). Similarly, CREsted, an integrated software and deep learning framework for modeling, analyzing, and designing enhancers, was recently developed, starting from single-cell ATAC-seq data. This framework also provides tools to interpret enhancer function by identifying TFBSs, predicting whether TFs act as activators or repressors, and uncovering enhancer grammar. The authors validated CREsted using diverse biological systems, including the mouse cortex, human immune cells, and zebrafish development, and showed that the platform can accurately predict enhancer activity and design cell type-specific synthetic enhancers validated *in vivo* (Kempynck et al., 2025).

Overall, although a universal grammar has yet to be defined, both experimental and computational evidences support the growing importance of motif syntax in shaping how TF bind, interact, and regulate CRE specificity and function across biological systems.

1.2 Gene regulation by promoters and enhancers

Transcription involves the generation of RNA from a DNA template. For protein-coding genes and numerous non-coding transcripts, this is carried out by RNA polymerase II (Pol II). Transcription proceeds through a defined sequence of steps: initiation, promoter-proximal pausing and pause release, productive elongation, and transcriptional termination (Archuleta et al., 2024). Transcription typically initiates at a precise genomic location known as the transcription start site (TSS), located at the 5′ end

of a gene. The TSS resides within the region known as the core promoter, a DNA sequence with a crucial regulatory role typically extending approximately 50 base pairs upstream and downstream of the TSS (Roeder, 1996). This region is responsible for recruiting the transcriptional machinery and orchestrating the early steps of the transcription cycle. Core promoters are, in principle, sufficient to direct basal transcription initiation; however, they typically exhibit low intrinsic activity. In this regard, promoter strength refers to the ability of a promoter to initiate transcription and drive gene expression, which can vary widely depending on its sequence and regulatory context (Haberle and Stark, 2018). In general, stronger promoters lead to higher levels of gene expression, while weaker promoters result in lower expression. The transcriptional output of core promoters can be modulated by proximal or distal regulatory elements, such as enhancer that can increase transcription levels.

1.2.1. Promoter architecture

The architecture of core promoters is highly diverse and context-dependent, defined by the presence or absence of specific sequence motifs such as the TATA box, initiator (Inr), motif ten element (MTE), and downstream promoter element (DPE). These elements are recognized by specific components of the general transcription factors (GTFs), particularly within the TFIID complex and serve as docking platforms for the assembly of the pre-initiation complex (PIC), thereby directing Pol II to the correct transcriptional start site (Vo Ngoc et al., 2017).

In vertebrates, a substantial number of promoters coincide with CpG islands (CGIs)—genomic regions characterized by high GC content and a dense distribution of CpG dinucleotides (Deaton and Bird, 2011). Promoters associated with CGIs generally lack classic core promoter elements and are frequently linked to broadly expressed housekeeping genes or essential developmental regulators. Although the precise mechanisms by which CGIs support core promoter function remain unclear, recent evidence suggests that CpG density itself drives promoter activity, independent of DNA methylation level (Hartl et al., 2019). Even when controlling for TF binding motifs, CpG-rich promoters exhibit stronger TF occupancy and increased transcriptional output. TFs such as NRF1 and GABPA bind their motif when located in CpG-rich regions, suggesting that CpG content not only contributes to chromatin accessibility but also boosts the binding efficiency of TFs to their target sequences (Hartl et al., 2019). Notably,

downstream of the TSS, periodic patterns of A/T-rich dinucleotides have been observed, resembling the sequence preferences of nucleosomal DNA (Haberle et al., 2014). This suggests that promoter nucleotide composition plays a role in organizing local chromatin architecture, particularly at motif-less promoters with broad initiation patterns.

1.2.2 Transcriptional bursting

Rather than being a continuous process, transcription occurs in episodic bursts defined by a frequency and an amplitude, with each burst representing a cluster of engaged RNA polymerases (Rodriguez and Larson, 2020). Live imaging using MS2 and PP7 loops has shown that core promoter motifs – such as the TATA-box, Inr, MTE, and DPE – influence transcriptional bursting. Studies in *Drosophila* embryos have demonstrated that distinct promoter architectures lead to different bursting dynamics (Pimmett et al., 2021). Enhancers also play a key role in regulating transcriptional bursting. Live imaging and quantitative analyses have revealed that enhancers primarily modulate the frequency of transcriptional bursts rather than their amplitude (Bartman et al., 2016; Fukaya, 2023; Fukaya et al., 2016). Notably, in mouse cells the position of an enhancer relative to its target promoter can control bursting frequency (Zuin et al., 2022). When enhancers are placed closer to the promoter, bursting becomes more frequent and uniform, whereas distant enhancers lead to more variable activity. Mathematical modeling suggests that enhancers act by shifting the promoter into a more active state where bursts occur more regularly, thus regulating transcriptional timing and variability at the single-cell level (Tunnermann et al., 2025). However, a recent study also proposes that transcriptional condensates, rather than the distance of the enhancer to the promoter, influences burst activation (Cheng et al., 2023). Finally, also TF cooperativity, by modulating the association probability and exchange of TFs at binding site, can strongly influence burst frequency and duration (Pomp et al., 2024).

1.2.3 Enhancers

First identified in 1981 (Banerji et al., 1981; Moreau et al., 1981), enhancers are segments of non-coding DNA that can be located at varying distances from the genes they regulate, often tens to hundreds of kilobases away in the mammalian genome. In addition, they can reside within intronic regions, upstream or downstream of a gene. Enhancers have initially been defined as regulatory elements capable of activating transcription

independently of their orientation, distance, or position relative to the target gene. More recently, a definition that emphasize their biological context had been introduced: in their native genomic, cellular, and organismal environment, enhancers recruit regulatory proteins to increase the likelihood of transcription at one or more distal genes through a *cis*-regulatory mechanism (Gasperini et al., 2020). In their endogenous setting, enhancers typically harbor multiple TFBSs or motifs, 4-6 on average (Vierstra et al., 2020). Enhancer activation is a multistep process that begins with TFs engaging inaccessible chromatin to make DNA accessible, a prerequisite for subsequent recruitment of cofactors, coactivators, chromatin remodelers, and the general transcription machinery (including Pol II) (Fuda et al., 2009). These interactions determine the activity of an enhancer in a tissue- and time-specific manner, and thus ultimately regulate gene expression patterns.

1.2.4 Enhancer-promoter communication

Enhancers are thought to regulate their target genes by physically interacting with their promoters, or at least by coming into close spatial proximity (Schoenfelder and Fraser, 2019). Since enhancers are often found far away from their target gene(s) in the linear genome, these enhancer–promoter contacts are shaped and constrained by genome topology (Figure 4). Genome topology refers to the three-dimensional (3D) organization of chromatin, as revealed by chromatin conformation capture technologies (among which 3C, 5C, Hi-C, Micro-C) (Dekker et al., 2002; Dostie et al., 2006; Hsieh et al., 2020; Lieberman-Aiden et al., 2009). Chromosomes occupy distinct territories and are organized into compartments, and at a finer scale, into topologically associating domains (TADs) – large, self-interacting regions that are typically conserved across cell types – as well as loops that facilitate regulatory interactions. In vertebrates, TADs and loops are shaped by cohesin-mediated loop extrusion, regulated by the DNA-binding protein CTCF, which anchors boundaries when bound in a convergent orientation (Figure 4) (Ganji et al., 2018; Nora et al., 2017). Depleting CTCF or cohesin disrupts these structures but only minimally affects gene expression (Hsieh et al., 2020; Rao et al., 2017). The interplay between topology and transcription remains complex and context-dependent. While structural loops are often dispensable for steady-state gene expression, certain configurations such as enhancer–promoter loops or gene-gene loops are functionally relevant in disease and development, respectively (Ibrahim and Mundlos, 2020; Levo et

al., 2022).

Although individual enhancers exhibit modular features as they can act as somewhat discrete units – they rarely operate in isolation. Instead, they often engage in non-linear, cooperative interactions among each other and target promoters: multiple enhancers can function redundantly, additively, or even synergistically, depending on cellular context and proximity in the 3D nuclear space (Bolt and Duboule, 2020; Bothma et al., 2015; Furlong and Levine, 2018).

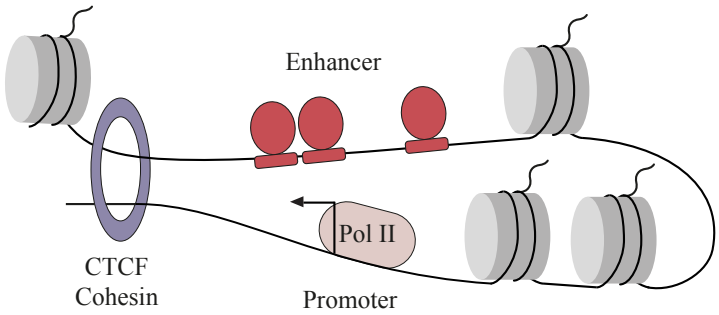


Figure 4. Schematic of an enhancer–promoter loop mediated by cohesin. The cohesin complex extrudes chromatin loops until it encounters CTCF bound to DNA, which acts as a barrier and anchors the loop. Red rectangle: TF binding motifs, red circles: TFs.

1.2.5 Common feature between enhancers and promoters

Enhancers share with promoters a common regulatory architecture, including tightly spaced divergent TSSs, similar frequencies of core promoter motifs, and phased nucleosome positioning around the TSS. However, a key difference lies in both RNA output and transcriptional directionality – while promoters primarily drive unidirectional transcription toward the gene, producing stable transcripts with limited antisense activity, enhancers typically exhibit bidirectional transcription and generate unstable, short-lived transcripts known as eRNAs (Core et al., 2014). The functional distinction between promoters and enhancers is becoming less clear, as many core promoters have been shown to exhibit enhancer activity in various episomal assays across *Drosophila*, mouse, and human cells, as well as in *Drosophila* embryos (Medina-Rivera et al., 2018; Mikhaylichenko et al., 2018; Neumayr et al., 2022; Zabidi et al., 2015) and being able to regulate regulating distant genes (Dao 2017). Also, enhancers can evolve into promoters and promoters into enhancers – reflecting the evolutionary plasticity of these elements (Carelli et al., 2018; Nguyen et al., 2016).

1.2.6 Identification and validation of enhancers

The identification of enhancers remains a major challenge in the field, largely due to their dispersed nature in the non-coding genome and their highly context-dependent and time-restricted activity. To systematically uncover putative enhancers, genome-wide assays have been developed that exploit biochemical features associated with enhancer activity. Techniques such as DNase-seq and ATAC-seq are commonly used to detect accessible chromatin regions, many of which correspond to open or bound enhancers (Minnoye et al., 2021). However, open chromatin does not necessarily indicate active regulatory function; therefore, it is useful to also look at other markers of activity. Indeed, these assays are often complemented by methods like ChIP-seq, that enables to study protein-DNA occupancy genome-wide, for TFs, coactivators, or specific histone modifications such as H3K27ac, which mark active enhancer states (Klemm et al., 2019). However, in multicellular organisms, enhancer discovery is complicated by cell-type heterogeneity, which masks spatiotemporal specificity when using bulk tissue. To overcome this, FACS-sorting strategies (Reddington et al., 2020) or single-cell approaches like scATAC-seq (Calderon et al., 2022; Cusanovich et al., 2018) are increasingly employed to preserve resolution. Enhancers can also be identified based on their transcriptional output, as the presence of eRNAs is often associated with enhancer activity (Mikhaylichenko et al., 2018). Chromatin conformation capture techniques offer an additional strategy by mapping physical contacts between enhancers and promoters (Chen et al., 2024; Pollex et al., 2024). Yet, both eRNAs and physical interactions are not sufficient to infer functionality also require experimental validation. This is typically achieved through reporter-based assays, in which candidate enhancers are placed upstream of minimal promoters driving reporter gene expression, such as massively parallel reporter assays and STARR-seq (Inoue and Ahituv, 2015; Kvon et al., 2014).

1.3 Epigenetic regulation of gene transcription

Epigenetics refers to chromatin modifications that, while reversible and dynamic, can also be inherited and do not alter the underlying DNA sequence. Due to such stability, epigenetic marks serve as a form of molecular memory, preserving gene expression patterns through cell division and ensuring the maintenance of tissue-specific functions. In some cases, these modifications can persist across generations, a phenomenon known as transgenerational epigenetic inheritance (Cavalli and Heard, 2019). Given their central

role in gene regulation, epigenetic mechanisms are essential for fundamental biological processes such as development, differentiation, and aging. Dysregulation of these pathways has been implicated in various diseases, including cancer and neurodegenerative disorders, highlighting their significance in both normal physiology and disease pathogenesis (Cavalli and Heard, 2019).

1.3.1 DNA methylation

Among well-characterized epigenetic modifications in mammals is DNA methylation (5-methylcytosine). It involves the covalent attachment of a methyl group to the fifth carbon of the cytosine ring, typically within CpG dinucleotides, and is commonly linked to gene silencing. It has been shown that it plays critical roles in several biological processes, including genomic imprinting, transposon repression, X chromosome inactivation, and the regulation of germline-specific genes (Greenberg and Bourc'his, 2019; Mattei et al., 2022).

Across species, DNA methylation shows varying degrees of conservation (Klughammer et al., 2023; Mattei et al., 2022). In mammals, however, 70-80% of CpGs are methylated (Meissner et al., 2008), and global DNA methylation dynamics are tightly regulated during embryogenesis and cell differentiation (Greenberg, 2021; Ziller et al., 2013).

DNA methylation is added and maintained by a family of enzymes called DNA methyltransferases (DNMTs). DNMT3A and DNMT3B are mainly responsible for adding new methylation marks during early development, while DNMT1 ensures that these marks are copied during DNA replication. It is a reversible modification that can be removed passively through replication-dependent dilution or actively by the ten-eleven translocation enzymes (including TET1, TET2 and TET3) (Mattei et al., 2022).

DNA methylation can modulate transcription through both direct and indirect mechanisms. Recent findings indicate that at most active enhancers, 5mC is not a determining factor for enhancer function, as TFs are often capable of binding independently of local DNA methylation levels (Kreibich et al., 2023). However, a subset of enhancers appears to be directly regulated by DNA methylation, which can restrict chromatin accessibility and hinder the binding of specific TFs, such as MAX-MYC (Kreibich et al., 2023). These observations align with earlier work showing that

methylation within TF binding motifs can impair TF occupancy, as demonstrated for CREB1 and NRF1 (Domcke et al., 2015). Alternatively, methylation can be interpreted by reader proteins such as the proteins of the Methyl-CpG-binding domain family (e.g., MeCP2, MBD1/2/4), which recruit chromatin-modifying complexes to reinforce repression (Baubec et al., 2013; Nan et al., 1998). However, knockout of all four MBD proteins in mouse ESCs resulted in only minor transcriptional changes, suggesting additional mechanisms mediate methylation-dependent repression (Kaluscha et al., 2022).

In summary, while DNA methylation is a key epigenetic mark associated with transcriptional repression, its role at CREs is context-dependent. Both direct effects on TF binding and indirect mechanisms involving reader proteins contribute to its regulatory function.

1.3.2 Histone post-translational modifications

Another key aspect of epigenetic regulation involves histone post-translational modifications (PTMs), which include acetylation, methylation, phosphorylation and ubiquitination, among others (Millán-Zambrano et al., 2022). These modifications can be present in both the terminal tails of histones and their globular core domains, and are regulated by three main classes of proteins: writers, which catalyze the addition of modifications; erasers, which remove them; and readers, which recognize and interpret specific marks to mediate downstream chromatin-based processes. The presence of several PTMs has been associated with both positive and negative modulation of transcriptional activity (Millán-Zambrano et al., 2022). However, the causal relationship between chromatin marks and transcription remains a central debate in epigenetics. While histone modifications may actively drive gene expression, others may fine-tune, maintain, or merely reflect transcriptional states. Epigenetic marks can act individually or in combination, but the functional consequences of these interactions – whether synergistic, antagonistic, or neutral – are not fully understood. The effects of chromatin modifications are highly context-dependent, varying by cell type, developmental stage, and genomic location (Cavalli and Heard, 2019).

1.3.3 Histone acetylation

Among these PTMs, histone acetylation has been extensively studied due to its strong association with transcriptional activation. It is deposited by histone acetyltransferases (HATs), which transfer an acetyl group from acetyl-CoA to lysine residues on histone tails (Lee and Workman, 2007). The discovery of histone acetylation dates back to the 1960s, when Allfrey et al. observed that acetylated histones reduced transcriptional repression in calf nuclei preparations (Allfrey et al., 1964). The first HAT was later identified in *Tetrahymena thermophila* macronuclei by Brownell and Allis in 1995, highlighting a conserved role for histone acetylation in chromatin regulation (Brownell and Allis, 1995). Acetylation neutralizes the positive charge of lysines, weakening histone-DNA interactions and promoting a chromatin environment permissive to transcription. Conversely, histone deacetylases (HDACs) remove acetyl groups, restoring chromatin compaction and repressing gene expression (Shahbazian and Grunstein, 2007). The balance between HAT and HDAC activity determines chromatin accessibility and is critical for cellular identity and response to environmental cues.

HATs are categorized into three major families: GNAT, MYST, and p300/CBP, each defined by distinct domain architectures and substrate specificities (Avvakumov and Côté, 2007; Lee and Workman, 2007; Liu et al., 2008; Marmorstein, 2001; Shahbazian and Grunstein, 2007). Members of the GNAT family contain a conserved ~160 amino acid HAT domain and a C-terminal bromodomain, which recognizes acetylated lysine residues. GNAT HATs are primarily involved in transcriptional activation and acetylate histones as well as non-histone proteins (Marmorstein, 2001). The MYST family, possesses a larger ~250 amino acid HAT domain, a zinc-finger binding domain, and an N-terminal chromodomain that interacts with methylated lysines, linking acetylation with other chromatin modifications (Avvakumov and Côté, 2007). The p300/CBP family, which includes the transcriptional coactivators p300 and CBP, features an even larger ~500 amino acid HAT domain, in addition to bromodomains and cysteine-histidine-rich motifs (TAZ, PHD, and ZZ domains) that facilitate protein-protein interactions and contribute to substrate specificity.

p300/CBP acts as a central integrator of transcriptional signals and is recruited by TFs to deposit acetylation at enhancers and promoters (Liu et al., 2008). Single-molecule tracking in human cells revealed that p300 binding to chromatin is not driven by individual TFs but requires the combinatorial input of multiple TF interaction domains.

Interestingly, while p300's presence depends on these multivalent TF assemblies, its enzymatic activity may actually reduce association with chromatin (Ferrie et al., 2024). Chemical inhibition studies have shown that p300 is involved in both RNA Pol II recruitment and transcriptional pause release (Narita et al., 2021). Co-condensation with TFs further links p300 to the regulation of transcriptional bursting and coordinated gene expression (Ma et al., 2021). Together, these findings indicate that p300 contributes to multiple stages of the transcription cycle. Interestingly, beyond its well-established activating role, p300 has also been implicated in gene silencing through its interaction with Polycomb complexes (Hunt et al., 2022).

Bromodomains (BRDs) are a class of evolutionarily conserved protein modules that function as readers of lysine acetylation, thereby translating histone acetylation into downstream regulatory outcomes. First identified in *Drosophila*, BRDs are found in a wide array of chromatin-associated proteins, HATs, chromatin remodeling complexes and transcription initiation factors (e.g., TAF1). By selectively recognizing acetylated lysine residues on histone tails, BRDs contribute to the precise targeting and activity of the enzymatic complexes they reside in. An example is the SWI/SNF ATP-dependent chromatin remodeler, whose catalytic subunits – SMARCA2 and SMARCA4 – harbor bromodomains, linking acetylation to nucleosome repositioning and chromatin accessibility regulation (Marmorstein and Zhou, 2014). Among BRD-containing factors, the BET (bromodomain and extra-terminal) family has emerged as a key player in transcriptional regulation. Comprising BRD2, BRD3, BRD4, and the testis-specific BRDT, BET proteins contain two conserved tandem bromodomains (BD1 and BD2) at the N-terminal that bind acetylated histones and an extra-terminal domain that mediates interactions with transcriptional co-regulators (Josling et al., 2012; Wang et al., 2023). BRD4, in particular, plays a central role in promoting transcriptional elongation by recruiting the P-TEFb complex to release paused RNA Polymerase II and by interacting with Mediator subunits to coordinate enhancer-driven gene expression (Bhagwat et al., 2016; Patel et al., 2013).

The best characterized histone acetylation mark is H3K27ac, deposited by the acetyltransferases p300 and CBP (Liu et al., 2008). It serves as a key marker of active enhancers and promoters, often co-occurring with H3K4me1 at enhancer elements (Preissl et al., 2023). Although widely used as a predictive marker for active enhancers, the causal role of H3K27ac in enhancer activation remains under debate. Targeting

transcriptional activators – such as p300 or CBP – to enhancers, leveraging CRISPR-based method, has been shown to induce the production of eRNAs, which is typically associated with increased mRNA expression from nearby promoters (Wang et al., 2022). However, in some instances, transcriptional upregulation has occurred without a corresponding increase in H3K27ac levels (Wang et al., 2022), suggesting that acetylation may not always be required for enhancer-driven transcription.

Recent work in mESCs demonstrated that the presence of H3K27ac lowers the concentration threshold of TFs required to induce gene expression during differentiation, supporting a model in which acetylation modulates the responsiveness of enhancers to TF input (Naqvi et al., 2023). Expanding on this, a recent programmable epigenome editing system enabled systematic deposition of individual or combinatorial chromatin modifications at defined loci in mESCs. When combined with single-cell expression profiling, this approach revealed that the transcriptional output driven by *de novo* H3K27ac is highly context-dependent. In some context, acetylation promoted transcription, while in others it had limited or no effect. These outcomes were influenced by the presence of specific TF motifs and likely reflect cooperativity between chromatin modifications and TF binding (Policarpi et al., 2024).

At the mechanistic level, recent cryo-electron microscopy studies revealed that OCT4 binding induces structural changes in nucleosomes, repositioning DNA to expose additional TF binding sites thereby facilitating the cooperative binding of OCT4 and downstream factors. H3K27 acetylation enhances this cooperativity by modulating interactions between OCT4's acidic patch and the histone H3 tails, resulting in altered DNA positioning and increased accessibility (Sinha et al., 2023). Additionally, inhibition of histone deacetylases (HDACs) has been shown to increase chromatin accessibility, particularly at retrotransposon regions, with increased TF binding – including factors such as YY1, MAX, and GABPA (Cusack et al., 2020).

While H3K27ac remains the most extensively characterized acetylation mark, it is not the only histone modification linked to active chromatin states. Other acetylated residues, such as H3K9ac and H3K14ac, also play important roles in shaping transcriptionally permissive chromatin environments (Gates et al., 2017; Karmodiya et al., 2012). These modifications, catalyzed respectively by histone acetyltransferases GCN5 and PCAF, are commonly enriched at active promoters and enhancers. Consistently, their presence has been strongly associated with elevated transcriptional output (Gates et al., 2017;

Karmodiya et al., 2012).

1.3.4 Histone methylation

Another well-characterized histone PTM is methylation. Histone methylation involves the covalent addition of one, two, or three methyl groups to lysine or arginine residues on histone tails, a process catalyzed by histone methyltransferases (HMTs) and reversed by histone demethylases (HDMs). Histone methyltransferases themselves are diverse in structure and specificity. They are typically classified into two main families: the SET-domain-containing enzymes, which include the majority of known HMTs such as SUV39H1 and members of the MLL family; and the non-SET-domain enzymes, composed of DOT1 and DOT1-like proteins, which uniquely methylates H3K79 within the globular domain of histone H3. Notably, HMTs generally exhibit high substrate specificity – for example, EZH2 methylates H3K27, while DOT1L exclusively targets H3K79 (Kizer et al., 2005; Piunti and Shilatifard, 2021).

The functional outcome of histone methylation is highly context-dependent and varies based on the specific residue modified, the degree of methylation, and the chromatin environment in which it occurs. For instance, H3K4me3 is enriched at the transcription start sites (TSS) of active promoters and strongly correlates with transcriptional activity (Benayoun et al., 2014; Talbert et al., 2019). However, increasing evidence suggests that H3K4me3 may be more a consequence than a cause of transcription, participating in a feedback loop that facilitates Pol II pause-release and elongation (Hu et al., 2017; Wang et al., 2023). Importantly, H3K4me3 can persist through transcriptionally inactive states, serving as a molecular memory of previous transcriptional activity (Millán-Zambrano et al., 2022).

In contrast, the mono-methylated form of the same residue, H3K4me1, is predominantly found at enhancers. H3K4me1 is deposited by MLL3 and MLL4 and is present at active enhancers. It has been proposed to prime enhancers for future activation rather than directly modulating enhancer activity (Calo and Wysocka, 2013; Dorighi et al., 2017). Indeed, recent findings revealed that MLL3 and MLL4's enhancer-binding function is separable from its methyltransferase activity, and while H3K4me1 levels are reduced without catalysis, gene expression remains largely unaffected. Other histone methylation marks such as H3K27me3 and H3K9me3 are associated with transcriptional

repression. EZH2, the catalytic subunit of Polycomb Repressive Complex 2 (PRC2), catalyzes the addition of the H3K27me3 mark, which mediates transcriptional silencing by recruiting PRC1 and promoting downstream chromatin compaction (Cao et al., 2002; Guo et al., 2021). In addition to its roles in developmental gene repression and non-canonical imprinting, this modification also supports the stable inheritance of silenced chromatin states across cell divisions (Escobar et al., 2019; Inoue et al., 2017a). Likewise, H3K9me3 marks constitutive heterochromatin and contributes to gene silencing by recruiting heterochromatin protein 1 (HP1), which drives chromatin compaction and restricts TF binding (Allshire and Madhani, 2018; Methot et al., 2021).

Although traditionally considered a relatively stable epigenetic mark, histone methylation displays greater plasticity than previously appreciated, particularly during development, cellular differentiation, and environmental responses. In embryonic stem cells, for example, bivalent loci marked at the same time by H3K4me3 and H3K27me3 poise developmental genes for activation or repression depending on lineage cues (Bernstein et al., 2006). Additionally, H3K36me3, deposited by SETD2 and localized to gene bodies of actively transcribed genes, has emerged as a key player in transcriptional elongation and co-transcriptional processes (Sun et al., 2020).

Taken together, histone modifications – including both acetylation and methylation – constitutes a highly versatile and context-specific mechanism for the epigenetic regulation of gene expression, contributing to both the activation and silencing of genes with temporal, spatial, and functional precision.

1.4 Studying how chromatin context shape cis-regulatory elements activity and gene expression at scale

A central hypothesis in gene regulation is that the activity of CREs, such as enhancers and promoters, is not solely determined by their DNA sequence but is significantly modulated by the chromatin context in which they reside. To rigorously test this hypothesis, and in addition to the CRISPR-base methods described in the previous sections, recent studies have employed high-throughput reporter assays that enable systematic integration of thousands of CREs across the genome. For example, Akhtar et al. developed the Thousands of Reporters Integrated in Parallel (TRIP) method to assess

position effects by randomly integrating barcoded reporters in the mESCs, revealing that gene expression varied by over three orders of magnitude depending on genomic location (Akhtar et al., 2013). Similarly, Maricque et al. used a lentiviral Massively Parallel Reporter Assay (MPRA) approach to show that identical regulatory sequences exhibit dramatically different activity levels when inserted into distinct chromatin environments, with open, transcriptionally permissive regions enhancing activity and heterochromatic domains suppressing it (Maricque et al., 2019). Complementing these findings, Hong and Cohen demonstrated that core promoters maintain their relative activity across genomic locations, but their absolute output is scaled multiplicatively by the chromatin environment, supporting a modular model of gene regulation (Hong and Cohen, 2022). Together, these studies conclusively demonstrate that chromatin context exerts a profound and consistent influence on CRE activity and gene expression. However, while these approaches offer genome-wide insights into how chromatin influences transcriptional output, they are limited in resolution. In particular, they cannot directly assess changes in accessibility and TF occupancy at the inserted CREs in a context-dependent manner, nor how these changes relate to transcriptional activity. Therefore, although these studies highlight the importance of chromatin context, they also underscore the need for assays that can integrate this information to fully uncover the mechanistic basis of CRE function *in vivo*.

1.5 Measuring chromatin accessibility

Chromatin accessibility refers to the extent to which DNA is exposed and available for interaction with regulatory proteins such as TFs. Accessible chromatin regions are generally free of nucleosomes, while inaccessible regions are often occupied by nucleosomes or other DNA-binding proteins. However, even within a homogeneous cell population, chromatin accessibility can vary from cell to cell at the same genomic locus, a feature that I will refer to as heterogeneity of accessibility. For example, an enhancer may be open and accessible in some cells, yet remain nucleosome-bound and inaccessible in others (Figure 5). This cell-to-cell variability complicates the study of accessibility-dependent gene regulation and the interactions between regulatory factors. Moreover, this can change over time, depending on the biological context.

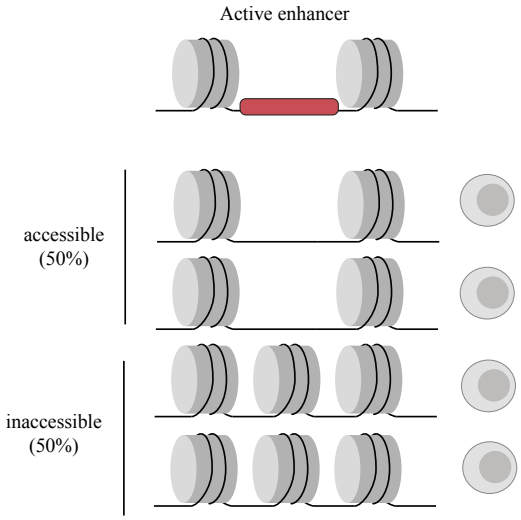


Figure 5. Accessibility heterogeneity. In an homogenous cell population at any given time, the same enhancer could be either in an open or a closed state, implicating that chromatin accessibility is inherently heterogenous.

To study accessibility across the genome, several experimental approaches have been developed (Minnoye et al., 2021). Bulk methods like DNase I hypersensitive site sequencing (DNase-seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq) are widely used (Figure 6). These methods rely on enzymatic cleavage of accessible DNA. DNase-seq uses the endonuclease DNase I, that preferentially cuts DNA in regions of open chromatin, generating fragments that can be sequenced to map accessible sites genome-wide. ATAC-seq instead uses a hyperactive Tn5 transposase that simultaneously fragments DNA and inserts sequencing adapters into accessible regions in a single step (Minnoye et al., 2021). Although powerful for identifying accessible sites across many cells, these techniques average chromatin signals across thousands or millions of cells, masking heterogeneity of accessibility at individual CREs (Figure 6).

Single-cell adaptations of these methods, such as single-cell ATAC-seq, have been developed (Klemm et al., 2019). These can detect differences in accessibility between individual cells, revealing that a significant fraction of accessible regions can vary from one cell to another. However, these methods face a major technical challenge: data sparsity. Since each cell provides only a limited number of sequencing reads, many regulatory regions may not be captured at all, making it difficult to distinguish whether a site is truly inaccessible or simply not captured (Figure 7).

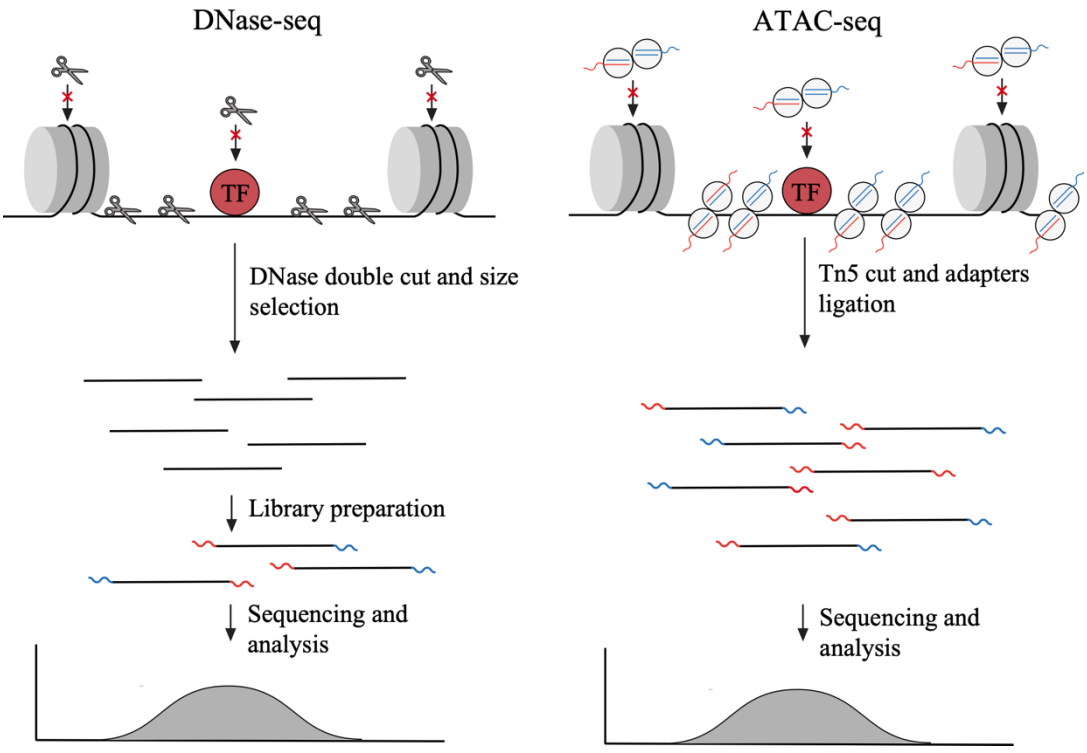


Figure 6. Schematic of bulk DNase-seq and ATAC-seq. DNase-seq uses the DNase I enzyme to double cut DNA at open chromatin regions, while protein-bound to the DNA block the enzyme's access for cleavage (indicated by red crosses). After digestion, accessible regions are size selected and further processed for library preparation and sequencing. The result is a bulk signal enriched at open chromatin regions, known as DNase hypersensitive sites (DHS). ATAC-seq uses a hyperactive Tn5 transposase to cut accessible DNA and insert sequencing adapters in one step. The result is a bulk signal enriched at open chromatin regions.

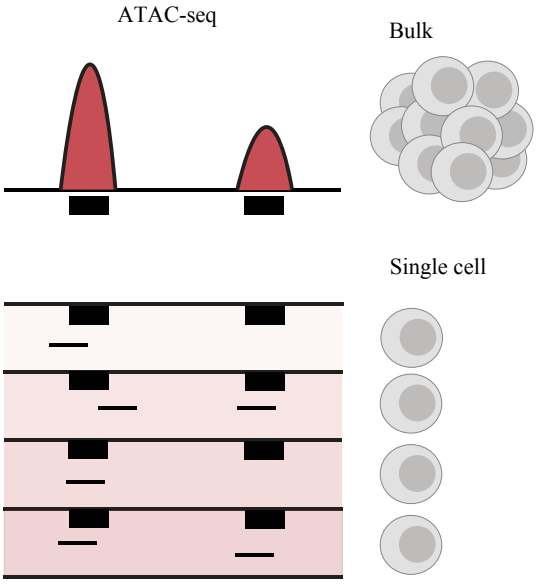


Figure 7. Schematic confronting bulk-assay and single-cell assay. Bulk assays**,** such as ATAC-seq, provide a relative enrichment of chromatin accessibility at a given locus, representing the average signal across a cell population. However, this approach cannot capture accessibility heterogeneity**.** Single-cell assays can inherently detect variability across a cell population. However, due to data sparsity, they are also limited in detecting accessibility heterogeneity, as it is impossible to distinguish false negatives from true negatives. CREs are depicted as black rectangle. Individual reads in the single-cell ATAC-seq are depicted as short black line.

Single-molecule approaches – such as advanced microscopy and genomic technologies – offer a complementary strategy to study gene regulation at single-molecule resolution (Krebs, 2021; Nguyen et al., 2023). Among those, Single Molecule Footprinting (SMF), provide a more direct and quantitative view of chromatin accessibility (Baderna et al., 2025; Krebs et al., 2017, p. 20; Kreibich et al., 2023; Sönmezer et al., 2021). SMF uses recombinant methyltransferases to label accessible cytosines on DNA. After bisulfite treatment, which converts unmethylated cytosines to thymines, sequencing reveals which cytosines were accessible and methylated (Kleinendorst et al., 2021). Importantly, SMF sequences all DNA molecules regardless of their accessibility state, capturing both open and nucleosome-bound regions (Figure 8). This allows for the detection of individual protein-DNA interactions and provides a base-pair resolution map (~7-14bp) of chromatin occupancy. Unlike bulk or single-cell methods, SMF does not suffer from signal averaging or sparse coverage, making it suited to measure heterogeneity of accessibility and TF binding across a population of cells at the single-molecule level (Figure 8).

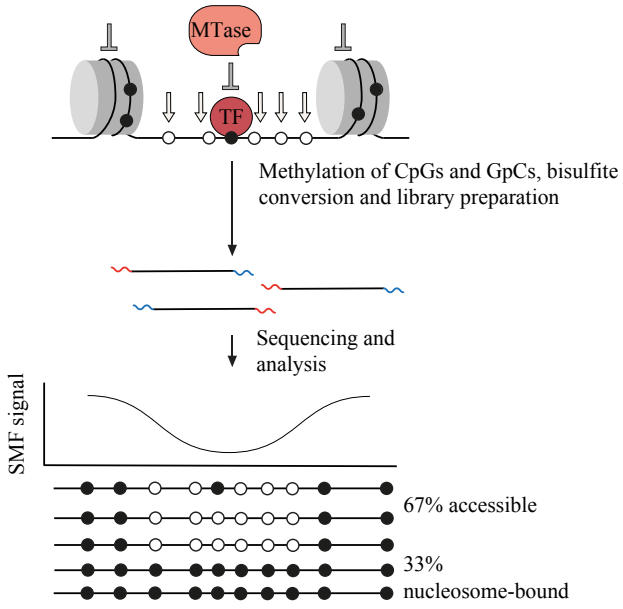


Figure 8. Schematic of single molecule footprinting. Starting from a homogenous pool of cells, nuclei are isolated and exposed to exogenous methyltransferases that methylates all accessible cytosines (not occupied by either a protein, such as a transcription factor (TF) or a nucleosome) in a GpC and CpG context. In the schematic, white and black circles represent methylated and unmethylated cytosines, respectively. Methylation is then detected by bisulfite conversion (unmethylated cytosines are converted to thymines) and sequencing, allowing inference and quantification of multiple footprints at single molecule resolution. At the bottom, shown is the SMF signal, plotted as the inverse frequency of bulk methylation (1-methylation (%)). All molecular states are captured enabling precise quantification of accessibility across the cell population – in the shown example, 67% of molecules were accessible and 33% nucleosome-bound.

1.6 AIMs

Chromatin accessibility at *cis*-regulatory elements (CREs) is influenced by both *cis*-encoded and *trans*-acting mechanisms. This includes short DNA motifs recognized by transcription factors (TFs), as well as chromatin post-translational modifications. Despite decades of research, dissecting the relative contribution of these regulatory features remains challenging. A key difficulty lies in the concurrent binding of TFs and the deposition of activating chromatin modifications at endogenous CREs, complicating efforts to distinguish their specific roles in chromatin accessibility and transcriptional activation.

To address this fundamental question, my PhD project involved the development of a chromatin-integrated reporter assay leveraging Recombination Mediated Cassette Exchange (RMCE) to insert CREs into predefined genomic landing pads where the influence of chromatin modification is minimal. Combined with Single Molecule Footprinting (SMF), this approach minimizes confounding genomic context effects, enabling the isolation of TF binding effects and chromatin context on chromatin opening. I investigated the intrinsic chromatin-opening potential of ubiquitously expressed transcription factors (TFs), which typically bind promoters, and pluripotency TFs, which often act at cell type–specific enhancers. I then tested whether this potential depends on promoter features like CpG content and chromatin marks such as H3K4me3 and H3K27ac.

2. MATERIALS AND METHODS

2.1 Experimental methods

2.1.1 Cell culture

Mouse embryonic stem cells (mESCs) were cultured on dishes coated with 0.2% gelatin (Sigma, #G-1890) in standard ES medium, consisting of Dulbecco's Modified Eagle Medium (DMEM), high glucose (Gibco, #41965039) supplemented with 15% FBS Embryomax (Millipore, #ES-009-B), 20 µg/ml leukemia inhibitory factor (LIF, prepared in house), 50 µM 2-mercaptoethanol (Merck, #M6250), 2 mM L-glutamine (Gibco, #A2916801), and 1× MEM Non-Essential Amino Acids (Gibco, t#11140050). Cells were maintained at 37 °C, with 5% CO2, with daily medium changes and passaged every other day.

2.1.2 Cell lines

159 DNMT triple knockout (TKO) lines (Domcke et al., 2015), were used to generate the reference bait-capture SMF dataset (Sönmezer et al., 2021) and ChIP-seq dataset upon p300 inhibition. TC-1 TKO mESCs with an RMCE landing pad (129S6/SvEvTac background; originally from A. Dean, NIH) were used for all the RMCE experiments.

2.1.3 Cell treatment

p300 complex inhibition was achieved by treating cells with 3 µM A-485 (Selleck Chemicals, #S8740), by adding the chemical in ES medium for 24 hours. DMSO was use as treatment control.

2.1.4 Generation of Dnmt1/3a/3b triple knockouts TC-1 line

Dnmt1/3a/3b triple knockouts (TKO) were produced in TC-1 using previously published guide RNAs (Domcke et al., 2015). Guide RNAs were cloned into the pX459 plasmid (Addgene #48139) and transfected into the target cells. Following puromycin selection, the resulting cell pool was screened for DNMT3a/b knockout using PCR combined with Restriction Fragment Length Polymorphism (RFLP) analysis and HpyCH4V digestion. Ninety-six individual clones were isolated and further screened

using the same approach. Confirmed triple mutants were further validated through amplicon sequencing, using the primers listed below. For the TKO clone used in the experiments described in this dissertation, digestion with a DNA methylation-sensitive restriction enzyme was used to confirm the absence of DNA methylation. Primers used as follow:

DNMT1_fwd: GCAGGAGGGACACAGTCATT (Tm: 63°C)

DNMT1_rev: AGGACTGCAACGTGCTTCTT

DNMT3a_fwd: TTGGCACCTCCCAGGGTTC (Tm: 63.6°C)

DNMT3a_rev: TCGGACAGTGAGTGGTGAGG

DNMT3b_fwd: CCCCCAGAGTCCAGTTCTTA (Tm: 64°C)

DNMT3b_rev: GGAAGGACCGGGCCTGAG

The generation of this TKO line was performed by Rozemarijn Kleinendorst.

2.1.5 Single Molecule Footprinting (SMF)

SMF was carried out as previously described (Kleinendorst et al., 2021). In brief, mESCs were trypsinized, centrifuged at 314 × g for 5 minutes, and washed once with cold PBS. For each SMF reaction, 0.25 × 10^6 cells were used. Cell pellets were lysed by resuspension in 1 ml of ice-cold lysis buffer (10 mM Tris-HCl, pH 7.4–7.6; 10 mM NaCl; 3 mM $MgCl_2$; 0.1 mM EDTA; 0.5% IGEPAL CA-630) and incubated on ice for 10 minutes. Nuclei were pelleted by centrifugation at 1,000 × g for 5 minutes at 4 °C, washed with 250 µl of ice-cold wash buffer (10 mM Tris (pH 7.4–7.6), 10 mM NaCl, 3 mM MgCl 2, 0.1 mM EDTA), and centrifuged again under the same conditions. The nuclei were resuspended in 94.5 µl of 1× M.GpC buffer and kept on ice.

GpC methylation was initiated by adding 150 µl of freshly prepared GpC methyltransferase mix (1×M.GpC buffer, 300 mM sucrose, 64 µM S-adenosylmethionine (SAM – NEB, #B9003S) and 50 µl of M.CviPI (NEB, #M0227L). The reaction was incubated at 37 °C for 7.5 minutes. Subsequently, 25 µl of additional M.CviPI and 4 µl SAM were added, and the reaction was continued for another 7.5 minutes at 37 °C.

For CpG methylation, 3.5 µl of 1 M $MgCl_2$, 15 µl of M.SssI, and 4 µl SAM were added sequentially, followed by a 7.5-minute incubation at 37 °C. The reaction was then stopped

with 300 µl of prewarmed stop solution (20 mM Tris-HCl pH 7.9, 600 mM NaCl, 1% SDS, 10 mM EDTA) and 6 µl of proteinase K. Samples were briefly vortexed and incubated overnight at 55 °C.

The following day, DNA was extracted using phenol–chloroform. To remove RNA contamination, 1 µl of RNase A (10 mg/ml) was added, and samples were incubated at 37 °C for 30 minutes. DNA concentration was determined using the Qubit 1X dsDNA High Sensitivity Assay, and purity was assessed using a Nanodrop spectrophotometer.

Library preparation and targeted enrichment of *cis*-regulatory elements were carried out using the SureSelect XT Mouse Methyl-Seq Kit (Agilent Technologies, version E0, April 2018, #G9651A), following the protocol detailed in Kleinendorst et al. (2021a). Enriched DNA was bisulfite converted using the EZ DNA Methylation-Gold Kit (Zymo, #D5005), followed by PCR amplification and indexing with components from the SureSelect XT Mouse Methyl-Seq Kit (Agilent, #G9651A). Final libraries were sequenced on an Illumina NextSeq500 platform using 150 bp paired-end reads.

For targeted SMF at the ectopic locus**,** genomic DNA was extracted and subjected to bisulfite conversion using the Qiagen Epitect Bisulfite Kit (Qiagen, #5910), following the manufacturer's instructions. Targeted PCR amplification was carried out using KAPA HiFi Uracil+ polymerase (Roche, #KK2802) with the following thermal cycling protocol: 95 °C for 4 minutes; 35 cycles of 98 °C for 20 seconds, 60 °C for 15 seconds, and 72 °C for 20 seconds; followed by a final extension at 75 °C for 5 minutes. Amplified DNA was purified using AMPure XP beads (Beckman Coulter), and libraries were prepared using the (New England Biolabs, #E7645L). Sequencing was performed on an Illumina MiSeq platform using 250 bp paired-end reads.

For the genome-wide p300 inhibition experiments, I performed the chemical inhibition using A-485 (and DMSO as control) in the cells and performed the Single Molecule Footprinting procedure. After, Rozemarijn Kleinendorst processed the sample from DNA extraction to library preparation and sequencing submission.

2.1.6 Cross-link ChIP-seq

Cross-link ChIP-seq was carried out as previously described (Trovato et al., 2024). In brief, for cell fixation, 10 million mESCs were cross-linked in 3 ml of pre-warmed ES medium containing 1% formaldehyde () for 10 minutes at room temperature. Cross-

linking was quenched with glycine to a final concentration of 125 mM, followed by a 10-minute incubation at room temperature. Cells were washed twice with ice-cold PBS supplemented with 10% FBS (Millipore, #ES-009-B), centrifuged at 200 × g for 5 minutes at 4 °C. After, fixed cell pellets were resuspended in 300 µl of Sonication Buffer 1 (50 mM Tris-HCl, pH 8.0; 0.5% SDS) and sonicated using a Bioruptor Pico for 15 cycles (30 s ON / 30 s OFF).

Sonicated chromatin was diluted to a final volume of 1.8 ml using Lysis Buffer (10 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1% Triton X-100; 1 mM EDTA; 0.5 mM EGTA; 0.1% sodium deoxycholate; 0.5% N-lauroylsarcosine) supplemented with 1× protease inhibitor cocktail (PIC, Merck #11836170001). Samples were centrifuged at full speed for 10 minutes at 4 °C, and the supernatants were transferred to fresh tubes and stored on ice or frozen in 10% glycerol. A 2.5% aliquot of the supernatant was reserved as input, and fragmentation (~150–500 bp) was assessed on a 2% agarose gel.

For immunoprecipitation (IP), 30 µl of Protein G beads (Thermo Fisher, #10003D) were prepared per IP by washing twice with 1 ml PBS-T (PBS + 0.01% Tween-20) at 4 °C, each for 1 minute with rotation. Beads were resuspended in PBS-T and incubated for 1 hour at room temperature with the H3K27ac antibody (Abcam, ab4729) on a rotator. Antibody-bound beads were washed once with PBS-T and twice with Lysis Buffer (supplemented with PIC). For each IP, 15 µg of chromatin were used. The antibody-bead complexes were added to chromatin, after resuspension in 30 µl Lysis Buffer and incubated overnight on a rotator at 4 °C. For normalization purposes, 1 µg of *Drosophila* Schneider 2 cell chromatin was added to each IP as spike-in. Of note, *Drosophila* chromatin was obtained with the same protocol described for mESCs.

The next day, immunoprecipitated complexes were washed twice for 5 minutes each with 1 ml of the following buffers: RIPA (10 mM Tris-HCl pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 1 mM EDTA, 140 mM NaCl), RIPA High Salt (same composition with 360 mM NaCl), and LiCl Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA). Beads were then rinsed with Buffer C (10 mM Tris-HCl, pH 8.0; 10 mM EDTA) and resuspended in 100 µl ChIP Elution Buffer.To digest RNA and protein, beads were treated with 2 µl RNase A (10 mg/ml, Sigma #R6513) at 37 °C for 30 minutes, followed by 1.5 µl Proteinase K (20 mg/ml, Sigma #124568) at 55 °C for 1 hour with shaking. Reverse cross-linking was performed by incubating the samples at 65 °C overnight. DNA was purified using 1.4×

SPRIselect beads and eluted in 50 µl ChIP Elution Buffer. Library preparation was carried out by using the NEBNext Ultra II DNA Library Preparation Kit (New England Biolabs, #E7645L) and sequenced on the Aviti Cloudbreak Low platform with 2×150 bp reads (250 million clusters per run).

2.1.7 DNA Library design and cloning

The library design is described in the Results section. The 250 bp sequences of interest were synthesized as a single-stranded oligonucleotide pool by Twist Bioscience. Each sequence was flanked by two synthetic adapter sequences to enable PCR amplification of the pooled library using primers, and subsequent cloning into the receiving plasmid. The oligo pool was PCR-amplified according to the manufacturer's instructions and cloned into a RMCE plasmid using the In-Fusion cloning system (Takara, #638947). This RMCE plasmid includes asymmetric Lox sites, a minimal promoter driving expression of a reporter gene (i.e., GFP), and flanking priming sites compatible with targeted SMF.

PCR primers used:

Fwd: TGATGAGCCAGTGATATCCATCG

Rev: AGCTGGCTAGCTTAATTAACTATCG

2.1.8 DNA Library insertion in the mouse genome

TC-1 ES cells previously engineered with a landing pad for targeted insertion via RMCE (Lienert et al., 2011) were used following the generation of the DNMT triple-knockout (TKO) background, as previously described in the Method section. To precisely map the genomic integration site of the landing pad, targeted locus amplification coupled with next-generation sequencing was carried out (Cergentis), revealing the insertion at mouse chromosome 2:130,098,005–130,098,011. For site-specific integration of DNA libraries into the mouse genome, RMCE was employed as previously described (Krebs et al., 2014, p. 20). TC-1 DNMT TKO embryonic stem cells were first selected with 250 µg/ml hygromycin (Roche, Switzerland) for 14 days. Subsequently, the Amaxa 4D-Nucleofector Core Unit (Lonza) was used to nucleofect 50 million cells with 300 µg of the plasmid DNA library and 180 µg of the pIC-CRE expression plasmid. Two days post-nucleofection, cells were subjected to selection with 3 µM ganciclovir (Selleckchem, #S1878) for 10 days.

2.1.9 Barcodes amplification

From the cell pool containing the inserted DNA library, genomic DNA (gDNA) was extracted using a phenol–chloroform protocol, while total RNA was isolated with the RNeasy Kit (Qiagen, #74104) following the manufacturer's instructions. On-column DNase treatment was included to ensure complete DNA degradation. Then, RNA was retro-transcribed using the PrimeScript RT Reagent Kit (Takara, #RR047A).

Barcodes were amplified from both gDNA and cDNA using two consecutive PCR reactions. PCR1 was carried out with 20 ng of gDNA or 100 ng of cDNA as input, using the following cycling conditions: 95 °C for 5 minutes; 20 cycles of 95 °C for 45 seconds, 70 °C for 20 seconds, and 72 °C for 30 seconds; 72 °C for 5 minutes as the final extension step. Primers used for PCR1 were:

Fwd: CTCTTTCCCTACACGACGCTCTTCCGATCTTCCTGCTGGAGTTCG

Rev: CTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTTTAAACTGTCGACCG

PCR2 was performed using 10 µL of PCR1 product as input, with the same cycling conditions. For indexing, NEBNext Multiplex Oligos for Illumina (NEB, #E7335L) were used as primers. The final library was sequenced on an Illumina NextSeq2000 platform using 100 bp single-end reads.

2.1.10 Barcode-to-CRE assignment

Following cloning of the CRE library, the plasmid pool was linearized using the restriction enzyme BaeI (NEB, #R0613S) to enable barcode-to-CRE assignment. The linearized library was then sequenced on a PromethION flow cell (R.10.4.1) using Nanopore technology, with library preparation and loading performed according to the manufacturer's instructions. I performed library preparation and loading, while Kasit Chatsirisupachai performed the analysis to assign each barcode to a CRE.

2.2 Computational analysis

2.2.1 ChIP-seq – Data Pre-processing and analysis

To remove Illumina adapter sequences and low-quality bases, raw sequencing reads were trimmed with TrimGalore (v0.6.7). Reads were then aligned to the *Mus musculus*

reference genome (BSgenome.Mmusculus.UCSC.mm10). The alignment step was performed using the QuasR R package, which utilizes Bowtie as the underlying aligner with the following parameters: -m 1 -e 70 -X 1000 -k 2 --best --strata. Only reads with unique mapping coordinates were retained. After, PCR duplicates were identified and removed using the MarkDuplicates tool from Picard (v3.1.0). Normalized signal coverage tracks were generated in bigWig format using deepTools bamCoverage (v3.5.0) with a bin size of 50 bp. For the H3K27Ac ChIP-seq experiments including exogenous *Drosophila* spike-in, reads were also aligned to the *Drosophila* reference genome (BSgenome.Dmelanogaster.UCSC.dm6) and processed identically. Normalization factors were then calculated by taking into account the number of uniquely aligned reads mapping to major chromosomes in both input and immunoprecipitated (IP) samples – for both mouse and *Drosophila* genomes – as follow: aligned reads × (target IP / spike-in IP) × (spike-in input / target input) (Patel et al., 2024). The resulting normalization factors were applied during bigWig file generation to scale the ChIP-seq signal accordingly.

To quantify changes in H3K27Ac levels between DMSO- and A-485-treated samples, read counts were obtained over 2 kb windows centered on regions pooled down during the SMF bait-capture step. This step was performed using the qCount function from the QuasR package. These count matrices were then analyzed using the DESeq2 Bioconductor package (Love et al., 2014) to identify differential regions. P-values were corrected for multiple testing using the Benjamini-Hochberg method, and log2 fold changes in H3K27Ac signal were used to evaluate the effect of p300 inhibition across the targeted genomic regions.

2.2.2 Bait-Capture SMF – Data Pre-processing

To remove Illumina adapter sequences and low-quality bases, raw sequencing reads were trimmed with TrimGalore (v0.6.7). Reads were then aligned to the *Mus musculus* reference genome (BSgenome.Mmusculus.UCSC.mm10). The alignment step was performed using the QuasR R package, which utilizes Bowtie as the underlying aligner with the following parameters: -m 1 -e 70 -X 1000 -k 2 --best --strata. Only uniquely reads with unique mapping coordinates were retained. After, PCR duplicates were identified and removed using the MarkDuplicates tool from Picard (v3.1.0).

2.2.3 Targeted SMF at the Ectopic Locus – Data Pre-processing

Amplicon SMF sequencing reads were initially processed with TrimGalore (v0.6.7) to remove Illumina adapter sequences and low-quality bases. Reads shorter than 20 bp after trimming were excluded. Residual plasmid backbone sequences were removed using a custom R script, through a strategy based on the position of the targeted SMF primers. Once the sequencing adapters were removed, SMF primers were matched in each read pair: the first nucleotide of each read had to correspond to the first nucleotide of the primers. Knowing that the backbone spans 42 bases and 129 bases from the forward and reverse primers respectively, these lengths were trimmed from the reads according to the identity of the primer matched. The resulting trimmed reads were aligned to the mouse reference genome (BSgenome.Mmusculus.UCSC.mm10) using the qAlign function from the QuasR package, with the following alignment parameters: -e 70 -X 1000 -k 2 --best --strata. Duplicate reads were identified based on identical cytosine methylation patterns and removed using custom R-based filtering functions.

2.2.4 SMF – Methylation Calling

Methylation calls at single-cytosine resolution were performed using the CallContextMethylation function from the development version of the SingleMoleculeFootprinting Bioconductor package, restricted to cytosines covered by at least 20 reads.

2.2.5 SMF – Footprint Detection and Quantification Using FootprintCharter

For each genomic region (tile), individual molecules covering the locus were extracted, and their binary methylation profiles were smoothed using a 40 bp sliding window. Pairwise Euclidean distances between molecules were then computed using the parDist function from the parallelDist R package. Unsupervised clustering was applied to the resulting distance matrix using the k-medoids algorithm via the pam function from the cluster package. Up to 12 clusters were generated per locus, with a minimum of five molecules per cluster. Clustering was performed on pooled data from paired samples to enable direct comparison of molecular state frequencies.

For each partition, the bulk methylation signal was calculated and used to detect footprints, defined as stretches of cytosines found unmethylated in at least 50% of the

molecules. TF footprints were defined as unmethylated stretches of cytosines spanning 12–75 bp, present in at least 50% of molecules in the cluster. Nucleosome footprints were defined as unmethylated stretches exceeding 120 bp. Any footprints supported by fewer than five molecules, or lacking at least one methylated cytosine on either side, were excluded from analysis to ensure robust detection.

This step of the analysis was performed by Guido Barzaghi.

2.2.6 SMF – Calculation of Chromatin Accessibility Frequency

For each TF motif, the fraction of molecules displaying continuous accessible regions – defined as the absence of detectable footprints – was quantified. Accessible stretches longer than 100 bp were used to classify molecules as active.

This step of the analysis was performed by Guido Barzaghi.

2.2.7 SMF – Single Locus Plots

Bulk SMF plots for individual loci were generated using the PlotAvgSMF function from the SingleMoleculeFootprinting Bioconductor package.

2.2.8 TFBS Annotation

Transcription factor binding sites (TFBSs) were identified by scanning the mouse reference genome using position weight matrices (PWMs) for vertebrate TFs obtained from the JASPAR database (Sandelin, 2004), accessed via the JASPAR2016 Bioconductor package (Mathelier et al., 2016). PWM-based motif scanning was performed using the TFBStools package (Tan and Lenhard, 2016), and matches with a score ≥10 were retained for downstream validation. To identify active binding sites, public ChIP-seq datasets were used to filter hits for 15 transcription factors: BANP, CTCF, E2F1, ESRRB, FOXD3, MAFK, MYC, NFYA, NRF1, OTX2, REST, STAT3, YY1, ZIC3, and ZFX. Predicted binding sites for OCT4, SOX2, and KLF4 were taken from Avsec et al. (2021). For consistency with other TFs, these motifs were re-scored using the corresponding JASPAR PWMs and validated against the ChIP-nexus datasets provided in the same study. In cases where PWM scanning produced overlapping motif predictions, only the match with the highest PWM score was retained for further analysis.

This annotation was performed by Guido Barzaghi. For a comprehensive list of the used ChIP-seq datasets, see TableS1 of (Baderna et al., 2025), as more then 150 datasets were integrated.

2.2.9 CRE annotation via ChromHMM

Genome-wide chromatin state annotations for mESCs were obtained from a 12-state ChromHMM model described in (Pintacuda et al., 2017). For downstream analysis, states were grouped into broader functional categories: CTCF (1_Insulator), intergenic (2_Intergenic), enhancer (4_Enhancer, 8_StrongEnhancer, 11_WeakEnhancer), and promoter (7_ActivePromoter).

2.2.10 Quantification of TF Motifs at CREs

The number of transcription factor motifs present at CREs was determined using the Arrange_TFBSs_clusters function from the SingleMoleculeFootprinting Bioconductor package. A curated list of bound motifs (see the 'TFBS Annotation' paragraph) was supplied to the TFBSs argument. To constrain motif clustering, the parameter max_cluster_width was set to 300 and the max_cluster_size was set to 10.

2.2.11 Statistical Testing and Correlation Analysis

Statistical comparisons of chromatin accessibility frequencies were performed using the Wilcoxon signed-rank test, as implemented in R. Pearson correlation coefficients were calculated using the cor function in R with the method set to "pearson".

2.2.12 Data Visualization

Heatmaps were generated with the pheatmap package in R (Gu et al., 2016). Boxplots, scatterplots, volcano plots, density plots, violin plots, and genomic track visualizations presented in the manuscript were created using the ggplot2 package in R.

3. RESULTS

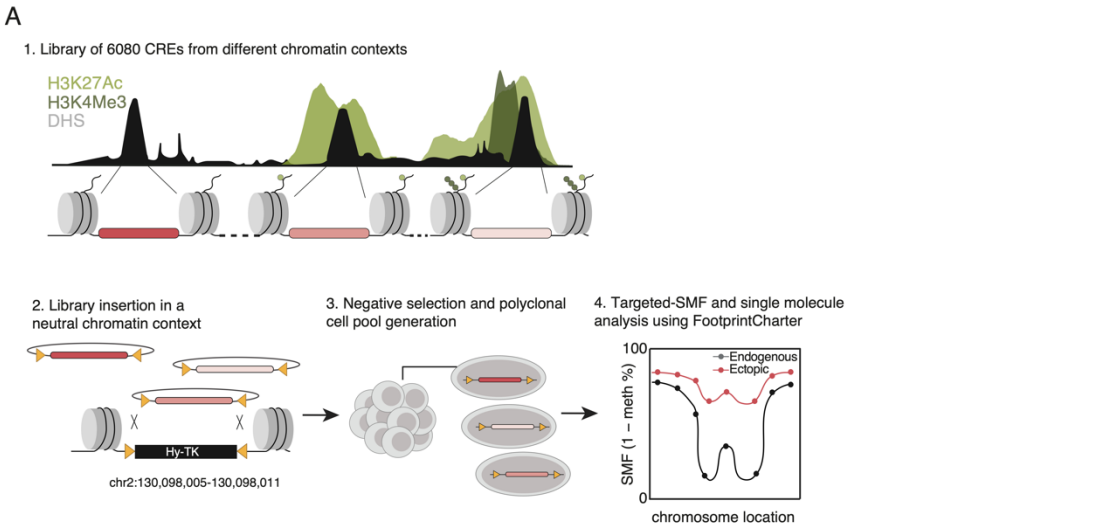
Transcription Factor (TF) binding events occur within a complex genomic environment, defined not only by specific partner TFs, but also shaped by the level of activating and repressing histone post-translational modifications (PTMs). TF binding and the deposition of activating chromatin marks often occur simultaneously at endogenous *cis*-regulatory elements (CREs), making it challenging to disentangle their individual contributions to chromatin accessibility. To disentangle the effect of TF binding on chromatin accessibility from the influence of the broader chromatin context, it is essential to study TF function within chromatin under controlled conditions where histone PTMs are either identical or potentially absent. To address this, I combined Recombination Mediated Cassette Exchange (RMCE) with Single Molecule Footprinting (SMF) (Krebs et al., 2017; Sönmezer et al., 2021; Kleinendorst et al., 2021; Krebs, 2021). RMCE enables the insertion of thousands of different DNA sequences at the same locus (or "landing pad"), while SMF marks chromatin accessibility along single DNA molecules through methylation of accessible cytosines with recombinant methyltransferases. Methylation is then detected by bisulfite conversion (i.e., unmethylated cytosines are converted to thymines) and sequencing, enabling inference of chromatin accessibility frequency at genomic loci.

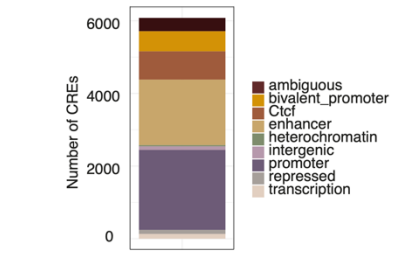
3.1 Library design and insertion

To assess the ability of CREs to recruit TFs and open chromatin using RMCE combined with SMF, I started by designing a library of elements to test. Using annotated open regulatory elements (DNAse hypersensitive sites, DHS) in mouse embryonic stem cells (mESCs) from published DNase-seq data (Domcke et al., 2015), I selected 6,080 elements, the majority of which are active regulatory regions such as promoters and enhancers, from different genomic context (Figure 9A, step 1). I further refined the selection by focusing on various categories of CREs, such as enhancers bound by pluripotency TFs (e.g., SOX2, OCT4, KLF4, MYC, MAX) as representative examples of cell-type-specific enhancers, and promoters bound by ubiquitously expressed TFs like NRF1, NFY, and YY1 (Figure 9B-C). Additionally, I included CTCF-bound loci as a positive control, as CTCF motifs have been shown to autonomously drive chromatin opening (Grand et al., 2024). SMF is agnostic to the identity of the TFs binding the DNA and therefore has to be complemented with orthogonal methods. For this reason, all the

selected CREs were required to have confirmed TF binding – based on ChIP-seq datasets – and to contain a sufficient number of GpC and CpG dinucleotides to enable high-resolution SMF and thus accurate measurements of chromatin accessibility, while minimizing the overrepresentation of CpG islands (Figure 9D). Finally, all selected regions were required to be covered in our reference bait-capture SMF dataset – previously generated in the lab by Can Sönmezer and Rozemarijn Kleinendorst (Sönmezer et al., 2021) – allowing for direct comparison of chromatin accessibility between ectopic and endogenous sequences.

The library was synthesized as a single stranded oligo pool, PCR amplified and cloned into a receiving plasmid flanked by asymmetric tandem lox sites, enabling its insertion into the genome through RMCE (Figure 9A, step 2). After, I integrated the library at a landing pad (hygromycin-thymidine kinase (Hy-TK) cassette also flanked by asymmetric tandem lox sites) that was previously used to study transcription regulation, TF binding, and the establishment of epigenetic marks (Butz et al., 2022; Grand et al., 2024; Hartl et al., 2019; Krebs et al., 2014; Lienert et al., 2011). Before integrating the library, I confirmed that the landing pad is located in a region lacking both activating and repressive chromatin marks – a locus that I then define as neutral – by leveraging a service provided by Cergentis (that couples targeted locus amplification with next generation sequencing) to genotype the precise location of the landing pad. The landing pad was mapped at mouse chr2:130,098,005-130,098,011 within a region that is gene poor, with neither active nor repressive chromatin marks, and with no active transcription (Figure 9E-F). After nucleofection, negative selection with ganciclovir was used to enrich successful insertion events (Figure 9A, step 3). Nucleofected mESCs were treated with ganciclovir. In cells where the insertion did not occur, the Hy-TK cassette was not replaced by the library and remained at the landing pad. As a result, thymidine kinase (TK) was still expressed and converted ganciclovir into its monophosphate form, which was subsequently phosphorylated by endogenous kinases into a toxic triphosphate. This toxic form was incorporated into DNA during replication, leading to chain termination and apoptosis. This negative selection approach avoids the possible impact on chromatin accessibility of active transcription from a selection marker at the ectopic site, which can occur with positive cell selection.

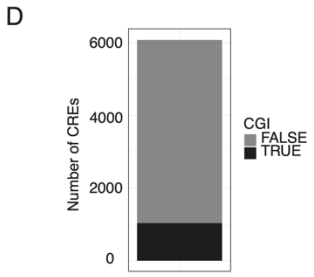


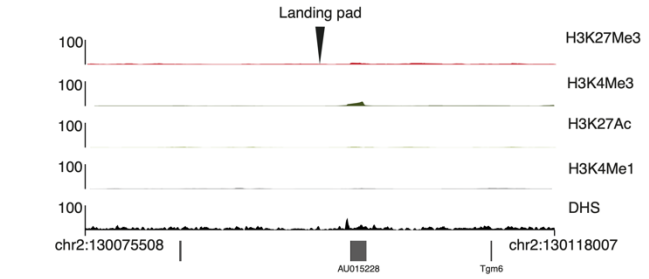


C

TF	Occurences
Klf4	3921
Sox2	1661
Oct4	1605
Ctcf	1445
Nanog	1131
Zfx	1052
Myc	904
Nrf1	717
Zic3	689
Bach1::Mafk	397
Yy1	388
Smad2::Smad3::Smad4	370
Stat3	368

TF	Occurences
Foxd3	300
Esrrb	227
Otx2	219
Nfya	216
Tbp	173
Nanog-Partner	122
E2f1	97
Tgif2	65
Rest	45
Tp53	36
Banp	7
Sox21	6
Tgif1	5





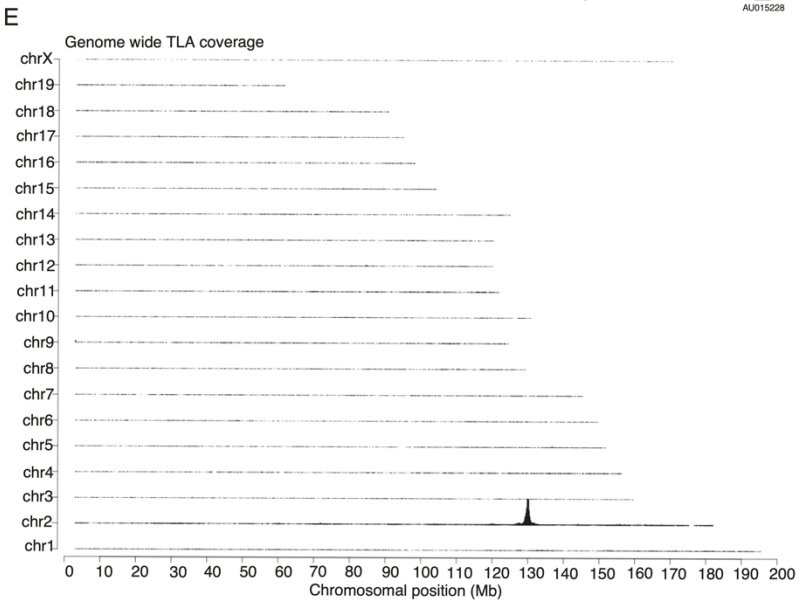


Figure 9. Design and composition of the CRE insertion library.
(A) Schematic overview of the experimental approach designed to investigate how TF binding and chromatin context influence CRE accessibility. A library of 6,080 DNA fragments covering genomic CREs is synthesized and cloned into a plasmid containing asymmetric tandem lox sites to allow RMCE. The pooled fragments are integrated into a predefined neutral genomic landing pad. Following negative selection, cells with successful recombination are enriched, generating a polyclonal population in which each cell carries a single insert. Accessibility of the inserted sequences is then measured using targeted SMF and compared to their corresponding endogenous loci.
(**B**) Stacked barplot showing the composition of the library based on ChromHMM annotation.
(**C**) Table summarizing TF identities and the number of their motif occurrences in the library.
(**D**) Stacked barplot showing the proportion of CpG islands present in the library.
(**E**) Targeted Locus Amplification (TLA) signal plotted across the mouse genome, with chromosomes and chromosomal position indicated. The landing pad is integrated at chr2:130,098,005–130,098,011.
(**F**) Genome browser view showing the levels of H3K27Me3, H3K4Me3, H3K27Ac, H3K4Me1, chromatin accessibility profiles (as DHS peaks) and gene annotation at the neutral landing pad (chr2:130,075,508–130,120,508; 40 kb window).

3.2 Optimization of SMF analysis for complex insert libraries using methylation-based PCR deduplication

Following the generation of the polyclonal cell pool carrying the DNA library, I assessed chromatin accessibility at the integration site. To this aim, I leveraged a targeted SMF approach, which uses primers designed to bind a synthetic sequence adjacent to the insert. This strategy enables for unambiguous distinction between the ectopic and endogenous loci (Figure 9A, step 4). After alignment to the mouse genome, I recovered 2,446 out of the 6,080 DNA fragments present in the library (Figure 10A). This recovery rate is consistent with expectations based on a simple simulation modelling the RMCE process (Figure 10B). The simulation takes into account the initial complexity of the library (i.e., 6,080 DNA fragments), the approximate insertion rate per nucleofection (~330), the insertion redundancy (i.e., out of 6,080 DNA fragments, the same one can be inserted multiple times across different nucleofections among the ~330 inserted, so not every insertion event lead to a new unique fragment being recovered) and 20% background noise (i.e, because the cloning efficiency is not 100%, plasmids without any DNA fragment can still be present in the population and get inserted, reducing the effective diversity of recovered fragments). It predicted an average recovery of approximately ~2,300 unique modules, placing the observed value within the expected range (Figure 10B).

In classic targeted amplicon-SMF experiments, PCR duplication is generally not a major concern. First, SMF quantifies protein occupancy based on the ratio of methylated to total molecules at each site, inherently correcting for amplification biases – provided that methylated and unmethylated are equally affected. Second, the target locus is present in all cells (typically in two copies, one per allele), ensuring high template abundance.

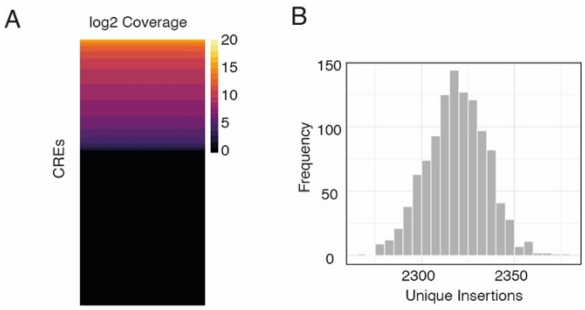


Figure 10. Quantification and modeling of CRE recovery following RMCE and SMF.
(A) Heatmap showing log2 transformed coverage values for the DNA fragments present in the library. Coverage values were obtained from aligned SMF data. Rows were sorted in descending order based on total coverage, and no clustering was applied. The color scale represents the log2 of the coverage.
(B) Simulation of expected unique insertion recovery after RMCE. Histogram showing the distribution of adjusted unique DNA fragment insertions across 1,000 simulated trials. Each simulation modeled 12 nucleofections with 330 insertions per nucleofection, sampled with replacement from a library of 6,080 elements. The simulation accounts for insertion redundancy (i.e., the same fragment can be inserted multiple times across nucleofections, so not all insertions are unique) and background noise (i.e., due to cloning efficiency not being 100%, some plasmids without a cloned DNA fragment may still be present in the plasmid pool and get inserted at the landing pad). A 20% background correction was applied to the total number of unique insertions in each trial. The resulting distribution represents the expected range of recovered unique elements under these experimental conditions.

However, this assumption breaks down in pooled RMCE-based assays, where each inserted element is present in only a small fraction of the cell population. As a result, the effective concentration of each DNA fragment is reduced, leading to a lower template complexity to start with. Under these conditions, PCR amplification frequently generates PCR duplicates, which limits the number of unique molecules per DNA fragments and constrains the subset of fragments with sufficient coverage for robust SMF analysis (Figure 11A). To overcome this limitation, together with Guido Barzaghi (former Ph.D. student in the lab), we developed a method to identify and remove PCR duplicates based on cytosine methylation patterns. Although bisulfite conversion typically has an efficiency of >95%, during this procedure two types of error can occur: unmethylated cytosines are not converted to thymines (false negative) and methylated cytosines are converted to thymines (false positive) (Figure 11B). This phenomenon introduces a low but consistent error rate that acts as a molecular fingerprint. Based on the library design, each 250 bp fragment contains, on average, approximately 40 cytosines in a non-methylable context and 30 cytosine in a methylable context (i.e., not targeted and targeted by the methyltransferase, respectively) (Figure 11C). This provides a sufficient number of bisulfite-sensitive positions to reliably distinguish individual molecules by their conversion patterns. The probability that two independent molecules exhibit the identical error pattern purely by chance is therefore extremely low (less than 1%). This allows us to distinguish true biological duplicates from technical PCR artifacts with high

confidence, reducing the risk of including PCR duplicates that could bias the measure of accessibility frequency during SMF analysis (Figure 11D). Using this strategy, I recovered ~400 well-covered DNA fragments from a total of ~2500 insertions.

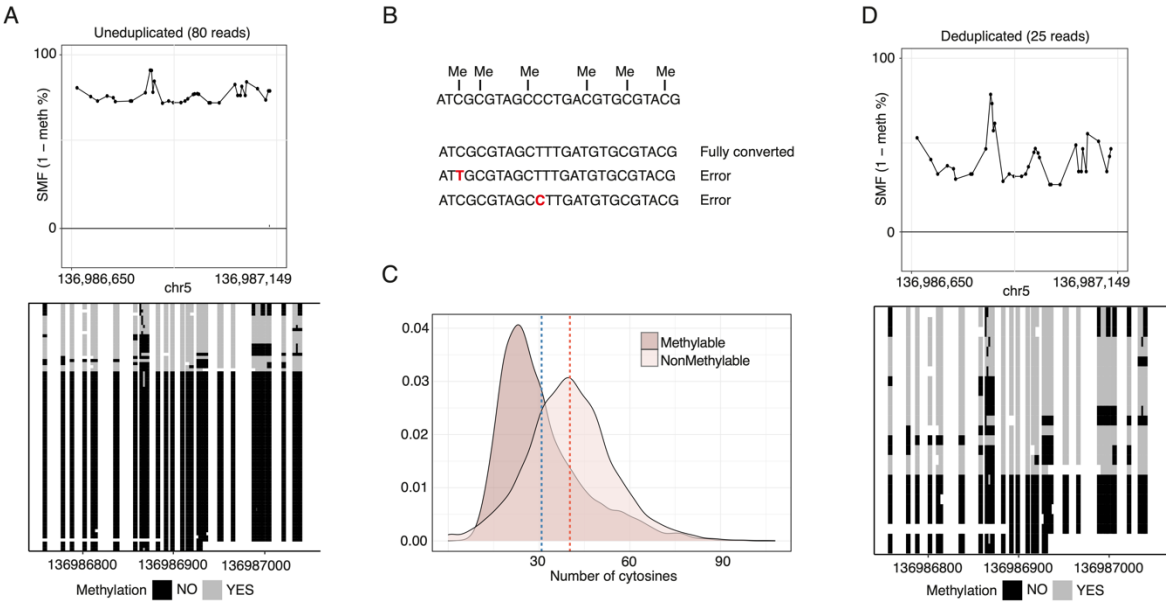


Figure 11. Validation and application of a methylation-based PCR duplicate removal strategy for SMF.
(A) Top: average SMF signal (1-methylation %) for a representative locus on chromosome 5, based on 80 reads with PCR duplicates retained. Each cytosine is represented as a black dot. Bottom: corresponding single-molecule stack plot displaying methylation status for each cytosine (black = unmethylated, gray = methylated).
(B) Schematic of bisulfite conversion and error types. Unmethylated cytosines are expected to convert to thymines, while methylated cytosines are protected. Errors introduced during bisulfite conversion (e.g., failed or inappropriate conversions) serve as molecular fingerprints for identifying unique molecules.
(C) Density plot showing the distribution of cytosine counts in methylable and non-methylable contexts across all 6,080 DNA fragments in the library. Blue and red dashed lines indicate the average of methylable and non-methylable cytosines, respectively.
(D) Top: average SMF signal (1-methylation %) for a representative locus on chromosome 5, after PCR deduplication (25 reads). Each cytosine is represented as a black dot. Bottom: corresponding single-molecule stack plot displaying methylation status for each cytosine (black = unmethylated, gray = methylated).

3.3 Benchmarking parallel chromatin accessibility quantification at ectopic loci using CTCF

A key pre-requisite of the system I implemented is the ability to directly compare the chromatin accessibility of ectopically inserted CREs with their endogenous counterparts. The reference endogenous SMF bait-capture covers over 60% of active promoters and enhancers in mESCs (Sönmezer et al., 2021). For the analysis of this dataset, Guido Barzaghi annotated TFBSs genome-wide (Mathelier et al., 2016) and classified the motifs for 18 TFs expressed in mESCs as bound or unbound, based on ChIP-seq and ChIP-nexus datasets that are publicly available (see 'TFBS annotation' in the Methods for details). Then, he quantified chromatin accessibility at single-molecule resolution using *FootprintCharter*, an unsupervised algorithm that he developed during his PhD (Barzaghi

et al., 2025). Unlike supervised approaches that require prior knowledge, this method does not rely on predefined annotation of motifs in the genome. Instead, it analyzes single-molecule methylation patterns directly: it clusters molecules based on their methylation profiles and identifies protected regions – interpreted as TF or nucleosome footprints – based on local SMF signals and footprint size (Figure 12A). This enables the estimation of chromatin occupancy frequencies across cell populations with high resolution and flexibility. This information is then complemented with the annotated TFBSs, allowing us to determine the accessibility of each individual motif. Importantly, footprints are not detected for all TFs and depend on the presence of cytosines near their binding sites (Kleinendorst et al., 2021; Sönmezer et al., 2021). Therefore, to ensure uniform analysis across all TFs, molecules were classified solely based on the absence or presence of nucleosome footprints, with those containing TF footprints categorized as accessible, unless specified otherwise.

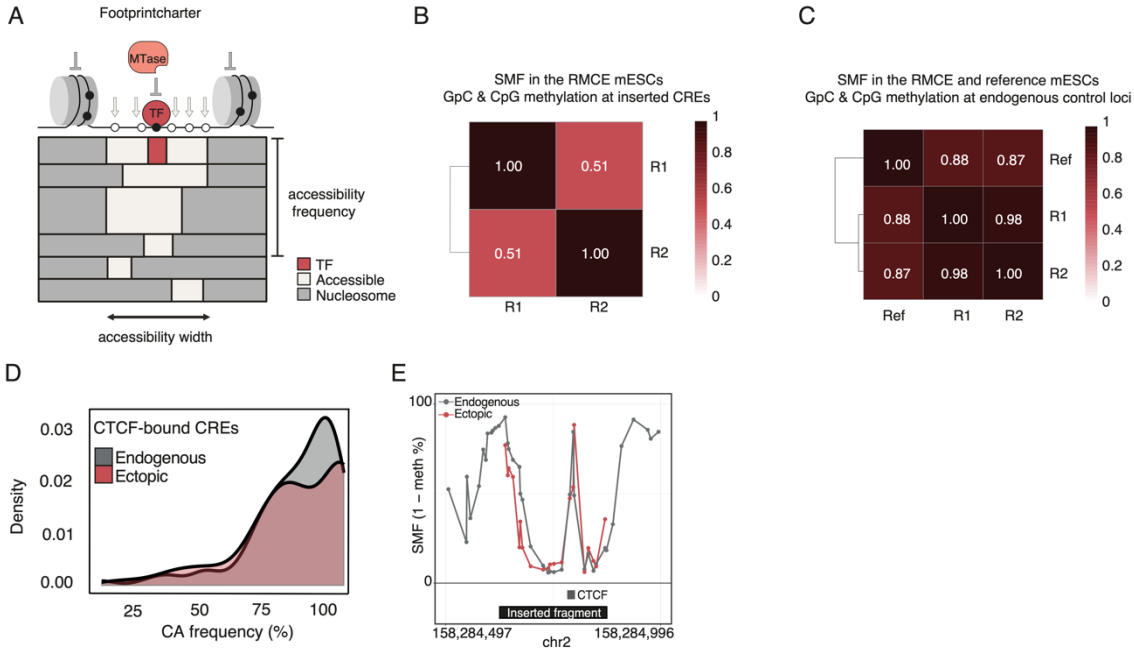


Figure 12. Quantification of chromatin accessibility at ectopic CREs using *FootprintCharter*.
(A) Schematic overview of the *FootprintCharter* algorithm. The method applies unsupervised partitioning to detect and quantify TF and nucleosome footprints from single-molecule methylation data. Chromatin accessibility frequencies are inferred based on the presence and width of protected regions, with molecules containing TF footprints classified as accessible.
(B) Heatmap showing correlation of GpC and CpG methylation at inserted CREs between two biological replicates (R1 and R2) in the RMCE mESCs. The Pearson correlation coefficient R is reported.
(C) Heatmap showing correlation of GpC and CpG methylation patterns at five endogenous control loci across two biological replicates (R1 and R2) in the RMCE mESCs and the reference mESC dataset. The Pearson correlation coefficient R is reported.
(D) Density plot showing the distribution of chromatin accessibility (CA) frequencies for CTCF-bound fragments (n=111) at endogenous (grey) and ectopic (red) loci. **(E)** Shown is the average SMF signal (1 – methylation %) at a representative CTCF-containing fragment (black rectangle) at its endogenous (grey) and the ectopic loci (red).
(E) Shown is the average SMF signal (1 – methylation %) at a representative CTCF-containing fragment (black rectangle) at its endogenous (grey) and the ectopic loci (red).

To enable direct comparison with the genome-wide SMF measurements, accessibility frequencies at the ectopic locus were quantified using *FootprintCharter*. This quantification step was carried out by Guido Barzaghi using the tool that he developed, while all upstream and downstream analyses were performed by me. With this strategy, I measured chromatin accessibility reproducibly across hundreds of distinct fragments in parallel when inserted at the ectopic locus (Figure 12B). In each experiment, I profiled chromatin accessibility at five endogenous control loci along with the ectopic locus. These endogenous loci are internal controls to confirm that the footprinting procedure was successful, as their signal should match the reference SMF data (Figure 11C). Therefore, any observed differences between the endogenous and ectopic loci can be attributed to biological factors rather than batch effects in the SMF protocol.

To further evaluate the capability of the method in detecting chromatin accessibility at the ectopic site, I first focused on CTCF-bound loci. This choice was based on recent findings showing that CTCF motifs alone are enough to drive chromatin opening when inserted at the same ectopic locus used in this study (Grand et al., 2024). A strong agreement was observed between ectopic and endogenous CTCF motif loci (Figure 12D), with a visible TF footprint and comparable levels of chromatin accessibility (Figure 12E). These results validate the method's ability to precisely quantify chromatin accessibility frequencies across hundreds of sequences in parallel integrated at a single ectopic site.

3.4 Chromatin accessibility is only partially re-established at the ectopic locus

I then analyzed all ectopic CREs that do not contain CTCF motifs and systematically assessed how their chromatin accessibility compared to that of their endogenous, sequence-identical counterparts. Notably, 97% of the ectopic inserts exhibited accessibility over the background, which was defined based on the accessibility level at unbound motifs (14%; calculated by Guido Barzaghi (Baderna et al., 2025)). Of these accessible sites, 42% showed comparable accessibility frequencies between the ectopic and endogenous loci (Figure 13A and Figure 13B), whereas the remaining 58% were less accessible when inserted ectopically (Figure 13A and Figure 13B). For instance, the *Phc3* promoter displayed nearly identical chromatin accessibility and TFs footprints in both contexts (Figure 13C), while the *Elp6* promoter only partially reproduced accessibility at

the ectopic locus (Figure 13D). These results suggest that while short (250 bp) CREs often retain functional information sufficient for chromatin opening, full recapitulation occurs in only about half of the cases. When grouped by CRE category, I observed a significant reduction in accessibility at both enhancers and promoters (Figure 13E).

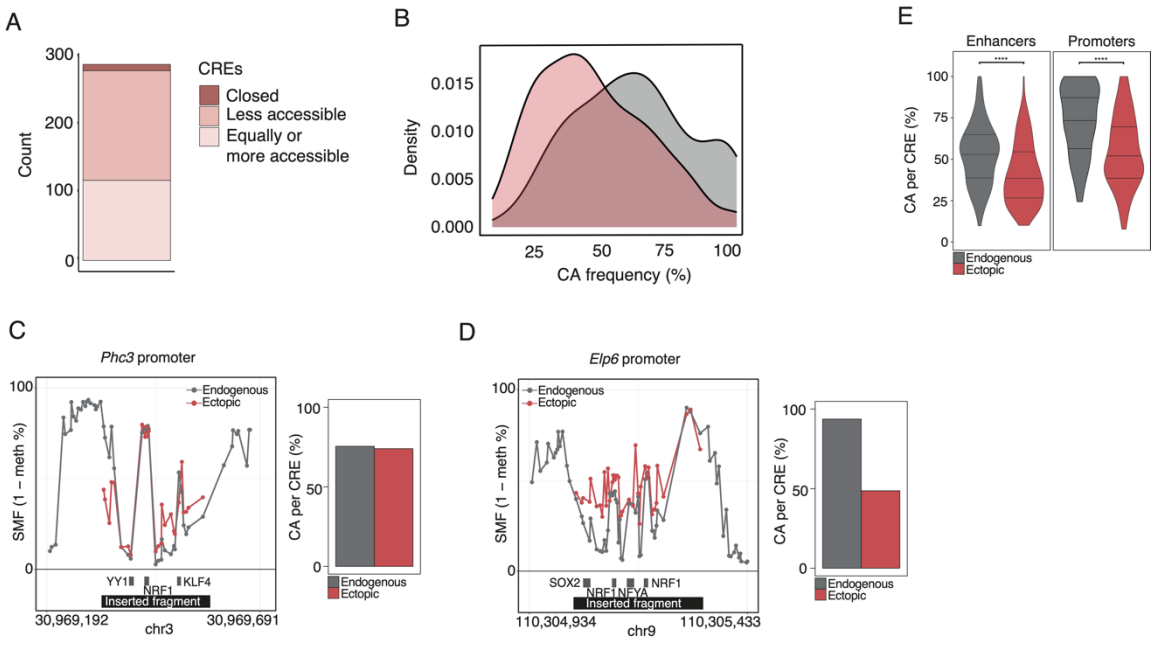


Figure 13. CREs generally lose accessibility when inserted at the ectopic locus.
(A) Stacked barplot displaying the distribution of DNA fragments (excluding CTCF-bound sites) across accessibility categories at the ectopic locus relative to their endogenous counterparts. Fragments are classified as closed, less accessible, or equally/more accessible based on a Fisher's exact test comparing the number of nucleosome-occupied versus accessible reads between endogenous and ectopic sites. Regions with accessibility below 14% – the background level observed at unbound TF motifs, as determined by *FootprintCharter* – are defined as closed.
(B) Density plot showing chromatin accessibility (CA) frequency distribution for endogenous (gray) and ectopic (red) loci (n=286).
(C–D) Representative examples of DNA fragments that either fully (*Phc3* promoter) or partially (*Elp6* promoter) recapitulate chromatin accessibility at the ectopic locus. Left: average SMF signal (1 – methylation %) for each promoter at the endogenous (gray) and ectopic (red) locations. Right: barplots showing chromatin accessibility frequencies at both loci as quantified by *FootprintCharter*.
(E) Violin plots comparing chromatin accessibility (CA) at endogenous and ectopic loci for enhancers (n = 144) and promoters (n = 99). Significance levels: ns = $p > 0.05$; **** = $p \leq 0.0001$.

To evaluate the contribution of specific TF motifs to chromatin accessibility at the ectopic site, I compared CREs' accessibility at ectopic versus endogenous loci, grouping them according to the TFs bound at their endogenous location (Figure 14A). With the exception of CTCF – which maintained nearly complete accessibility at both genomic contexts – all other TF-associated CREs exhibited reduced accessibility when inserted at the ectopic locus (Figure 14B and 14C). For instance, fragments containing NFY and NRF1 motifs showed a median of 68% and 60% accessibility at the ectopic site,

respectively, compared to accessibility levels exceeding a median of 80% at their native loci (Figure 14B and 14C). A similar trend was observed for pioneer TFs such as SOX2 and KLF4, whose motif-containing fragments also showed decreased accessibility in the ectopic context (Figure 14B and 14C). Next, I asked whether the number of TF motifs within a CRE could explain the observed differences in chromatin accessibility as measured ectopically. To address this, I grouped CREs based on the number of motifs present within each element. However, this analysis did not reveal a clear correlation between motif number and accessibility frequency at either ectopic or endogenous loci. It is important to note that the number of observations in each category was limited (in order: 144, 88, 36), which may have impaired my ability to detect subtle trends (Figure 14D).

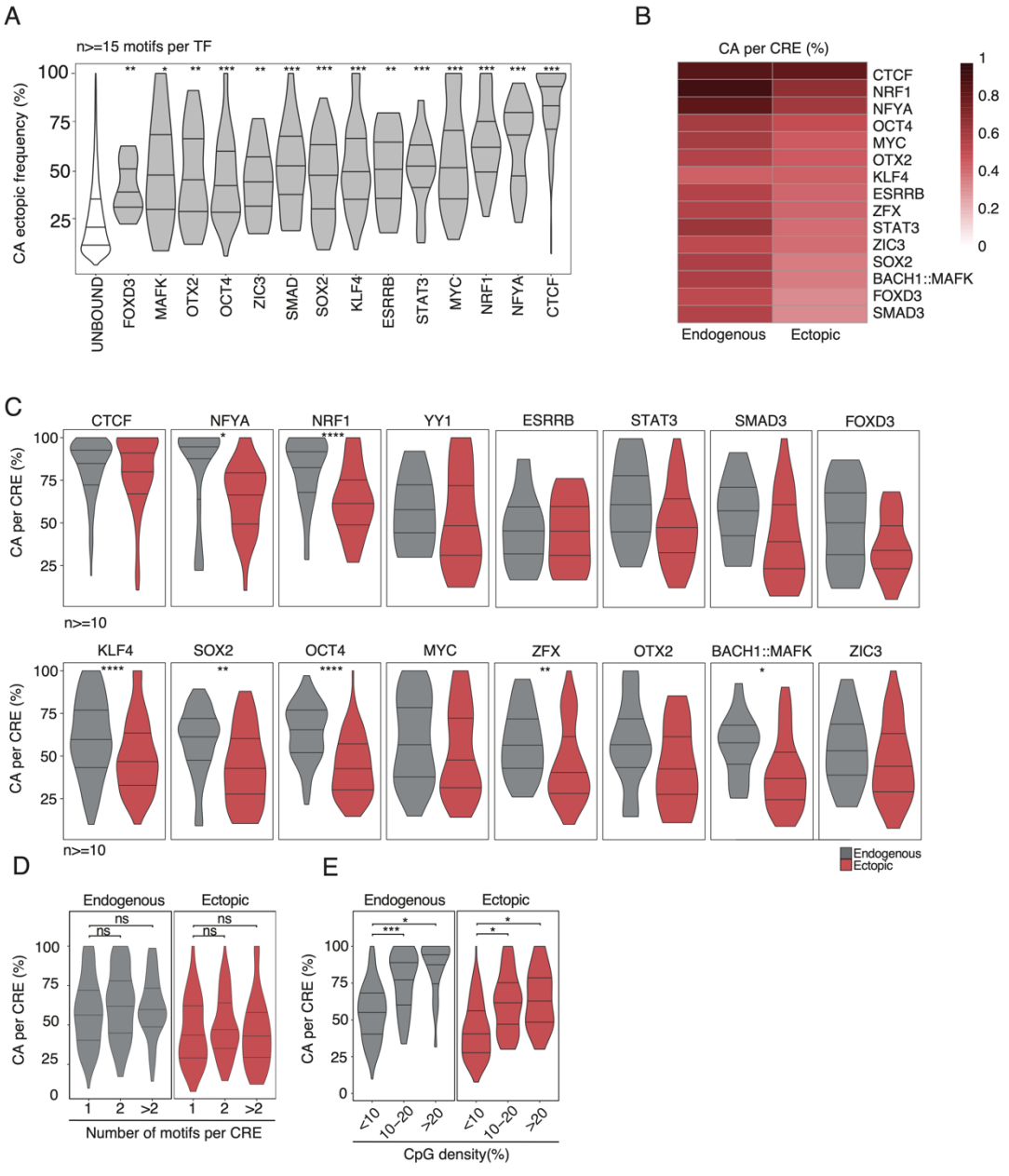


Figure 14. TF find and bind to CREs and open chromatin at the ectopic locus, typically at lower levels than observed at endogenous sites.
(A) Violin plots showing chromatin accessibility frequencies (CA) at ectopic CREs, grouped by the presence of specific TF motifs within the DNA fragments and compared to accessibility at unbound motifs. TF motif presence was defined based on ChIP-seq data and motif annotations. Significance levels: ns = $p > 0.05$; * = $p \leq 0.05$; ** = $p \leq 0.01$; *** = $p \leq 0.001$; **** = $p \leq 0.0001$.
(B) Heatmap displaying the median chromatin accessibility values at both endogenous and ectopic loci. TF motif presence was determined using ChIP-seq data and motif annotations.
(C) Violin plots showing chromatin accessibility (CA) frequencies at CREs, comparing endogenous loci (grey) and the ectopic locus (red), grouped by the presence of specific TF motifs within the DNA fragments (n>=10). Motif presence was defined based on ChIP-seq data and motif annotations. Significance levels: ns = $p > 0.05$; * = $p \leq 0.05$; ** = $p \leq 0.01$; *** = $p \leq 0.001$; **** = $p \leq 0.0001$.
(D) Violin plots displaying chromatin accessibility frequencies (CA) at CREs measured at endogenous loci (grey) and at the ectopic locus (red), grouped by the number of TF motifs present within each CRE (n=286). Significance levels: ns = $p > 0.05$; * = $p \leq 0.05$; ** = $p \leq 0.01$; *** = $p \leq 0.001$; **** = $p \leq 0.0001$.
(E) Violin plots displaying chromatin accessibility (CA) frequencies at CREs measured at endogenous loci (grey) and the ectopic locus (red), stratified by CpG density (n=286).

CREs and particularly promoter elements have increased density of CpG dinucleotides that could facilitate displacement of nucleosomes (Hartl et al., 2019). I thus asked whether CpG density could explain the differential ability of certain fragments to maintain chromatin accessibility. When CREs were grouped based on their number of CpGs (<10, 10-20, >20), I observed a stronger scaling between CpG density and the frequency of accessibility at the endogenous loci, than at the ectopic site, where there was any difference between the 10-20 and the >20 category (Figure 14E). This result suggests that CpG density may not directly contribute to enhanced chromatin accessibility, illustrating the importance of studying regulatory elements outside of their naturally occurring contexts. Also, any possible influence of DNA methylation on chromatin accessibility at the endogenous locus compared to the ectopic one was excluded, as all the experiments were performed in an isogenic mESC line lacking all three DNA methyltransferases (Sönmezer et al., 2021).

Collectively, these findings indicate that chromatin accessibility frequencies are only partly determined by the DNA sequence itself via TF motifs. With the notable exception of the insulator protein CTCF, the presence of the majority of the TF motifs in a 250 bp context is insufficient to fully recapitulate endogenous chromatin accessibility, even for TFs like NFY and NRF1, which typically exhibit high chromatin accessibility frequency at their endogenous sites. Also, the data suggest that additional non-CpG-driven mechanisms operate at the endogenous loci to drive higher accessibility frequency of CREs.

3.5 Increasing the size of inserted fragment seems to restore full accessibility at the ectopic site for a single CRE

Although the sequences inserted ectopically are genetically identical to their endogenous counterparts, one key difference is that only 250 bp are introduced at the ectopic site. In their endogenous context, additional TF motifs in the surroundings could guide the binding of more TFs (Figure 15A). These TFs, in turn, may influence chromatin accessibility either through direct binding or by recruiting co-factors that modify the local chromatin environment – for example, through the addition of histone PTMs. In this case, the observed decrease in chromatin accessibility at the ectopic site could be partially explained by the size of the inserted fragment. To test this hypothesis, I selected a CRE with ~25% loss of chromatin accessibility, which is bound by a pair of NRF1 motifs (Figure 15B). I cloned fragments spanning the motifs in sizes of 500 bp, 1 kb, 2 kb, 3 kb and 5kb, that I amplified from a mouse genetic background distinct from that of the recipient cells, containing single nucleotide variations (SNVs) (Figure 15C). I then generated 5 different cell pools, each containing only one size insert. After insertion, targeted SMF was carried out using primers that anneal to the flanking region of the TFBSs, amplifying always the same ~500bp, ensuring comparability across experiments. Since the primers amplify both endogenous and inserted elements, the reads are distinguished based on the SNVs present in the inserts compared to the endogenous counterpart, allowing for unambiguous distinction between the ectopic and endogenous loci. For technical reasons, one of the two biological replicates for some of the fragment size was not useable. From preliminary analysis, I observed that starting from the 1 kb fragment size, chromatin accessibility at the ectopic site reached levels comparable to the endogenous locus (Figure 15D and Figure 15E), as well as TF occupancy (Figure 15F and Figure 15G). These results suggest that the additional sequence context included in the larger fragment is sufficient to restore full chromatin accessibility, at least for this specific NRF1-bound CRE. While the 500 bp fragment showed reduced accessibility, the progressive increase in fragment length was associated with a recovery of accessibility, and TF occupancy, pointing toward a size-dependent effect.

These preliminary results suggest that, at least for this specific NRF1-bound CRE within a 1 kb region, the underlying DNA sequence alone is sufficient to restore full chromatin accessibility at the ectopic locus. If this is due to the presence of the additional NRF1 and KLF4 motifs (Figure 15A) remains yet to be determined. To build on this finding, it would be interesting to repeat the experiment using additional CREs with

varying features (e.g., different TF binding motifs) and from different chromatin contexts, complementing with the deletion of TF motifs in the fragments, or TF perturbation, to show causality.

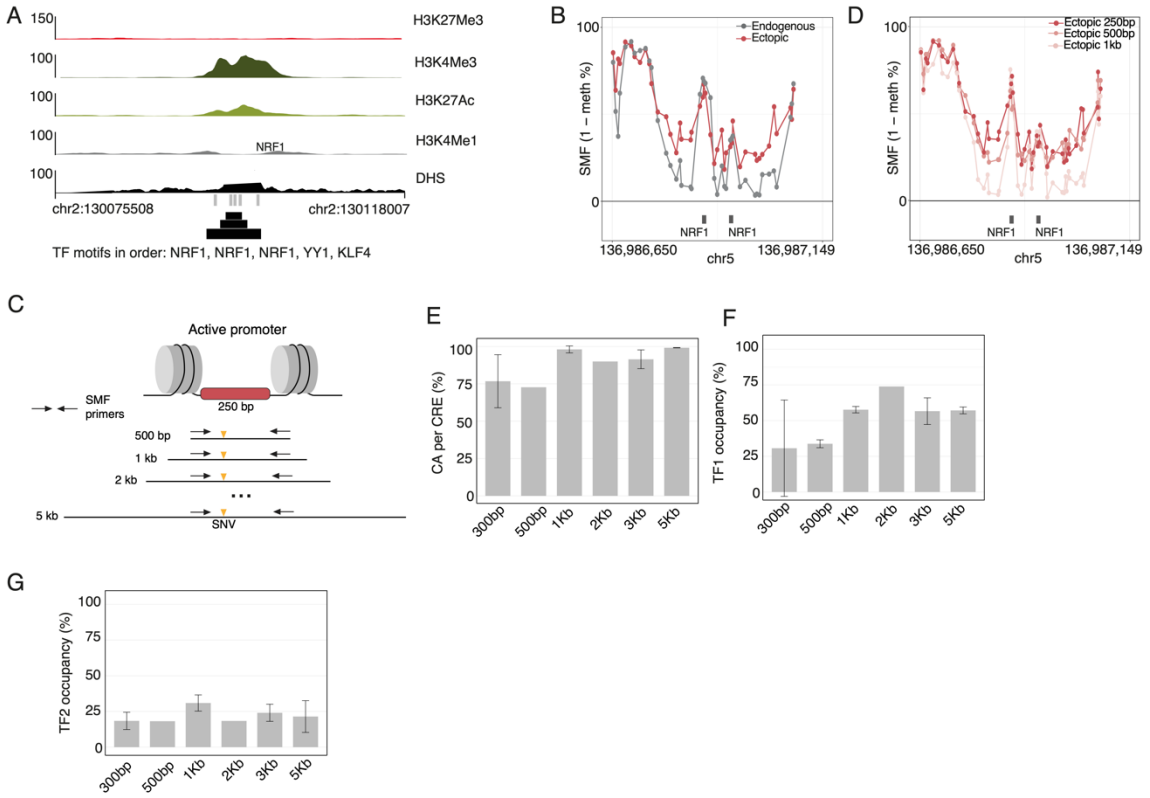


Figure 15. Increasing the size of a single inserted fragment restores chromatin accessibility at the ectopic locus.
(A) Genome browser view showing the levels of H3K27Me3, H3K4Me3, H3K27Ac, H3K4Me1, chromatin accessibility profiles (as DHS peaks), TF motifs annotation (grey rectangles) and the insert of 250bp, 500bp and 1kb (black rectangles of increasing size).
(B) Schematic representation of the experimental setup. Fragments of varying lengths (500 bp to 5 kb) encompassing a CRE with two NRF1 binding sites were amplified from a different mouse strain containing unique SNVs (indicated by orange triangles). These fragments were inserted at the ectopic site, and SMF was performed using the same primer pair, positioned in the flanking ~500bp region.
(C) Average SMF signal (1 – methylation %) at the NRF1-bound CRE in its endogenous context (grey) and at the ectopic locus (red).
(D) Average SMF signal (1 – methylation %) at NRF1 binding sites and the surrounding ~500 bp region at the ectopic locus, shown for inserted fragments of 250 bp, 500 bp and 1 kb.
(E-G) Quantification of chromatin accessibility (CA, panel E) and NRF1 occupancy across different fragment sizes (F, G). TF occupancy is shown for two NRF1 binding events (TF1 and TF2) across fragments of varying lengths. Error bars represent the standard deviation across biological replicates, where applicable.

3.6 Chromatin accessibility is enhanced by the presence of H3K27Ac

Another key difference is that in contrast to the endogenous context, activating histone marks – such as H3K4me3 and H3K27Ac – are not present at the ectopic integration locus (Figure 9A). To explore how these modifications relate to chromatin accessibility, I grouped CREs by their endogenous ChIP-seq signal for H3K4me3 and H3K27Ac (Figure

16A and Figure 16B). At endogenous loci, higher levels of these marks correlated with increased chromatin accessibility (Figure 16C). However, when comparing these same marks at the endogenous site to accessibility at the ectopic insertion, only H3K4me3 showed a positive correlation (Figure D, left), while H3K27Ac did not (Figure 16D, right). For example, among two NRF1-bound promoters, the *Fam277b* promoter – characterized by low endogenous H3K27Ac – closely recapitulated its accessibility pattern at the ectopic locus (Figure 16E). Conversely, the *Rsrp1* promoter, which displays high acetylation in its native context, showed only partial accessibility when ectopically inserted (Figure 16F). These findings support a model in which H3K27Ac enhances chromatin accessibility in the endogenous context, an effect that is diminished or absent at the ectopic site where histone modifications are lacking.

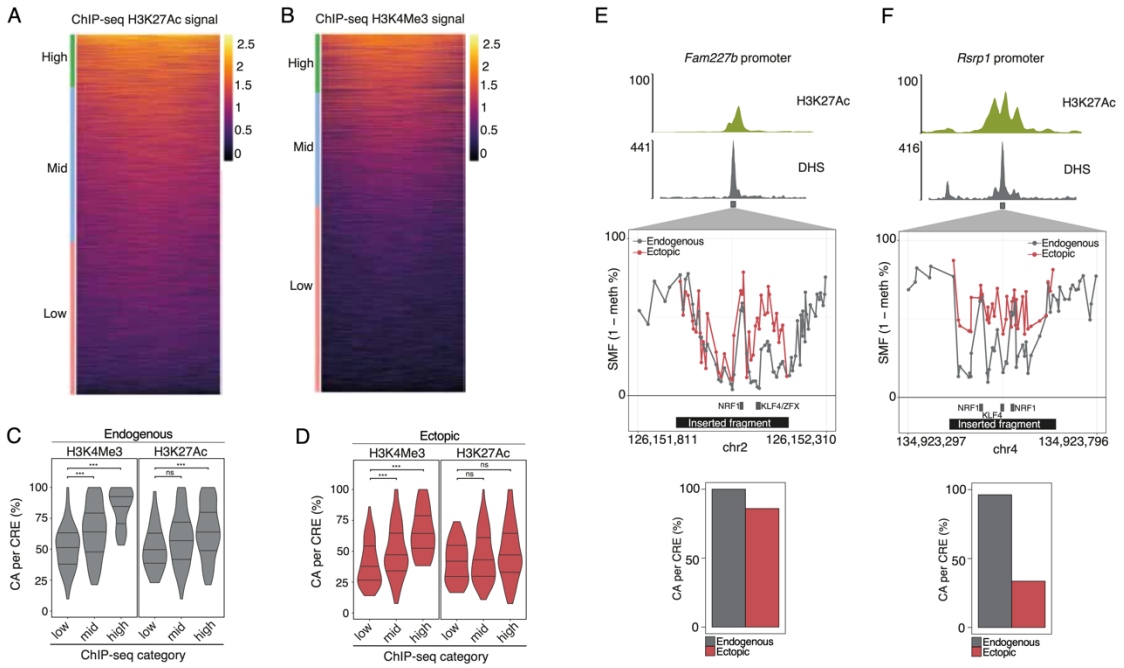


Figure 16. Chromatin accessibility at ectopic sites does not scale with endogenous H3K27Ac levels.

(A) H3K27Ac ChIP signal, ranked from highest to lowest. ChIP signal (log10) was computed over 2 kb windows surrounding TFBSs enriched during SMF bait capture. Enrichment categories were defined as: low (<100), mid (100–500), high (>500). ChIP values were normalized to input and spike-in controls.

(B) Same as (A), but for H3K4Me3 ChIP-seq signal. Enrichment categories were defined as: low (<50), mid (50–300), and high (>300).

(C-D) Violin plots illustrating chromatin accessibility (CA) frequencies at endogenous loci (grey) and ectopic loci (red), grouped by H3K4me3 (left) or H3K27Ac (right) ChIP-seq signal levels at the endogenous site (n = 286). Significance levels: ns = $p > 0.05$; * = $p \leq 0.05$; ** = $p \leq 0.01$; *** = $p \leq 0.001$; **** = $p \leq 0.0001$.

(E) Representative example of a CRE (*Fam227b* promoter) displaying similar chromatin accessibility at both endogenous and ectopic loci. Top: Genome browser tracks showing differential levels of H3K4me3 and H3K27Ac at the DHS region. Middle: average SMF signal (1 – methylation %) comparing the endogenous and ectopic loci. Bottom: chromatin accessibility (CA) frequencies at both loci, quantified using *FootprintCharter*.

(F) Representative example of a CRE (*Rsp1* promoter) showing partial recapitulation of chromatin accessibility at the ectopic locus. Displayed as in panel (E).

To experimentally test this hypothesis, I globally reduced H3K27Ac levels by inhibiting the histone acetyltransferase p300, a factor known to acetylate enhancers in mESCs (Krebs et al., 2011), using the small-molecule inhibitor A-485 (Lasko et al., 2017). Following 24 hours of treatment, H3K27Ac levels were assessed by ChIP-seq, with 5,823 CREs that showed a reduction of H3K27Ac in a significant manner (Figure 17A and Figure 17B). Under the same treatment conditions, chromatin accessibility at these CREs was measured using bait-capture SMF (Figure 17C). The reduction in chromatin accessibility was proportional to the loss of acetylation (Figure 17D and Figure 17F), indicating that H3K27Ac promotes the frequency of chromatin accessibility at these loci.

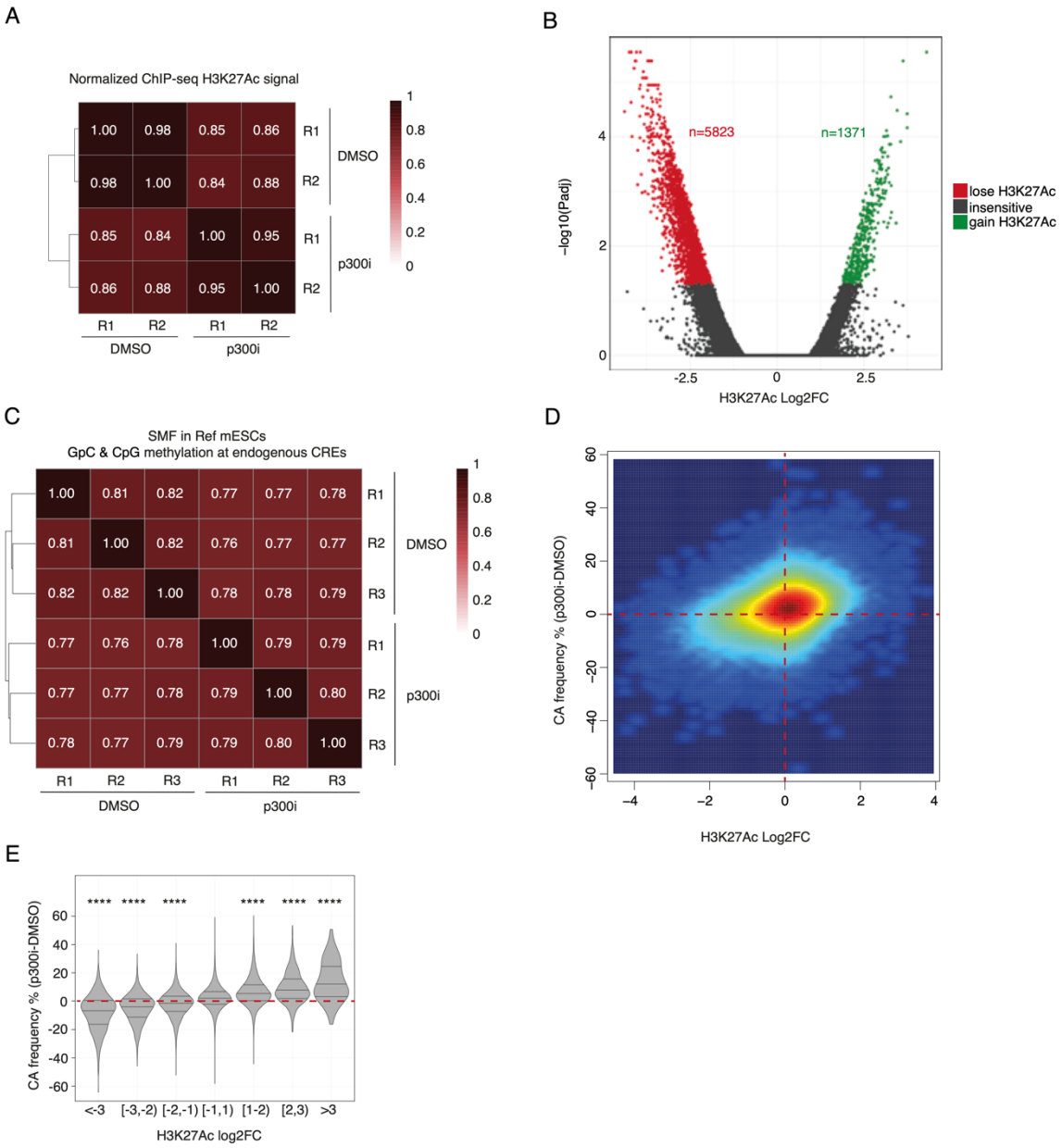


Figure 17. The loss of accessibility scales with the loss of H3K27Ac endogenously.
(A) Pearson correlation heatmap showing H3K27Ac ChIP-seq signal across biological replicates treated with either DMSO or p300 inhibitor. Signal intensities were normalized to input and spike-in controls. Values were calculated over 2 kb windows overlapping genomic regions captured during SMF bait enrichment. The dendrogram represents hierarchical clustering using the complete-linkage method, based on similarity in H3K27Ac binding profiles. Branch lengths reflect relative distances between clusters, and Pearson correlation coefficients R are indicated.
(B) Volcano plot displaying log2 fold changes (Log2FC, x-axis) in H3K27Ac signal and corresponding –log10 p-values (y-axis) following p300 inhibition (p300i), assessed at endogenous loci. Log2FC values were calculated using DESeq2 based on H3K27Ac ChIP-seq counts normalized to input and spike-in controls in DMSO- and p300i-treated samples.
(C) Pearson correlation heatmap of cytosine methylation rates in CpG and GpC contexts across replicates treated with DMSO or p300 inhibition (p300i), as measured by SMF. Analysis was performed on regions corresponding to SMF bait-enriched loci in DNMT triple-knockout mESCs (Ref). Visualization format matches that of panel (A).
(D) Scatterplot showing the relationship between changes in chromatin accessibility following p300 inhibition (CA p300i – CA DMSO; y-axis) and corresponding log2 fold changes in H3K27Ac signal (x-axis) at matched endogenous loci (n=73,529).
(E) Violin boxplots showing that changes in chromatin accessibility scale with changes in H3K27Ac levels following p300 inhibition. Differences in accessibility (CA p300i – CA DMSO) are grouped according to H3K27Ac log2 fold change intervals across endogenous loci (n = 73,529). Statistical significance was calculated over the [-1,1] H3K27Ac log2F class (****$p \leq 0.0001$).

A key implication of this model is that reducing H3K27Ac levels should impact chromatin accessibility specifically at endogenous loci, while leaving ectopic inserts unaffected due to their insertion in a chromatin-neutral context. To test this, I examined p300-sensitive CREs (i.e., endogenous CREs that showed reduced accessibility upon inhibition of p300) and compared them to their corresponding ectopic counterparts. The p300 inhibition experiment was repeated in cell lines harboring the ectopic insert, and chromatin accessibility was assessed using targeted SMF, as before (Figure 18A and Figure 18B). CREs that were not p300-sensitive – such as CTCF-containing regions – showed similar accessibility under both DMSO control and p300-inhibited conditions, independent of the genomic context (Figure 18C and Figure 18D, left panel). In contrast, CREs that lost accessibility at their endogenous location upon p300 inhibition did not exhibit this reduction at the ectopic site (Figure 18D, right panle). Clear examples are the *Mctp2* promoter and the *Xpa* enhancer. The first one, which exhibits low levels of endogenous H3K27Ac, remained unchanged at both the endogenous and ectopic locus following inhibition (Figure 18E). The second one showed a ~20% drop in accessibility frequency at the endogenous locus in response to inhibition, while no change was detected at the ectopic insertion (Figure 18F). These results support the conclusion that H3K27Ac facilitates chromatin accessibility at endogenous CREs, whereas accessibility at ectopic sites is largely driven by TF binding alone.

In conclusion, our data indicate the frequency of chromatin accessibility at CREs is enhanced by the deposition of histone marks like H3K27Ac.

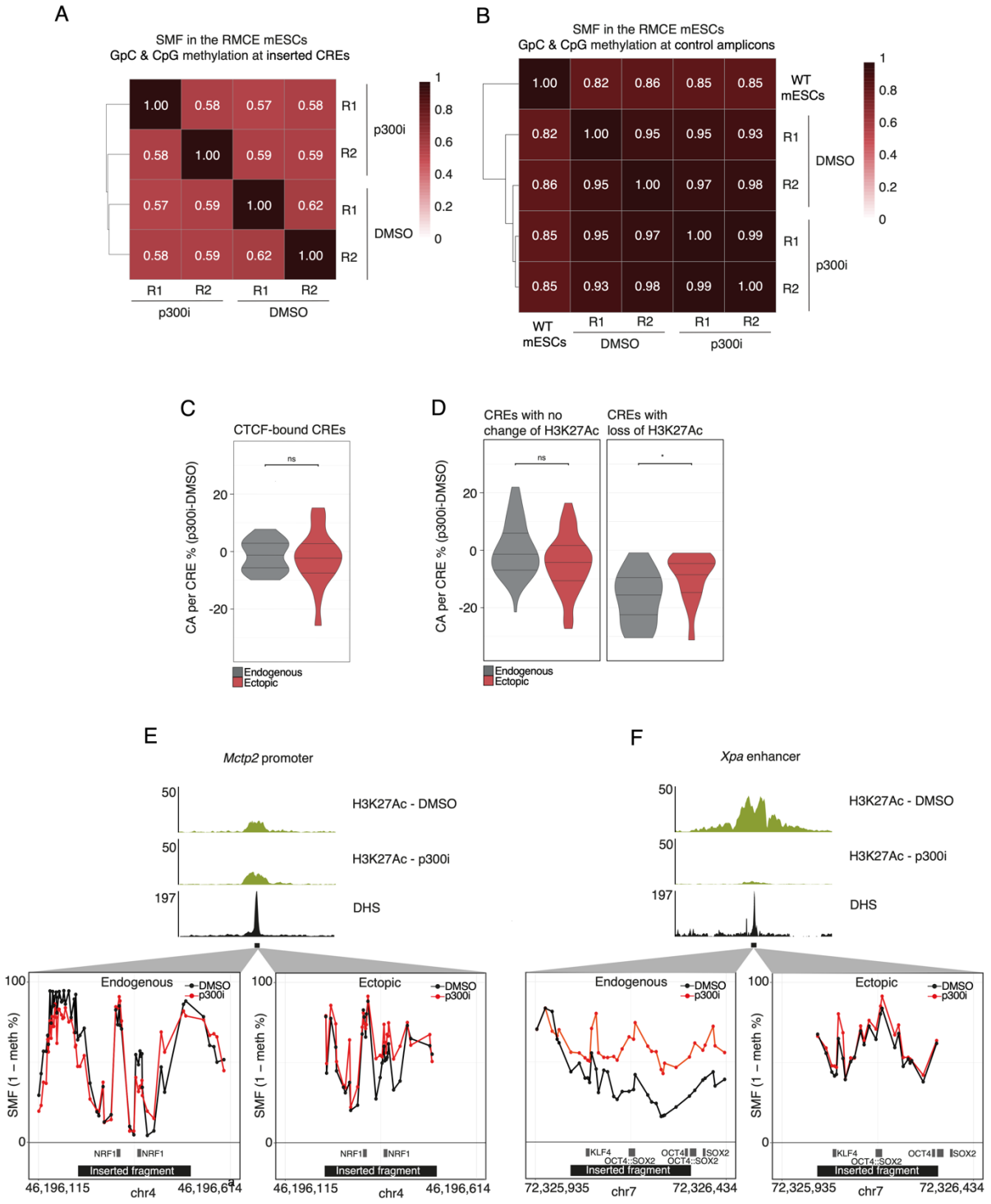


Figure 18. p300 inhibition affects chromatin accessibility endogenously but not ectopically.
(A) Heatmap showing Pearson correlation of GpC and CpG methylation rates at inserted CREs across biological replicates under DMSO and p300 inhibition (p300i) conditions in RMCE mESCs. Pearson correlation coefficients R are indicated.
(B) Heatmap showing Pearson correlation of GpC and CpG methylation patterns at five endogenous control loci across biological replicates in RMCE mESCs treated with DMSO or p300 inhibitor (p300i), as well as a reference mESC dataset. Pearson correlation coefficients R are indicated.

(C) Violin boxplots showing the change in chromatin accessibility (CA p300i – CA DMSO) at CTCF-containing loci, comparing endogenous (grey) and ectopic (red) sites following p300 inhibition (n=23). Statistical significance is indicated as follows: ns ($p > 0.05$), * ($p \leq 0.05$), ** ($p \leq 0.01$).
(D) CREs were grouped based on significant changes in H3K27Ac signal at endogenous sites upon p300i treatment (n=56 and n=13 for the p300i insensitive and sensitive classes, respectively). Data is presented as in panel (C).
(E) Representative example of a promoter that does not exhibit H3K27Ac reduction upon p300 inhibition (p300i), along with its chromatin accessibility profiles at the endogenous (left) and ectopic (right) loci. Top panel: genome browser tracks showing H3K27Ac ChIP-seq signal under DMSO and p300i conditions, alongside chromatin accessibility profiles (DHS peaks). Bottom panel: average SMF signal (1 – methylation%) for DMSO-treated cells (grey) and p300i-treated cells (red).
(F) Example of a p300-regulated enhancer displaying changes in chromatin accessibility upon p300 inhibition at both the endogenous (left) and ectopic (right) loci. Representation is consistent with panel (E).

3.7 Exploratory Applications

3.7.1 Proof of Concept: simultaneous profiling of chromatin accessibility and transcription for multiple CREs at an ectopic site

The method I developed enables the measurement of chromatin accessibility driven by hundreds of CREs inserted in an ectopic genomic locus. I then asked whether this could be coupled with a parallel reporter assay (PRA) to simoultaneously quantify transcriptional activity, therby providing a direct link between chromatin accessibility and gene expression.

In addition to a synthetic flanking sequence that enables targeted SMF of the inserted DNA libraries, the receiving plasmid includes a minimal promoter downstream of each CRE, driving the expression of a reporter gene coupled with a unique barcode. In the plasmid pool, each CRE is associated with at least five different barcodes to ensure that the barcode sequence itself does not influence transcriptional output (Figure 19A, step 1). To assign each barcode to its corresponding CRE, after cloning the pool of CREs, the plasmid library was sequenced using Nanopore technology. The analysis of the Nanopore data and assignment of barcodes to CREs was performed by Kasit Chatsirisupachai, a postdoctoral fellow in the lab.

To evaluate the ability of the assay to simultaneously quantify chromatin accessibility and TF binding via SMF, alongside transcriptional activity via PRA, I inserted a subset of 80 CREs from the previously described library at the ectopic locus using RMCE. As a positive control, I included three strong promoters (i.e., *Cmv*, *Pwp2 and Sxn3*) previously validated in PRA assays (Hartl et al., 2019), which were expected to drive high, robust, and reproducible levels of reporter gene expression. From the same tissue plates, I collected cells for both SMF and transcriptomic analysis. For the latter, total RNA and

genomic DNA (used for transcription normalization) were extracted, and the RNA was subsequently reverse transcribed. I then amplified the barcodes using specific primers and quantified their abundance via high-throughput sequencing to measure the transcriptional activity of each CRE (Figure 19A, step 2). By normalizing RNA barcode frequency to DNA barcode frequency, I obtained relative and reproducible expression levels for 23 constructs (Figure 19B).

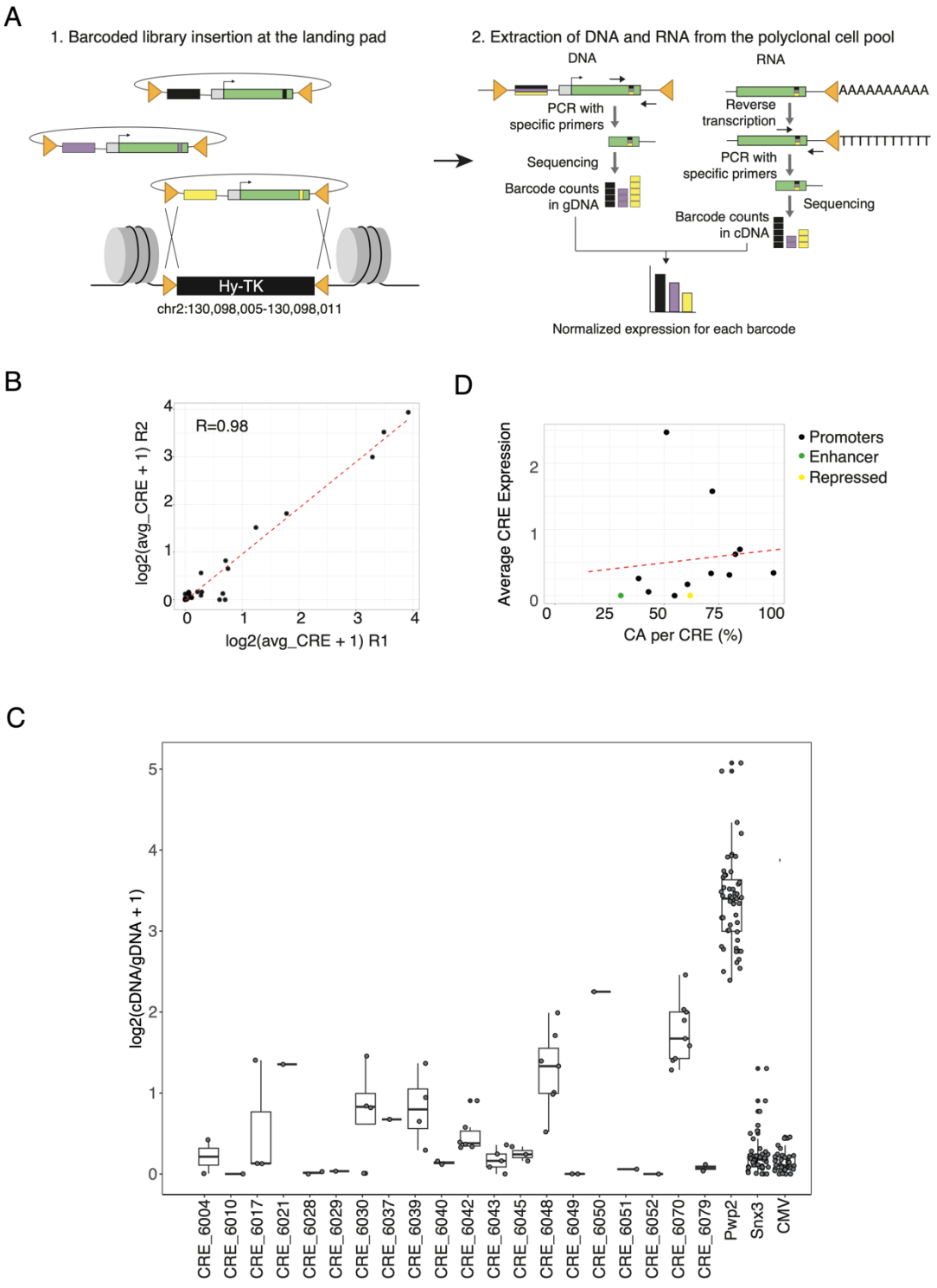


Figure 19. Simultaneous quantification of chromatin accessibility and transcriptional activity at an ectopic locus.
(A) Outline of the experimental strategy. After RMCE and polyclonal cell pool generation, genomic DNA (gDNA) and total RNA will be isolated, and the latter retrotranscribed. Barcodes will be amplified using specific primers and their abundance will be determined with high-throughput sequencing. The ratio between the barcode counts in the gDNA and RNA will give a measure of the CRE activity.
(B) Scatterplot showing the reproducibility of the average reporter gene expression (y-axis) for individual CREs across two biological replicates. Pearson's correlation coefficient R is annotated.
(C) Boxplot showing transcriptional activity measured via parallel reporter assay (PRA) for 23 inserted CREs, including three previously validated control promoters (*Pwp2*, *Snx3*, *Cmv*). Reporter activity is represented as log2(RNA barcode counts / DNA barcode counts + 1) for each CRE. Each dot represents the expression level of a unique barcode corresponding to that specific CRE.
(D) Scatter plot displaying the relationship between average chromatin accessibility (CA per CRE, x-axis) and average reporter gene expression (y-axis) for individual CREs. Each dot represents one CRE, colored by CRE type. A linear regression (dashed red line) is annotated.

Notably, one of the three control promoters, *Pwp2*, drove very high transcription levels compared to the other control promoters and tested CREs (Figure 19C). As a result, it took up a large fraction of the total sequencing reads, thereby reducing the dynamic range for detecting lower-expression constructs. Despite this limitation, transcriptional activity was reproducibly detected for 23 CREs (Figure 19C). When comparing average CRE accessibility to expression levels, I observed a weak overall correlation (Figure 19D). However, this analysis was performed solely to validate the robustness and sensitivity of the assay in capturing both chromatin accessibility and transcriptional output at an ectopic locus.

3.7.2 Proof of Concept: testing the contribution of motif syntax to TF occupancy

A key advantage of the chromatin-integrated reporter assay I established is that the landing pad is located in a chromatin-neutral environment. This allows the intrinsic ability of TFs to open chromatin to be assessed independently of chromatin context. Building on this rationale, I asked whether the system could also be leveraged to test whether motif syntax influences TF binding and cooperativity.

To address this, a locus containing a pair of NFYA motifs that re-establish chromatin accessibility when inserted at the landing pad was selected from the previously described library (Figure 20A). Based on this locus, I designed synthetic DNA sequences in which motif configuration was systematically altered: inter-motif distances were varied (20, 25, 40, 70, 100 bp), motif orientation was inverted, and individual motifs were mutated (through randomization of the motif to abolish TF affinity and binding).

I cloned the library in the RMCE plasmid and performed the insertion of the pool at the landing pad. After target SMF, I then investigated how each of these modifications affected TF binding. Notably, mutation of the upstream motif led to a complete loss of binding at both TFBSs, while mutating the downstream motif still allowed detection of a small fraction of TF-bound molecules at both sites (Figure 20B), suggesting incomplete disruption of TF affinity. This directional effect points to a potential hierarchy or dependency within the motif pair. Next, I examined the effect of spacing between motifs.

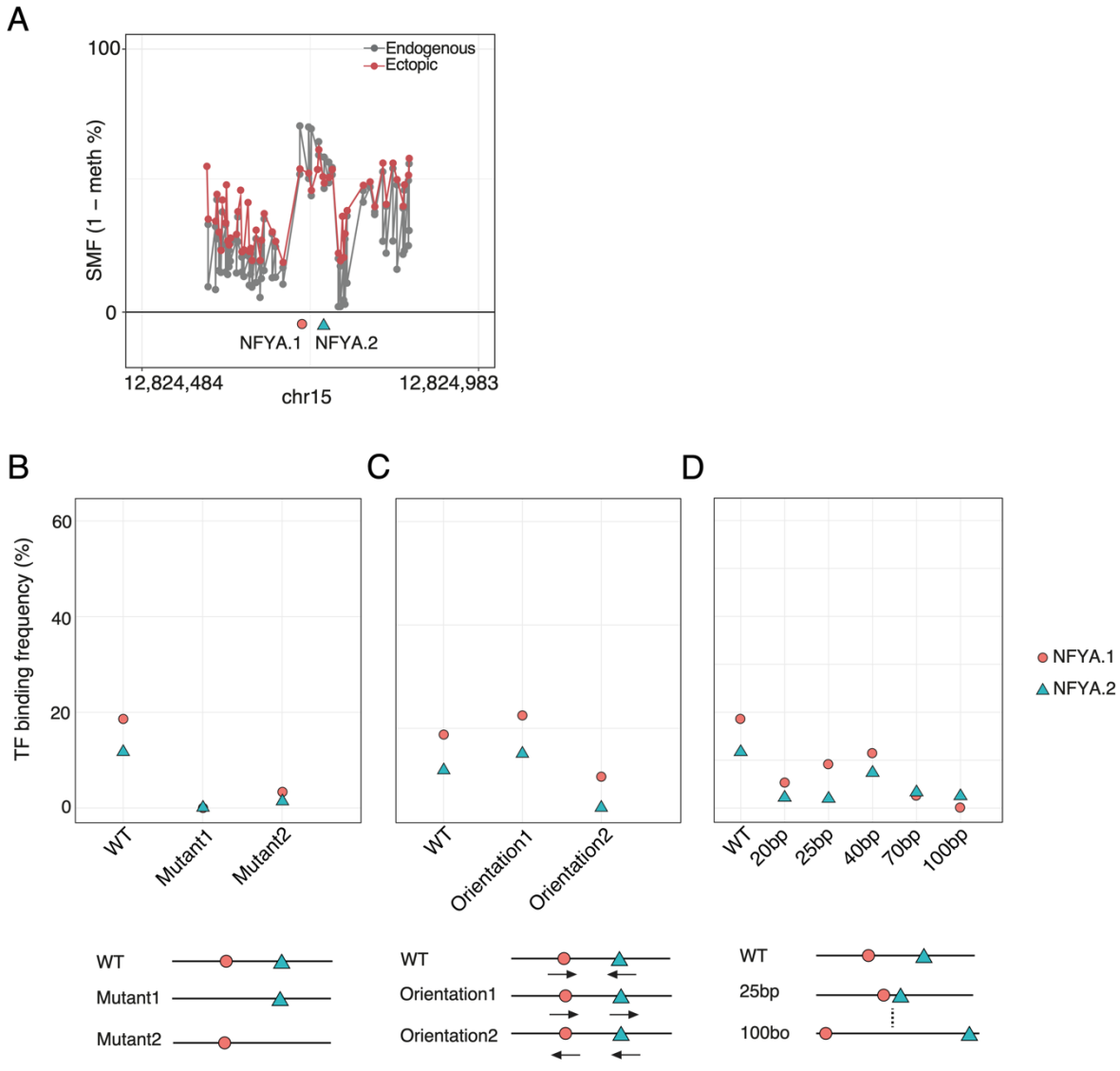


Figure 20. Testing the contribution of motif syntax on the binding of an NFYA pair.
(A) Average SMF signal (1 – methylation %) at the endogenous NFYA-bound locus (grey) and the ectopic insertion site (red), showing two NFYA motifs (NFYA.1 and NFYA.2).
(B–D) Quantification of TF binding frequency (%) at NFYA.1 (red) and NFYA.2 (blue) for synthetic variants inserted at the ectopic locus. (B) TF binding frequency in wild-type (WT) and motif affinity mutant constructs. (C) TF binding frequency across orientation-inverted constructs. (D) TF occupancy across constructs with increasing inter-motif distances (20 bp, 25 bp, 40 bp, 75 bp, 100 bp). Schematic representations of each construct are shown below the corresponding plots.

The construct with a 40 bp inter-motif distance – matching the endogenous sequence – most closely resembled the wild-type binding pattern. Shorter distances reduced TF binding, while greater distances abolished it entirely (Figure 20C), indicating that precise spacing is required for cooperative TF binding at this locus. Lastly, motif orientation was found to impact TF binding asymmetrically. Inverting one motif had no effect, but inverting the other significantly reduced binding (Figure 20D). Together, these findings suggest that motif distance, orientation, and TFBS presence all influence the cooperative interaction and binding efficiency of TFs.

Overall, this pilot dataset validates the sensitivity of the assay in testing how motif syntax contributes to TF occupancy. Building on this approach, and considering recent community efforts to develop tools such as neural networks that learn regulatory rules from genomic data, this system could also be used to experimentally test the validity of such rules at scale.

4. DISCUSSION AND PERSPECTIVES

During my Ph.D., I developed a method to study the intrinsic ability of short *cis*-regulatory elements (CREs) to establish chromatin accessibility when inserted into a neutral ectopic locus in mouse embryonic stem cells. This assay allows for parallel measurement of chromatin accessibility across hundreds of sequences in a controlled chromatin context. Using single-molecule footprinting (SMF), I profiled hundreds of fragments (250 bp in length) spanning full CREs, including both promoters and enhancers. These sequences are bound by a range of transcription factors, (TFs) including both general and pluripotency-associated TFs. I included CTCF-bound regions as positive controls.

As expected, CTCF-containing sequences (used as a positive control) autonomously re-established chromatin accessibility at identical levels when inserted at the ectopic locus. On the other hand, sequences bound by other TFs, such as NRF1 and NFY, and pluripotency factors like KLF4 and OCT4, only partially reproduce their endogenous accessibility frequency when tested in isolation at the ectopic site. Furthermore, while CpG density correlated with accessibility at endogenous loci, this effect was largely lost at the ectopic site. Stratification of endogenous CREs by their H3K27Ac ChIP-seq signal revealed a positive correlation between acetylation levels and chromatin accessibility frequency, suggesting a link between H3K27Ac and accessibility at native loci. However, this correlation was lost at the ectopic site, indicating that H3K27Ac-dependent effects are specific to the endogenous chromatin context. To directly test if H3K27Ac enhances chromatin accessibility, I chemically inhibited the histone acetyltransferase p300, which deposits H3K27Ac at enhancers in mESCs. p300 inhibition led to a reduction in chromatin accessibility at endogenous loci, with the extent of the decrease in accessibility scaling with the degree of H3K27Ac loss. In line with the initial hypothesis, the effect of p300 inhibition was confined to the endogenous loci, while chromatin accessibility at the ectopic site remained unaffected. For instance, the *Xpa* enhancer – highly acetylated within its endogenous context – exhibited a clear reduction in accessibility only at its native locus but not at the ectopic site following p300 inhibition. This supports the idea that H3K27Ac contributes to chromatin accessibility in a context-dependent manner.

While no association was found between the number of TF motifs and accessibility at ectopic sites in this library, our recent work demonstrated that chromatin accessibility at endogenous enhancers emerges from the combined binding of multiple TFs, rather than

the action of few with a specialized function (Baderna et al., 2025). This is in alignment with recent data showing that most individual TF motifs are insufficient to open chromatin when binding on their own (Grand et al., 2024). The data presented in the paper support a 'cumulative model' in which TFs act in an additive manner to sustain chromatin accessibility over time (Baderna et al., 2025). This model complements the 'pioneer' one, where specific TFs initiate chromatin opening by engaging directly nucleosomal DNA. In this framework, maintenance of an open state may involve broader TF cooperation, potentially facilitated by the presence of destabilized nucleosomes (Brahma and Henikoff, 2020), which are indistinguishable from canonical ones in the current data. To directly test this cumulative model, a dedicated library has been designed, and RMCE combined with SMF profiling will be conducted in upcoming experiments.

Although preliminary, my results suggest that for at least one specific NRF1-bound CRE within a 1 kb region, the underlying DNA sequence alone is sufficient to fully restore chromatin accessibility at the ectopic locus. To build on this finding, it would be interesting to repeat the experiment using additional CREs with varying features (e.g., TF binding motifs) and from different chromatin contexts. With the strategy described here, however, the throughput would remain limited to a few dozen CREs, since each fragment size for each CRE would require the generation of individual cell lines. In this regard, to scale the approach and test many CREs in parallel, a potential solution would be to use Cas9-based enrichment combined with Nanopore sequencing after RMCE – though this remains technically challenging at moment. In principle, this method would enable targeted capture of the landing site and provide both high coverage (which is necessary to analyze accessibility at single-molecule resolution) and long-read data across a library of large fragments (~1–5 kb).

Studying regulatory elements outside their native genomic context revealed that they have higher accessibility frequencies in the presence of the activating histone mark H3K27Ac endogenously. This supports a model in which histone PTMs are not strictly required to open chromatin, but their presence increases the likelihood of CRE to be accessible in a cell population. Many of the CREs I tested were able to recruit TFs; however, the data suggests that they were not capable to efficiently recruit the histone acetyltransferase p300 and establish H3K27Ac at the ectopic locus. Structural analyses of p300 interaction domains suggest that its effective recruitment depends on interactions with multiple TFs (Ferrie et al., 2024). This implies that the 1–5 TF binding sites typically

present in our 250 bp CREs may be inadequate for p300 recruitment, and that cooperative interactions between adjacent CREs or additional flanking elements may be necessary. This is also consistent with my preliminary results showing that for some fragments, the regulatory information beyond the 250 bp core may be required, possibly such as secondary TF motifs or sequences capable of recruiting co-factors like the histone acetyltransferase p300. In light of this, it would be valuable to leverage this RMCE-SMF system in combination with ectopic ChIP-seq or qPCR to define the minimal TF modules sufficient to recruit p300 and drive H3K27 acetylation at an ectopic locus. Additionally, approaches leveraging CRISPR-based systems fused to transcriptional activators – such as p300 or CBP – could be used to deposit acetylation at loci that naturally lack it, or even at the ectopic site, before measuring chromatin accessibility frequency with SMF. This would provide an orthogonal strategy to complement the finding that H3K27ac enhances chromatin accessibility.

The observation that H3K27Ac enhances chromatin accessibility is consistent with previous findings showing that histone acetylation can lower the TFs concentration required for them to be active, as demonstrated for SOX9 (Naqvi et al., 2023). This provides a mechanistic basis for how chromatin context may influence CREs' function. Importantly, this model does not exclude the possibility that H3K27Ac may be instructive for accessibility at specific loci. Recent studies have shown that the transcriptional output driven by de novo H3K27Ac is highly context-dependent – promoting transcription in some settings, while having minimal or no effect in others. These observations suggest that this model could be extended to include a layer of context dependency, where the influence of acetylation on chromatin accessibility and transcription varies according to local genomic features or cellular conditions.

On the same line, recent studies showed that the activity of transcriptional regulators (TRs), other then and including TFs, can be further influenced by the genomic environment in which their target CREs are embedded in. TRs have been found to form transient high-concentration foci in the nucleus at cluster of enhancers (CE) (also named Super-Enhancers), which are important in the control of lineage-specific genes (Basu et al., 2020; Boehning et al., 2018; Boija et al., 2018; Sabari et al., 2018). These loci are bound by unusually high levels of TFs and GTFs (Boehning et al., 2018; Boija et al., 2018; Cho et al., 2018; Chong et al., 2018; Guo et al., 2019). It has been hypothesized that this clustering of CREs could lead to the formation of transient membrane-less

organelles that could help maintain such high concentrations of TRs and thus sustain high expression levels. These findings suggest that CREs could be exposed to various concentrations of TFs depending on their genomic location and argue that this phenomenon could be important for the regulation of cell identity genes. However, the mechanisms by which changes in TR availability influence transcription remain largely undefined. To explore these positional effects more mechanistically, one could insert identical reporter constructs into multiple genomic landing pads selected for their distinct chromatin states or levels of transcriptional activity, as well as the diversity and concentration of transcription regulators they recruit. Complementing this approach, chromatin conformation capture experiments could be employed to investigate whether differences in transcription factor binding and chromatin accessibility frequencies might also stem from variations in chromatin looping—such as enhancer–enhancer or enhancer–promoter interactions.

An aspect that I have not addressed in this study is whether the ability of specific TF modules to open chromatin depends on the activity of a particular chromatin remodeler complex, or whether functional redundancy exists, particularly in CREs bound by a diverse set of TFs. It was shown that different TFs rely on distinct remodeling complexes for stable binding. For example, OCT4, SOX2, NANOG, ESRRB, and REST require the continuous activity of the SWI/SNF complex, whereas NRF1 and CTCF appear unaffected by its inhibition (Iurlaro et al., 2021). In contrast, CTCF, together with NFYA and YY1, depends on the ISWI complex to engage with its binding motifs (Barisic et al., 2019). To investigate this further, one possibility would be to combine the RMCE-SMF system with pharmacological perturbation of chromatin remodelers. BRM014, a selective inhibitor of BRG1 (i.e., the catalytic subunit of the SWI/SNF complex) offers a useful tool for this purpose (Iurlaro et al., 2021). Since the library includes CREs bound by both SWI/SNF-dependent and -independent TFs, such an experiment could reveal if chromatin accessibility driven by TFs requires SWI/SNF activity and whether compensatory mechanisms exist, potentially involving other TFs that recruit alternative remodeling complexes.

A recent study has implemented a framework similar to the one described in this dissertation, linking transcriptional output to TF binding and chromatin dynamics using a library of reporter constructs composed of a CRE containing varying numbers of Tet operator motifs (Doughty et al., 2024). Their results show that the transcriptional response

to increasing motif number is nonlinear and driven by cooperative TF binding, which is in turn enabled by the active eviction of nucleosomes through chromatin remodeler recruitment. Now that the transcriptional reporter system established here is functional, a similar approach could be extended to real CREs. This would allow for systematic investigation of how natural combinations of TF motifs – rather than synthetic repeats – shape chromatin accessibility and transcriptional output, bridging motif syntax with functional enhancer activity in a more biologically relevant context.

5. REFERENCES

Akhtar, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F.A., van Lohuizen, M., van Steensel, B., 2013. Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel. Cell 154, 914–927. https://doi.org/10.1016/j.cell.2013.07.018

Alerasool, N., Leng, H., Lin, Z.-Y., Gingras, A.-C., Taipale, M., 2022. Identification and functional characterization of transcriptional activators in human cells. Molecular Cell 82, 677-695.e7. https://doi.org/10.1016/j.molcel.2021.12.008

Allfrey, V.G., Faulkner, R., Mirsky, A.E., 1964. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proc. Natl. Acad. Sci. U.S.A. 51, 786–794. https://doi.org/10.1073/pnas.51.5.786

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., Zeitlinger, J., 2021. Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat Genet 53, 354–366. https://doi.org/10.1038/s41588-021-00782-6

Avvakumov, N., Côté, J., 2007. The MYST family of histone acetyltransferases and their intimate links to cancer. Oncogene 26, 5395–5407. https://doi.org/10.1038/sj.onc.1210608

Baderna, V., Barzaghi, G., Kleinendorst, R., Chatsirisupachai, K., Moniot-Perron, L., Schopp, M., Hochepied, T., Libert, C., Odom, D.T., Zaugg, J.B., Krebs, A., 2025. Cumulative TF binding and H3K27 Acetylation drive enhancer activation frequency. https://doi.org/10.1101/2025.03.26.645413

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.-F., Coburn, D., Newburger, D.E., Morris, Q., Hughes, T.R., Bulyk, M.L., 2009. Diversity and Complexity in DNA Recognition by Transcription Factors. Science 324, 1720–1723. https://doi.org/10.1126/science.1162327

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

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

Bartman, C.R., Hsu, S.C., Hsiung, C.C.-S., Raj, A., Blobel, G.A., 2016. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol Cell 62, 237–247. https://doi.org/10.1016/j.molcel.2016.03.007

Barzaghi, G., Krebs, A.R., Zaugg, J.B., 2025. FootprintCharter: unsupervised detection and quantification of footprints in single molecule footprinting data.

Basu, S., Mackowiak, S.D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., Geertsema, H., Ali, S., Jerković, I., Ewers, H., Mundlos, S., Meissner, A., Ibrahim, D.M., Hnisz, D., 2020. Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. Cell 181, 1062-1079.e30. https://doi.org/10.1016/j.cell.2020.04.018

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

Bhagwat, A.S., Roe, J.-S., Mok, B.Y.L., Hohmann, A.F., Shi, J., Vakoc, C.R., 2016. BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. Cell Rep 15, 519–530. https://doi.org/10.1016/j.celrep.2016.03.054

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., Darzacq, X., Zweckstetter, M., 2018. RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol 25, 833–840. https://doi.org/10.1038/s41594-018-0112-y

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., Abraham, B.J., Afeyan, L.K., Guo, Y.E., Rimel, J.K., Fant, C.B., Schuijers, J., Lee, T.I., Taatjes, D.J., Young, R.A., 2018. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell 175, 1842-1855.e16. https://doi.org/10.1016/j.cell.2018.10.042

Bolt, C.C., Duboule, D., 2020. The regulatory landscapes of developmental genes. Development 147, dev171736. https://doi.org/10.1242/dev.171736

Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., Levine, M., 2015. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. eLife 4, e07956. https://doi.org/10.7554/eLife.07956

Brownell, J.E., Allis, C.D., 1995. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proc. Natl. Acad. Sci. U.S.A. 92, 6364–6368. https://doi.org/10.1073/pnas.92.14.6364

Bulyk, M.L., Drouin, J., Harrison, M.M., Taipale, J., Zaret, K.S., 2023. Pioneer factors — key regulators of chromatin and gene expression. Nat Rev Genet 24, 809–815. https://doi.org/10.1038/s41576-023-00648-z

Butz, S., Schmolka, N., Karemaker, I.D., Villaseñor, R., Schwarz, I., Domcke, S., Uijttewaal, E.C.H., Jude, J., Lienert, F., Krebs, A.R., De Wagenaar, N.P., Bao, X., Zuber, J., Elling, U., Schübeler, D., Baubec, T., 2022. DNA sequence and chromatin modifiers cooperate to confer epigenetic bistability at imprinting control regions. Nat Genet 54, 1702–1710. https://doi.org/10.1038/s41588-022-01210-z

Calderon, D., Blecher-Gonen, R., Huang, X., Secchia, S., Kentro, J., Daza, R.M., Martin, B., Dulja, A., Schaub, C., Trapnell, C., Larschan, E., O'Connor-Giles, K.M., Furlong, E.E.M., Shendure, J., 2022. The continuum of *Drosophila* embryonic development at single-cell resolution. Science 377, eabn5800. https://doi.org/10.1126/science.abn5800

Carelli, F.N., Liechti, A., Halbert, J., Warnefors, M., Kaessmann, H., 2018. Repurposing of promoters and enhancers during mammalian evolution. Nat Commun 9, 4066. https://doi.org/10.1038/s41467-018-06544-z

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

Chen, Z., Snetkova, V., Bower, G., Jacinto, S., Clock, B., Dizehchi, A., Barozzi, I., Mannion, B.J., Alcaina-Caro, A., Lopez-Rios, J., Dickel, D.E., Visel, A., Pennacchio, L.A., Kvon, E.Z., 2024. Increased enhancer–promoter interactions during developmental enhancer activation in mammals. Nat Genet 56, 675–685. https://doi.org/10.1038/s41588-024-01681-2

Cheng, L., De, C., Li, J., Pertsinidis, A., 2023. Mechanisms of transcription control by distal enhancers from high-resolution single-gene imaging. bioRxiv 2023.03.19.533190. https://doi.org/10.1101/2023.03.19.533190

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., Cisse, I.I., 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science 361, 412–415. https://doi.org/10.1126/science.aar4199

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., Tjian, R., 2018. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science 361, eaar2555. https://doi.org/10.1126/science.aar2555

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., Zaret, K.S., 2002. Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. Molecular Cell 9, 279–289. https://doi.org/10.1016/S1097-2765(02)00459-8

Contreras-Martos, S., Piai, A., Kosol, S., Varadi, M., Bekesi, A., Lebrun, P., Volkov, A.N., Gevaert, K., Pierattelli, R., Felli, I.C., Tompa, P., 2017. Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. Sci Rep 7, 4676. https://doi.org/10.1038/s41598-017-04611-x

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., Lis, J.T., 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet 46, 1311–1320. https://doi.org/10.1038/ng.3142

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

Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R.S., Stern, D.L., 2015. Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. Cell 160, 191–203. https://doi.org/10.1016/j.cell.2014.11.041

Crocker, J., Preger-Ben Noon, E., Stern, D.L., 2016. The Soft Touch, in: Current Topics in Developmental Biology. Elsevier, pp. 455–469. https://doi.org/10.1016/bs.ctdb.2015.11.018

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

Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., Lee, C., Regalado, S.G., Read, D.F., Steemers, F.J., Disteche, C.M., Trapnell, C., Shendure, J., 2018. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell 174, 1309-1324.e18. https://doi.org/10.1016/j.cell.2018.06.052

Cusanovich, D.A., Pavlovic, B., Pritchard, J.K., Gilad, Y., 2014. The Functional Consequences of Variation in Transcription Factor Binding. PLoS Genet 10, e1004226. https://doi.org/10.1371/journal.pgen.1004226

De Almeida, B.P., Reiter, F., Pagani, M., Stark, A., 2022. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. Nat Genet 54, 613–624. https://doi.org/10.1038/s41588-022-01048-5

De Almeida, B.P., Schaub, C., Pagani, M., Secchia, S., Furlong, E.E.M., Stark, A., 2024. Targeted design of synthetic enhancers for selected tissues in the Drosophila embryo. Nature 626, 207–211. https://doi.org/10.1038/s41586-023-06905-9

De Boer, C.G., Taipale, J., 2024. Hold out the genome: a roadmap to solving the cis-regulatory code. Nature 625, 41–50. https://doi.org/10.1038/s41586-023-06661-w

Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. Genes Dev. 25, 1010–1022. https://doi.org/10.1101/gad.2037511

Dekker, J., Rippe, K., Dekker, M., Kleckner, N., 2002. Capturing Chromosome Conformation 295.

DelRosso, N., Tycko, J., Suzuki, P., Andrews, C., Aradhana, Mukund, A., Liongson, I., Ludwig, C., Spees, K., Fordyce, P., Bassik, M.C., Bintu, L., 2023. Large-scale

mapping and mutagenesis of human transcriptional effector domains. Nature 616, 365–372. https://doi.org/10.1038/s41586-023-05906-y
Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., Schübeler, D., 2015. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature 528, 575–579. https://doi.org/10.1038/nature16462
Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., Green, R.D., Dekker, J., 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16, 1299–1309. https://doi.org/10.1101/gr.5571506
Doughty, B.R., Hinks, M.M., Schaepe, J.M., Marinov, G.K., Thurm, A.R., Rios-Martinez, C., Parks, B.E., Tan, Y., Marklund, E., Dubocanin, D., Bintu, L., Greenleaf, W.J., 2024. Single-molecule states link transcription factor binding to gene expression. Nature 636, 745–754. https://doi.org/10.1038/s41586-024-08219-w
Erceg, J., Saunders, T.E., Girardot, C., Devos, D.P., Hufnagel, L., Furlong, E.E.M., 2014. Subtle Changes in Motif Positioning Cause Tissue-Specific Effects on Robustness of an Enhancer's Activity. PLOS Genetics 10, e1004060. https://doi.org/10.1371/journal.pgen.1004060
Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., Levine, M.S., 2015. Suboptimization of developmental enhancers. Science 350, 325–328. https://doi.org/10.1126/science.aac6948
Ferrie, J.J., Karr, J.P., Graham, T.G.W., Dailey, G.M., Zhang, G., Tjian, R., Darzacq, X., 2024. p300 is an obligate integrator of combinatorial transcription factor inputs. Molecular Cell 84, 234-243.e4. https://doi.org/10.1016/j.molcel.2023.12.004
Fuda, N.J., Ardehali, M.B., Lis, J.T., 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192. https://doi.org/10.1038/nature08449
Fukaya, T., 2023. Enhancer dynamics: Unraveling the mechanism of transcriptional bursting. Science Advances 9, eadj3366. https://doi.org/10.1126/sciadv.adj3366
Fukaya, T., Lim, B., Levine, M., 2016. Enhancer Control of Transcriptional Bursting. Cell 166, 358–368. https://doi.org/10.1016/j.cell.2016.05.025
Furlong, E.E.M., Levine, M., 2018. Developmental enhancers and chromosome topology. Science 361, 1341–1345. https://doi.org/10.1126/science.aau0320
Gasperini, M., Tome, J.M., Shendure, J., 2020. Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat Rev Genet 21, 292–310. https://doi.org/10.1038/s41576-019-0209-0
Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.-J., Tsai, S.Y., Li, W., Foulds, C.E., O'Malley, B.W., 2017. Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. J Biol Chem 292, 14456–14472. https://doi.org/10.1074/jbc.M117.802074
Gillespie, M.A., Palii, C.G., Sanchez-Taltavull, D., Shannon, P., Longabaugh, W.J.R., Downes, D.J., Sivaraman, K., Espinoza, H.M., Hughes, J.R., Price, N.D., Perkins, T.J., Ranish, J.A., Brand, M., 2020. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. Molecular Cell 78, 960-974.e11. https://doi.org/10.1016/j.molcel.2020.03.031
Grand, R.S., Pregnolato, M., Baumgartner, L., Hoerner, L., Burger, L., Schübeler, D., 2024. Genome access is transcription factor-specific and defined by nucleosome position. Molecular Cell 84, 3455-3468.e6. https://doi.org/10.1016/j.molcel.2024.08.009

Greenberg, M.V.C., 2021. Get Out and Stay Out: New Insights Into DNA Methylation Reprogramming in Mammals. Front. Cell Dev. Biol. 8. https://doi.org/10.3389/fcell.2020.629068
Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 20, 590–607. https://doi.org/10.1038/s41580-019-0159-6
Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313
Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R., Zaret, K.S., 1996. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev. 10, 1670–1682. https://doi.org/10.1101/gad.10.13.1670
Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., Stark, A., 2019. Transcriptional cofactors display specificity for distinct types of core promoters. Nature 570, 122–126. https://doi.org/10.1038/s41586-019-1210-7
Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., Van IJcken, W.F.J., Armant, O., Ferg, M., Strähle, U., Carninci, P., Müller, F., Lenhard, B., 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. Nature 507, 381–385. https://doi.org/10.1038/nature12974
Haberle, V., Stark, A., 2018. Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol 19, 621–637. https://doi.org/10.1038/s41580-018-0028-8
Hartl, D., Krebs, A.R., Grand, R.S., Baubec, T., Isbel, L., Wirbelauer, C., Burger, L., Schübeler, D., 2019. CG dinucleotides enhance promoter activity independent of DNA methylation. Genome Res. 29, 554–563. https://doi.org/10.1101/gr.241653.118
Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., Sharp, P.A., 2017. A Phase Separation Model for Transcriptional Control. Cell 169, 13–23. https://doi.org/10.1016/j.cell.2017.02.007
Hong, C.K.Y., Cohen, B.A., 2022. Genomic environments scale the activities of diverse core promoters. Genome Res. 32, 85–96. https://doi.org/10.1101/gr.276025.121
Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., Darzacq, X., 2020. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol Cell 78, 539-553.e8. https://doi.org/10.1016/j.molcel.2020.03.002
Hughes, T.R. (Ed.), 2011. A Handbook of Transcription Factors, Subcellular Biochemistry. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-90-481-9069-0
Hunt, G., Boija, A., Mannervik, M., 2022. p300/CBP sustains Polycomb silencing by non-enzymatic functions. Molecular Cell 82, 3580-3597.e9. https://doi.org/10.1016/j.molcel.2022.09.005
Ibrahim, D.M., Mundlos, S., 2020. The role of 3D chromatin domains in gene regulation: a multi-facetted view on genome organization. Current Opinion in Genetics & Development, Genome Architecture and Expression 61, 1–8. https://doi.org/10.1016/j.gde.2020.02.015
Inoue, F., Ahituv, N., 2015. Decoding enhancers using massively parallel reporter assays. Genomics 106, 159–164. https://doi.org/10.1016/j.ygeno.2015.06.005
Iurlaro, M., Stadler, M.B., Masoni, F., Jagani, Z., Galli, G.G., Schübeler, D., 2021. Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat Genet 53, 279–287. https://doi.org/10.1038/s41588-020-00768-w
Jindal, G.A., Bantle, A.T., Solvason, J.J., Grudzien, J.L., D'Antonio-Chronowska, A., Lim, F., Le, S.H., Song, B.P., Ragsac, M.F., Klie, A., Larsen, R.O., Frazer, K.A., Farley,

E.K., 2023. Single-nucleotide variants within heart enhancers increase binding affinity and disrupt heart development. Developmental Cell 58, 2206-2216.e5. https://doi.org/10.1016/j.devcel.2023.09.005
Jindal, G.A., Farley, E.K., 2021. Enhancer grammar in development, evolution, and disease: dependencies and interplay. Developmental Cell 56, 575–587. https://doi.org/10.1016/j.devcel.2021.02.016
Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T., Taipale, J., 2013. DNA-Binding Specificities of Human Transcription Factors. Cell 152, 327–339. https://doi.org/10.1016/j.cell.2012.12.009
Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., Taipale, J., 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature 527, 384–388. https://doi.org/10.1038/nature15518
Josling, G.A., Selvarajah, S.A., Petter, M., Duffy, M.F., 2012. The Role of Bromodomain Proteins in Regulating Gene Expression. Genes 3, 320–343. https://doi.org/10.3390/genes3020320
Kaluscha, S., Domcke, S., Wirbelauer, C., Stadler, M.B., Durdu, S., Burger, L., Schübeler, D., 2022. Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. Nat Genet 54, 1895–1906. https://doi.org/10.1038/s41588-022-01241-6
Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., Tora, L., 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics 13, 424. https://doi.org/10.1186/1471-2164-13-424
Kempynck, N., Winter, S.D., Blaauw, C.H., Konstantakos, V., Dieltiens, S., Ekşi, E.C., Bercier, V., Taskiran, I.I., Hulselmans, G., Spanier, K., Christiaens, V., Bosch, L.V.D., Mahieu, L., Aerts, S., 2025. CREsted: modeling genomic and synthetic cell type-specific enhancers across tissues and species. https://doi.org/10.1101/2025.04.02.646812
Kim, S., Wysocka, J., 2023. Deciphering the multi-scale, quantitative cis-regulatory code. Molecular Cell 83, 373–392. https://doi.org/10.1016/j.molcel.2022.12.032
Kleinendorst, R.W.D., Barzaghi, G., Smith, M.L., Zaugg, J.B., Krebs, A.R., 2021. Genome-wide quantification of transcription factor binding at single-DNA-molecule resolution using methyl-transferase footprinting. Nat Protoc 16, 5673–5706. https://doi.org/10.1038/s41596-021-00630-1
Klemm, S.L., Shipony, Z., Greenleaf, W.J., 2019. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 20, 207–220. https://doi.org/10.1038/s41576-018-0089-8
Klughammer, J., Romanovskaia, D., Nemc, A., Posautz, A., Seid, C.A., Schuster, L.C., Keinath, M.C., Lugo Ramos, J.S., Kosack, L., Evankow, A., Printz, D., Kirchberger, S., Ergüner, B., Datlinger, P., Fortelny, N., Schmidl, C., Farlik, M., Skjærven, K., Bergthaler, A., Liedvogel, M., Thaller, D., Burger, P.A., Hermann, M., Distel, M., Distel, D.L., Kübber-Heiss, A., Bock, C., 2023. Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. Nat Commun 14, 232. https://doi.org/10.1038/s41467-022-34828-y
Kornberg, R.D., Lorch, Y., 1999. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. Cell 98, 285–294. https://doi.org/10.1016/S0092-8674(00)81958-3
Krebs, A.R., 2021. Studying transcription factor function in the genome at molecular resolution. Trends in Genetics 37, 798–806. https://doi.org/10.1016/j.tig.2021.03.008

Krebs, A.R., Dessus-Babus, S., Burger, L., Schübeler, D., 2014. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. eLife 3, e04094. https://doi.org/10.7554/eLife.04094
Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., Schübeler, D., 2017. Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Molecular Cell 67, 411-422.e4. https://doi.org/10.1016/j.molcel.2017.06.027
Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., Tora, L., 2011. SAGA and ATAC Histone Acetyl Transferase Complexes Regulate Distinct Sets of Genes and ATAC Defines a Class of p300-Independent Enhancers. Molecular Cell 44, 410–423. https://doi.org/10.1016/j.molcel.2011.08.037
Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S., Krebs, A.R., 2023. Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. Molecular Cell 83, 787-802.e9. https://doi.org/10.1016/j.molcel.2023.01.017
Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., Stark, A., 2014. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature 512, 91–95. https://doi.org/10.1038/nature13395
Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., Weirauch, M.T., 2018. The Human Transcription Factors. Cell 172, 650–665. https://doi.org/10.1016/j.cell.2018.01.029
Lasko, L.M., Jakob, C.G., Edalji, R.P., Qiu, W., Montgomery, D., Digiammarino, E.L., Hansen, T.M., Risi, R.M., Frey, R., Manaves, V., Shaw, B., Algire, M., Hessler, P., Lam, L.T., Uziel, T., Faivre, E., Ferguson, D., Buchanan, F.G., Martin, R.L., Torrent, M., Chiang, G.G., Karukurichi, K., Langston, J.W., Weinert, B.T., Choudhary, C., De Vries, P., Kluge, A.F., Patane, M.A., Van Drie, J.H., Wang, C., McElligott, D., Kesicki, E., Marmorstein, R., Sun, C., Cole, P.A., Rosenberg, S.H., Michaelides, M.R., Lai, A., Bromberg, K.D., 2017. Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. Nature 550, 128–132. https://doi.org/10.1038/nature24028
Lee, K.K., Workman, J.L., 2007. Histone acetyltransferase complexes: one size doesn't fit all. Nat Rev Mol Cell Biol 8, 284–295. https://doi.org/10.1038/nrm2145
Lenhard, B., Sandelin, A., Carninci, P., 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet 13, 233–245. https://doi.org/10.1038/nrg3163
Levo, M., Raimundo, J., Bing, X.Y., Sisco, Z., Batut, P.J., Ryabichko, S., Gregor, T., Levine, M.S., 2022. Transcriptional coupling of distant regulatory genes in living embryos. Nature 605, 754–760. https://doi.org/10.1038/s41586-022-04680-7
Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science 326, 289–293. https://doi.org/10.1126/science.1181369
Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., Schübeler, D., 2011. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 43, 1091–1097. https://doi.org/10.1038/ng.946
Lim, F., Solvason, J.J., Ryan, G.E., Le, S.H., Jindal, G.A., Steffen, P., Jandu, S.K., Farley, E.K., 2024. Affinity-optimizing enhancer variants disrupt development. Nature 626, 151–159. https://doi.org/10.1038/s41586-023-06922-8

Liu, X., Wang, L., Zhao, K., Thompson, P.R., Hwang, Y., Marmorstein, R., Cole, P.A., 2008. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. Nature 451, 846–850. https://doi.org/10.1038/nature06546
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. https://doi.org/10.1186/s13059-014-0550-8
Ma, L., Gao, Z., Wu, J., Zhong, B., Xie, Y., Huang, W., Lin, Y., 2021. Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics. Molecular Cell 81, 1682-1697.e7. https://doi.org/10.1016/j.molcel.2021.01.031
Maniatis, T., 1995. Virus Induction of Human IFNP Gene Expression Requires the Assembly of an Enhanceosome.
Maricque, B.B., Chaudhari, H.G., Cohen, B.A., 2019. A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. Nat Biotechnol 37, 90–95. https://doi.org/10.1038/nbt.4285
Marmorstein, R., 2001. Structure of histone acetyltransferases. J Mol Biol 311, 433–444. https://doi.org/10.1006/jmbi.2001.4859
Marmorstein, R., Zhou, M.-M., 2014. Writers and readers of histone acetylation: structure, mechanism, and inhibition. Cold Spring Harb Perspect Biol 6, a018762. https://doi.org/10.1101/cshperspect.a018762
Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A.W., Parcy, F., Lenhard, B., Sandelin, A., Wasserman, W.W., 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 44, D110–D115. https://doi.org/10.1093/nar/gkv1176
Mattei, A.L., Bailly, N., Meissner, A., 2022. DNA methylation: a historical perspective. Trends in Genetics 38, 676–707. https://doi.org/10.1016/j.tig.2022.03.010
Medina-Rivera, A., Santiago-Algarra, D., Puthier, D., Spicuglia, S., 2018. Widespread Enhancer Activity from Core Promoters. Trends in Biochemical Sciences 43, 452–468. https://doi.org/10.1016/j.tibs.2018.03.004
Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., Lander, E.S., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766–770. https://doi.org/10.1038/nature07107
Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., Furlong, E.E.M., 2018. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. 32, 42–57. https://doi.org/10.1101/gad.308619.117
Millán-Zambrano, G., Burton, A., Bannister, A.J., Schneider, R., 2022. Histone post-translational modifications — cause and consequence of genome function. Nat Rev Genet 23, 563–580. https://doi.org/10.1038/s41576-022-00468-7
Minnoye, L., Marinov, G.K., Krausgruber, T., Pan, L., Marand, A.P., Secchia, S., Greenleaf, W.J., Furlong, E.E.M., Zhao, K., Schmitz, R.J., Bock, C., Aerts, S., 2021. Chromatin accessibility profiling methods. Nat Rev Methods Primers 1, 10. https://doi.org/10.1038/s43586-020-00008-9
Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., Chambon, P., 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res 9, 6047–6068. https://doi.org/10.1093/nar/9.22.6047

Morgunova, E., Taipale, J., 2017. Structural perspective of cooperative transcription factor binding. Current Opinion in Structural Biology 47, 1–8. https://doi.org/10.1016/j.sbi.2017.03.006
Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386–389. https://doi.org/10.1038/30764
Naqvi, S., Kim, S., Hoskens, H., Matthews, H.S., Spritz, R.A., Klein, O.D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J.K., Wysocka, J., 2023. Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. Nat Genet 55, 841–851. https://doi.org/10.1038/s41588-023-01366-2
Narita, T., Higashijima, Y., Kilic, S., Liebner, T., Walter, J., Choudhary, C., 2023. Acetylation of histone H2B marks active enhancers and predicts CBP/p300 target genes. Nat Genet 55, 679–692. https://doi.org/10.1038/s41588-023-01348-4
Narita, T., Ito, S., Higashijima, Y., Chu, W.K., Neumann, K., Walter, J., Satpathy, S., Liebner, T., Hamilton, W.B., Maskey, E., Prus, G., Shibata, M., Iesmantavicius, V., Brickman, J.M., Anastassiadis, K., Koseki, H., Choudhary, C., 2021. Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. Molecular Cell 81, 2166-2182.e6. https://doi.org/10.1016/j.molcel.2021.03.008
Neumayr, C., Haberle, V., Serebreni, L., Karner, K., Hendy, O., Boija, A., Henninger, J.E., Li, C.H., Stejskal, K., Lin, G., Bergauer, K., Pagani, M., Rath, M., Mechtler, K., Arnold, C.D., Stark, A., 2022. Differential cofactor dependencies define distinct types of human enhancers. Nature 606, 406–413. https://doi.org/10.1038/s41586-022-04779-x
Nguyen, T.A., Jones, R.D., Snavely, A.R., Pfenning, A.R., Kirchner, R., Hemberg, M., Gray, J.M., 2016. High-throughput functional comparison of promoter and enhancer activities. Genome Res. 26, 1023–1033. https://doi.org/10.1101/gr.204834.116
Nguyen, T.D., Chen, Y.-I., Chen, L.H., Yeh, H.-C., 2023. Recent Advances in Single-Molecule Tracking and Imaging Techniques. Annual Rev. Anal. Chem. 16, 253–284. https://doi.org/10.1146/annurev-anchem-091922-073057
Patel, L.A., Cao, Y., Mendenhall, E.M., Benner, C., Goren, A., 2024. The Wild West of spike-in normalization. Nat Biotechnol 42, 1343–1349. https://doi.org/10.1038/s41587-024-02377-y
Patel, M.C., Debrosse, M., Smith, M., Dey, A., Huynh, W., Sarai, N., Heightman, T.D., Tamura, T., Ozato, K., 2013. BRD4 coordinates recruitment of pause release factor P-TEFb and the pausing complex NELF/DSIF to regulate transcription elongation of interferon-stimulated genes. Mol Cell Biol 33, 2497–2507. https://doi.org/10.1128/MCB.01180-12
Perissi, V., Jepsen, K., Glass, C.K., Rosenfeld, M.G., 2010. Deconstructing repression: evolving models of co-repressor action. Nat Rev Genet 11, 109–123. https://doi.org/10.1038/nrg2736
Pimmett, V.L., Dejean, M., Fernandez, C., Trullo, A., Bertrand, E., Radulescu, O., Lagha, M., 2021. Quantitative imaging of transcription in living Drosophila embryos reveals the impact of core promoter motifs on promoter state dynamics. Nat Commun 12, 4504. https://doi.org/10.1038/s41467-021-24461-6
Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., Brockdorff, N., 2017. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Molecular Cell 68, 955-969.e10. https://doi.org/10.1016/j.molcel.2017.11.013

Policarpi, C., Munafò, M., Tsagkris, S., Carlini, V., Hackett, J.A., 2024. Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications. Nat Genet 56, 1168–1180. https://doi.org/10.1038/s41588-024-01706-w

Pollex, T., Rabinowitz, A., Gambetta, M.C., Marco-Ferreres, R., Viales, R.R., Jankowski, A., Schaub, C., Furlong, E.E.M., 2024. Enhancer–promoter interactions become more instructive in the transition from cell-fate specification to tissue differentiation. Nat Genet 56, 686–696. https://doi.org/10.1038/s41588-024-01678-x

Pomp, W., Meeussen, J.V.W., Lenstra, T.L., 2024. Transcription factor exchange enables prolonged transcriptional bursts. Molecular Cell 84, 1036-1048.e9. https://doi.org/10.1016/j.molcel.2024.01.020

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

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., Huang, X., Shamim, M.S., Shin, J., Turner, D., Ye, Z., Omer, A.D., Robinson, J.T., Schlick, T., Bernstein, B.E., Casellas, R., Lander, E.S., Aiden, E.L., 2017. Cohesin Loss Eliminates All Loop Domains. Cell 171, 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026

Reddington, J.P., Garfield, D.A., Sigalova, O.M., Karabacak Calviello, A., Marco-Ferreres, R., Girardot, C., Viales, R.R., Degner, J.F., Ohler, U., Furlong, E.E.M., 2020. Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. Dev Cell 55, 648-664.e9. https://doi.org/10.1016/j.devcel.2020.10.009

Rodriguez, J., Larson, D.R., 2020. Transcription in Living Cells: Molecular Mechanisms of Bursting. Annu. Rev. Biochem. 89, 189–212. https://doi.org/10.1146/annurev-biochem-011520-105250

Roeder, R.G., 1996. The role of general initiation factors in transcription by RNA polymerase II. Trends in Biochemical Sciences 21, 327–335. https://doi.org/10.1016/S0968-0004(96)10050-5

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., Li, C.H., Guo, Y.E., Day, D.S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T.I., Cisse, I.I., Roeder, R.G., Sharp, P.A., Chakraborty, A.K., Young, R.A., 2018. Coactivator condensation at super-enhancers links phase separation and gene control. Science 361, eaar3958. https://doi.org/10.1126/science.aar3958

Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O., Cramer, P., Kivioja, T., Taipale, J., 2022. Sequence determinants of human gene regulatory elements. Nat Genet 54, 283–294. https://doi.org/10.1038/s41588-021-01009-4

Sandelin, A., 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Research 32, 91D – 94. https://doi.org/10.1093/nar/gkh012

Scardigli, R., Bäumer, N., Gruss, P., Guillemot, F., Le Roux, I., 2003. Direct and concentration-dependent regulation of the proneural gene *Neurogenin2* by Pax6. Development 130, 3269–3281. https://doi.org/10.1242/dev.00539

Schick, S., Grosche, S., Kohl, K.E., Drpic, D., Jaeger, M.G., Marella, N.C., Imrichova, H., Lin, J.-M.G., Hofstätter, G., Schuster, M., Rendeiro, A.F., Koren, A., Petronczki, M., Bock, C., Müller, A.C., Winter, G.E., Kubicek, S., 2021. Acute BAF perturbation

causes immediate changes in chromatin accessibility. Nat Genet 53, 269–278. https://doi.org/10.1038/s41588-021-00777-3
Schoenfelder, S., Fraser, P., 2019. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet 20, 437–455. https://doi.org/10.1038/s41576-019-0128-0
Shahbazian, M.D., Grunstein, M., 2007. Functions of Site-Specific Histone Acetylation and Deacetylation. Annual Review of Biochemistry 76, 75–100. https://doi.org/10.1146/annurev.biochem.76.052705.162114
Sinha, K.K., Bilokapic, S., Du, Y., Malik, D., Halic, M., 2023. Histone modifications regulate pioneer transcription factor cooperativity. Nature 619, 378–384. https://doi.org/10.1038/s41586-023-06112-6
Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., Mann, R.S., 2011. Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. Cell 147, 1270–1282. https://doi.org/10.1016/j.cell.2011.10.053
Sönmezer, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., Krebs, A.R., 2021. Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. Molecular Cell 81, 255-267.e6. https://doi.org/10.1016/j.molcel.2020.11.015
Spitz, F., Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13, 613–626. https://doi.org/10.1038/nrg3207
Tan, G., Lenhard, B., 2016. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. Bioinformatics 32, 1555–1556. https://doi.org/10.1093/bioinformatics/btw024
The ENCODE (ENCyclopedia Of DNA Elements) Project, 2004. 306.
The ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. https://doi.org/10.1038/nature11247
Trivedi, R., Nagarajaram, H.A., 2022. Intrinsically Disordered Proteins: An Overview. Int J Mol Sci 23, 14050. https://doi.org/10.3390/ijms232214050
Trovato, M., Bunina, D., Yildiz, U., Fernandez-Novel Marx, N., Uckelmann, M., Levina, V., Perez, Y., Janeva, A., Garcia, B.A., Davidovich, C., Zaugg, J.B., Noh, K.-M., 2024. Histone H3.3 lysine 9 and 27 control repressive chromatin at cryptic enhancers and bivalent promoters. Nat Commun 15, 7557. https://doi.org/10.1038/s41467-024-51785-w
Tunnermann, J., Roth, G., Cramard, J., Giorgetti, L., 2025. Supplementary Information: Enhancer control of promoter activity and variability via frequency modulation of clustered transcriptional bursts.
Tycko, J., DelRosso, N., Hess, G.T., Aradhana, Banerjee, A., Mukund, A., Van, M.V., Ego, B.K., Yao, D., Spees, K., Suzuki, P., Marinov, G.K., Kundaje, A., Bassik, M.C., Bintu, L., 2020. High-Throughput Discovery and Characterization of Human Transcriptional Effectors. Cell 183, 2020-2035.e16. https://doi.org/10.1016/j.cell.2020.11.024
Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., Rynes, E., Reynolds, A., Nelson, J., Johnson, A., Frerker, M., Buckley, M., Kaul, R., Meuleman, W., Stamatoyannopoulos, J.A., 2020. Global reference mapping of human transcription factor footprints. Nature 583, 729–736. https://doi.org/10.1038/s41586-020-2528-x
Vo Ngoc, L., Wang, Y.-L., Kassavetis, G.A., Kadonaga, J.T., 2017. The punctilious RNA polymerase II core promoter. Genes Dev 31, 1289–1301. https://doi.org/10.1101/gad.303149.117

Wang, K., Escobar, M., Li, J., Mahata, B., Goell, J., Shah, S., Cluck, M., Hilton, I.B., 2022. Systematic comparison of CRISPR-based transcriptional activators uncovers gene-regulatory features of enhancer-promoter interactions. Nucleic Acids Res 50, 7842–7855. https://doi.org/10.1093/nar/gkac582

Wang, Z.-Q., Zhang, Z.-C., Wu, Y.-Y., Pi, Y.-N., Lou, S.-H., Liu, T.-B., Lou, G., Yang, C., 2023. Bromodomain and extraterminal (BET) proteins: biological functions, diseases and targeted therapy. Sig Transduct Target Ther 8, 420. https://doi.org/10.1038/s41392-023-01647-6

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., Stark, A., 2015. Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. Nature 518, 556–559. https://doi.org/10.1038/nature13994

Zeitlinger, J., 2020. Seven myths of how transcription factors read the cis-regulatory code. Current Opinion in Systems Biology 23, 22–31. https://doi.org/10.1016/j.coisb.2020.08.002

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., Cramer, P., Taipale, J., 2018. The interaction landscape between transcription factors and the nucleosome. Nature 562, 76–81. https://doi.org/10.1038/s41586-018-0549-5

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., Gnirke, A., Meissner, A., 2013. Charting a dynamic DNA methylation landscape of the human genome. Nature 500, 477–481. https://doi.org/10.1038/nature12433

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., Giorgetti, L., 2022. Nonlinear control of transcription through enhancer–promoter interactions. Nature 604, 571–577. https://doi.org/10.1038/s41586-022-04570-y