OVERCOMING THE CHROMATIN BARRIER:

A HEMATOPOIETIC PIONEER TRANSCRIPTION FACTOR PAVES THE WAY FOR SITE-SPECIFIC ATP-DEPENDENT NUCLEOSOME REMODELING

Megan Anna Frederick

A DISSERTATION

in

Immunology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

Kenneth S. Zaret

Joseph Leidy Professor of Cell and Developmental Biology

Graduate Group Chairperson

Taku Kambayashi

Associate Professor of Pathology and Laboratory Medicine

Dissertation Committee

Warren S. Pear, Gaylord P. and Mary Louise Harnwell Professor

Kristen W. Lynch, Benjamin Rush Professor of Biochemistry

Melike Lakadamyali, Associate Professor of Physiology

Ben E. Black, Eldridge Reeves Johnson Foundation Professor of Biochemistry and Biophysics

Dedicated to the memory of my Papa, George Dirschka. I will always be guided by your kindness, humor, and curiosity.

ACKNOWLEDGMENT

Thank you to my thesis advisor, Dr. Ken Zaret. Ken's persistence, optimism, and unwavering enthusiasm has helped me stay motivated and push through the most difficult parts of my project. Ken always prioritizes his trainees' scientific development and takes the time to be involved in every aspect of our training. I have also had so much fun in the lab and will always appreciate the effort Ken put into lab outings and parties (and virtual movie nights during the worst of the pandemic). It has truly been a privilege to be a part of the Zaret Lab and I feel very fortunate to have had such a wonderful environment to conduct my doctoral research.

Thank you to my thesis committee Dr. Warren Pear, Dr. Kristen Lynch, Dr. Melike Lakadamyali, and Dr. Ben Black who provided essential feedback and encouragement throughout my PhD.

I must also thank the greater scientific community at Penn. Especially Penn Biomedical Graduate Studies, the Immunology Graduate Group, the Department of Cell & Developmental Biology, the Epigenetics Institute, and the Institute for Regenerative Medicine. These groups all provided a supportive training environment for students and helped to foster a community amongst the scientists at Penn. Thank you to Dr. Wei Tong for running the Hematopoiesis T32 which supported me and provided excellent additional training opportunities. Thank you to Mary Taylor, Anne Levy, and Thalia Mangan for taking on the demanding job of IGG Coordinator and for always being so supportive and helpful. In addition to our coordinators, thank you to Dr. David Allman, Dr. Paula Oliver, Dr. Igor Brodsky, and Dr. Taku Kambayashi of IGG for running a wonderful graduate group that I am so grateful to be a part of. Thank you to Dr. Golnaz Vahedi for being a wonderful first-year advisor and rotation advisor for my first prelim. Finally, thank you to my IGG cohort. It has been a privilege to share this experience with you all and it is incredible seeing you all flourish in and after your PhD.

I am lucky to have had many wonderful mentors before coming to Penn as well. I would not be here without the instruction and encouragement of my professors at Rollins College. A special thank you to Dr. Susan Walsh and Dr. Eileen Gregory for making a huge difference in my life. Their unwavering enthusiasm and dedication to their students is unparalleled. Thank you to Dr. Matthew Pipkin at the Scripps Research Institute who personally trained me to be a diligent molecular biologist and always encouraged me to think bigger.

I am so fortunate to have trained in an incredible lab environment and I am forever grateful for my wonderful colleagues in the Zaret Lab. Thank you to Dario, Jungsun, Maki, Oscar, Tony, Ryan, Jon, Jingchao, Alyssa, Kenji, Greg, Melinda and Ann for the advice, chats, guidance, and enthusiasm that has helped me so much over the years. Thank you to my fellow graduate students Kate, Kelsey, Meilin, Jessica, Justin, Jailynn, Andy, Edgar, and Claire who have been amazing friends though the ups and downs of grad school. It has been so inspiring for me to learn from Meilin, Kate, Jessica, and Kelsey and to see you all transition to post PhD life. A special shout out to Kelsey who, as my baymate for 5 years, has literally been by my side the entire time and made coming to lab a joy. Thank you to Naomi who I will never be able to thank enough for supporting me (as well as the rest of the lab).

Thank you to my family and friends who have convinced me that I can accomplish anything I put my mind to. Thank you to my loving, supportive, and fun family- Mom, Dad, Grandma, Grandy, Jocelyn, Giovanni, Olivia, George, Diane, Walter, Leo, Diane Jr., Ari, Karen, Austin, Rachel, Tyler, Peg, Aunts, Uncles and Cousins. Thank you to my dear sister-friends, Kiersten and Mallory, who I have grown up with and I am so proud of. Thank you to all the amazing friends I have made throughout this journey who have been an incredible source of joy through the years. Thank you to my darling husband, Max Berman Ferretti, who provides me with endless love and support. Thank you for moving to Philadelphia and for keeping me grounded during graduate school and lockdown. I truly am amazed by how wonderful it is to share my life with someone so supportive, patient, kind, and good humored (and a doctor!). Finally, thank you to Mason the cat, who it seems has unwittingly dedicated his life to stewarding Max and I through graduate school. I guess it’s now our turn to help him get his cat PhD.

ABSTRACT

OVERCOMING THE CHROMATIN BARRIER:

A HEMATOPOIETIC PIONEER TRANSCRIPTION FACTOR PAVES THE WAY FOR SITE-SPECIFIC ATP-DEPENDENT NUCLEOSOME REMODELING

Megan Anna Frederick

Kenneth S. Zaret

Pioneer transcription factors engage nucleosomal DNA in chromatin to initiate gene regulatory events that control cell fate. To determine how different pioneer transcription factors initiate the formation of a locally accessible environment within silent, compacted chromatin and collaborate with an ATP-dependent chromatin remodeler, we generated nucleosome arrays *in vitro* with a central nucleosome that can be targeted by the hematopoietic ETS factor PU.1 and bZIP factors C/EBPα, and C/EBPβ. Each class of factor can expose target nucleosomes on linker histone-compacted arrays, but with different hypersensitivity patterns, as discerned by long-read sequencing. The DNA binding domain of PU.1 is sufficient for mononucleosome binding but requires an additional intrinsically disordered domain to bind and open compacted chromatin. The canonical mammalian SWI/SNF (BAF) complex, cBAF, was unable to act upon two forms of locally open chromatin, in the presence of linker histone, unless cBAF was enabled by the acidic- and glutamine-enriched transactivation domain of PU.1. However, cBAF complexes potentiate the nucleosome binding DBD of PU.1 to weakly open chromatin in the absence of the PU.1 unstructured domain. Together our findings provide a mechanism for how pioneer factors initially target chromatin structures to provide specificity for action by nucleosome remodelers that further open local domains.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION

CHAPTER 2: A PIONEER FACTOR LOCALLY OPENS COMPACTED CHROMATIN TO ENABLE TARGETED ATP-DEPENDENT NUCLEOSOME REMODELING

APPENDIX A: CONSTRUCTION OF NUCLEOSOME ARRAY TEMPLATES CONTAINING NATURAL NUCLEOSOME SEQUENCES TARGETED BY DIVERSE REPROGRAMMING FACTORS

APPENDIX B: PU.1 AND C/EBPA HYBRID TRANSCRIPTION FACTORS

CHAPTER 1: INTRODUCTION

1.1 Pioneer Factors bind nucleosomes and enable cooperative interactions within silent chromatin

Within the human body there are hundreds of different cell types with diverse biological roles, and yet they contain identical genomes. The ability of the cells to have different identities and carry out different functions is largely due to their cell type-specific gene expression pattern. Once cell identity is established, the cell must maintain its identity to ensure its proper function. This is accomplished in part by cell-type specific higher order chromatin structure. The fundamental repeating unit of chromatin is the nucleosome, which consists of ~147 bp of DNA wrapped nearly twice around an octamer of core histone proteins. Linker histones, which are nearly as abundant as core histones, bind to nucleosomes and stabilize condensed, repressive states. Within the genome, long arrays of nucleosomes are then further organized into higher-order structures. Active chromatin is ‘open’, possess active histone modifications, and readily allows access of any transcription factors and transcriptional machinery (Kundaje et al., 2015) (Figure 1A). However, half of the genome is in low signal or ‘naïve’ domains, not enriched for either active or repressive histone modifications (Boettiger et al., 2016; Kundaje et al., 2015), but associated with linker histone which represses transcription (Shimada et al., 2019). In contrast to active and low signal chromatin, actively repressed heterochromatin domains are further compacted by proteins that bind repressive histone modifications, including H3K9 or H3K27 methylation, and strongly repress transcription (Becker et al., 2017; Soufi et al., 2012). This chromatin compaction serves the important role of maintaining and stabilizing cell fate by preventing access of DNA binding proteins thereby preventing aberrant gene expression. However, during cell fate change, including cellular development or differentiation, chromatin structures must be reorganized to establish cell-type specific gene expression.

In the context of chromatin, lineage specific transcription factors recognize specific DNA motifs and direct cell fate changes. Thus, how transcription factors access their target sites within compacted chromatin to initiate new gene expression networks is a fundamental biological question. Indeed, expression of specific transcription factors is indispensable for cellular development and is sufficient to drive cellular reprograming (Takahashi and Yamanaka, 2006) or transdifferentiation (Davis et al., 1987) (Figure 1B). These developmentally indispensable transcription factors are thought to act as “pioneers”: a special subset of transcription factors that scan the genome (Lerner et al., 2020; Sekiya et al., 2009) and directly bind nucleosomal DNA within compacted chromatin (Figure 2A) (Soufi et al., 2015; Meers et al., 2019; Cernilogar et al., 2019) resulting in exposure of the targeted nucleosome (Figure 2B) (Cirillo et al., 2002; Mivelaz et al., 2020) allowing additional gene regulatory factors to bind the chromatin and initiate gene expression or further gene repression (Figure 2C) (Chen et al., 2022; Mivelaz et al., 2020; Sekiya and Zaret, 2007; Tang et al., 2022). Thus, understanding the mechanisms by which pioneer transcription factors can bind and access closed chromatin is critical to understanding the molecular basis by which cellular fate change is initiated.

1.2 Pioneer factor control of hematopoiesis

Hematopoiesis, the specification of blood cells from hematopoietic stem cells, is tightly regulated by expression of lineage-specific transcription factors (Van Oevelen et al., 2015a), and ectopic expression of specific combinations of transcription factors can convert somatic cells to hematopoietic cell lineages (Feng et al., 2008; Pulecio et al., 2014; Riddell et al., 2014; Szabo et al., 2010; Xie et al., 2004a) (Figure 3). Cell fate control is a fundamental process in biology and *de novo* generation of clinically relevant cell types has enormous implications for research and therapeutics. During differentiation or reprogramming, genes that had been hidden in compacted chromatin must be made accessible in a lineage-specific manner (Iwafuchi-Doi and Zaret, 2014a; Zaret and Carroll, 2011; Zaret and Mango, 2016; Zaret, 2014). Concordant with their indispensable role in development and cellular reprogramming, a subset of hematopoietic transcription factors have been suggested to act as pioneer factors (Barozzi et al., 2014a; Boller et al., 2016a; Fuglerud et al., 2017a; Iwafuchi-Doi and Zaret, 2016; Li et al., 2018; Van Oevelen et al., 2015a; Wang et al., 2018; Zaret, 2014).

PU.1, encoded by the *Spi1* gene, is an ETS-family transcription factor with multiple roles in hematopoiesis and positively regulates many genes in the macrophage, granulocyte, dendritic cell, and B-cell lineages. Expressed at high levels in monocytes/macrophages, or moderate levels in B cells, its action is indispensable for generation of all known hematopoietic precursors that have lymphoid developmental potentials (Scott et al., 1994a; Singh et al., 1999) (Figure 3). Thus, PU.1 must localize to different genomic loci to govern different hematopoietic cell-fates. Factors that affect PU.1 binding site selection include motif-binding affinity, PU.1 concentration, and cooperative binding with other transcription factors (Pham et al., 2013). For example, PU.1 collaborates with other lineage-determining factors C/EBPα/β or IRF4/8 to specific macrophage or B-cell development, respectively (Heinz et al., 2013, 2010; Ostuni et al., 2013).

The hematopoietic transcription factors PU.1, C/EBPα, and C/EBPβ have been suggested to be pioneer factors. All three factors are required to drive normal hematopoiesis: PU.1 is required for development of both the myeloid and lymphoid lineages (Decker et al., 2009; Scott et al., 1994b), C/EBPα is required for the development of granulocyte-macrophage progenitors, and C/EBPβ is required for normal macrophage function (Chen et al., 1997; Friedman, 2007a; Heath et al., 2004; Pundhir et al., 2018). PU.1, C/EBPα, and C/EBPβ colocalize genome-wide and co-expression of C/EBPα/β causes PU.1 to localize preferentially to macrophage specific cis-regulatory regions over B-cell specific regions (Heinz et al., 2010; Tagore et al., 2015). Indeed, ectopic expression of C/EBPα in pre-B lymphocytes drives transdifferentiation to the macrophage lineage in a PU.1 and C/EBPβ dependent manner (Bussmann et al., 2009; Di Stefano et al., 2014a; Van Oevelen et al., 2015a). In addition, ectopic expression of PU.1 and C/EBPα/β in fibroblasts, a more distantly related cell type, drives transdifferentiation to macrophage-like cells (Feng et al., 2008) (Figure 3). Furthermore, genomic experiments have shown that binding by the transcription factors PU.1 and C/EBPα precedes deposition of activating histone marks, suggesting that both factors target unprogrammed chromatin prior to recruitment of additional regulatory proteins that further activate the target gene (Ghisletti et al., 2010; Van Oevelen et al., 2015a). PU.1 binding is correlated with nucleosome depletion and promotion of histone modifications (Barozzi et al., 2014b; Heinz et al., 2010). Yet fundamental questions remain regarding the mechanistic basis by which these pioneer factors access closed chromatin.

1.3 Linker histone and silent chromatin

The fundamental unit of chromatin is the nucleosome, consisting of 146 bp of DNA wrapped almost twice around a histone octamer core comprised of an $(H3\text{-}H4)_2$

tetramer and two H2A-H2B dimers. A fifth histone, linker histone H1, binds to nucleosomal core and linker DNA, forming the chromatosome, which alters the linker DNA angle to further compact mononucleosomes and arrays of nucleosomes (Bednar et al., 2017; Fan et al., 2005; Lu et al., 2009). Linker histones are abundant chromatin architectural proteins found in all eukaryotes, with a stoichiometry of ~0.8 total linker histones per nucleosome in most tissues (Woodcock et al., 2006). Linker histones are modularly structured with a short ~35 residue unstructured N-terminal domain, a central globular, winged helix domain, and a ~100 residue unstructured C-terminal domain (Allan et al., 1980). Linker histones bind to chromatin fibers through interaction of the globular domain with nucleosomal sites, and the C-terminal domain with linker DNA (Bednar et al., 2017). Most of the genome is composed of such naïve, silent chromatin that is refractory to chromatin regulators and transcriptional machinery (Adams and Workman, 1995; Bustin et al., 2005; Hill and Imbalzano, 2000; Horn et al., 2002; Mirny, 2010; Ramachandran et al., 2003; Schwarz and Hansen, 1994; Shimada et al., 2019) yet lacks enrichment of apparent histone modifications (Boettiger et al., 2016; Soufi et al., 2012).

Linker histone is a key regulator of chromatin compaction and gene silencing *in vivo*. Transcriptionally inactive regions contain higher linker histone densities and more dense chromatin clusters than transcriptionally active regions of the genome (Braunschweig et al., 2009; Cao et al., 2013; Gilbert et al., 2004; Izzo et al., 2013; Ou et al., 2017; Ricci et al., 2015). Recent studies have found that depletion of linker histone to low levels *in vivo*, leads to global chromatin decompaction, loss of repressive histone modifications, and derepression of specific genes and repetitive elements (Healton et al., 2020; Willcockson et al., 2021; Zhang et al., 2012). Concordant with *in vivo* observations, linker histone is sufficient to physically compact nucleosome arrays and inhibit transcription *in vitro* (Li et al., 2010; Shimada et al., 2019). Thus, understanding mechanisms for disruption of linker histone compacted chromatin is key to understanding the basis of gene regulation.

FoxA and Gata4 are transcription factors required for hepatic specification of the gut endoderm (Gualdi et al., 1996), as well as a reprogramming transcription factors required for the fibroblast to hepatocyte conversion (Huang et al., 2011). *In vivo* footprinting of the early endoderm revealed that FoxA and Gata4 sites are already occupied in the liver-specific *albumin* enhancer, prior to expression of these genes, within silent chromatin (Gualdi et al., 1996; McPherson et al., 1993). Recombinant FoxA1 bound nucleosomes reconstituted *in vitro* with the same albumin enhancer target sequence (Cirillo and Zaret, 1999). Excess FoxA1 competed linker histone off of the nucleosome and enabled Gata4 nucleosome binding (Cirillo et al., 1998). Additionally, FoxA, and to a lesser extent Gata4, could induce DNase I hypersensitivity proximal to their binding sites within H1-compacted 13x nucleosome arrays containing the Alb1 enhancer sequence, indicating that these factors induced local chromatin opening restricted to the targeted nucleosome in an ATP independent manner (Cirillo et al., 2002). Together, these data defined the distinguishing features of FoxA and Gata4 as pioneer factors: they are able to target nucleosomes within silent chromatin to endow competence for additional gene regulatory factors (Sekiya and Zaret, 2007) and initiate new gene expression networks (Iwafuchi et al., 2020; Iwafuchi-Doi et al., 2016a). While diverse pioneer factors are now known to directly bind mononucleosomes, it is not clear whether pioneer factors, other than FoxA1, expose nucleosomes within H1-compacted chromatin and initiate changes to the underlying higher-order chromatin structure.

Few additional studies have investigated pioneer factor binding in the context of the chromatosome or H1-compacted nucleosome arrays, but these handful of studies indicated that transcription factors play important roles in regulating linker histone chromatin. In addition to FoxA, several other factors have been implicated in H1 eviction, including HMG proteins, PARP1, and Oct4, by competing with linker histone for binding to nucleosomes (Cirillo et al., 1998; Echigoya et al., 2020; Krishnakumar and Kraus, 2010; Postnikov and Bustin, 2016). In contrast, binding of the transcription factor Gal4 DNA-binding domain to a 601 chromatosome induced partial unwrapping by causing the H1 C-terminal domain to release the linker DNA, but without evicting H1 from the nucleosome dyad (Burge et al., 2022). Thus, it is possible for transcription factors to disrupt the chromatosome structure and liberate linker DNA without linker histone eviction (Burge et al., 2022). Finally, NFκB can directly bind H1-compacted and recruit the H1 chaperone NAP1 to evict linker histone H1 (Shimada et al., 2019). Together, these studies indicate that there may be diverse ways by which pioneer factors can directly or indirectly disrupt H1-compacted chromatin.

1.4 Pioneer transcription factor domains involved in hierarchical chromatin binding

A fundamental property of a pioneer factor is the ability to bind directly to nucleosomal DNA. As the principal packing elements of DNA, the nucleosome is the primary determinant of DNA accessibility to gene regulatory factors. Within a nucleosome, part of the surface of DNA forms contacts with the globular domains of the core histones, restricting access of transcription factors and other DNA binding proteins (Luger et al., 1997). Thus, the structure of nucleosomal DNA, including curved DNA and adjacent DNA gyres, can be inherently restrictive to transcription factor access and other gene regulatory factors (Luger et al., 1997). Despite these constraints, a wide variety of protein and protein complexes can directly bind and alter the nucleosome structure, including epigenetic modifiers and chromatin remodelers (Han et al., 2020; McGinty et al., 2014; Zhu et al., 2018). Among these nucleosome binding factors, only transcription factors bind nucleosomes with specificity by recognizing unique DNA sequences (Fernandez Garcia et al., 2019; Soufi et al., 2015). Thus, nucleosome binding is a crucial characteristic of pioneer factors that enables gene specific targeting of chromatin and recruitment of chromatin regulators (Gutiérrez et al., 2007; Li et al., 2010).

Studies have shown that transcription factors interact with nucleosomes through a variety of mechanisms. The accessibility of specific transcription factors motifs on the nucleosome depends on their rotational and translational positioning along the DNA. Thus, most transcription factors possess position-specific binding preferences on the nucleosome including periodic (Zhu et al., 2018) and dyad binding to nucleosomal DNA (Dodonova et al., 2020; Li et al., 2019). Other factors prefer to bind within entry/exit sites on the nucleosome, which may be less restrictive to transcription factor binding due to fewer DNA-histone contacts and nucleosome breathing (Michael et al., 2020; Mivelaz et al., 2020; Zhu et al., 2018).

The structure of the DNA-binding domain itself affects the ability of a transcription factors to bind nucleosomes. Pioneer transcription factors tend to have DNA-binding domains that engage their specific motif on the DNA helix without clashing with DNA-core histone interactions (Fernandez Garcia et al., 2019). Additionally, DNA-binding domains containing short anchoring α helices, which facilitate binding to the surfaces of the DNA helix that do not contact core histones, have been shown to interact most strongly with nucleosomes *in vitro.* For example, PU.1 has short α-helix that binds only the DNA's outer surface, while C/EBPα, the weaker nucleosome binder, possesses two α-helices that bind either side of the DNA helix perpendicularly to the core, but extend past the DNA circumference to potentially clash with core histones. However, the DNA-binding domain of C/EBPα, like many other transcription factors, is partly disordered until undergoing a disorder-order transition upon binding its cognate DNA sequence (Miller, 2009; Weiss, 2001). Potentially such factors bind partial motifs on the nucleosome, allowing part of their DNA-binding domain to remain disordered and flexible, and thus reducing any potential clashes with the nucleosome core (Luzete-Monteiro and Zaret, 2022). Indeed, several transcription factors have been reported to bind partial motifs on nucleosomal DNA to adapt to structural constraints (Meers et al., 2019; Michael et al., 2020; Roberts et al., 2021; Soufi et al., 2015). For example Oct4 is able to bind partial motifs when part of its motif is hidden in the nucleosome core likely due to its flexible DNA-binding domain (Li et al., 2019; Michael et al., 2020; Roberts et al., 2021; Soufi et al., 2015).

Classically, transcription factors have been thought of in a simple modular manner. The specificity of transcription factors is attributed to their structured DNA-binding domains, which recognize specific DNA motifs and govern their specific binding, and their transactivation domains, which contribute to gene regulation through additional protein-protein interactions with other co-factors and transcriptional machinery that otherwise lack specific DNA binding (Brent’ and Ptashne, 1985). However, transcription factors with the same or similar DNA-binding domains often differ substantially in their genome localization and function, indicating that non-DNA-binding domains are important for the targeting of specific genomic loci and are vital to their specific biological roles (Brodsky et al., 2020; Chen et al., 2021; Garcia et al., 2021; Kribelbauer et al., 2019). While the DNA-binding domains have been extensively studied, much less is known about the non-DNA binding domains in flanking regions.

While the DNA-binding domains of many pioneer transcription factors have been found to be sufficient for mononucleosome binding (Dodonova et al., 2020; Donovan et al., 2019; Michael et al., 2020; Tanaka et al., 2020; Zhu et al., 2018), several studies have found that the non-DNA binding domains are required for perturbations of higher-order chromatin structure *in vivo* (Boller et al., 2016a; Iwafuchi et al., 2020; Ungerbäck et al., 2018; Wang et al., 2020). There is growing evidence that non-DNA binding domains of pioneer transcription factors are key to their special abilities to bind closed chromatin and initiate chromatin opening, by facilitating interactions with chromatin associated proteins, including core histones (Fuglerud et al., 2017b; Iwafuchi et al., 2020) or nucleosome remodelers (Boller et al., 2016b; Iwafuchi et al., 2020; Minderjahn et al., 2020; Ungerbäck et al., 2018). For example, the pioneer factor FoxA possesses a short α-helical region outside of its DNA-binding domain that interacts directly with core histones and is necessary for chromatin opening *in vitro* and *in vivo*. Thus, it is possible that, in addition to possessing a DNA-binding domain capable of binding nucleosomal DNA, pioneer factors also possess non-DNA binding domains that further enable cooperative events in chromatin. Identifying these critical, but thus far unheralded, domains is a key goal of this work.

Intrinsically disordered regions are abundant in core histones, linker histone, histone chaperones, transcription factors, chromatin remodeling proteins and architectural proteins. Further, these regions have now been shown to play a key role in the ability of these factors to interact with and regulate chromatin states (Boija et al., 2018; Burge et al., 2022; Canzio et al., 2011; Desai et al., 2015; Heidarsson et al., 2022; Kan et al., 2007; Liu et al., 2020; Munari et al., 2012; Musselman and Kutateladze, 2021). Most transcription factors also possess long intrinsically disordered regions which contain linkers and transactivation domains. The intrinsically disordered regions of non-DNA-binding domains are capable of forming lock-and-key type interactions when folding upon substrate binding (Habchi et al., 2014; Minezaki et al., 2006). Intrinsically disordered regions can additionally interact by more promiscuous, long-range electrostatic attractions (Borgia et al., 2018; Protter et al., 2018), and there is growing evidence that such weak, “fuzzy” interactions are important for a wide spectrum of affinities and specificities (Arbesú et al., 2018; Brzovic et al., 2011; Danilenko et al., 2019; Gao et al., 2018). For instance, transcription factors EBF1, a B-cell pioneer factor, and EWS-FLI1, the oncogenic fusion protein, recruit the mammalian SWI/SNF ATP-dependent chromatin remodeler, BAF, via a prion-like intrinsically disordered region (Boulay et al., 2017; Wang et al., 2020). In this work I find that the disordered regions of pioneer factors are critical for accessing H1 compacted chromatin regions, which allows access by remodeling complexes to further open the targeted chromatin region.

1.5 Pioneer factors collaborate with chromatin remodeling complexes

While *in vitro* studies have indicated that pioneer transcription factors are sufficient to bind and locally open nucleosome and nucleosome arrays, *in vivo* time course studies have found differing results. Shortly after induction, C/EBPα; EBF1, a B cell transcription factor; or Pax7, a pituitary transcription factor; bind to closed chromatin several hours prior to chromatin opening, as assessed by the ATAC-seq assay (Li et al., 2018; Mayran et al., 2018; Sardina et al., 2018). At later timepoints, additional factors are recruited and the sites become accessible by ATAC-seq (Mayran et al., 2018; Sardina et al., 2018). These findings imply that, in vivo, pioneer factors alone cannot fully open closed chromatin. However, ATAC-seq methods may only sensitive to chromatin opening that occurs as a result of additional factors acting to establish a new regulatory region after a pioneer factor has bound a specific genomic locus, as opposed to the local chromatin opening observed *in vitro* which is restricted to a single targeted nucleosome (Cirillo et al., 2002). This could be due to key differences between the assays, where ATAC-seq requires Tn5 to integrate at 2 nearby sites (Buenrostro et al., 2013), while *in vitro* DNase hypersensitivity assays detect enzymatic cleavage at a single accessible site (Cirillo et al., 2002). Thus, these time course studies highlight that there are steps between pioneer factor binding and overt chromatin opening in vivo, which requires the recruitment of secondary factors.

Nucleosome remodeling complexes, which use ATP to slide, disassemble or eject nucleosomes, play a key role in establishment and maintenance of accessible chromatin (Iurlaro et al., 2021; Schick et al., 2021). The four major families of nucleosome remodeler complexes in mammals are SWI/SNF (switch/sucrose non-fermentable), ISWI (imitation switch), CHD (chromodomain helicase DNA binding) and INO80 (Clapier et al., 2017). The assembly remodelers, ISWI and CHD, are primarily involved with the nucleosome assembly and INO80 is primarily 'edits' nucleosomes by removing and replacing conical or variant histone (Clapier et al., 2017). The SWI/SNF family, however, primarily regulates chromatin access by sliding nucleosomes along DNA and evicting nucleosome core components. Thus, while assembly remodelers promote gene silencing through the formation of tightly packed nucleosome arrays, the SWI/SNF family can promote gene activation by exposing sites for transcription machinery. Indeed, inhibition of the mammalian SWI/SNF ATP-dependent nucleosome remodeler, BAF, results in rapid loss of enhancer chromatin accessibility in embryonic stem cells, which is re-established upon reinitiation of BAF activity (Iurlaro et al., 2021), indicating that SWI/SNF nucleosome remodelers continuously maintain enhancer accessibility. However, left unanswered by these studies was the mechanism by which SWI/SNF nucleosome remodelers, which lack specific DNA-binding motifs, are initially recruited to particular sites in the genome to establish chromatin accessibility. For instance, do SWI/SNF complexes rely on factors that remain associated with chromatin, like transcription factors or histone modifications, for recruitment?

In vitro studies have indicated that transcription binding modulates nucleosome remodeler activity. An early *in vitro* study found that the purified Drosophila pioneer factor GAGA (GAF) could disrupt the structure of chromatin reconstituted with Drosophila embryo extracts in an ATP-dependent manner that was attenuated by the addition of linker histone, indicating that GAF acted in conjunction with ATP-dependent nucleosome remodelers present in the crude extract to initiate chromatin opening. Recently, stable chromatin binding of GAF was found to be independent of the presence of nucleosome remodelers *in vivo* (Tang et al., 2022). Critically, *in vitro* studies utilizing chromatin assemblies with purified DNA and histones have now identified a wide array of purified recombinant transcription factors that can directly bind mononucleosomes (Cirillo et al., 2002; Li et al., 2010) and nucleosome arrays in the absence of nucleosome remodelers. A recent *in vitro* single-molecule total internal reflection fluorescence microscopy study showed that the yeast pioneer factor Rap1 is able to target its specific binding site on a central nucleosome within a salt compacted 13x nucleosome array. To access this nucleosome, Rap1 makes the targeted nucleosome accessible by interrupting nucleosome stacking (Mivelaz et al., 2020). Mononucleosome remodeling assays indicated that the directionality of nucleosome remodeling by RSC, the yeast homologue to the mammalian SWI/SNF complex, was biased in the presence of Rap1 to move the nucleosome core away from a Rap1 binding site, enabling Rap1 to stably bind free DNA (Mivelaz et al., 2020).

The synthetic transcription factor Gal4-VP16 enabled SWI/SNF mediated histone eviction within a nucleosome array reconstituted *in vitro* in the absence of linker histone (Gutiérrez et al., 2007). In this assay, the transactivation domain of Gal4-VP16 was required to enable targeted eviction of a nucleosome adjacent to the GAL4 binding site. The observations lead the authors to postulate that the ability of a transcription factor to specifically target SWI/SNF to a nucleosome was a prerequisite for remodeling and that the accessibility of the transcription factor binding site may be the first determinant in the sequence of events leading to nucleosome remodeling (Gutiérrez et al., 2007). Similarly, the mammalian retinoic acid receptor, RAR/RXR, can recruit SWI/SNF to nucleosome arrays, which lack linker histone but possess native acetylated core histones, and remodel nucleosome array structure, subsequently allowing an additional transcription factor, NF1, to bind its cognate site (Li et al., 2010). However, GAL4-V16 binding was insufficient to enable SWI/SNF to remodel nucleosomes within linker histone compacted chromatin (Shimada et al., 2019). In this case SWI/SNF could only access and remodel the underlying chromatin in the presence of histone chaperone NAP1, which facilitates the removal of linker histone. Taken together, *in vitro* studies have determined that transcription factors enable site specific nucleosome remodeling, yet it is unclear how ATP-dependent nucleosome remodeling is initiated within closed, H1-compacted chromatin.

In vivo studies have detailed reciprocity between pioneer factors and chromatin remodelers, where rapid depletion of Oct4 leads to loss of chromatin accessibility and BAF complex binding at some sites (Friman et al., 2019; King and Klose, 2017), and rapid depletion of BAF complexes leads to loss of chromatin accessibility and reduction of transcription factor binding at other sites (Boulay et al., 2017; Sandoval et al., 2018; Iurlaro et al., 2021; Xiao et al., 2022). Such studies indicate that maintenance of an open chromatin environment is a dynamic process that requires the sustained action of regulatory proteins. Indeed, continuous binding of pioneer factors to nucleosomes within regulatory regions can be required to maintain an open chromatin environment (Iwafuchi-Doi et al., 2016a; Roberts et al., 2021). While pioneer factor chromatin binding and BAF complex recruitment are both an integral part of tissue-specific gene expression, little is known about the individual steps in the process of chromatin opening and the mechanism by which pioneer factors may enable BAF complex activity, or vice-versa.

1.6 Open questions of this thesis

During cell fate changes, silent genes in compacted chromatin must be accessed to establish new genetic networks. Pioneer transcription factors and ATP-dependent SWI/SNF nucleosome remodeling complexes play key roles in initiating and maintaining accessible chromatin states at gene regulatory regions. Despite significant interest in understanding the interplay between pioneer transcription factors and chromatin remodeling complexes, *in vivo* genomic studies lack the resolution to clearly define how each of these factors individually contribute to initial chromatin opening. Furthermore, the fundamental mechanism by which nucleosome remodeling enzymes initially access compacted chromatin in a site-specific manner has yet to be answered.

There is evidence that C/EBPα, C/EBPβ, and PU.1 interact physically with SWI/SNF chromatin remodelers (Kowenz-Leutz and Leutz, 1999; Pedersen et al., 2001; Ungerbäck et al., 2018), and induction of PU.1 expression has been correlated with recruitment of SWI/SNF complexes to macrophage specific cis-regulatory regions (McAndrew et al., 2016). Further, truncation or deletion of the non-DNA binding domains of C/EBPα and PU.1 results in the loss of binding to closed chromatin sites (Jakobsen, 2019; Ungerbäck et al., 2018). In the case of PU.1, the loss of closed chromatin binding is correlated with loss of chromatin accessibility and loss of SWI/SNF binding in vivo (Minderjahn et al., 2020; Ungerbäck et al., 2018). In the current *in vitro* study, I set out to determine if these factors possessed the intrinsic ability to bind and locally open linker histone compacted chromatin and, if so, to determine the molecular features of these factors that enable such activity, including the ability to enable a SWI/SNF ATP-dependent chromatin remodeler.

1.7 Figures

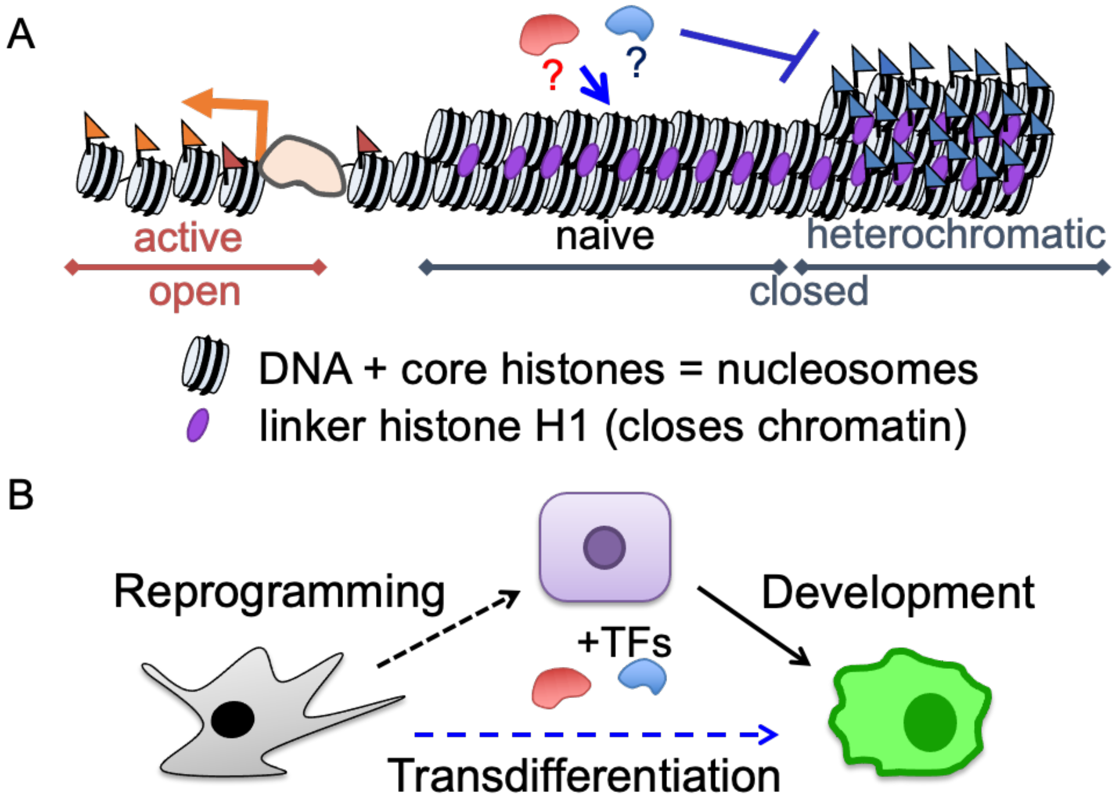


Figure 1 - 1 Chromatin compaction maintains cell fate by preventing access to DNA binding proteins.

(A) Cell type specific gene expression is governed in part by cell-type specific chromatin structure, where long arrays of nucleosomes, consisting of DNA wrapped nearly twice around an octamer of core histones, are organized into higher-order structures. Some regions of the genome possess loosely packed nucleosomes and activating histone modifications (red flags) and transcription factors can readily bind. Other regions have tightly packed nucleosomes, associated with linker histone which closes chromatin, and are refractory to DNA binding factors. Heterochromatic domains possess repressive histone modifications (blue flags) and repressed by association of additional heterchromatin proteins. (B) Transcription factors are necessary for cellular development and sufficient for cellular reprogramming or transdifferentiation.

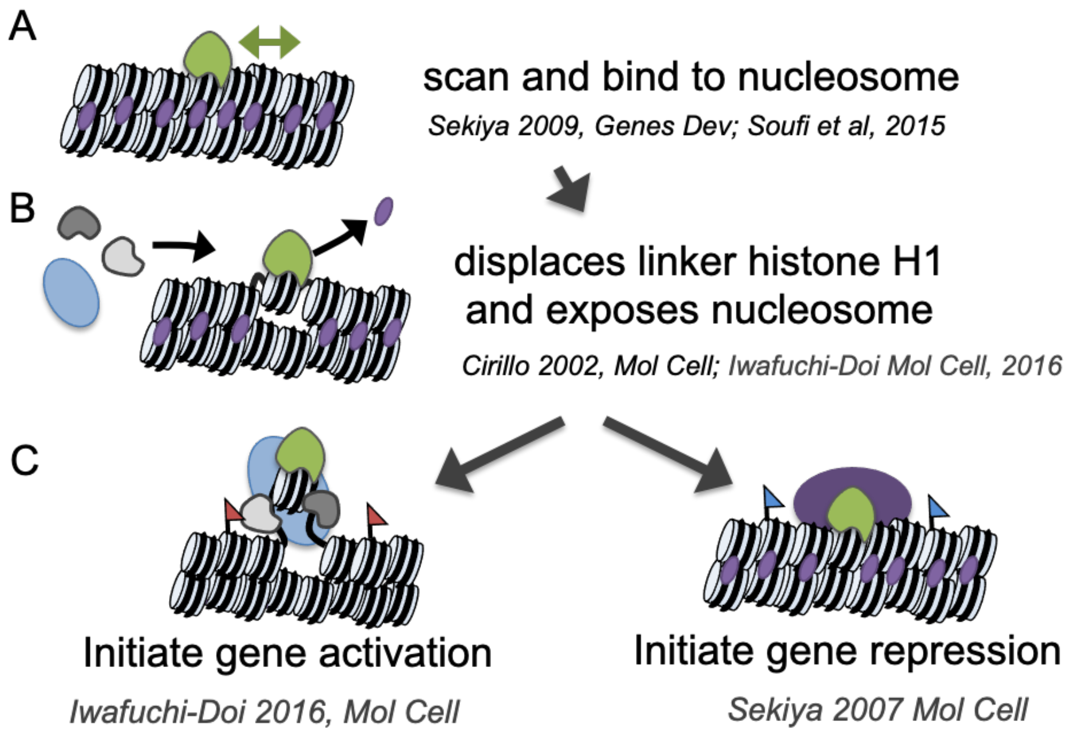


Figure 1 - 2 Initial targeting of pioneer factor and subsequent events.

(A) Pioneer transcription factors can scan compacted chromatin and bind their target sites on a nucleosome. (B) Pioneer factors binding exposes the targeted nucleosome by evicting linker histone, and (C) enables other gene regulatory factors to access the chromatin, including transcription factors, chromatin modifiers, and remodelers, initiating further gene regulatory events leading to gene activation or further gene repression. Other transcription factor binding and chromatin modifications could stabilize pioneer factor binding to the target sites.

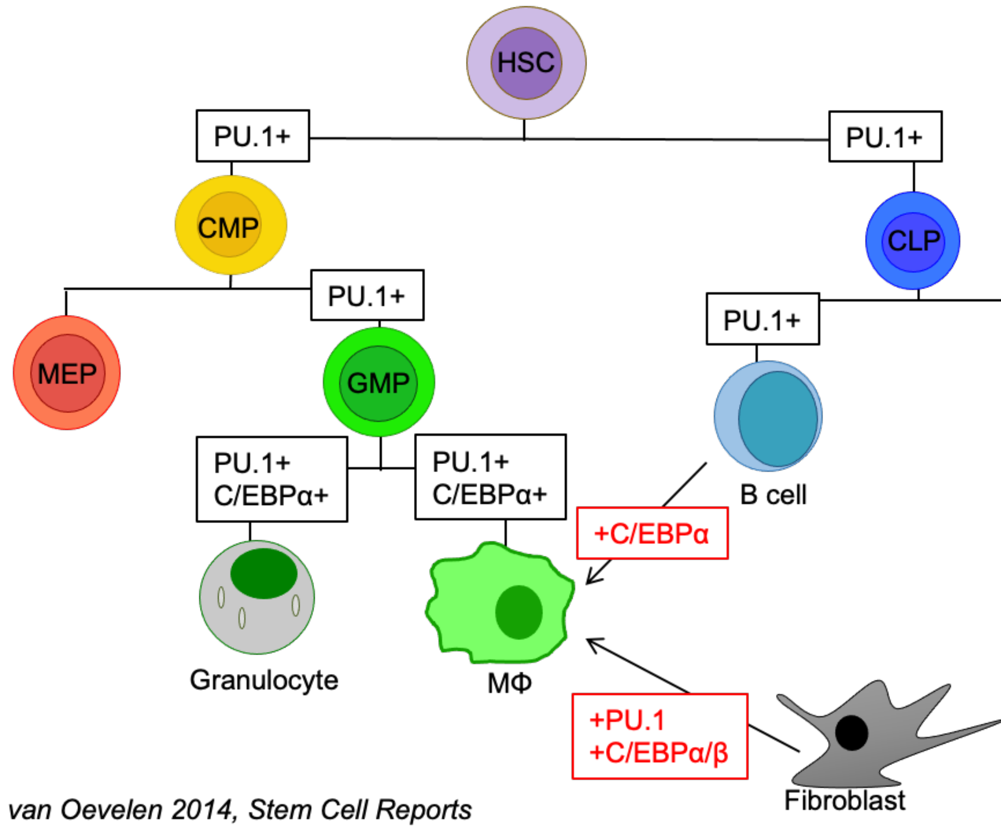


Figure 1 - 3 Hematopoietic TFs PU.1, C/EBPα, and C/EBPβ drive blood cell specification.

All three factors are required to drive normal hematopoiesis: PU.1 is required for development of both the myeloid and lymphoid lineages (Decker et al., 2009; Scott et al., 1994b), C/EBPα/β are required for the development of granulocyte-macrophage progenitors (Chen et al., 1997; Friedman, 2007a; Heath et al., 2004; Pundhir et al., 2018). Ectopic expression of C/EBPα in pre-B lymphocytes drives transdifferentiation to the macrophage lineage in a PU.1 and C/EBPβ dependent manner (Bussmann et al., 2009; Di Stefano et al., 2014a; Van Oevelen et al., 2015a). In addition, ectopic expression of PU.1 and C/EBPα/β in fibroblasts, a more distantly related cell type, drives transdifferentiation to macrophage-like cells (Feng et al., 2008). Simplified cartoon of blood cell specification adapted from Van Oevelen et al. (Van Oevelen et al., 2015a).

CHAPTER 2: A PIONEER FACTOR LOCALLY OPENS COMPACTED CHROMATIN TO ENABLE TARGETED ATP-DEPENDENT NUCLEOSOME REMODELING

2.1 Preface

The manuscript presented in this chapter was officially accepted for publication in *Nature Structural and Molecular Biology* November 3, 2022. It has been reformatted here in accordance with University of Pennsylvania dissertation formatting guidelines. Nanopore sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE216065.

Authors: Megan A. Frederick[1-4], Kaylyn E. Williamson[5], Meilin Fernandez Garcia[1-3,6], Max B. Ferretti[6], Ryan L. McCarthy[1-3], Greg Donahue[1-3], Edgar Luzete Monteiro[1-3,7], Naomi Takenaka[1-3], Janice Reynaga[1-3], Cigall Kadoch[5], Kenneth S. Zaret[1-3]

Affiliations:

[1]Institute for Regenerative Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[2]Penn Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[3]Dept. Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[4]Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[5]Dept. of Pediatric Oncology, Dana–Farber Cancer Institute and Harvard Medical School, Boston, MA, USA

[6]Dept. of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[7]Dept. of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA.

2.2 Respective Contributions

The majority of the experiments and analyses presented in this chapter were designed and performed by myself with the guidance of my thesis advisor Kenneth S. Zaret. I performed protein expression and purification, design and generation of all end-labeled nucleosome array templates, nucleosome array reconstitution, transcription factor binding reactions, nuclease hypersensitivity assays, nucleosome remodeling assays, and nanopore sequencing sample preparation, sequencing, and analysis. Kaylyn E. Williamson purified the human cBAF complex. Meilin Fernandez Garcia and Janice Reynaga performed the EMSAs of the PU.1 and C/EBPα. Wild-type and DNA-binding domains binding to Cx3cr1 mononucleosomes. Meilin Fernandez Garcia reconstituted the Cx3cr1 mononucelosomes. Edgar Luzete Monteiro purified the PU.1 and C/EBPα DNA-binding domains under the supervision of Meilin Fernandez Garcia. I carried out the Nanopore sequencing end-point analysis with support and guidance from Max B. Ferretti, Greg Donahue and Ryan L. McCarthy. Max B. Ferretti wrote the scripts to normalize the nanopore sequencing data and performed the computational analysis of the translational positions of nucleosomes within the Cx3cr1 nucleosome array. Naomi Takenaka generated the purified histone octamers.

2.3 Abstract

To determine how different pioneer transcription factors form a targeted, accessible nucleosome within compacted chromatin and collaborate with an ATP-dependent chromatin remodeler, we generated nucleosome arrays *in vitro* with a central nucleosome containing binding sites for the hematopoietic ETS factor PU.1 and bZIP factors C/EBPα, and C/EBPβ. Our long-read sequencing reveals that each factor can expose a targeted nucleosome on linker histone-compacted arrays, but with different nuclease sensitivity patterns. The DNA binding domain (DBD) of PU.1 binds mononucleosomes but requires an additional intrinsically disordered domain to bind and open compacted chromatin. The canonical mammalian SWI/SNF (cBAF) remodeler was unable to act upon two forms of locally open chromatin unless cBAF was enabled by a separate transactivation domain of PU.1. cBAF potentiates the PU.1 DBD to weakly open chromatin in the absence of the PU.1 disordered domain. Our findings reveal a hierarchy by which chromatin is opened and show that pioneer factors can provide specificity for action by nucleosome remodelers.

2.4 Introduction

Chromatin compaction by linker histone maintains and stabilizes cell fate by preventing access of chromatin regulators and the transcriptional machinery (Adams and Workman, 1995; Bustin et al., 2005; Hill and Imbalzano, 2000; Horn et al., 2002; Li et al., 2010; Mirny, 2010; Ramachandran et al., 2003; Schwarz and Hansen, 1994; Shimada et al., 2019). During cell fate changes, new genetic networks are initiated by regulatory proteins that access DNA control regions embedded in linker histone-bound, compacted, and silent chromatin. Key among such proteins are pioneer transcription factors, which can bind nucleosomal DNA and locally engage other proteins to enable chromatin opening or further compact the chromatin (Adams and Workman, 1995; Iwafuchi-Doi and Zaret, 2014b; Mirny, 2010; Sekiya and Zaret, 2007). Low-signal chromatin harboring linker histone, transcriptionally silent and lacking enrichment for known histone modifications, is most frequently targeted by pioneer transcription factors during cellular reprogramming (Soufi et al., 2012). Thus, in vitro studies, which have identified diverse pioneer transcription factors that directly bind mononucleosomes, have utilized histones lacking modifications (Fernandez Garcia et al., 2019; Zhu et al., 2018). However, less is understood about how diverse pioneer factors engage and initiate the opening of compacted chromatin, and how they work with nucleosome remodeling complexes.

Time course studies of chromatin opening *in vivo* have reported that, shortly after being ectopically induced, various pioneer transcription factors bind closed chromatin rapidly, but require the recruitment of additional co-factors to induce wider changes to the underlying chromatin structure (Cernilogar et al., 2019; Mayran et al., 2019). A foundational *in vitro* study found that the purified Drosophila pioneer factor GAGA could disrupt the structure of chromatin reconstituted with Drosophila embryo extracts in an ATP-dependent manner that was attenuated by the addition of linker histone, indicating that GAGA acted in conjunction with ATP-dependent chromatin remodelers to initiate chromatin opening (Tsukiyama et al., 1994). The initial, stable chromatin binding of GAF was later found to be independent of the presence of nucleosome remodelers *in vivo* (Tang et al., 2022). *In vitro* studies utilizing nucleosome arrays assembled with purified DNA and histones identified recombinant transcription factors that can bind H1-compacted arrays (Cirillo et al., 2002; Li et al., 2010). DNase I hypersensitivity and footprinting assays determined that the liver specific pioneer transcription factor FOXA1 not only binds H1-compacted nucleosome arrays derived from natural nucleosome sequences of the *Alb* gene, but also exposes an underlying nucleosome in a target sequence-dependent manner (Cirillo et al., 2002; Iwafuchi et al., 2020). By contrast, linker histone H1 impedes the action of ATP-dependent, SWI/SNF nucleosome remodelers (Hill and Imbalzano, 2000; Horn et al., 2002; Ramachandran et al., 2003). However, it is not clear whether pioneer factors other than FOXA1 are sufficient to access their sites in H1-compacted chromatin and expose a targeted nucleosome like FOXA1. Furthermore, it is unclear if locally opening the chromatin is sufficient to enable the action of a nucleosome remodeler, or whether direct recruitment by the pioneer factor is necessary.

In vivo studies have detailed reciprocity between pioneer factors and chromatin remodelers, where rapid depletion of OCT4 leads to loss of chromatin accessibility and BAF complex binding at some sites (Friman et al., 2019; King and Klose, 2017), and rapid depletion of BAF complexes leads to loss of chromatin accessibility and reduction of transcription factor binding at other sites (Boulay et al., 2017; Sandoval et al., 2018; Iurlaro et al., 2021; Xiao et al., 2022). The pioneer transcription factor PU.1 (SPI1) (Fernandez Garcia et al., 2019; Heinz et al., 2010) physically interacts with components of the SWI/SNF chromatin remodeling complex (Hosokawa et al., 2018), and truncation or deletion of the non-DNA binding domains of PU.1 results in the loss of binding to closed chromatin sites, which is correlated with loss of chromatin accessibility and loss of SWI/SNF binding *in vivo* (Minderjahn et al., 2020; Ungerbäck et al., 2018). Such studies indicate that maintenance of an open chromatin environment is a dynamic process that requires the sustained action of regulatory proteins. Indeed, continuous binding of pioneer factors to nucleosomes within regulatory regions can be required to maintain an open chromatin environment (Iwafuchi-Doi et al., 2016a; Roberts et al., 2021). While pioneer factor chromatin binding and BAF complex recruitment are both an integral part of tissue-specific gene expression, little is known about the individual steps in the process of chromatin opening and the mechanism by which pioneer factors may enable BAF complex activity, or vice-versa. Given redundancies and complexities *in vivo*, we took an *in vitro* approach with purified components previously studied *in vivo* or in extracts to reveal the interplay between a pioneer factor and a chromatin remodeler.

2.5 Results

Nucleotide resolution of accessibility in nucleosome arrays

PU.1 is a master regulator required for hematopoietic stem cell renewal and collaborates with other factors to give rise to myeloid and lymphoid lineages. Macrophage development from hematopoietic stem cells requires the coordinated expression of PU.1 with the myeloid transcription factors C/EBPα (CEBPA) and C/EBPβ (CEBPB) (Heinz et al., 2010), and ectopic expression of PU.1 and C/EBPα/β converts fibroblasts to the macrophage lineage (Feng et al., 2008). Previously, we identified a nucleosome targeted by PU.1, C/EBPα, and C/EBPβ *in vivo* (Extended Data Fig. 1a) and reconstituted this “Cx3cr1” mononucleosome *in vitro* (Fernandez Garcia et al., 2019). We found that all three transcription factors can directly bind Cx3cr1 mononucleosomes, with varying affinities in the nanomolar range (Fernandez Garcia et al., 2019). Here, we inserted the *Cx3cr1* mononucleosome sequence centrally within a 13X nucleosome array (Extended Data Fig. 1b) and 5’ end-labeled the DNA template with the fluorescent nucleotide, dCTP-Cy5 (Fernandez Garcia et al., 2019). "Cx3cr1 nucleosome arrays" were assembled using human recombinant purified histones and the Cy5 DNA templates, and validated as previously described (Cirillo et al., 2002) (Extended Data Figs. 1 and 2).

To map nucleosome positions along the 2733 bp Cx3cr1 nucleosome array, we performed partial MNase digests of free DNA and extended Cx3cr1 nucleosome arrays, purified the DNA, performed long-read Nanopore sequencing, and then mapped the endpoints of each aligned fragment (Extended Data Fig. 2a-c). We subtracted the free DNA MNase cleavage signals from the nucleosome array MNase signals, to correct for MNase DNA sequence preferences (Extended Data Fig. 2c) (Drew and Travers, 1985). The resulting plot yielded patterns of protection, indicating the positions of phased nucleosomes, and MNase hypersensitivity, indicating the positions of linker DNA (Extended Data Fig. 2d). To mimic a naïve, silent chromatin state (Fyodorov et al., 2018), we further compacted nucleosome arrays with linker histone (H1). As previously reported (Cirillo et al., 2002; Shimada et al., 2019), a 1:1 molar ratio of H1 to nucleosomes was sufficient to quantitatively compact the Cx3cr1 nucleosome arrays, as seen by a marked overall resistance to DNase I digestion and increased mobility on a native agarose, gel relative to the extended nucleosome arrays (Extended Data Fig. 1c and d).

PU.1, C/EBP elicit different access to compacted chromatin

Incubation of PU.1 and C/EBPs with the H1-compacted nucleosome arrays each resulted in induced DNase cleavages within the Cx3cr1 nucleosome (Fig. 1a-d, Extended Data Figs. 3a-c). However, the factors' patterns of chromatin opening differed markedly. PU.1 generated a strong DNase hypersensitive signal underlying its binding site within the Cx3cr1 nucleosome core and into 3’ flanking linker DNA. By contrast, C/EBPα and C/EBPβ each elicited hypersensitivity, but less intensely than PU.1, within the entry and exit regions of DNA on the nucleosome as well as in the 5’ linker region

(Fig. 1b-d). IRF3, which specifically binds the Cx3cr1 free DNA, but fails to bind Cx3cr1 mononucleosomes (Fernandez Garcia et al., 2019), did not change accessibility of the H1-compacted array (Extended Data Fig. 3d). We found that incubating PU.1 together with C/EBPα led to a synergistic, rather than additive, increase in the DNase I hypersensitive signal (Extended Data Fig. 4a-c), indicating that two pioneer factors can cooperate to open chromatin. Importantly, the observed transcription factor-induced hypersensitivity required the specific transcription factor binding motifs within the central Cx3cr1 nucleosome (Fig. 1e). We conclude that the different pioneer factors elicit different opening patterns on H1-compacted chromatin, in a target site-dependent manner.

To address whether the relative positions of binding sites on the Cx3cr1 nucleosome determined the differences in DNase I cleavage patterns, we swapped the two PU.1 and one C/EBP binding motifs in two ways (Fig. 1f). Swapping motif positions had little effect on the patterns of nucleosome exposure (Fig. 1f). Thus, the transcription factor itself, rather than its precise binding position on the nucleosome, dictates a particular chromatin opening pattern. We note that a prior study using the 601 strong nucleosome positioning sequence only observed PU.1 binding at the entry and exit points of the nucleosome (Minderjahn et al., 2020). Such findings conflict with those observed with nucleosomes made with natural DNA (Fernandez Garcia et al., 2019) and binding-selected sequences (Zhu et al., 2018) that allow factors to bind at multiple locations on the nucleosome.

We found that the DNA-binding domains of PU.1 and C/EBPα were sufficient to bind a Cx3cr1 mononucleosome (Fig. 2a, b, Extended Data Fig. 5a-d), yet were insufficient to efficiently elicit DNase I hypersensitivity on compacted nucleosome arrays (Fig. 2c, d). These *in vitro* data are concordant with previous reports that nucleosome binding is encoded within the DNA binding domains of pioneer factors (Cirillo et al., 1998; Dodonova et al., 2020; Michael et al., 2020; Roberts et al., 2021; Zhu et al., 2018), but non-DNA binding domains are required for perturbations of higher-order chromatin structure *in vivo* (Iwafuchi et al., 2020; Ungerbäck et al., 2018; Wang et al., 2020).

PU.1 enables further chromatin opening by cBAF nucleosome remodeler

We determined that purified cBAF (Mashtalir et al., 2020, 2018) (Extended Data Fig. 6a) plus ATP could move nucleosomes to expose a restriction site on extended Cx3cr1 nucleosome arrays, without H1, proximal to the Cx3cr1 nucleosome (Extended Data Fig. 6b, c). Consistent with previous studies (Hill and Imbalzano, 2000; Horn et al., 2002; Ramachandran et al., 2003), H1-compacted arrays were refractory to nucleosome remodeling by cBAF in the presence of ATP (Fig. 3a, no change between lanes 2-3 vs 6-7). PU.1 incubated with cBAF complex, without the addition of ATP, showed no difference in eliciting hypersensitivity compared to PU.1 alone (Fig. 3a, no change between lanes 8-9 vs 10-11). However, when PU.1 was incubated with cBAF in the presence of ATP, we observed a 3' expansion of hypersensitivity into the neighboring 3' nucleosome (Fig, 3a, lanes 12-13, denoted by a red arrow), compared to PU.1 alone (lanes 8-9), as detailed by Nanopore sequencing (Fig. 3b and Extended Fig. 6d and e). Thus, PU.1 enabled cBAF remodeler action within the compacted chromatin in an ATP-dependent manner.

There were no differences in hypersensitivity patterns when the cBAF complex was added to C/EBPα and H1-compacted nucleosome arrays (Fig. 3a, lanes 14-19). Notably, the ability of C/EBPα to open the chromatin, but not enable BAF activity, indicated that the local chromatin opening pattern of C/EBP factors, per se, was insufficient to enable cBAF activity.

The intrinsically disordered region of PU.1 opens chromatin

We performed a deletion analysis of PU.1 to determine the molecular features that enable initial nucleosome exposure and subsequent cBAF remodeling (Fig. 4a,b). Between the N-terminal acidic (AQ) domain and the DNA-binding domain of PU.1 is a highly anionic domain with similarity to a “PEST” domain. However, the PEST domain does not target PU.1 for metabolic turnover (Perkel and Atchison, 1998), though there is evidence that it controls the dynamics of PU.1 homodimerization (Xhani et al., 2020). Using two different methods to map protein structure (Dosztányi et al., 2005a; Piovesan et al., 2021) (Fig. 4a), we found that the PEST domain is predicted to be highly disordered and thus we refer to it as an intrinsically disordered region (IDR).

We interrogated DNA and chromatin binding by the PU.1 deletion mutants in a series of electromobility shift assays (Extended Data Fig. 7). The **Δ**AQ and **Δ**IDR mutants of PU.1 bound Cx3cr1 free DNA and mononucleosomes as well as WT PU.1 (Extended Data Fig. 7a-c). The **Δ**AQ truncation mutant bound extended arrays and H1-compacted arrays as well as WT PU.1, in a target site-dependent manner (Extended Data Fig. 7e, g. Compare lanes 2-3 to lanes 14-15, respectively). However, the **Δ**IDR mutant bound extended arrays less efficiently (Extended Data Fig. 7e, lanes 8-9) and failed to shift H1-compacted arrays (Extended Data Fig. 7g, lanes 8-9). Thus, the IDR domain of PU.1 is not needed for nucleosome binding, but facilitates PU.1 engagement with compacted chromatin.

The N-terminal ΔAQ truncation mutant, with the IDR of PU.1 tethered to the PU.1 DNA-binding domain, was sufficient to robustly expose the Cx3cr1 nucleosome within H1-compacted arrays (Extended Data Fig. 8a, b, lanes 11-12). By contrast, the ΔIDR mutant did not elicit chromatin opening (Extended Data Fig. 8a, b lanes 7-8; Extended Data Fig. 8c, lanes 14, 15), comparable to the low range observed with the DBD alone (Extended Data Fig. 8a, b lanes 5-6; Extended Data Fig. 8c, lanes 10-11). We conclude that the IDR domain is essential for PU.1 to interact with H1-compacted chromatin and for the intrinsic chromatin opening function of PU.1.

BAF does not act at open chromatin sites without recruitment

Strikingly, robust chromatin opening by the N-terminal ΔAQ truncation mutant (Fig. 4c, lanes 14-15) was insufficient to enable cBAF activity (Fig. 4c compare lanes 14-15 to lanes 16-17, respectively). Thus, the acidic domain interacts with (Minderjahn et al., 2020) or positions a unique orientation of cBAF complexes required for them to expand hypersensitivity within H1-compacted chromatin.

Interestingly, we found that incubation of cBAF and ATP with the ΔIDR deletion mutant, which can bind mononucleosomes (Extended Data Fig. 7b) yet is deficient for nucleosome exposure in compacted chromatin, resulted in a weak hypersensitive signal within the Cx3cr1 nucleosome of the compacted nucleosome array. Thus, residual nucleosome binding by the ΔIDR deletion mutant, which retains the N-terminal AQ region and the DBD, enabled cBAF to elicit chromatin opening weakly (Fig. 4c lanes 10-13). The DNA-binding domain alone, which like the ΔIDR deletion mutant can bind mononucleosomes (Extended Data Fig. 5b) and is deficient for nucleosome exposure in compacted chromatin (Extended Data Fig. 8c), did not enable cBAF activity (Extended

Data Fig. 8c, compare lanes 10-11 to lanes 12-13). Together, these data indicate that the N-terminal AQ region is necessary for PU.1 to enable cBAF to engage and remodel H1-compacted chromatin in vitro.

2.6 Discussion

We find that to act on linker histone-compacted chromatin, cBAF chromatin remodeling complexes require initial chromatin opening by PU.1, mediated by PU.1's IDR and acidic domains. The lack of cBAF complex activity on the chromatin opened by C/EBPα and the PU.1 **Δ**AQ truncation mutant, each of which elicit different locally open chromatin patterns, demonstrate that mere local opening is not sufficient for cBAF activity and that cBAF complexes must also be stabilized by, positioned, or recruited to the compacted chromatin by the pioneer factor. Presumably, the necessity for the pioneer factor provides a specificity check for ATP-dependent remodeling at particular sites in chromatin.

Furthermore, our finding that cBAF can weakly promote chromatin opening in cooperation with a PU.1 deletion mutant deficient for chromatin opening (Fig. 4c, lanes 12-13) indicates that the chromatin remodeler can "feed forward" weak interactions with chromatin by a nucleosome-binding transcription factor. These results can explain *in vivo* findings that the loss of SWI/SNF results in a subset of sites where transcription factor binding is lost (Iurlaro et al., 2021; Xiao et al., 2022). However, our *in vitro* studies indicate that SWI/SNF activity is not required for the wild-type pioneer factors to initially access and locally open compacted chromatin. Our *in vitro* studies provide a mechanistic understanding of how pioneer factors and nucleosome remodeling complexes together play fundamental roles in enabling alterations of DNA-nucleosome contacts and chromatin accessibility, both necessary determinants of gene expression.

2.7 Main Figures

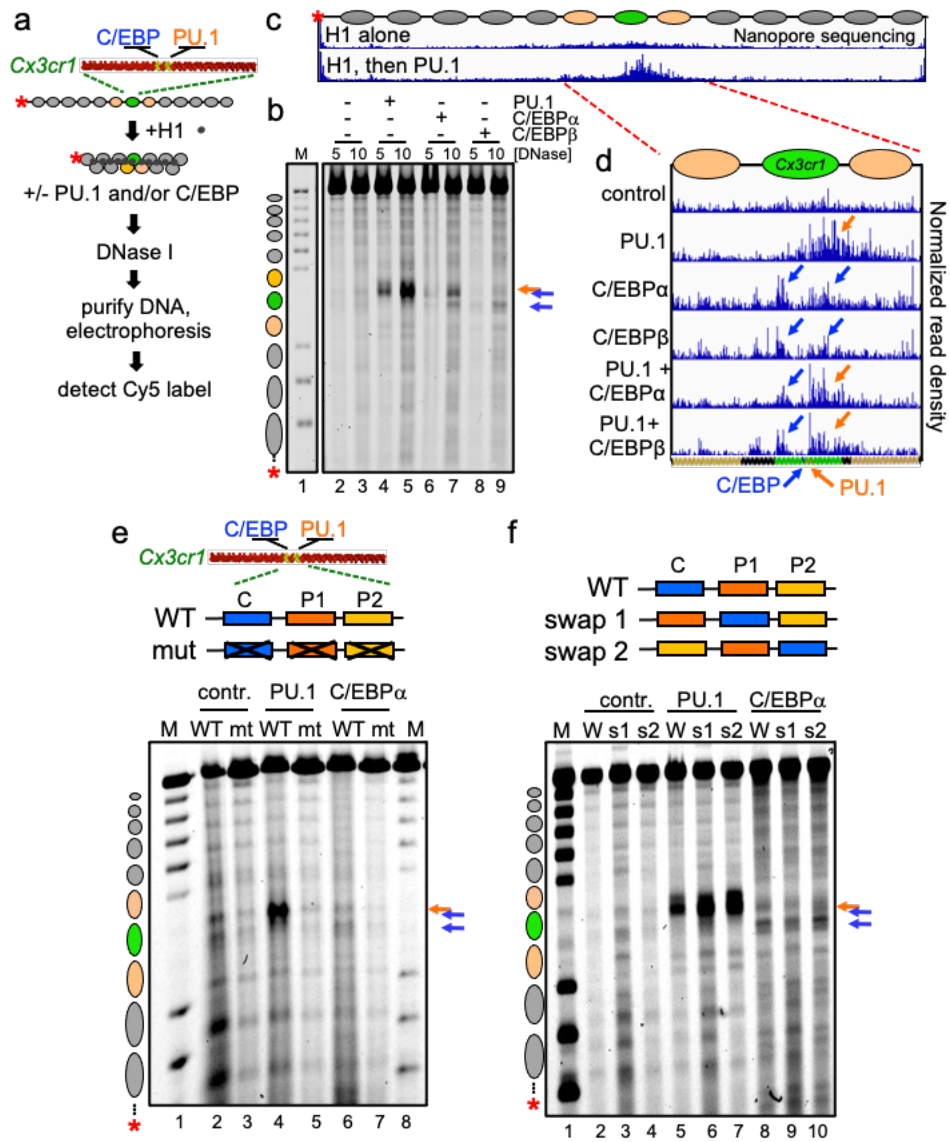


Figure 2 - 1 Different patterns of chromatin opening by PU.1 and C/EBP.

a, Schematic of chromatin assembly, compaction, transcription factor binding, and DNase I digestion assays.

b, DNase I digestion analysis of TFs binding to H1-compacted nucleosome arrays visualized by gel electrophoresis at two DNase I concentrations (in ng/uL). Distinct hypersensitive sites are observed, relative to no TF control (lanes 2-3), upon PU.1 (lanes 4-5, orange arrow), C/EBPα (lanes 6-7, blue arrows), and C/EBPβ (lanes 8-9, blue arrows). Lane 1 is a DNA marker (M), partial EcoRI digest of end-labeled array fragment with fragments in basepairs corresponding to (from bottom to top): 834, 1028, 1658, 1866, 2074, 2282, 2490, and 2733. Positions of nucleosomes on reconstituted arrays are indicated.

c and **d**, Nanopore sequencing endpoint analysis of DNase I digested (10 ng/uL) H1-compacted arrays from n = 2 biological replicates in extended data figure 3b . Ovals indicate the translational positions of nucleosomes determined in Extended Data Fig. 2d. Plots show normalized read density on the y axis. For each plot, the maximum value is set to 0.4% of reads.

e, DNase I digestion (10 ng/uL) of H1-compacted wild-type Cx3cr1 nucleosome arrays (WT) or motif mutant Cx3cr1 nucleosome arrays (mt) incubated alone (contr., lanes 2-3), with PU.1 (lanes 4-5) or C/EBPα (lanes 6-7). Representative gel image of 3 biological replicates. Lane 1 is a DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 834, 1028, 1658, 1866, 2074, 2282, 2490, and 2733. **f**, DNase I digestion of H1-compacted wild-type Cx3cr1 nucleosome arrays (WT) or PU.1 and C/EBPα motif swap arrays (swap 1 or swap 2) incubated alone (contr., lanes 2-4), with PU.1 (lanes 5-7) or C/EBPα (lanes 8-10). Representative gel image of 2 biological replicates. Lane 1 is a DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 624, 834, 1028, 1658, 1866, 2074, 2282, 2490, and 2733.

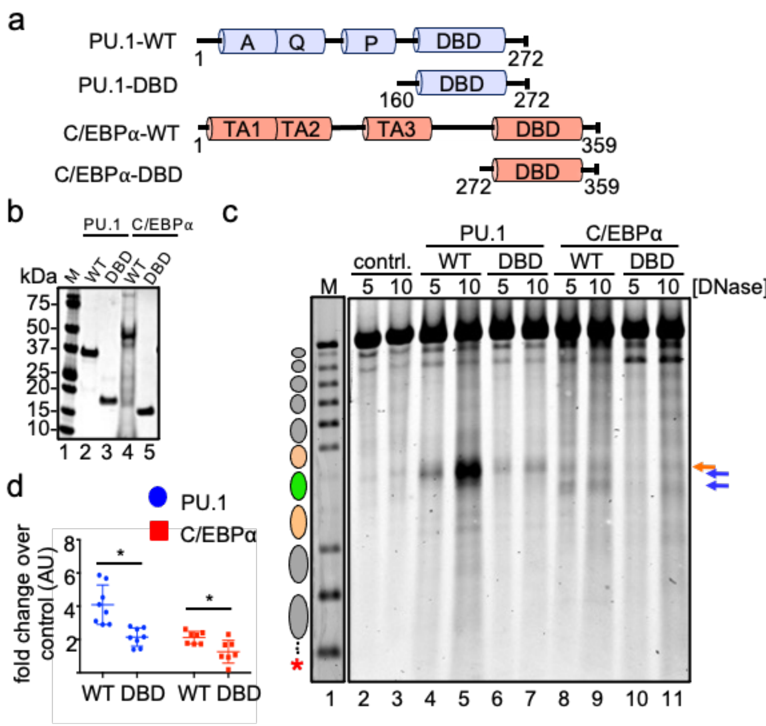


Figure 2 - 2 Domains outside of DBDs are needed to open chromatin.

a, Illustration of wild-type (WT) and DNA-binding domain (DBD) recombinant PU.1 and C/EBPα.

b, Coomassie stained SDS-PAGE of 1 ug of each indicated purified protein.

c, Representative gel image of DNase I digestion (5 and 10 ng/uL) analysis of WT and DBD TF opening of H1-compacted chromatin. Lane 1 is a partial EcoRI digest DNA marker (M) with fragments in base pairs corresponding to (from bottom to top): 624, 834, 1028, 1658, 1866, 2074, 2282, 2490, 2686, and 2733.

d, Quantified Cy5 signal in each lane normalized to no TF control from n=4 independent biological replicates represented in **c**, for PU.1 (blue) p = 0.0012 and for C/EBPα (red) p = 0.0079 . Data are mean ± SD and significance was determined by unpaired, two-tailed Student's t-test.

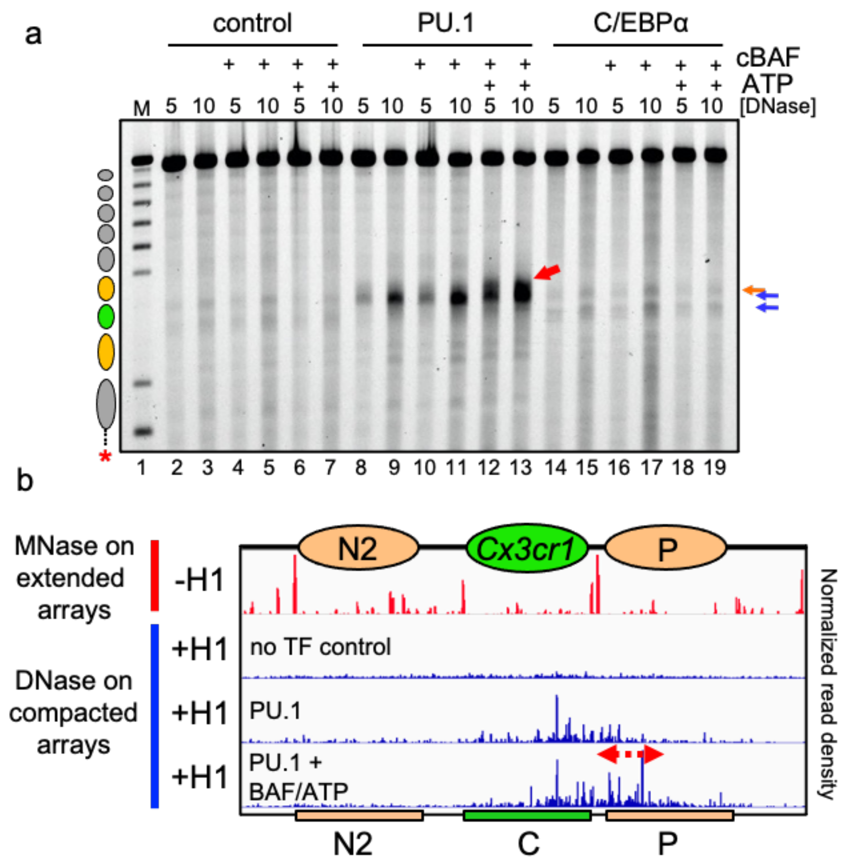


Figure 2 - 3 PU.1 Enables further chromatin opening by cBAF.

a, DNase I digestion analysis of H1-compacted nucleosome arrays incubated with PU.1 (lanes 8-13) or C/EBPα (lanes 14-19), in the presence of cBAF alone (lanes 10-11 and lanes 16-17, respectively) or cBAF and ATP (lanes 12-13 and lanes 18-19, respectively) representative of 2 biological replicates. Lane 1 is a partial EcoRI digest DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 834, 1028, 1658, 1866, 2074, 2282, 2490, 2686, and 2733.

b, Nanopore analysis of MNase digested extended arrays (-H1) and DNase I digested compacted arrays (+H1) alone (no TF control), with PU.1 alone, or with PU.1, cBAF and

ATP from n = 2 biological replicates in extended data figure 6e. Ovals indicate the translational positions of nucleosomes determined in Extended Data Fig. 2d. Plots show normalized read density on the y axis. For each plot, the maximum value is set to 0.4% of reads.

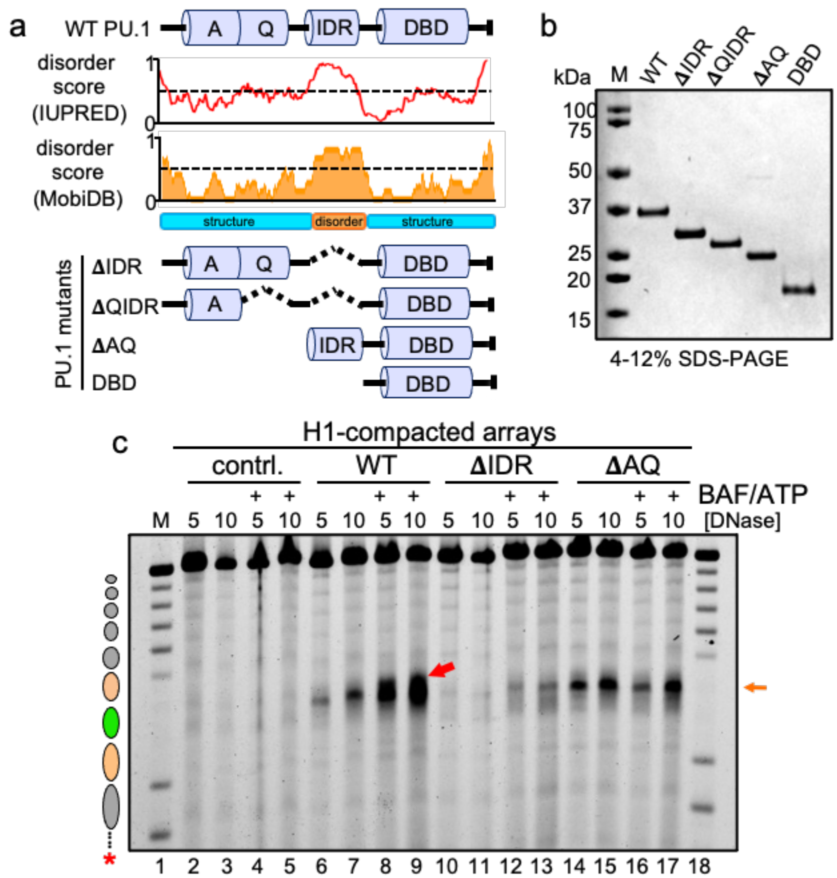


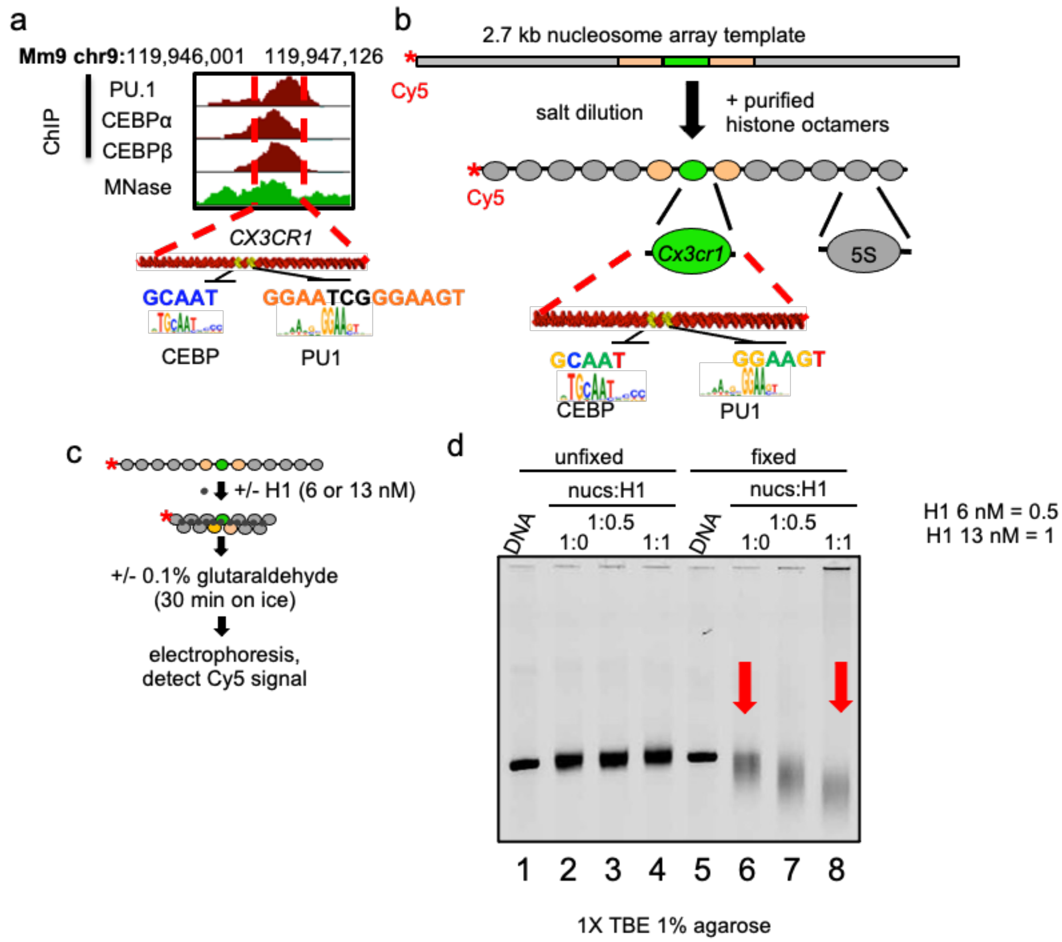
Figure 2 - 4 cBAF action requires AQ and IDR domains of PU.1.

a, Predicted disorder tendency of WT PU.1. The red and orange lines indicate the predicted disorder tendency along the WT PU.1 protein (as calculated by IUPred and MobiDB, respectively) with values above 0.5 (indicated by the dotted line) considered disordered. Shown are the positions of the Acidic domain (A), Q-rich domain (Q), Intrinsically disordered region (IDR) and DNA-binding domain (DBD). The deletion mutant series of PU.1 are indicated below as **Δ**IDR, **Δ**QIDR, **Δ**AQ, and DBD.

b, SDS-PAGE analysis of 1 µg of each protein, stained with Coomassie blue. Gel image representative of 2 replicates.

c, DNase I digestion analysis H1-compacted nucleosome arrays incubated alone (contrl., lanes 2-5), WT PU.1 (lanes 6-9), **Δ**IDR deletion (lanes 10-13), and **Δ**AQ truncation (lanes 14-17) with or without the addition of BAF and ATP. Representative gel image of 2 biological replicates. Lanes 1 and 18 are a DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 624, 834, 1028, 1658, 1866, 2074, 2282, 2490, and 2733.

2.8 Extended Data Figures



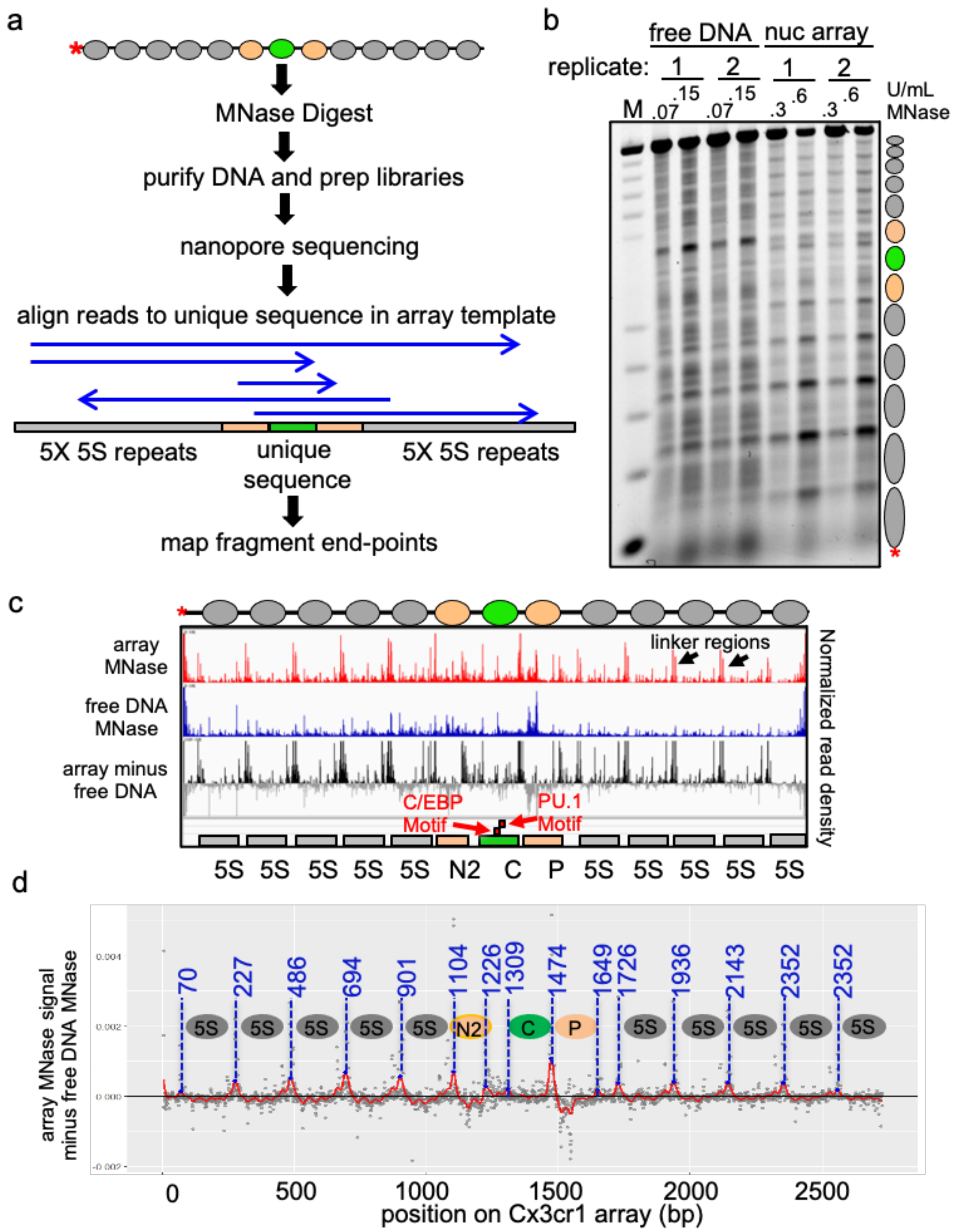
Extended Data 2 - 1 Generation of H1-compacted Cx3cr1 nucleosome arrays.

a, PU.1 and C/EBP ChIP-seq profiles (red) in macrophages and MNase-seq profile (green) in fibroblasts near the Cx3cr1 gene within the displayed region of the mouse genome with TF motifs indicated.

b, Diagram of the dCTP-Cy5 end-labeled Cx3cr1 nucleosome array.

c, Schematic of linker histone-mediated chromatin compaction.

d, Representative EMSA of linker histone-mediated chromatin compaction by native agarose gel electrophoresis. Free DNA (lanes 1 and 5) and nucleosome arrays with nucleosome:linker histone ratios of 0 (lanes 2 and 6), 0.5 (lanes 3 and 7) and 1.0 (lanes 4 and 8) are shown. Representative gel image of 2 biological replicates.



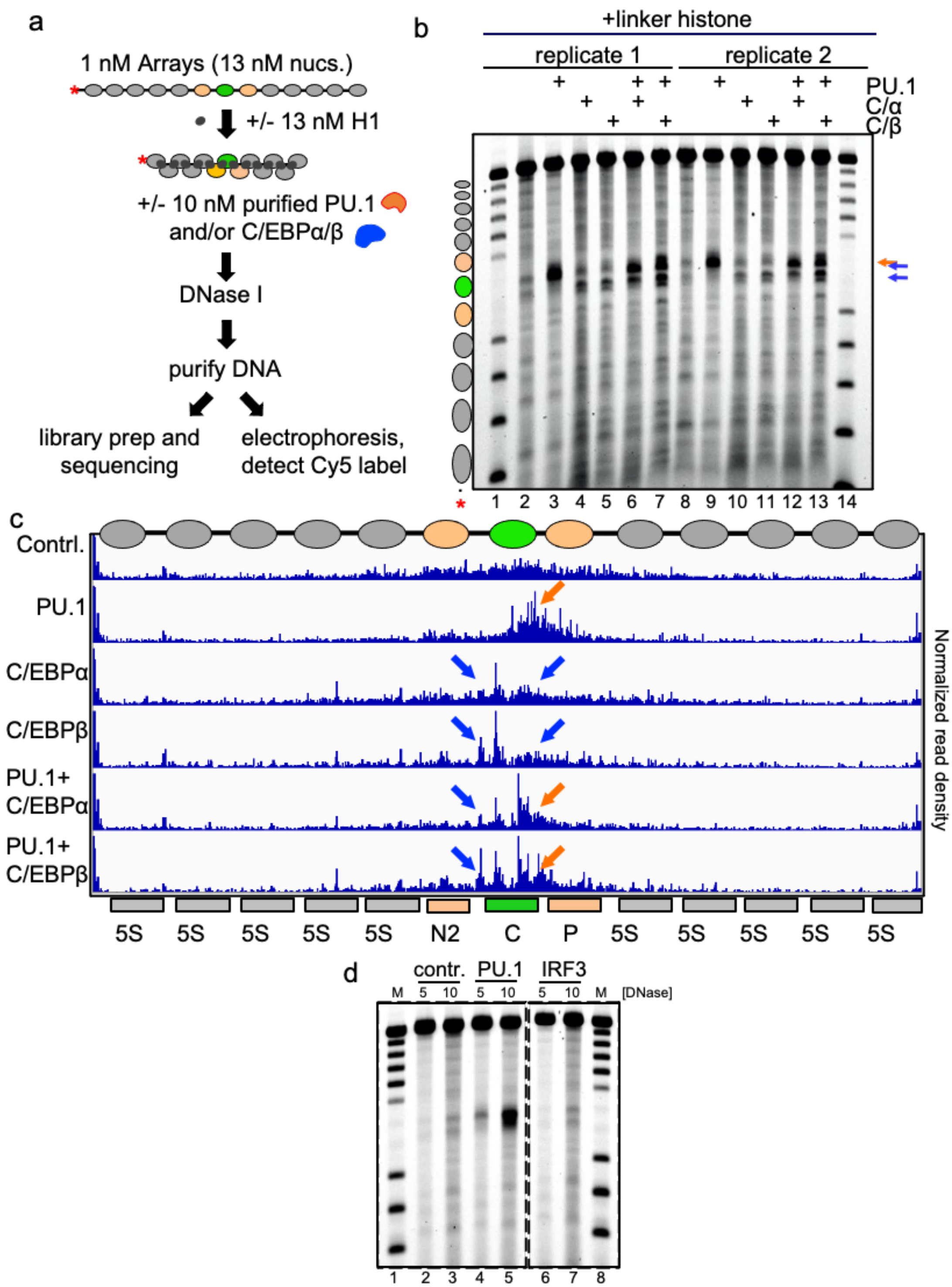
Extended Data 2 - 2 Nanopore sequencing to determine translational position of nucleosomes in the Cx3cr1 array.

a, Schematic of Nanopore sequencing and endpoint analysis pipeline.

b, MNase digestion analysis of free DNA and extended nucleosome arrays at two MNase concentrations (U/mL) visualized by gel electrophoresis. Two biological replicates are shown on the same gel. Lane 1 is a partial EcoRI digest DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 208, 416, 624, 834, 1028, 1658, 1866, 2074, 2282, 2490, 2733.

c, IGV visualization of Nanopore sequencing endpoint analysis of MNase digested free DNA (0.07 U/mL MNase), extended nucleosome arrays (0.3 U/mL) and free DNA signal subtracted from array signal to account for MNase site bias. n = 2 biological replicates in **b**. Plots show normalized read density on the y axis. For each plot, the maximum value is set to 0.4% of reads.

d, Determination of nucleosome translational positions.



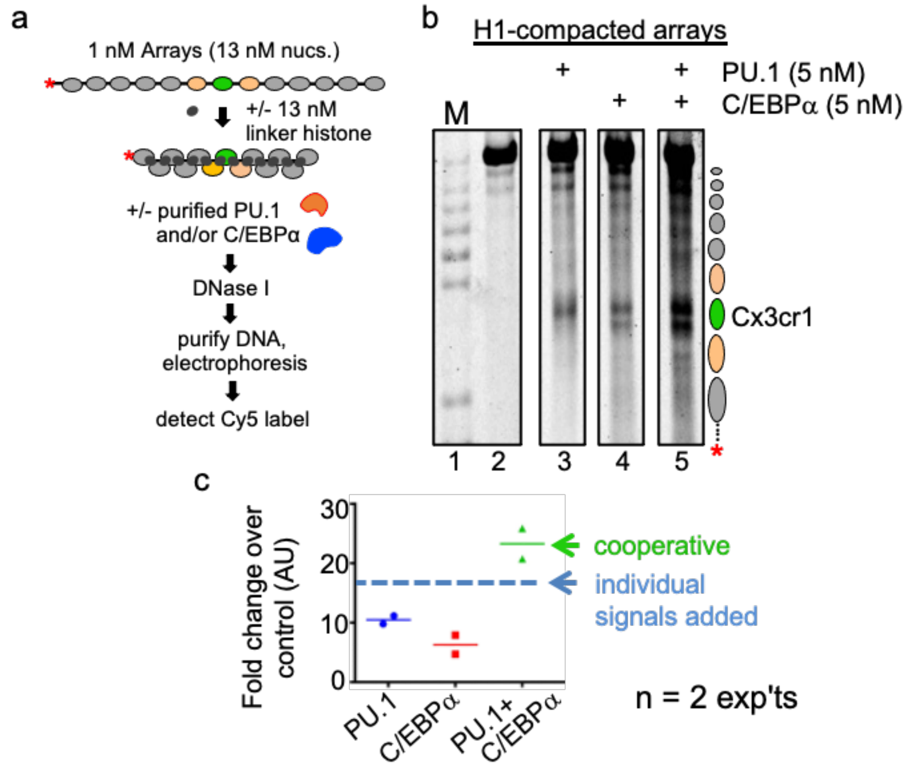
Extended Data 2 - 3 Nanopore sequencing of TFs incubated with H1-compacted arrays.

a, Schematic of DNase I digestion analysis.

b, DNase I digestion analysis of TFs binding to H1-compacted nucleosome arrays visualized by gel electrophoresis (10 ng/uL DNase I). Two biological replicates are shown on the same gel. Lanes 1 and 14 are a partial EcoRI digest DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 416, 624, 834, 1028, 1658, 1866, 2074, 2282, 2490, 2733.

c, IGV visualization of Nanopore endpoint analysis of DNase I digested H1-compacted nucleosome arrays incubated with indicated transcription factor(s) from n = 2 biological replicates in **b**. Plots show normalized read density on the y axis. For each plot, the maximum value is set to 0.4% of reads.

d, DNase I digestion analysis of indicated TFs binding to H1-compacted nucleosome arrays visualized by gel electrophoresis. Gel image is representative of 2 biological replicates. Lanes 1 and 8 are a partial EcoRI digest DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 624, 834, 1028, 1658, 1866, 2074, 2282, 2490, 2733. All lanes shown are from the same gel.

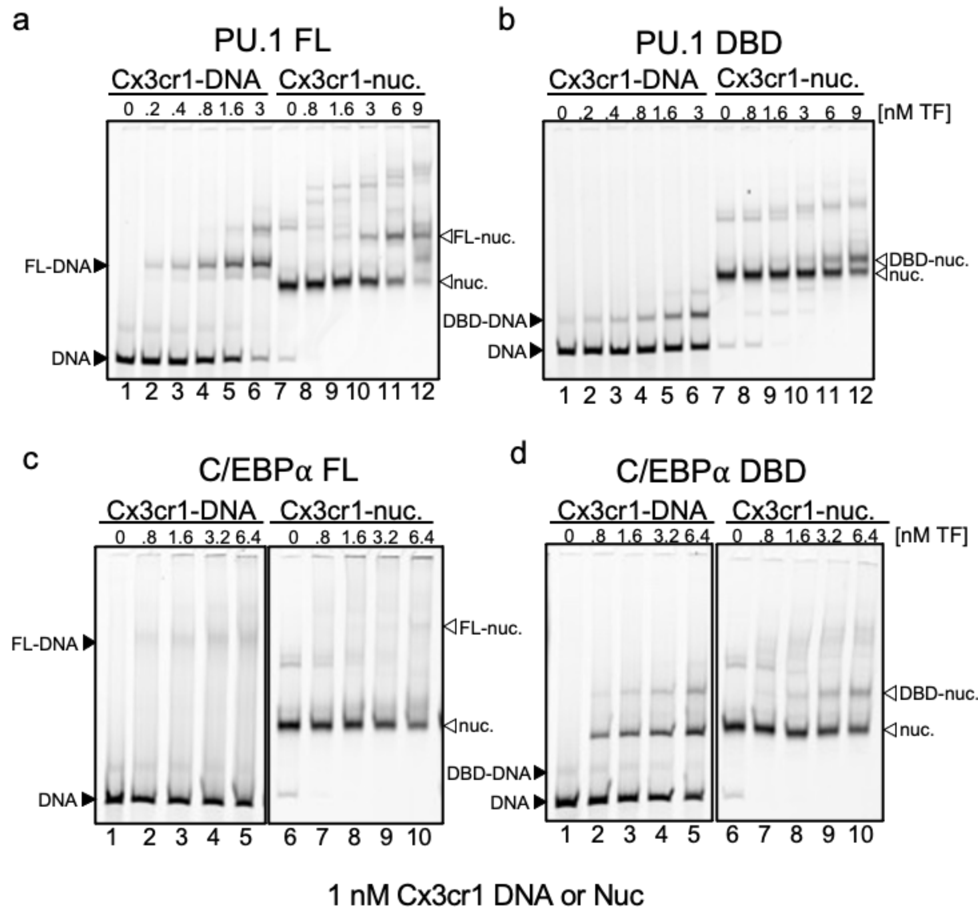


Extended Data 2 - 4 PU.1 and C/EBPα open chromatin cooperatively.

a, Schematic of DNase I digestion analysis.

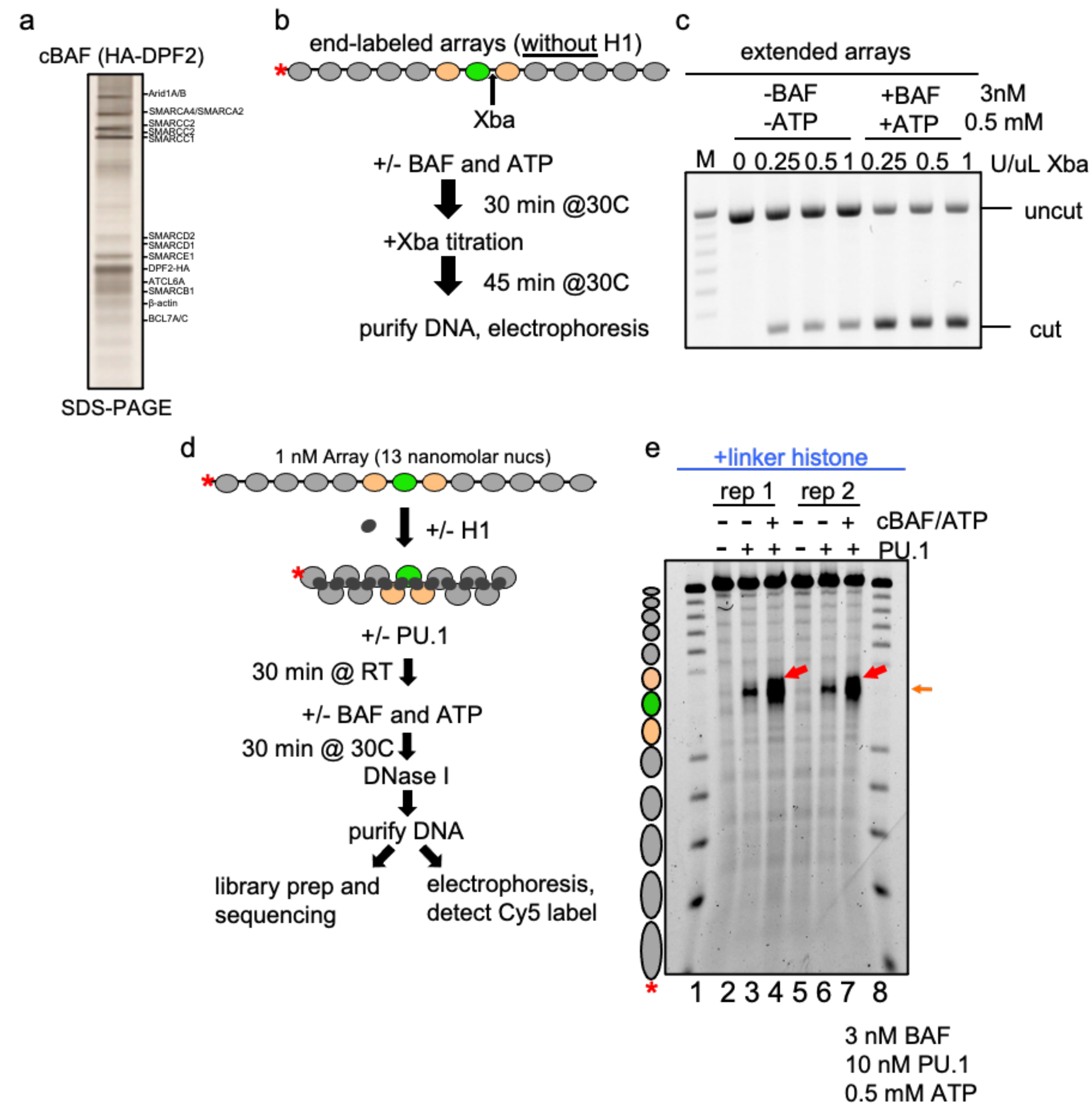
b, DNase I digestion analysis of TFs binding to H1-compacted nucleosome arrays visualized by gel electrophoresis. Gel image representative of 2 bioligcal replicates.

c, Quantified Cy5 signal in each lane normalized to no TF control from n = 2 biological replicates in **b**.



Extended Data 2 - 5 DNA-binding domains of PU.1 and C/EBPα are sufficient for nucleosome binding.

a-d, Representative EMSAs from 2 biological repeats showing the affinity of increasing amounts of **a**, Full-length PU.1 (PU.1 FL), **b**, PU.1-DBD, **c**, Full-length C/EBPα (C/EBPα FL), **d**, C/EBP-DBD to Cy5-labeled Cx3cr1 free DNA (black arrows) or mononucleosomes (white arrows).



Extended Data 2 - 6 cBAF readily remodels extended nucleosome arrays but requires PU.1 to access H1-compacted arrays.

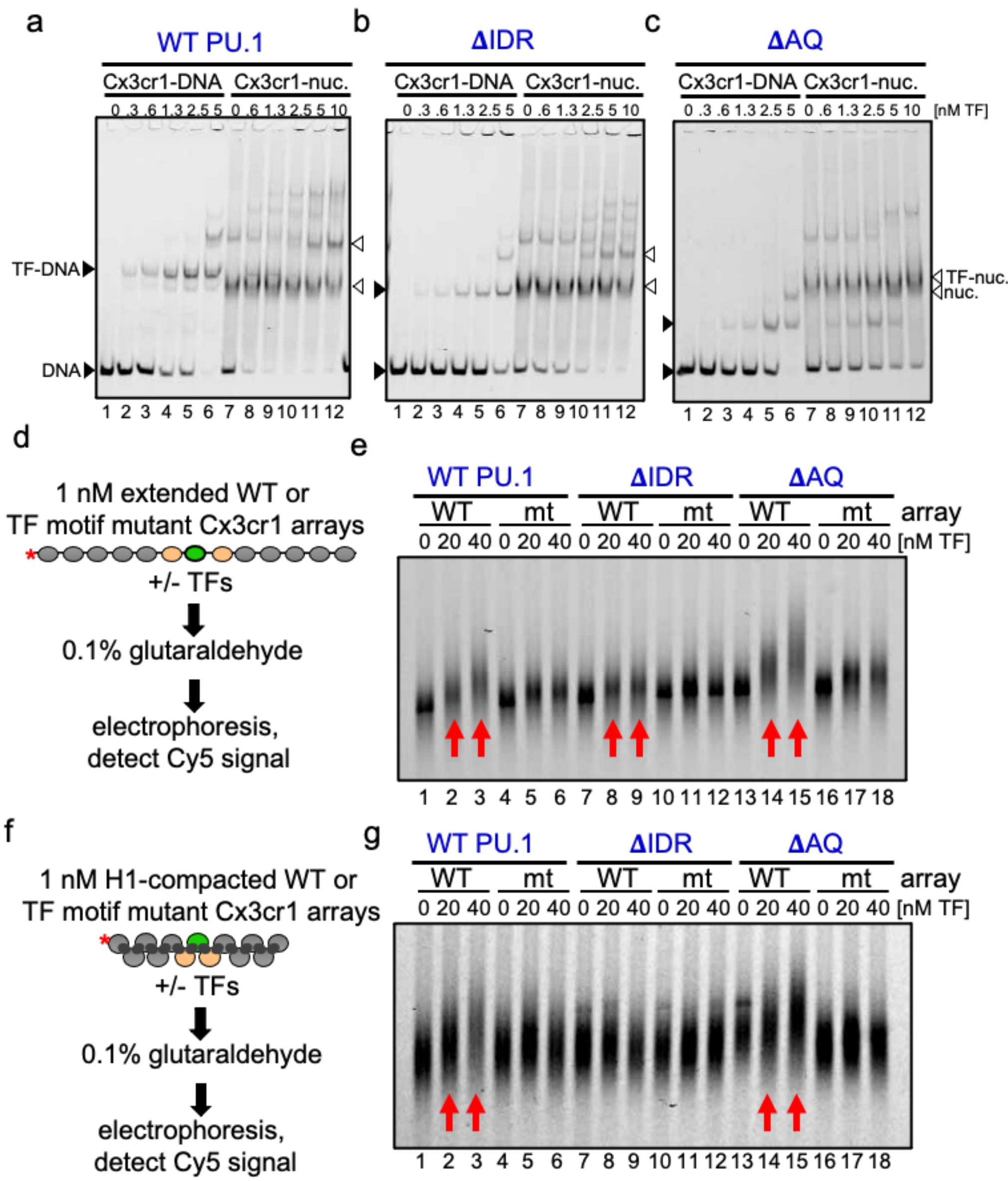
a, silver-stain of affinity-purified cBAF complexes from mammalian HEK293T cells expressing HA-DPF2.

b, Schematic of XbaI accessibility assay with extended arrays and cBAF remodeling complex.

c, Agarose gel visualization of XbaI accessibility assay of extended nucleosome arrays (no linker histone) incubated without and with cBAF plus ATP. Gel image representative of 2 biological replicates. Lanes 1 is a partial EcoRI digest DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 1658, 1866, 2074, 2282, 2490, 2733.

d, Schematic of DNase I digestion analysis of transcription factors and cBAF with H1-compacted arrays.

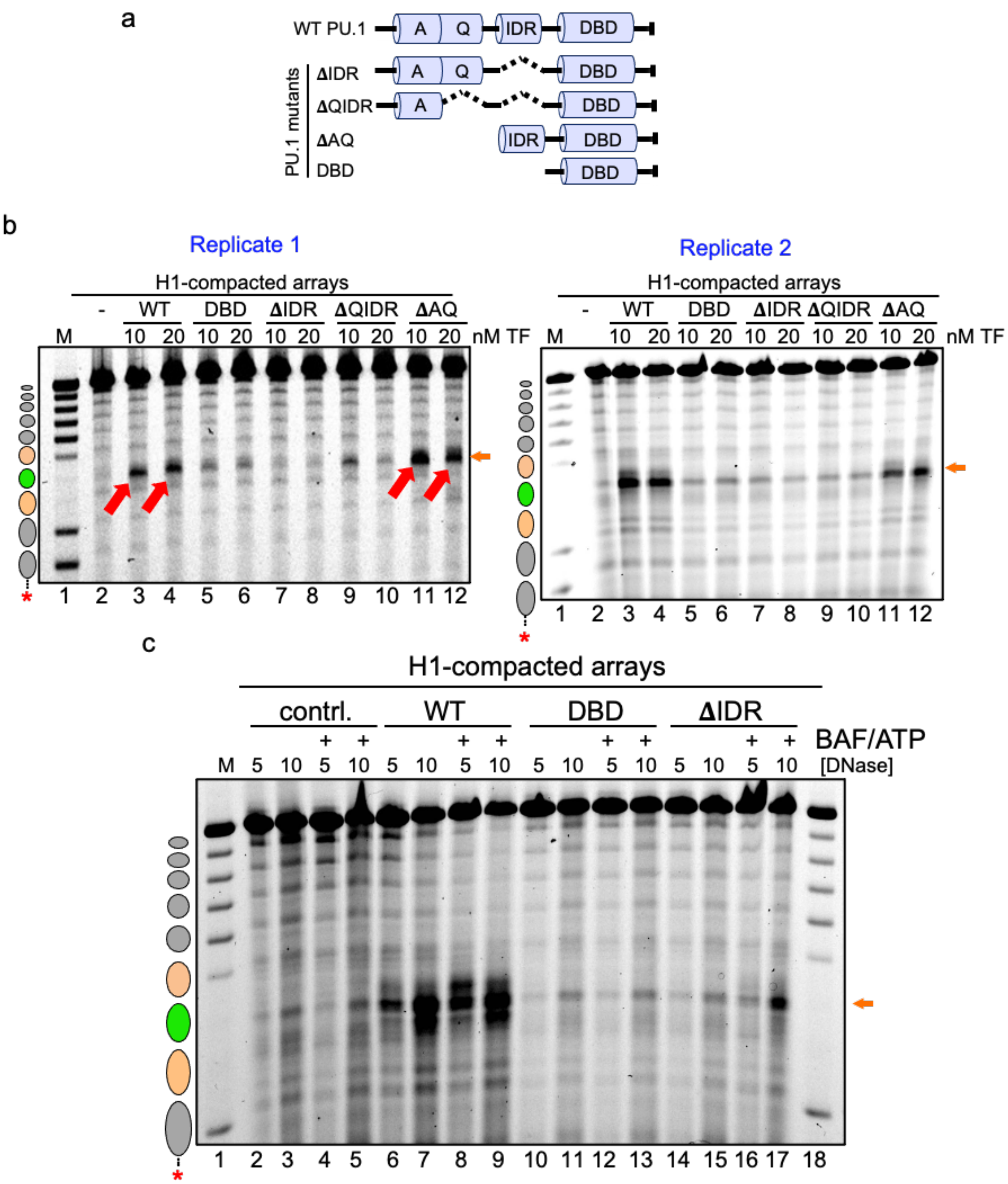
e, DNase I digestion analysis H1-compacted nucleosome arrays incubated alone (no TF, lanes 2 and 5), PU.1 alone (lanes 3 and 6), and PU.1 with cBAF and ATP (lanes 4 and 7). The same samples were used to perform nanopore sequencing shown in Fig. 3b. Lanes 1 and 8 are a partial EcoRI digest DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 208, 416, 624, 834, 1028, 1658, 1866, 2074, 2282, 2490, 2733.



Extended Data 2 - 7 Hierarchy of chromatin binding by PU.1 WT and mutants.

Representative EMSAs of 2 biological repeats showing the affinity of increasing amounts of **a**, WT PU.1, **b**, ΔIDR, **c**, ΔAQ to Cy5-labeled Cx3cr1 free DNA (black arrows) or mononucleosomes (white arrows). **d**, Schematic of EMSA performed with transcription factors and extended nucleosome arrays **e**, Representative EMSA of 2 biological repeats showing the affinity of increasing amounts of WT PU.1 (lanes 1-6), ΔIDR (lanes

7-12), and **Δ**AQ (lanes 13-18) to Cy5-labeled extended WT or TF motif mutant Cx3cr1 nucleosome arrays. **f**, Schematic of EMSA performed with transcription factors and H1-compacted nucleosome arrays. **g**, Representative EMSA of 2 biological replicates showing the affinity of increasing amounts of WT PU.1 (lanes 1-6), **Δ**IDR (lanes 7-12), and **Δ**AQ (lanes 13-18) to Cy5-labeled H1-compacted WT or TF motif mutant Cx3cr1 nucleosome arrays.



Extended Data 2 - 8 The IDR domain of PU.1 is most crucial for chromatin opening.

a, Illustration of the PU.1 deletion mutant series. Shown are the positions of the acidic domain (A), Q-rich domain (Q), intrinsically disordered region (IDR) and DNA-binding domain (DBD).

b, DNase I digestion analysis of two transcription factor concentrations binding to H1-compacted nucleosome arrays visualized by gel electrophoresis (10 ng/uL DNase I). Two biological repeats are shown. Lanes 1 is a partial EcoRI digest DNA marker (M) with fragments in basepairs corresponding to (from bottom to top): 1028, 1658, 1866, 2074, 2282, 2490, 2733.

c, DNase I digestion analysis of H1-compacted nucleosome arrays incubated with no TFs (contrl., lanes 2-5), WT PU.1 (lanes 6-9), DBD (lanes 10-13), and **Δ**IDR deletion (lanes 14-17) with or without the addition of cBAF and ATP. Gel image representative of two biological replicates.

2.9 Materials and Methods

Protein Expression and Purification

The bacterial expression plasmids: pET-28b-PU.1, pET-28b-C/EBPα, pET-28b-C/EBPβ encode the mouse PU.1, C/EBPα, and C/EBPβ fused to an N-terminal 6X histidine tag, respectively(Fernandez Garcia et al., 2019). pET-28b-PU.1-DBD (encodes PU.1-DBD, residues 160-272) and pET-28b-C/EBPα-DBD (encodes C/EBPα-DBD, residues 272-359) were generated by PCR from the respective full-length construct, introducing NdeI and XhoI restriction sites for insertion into the pET-28b plasmid. The remaining mutant PU.1 constructs, pET-28b-PU.1-**Δ**IDR (encodes residues 1-115, 161-272 of PU.1), pET-28b-PU.1-**Δ**QIDR (encodes residues 1-73, 101-272 of PU.1), pET-28b-PU.1-**Δ**AQ (encodes residues 116-272 of PU.1), were generated by infusion cloning of gene blocks (IDT) using XhoI NdeI digested pET-28b.

The histidine-tagged proteins were expressed in *E. Coli* Rosetta (DE3) pLysS (Novagen # 70956-3). Transformed cells were grown at 37°C to a density of 0.5-0.7 at an absorbance of 600 nm and protein expression was induced with 1 mM IPTG at 16°C for 16 hr for PU.1 and PU.1 mutant constructs; 1 mM IPTG at 37°C for 4 hr CEBPβ; 2 mM IPTG at 37°C for 4 hr CEBPα and C/EBPα-DBD.

The recombinant proteins were purified over Ni-NTA resin under denaturing conditions (20 mM Tris-HCl pH 8.0, 0.5M NaCl, 6M Urea) with 5 mM imidazole and 20-300 mM single step imidazole changes follow by 4M and 2M urea step dialysis. The recombinant human full length histones H2A, H2B, H3, and H4 were expressed and purified as described previously(Tanaka et al., 2004). Recombinant human IRF3 was purchased (Active Motif 31544).

Human cBAF Complex Purification

Human cBAF complexes were purified as described previously(Mashtalir et al., 2020, 2018), with several modifications. Briefly, HEK-293T cells were lentivirally transduced with HA-DPF2 for the specific capture of cBAF complexes (gift of Gerald Crabtree, Stanford University). Pellets were washed in cold phosphate-buffered saline (PBS) (GIBCO). Cell suspension was then centrifuged for 5 minutes at 4 degrees C at 4000 g. Cell pellets were resuspended in HB (10 mM Tris-HCl, pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 1 mM PMSF) and homogenized in a glass Dounce homogenizer. The suspension was pelleted for 30 minutes at 4 degrees C at 4000 g and nuclear pellets were resuspended in pre-extraction buffer (50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 1 mM EDTA, 1 mM, 0.1% (v/v) NP40, 1 mM DTT, 1 mM PMSF and protease inhibitor cocktail. After pelleting for 10 minutes at 4 degrees C at 4000 g, chromatin was resuspended in high-salt buffer (HSB; 50 mM Tris-HCl, pH 7.5, 300 mM KCl, 1 mM $MgCl_2$, 1 mM EDTA, 1 mM, 1% NP40, 1 mM DTT, 1 mM PMSF and protease inhibitor cocktail) and incubated with rotation for one hour at 4 degrees C. Homogenates were centrifuged at 20,000 rpm in a SW32Ti rotor (Beckman Coulter) for 1 hour at 4 degrees C and the supernatant collected. Nuclear extracts were filtered with a 0.45-mm filter and rotated overnight at 4 degrees C with HA magnetic resin (Thermo Fisher Scientific). The beads were washed in HSB and proteins were eluted by incubating the beads 4 times for 1.5 hours with 1 mg/mL of HA peptide in HSB. Eluted protein complexes were separated by density using 10%–30% glycerol gradients and fractions with confirmed purified expression of full cBAF complex were pooled and concentrated using the appropriate molecular weight cutoff protein concentrator (Pierce). Samples were snap frozen and kept at -80°C until needed.

Generation of Nucleosome Array Templates

The 152 bp Cx3cr1 mononucleosome sequence that corresponds to the genomic location: mm9 chr9:119946611-119946762 (Fernandez Garcia et al., 2019) was extended on the 5' side by 36 bp and on the 3' side by 37 bp to incorporate linker DNA and BfuAI and XbaI flanking sequences:

ACCTGCGTTACAGCATCCACTCAGTATCCCTTGAGCCCCGCGTGCAAAGCCCAGG
AGCCCTTGCTTAGGTGCAGGGCCTCTCGGCTGCTGATCTTCAGCTGGTTGCTGAG
AGTTGCAGCATTGCTGAGTCTTAGCAATGGATACTTCCCGATTCCCCTCACAAAAT
AGGTCAGTCTGTCTGGCTAGTTCTGTACTTGCAGACACAGGGCATGTGGGGTTCCT
ATTTTTCTAGCTCCCAGGCTTCTGTCTGCTTCCTTCGTTTAGTATGTCTA

The DNA sequence was generated by PCR of genomic mouse DNA with BfuAI and XbaI restriction sites cloned into the p208-10.N1N2(Cirillo et al., 2002) to generate p208-10-Cx3cr1. Mutant Cx3cr1 arrays were generated by infusion cloning using linearized BfuAI XbaI digested p208-10-Cx3cr1 plasmids. Geneblocks for cloning were obtained from IDT:

TF mutant motif Cx3cr1:

AAAATAACCTGCGTTACAGCATCCACTCAGTATCCCTTGAGCCCCGCGTGCAAAGC
CCAGGAGCCCTTGCTTAGGTGCAGGGCCTCTCGGCTGCTGATCTTCAGCTGGTTG
CTGAGAGTTGCAGCATTGCTGAGTCCCGTATAGAGATCGTATAGCGTATAGAGCAC
AAAAATAGGTCAGTCTGTCTGGCTAGTTCTGTACTTGCAGACACAGGGCATGTGGG
GTTCCTATTTTTCTAGCTCCCAGGCTTCTGTCTGCTTCCTTCGTTTAGTATGTCTAGA
CTACAGTTATTGGTT

Swap1 Cx3cr1:

AAAATAACCTGCGTTACAGCATCCACTCAGTATCCCTTGAGCCCCGCGTGCAAAGC
CCAGGAGCCCTTGCTTAGGTGCAGGGCCTCTCGGCTGCTGATCTTCAGCTGGTTG
CTGAGAGTTGCAGCATTGCTGAGTCTACTTCCCGAATTAGCAATGGTTCCCCTCAC
AAAAATAGGTCAGTCTGTCTGGCTAGTTCTGTACTTGCAGACACAGGGCATGTGGG
GTTCCTATTTTTCTAGCTCCCAGGCTTCTGTCTGCTTCCTTCGTTTAGTATGTCTAGA
CTACAGTTATTGGTT

Swap2 Cx3cr1:

AAAATAACCTGCGTTACAGCATCCACTCAGTATCCCTTGAGCCCCGCGTGCAAAGC
CCAGGAGCCCTTGCTTAGGTGCAGGGCCTCTCGGCTGCTGATCTTCAGCTGGTTG
CTGAGAGTTGCAGCATTGCTGAGTCTACTTCCCGAATTAGCAATGGTTCCCCTCAC
AAAAATAGGTCAGTCTGTCTGGCTAGTTCTGTACTTGCAGACACAGGGCATGTGGG
GTTCCTATTTTTCTAGCTCCCAGGCTTCTGTCTGCTTCCTTCGTTTAGTATGTCTAGA
CTACAGTTATTGGTT

Plasmids were amplified in E.coli, purified, and digested with MluI, SmaI and HhaI to release array fragments. Fragments were purified by ethanol precipitation then labeled with dCTP-Cy5 (Cytiva PA55021) by end-repair with Klenow fragment DNA polymeraase (3'-5' exo-) (NEB). Cy5-DNA labeling reactions were carried out at final concentrations of 1 μM DNA, 4 mM Cy5-dCTP and 0.5 U/μL Klenow fragment in the presence of excess 4mM dATP, 4 mM dTTP, and 4 mM dGTP in NEB buffer 2 (50mMNaCl, 10mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT). The reaction was incubated at 37C for 1 hr. Labeling reactions were purified by phenol:chloroform extraction followed by FPLC purification of the 2.7 kb MluI-SmaI fragment using a Capto HiRes Q anion exchange column (Cytiva 29275878).

Nucleosome Array and Mononucleosome Reconstitution

Nucleosome arrays were assembled as previously described(Cirillo et al., 2002). Assembly reactions containing 2 µg of end-labeled DNA fragments and core histones at a 1:1.0 molar ratio of octamers to nucleosomal sites, and 2 M NaCl in a total of 10 µl were incubated at 37°C for 15 min, serially diluted by adding 3.3, 6.7, 5, 3.6, 4.7, 6.7, 10, 30, and 20 µl of 50 mM HEPES (pH 7.5), 1 mM EDTA, 5 mM DTT, and 0.5 mM PMSF in 15 min incubation steps at 30°C, and brought to 0.1 M NaCl by adding 100 µl of 10 mM Tris·HCl (pH 7.5), 1 mM EDTA, IGEPAL, 5 mM DTT, 0.5 mM PMSF, and 20% glycerol, followed by incubation at 30°C for 15 min. Array saturation was determined by EcoRI digestion, performed with a concentration of 2 nM nucleosome array or free DNA in 20 mM HEPES (pH 7.5), 50 mM KCl, 1% glycerol, 5 mM DTT, 100 µg/ml BSA, and 2 U/µl EcoRI. Following a 2 hr incubation at 37°C, digestion products were resolved on 4% polyacrylamide 0.5X TBE gels and analyzed as described(Cirillo et al., 2002).

Cx3cr1 mononucleosomes were reconstituted by dialysis with the Cy5 end-labeled 162 bp Cx3cr1 DNA fragment and recombinant human histones as described previously (Fernandez Garcia et al., 2019).

TF Binding Reactions and Enzymatic Assays

Binding reactions for the DNase I digestion, MNase, and Restriction enzyme assays were carried out with 1 nM (1.5 ng/µl) of nucleosome array (13 nM of nucleosomes) in 6 mM HEPES (pH 7.5), 35 mM KCl, 1.5% glycerol, 2 mM DTT, 250 µg/ml BSA. In experiments with compacted nucleosome arrays, 13 nM purified histone H1 (Calbiochem) was incubated with the nucleosome arrays at room temperature for 1 hr. Purified transcription factors were then incubated with the nucleosome arrays at room temperature for 1 hr. For cBAF remodeling assays 10nM of indicated transcription factors were incubated with H1-compacted arrays at room temperature followed by incubation with cBAF (6 nM) and ATP (0.5mM) for 30 minutes at 30°C.

DNase digestions were carried out by 2.5-10 ug/mL of DNase I diluted in 50 mM $MgCl_2$, followed by incubation at room temperature for 1 min. MNase digestion was performed by addition of .075-0.6 U/mL MNase diluted in 30 mM CaCl2, followed by incubation at room temperature for 1 min. XbaI restriction enzyme digestion was performed by incubation with 0.25-1 U/mL XbaI for 45 minutes at 30°C. Reactions were terminated by addition of 1 volume of stop buffer (20 mM Tris pH 7.5, 50 mM EDTA, 1% SDS, 0.5 mg/mL tRNA, and 0.2 mg/mL proteinase K) and incubated at 50C for 15 min. Digestion products are purified by ethanol precipitation and electrophoresed on 1% Agarose 1X TBE gels.

Electromobility shift assays with end-labeled 162 bp Cx3cr1-DNA or mononucleosomes(Fernandez Garcia et al., 2019) incubated with recombinant proteins were performed in DNA-binding buffer (10 mM Tris-HCl pH7.5, 1 mM $MgCl_2$, 10 mM ZnCl2, 1 mM DTT, 50 mM KCl, 3 mg/ml BSA, 5% Glycerol) at room temperature for 30 min. Free and bound DNA were separated on 4% non-denaturing polyacrylamide gels run in 0.5X Tris–borate–EDTA. Electromobility shift assays with extended or H1-compacted nucleosome arrays incubated with recombinant proteins were performed in 6 mM HEPES (pH 7.5), 35 mM KCl, 1.5% glycerol, 2 mM DTT for 1 hour, fixed on ice with 0.1% glutaraldehyde for 30 minutes, then electrophoresed on 1% agarose gel run in 0.5X Tris–borate–EDTA.

Gels were visualized using with an Amersham Typhoon RGB Biomolecular Imager using Cy5 fluorescence setting (excitation at 633 nm and emission filter 670 BP 30). All experiments were performed at least twice.

Nanopore Sequencing Library Prep

Partially MNase-digested chromatin retaining contiguous oligonucleosomes was used for generation of libraries for nanopore sequencing. Purified DNA was end-repaired and A-tailed with the NEBnext Ultra II end repair module (New England Biolabs), purified with Ampure XP beads (Beckman Coulter) and ligated to AMX adapter from the Oxford Nanopore 1D sequencing Kit (SQK-LSK108, Oxford Nanopore Technologies) using NEB Blunt/TA ligase master mix. Excess adapters were removed by adding 0.5 volumes of Ampure XP beads and washing with ABB buffer from the 1D sequencing kit. Purified library was eluted and loaded to a MinION R9.4 flowcell according to the manufacturer's instructions and run for 24 h.

Nanopore Sequencing Analysis

The fast5 raw data files from the Nanopore sequencer were basecalled and converted to fastq format with Guppy Software version 6.1.2 (Oxford Nanopore Technologies). Reads were mapped to the *Cx3cr1 Array* reference (see supplementary sequence file) with minimap2, a mapping algorithm specifically designed to analyze nanopore sequencing reads. Bam files produced by minimap2 were imported into R version 4.1.1("R: The R Project for Statistical Computing," n.d.) then filtered and normalized using tidyverse packages version 1.1.4(Wickham et al., 2019). Specifically, reads were excluded if they were identified as multimappers by minimap2 or were located entirely within the non-unique regions of the nucleosome array DNA sequence. The ends of each read from an array digestion identify the precise nucleotide at which a nuclease was able to access and cleave the DNA in the array. We quantified the endpoints of each read using the hash package version 2.2.6.1 and then normalized each dataset to the total number of mapped reads. To determine nucleosome translational positions, the data was visualized using the ggplot2 package and the smoothed cleavage traces were produced with the kernel smoothing function in R.

Protein disorder prediction

Disorder tendency of PU.1 was calculated using the IUPRED tool(Dosztányi et al., 2005b) and MobiDB(Piovesan et al., 2021), as recommended by the programs.

Statistics

Unpaired, two-tailed, Student's t-tests were perform using GraphPad Prism 7 software to test differences in chromatin opening efficiencies between full-length transcription factors and their DNA-binding domains.

Data and materials availability

All cloned reagents described are available from the authors. Nanopore FASTq and mapped bigwig files have been deposited to the NCBI GEO under accession number GSE216065.

Code availability

All code for data analysis and visualization is available from the corresponding author on request.

CHAPTER 3: PERSPECTIVES AND FUTURE DIRECTIONS

3.1 Summary of Major Conclusions

The goal of my thesis work was to investigate the basis by which different hematopoietic transcription factors engage compact chromatin structures and initiate chromatin opening. I focused my project on the hematopoietic ETS factor PU.1 and the bZIP factors C/EBPα, and C/EBPβ, which are required for the development of macrophages from hematopoietic stem cells (Friedman, 2007b; Scott et al., 1994a). In addition, ectopic expression of PU.1 and C/EBPα/β converts fibroblasts to the macrophage lineage (Feng et al., 2008; Van Oevelen et al., 2015b). These developmentally indispensable transcription factors are thought to act as “pioneers”: a special subset of TFs that can directly bind nucleosomal DNA within compacted chromatin and initiate chromatin accessibility by exposing the targeted nucleosome. Previously, our lab identified nucleosomes targeted by PU.1, C/EBPα, and C/EBPβ *in vivo* and reconstituted these nucleosomes *in vitro*. We determined that all three factors can directly bind nucleosomes with varying affinities in the nanomolar range. Together these data indicated that the presence of a short alpha-helix for sequence recognition in the DNA binding domain predicted the ability of TFs to strongly bind mononucleosomes (Fernandez Garcia et al., 2019).

Many other studies have focused on pioneer transcription factor binding to nucleosomes and have found that such binding leads to nucleosomal perturbations, including DNA distortions and repositioning of core histones, which would likely interfere with chromatin compaction. Indeed, the binding of the pioneer factors FoxA and Rap1 to linker histone H1- or salt-compacted nucleosome arrays, respectively, *in vitro* is sufficient to induce local chromatin opening (Cirillo et al., 2002; Mivelaz et al., 2020). *In vivo*, pioneer transcription factor binding precedes overt chromatin opening, which presumably only occurs after additional co-factors and nucleosome remodelers have been recruited (Li et al., 2018; Mayran et al., 2018; Sardina et al., 2018). Therefore, there was a gap in understanding how nucleosome perturbations that occur upon pioneer factor binding lead to the cooperative engagement of other transcription factors and nucleosome remodelers. Despite significant interest in understanding the interplay between pioneer transcription factors and chromatin remodeling complexes, *in vivo* genomic studies lack the resolution to clearly define how each of these factors individually contribute to initial chromatin opening. The fundamental mechanism by which nucleosome remodeling enzymes initially access compacted chromatin in a site-specific manner had yet to be answered. Can nucleosome remodelers target any local open domain in chromatin, or must they be guided by a pioneer transcription factor?

To directly answer these fundamental questions, I took an *in vitro* approach, which required the generation of a 13X nucleosome array with a central natural nucleosomal "Cx3cr1" sequence targeted by PU.1, C/EBPα, and C/EBPβ both *in vivo* and *in vitro* (Fernandez Garcia et al., 2019). We are the first to use long-read Nanopore sequencing to map nucleosome positions at base-pair resolution within *in vitro* generated nucleosome arrays. We further compacted these "Cx3cr1" nucleosome arrays with linker histone H1, to reconstitute a naïve, compacted chromatin state as found *in vivo*. Using long-read Nanopore sequencing, we determined how individual, and combinations of, purified pioneer factors expose targeted nucleosomes within H1-compacted chromatin *in vitro*. Using this method, we discovered that different classes of pioneer transcription factors can expose targeted nucleosomes in compacted chromatin in different structural ways, reflected by hypersensitivity patterns. We found that PU.1 and C/EBPα together led to a synergistic increase in DNase I hypersensitive signal in H1-compacted nucleosome arrays, indicating that the two pioneer factors cooperate to locally open chromatin. Thus, different pioneer factors elicit different opening patterns on H1-compacted chromatin, in a target site-dependent manner.

Further, we determined that local chromatin opening by PU.1 is sufficient to enable the canonical mammalian ATP-dependent SWI/SNF complex, cBAF, within compacted chromatin. In contrast, the cBAF complex is inactive in the presence of C/EBPα, despite C/EBPα inducing local chromatin opening. Through a deletion analysis, we identified an intrinsically disordered region (IDR) of PU.1 which is, but required for PU.1 to bind H1-compacted chromatin and initiate accessibility, but is dispensable for nucleosome binding. Mere chromatin opening by the PU.1 **Δ**AQ truncation, consisting of the IDR and DNA-binding domains, was not sufficient to enable BAF activity, which requires additional domains within the N-terminus of PU.1. Thus, we found that the chromatin remodeler cBAF needs specific engagement with a transcription factor before it is able to alter the chromatin state. This explains how the remodeling complex would be restricted, *in vivo*, from acting at any local open region. Based on all of the above, our chromatin reconstitution studies provide a mechanistic understanding of how pioneer factors and nucleosome remodeling complexes interact to initiate new gene regulatory events within compacted chromatin.

3.2 A hierarchical model of pioneer transcription factor chromatin binding and nucleosome remodeler recruitment

My thesis research builds upon previous data from our lab, which found that PU.1 is a strong nucleosome binder, while C/EBPα shows a weaker nucleosome affinity (Fernandez Garcia et al., 2019). During B cell reprogramming, knockdown of PU.1 results in a decreased C/EBPα binding to pre-existing and de novo enhancers, suggesting that in the absence of PU.1, C/EBPα targeting to closed chromatin might be driven by cooperative interactions (Van Oevelen et al., 2015b). Thus, the study revealed that nucleosome binding positively correlates with reprogramming potential and suggested a hierarchical model where diverse pioneer transcription factors with different chromatin binding potential interact to drive new cell fates. Consistent with this model, I found that PU.1, C/EBPα, and C/EBPβ have the intrinsic ability to initiate accessibility of linker histone H1-compacted chromatin *in vitro* with different efficiencies: PU.1 elicits stronger DNase I hypersensitivity compared to C/EBPα and C/EBPβ. Despite both factors initiating local chromatin opening, only chromatin opening by PU.1 enabled further cBAF remodeling. These data suggest an additional component to this hierarchy model, where pioneer transcription factors with greater reprogramming potential may, in addition to more efficient nucleosome binding and local chromatin opening, more efficiently enable cooperative actions that lead to further chromatin opening.

A previous *in vitro* study found that the purified Drosophila pioneer factor GAF could disrupt the structure of chromatin reconstituted with Drosophila embryo extracts in an ATP-dependent manner, indicating that GAF acted in conjunction with ATP-dependent chromatin remodelers to initiate chromatin opening (Tsukiyama et al., 1994). Consistent with other *in vivo* studies (Chronis et al., 2017), it has been shown that remodeler activity was necessary to maintain accessibility at GAF targets *in vivo* (Judd et al., 2021), though the initial, stable chromatin binding of GAF is independent of the presence of nucleosome remodelers *in vivo* (Tang et al., 2022). Together, the authors proposed that remodeler recruitment by GAF and other transcription factors is a fundamental biochemical criterion for a pioneer transcription factor. The data I present in Chapter 2 clearly support the concept that pioneer transcription factors enable site-specific nucleosome remodeling activity. However, the data further support a model that diverse pioneer transcription factors initiate local accessibility within compacted chromatin *in vitro* independently of ATP-dependent chromatin remodelers. Further, CEBPα can bind nucleosomes and locally open chromatin but could not enable cBAF activity. Instead, CEBPα could further cooperatively open chromatin with PU.1. In cell reprogramming and development, different pioneer factors colocalize and it has been observed that co-binding of transcription factors to closed chromatin is required for subsequent chromatin opening *in vivo* (Cernilogar et al., 2019; Mayran et al., 2019).

I postulate that that enabling cooperative chromatin opening, and thus synergistically increasing chromatin accessibility, is an important component of pioneer transcription factor activity *in vivo*. Different pioneer transcription factors likely enable a variety of cooperative interactions within compacted chromatin, and do not necessarily enable nucleosome remodelers. This hypothesis could be tested *in vivo* using a model where pre-B cells, already expressing PU.1, are reprogrammed to macrophages with inducible expression of C/EBPα (Bussmann et al., 2009; Di Stefano et al., 2014a; Van Oevelen et al., 2015a; Xie et al., 2004b). During pre-B cell to macrophage reprogramming, C/EBPα engages pre-existing enhancers bound by PU.1, as well as a subset of *de novo* enhancers that are initially targeted by C/EBPα, followed by binding of PU.1. It would be interesting to assess the dynamics of transcription factor and BAF recruitment at these *de novo* enhancers by performing a time course Chip-seq experiment. Based on the data presented in Chapter 2, I expect BAF to be absent at timepoints when only C/EBPα is bound to de novo enhancers, and to be present at *de novo* enhancers once PU.1 is bound.

3.3 Investigating the role of non-DNA binding pioneer transcription factor domains in facilitating interactions with compacted chromatin

Whereas intense focus has been on the DNA-binding domains of transcription factors, much less is known about the intrinsically disordered domains in flanking regions and their contributions to the general mechanism of pioneer transcription factor chromatin binding and opening. Further, while domains outside of the DNA-binding domain are clearly important to interact with and recruit nucleosome remodelers (Boulay et al., 2017; Minderjahn et al., 2020; Wang et al., 2020), there is growing evidence that transcription factors interact with histones directly through specific domains (Cirillo et al., 2002; Echigoya et al., 2020) that are have critical roles in initiating cell-type specific gene expression (Fuglerud et al., 2017b; Iwafuchi et al., 2020). Advancement in structural biology techniques has led an increase in the number of mononucleosome (Roberts et al., 2021; Takizawa et al., 2018) and transcription factor bound-mononucleosome structures (Dodonova et al., 2020; Michael et al., 2020; Tanaka et al., 2020) and has provided new insight into our understanding of transcription factor-chromatin interactions ay high resolution. However, such structural studies are currently limited to the structured DNA-binding domains of pioneer transcription factors and there are no published structures of full-length sequence-specific transcription factors bound to mononucleosomes or more complex chromatin substrates. This has limited our understanding of the biological functions of transcription factors, as there is now ample evidence that the DNA-binding domains of transcription factors are insufficient to engage higher-order chromatin structures.

In the case of PU.1, we identified that the IDR was required for both binding and local chromatin opening of H1-compacted arrays. Previously, the C-terminal domain of FoxA was found to be required both for histone interactions and chromatin opening. This suggests that local chromatin opening and binding may require both protein-DNA interactions, mediated by pioneer transcription factor DNA-binding domains, and protein-histone interactions, mediated by pioneer factor domains outside of the DBD. Indeed, other pioneer transcription factors have been shown to interact with histones: mass spectrometry indicated histones as an interacting partner of the pluripotency transcription factor Oct4 (Pardo et al., 2010), and the hematopoietic transcription factor c-Myb was shown to interact with purified core histones in pull-down assays using whole cell lysates (Fuglerud et al., 2017b). Thus, the presence of histone binding domains within the protein structures of PU.1, C/EBPα, and C/EBPβ might explain the unique ability of pioneer transcription factors to localize to closed chromatin, while non-pioneer transcription factors do not. Thus, I hypothesize that PU.1, C/EBPα and C/EBPβ possess histone interaction domains, which are required for efficient nucleosome binding and chromatin opening. This would require incubating purified PU.1, C/EBPα, or C/EBPβ with Cx3cr1 mononucleosomes in the presence of formaldehyde as a crosslinker and performing crosslinking mass spectrometry (Sinz, 2018; Sutherland et al., 2008). Purified FoxA and the Alb1 nucleosome can be used as a positive control for the formaldehyde crosslinking (Iwafuchi et al., 2020). The resulting crosslinked material could be concentrated, run on a SDS-PAGE gel and visualized by Coomassie staining. Higher molecular weight bands, compared to cross-linked mononucleosomes alone as a negative control, are indicative of histone-transcription factor crosslinks and will be cut from the gel and trypsin digested. The digested bands could then be analyzed by mass spectrometry to identify cross-linked peptides and linked amino acid residues. Finally, any transcription factor residues that crosslink to core histone could be mutagenized and tested for impairing chromatin binding and opening.

Different mechanisms have been suggested for the action of pioneer transcription factors to induce local chromatin changes in DNA accessibility. For example, comparison of FoxA with linker histone showed that the FoxA DBD contains a winged-helix domain resembling linker histone. Of course, linker histone and FoxA execute opposing functions, as linker histone compacts nucleosomes while FoxA induces local chromatin opening (Cirillo et al., 2002; Osipova et al., 1980). This could occur by FoxA displacing linker histone (Cirillo et al., 1998; Iwafuchi-Doi et al., 2016b) possibly due to differences in the charges of the flanking the globular domain of linker histone and the DBD of FoxA which are predominately positively and negatively charged, respectively. However, binding of the Gal4 DNA-binding domain to a Windom 601 chromatosome induced partial unwrapping by causing the linker histone C-terminal domain to release the linker DNA, but without evicting linker histone from the nucleosome dyad (Burge et al., 2022). Thus, it’s also possible to disrupt chromatosome structure, without completely evicting linker histone. An immediate next step of my project is to assess whether linker histone is being evicted by PU.1, C/EBPα or C/EBPβ. This can be accomplished through an *in vitro* linker histone H1 ChIP assay followed by qPCR to assess specific loss of linker histone at the Cx3cr1 nucleosome after incubating the different transcription factors with H1-compacted nucleosome arrays followed by MNase digestion.

A compelling hypothesis, reinforced by the commonly observed promiscuity of disordered interactions, is that the disordered regions in pioneer transcription factors interact with the linker histone to facilitate its eviction. Recently, several studies were published using single molecule florescence microscopy to assess the *in vitro* binding kinetics of linker histone H1. These studies found that a histone chaperone, prothymosin α, forms a highly disordered complex with linker histone H1 (Borgia et al., 2018; Sottini et al., 2020). Heidarson et al. found that, by forming a transient ternary complex via electrostatic interactions with the positively charged disordered domains of H1, the highly negatively charged and disordered histone chaperone could efficiently invade the chromatosome and accelerate linker histone H1 dissociation from the nucleosome by a competitive substitution mechanism (Heidarsson et al., 2022). Given the great abundance of highly negatively charged disordered domains in nuclear proteins, including transcription factors like PU.1 and FoxA, it is of great interest to explore if transcription factors are able to form such complexes with linker histone to facilitate its eviction through a similar competitive substitution mechanism (Cirillo et al., 1998; Iwafuchi-Doi et al., 2016b). The specific experiment would be to measure the potential binding kinetics of linker histone and PU.1 WT and PU.1 deletions to determine if PU.1 can bind to linker histone and form a ternary complex (Borgia et al., 2018; Sottini et al., 2020) and if any potential binding is mediated by the IDR. This would require reconstitution of Cx3cr1 mononucleosome with extended linker DNA to form Cx3cr1 chromatosomes to determine if PU.1 can facilitate H1 dissociation from the nucleosome.

3.4 Reconstitution of complex heterochromatin substrates

Distinct forms of silent chromatin repress genes by different mechanisms: naïve domains lack apparent histone modifications and are compacted by linker histone, while heterochromatin domains that are marked by repressive histone modifications H3K9me3 or H3K27me3 and associated with additional heterochromatin associated proteins. For example, the heterochromatin associated protein, HP1, possesses chromodomains that enable specific binding to H3K9me3 modified chromatin. Pioneer transcription factors do not bind all of their potential DNA motifs throughout the genome, and their binding may be prevented by heterochromatin. For example, during reprogramming of human fibroblasts to pluripotency stem cells, the pioneer factors Oct4, Sox2, and Klf4 target silent pluripotency genes in naïve domains, but not in H3K9me3-marked heterochromatin (Soufi et al., 2012). In addition, various studies have indicated that diverse pioneer transcription factors may engage with one form of heterochromatin over another (Boller et al., 2016b; Johnson et al., 2018; Mayran et al., 2018). While I have shown that pioneer transcription factors can engage H1-compacted structures, the components of heterochromatin that selectively exclude pioneer transcription factors are unknown.

In vitro studies have defined how dynamics in the conformation of the HP1 chromodomain govern the ability to oligomerize on chromatin and promote chromatin compaction of H3K9me3-marked nucleosome arrays (Bryan et al., 2017; Kilic et al., 2018, 2015). However, there is a lack of robust biochemical tools to study both how these heterochromatin structures are established and how pioneer factors interact with different heterochromatin structures. To improve our understanding, it will be important to develop more complex heterochromatin structures *in vitro* and determine which heterochromatin associated proteins are responsible for the observed inability of pioneer factors to engage H3K9me3 marked domains of the genome. My studies provide a platform to systematically test the ability of different pioneer factors to interact with different forms of heterochromatin. This could be accomplished by purifying histone octamers containing H3k9me3 modifications to assemble nucleosome arrays, as well as adding HP1 and other heterochromatin associated proteins to determine if they are sufficient to impair pioneer factor activity. This project will set the groundwork for studying pioneer factor heterochromatin interactions *in vitro* and establish models for testing functionality *in vivo.*

APPENDIX A: CONSTRUCTION OF NUCLEOSOME ARRAY TEMPLATES CONTAINING NATURAL NUCLEOSOME SEQUENCES TARGETED BY DIVERSE REPROGRAMMING FACTORS

This appendix contains unpublished details of the design and construction of nucleosome arrays containing a central NRCAM, Cx3cr1 (Fernandez Garcia et al., 2019), or LIN28B mononucleosome (Soufi et al., 2015) sequences, as well as design as NRCAM, LIN28B and Cx3cr1 trinucleosome sequences. The constructs were prepared by me with guidance from Dr. Zaret.

I. Objective

A subset of eukaryotic transcription factors (TFs) possesses the ability to reprogram terminally differentiated cells into another cell type. Ectopic expression of four transcription factors, Oct4, Sox2, Klf4 and cMyc, was demonstrated to reprogram human fibroblasts into induced pluripotent stem cells (Takahashi and Yamanaka, 2006). Subsequently, additional cocktails of specific TFs have been demonstrated to drive cell-type conversions in fibroblasts (Figure A-1), including: FoxA, Gata4 and Hnf1a for hepatocyte transdifferentiation (Huang et al., 2011; Sekiya and Suzuki, 2011), PU.1 and C/EBPα/β for macrophage transdifferentiation (Di Stefano et al., 2014b; Feng et al., 2008) and Ascl1, Mytl1 and Brn2 for neuronal cell transdifferentiation (Vierbuchen et al., 2010; Wapinski et al., 2013). Genetic and genomic studies have shown that several of these TFs act as “pioneer” factors, meaning they are able to target developmentally silent genes within nucleosome dense chromatin to initiate changes in gene expression profiles during cellular reprogramming (Heinz et al., 2010; McPherson et al., 1993; Soufi et al., 2015; Wapinski et al., 2013).

Previously, Soufi et al. and Fernandez Garcia et al. identified natural nucleosomal sequences containing motif-enriched sequences targeted by pioneer factors *in vivo* (Figure A1) and used these sequences to reconstitute nucleosomes *in vitro* to test the ability of the different transcription factors to directly bind nucleosomes. Purified recombinant Oct4, Sox2, Klf4, Ascl1/E12α, PU.1 and C/EBPα bound *in vitro* generated mononucleosomes with a range of nucleosome binding affinities that correlated with their reprogramming potential (Fernandez Garcia et al., 2019; Soufi et al., 2015). The goal of this study is to design and construct 6 different plasmids containing the LIN28B, NRCAM, and Cxcr1 nucleosome sequences (Figure A-2a). Generating nucleosome arrays containing the targeted nucleosome sequences will allow study of how these diverse transcription factors interact with condensed chromatin *in vitro.*

II. Reagents

DNA Sequences and Primers

LIN28B mononucleosome insert sequence (Plasmids Name: p208-10-LIN28B-MN)

hg18 genomic coordinates: chr6:105637973-105638197:

CAGTGTTTTACATCAGTTGGAGCATAAGTTAAGTGGTATTAACATATCCT

CAGTGGTGAGTATTAACATGGAACTTACTCCAACAATACAGATGCTGAAT

AAATGTAGTCTAAGTGAAGGAAGAAGGAAAGGTGGGAGCTGCCATCACTC

AGAATTGTCCAGCAGGGATTGTGCAAGCTTGTGAATAAAGACACATACTT

CATGTAGTCAGAAGAGTGGTCTGAA

Primers to amplify LIN28B mononucleosome insert sequence	
LIN28B-mononuc. Forward BfuAI site	CGCGCGACCTGCGTTACAGCATCCACTCAG TATCCCTTGACAGTGTTTTACATCAGTTGG
LIN28B-Trinuc. Reverse XbaI site	GCGTCTAGACATACTAAACTTCAGACCACT CTTCTGACT

LIN28B Trinucleosome insert sequence

hg18 genomic coordinates: chr6:105637797-105638373

TATATTCTCATTTCATGCCTAATGTCTTATTCTGTCAATTATGGATATGT

TGAGGTTTAAAAAAATTACTTGATTAAAAATAAAACATATAACGTTGGCA

TTTATGATGGGTTTTGGAGATTTTTTATGATGTGTGTTCTTACTAAGTTA

AGTGAAAGGATAAAAGGCCATTTAGGCAGTGTTTTACATCAGTTGGAGCA

TAAGTTAAGTGGTATTAACATATCCTCAGTGGTGAGTATTAACATGGAAC

TTACTCCAACAATACAGATGCTGAATAAATGTAGTCTAAGTGAAGGAAGA

AGGAAAGGTGGGAGCTGCCATCACTCAGAATTGTCCAGCAGGGATTGTGC

AAGCTTGTGAATAAAGACACATACTTCATGTAGTCAGAAGAGTGGTCTGA

AGCAAAATATTCAAAGCCTATTGAAGAAATCATTTTAGAATTTTTCTACA

AGTTTTTTCTAATAAAGAATTTGAAAAGGGCAAATTTTATCAGTACCAAA

TATCTAAAATACGAATGTAGATTTATCAAACCATGAATGCAGTATTGATG

TACTGTAACAAACGATTTGGTAAAACT

Primers to amplify LIN28B Trinucleosome insert sequence (Plasmids Name: p208-10-LIN28B-TN)

LIN28B-Trinuc. Forward NheI site	GCAGCTAGCGCCCAGAGAAAGCTCAAGGTT
LIN28B-Trinuc. Reverse SpeI site	GCAACTAGTCTAAGTCACTCTCCATCCCTGG

NRCAM monoucleosome insert sequence (Plasmids Name: p208-10-NRCAM-MN)

mm9 genomic coordinates: chr12:45445248-45445472

CCAACAGATCCCCCATCAAGGAGTGGCACAGTATCAATTACTTCTGAAAC

AGATGACTCCCAGCAGCTGCTGCCTGTGGCCCACAGGGCTTCCTGCCCTG

CATGACAGCTGCACATCACATCCTGTGGTCATACTACTTCAGCCGCTTCT

ACGGCCAGATACAAAAGTGGGTGGGGAACATAGGCAAGCCTACCAGTTCC

AAGGTGTTATAAATCCCAGACGCAT

Primers to amplify NRCAM mononucleosome insert sequence

NRCAM-mononuc. Forward BfuAI site	GCGCGACCTGCGTTACAGCATCCACTCAG TATCCCTTGACCAACAGATCCCCCATCAAG
MRCAM-mononuc. Reverse XbaI site	GCGTCTAGACATACTAAACATGCGTCTGG GATTTATAAC

NRCAM trinucleosome insert sequence. (Plasmid Name: p208-10-NRCAM-TN)

mm9 genomic coordinates: chr12:45445072-45445648

ATTTACCACGGACAACCGGTTATTACTGAGGAGAGTGGAAATAAGCTCTG

TCGTGGCTGCCTCTGTCTGCTCTGGCCATATTCTAAGGGTGGTTTGAATC

CTTCTCCCTCTCTTGTCTCAACCATACAGTTAGGAGCAGGGCCAGTCATG

AATATGCATGCGTCACACACCTGTGACCAACAGATCCCCCATCAAGGAGT

GGCACAGTATCAATTACTTCTGAAACAGATGACTCCCAGCAGCTGCTGCC

TGTGGCCCACAGGGCTTCCTGCCCTGCATGACAGCTGCACATCACATCCT

GTGGTCATACTACTTCAGCCGCTTCTACGGCCAGATACAAAAGTGGGTGG

GGAACATAGGCAAGCCTACCAGTTCCAAGGTGTTATAAATCCCAGACGCA

TTTGTTATTTGCCATAGGAGTACCTTCTGTGATGTCTCTTTTCATCTTCA

CCAAACAAATGTGTCAGAAGCTGGCCCTTGCTTTTTGTTGTTGCTGTGTG

TTTCTTGTTTTGTTTTGATTTTGATTTTGTTTGTTTGTTTTCTGAAAAGG

GACAGATAATGTGATAAGCTTTATAAG

Primers to amplify NRCAM trinucleosome insert sequence

NRCAM-trinuc. Forward NheI site	GCAGCTAGCATTTACCACGGACAACCGGT
NRCAM-trinuc. Reverse SpeI site	GCAACTAGTCTTATAAAGCTTATCACATTA TCTGTC

Cx3cr1 monoucleosome insert sequence (Plasmid Name: p208-10-Cx3cr1-MN)

mm9 genomic coordinates: chr9:119946575-119946799

GCCCCGCGTGCAAAGCCCAGGAGCCCTTGCTTAGGTGCAGGGCCTCTCGG

CTGCTGATCTTCAGCTGGTTGCTGAGAGTTGCAGCATTGCTGAGTCTTAG

CAATGGATACTTCCCGATTCCCCTCACAAAAATAGGTCAGTCTGTCTGGC

TAGTTCTGTACTTGCAGACACAGGGCATGTGGGGTTCCTATTTTTCTAGC

TCCCAGGCTTCTGTCTGCTTCCTTC

Primers to amplify Cx3cr1 mononucleosome insert sequence

Cx3cr1-mononuc. Forward BfuAI site	GCGCGACCTGCGTTACAGCATCCACTCAGT ATCCCTTGAGCCCCGCGTGCAAAGCCCAG

Cx3cr1-mononuc. Reverse XbaI site	GCGTCTAGACATACTAAACGAAGGAAGCAG ACAGAAGCCTGG

Cx3cr1 Trinucleosome insert sequence (Plasmid Name: p208-10-Cx3cr1-TN)

mm9 genomic coordinates: chr9:119946399-119946975

GCCCAGAGAAAGCTCAAGGTTCCTTCTCACTAGGGGGGAGAAAGAGGTCA

TTACATTGTAGGTGCTGGATGAGGAGGTCCTATGAGGAGTTTTCCCATAT

GGCTGCCTCAGGGAGAAACAAGGTCCTGGGGAAGTCTGGATTCCTGTTTT

CCCGGCAGCCTGAAGAATCTAGAGTCGCCCCGCGTGCAAAGCCCAGGAGC

CCTTGCTTAGGTGCAGGGCCTCTCGGCTGCTGATCTTCAGCTGGTTGCTG

AGAGTTGCAGCATTGCTGAGTCTTAGCAATGGATACTTCCCGATTCCCCT

CACAAAAATAGGTCAGTCTGTCTGGCTAGTTCTGTACTTGCAGACACAGG

GCATGTGGGGTTCCTATTTTTCTAGCTCCCAGGCTTCTGTCTGCTTCCTT

CATGGAGAACTGCTTGCTAAGAATTTTACCACAGTGGGAGCTTAAGAAAC

AGCAAGGCTTAAAAATAAACCTCTAAGATGGCCAGTAACACTGGGTACGA

GTGTGTCTAGTGTGTCTCTATTCAACCCAGTGCTTGAGAGGCAGAGGCAA

GAGGCCCAGGGATGGAGAGTGACTTAG

Primers to amplify Cx3cr1 trinucleosome insert sequence

Cx3cr1-trinuc. Forward <u>NheI site</u>	GCA<u>GCTAGC</u>GCCCAGAGAAAGCTCAAGGTT
Cx3cr1-trinuc. Reverse <u>SpeI site</u>	GCA<u>ACTAGT</u>CTAAGTCACTCTCCATCCCTGG

III. Protocol

Goal

To study how these factors interact with high order chromatin structures, I used the Cx3cr1, NRCAM, and LIN28B mononucleosomes sequences to generate 13X nucleosome array sequences containing a central Cx3cr1, NRCAM, and LIN28B mononucleosomes or trinucleosome sequence (Figure A-2).

Approach

I note that Soufi et al. generated the 162 bp LIN28B nucleosomal sequence by PCR (Soufi et al., 2015). Fernandez Garcia et al generated the 162 bp NRCAM and Cx3cr1 sequences by subcloning 152bp of endogenous sequences with BamHI restriction sites into pUC19 plasmids (Fernandez Garcia et al., 2019). Thus, the endogenous LIN28B sequence is 162 bp and the NRCAM and Cx3cr1 sequences are 152 bp.

To prepare the Cx3cr1 and NRCAM mononucleosome inserts, the 152bp sequences were extended on the 5' side by 36 bp and on the 3' side by 37 bp (31 bp and 32 bp for the LIN28B sequence, respectively) to incorporate linker DNA and BfuAI and XbaI flanking sequences (Figure A-2). To prepare the Cx3cr1 and NRCAM trinucleosome inserts, the 152bp sequences were extended on the 5' end by 212 bp and on the 3' end by 213 bp (207 bp and 208 bp for the LIN28B sequence, respectively) to incorporate linker DNA and BfuAI and XbaI flanking sequences (Figure A-2) to accommodate two additional nucleosomes and NheI and SpeI flanking sequences. The DNA sequences were generated by PCR of genomic mouse DNA for NRCAM and Cx3cr1 and genomic human DNA for LIN28B with BfuAI/XbaI or NheI/SpeI restriction sites cloned into the p208-10.N1N2 (Figure A-2a, "Alb1 Array").

IV. Conclusions

Six constructs were generated in total: p208-10-Cx3cr1-MN, p208-10-NRCAM-MN, p208-10-LIN28B-MN, p208-10-Cx3cr1-TN, p208-10-NRCAM-TN, and p208-10-LIN28B-TN. All plasmids were verified by sanger sequencing and restriction digestion.

V. Figures

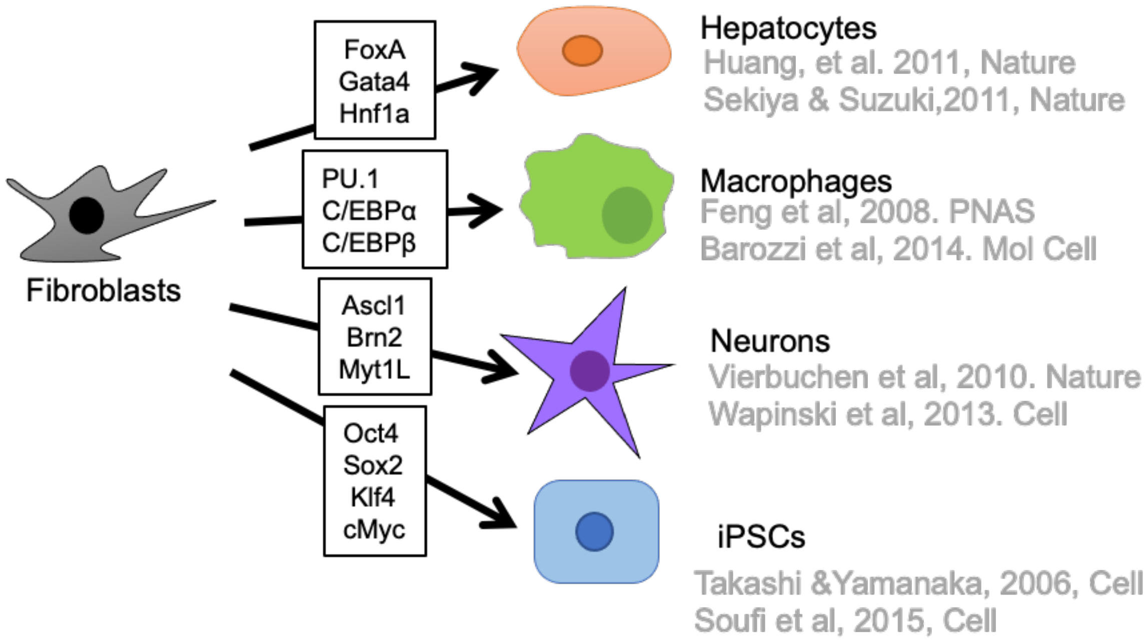


Figure A - 1 Transcription factor mediated cellular reprogramming and transdifferentiation.

Forced expression of pluripotency transcription factors mediate reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs). Likewise, fibroblast can be driven to a new somatic cell state by forced expression of different cocktails of lineage specific transcription factors in a process referred to a transdifferentiation.

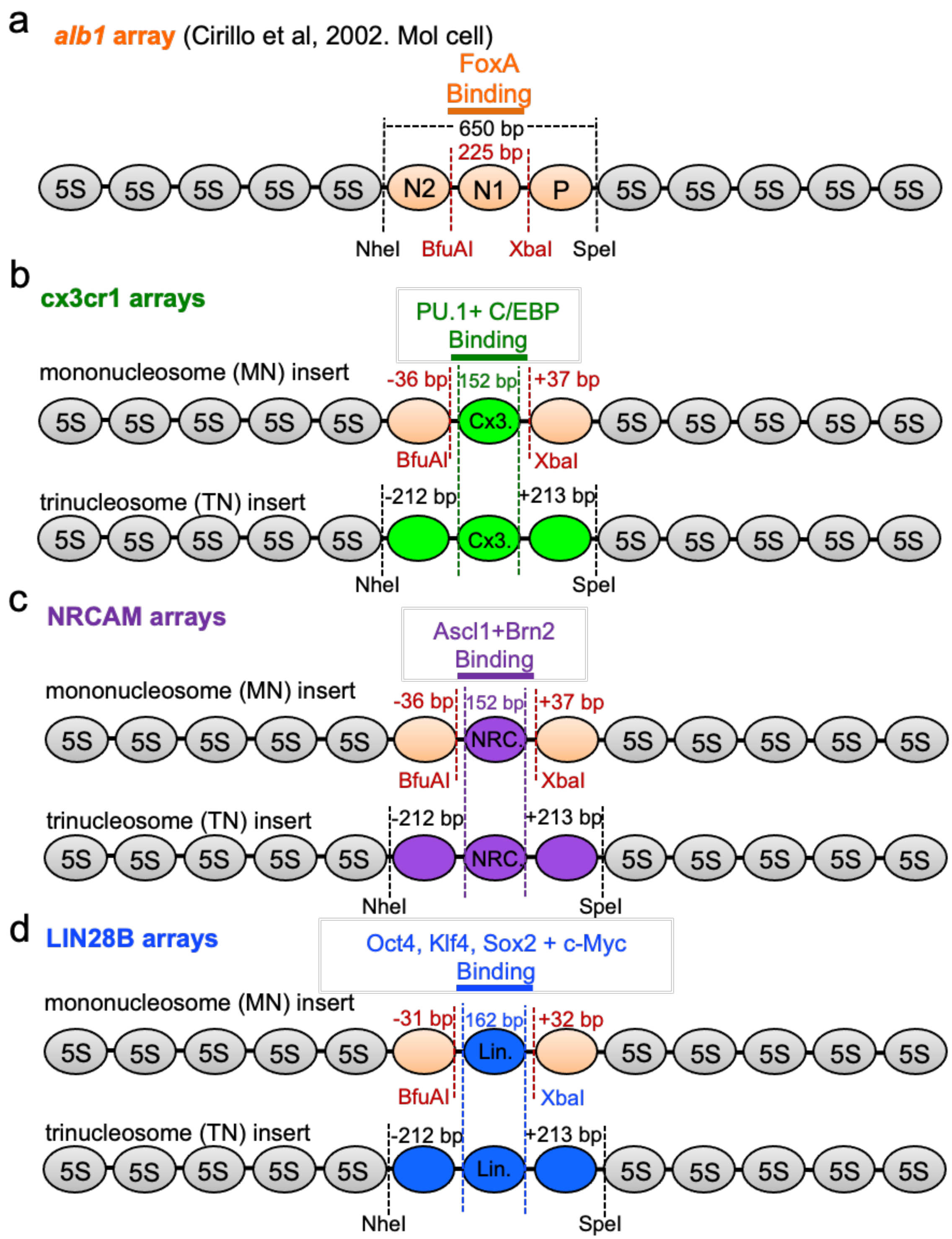


Figure A - 2 Design of nucleosome array sequence constructs.

a, Diagram of the Alb1 nucleosome array encoded by the p208-10.N1N2 plasmids with relevant restriction sites indicated.. The central “N1” nucleosome possesses FoxA and Gata4 binding sites. b-d, example diagram of the Cx3cr1 (b), NRCAM (c), and LIN28B (d) array constructs. For each mononucleosome sequence, two array plasmids were generated containing a mononucleosomes (MN) or trinucleosome (TN) insert. To prepare the Cx3cr1 and NRCAM mononucleosome inserts, the 152bp sequences were extended on the 5’ side by 36 bp and on the 3’ side by 37 bp (31 bp and 32 bp for the LIN28B sequence, respectively) to incorporate linker DNA and BfuAI and XbaI flanking sequences (Figure A2c). To prepare the Cx3cr1 and NRCAM trinucleosome inserts, the 152bp sequences were extended on the 5’ end by 212 bp and on the 3’ end by 213 bp (207 bp and 208 bp for the LIN28B sequence, respectively) to incorporate linker DNA and BfuAI and XbaI flanking sequences to accommodate two additional nucleosomes and NheI and SpeI flanking sequences.

APPENDIX B: PU.1 AND C/EBPA HYBRID TRANSCRIPTION FACTORS

This appendix contains unpublished analyses which compare the ability of WT PU.1, WT C/EBPα, and N-terminal swap hybrid transcription factors to bind short dsDNA probes and locally open H1-compacted nucleosome arrays. All experiments were performed by me under the guidance of Ken Zaret.

I. Objective

I have found that DBDs of PU.1 and CEBPα are insufficient for efficient chromatin opening and hypothesize that PU.1 and C/EBPα possess specific domains that enable chromatin opening. Furthermore, I determined that changing the underlying DNA sequence did not change the different patterns of chromatin opening observed. Thus, I hypothesize that the different patterns of chromatin opening are due to intrinsic properties of the non-DNA binding domains of PU.1 and C/EBPα. I tested how non-DBD regions of pioneer transcription factors control chromatin opening by swapping the N-terminal of PU.1 and C/EBPα and testing their ability to open compacted chromatin *in vitro*. I will assess if the hybrid TFs elicit similar or different chromatin opening patterns compared to their wild-type counterparts.

II. Reagents

Primers

Primers to amplify CEBPα N-terminal domain (residues 1-271)	
Forward Primer	CGCGCGGCAGCCATA**GAGTCGGGCCGACTTCTACGAGG**
Reverse Primer	TCCCCGTGCAGCATA**ACCGGCACCGCTGCCACC**

Primers to amplify PU.1 N-terminal domain (residues 1-159)	
Forward Primer	cgcgcggcagccaTA**TTACAGGCGTGCAAAATGGAAG**
Reverse Primer	TCCCCGTGCAGCATA**AAGACCTGGCCAGGCTC**

Primers for SDM Ca N-term- PU.1 DBD	
Forward Primer	CTGCACGGGGAGACAGGCAG
Reverse Primer	ACCGGCACCGCTGCCACC

Primers for SDM PU.1 N-term- Ca DBD	
Forward Primer	GCGGGCAAAGCCAAGAAG
Reverse Primer	AAGACCTGGCCCAGGCTC

dsDNA probes

(Fernandez Garcia et al., 2019)

“Igk (2-4)” 23 bp dsDNA with specific PU.1 motif, but no specific C/EBP motif

GAATAAAAGGAAGTGAAACC

"synt CEBP" dsDNA with specific <u>C/EBP motif</u>, but no specific PU.1 motif

GAACTGGAT<u>TGCGCAA</u>TAG

Plasmids

pET-28b-PU.1-DBD (encodes PU.1-DBD, residues 160-272)

pET-28b-C/EBPα-DBD (encodes C/EBPα-DBD, residues 272-359)

Cloning Kit

in-fusion cloning kit (Takara Bio cat # 638947)

III. Results

Goal

To construct PU.1 and C/EBPα N-terminal swap plasmids, purify proteins, and subsequently test their ability to specifically bind to short dsDNA compared to WT PU.1 and C/EBPα. Determine if the hybrid TFs elicit similar or different chromatin opening patterns compared to their wild-type counterparts.

Approach

The IDR of PU.1 (Figure B1a) is required for PU.1 to locally open H1-compacted chromatin. Using the same disorder prediction algorithm, MobidDB (Piovesan et al., 2021), I found the N-terminal domain of C/EBPα contains three intrinsically disorder regions (Figure B1a). The N-terminal domain or PU.1 and C/EBPα were inserted by in-fusion cloning (Takara Bio cat # 638947) into NdeI linearized pET-28b-PU.1-DBD (encodes PU.1-DBD, residues 160-272) or pET-28b-C/EBPα-DBD (encodes C/EBPα-DBD, residues 272-359) vectors, respectively, followed by site directed mutagenesis to join the n-terminal domain segments to the DNA-binding domain segment (NEB, cat # E0554S) (Figure B1b). Recombinant proteins were purified (Figure B1c) from E. coli under denaturing conditions as described in Chapter 3. Purified recombinant proteins were tested for their ability to bind short dsDNA probes with or without their specific binding motif by electromobility shift assay (Figure B1d-g). We next assessed the ability of full-length, DBD only, and N-terminal swap TFs to initiate nucleosome exposure within H1-compacted arrays.

Results

The hybrid transcription factors bind short dsDNA with specificity and similar affinity compared to their wild-type counterparts. Strikingly, the C/EBPα-ETS hybrid TF and wildtype PU.1 exhibit identical intensities and patterns of DNase I hypersensitivity.

IV. Conclusions

Compared to their full-length counterparts, the DBDs of PU.1 and C/EBPα possess reduced chromatin opening abilities compared to their wild-type counter parts, indicating that domains outside of the DBD play a role in chromatin opening. Strikingly, the C/EBPα-ETS hybrid TF and wildtype PU.1 exhibit identical patterns of DNase I hypersensitivity. While, the DNA-binding domain determine the pattern of nucleosome exposure observed, non-DNA binding domains are required for targeting and exposure of individual nucleosome within compacted chromatin.

As discussed in Chapter 1, the DNA-binding domains of various pioneer transcription factors are sufficient to bind nucleosomes and elicit structural changes (Dodonova et al., 2020; Michael et al., 2020; Zhu et al., 2018). The results presented in this appendix underscore the role of the DNA-binding domain in determining the pattern of nucleosome exposure. However, as shown in Chapter 2, the DNA-binding domains of PU.1 and C/EBPα could not generate DNase I hypersensitive sites in H1-compacted chromatin like their full-length counterparts. Together, these results suggest that the non-DNA-binding domains are important to enable nucleosome targeting by the DNA-binding domains within H1-compacted chromatin. As discussed in Chapter 3, this could possibly be mediated by intrinsically disordered domains, which I hypothesize mediate interactions with core histones and/or linker histone. Indeed, the IDR of PU.1 was required for chromatin opening.

Intriguingly, I do not observe chromatin opening by the PU.1-bZip mutant which consists of the C/EBPα DNA-binding domain tethered to the N-terminal non-DNA binding domain of PU.1. This indicates, that unlike the PU.1 DNA-binding domains, the PU.1 IDR is insufficient to enable the bZIP DNA-binding domain to elicit chromatin opening. Comparing lanes 4 and 6 from Figure B-2, I observe that C/EBPα-ETS opens chromatin more potently than PU.1. Thus, I postulate that the N-terminal of C/EBPα is more efficient at targeting the DNA-binding domain of PU.1 than the N-terminal of PU.1. Perhaps the N-terminal domain of C/EBPα, which is predicted to contain more overall disorder compared to the N-terminal domain of PU.1 (Figure B-1a), has a greater affinity for histones. As previously published, C/EBPα is less efficient at nucleosome binding compared to PU.1 and other pioneer transcription factors due to the structure of its DNA-binding domain (Fernandez Garcia et al., 2019). However, C/EBPα is able to elicit hypersensitivity within H1-compacted nucleosome arrays. Therefore, having a non-DNA-binding domain that is efficient at targeting H1-compacted arrays is likely crucial for a weaker nucleosome binding pioneer transcription factor to elicit chromatin opening.

V. Figures

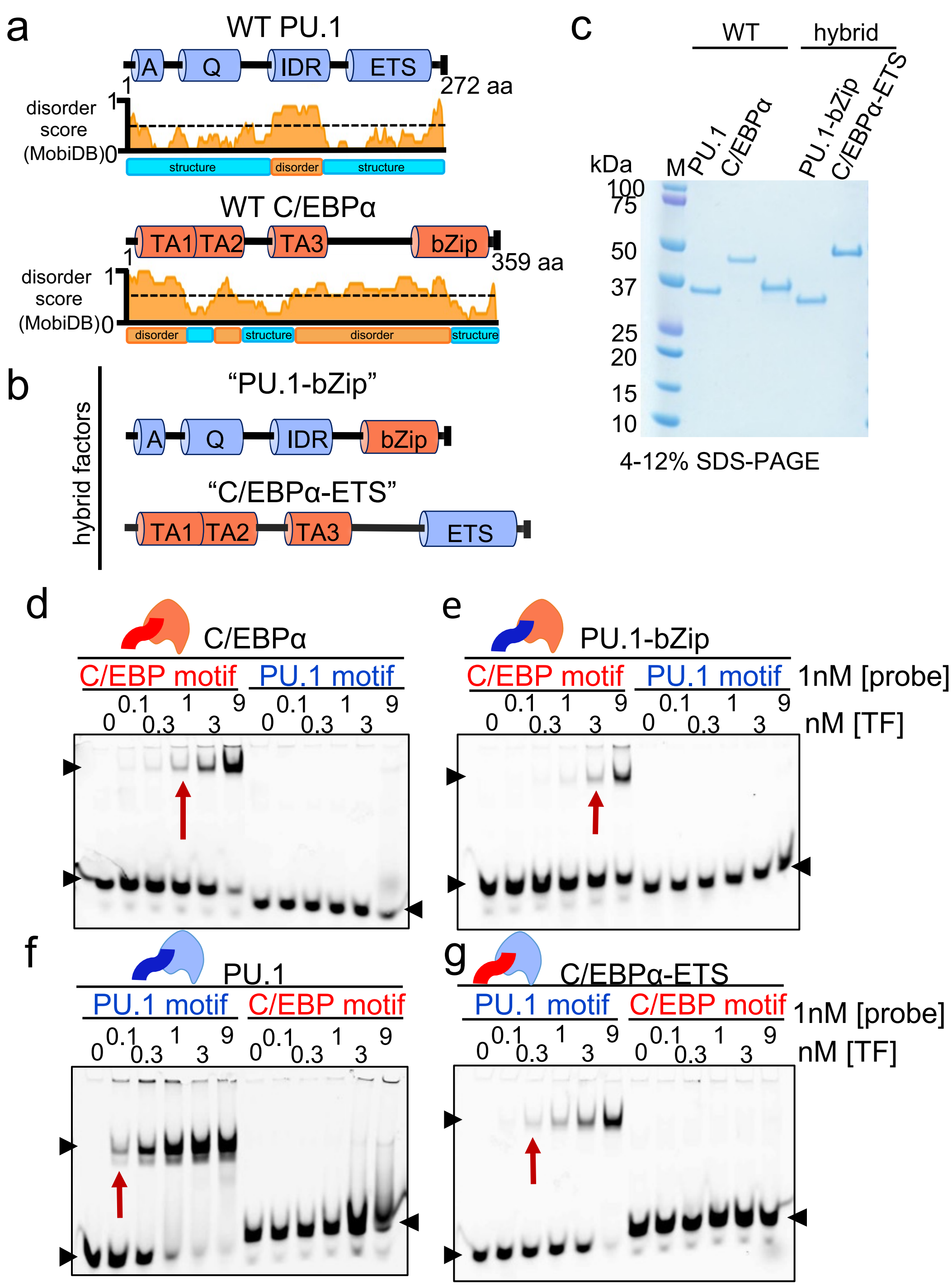


Figure B - 1 Purified recombinant hybrid transcription factors bind dsDNA probes with specificity.

Predicted disorder tendency of WT PU.1. The red and orange lines indicate the predicted disorder tendency along the WT PU.1 protein and WT C/EBPα protein (as calculated by MobiDB, respectively) with values above 0.5 (indicated by the dotted line) considered disordered. Shown are the positions of the Acidic domain (A), Q-rich domain (Q), Intrinsically disordered region (IDR) and ETS DNA-binding domain (ETS), Transactivation domains 1, 2, and 3 (TA1, TA2, TA3), and bZip DNA-Binding domain (bZip). b, Illustration of hybrid transcription factors. PU.1-bZip is composed of the N-terminal domains of PU.1 and the bZip DNA-binding domain of C/EBPα. C/EBPα-ETS is composed of the N-terminal domain of C/EBPα fused to the ETS DNA-binding domain of PU.1 c, SDS-PAGE analysis of 1 µg of each protein, stained with Coomassie blue. d-g Representative EMSAs showing the affinity of increasing amounts of **d**, PU.1, **e**, PU.1-bZip, **f**, C/EBPα, **g**, C/EBP-ETS to Cy5-labeled dsDNA (black arrows) containing the specific PU.1 motif or C/EBP motif.

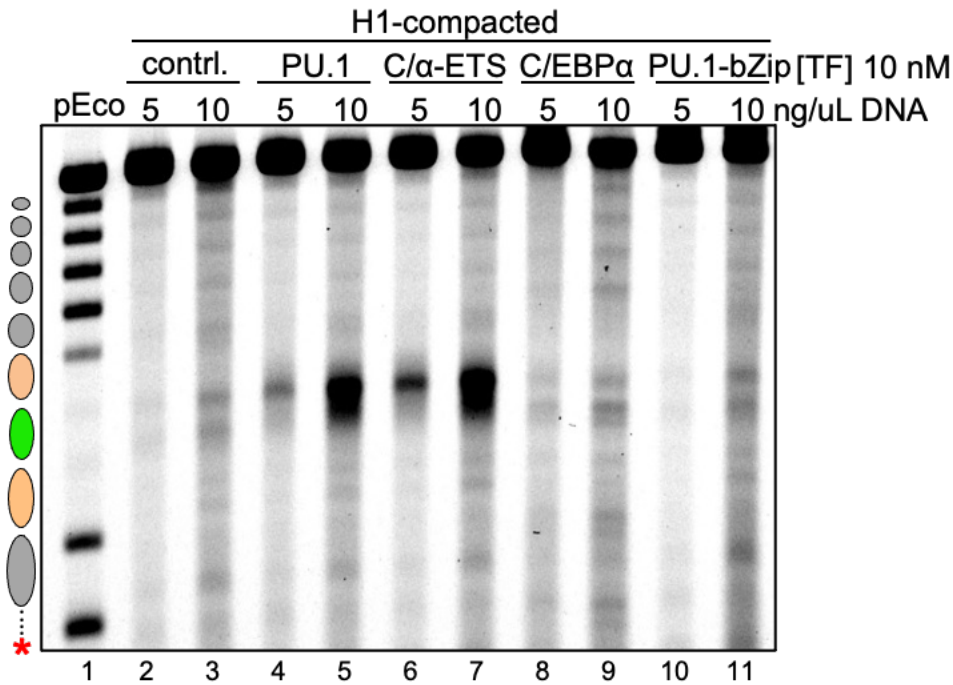


Figure B - 2 PU.1 and C/EBPα-ETS produce the same chromatin opening pattern. DNase I digestion analysis H1-compacted nucleosome arrays incubated alone (control., lanes 2-3), WT PU.1 (lanes 4-5), C/EBPα-ETS (C/a-ETS) (lanes 6-7), and C/EBPα (lanes 8-9), and PU.1-bZip (lanes 10-11).

APPENDIX C: NANOPORE SEQUENCING ANALYSIS OF NUCLEOSOME ARRAYS

This appendix provides a detailed overview of the nanopore sequencing protocol and data analysis used in Chapter 2. It was specifically designed for the preparation and analysis of *in vitro* generated nucleosome arrays. This protocol is based on protocols published by the Zaret Lab (Cirillo and Zaret, 2004) and Oxford Nanopore, but was adapted for sequencing of *in vitro* generated nucleosome array nuclease digests. All adjustments to the published library prep protocols were designed by myself, with input from Dr. Kelsey Kaeding and Dr. Ryan McCarty. The Barcoding, Basecalling and alignment protocols are based on protocols published by Oxford Nanopore, but were adapted to perform the downstream End-point analysis of nuclease digested nucleosome arrays. The End-point analysis pipeline was developed by myself, Dr. Max Ferretti, and Greg Donahue. All experiments and analyses were performed with guidance from Dr. Zaret.

I. Objective

Adapt nanopore sequencing to assess nuclease digestion patterns of *in vitro* generated nucleosome arrays.

II. Reagents

Preparation of nucleosome arrays for Nanopore sequencing

1. End-labeled nucleosome arrays
2. Linker histone Calbiochem Cat No. 10223549001
3. MNase Worthington Biochemicals Cat No. LS004797

4. DNase Worthington Biochemicals Cat No. LS006342

Nanopore sequencing library preparation

1. Input DNA- 5 uL of nuclease digested DNA
2. Nuclease free water
3. Low retention Eppendorf tubes
4. 0.2 mL thin walled PCR tubes
5. FLO-MIN106 flow cells
6. Native barcoding expansion 1-12 Cat No. EXP-NBD103
7. Ligation sequencing kit Cat No. SQK-LSK109
8. FLOW Cell Priming kit Cat No. EXP-FLP001
9. Agencourt AMPure XP beads
10. NEBNext End Repair / dA-tailing module Cat No. E7546
11. NEBNext Quick Ligation Module Cat No. E6056
12. NEBNEXT FFPE Repair Mix Cat No. M6630
13. NEB Blunt/TA Ligase Master Mix Cat No. M0367
14. Freshly prepared 70% ethanol in nuclease free water

III. Protocol

Preparation of nucleosome arrays for Nanopore sequencing

Day 1: nuclease digest of extended nucleosome arrays (I) *OR* H1-compacted nucleosome arrays (II).

I. DNase or MNase Digestion of extended nucleosome arrays (~2 hours)

1. Set up 80 uL reactions. The concentration of array will be ~1nM or ~40 ng/uL:

order	Component	Stock conc.	Final conc.	Volume
1	ddH_20	-	-	45.6 ul
2	10X Hepes, pH 7.5/KCL	60 mM Hepes and 350 mM KCl	6 mM Hepes 35 mM KCl	8 uL
3	BSA	10 mg/mL	0.25 mg/mL	2 uL
4	DTT	1M	5mM	.4 uL
5	Glycerol	30%	1.5%	4 uL
6	Nucleosome array	6.7 nM	1nM	12 uL
7	Nuclease (1 minute incubation)	varies	varies	8 uL
8	STOP Buffer	-	-	80 uL

2. Incubate 30 minutes at RT
3. Prepare Stop Buffer
4. Prepare DNase in 50 mM MgCl2. I typically use 1-5 ng/uL DNase for extended (final concentration). If digesting with MNase, prepare MNase dilution in 30 mM CaCl2, typically .1-.6 U/mL of MNase for extended nucleosome arrays.
5. Digest each reaction for 1 min. I typically digest four samples at a time and stagger the addition of DNase every 15 seconds
6. After 1 minute digestion add 40 uL Stop buffer. Proceed to Ethanol Precip.

II. Chromatin Opening Assay with H1-compacted nucleosome arrays (~4 hours)

1. Set up 40 uL reactions. The concentration of array will be ~1nM or ~20 ng/uL.
2. Dilute proteins (linker histone and TFs) in “EcoDB” protein dilution buffer (Cirillo Mol Cell 2002). Below is the recipe to make 10 mL:

EcoDB Protein Dilution Buffer	
reagent	volume
1 M KCL	1 mL
1 M Hepes	200 uL
100% Glycerol	2 mL
10% Igepal	100 uL
1M MgCl2	10 uL
ddH20	6.690 mL

3. Below is an example of setting up the reaction:

order	Component	Stock conc.	Final conc.	Volume
1	ddH_20	-	-	37.6 uL
2	10X Hepes, pH 7.5/KCL	60 mM Hepes and 350 mM KCl	6 mM Hepes 35 mM KCl	8 uL
3	BSA	10 mg/mL	0.25 mg/mL	2 uL

4	DTT	1M	5mM	.4 uL
5	Glycerol	30%	1.5%	4 uL
6	Nucleosome array	6.7 nM	1nM	12 uL
7	Linker histone H1 (1 hr incubation)	260 nM	13 nM	4ul
8	Transcription factor (Incubation time varies)	varies	varies	4uL
9	Nuclease (1 minute incubation)	varies	varies	8 uL
10	STOP Buffer	-	-	81 uL

4. Incubate 1 hr after the addition of linker histone H1
5. Add transcription factor (usually aim for a final concentration of 10-40 nM, but individual TFs should be titrated to determine optimal concentration for the experiment). Optimal time of incubation should be titrated as well. I incubate PU.1 and C/EBPs for 1 hr, but Lisa Cirillo incubated with FoxA and other TFs for 2 hours. Incubation time can make a big difference!!!
6. Prepare Stop Buffer
7. Prepare DNase in 50 mM $MgCl_2$. I typically us 5-10 ng/uL DNase (final concentration). In this case the 10X DNase concentration would be 50-100ng/uL.

8. Digest each reaction for 1 min. I typically digest four samples at a time and stagger the addition of DNase every 15 seconds
9. After minute digestion add 40 uL **Stop buffer** (recipe below) and proceed to ethanol precipitation:

Stop Buffer	
Reagent	Volume
1M Tris, ph 7.5	20 uL
tRNA 50 mg/mL	5 uL
0.5 M EDTA ph 8.0	125 uL
10% SDS	125 uL
Proteinase K 20 ug/uL	12.5 uL
ddH20	712.5uL

III. Ethanol precipitation of digested nucleosome arrays (overnight)

1. Add 200 uL of ddH20 to each tube to bring the volume to 280uL. This prevents the SDS in the Stop buffer from precipitating. Then add 28uL 3M Sodium Acetate and 1 mL of 100% Ethanol.
2. Precipitate overnight at -20°C
3. Prepare 80% ethanol (1 mL per sample) and store at -20C until needed. I suggest doing this the night before so it is ready to go in the morning

I. Day 2: Resuspend DNA, run and analyze gel (~5 hrs).

4. Spin down at max speed in a microcentrifuge for 30 min.
5. While tubes are spinning down, prepare a 1% Agarose gel made with filtered 1X TBE. Caste a midi gel (10X12 cm) with the 24 well comb. It is important to caste a thin gel as this will improve Cy5 signal. It is recommended to prepare 75 mL of agarose, which produces a gel 6.25 mm thick.
6. Using a needle, aspirate the supernatant taking care to not disturb the pellet. It's important to aspirate as much of the supe as possible.
7. Add a mL of ice cold 80% ethanol to each tube.
8. Spin at max speed for 15 minutes 4C
9. Using a needle, aspirate the supernatant taking care to not disturb the pellet. It's important to aspirate as much of the supe as possible. Aspirate liquid from tube walls to prevent it from pooling at the bottom of the tube.
10. Dry pellet for 20 minutes (uncapped on benchtop).
11. Add 20 uL 1X Ficoll to the bottom of each tube to cover the pellet. I let the pellets sit like this for 10 minutes, then resuspend the pellets. Quick spin each tube. You will save 10 ul of each sample to prepare sequencing libraries and run 10 ul of each sample on a gel.
12. Pipette 10uL of each sample into a fresh, labeled Eppendorf tube. Store samples at -20C until ready to prep libraries.
13. Load 10 ul of each sample onto 1% Agarose gel made with 1X TBE. I load the gel while running at 80V for ten minutes, then run at 100V for ~3 hours. Image gel on Typhoon.

1D Native Barcoding Library Prep For Nanopore Sequencing Of Nuclease Digested Nucleosome Arrays

<u>Library Prep Workflow:</u>

-Repair the DNA and prepare the DNA ends for adapter attachment

-Attach Native barcodes supplied in the kit to the DNA ends

-Prime the flow cell, and load your DNA library into the flow cell

Before you begin: to login to the Miniknow software to perform a sequencing run, you must have login credentials for Oxford Nanopore. You must have these credentials to access protocols and other support.

Helpful link to Oxford nanopore protocols for library prep:

https://community.nanoporetech.com/docs/prepare/library_prep_protocols/experiment-companion-minknow/v/mke_1013_v1_revco_11apr2016

I. Preparing input DNA

a. Note: For the final step you will want to pool your samples and that total mass of DNA should not exceed ~ 200 fmol. This is because the moles of DNA fragments are important. Fragments resulting from nuclease digestion of nucleosome arrays are generally small (< 3 kb) so you will need less mass than what is anticipated for genomic DNA. If the average fragment is ~ 2000 bp, then 60 ng of DNA is around 45 fmol. I barcode less than or equal to 12 samples at a time and start with 30 ng/sample.

b. Note: The library prep protocol published by Oxford Nanopore is written for repairing and barcoding an excess of genomic DNA. Because I have limited amounts of *in vitro* derived sample, I only repair and barcode what I will need for the adaptor ligation and sequencing run.

1. Dilute sample to 48 uL final volume with nuclease free water in a 0.2 mL PCR tube. For example, add 43 ul of nuclease free water to 0.2 ml PCR tube and 5 uL of sample DNA (30 ng)
2. Mix thoroughly by inversion
3. Spin down briefly

II. DNA Repair and End Prep (~90 minutes)

4. Prepare the NEBNext FFPE DNA Repair Mix and NEBNext End repair/ dA-tailing module reagents according to NEB protocols
5. In a 0.2 ml thin walled PCR tube, mix the following. Do not add Ultra II End-Prep Enzyme Mix until after the first 20C incubation

Reagent	**Volume**
DNA	48 ul
NEBNext FFPE DNA Repair Buffer	3.5 ul
NEBNext FFPE DNA Repair Mix	2 ul
Ultra II End-Prep Reaction Buffer	3.5 ul
Ultra II End-Prep Enzyme Mix	3 ul (add after first 20C incubation)
Total	**60 ul**

6. Mix by gently flicking the tube and spin down
7. Using a thermal cycler:
    a. incubate for 30 minutes at 20°C

b. Add the Ultra II End-Prep Enzyme Mix. Mix by gently flicking the tube and spin down. incubate for 30 minutes at 20°C
c. incubate for 5 minutes at 65°C

8. Resuspend AMPure XP beads by vortexing

9. Transfer the sample to a clean low retention 1.5 mL tube

10. Add 60 ul of resuspended AMPure XP beads to the end-prep reaction and mix by pipetting

11. Incubate on a rotation mixer for 5 mins at RT

12. Prepare 500 uL of fresh 70% ethanol in nuclease-free water per sample.

13. Spin down the sample and pellet on a magnet. Keep the tube on the magnet, and pipette of the supernatant.

14. Keep the tube on the magnet and wash the beads with 200 uL of freshly prepared 70% ethanol without disturbing the pellet. Remove the 70% ethanol using a pipette and discard.

15. Repeat the wash with 70% ethanol, pipette off and discard.

16. Spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for ~30 seconds, but do not dry the pellet to the point of cracking.

17. Remove the tube from the magnetic rack and resuspend the pellet in 22.5 ul of nuclease-free water.

18. Incubate for 2 minutes at room temperature.

19. Pellet the beads on a magnet until the eluate is clear and colorless.

20. Remove and retain 22.5 ul of eluate into a clean 1.5 ml low retention Eppendorf

III. Native Barcode Ligation (~110 minutes)

21. Thaw the native barcodes at room temperature – one barcode per sample. Mix the barcodes by pipetting and place them on ice.
22. Select a unique barcode for every sample to be run together on the same flow cell, from the provided 24 barcodes. 24 is the maximum with kits EXP-NBD104 (1-12) and EXP-NBD114 (1-24)
23. Add the reagents in the order given below, and mix by flicking the tube in between each addition.

Reagent	**Volume**
end-prepped DNA	22.5 ul
Native Barcode	2.5 ul
Blunt/TA Ligase Master Mix	25 ul
Total	**50 ul**

24. Mix by gently flicking the tube, and spin down.
25. Incubate the reaction for 10 minutes at room temperature.
26. Prepare the AMPure XP beads for use; resuspend by vortexing.
27. Add 50 ul of resuspended AMPure XP beads to the reaction and mix by pipetting.
28. Incubate on a rotating mixer for 5 mintues at room temperature.
29. Prepare 500 ul of fresh 70% ethanol in nuclease-free water per sample.
30. Spin down the sample and pellet on a magnet. Keep the tube on the magnet and pipette off the supernatant.

31. Keep on the magnet, wash beads with 200 ul of freshly prepared 70% ethanol without disturbing the pellet. Remove the 70% ethanol using a pipette and discard.

32. Repeat the wash, pipette off the ethanol, and discard.

33. Spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for ~30 seconds, but do not dry the pellet to the point of cracking.

34. Remove the tube from the magnetic rack, and resuspend the pellet in 20 uL of nuclease-free water.

35. Incubate for 2 minutes at room temperature.

36. Pellet the beads on a magnet until the eluate is clear and colorless.

37. Remove and retain 20 ul* of eluate into a clean 1.5 ml low retention Eppendorf tube.

   a. After removing the supernatant, discard the beads.

   b. Before performing this step note that the next step is to pooled samples prior to adaptor ligation, which will require the volume of pooled samples to be 50 uL. IF you have many barcoded samples it is better to do another round of beads to concentrate your sample instead of trying to elute DNA in a very small volume. *this volume can vary depending on your sample number, but I don't recommend using a smaller volume

   c. Remember you want to have around 200 fmols of material for the adapter ligation, so pool your samples accordingly!

IV. Optional concentration of pooled sample volume (~30 minutes)

38. Optional: If smaller volume of samples is required, after step 38, pool EQUAL volumes of sample (in this example each sample should be 20 uL, so pool 20 uL of each sample). Take note of your final volume
39. Resuspend Ampure beads by vortexing. Add 2.5X ampure beads (if your pooled library is 200 uL, add 500 uL of beads).
40. Prepare 500 ul of fresh 70% ethanol in nuclease-free water per sample.
41. Spin down the sample and pellet on a magnet. Keep the tube on the magnet and pipette off the supernatant.
42. Keep on the magnet, wash beads with 200 ul of freshly prepared 70% ethanol without disturbing the pellet. Remove the 70% ethanol using a pipette and discard.
43. Repeat the wash, pipette off the ethanol, and discard.
44. Spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for ~30 seconds, but do not dry the pellet to the point of cracking.
45. Remove the tube from the magnetic rack, and resuspend the pellet in 50 uL of nuclease-free water.
46. Incubate for 2 minutes at room temperature.
47. Pellet the beads on a magnet until the eluate is clear and colorless.
48. Remove and retain 50 uL of eluate into a clean 1.5 ml low retention Eppendorf tube.

V. Adapter Ligation and Clean-up (~30 minutes)

49. Pool equal amounts of sample.

50. Thaw one tune of barcode adapter mix (BAM 1D) at RT. Spin down, mix by pipetting, and place on ice.

51. Thaw elution buffer (EB) and NEBNext Quick Ligation Reaction Buffer (5x) at room temperature, mix by vortexing, spin down, and place on ice. Check that the tubes are clear of any precipitate.

52. Spin down the T4 Ligase and place on ice.

53. To retain all DNA fragments, thaw one tube of Short fragment buffer at room temperature, mix by vortexing, spin down, and place on ice.

54. Taking the barcoded and pooled DNA, perform adapter ligation as follows – mixing and flicking the tube between each sequential addition.

Reagent	**Volume**
~200 fmols pooled library	50 ul
Barcode Adapter Mix (BAM 1D)	20 ul
NEBNext Quick Ligation Reaction Buffer (5x)	20 ul
Quick T4 DNA Ligase	10 ul
Total	100 ul

55. Mix by gently flicking the tube and spin down.

56. Incubate the reaction for 10 minutes at room temperature.

57. Prepare the AMPure XP beads by voretxing.

58. Add 40 ul of resuspended AMPure XP beads to the adapter ligation reaction from the previous step and mix by pipetting.

59. Incubate on a rotating mixer for 5 minutes at RT.

60. Place on the magnetic rack, allow beads to pellet and pipette off supernatant.

61. Wash the beads by adding 250 ul Short fragment buffer. Flick the beads to resuspend, then return to the magnetic rack and allow the beads to pellet.
62. Remove the supernatant using a pipette and discard.
63. Repeat the wash with 250 ul Short fragment buffer, return to magnet, and pipette off supernatant and discard.
64. Spin down and place the tube back on the magnet. Pipette off any residual supernatant. Allow to dry for ~30 seconds, do not dry the pellet to the point of cracking.
65. Remove the tube from the magnetic rack and resuspend the pellet in 15 ul Elution Buffer (EB).
66. Incubate for 10 minutes at room temperature.
67. Pellet the beads until the eluate is clear and colorless.
68. Remove and retain 15ul of the supernatant into a clean low-retention 1.5 ml Eppendorf tube.
69. The prepared library is ready for loading into the MinION flow cell.

VI. Priming and Loading the SpotON flow cell

See useful video here: https://www.youtube.com/watch?v=CC11Jlydqrc

70. Thaw at room temperature and then place on ice:
    a. Sequencing buffer (SQB) – vortex, spin down and place on ice
    b. Loading beads (LB) – vortex, spin down, and place on ice
    c. Flush tether (FLT) – mix by pipetting, spin down and place on ice
    d. One tube of Flush Buffer (FLB) – spin down and place on ice
71. Open the lid of the minION and slide the flow cell's priming port cover clockwise so that the priming port is visible.

72. After opening the priming port, check for small bubbles under the cover. Draw back a small volume to remove any bubble (a few uls).
    a. Set a p1000 pipette to 200 ul
    b. Insert the tip into the priming port
    c. Turn the wheel until the dial shows 220-230 ul and, or until you can see a small column of buffer entering the pipet tip.
73. Prepare the flow cell *priming mix*
    a. Add 30 ul thawed and mixed flush tether (FLT) directly to the tube of thawed and mixed flush buffer (FLB) and mix by pipetting up and down
74. Load 800 ul of the *priming mix* into the flow cell via the priming port, avoiding the introduction of air bubbles. Wait for 5 minutes.
75. Thoroughly mix the contents of the LB tube by pipetting – use immediately
76. Prepare the library for loading

Reagent	**Volume**
Sequencing Buffer (SQB)	37.5 ul
Loading Beads (LB)	25.5 ul
DNA library	12 ul
Total	**75 ul**

77. Gently lift the SpotON sample port cover to make the spotON sample port accessible
78. SLOWLY load 200 ul of the *priming mix* into the flow cell via the priming port – avoiding air bubbles.
79. Mix the prepared library gently by pipetting up and down just prior to loading.

80. SLOWLY add 75 ul of sample to the flow cell via the SpotON sample port in a dropwise fashion. Ensure each drop flows into the port before adding the next

81. Gently replace the spotON sample port cover, making sure the bung enters the SpotON port, close the priming port and replace the MinION lid.

VII. Perform the sequencing run following the "Starting a Sequencing Run" section of the protocol

Link:

https://community.nanoporetech.com/docs/prepare/library_prep_protocols/experiment-companion-minknow/v/mke_1013_v1_revco_11apr2016/starting-a-sequencing-run-on-minion

1. Open Miniknow and Login
2. Select the sequencing device connected to the computer.
3. Type in the experiment name, sample ID and choose flow cell type from the drop down menu.
4. Select the kit used from the Kit selection menu.
5. Select run options. I generally sequence for 24 hours.
6. Turn off basecalling, barcoding and alignment options. They will be performed with Guppy after the sequencing run is complete. The only output you need are the raw FAST5 files
7. Select the output data location and select FAST5 for output format
8. Click 'Start' to run the experiment.

VIII. Washing the flow cell

Washing the flow cell allows for sequential runs of multiple sequencing libraries. This allows for the same flow cell to be used multiple times.

See this helpful video ! ***https://www.youtube.com/watch?v=lRTdK0kl9v4***

82. Prepare the flow cell for washing

83. Thaw solution A

- **a.** IF immediately starting another run, also thaw solution B
- **b.** IF storing the flow cell, thaw solution S

84. Stop the MinKNOw if you haven't yet

85. Open the priming port cover of the flow cell to check that the buffer is continuous. Carefully remove any bubbles by drawing up a few ul of buffer through the priming port.

86. Ensure that the SpotON sample port is closed and that the priming port is open for the washing steps.

87. Using a P1000 add 150 ul of Solution A through the priming port.

88. Wait 10 minutes.

- ***a.*** *To Run a second library immediately (SKIP if storing the flow cell)*
  - **i.** Add 150 ul of Solution B through the priming port of the flow cell.
  - **ii.** Follow instructions in the priming and loading the SpotON flow cell section of the protocol

iii. If you stopped data acquisition, start a new minKNOW experiment by naming the run and selection the appropriate experimental script.

b. *To store the flow cell for later use*

i. SLOWLY add 500 ul of storage buffer through the priming port of the flow cell

ii. Close the priming port and remove all the buffer from the waste section of the flow cell through either of the waste ports (directly across from the priming port)

iii. The flow cell can now be stored at 4C.

iv. When you wish to reuse the flow cell from storage, allow to warm to RT for ~5 minutes.

89. On the next experiment, it is advisable to adjust the starting voltage due to voltage drift during the course of the previous run.

Nanopore Sequencing End-Point Analysis Of Nuclease Digested Nucleosome Arrays

For the full Guppy handbook go to the Oxford Nanopore Website:
https://community.nanoporetech.com/docs/prepare/library_prep_protocols/Guppy-protocol/v/gpb_2003_v1_revam_14dec2018

You will need an account associated with the Zaret lab to log in. I recommend reading through the protocol to familiarize yourself with Guppy and the various options. Guppy has now integrated minimap2 to perform the alignment. Because this is a newer feature it is not in the protocol below, but is an option to streamline for future analyses. As written, Basecalling and Barcoding are performed on the Zaret Lab “Alien” computer and Alignment is performed on the HPC.

I. Basecalling with Guppy

1. First step is to take the raw FAST5 files produced by the flowcell and basecall them. There are different configuration files that can be used for each flowcell/kit combination. In addition, there is a config file for each mode: normal, high accuracy and super-high accuracy (that last one is only available for some kits).
2. High accuracy is the way to go, but you *must* use the Alienware computer with the NVIDIA geforce rtx 3090 graphics card (GPU). Plan on it taking anywhere from 2 hours to several hours, depending on the size of your FAST5 files. The normal mode is possible with a CPU, but basecalling a single run can take over a week.
3. To run basecall with Guppy you need the following commands:
   `Guppy_basecaller -i input_fast5_directory -s output_directory -c configuration_file.cfg –device auto`
   i. The “i” option: Guppy uses files in *fast5* format as input option -i or –input_path.

ii. The output directory is -s or--save_path

iii. Apart from the input fast5 files guppy requires a configuration file that sets the basecalling parameters depending on which flowcell and library preparation kit was used to produce the data. To start the basecalling specify the configuration file with option -c or –config

iv. Basecalling with guppy requires a GPU, which we have in the Zaret lab. However, Guppy will default to CPU usage unless the GPU device is specified using option --device

4. Here is an example (putting sudo in front grants the program read/write permission and is necessary):
```
sudo /home/zaretlab/ont-guppy/bin/guppy_basecaller -i ./
Nanopore_seq_run _fast5_allSample/ -s
Nanopore_seq_run_20210930_GuppyOut/ -c
/opt/ont/guppy/data/dna_r9.4.1_450bps_hac.cfg --device auto
```

II. Barcoding with Guppy

1. The basecaller converts the raw voltage signal stored in the fast5 files and outputs fastq files. The files that you will use for Barcoding will be located in the “Pass” folder. Some fail QC. Don’t use these.

2. Next, you need to figure out which reads belong to which samples using the barcoder.

3. To run barcoder you need the following commands:
```
guppy_barcoder --input_path input_directory --save_path
output_directory --config config_file.cfg --
require_barcodes_both_ends --trim_barcodes --barcode_kits "EXP-
NBD104 EXP-NBD114" --device auto
```

i. Input directory will be your folder with Fastq files produced by the basecalling that passed QC
ii. Specify that reads must have the barcode present at both ends with option
--require_barcodes_both_ends
iii. Enable barcode trimming with option --trim_barcodes
iv. Specify barcode kits used with --barcode_kits

For Example:

```
sudo /home/zaretlab/ont-guppy/bin/guppy_barcoder --input_path
Nanopore_seq_run_20210930_GuppyOut/pass --save_path
Nanopore_seq_run_20210930_GuppyOut/pass/barcoding --config
/opt/ont/guppy/data/dna_r9.4.1_450bps_hac.cfg --
require_barcodes_both_ends --trim_barcodes --barcode_kits "EXP-
NBD104 EXP-NBD114" --device auto
```

III. Aligning with minimap2

Helpful Link: https://lh3.github.io/minimap2/minimap2.html

1. Guppy will output a folder for each barcode containing multiple fastq files per barcode, which should be pooled to one fastq file per barcode. For example, move into the directory containing all the barcode folders and use this loop:

```
for dir in *; do cat $dir/*.fastq > $dir.merged.fastq; done;
```

2. Once all your pooled fastq files have moved to a single directory, proceed to alignment with minimap2

3. Alignment example for just one file:
    a. Make a directory for the aligned files:
    `mkdir New_barcode_directory_aligned`
    b. Run minimap2 with this command:
    `Minimap2 -a ref.fastq input_file.fastq --sam-hit-only --secondary=no > output.sam`

        i. -a tells minimap what sequence to align against (reference sequence)

        ii. --sam-hit-only only mapped reads are in the output

        iii. --secondary=no by default minimap2 outputs secondary alignment, use this option so there are no secondary alignments

Example:

```
minimap2 -a /home/meganfr/reference/Cx3cr1_11.fa
input_fastq/home/meganfr/Sept_2020/TF_CO_fastq_pass/Demulti/barco
de01.fastq --sam-hit-only --secondary=no >
/home/meganfr/Sept_2020/TF_CO_fastq_pass/Demulti/align3/barcode01
.sam
```

4. That's how it works, but I use a loop to align all the stuff in one command:

```
cd /home/meganfr/Sept_2020/TF_CO_fastq_pass/Demulti
for i in `ls *.fastq`
do
 filename="${i%*.fastq}"
```

```
 echo $filename

mkdir /home/meganfr/Sept_2020/TF_CO_fastq_pass/Demulti/align3
minimap2 -a /home/meganfr/reference/Cx3cr1_11.fa
/home/meganfr/Sept_2020/TF_CO_fastq_pass/Demulti/${filename}.fast
q --sam-hit-only --secondary=no >
/home/meganfr/Sept_2020/TF_CO_fastq_pass/Demulti/align3/${filenam
e}.sam

done
```

You can run this on the lab computer or just on your own computer. If you want to do it on the cluster, make the loop above into a text file and name it minimapJob.sh then submit with:

```
j=minimapJob.sh

bsub -M 50000 -n 12 -R "span[hosts=1] rusage [mem=50000]" -J $j -o
log.$j.out -e log.$j.err sh $j.sh
```

IV. Make bed files

1. The sam files just tell you where each base of each read aligns in the genome. For downstream analysis we just need to know the location/length of the ends of each read (a read ending means the nuclease cut the array). Bed files contain all the info we need and are a lot easier to work with.

2. This is the script I use to make bedfiles using samtools (sam to bam) and bedtools (bam to bed). Move into the directory containing the sam files you want to convert and run this loop:

```
cd /home/meganfr/Jan2022/barcoded_3/align
for i in `ls *.sam`
do
```

```
 filename="${i%*.sam}"
 echo $filename

 cd /home/meganfr/samtools-1.9
 ./samtools view -S -b
/home/meganfr/Jan2022/barcoded_3/align/$filename.sam >
/home/meganfr/Jan2022/barcoded_3/align/$filename.bam

 cd ../bedtools2
 ./bin/bamToBed -i
/home/meganfr/Jan2022/barcoded_3/align/$filename.bam >
/home/meganfr/Jan2022/barcoded_3/align/$filename.bed

done
```

3. Now you've got bed files!

V. Make bedgraph files, normalize and visualize

1. Next up, we need to look at each bed file and figure out where the ends of each read are. These are the places where we have enzymatic cleavage. We need to add up all the read endpoints for each sample. Shout out to Greg Donahue who wrote the preliminary End-point analysis and Max Ferretti who re-wrote a streamlined pipeline in R which is user friendly. The scripts to do this are on the Zaret Lab computer, but can also be found here: https://github.com/maximus-sci/EndpointAnalyzer.
2. **The first script is "Endpoint_Analyzer"** and it requires R and R studio.
    i. In the first few lines you must specify the location of the input bed files (optionally, the output location for the filtered and normalized bedgraph files) and the reference sequence information.
    ii. Options within the script allow you to filter out reads based on length, quality, and location.
        1. The exclusions regions are to get rid of reads that map entirely within the repetitive regions of the nucleosome array (by their nature these reads can't be properly aligned. Sometimes minimap2 will still give it a go, though, so it's best to get rid of them). There is also a quality score filter, which I set to "2" to remove multimappers (0 and 1).

2. There is an option to exclude read sizes, but in my published NSMB paper the parameters are 0 and 2733, in which case no reads are excluded by size. This could be useful in the future though if looking at particular read sizes is of interest (perhaps to distinguish between free and nucleosomal DNA).

iii. The script itself is commented to tell you what each part is doing. In summary, first it reads in all the bed files and applied the filters. Next, it goes through and counts up all the read endpoints using a hash dictionary and normalizes these to the fraction of reads at each position to create a normalized density. Finally, a shell script is written that will convert the bedgraph files into bigWigs for visualization in a genome browser.

3. **Bedgraph_normalizer_relToCntrl.** This one was written in Python3 by Max Ferretti. You will use this to subtract the signal of one bedgraph file from another bedgraph file. This is useful for accounting for taking into account sequence bias for DNase or MNase by subtracting the free DNA signal from the nucleosome array signal. For example, this is how extended data figure 2C was generated in Frederick et al 2022 NSMB.
    i. Take the script and paste a copy of it in the directory containing your normalized bedgraphs produced by the first script.
    ii. In the command line, navigate to the folder and run the script with: `python Bedgraph_normalizer_relToCntrl,v1.py`
    iii. Tell it which file is the control and then it'll make a new set of bedgraphs.

4. **Bedgraph_smoother.** This is an R script that should be used in R studio. It uses tidyverse syntax (mostly dplyr) to generate plots of bedgraph reads with a nice, smoothed line for the read density. This uses the kernel smoothing function in R and makes a nice curve. To get this to work, you'll probably need at least some familiarity with R and tidyverse. For reference, this is what was used to generate extended figure 2d in Frederick et al 2022 NSMB

REFERENCES

Adams, C.C., Workman, J.L., 1995. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. Molecular and Cellular Biology 15, 1405–21. https://doi.org/10.1128/mcb.15.3.1405

Allan, J., Hartman, P.G., Crane-Robinson, C., Aviles, F.X., 1980. The structure of histone H1 and its location in chromatin. Nature 288, 675–679. https://doi.org/10.1038/288675a0

Arbesú, M., Iruela, G., Fuentes, H., Teixeira, J.M.C., Pons, M., 2018. Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains. Front. Mol. Biosci. 5. https://doi.org/10.3389/fmolb.2018.00039

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., Natoli, G., 2014a. Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers. Molecular Cell. https://doi.org/10.1016/j.molcel.2014.04.006

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., Natoli, G., 2014b. Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers. Molecular Cell. https://doi.org/10.1016/j.molcel.2014.04.006

Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A., Zaret, K.S., 2017. Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. Molecular Cell 68, 1023-1037.e15. https://doi.org/10.1016/j.molcel.2017.11.030

Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., Menoni, H., Papin, C., Skoufias, D.A., Kurumizaka, H., Lavery, R., Hamiche, A., Hayes, J.J., Schultz, P., Angelov, D., Petosa, C., Dimitrov, S., 2017. Structure and dynamics of a 197 base-pair nucleosome in complex with linker histone H1. Mol Cell 66, 384-397.e8. https://doi.org/10.1016/j.molcel.2017.04.012

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., Zhuang, X., 2016. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529, 418–422. https://doi.org/10.1038/nature16496

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

Boller, S., Ramamoorthy, S., Akbas, D., Nechanitzky, R., Burger, L., Murr, R., Schübeler, D., Grosschedl, R., 2016a. Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. Immunity. https://doi.org/10.1016/j.immuni.2016.02.021

Boller, S., Ramamoorthy, S., Akbas, D., Nechanitzky, R., Burger, L., Murr, R., Schübeler, D., Grosschedl, R., 2016b. Pioneering Activity of the C-Terminal

Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. Immunity. https://doi.org/10.1016/j.immuni.2016.02.021

Borgia, A., Borgia, M.B., Bugge, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D., Kragelund, B.B., Best, R.B., Schuler, B., 2018. Extreme disorder in an ultrahigh-affinity protein complex. Nature 555, 61–66. https://doi.org/10.1038/nature25762

Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M.E., Rengarajan, S., Volorio, A., McBride, M.J., Broye, L.C., Zou, L., Stamenkovic, I., Kadoch, C., Rivera, M.N., 2017. Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. Cell 171, 163-178.e19. https://doi.org/10.1016/j.cell.2017.07.036

Braunschweig, U., Hogan, G.J., Pagie, L., van Steensel, B., 2009. Histone H1 binding is inhibited by histone variant H3.3. EMBO J 28, 3635–3645. https://doi.org/10.1038/emboj.2009.301

Brent', R., Ptashne, M., 1985. A Eukaryotic Transcriptional Activator Bearing the DNA Specificity of a Pmkaryotic Repressor 8.

Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D.K., Carmi, M., Barkai, N., 2020. Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. Molecular Cell 79, 459-471.e4. https://doi.org/10.1016/j.molcel.2020.05.032

Bryan, L.C., Weilandt, D.R., Bachmann, A.L., Kilic, S., Lechner, C.C., Odermatt, P.D., Fantner, G.E., Georgeon, S., Hantschel, O., Hatzimanikatis, V., Fierz, B., 2017. Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions. Nucleic Acids Res 45, 10504–10517. https://doi.org/10.1093/nar/gkx697

Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., Hahn, S., 2011. The acidic transcription activator Gcn4 binds the Mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. Mol Cell 44, 942–953. https://doi.org/10.1016/j.molcel.2011.11.008

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

Burge, N.L., Thuma, J.L., Hong, Z.Z., Jamison, K.B., Ottesen, J.J., Poirier, M.G., 2022. H1.0 C Terminal Domain Is Integral for Altering Transcription Factor Binding within Nucleosomes. Biochemistry acs.biochem.2c00001. https://doi.org/10.1021/acs.biochem.2c00001

Bussmann, L.H., Schubert, A., Vu Manh, T.P., De Andres, L., Desbordes, S.C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., Ballestar, E., Graf, T., 2009. A Robust and Highly Efficient Immune Cell Reprogramming System. Cell Stem Cell. https://doi.org/10.1016/j.stem.2009.10.004

Bustin, M., Catez, F., Lim, J.H., 2005. The dynamics of histone H1 function in chromatin. Molecular Cell 17, 617–620. https://doi.org/10.1016/j.molcel.2005.02.019

Canzio, D., Chang, E.Y., Shankar, S., Kuchenbecker, K.M., Simon, M.D., Madhani, H.D., Narlikar, G.J., Al-Sady, B., 2011. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. Mol Cell 41, 67–81. https://doi.org/10.1016/j.molcel.2010.12.016

Cao, K., Lailler, N., Zhang, Y., Kumar, A., Uppal, K., Liu, Z., Lee, E.K., Wu, H., Medrzycki, M., Pan, C., Ho, P.-Y., Jr, G.P.C., Dong, X., Bock, C., Bouhassira, E.E., Fan, Y., 2013. High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. PLOS Genetics 9, e1003417. https://doi.org/10.1371/journal.pgen.1003417

Cernilogar, F.M., Hasenöder, S., Wang, Z., Scheibner, K., Burtscher, I., Sterr, M., Smialowski, P., Groh, S., Evenroed, I.M., Gilfillan, G.D., Lickert, H., Schotta, G., 2019. Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. Nucleic Acids Research 47, 9069–9086. https://doi.org/10.1093/nar/gkz627

Chen, H., Kharerin, H., Dhasarathy, A., Kladde, M., Bai, L., 2022. Partitioned usage of chromatin remodelers by nucleosome-displacing factors. Cell Reports 40. https://doi.org/10.1016/j.celrep.2022.111250

Chen, X., Liu, W., Ambrosino, C., Ruocco, M.R., Poli, V., Romani, L., Quinto, I., Barbieri, S., Holmes, K.L., Venuta, S., Scala, G., 1997. Impaired generation of bone marrow B lymphocytes in mice deficient in C/EBPbeta. Blood.

Chen, Y., Cattoglio, C., Dailey, G., Zhu, Q., Tjian, R., Darzacq, X., 2021. Mechanisms Governing Target Search and Binding Dynamics of Hypoxia-Inducible Factors (preprint). Molecular Biology. https://doi.org/10.1101/2021.10.27.466110

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

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, 279–89. https://doi.org/10.1016/S1097-2765(02)00459-8

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., Zaret, K.S., 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO Journal 17, 244–254. https://doi.org/10.1093/emboj/17.1.244

Cirillo, L.A., Zaret, K.S., 2004. Preparation of Defined Mononucleosomes, Dinucleosomes, and Nucleosome Arrays In Vitro and Analysis of Transcription Factor Binding. Methods in Enzymology. https://doi.org/10.1016/S0076-6879(03)75009-5

Cirillo, L.A., Zaret, K.S., 1999. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. Molecular Cell. https://doi.org/10.1016/S1097-2765(00)80225-7

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

Danilenko, N., Lercher, L., Kirkpatrick, J., Gabel, F., Codutti, L., Carlomagno, T., 2019. Histone chaperone exploits intrinsic disorder to switch acetylation specificity. Nat Commun 10, 3435. https://doi.org/10.1038/s41467-019-11410-7

Davis, R.L., Weintraub, H., Lassar, A.B., 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51, 987–1000. https://doi.org/10.1016/0092-8674(87)90585-X

Decker, T., Pasca di Magliano, M., McManus, S., Sun, Q., Bonifer, C., Tagoh, H., Busslinger, M., 2009. Stepwise Activation of Enhancer and Promoter Regions of the B Cell Commitment Gene Pax5 in Early Lymphopoiesis. Immunity. https://doi.org/10.1016/j.immuni.2009.01.012

Desai, M.A., Webb, H.D., Sinanan, L.M., Scarsdale, J.N., Walavalkar, N.M., Ginder, G.D., Williams, D.C., 2015. An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. Nucleic Acids Res 43, 3100–3113. https://doi.org/10.1093/nar/gkv168

Di Stefano, B., Sardina, J.L., Van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., Graf, T., 2014a. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature. https://doi.org/10.1038/nature12885

Di Stefano, B., Sardina, J.L., Van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., Graf, T., 2014b. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature. https://doi.org/10.1038/nature12885

Dodonova, S.O., Zhu, F., Dienemann, C., Taipale, J., Cramer, P., 2020. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature 580, 669–672. https://doi.org/10.1038/s41586-020-2195-y

Donovan, B.T., Chen, H., Jipa, C., Bai, L., Poirier, M.G., 2019. Dissociation rate compensation mechanism for budding yeast pioneer transcription factors. eLife 8, e43008. https://doi.org/10.7554/eLife.43008

Dosztányi, Z., Csizmók, V., Tompa, P., Simon, I., 2005a. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. Journal of Molecular Biology 347, 827–839.

Dosztányi, Z., Csizmók, V., Tompa, P., Simon, I., 2005b. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. Journal of Molecular Biology 347, 827–839.

Drew, H.R., Travers, A.A., 1985. DNA bending and its relation to nucleosome positioning. Journal of Molecular Biology 186, 773–790. https://doi.org/10.1016/0022-2836(85)90396-1

Echigoya, K., Koyama, M., Negishi, L., Takizawa, Y., Mizukami, Y., Shimabayashi, H., Kuroda, A., Kurumizaka, H., 2020. Nucleosome binding by the pioneer transcription factor OCT4. Sci Rep 10, 11832. https://doi.org/10.1038/s41598-020-68850-1

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., Skoultchi, A.I., 2005. Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation. Cell 123, 1199–1212. https://doi.org/10.1016/j.cell.2005.10.028

Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., Graf, T., 2008. PU.1 and C/EBP alpha/beta convert fibroblasts into macrophage-like cells. Proceedings of the National Academy of Sciences 105, 6057–62. https://doi.org/10.1073/pnas.0711961105

Fernandez Garcia, M., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H., Zaret, K.S., 2019. Structural Features of Transcription Factors Associating with Nucleosome Binding. Molecular Cell 75, 921–932. https://doi.org/10.1016/j.molcel.2019.06.009

Friedman, A.D., 2007a. Transcriptional control of granulocyte and monocyte development. Oncogene. https://doi.org/10.1038/sj.onc.1210764

Friedman, A.D., 2007b. Transcriptional control of granulocyte and monocyte development, Oncogene. https://doi.org/10.1038/sj.onc.1210764
Friman, E.T., Deluz, C., Meireles-Filho, A.C., Govindan, S., Gardeux, V., Deplancke, B., Suter, D.M., 2019. Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle. eLife 8, e50087. https://doi.org/10.7554/eLife.50087
Fuglerud, B.M., Lemma, R.B., Wanichawan, P., Sundaram, A.Y.M., Eskeland, R., Gabrielsen, O.S., 2017a. A c-Myb mutant causes deregulated differentiation due to impaired histone binding and abrogated pioneer factor function. Nucleic Acids Research. https://doi.org/10.1093/nar/gkx364
Fuglerud, B.M., Lemma, R.B., Wanichawan, P., Sundaram, A.Y.M., Eskeland, R., Gabrielsen, O.S., 2017b. A c-Myb mutant causes deregulated differentiation due to impaired histone binding and abrogated pioneer factor function. Nucleic Acids Research. https://doi.org/10.1093/nar/gkx364
Fyodorov, D.V., Zhou, B.-R., Skoultchi, A.I., Bai, Y., 2018. Emerging roles of linker histones in regulating chromatin structure and function. Nat Rev Mol Cell Biol 19, 192–206. https://doi.org/10.1038/nrm.2017.94
Gao, A., Shrinivas, K., Lepeudry, P., Suzuki, H.I., Sharp, P.A., Chakraborty, A.K., 2018. Evolution of weak cooperative interactions for biological specificity. Proc. Natl. Acad. Sci. U.S.A. 115. https://doi.org/10.1073/pnas.1815912115
Garcia, D.A., Johnson, T.A., Presman, D.M., Fettweis, G., Wagh, K., Rinaldi, L., Stavreva, D.A., Paakinaho, V., Jensen, R.A.M., Mandrup, S., Upadhyaya, A., Hager, G.L., 2021. An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. Molecular Cell. https://doi.org/10.1016/j.molcel.2021.01.013
Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L., Ragoussis, J., Natoli, G., 2010. Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. Immunity. https://doi.org/10.1016/j.immuni.2010.02.008
Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., Bickmore, W.A., 2004. Chromatin Architecture of the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. Cell 118, 555–566. https://doi.org/10.1016/j.cell.2004.08.011
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
Gutiérrez, J.L., Chandy, M., Carrozza, M.J., Workman, J.L., 2007. Activation domains drive nucleosome eviction by SWI/SNF. EMBO J 26, 730–740. https://doi.org/10.1038/sj.emboj.7601524
Habchi, J., Tompa, P., Longhi, S., Uversky, V.N., 2014. Introducing Protein Intrinsic Disorder. Chem. Rev. 114, 6561–6588. https://doi.org/10.1021/cr400514h
Han, Y., Reyes, A.A., Malik, S., He, Y., 2020. Cryo-EM structure of SWI/SNF complex bound to a nucleosome. Nature 579, 452–455. https://doi.org/10.1038/s41586-020-2087-1
Healton, S.E., Pinto, H.D., Mishra, L.N., Hamilton, G.A., Wheat, J.C., Swist-Rosowska, K., Shukeir, N., Dou, Y., Steidl, U., Jenuwein, T., Gamble, M.J., Skoultchi, A.I., 2020. H1 linker histones silence repetitive elements by promoting both histone

H3K9 methylation and chromatin compaction. Proceedings of the National Academy of Sciences 117, 14251–14258. https://doi.org/10.1073/pnas.1920725117
Heath, V., Suh, H.C., Holman, M., Renn, K., Gooya, J.M., Parkin, S., Klarmann, K.D., Ortiz, M., Johnson, P., Keller, J., 2004. C/EBPα deficiency results in hyperproliferation of hematopoietic progenitor cells and disrupts macrophage development in vitro and in vivo. Blood. https://doi.org/10.1182/blood-2003-11-3963
Heidarsson, P.O., Mercadante, D., Sottini, A., Nettels, D., Borgia, M.B., Borgia, A., Kilic, S., Fierz, B., Best, R.B., Schuler, B., 2022. Release of linker histone from the nucleosome driven by polyelectrolyte competition with a disordered protein. Nat. Chem. 14, 224–231. https://doi.org/10.1038/s41557-021-00839-3
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell 38, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004
Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orozco, L.D., Glass, C.K., 2013. Effect of natural genetic variation on enhancer selection and function. Nature 503, 487–492. https://doi.org/10.1038/nature12615
Hill, D.A., Imbalzano, A.N., 2000. Human SWI/SNF nucleosome remodeling activity is partially inhibited by linker histone H1. Biochemistry 39, 11649–56. https://doi.org/10.1021/bi001330z
Horn, P.J., Carruthers, L.M., Logie, C., Hill, D.A., Solomon, M.J., Wade, P.A., Imbalzano, A.N., Hansen, J.C., Peterson, C.L., 2002. Phosphorylation of linker histones regulates ATP-dependent chromatin remodeling enzymes. Nat Struct Mol Biol 9, 263–267. https://doi.org/10.1038/nsb776
Hosokawa, H., Ungerbäck, J., Wang, X., Matsumoto, M., Nakayama, K.I., Cohen, S.M., Tanaka, T., Rothenberg, E.V., 2018. Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. Immunity 48, 1119-1134.e7. https://doi.org/10.1016/j.immuni.2018.04.024
Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., Hui, L., 2011. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature. https://doi.org/10.1038/nature10116
Iurlaro, M., Stadler, M.B., Masoni, F., Jagani, Z., Galli, G.G., Schübeler, D., 2021. Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nature Genetics 53, 279–287. https://doi.org/10.1038/s41588-020-00768-w
Iwafuchi, M., Cuesta, I., Donahue, G., Takenaka, N., Osipovich, A.B., Magnuson, M.A., Roder, H., Seeholzer, S.H., Santisteban, P., Zaret, K.S., 2020. Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. Nat Genet 52, 418–427. https://doi.org/10.1038/s41588-020-0591-8
Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., Zaret, K.S., 2016a. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Molecular Cell. https://doi.org/10.1016/j.molcel.2016.03.001

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., Zaret, K.S., 2016b. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Molecular Cell 62, 79–91. https://doi.org/10.1016/j.molcel.2016.03.001

Iwafuchi-Doi, M., Zaret, K.S., 2016. Cell fate control by pioneer transcription factors. Development. https://doi.org/10.1242/dev.133900

Iwafuchi-Doi, M., Zaret, K.S., 2014a. Pioneer transcription factors in cell reprogramming. Genes and Development. https://doi.org/10.1101/gad.253443.114

Iwafuchi-Doi, M., Zaret, K.S., 2014b. Pioneer transcription factors in cell reprogramming. Genes and Development 28, 2679–92. https://doi.org/10.1101/gad.253443.114

Izzo, A., Kamieniarz-Gdula, K., Ramírez, F., Noureen, N., Kind, J., Manke, T., van Steensel, B., Schneider, R., 2013. The Genomic Landscape of the Somatic Linker Histone Subtypes H1.1 to H1.5 in Human Cells. Cell Reports 3, 2142–2154. https://doi.org/10.1016/j.celrep.2013.05.003

Jakobsen, J.S., 2019. Mutant CEBPA directly drives the expression of the targetable tumor-promoting factor CD73 in AML. https://doi.org/10.1126/sciadv.aaw4304

Johnson, J.L., Georgakilas, G., Petrovic, J., Kurachi, M., Cai, S., Harly, C., Pear, W.S., Bhandoola, A., Wherry, E.J., Vahedi, G., 2018. Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. Immunity 48, 243-257.e10. https://doi.org/10.1016/j.immuni.2018.01.012

Judd, J., Duarte, F.M., Lis, J.T., 2021. Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. Genes Dev 35, 147–156. https://doi.org/10.1101/gad.341768.120

Kan, P.-Y., Lu, X., Hansen, J.C., Hayes, J.J., 2007. The H3 Tail Domain Participates in Multiple Interactions during Folding and Self-Association of Nucleosome Arrays. Molecular and Cellular Biology 27, 2084–2091. https://doi.org/10.1128/MCB.02181-06

Kilic, S., Bachmann, A.L., Bryan, L.C., Fierz, B., 2015. Multivalency governs HP1α association dynamics with the silent chromatin state. Nat Commun 6, 7313. https://doi.org/10.1038/ncomms8313

Kilic, S., Felekyan, S., Doroshenko, O., Boichenko, I., Dimura, M., Vardanyan, H., Bryan, L.C., Arya, G., Seidel, C.A.M., Fierz, B., 2018. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. Nat Commun 9, 235. https://doi.org/10.1038/s41467-017-02619-5

King, H.W., Klose, R.J., 2017. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. eLife 6, e22631. https://doi.org/10.7554/eLife.22631

Kowenz-Leutz, E., Leutz, A., 1999. A C/EBPβ isoform recruits the SWI/SNF complex to activate myeloid genes. Molecular Cell. https://doi.org/10.1016/S1097-2765(00)80384-6

Kribelbauer, J.F., Rastogi, C., Bussemaker, H.J., Mann, R.S., 2019. Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. Annual Review of Cell and Developmental Biology 35, 357–379. https://doi.org/10.1146/annurev-cellbio-100617-062719

Krishnakumar, R., Kraus, W.L., 2010. PARP-1 Regulates Chromatin Structure and Transcription through a KDM5B-Dependent Pathway. Molecular Cell 39, 736–749. https://doi.org/10.1016/j.molcel.2010.08.014

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., Jager, P.L.D., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. https://doi.org/10.1038/nature14248

Lerner, J., Gomez-Garcia, P.A., McCarthy, R.L., Liu, Z., Lakadamyali, M., Zaret, K.S., 2020. Two-Parameter Mobility Assessments Discriminate Diverse Regulatory Factor Behaviors in Chromatin. Molecular Cell 79, 677-688.e6. https://doi.org/10.1016/j.molcel.2020.05.036

Li, G., Margueron, R., Hu, G., Stokes, D., Wang, Y.-H., Reinberg, D., 2010. Highly Compacted Chromatin Formed In Vitro Reflects the Dynamics of Transcription Activation In Vivo. Molecular Cell 38, 41–53. https://doi.org/10.1016/j.molcel.2010.01.042

Li, R., Cauchy, P., Ramamoorthy, S., Boller, S., Chavez, L., Grosschedl, R., 2018. Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. Genes and Development. https://doi.org/10.1101/gad.309583.117

Li, S., Zheng, E.B., Zhao, L., Liu, S., 2019. Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. Cell Reports 28, 2689-2703.e4. https://doi.org/10.1016/j.celrep.2019.07.103

Liu, Y., Zhou, K., Zhang, N., Wei, H., Tan, Y.Z., Zhang, Z., Carragher, B., Potter, C.S., D'Arcy, S., Luger, K., 2020. FACT caught in the act of manipulating the nucleosome. Nature 577, 426–431. https://doi.org/10.1038/s41586-019-1820-0

Lu, X., Hamkalo, B., Parseghian, M.H., Hansen, J.C., 2009. Chromatin Condensing Functions of the Linker Histone C-Terminal Domain Are Mediated by Specific Amino Acid Composition and Intrinsic Protein Disorder. Biochemistry 48, 164–172. https://doi.org/10.1021/bi801636y

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389, 251–260. https://doi.org/10.1038/38444

Luzete-Monteiro, E., Zaret, K.S., 2022. Structures and consequences of pioneer factor binding to nucleosomes. Current Opinion in Structural Biology 75, 102425. https://doi.org/10.1016/j.sbi.2022.102425

Mashtalir, N., D'Avino, A.R., Michel, B.C., Luo, J., Pan, J., Otto, J.E., Zullow, H.J., McKenzie, Z.M., Kubiak, R.L., St. Pierre, R., Valencia, A.M., Poynter, S.J., Cassel, S.H., Ranish, J.A., Kadoch, C., 2018. Modular Organization and

Assembly of SWI/SNF Family Chromatin Remodeling Complexes. Cell 175, 1272–1288. https://doi.org/10.1016/j.cell.2018.09.032
Mashtalir, N., Suzuki, H., Farrell, D.P., Sankar, A., Luo, J., Filipovski, M., D’Avino, A.R., St. Pierre, R., Valencia, A.M., Onikubo, T., Roeder, R.G., Han, Y., He, Y., Ranish, J.A., DiMaio, F., Walz, T., Kadoch, C., 2020. A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. Cell 183, 802–817. https://doi.org/10.1016/j.cell.2020.09.051
Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., Drouin, J., 2018. Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat Genet 50, 259–269. https://doi.org/10.1038/s41588-017-0035-2
Mayran, A., Sochodolsky, K., Khetchoumian, K., Harris, J., Gauthier, Y., Bemmo, A., Balsalobre, A., Drouin, J., 2019. Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. Nat Commun 10, 3807. https://doi.org/10.1038/s41467-019-11791-9
McAndrew, M.J., Gjidoda, A., Tagore, M., Miksanek, T., Floer, M., 2016. Chromatin remodeler recruitment during macrophage differentiation facilitates transcription factor binding to enhancers in mature cells. Journal of Biological Chemistry. https://doi.org/10.1074/jbc.M116.734186
McGinty, R.K., Henrici, R.C., Tan, S., 2014. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature 514, 591–596. https://doi.org/10.1038/nature13890
McPherson, C.E., Shim, E.Y., Friedman, D.S., Zaret, K.S., 1993. An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. Cell. https://doi.org/10.1016/0092-8674(93)80079-T
Meers, M.P., Janssens, D.H., Henikoff, S., 2019. Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. Molecular Cell 75, 562-575.e5. https://doi.org/10.1016/j.molcel.2019.05.025
Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker, R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R., Weiss, J., Matsumoto, S., Burger, L., Schübeler, D., Thomä, N.H., 2020. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science 368, 1460–1465. https://doi.org/10.1126/science.abb0074
Miller, M., 2009. The Importance of Being Flexible: The Case of Basic Region Leucine Zipper Transcriptional Regulators. Curr Protein Pept Sci 10, 244–269.
Minderjahn, J., Schmidt, A., Fuchs, A., Schill, R., Raithel, J., Babina, M., Schmidl, C., Gebhard, C., Schmidhofer, S., Mendes, K., Ratermann, A., Glatz, D., Nützel, M., Edinger, M., Hoffmann, P., Spang, R., Längst, G., Imhof, A., Rehli, M., 2020. Mechanisms governing the pioneering and redistribution capabilities of the non-classical pioneer PU.1. Nat Commun 11, 402. https://doi.org/10.1038/s41467-019-13960-2
Minezaki, Y., Homma, K., Kinjo, A.R., Nishikawa, K., 2006. Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation. Journal of Molecular Biology 359, 1137–1149. https://doi.org/10.1016/j.jmb.2006.04.016
Mirny, L.A., 2010. Nucleosome-mediated cooperativity between transcription factors. Proceedings of the National Academy of Sciences 107, 22534–9. https://doi.org/10.1073/pnas.0913805107

Mivelaz, M., Cao, A.-M., Kubik, S., Zencir, S., Hovius, R., Boichenko, I., Stachowicz, A.M., Kurat, C.F., Shore, D., Fierz, B., 2020. Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. Molecular Cell 77, 488-500.e9. https://doi.org/10.1016/j.molcel.2019.10.025

Munari, F., Soeroes, S., Zenn, H.M., Schomburg, A., Kost, N., Schröder, S., Klingberg, R., Rezaei-Ghaleh, N., Stützer, A., Gelato, K.A., Walla, P.J., Becker, S., Schwarzer, D., Zimmermann, B., Fischle, W., Zweckstetter, M., 2012. Methylation of Lysine 9 in Histone H3 Directs Alternative Modes of Highly Dynamic Interaction of Heterochromatin Protein hHP1β with the Nucleosome *. Journal of Biological Chemistry 287, 33756–33765. https://doi.org/10.1074/jbc.M112.390849

Musselman, C.A., Kutateladze, T.G., 2021. Characterization of functional disordered regions within chromatin-associated proteins. iScience 24, 102070. https://doi.org/10.1016/j.isci.2021.102070

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., Natoli, G., 2013. Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell 152, 157–171. https://doi.org/10.1016/j.cell.2012.12.018

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., O'Shea, C.C., 2017. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science 357, eaag0025. https://doi.org/10.1126/science.aag0025

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., Choudhary, J., 2010. An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. Cell Stem Cell. https://doi.org/10.1016/j.stem.2010.03.004

Pedersen, T., Kowenz-Leutz, E., Leutz, A., Nerlov, C., 2001. Cooperation between C/EBPα TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. Genes and Development. https://doi.org/10.1101/gad.209901

Perkel, J.M., Atchison, M.L., 1998. A Two-Step Mechanism for Recruitment of Pip by PU.1. The Journal of Immunology 160, 241–252.

Pham, T.-H., Minderjahn, J., Schmidl, C., Hoffmeister, H., Schmidhofer, S., Chen, W., Längst, G., Benner, C., Rehli, M., 2013. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. Nucleic Acids Research 41, 6391–6402. https://doi.org/10.1093/nar/gkt355

Piovesan, D., Necci, M., Escobedo, N., Monzon, A.M., Hatos, A., Mičetić, I., Quaglia, F., Paladin, L., Ramasamy, P., Dosztányi, Z., Vranken, W.F., Davey, N.E., Parisi, G., Fuxreiter, M., Tosatto, S.C.E., 2021. MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Research 49, D361–D367. https://doi.org/10.1093/nar/gkaa1058

Postnikov, Y.V., Bustin, M., 2016. Functional interplay between histone H1 and HMG proteins in chromatin. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, Histone H1 in gene expression and development 1859, 462–467. https://doi.org/10.1016/j.bbagrm.2015.10.006

Protter, D.S.W., Rao, B.S., Van Treeck, B., Lin, Y., Mizoue, L., Rosen, M.K., Parker, R., 2018. Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. Cell Reports 22, 1401–1412. https://doi.org/10.1016/j.celrep.2018.01.036

Pulecio, J., Nivet, E., Sancho-Martinez, I., Vitaloni, M., Guenechea, G., Xia, Y., Kurian, L., Dubova, I., Bueren, J., Laricchia-Robbio, L., Belmonte, J.C.I., 2014. Conversion of human fibroblasts into monocyte-like progenitor cells. Stem Cells. https://doi.org/10.1002/stem.1800

Pundhir, S., Bratt Lauridsen, F.K., Schuster, M.B., Jakobsen, J.S., Ge, Y., Schoof, E.M., Rapin, N., Waage, J., Hasemann, M.S., Porse, B.T., 2018. Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors. Cell Reports. https://doi.org/10.1016/j.celrep.2018.05.012

R: The R Project for Statistical Computing [WWW Document], n.d. URL https://www.r-project.org/ (accessed 7.12.22).

Ramachandran, A., Omar, M., Cheslock, P., Schnitzler, G.R., 2003. Linker histone H1 modulates nucleosome remodeling by human SWI/SNF. J Biol Chem 278, 48590–48601. https://doi.org/10.1074/jbc.M309033200

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., Cosma, M.P., 2015. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. Cell 160, 1145–1158. https://doi.org/10.1016/j.cell.2015.01.054

Riddell, J., Gazit, R., Garrison, B.S., Guo, G., Saadatpour, A., Mandal, P.K., Ebina, W., Volchkov, P., Yuan, G.C., Orkin, S.H., Rossi, D.J., 2014. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. Cell. https://doi.org/10.1016/j.cell.2014.04.006

Roberts, G.A., Ozkan, B., Gachulincová, I., O'Dwyer, M.R., Hall-Ponsele, E., Saxena, M., Robinson, P.J., Soufi, A., 2021. Dissecting OCT4 defines the role of nucleosome binding in pluripotency. Nat Cell Biol 23, 834–845. https://doi.org/10.1038/s41556-021-00727-5

Sandoval, G.J., Pulice, J.L., Pakula, H., Schenone, M., Takeda, D.Y., Pop, M., Boulay, G., Williamson, K.E., McBride, M.J., Pan, J., St Pierre, R., Hartman, E., Garraway, L.A., Carr, S.A., Rivera, M.N., Li, Z., Ronco, L., Hahn, W.C., Kadoch, C., 2018. Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis. Mol Cell 71, 554-566.e7. https://doi.org/10.1016/j.molcel.2018.06.040

Sardina, J.L., Collombet, S., Tian, T.V., Gómez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., Gut, I.G., Heath, S., Aranda, S., Di Croce, L., Hochedlinger, K., Thieffry, D., Graf, T., 2018. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell 23, 727-741.e9. https://doi.org/10.1016/j.stem.2018.08.016

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

Schwarz, P.M., Hansen, J.C., 1994. Formation and Stability of Higher Order Chromatin Structures: Contributions of the Histone Octamer. Journal of Biological Chemistry 269, 16284–9.

Scott, E.W., Simon, M.C., Anastasi, J., Singh, H., 1994a. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science. https://doi.org/10.1126/science.8079170

Scott, E.W., Simon, M.C., Anastasi, J., Singh, H., 1994b. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science. https://doi.org/10.1126/science.8079170

Sekiya, S., Suzuki, A., 2011. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature. https://doi.org/10.1038/nature10263

Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A.V., Zaret, K.S., 2009. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. 23, 804–809. https://doi.org/10.1101/gad.1775509

Sekiya, T., Zaret, K.S., 2007. Repression by Groucho/TLE/Grg Proteins: Genomic Site Recruitment Generates Compacted Chromatin In Vitro and Impairs Activator Binding In Vivo. Molecular Cell 28, 291–303. https://doi.org/10.1016/j.molcel.2007.10.002

Shimada, M., Chen, W.-Y., Nakadai, T., Onikubo, T., Guermah, M., Rhodes, D., Roeder, R.G., 2019. Gene-Specific H1 Eviction through a Transcriptional Activator→p300→NAP1→H1 Pathway. Molecular Cell 74, 268-283.e5. https://doi.org/10.1016/j.molcel.2019.02.016

Singh, H., Dekoter, R.P., Walsh, J.C., 1999. PU.1, a Shared Transcriptional Regulator of Lymphoid and Myeloid Cell Fates. Cold Spring Harbor Symposia on Quantitative Biology 64, 13–20. https://doi.org/10.1101/sqb.1999.64.13

Sinz, A., 2018. Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein–Protein Interactions: Where Are We Now and Where Should We Go from Here?, Angewandte Chemie - International Edition. https://doi.org/10.1002/anie.201709559

Sottini, A., Borgia, A., Borgia, M.B., Bugge, K., Nettels, D., Chowdhury, A., Heidarsson, P.O., Zosel, F., Best, R.B., Kragelund, B.B., Schuler, B., 2020. Polyelectrolyte interactions enable rapid association and dissociation in high-affinity disordered protein complexes. Nat Commun 11, 5736. https://doi.org/10.1038/s41467-020-18859-x

Soufi, A., Donahue, G., Zaret, K.S., 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151, 994–1004. https://doi.org/10.1016/j.cell.2012.09.045

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., Zaret, K.S., 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell. https://doi.org/10.1016/j.cell.2015.03.017

Sutherland, B.W., Toews, J., Kast, J., 2008. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions, Journal of Mass Spectrometry. https://doi.org/10.1002/jms.1415

Szabo, E., Rampalli, S., Risueño, R.M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M., Bhatia, M., 2010. Direct conversion of human fibroblasts to multilineage blood progenitors. Nature. https://doi.org/10.1038/nature09591

Tagore, M., McAndrew, M.J., Gjidoda, A., Floer, M., 2015. The Lineage-Specific Transcription Factor PU.1 Prevents Polycomb-Mediated Heterochromatin Formation at Macrophage-Specific Genes. Molecular and Cellular Biology. https://doi.org/10.1128/mcb.00027-15

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell. https://doi.org/10.1016/j.cell.2006.07.024

Takizawa, Y., Tanaka, H., Machida, S., Koyama, M., Maehara, K., Ohkawa, Y., Wade, P.A., Wolf, M., Kurumizaka, H., 2018. Cryo-EM structure of the nucleosome containing the ALB1 enhancer DNA sequence. Open Biol 8, 170255. https://doi.org/10.1098/rsob.170255

Tanaka, H., Takizawa, Y., Takaku, M., Kato, D., Kumagawa, Y., Grimm, S.A., Wade, P.A., Kurumizaka, H., 2020. Interaction of the pioneer transcription factor GATA3 with nucleosomes. Nat Commun 11, 1–10. https://doi.org/10.1038/s41467-020-17959-y

Tanaka, Y., Tawaramoto-Sasanuma, M., Kawaguchi, S., Ohta, T., Yoda, K., Kurumizaka, H., Yokoyama, S., 2004. Expression and purification of recombinant human histones. Methods 33, 3–11. https://doi.org/10.1016/j.ymeth.2003.10.024

Tang, X., Li, T., Liu, S., Wisniewski, J., Zheng, Q., Rong, Y., Lavis, L.D., Wu, C., 2022. Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. Nat Struct Mol Biol 29, 665–676. https://doi.org/10.1038/s41594-022-00800-z

Tsukiyama, T., Becker, P.B., Wu, C., 1994. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 367, 525–532. https://doi.org/10.1038/367525a0

Ungerbäck, J., Hosokawa, H., Wang, X., Strid, T., Williams, B.A., Sigvardsson, M., Rothenberg, E.V., 2018. Pioneering, chromatin remodeling, and epigenetic constraint in early T-cell gene regulation by SPI1 (PU.1). Genome Research 28, 1508–1519. https://doi.org/10.1101/gr.231423.117

Van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J.L., Thieffry, D., Beato, M., Shi, Y., Bonifer, C., Graf, T., 2015a. C/EBPα Activates Pre-existing and de Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. Stem Cell Reports. https://doi.org/10.1016/j.stemcr.2015.06.007

Van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J.L., Thieffry, D., Beato, M., Shi, Y., Bonifer, C., Graf, T., 2015b. C/EBPα Activates Pre-existing and de Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. Stem Cell Reports. https://doi.org/10.1016/j.stemcr.2015.06.007

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., Wernig, M., 2010. Direct conversion of fibroblasts to functional neurons by defined factors. Nature 463, 1035–1041. https://doi.org/10.1038/nature08797

Wang, D., Diao, H., Getzler, A.J., Rogal, W., Frederick, M.A., Milner, J., Yu, B., Crotty, S., Goldrath, A.W., Pipkin, M.E., 2018. The Transcription Factor Runx3 Establishes Chromatin Accessibility of cis-Regulatory Landscapes that Drive Memory Cytotoxic T Lymphocyte Formation. Immunity. https://doi.org/10.1016/j.immuni.2018.03.028

Wang, Y., Zolotarev, N., Yang, C.-Y., Rambold, A., Mittler, G., Grosschedl, R., 2020. A Prion-like Domain in Transcription Factor EBF1 Promotes Phase Separation and Enables B Cell Programming of Progenitor Chromatin. Immunity 53, 1151-1167.e6. https://doi.org/10.1016/j.immuni.2020.10.009

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., Drechsel, D., Martynoga, B., Castro, D.S., Webb, A.E., Südhof, T.C., Brunet, A., Guillemot, F., Chang, H.Y., Wernig, M., 2013. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell 155, 621–635. https://doi.org/10.1016/j.cell.2013.09.028

Weiss, M.A., 2001. Floppy SOX: Mutual Induced Fit in HMG (High-Mobility Group) Box-DNA Recognition. Molecular Endocrinology 15, 353–362. https://doi.org/10.1210/mend.15.3.0617

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.Da., François, R., Grolemund, G., Hayes, A., Henry, L.A., Hester, J.J., Kuhn, M., Pedersen, T.Q., Miller, E., Bache, S., Müller, K., Ooms, J.C.L., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C.O., Woo, K.H., Yutani, H., 2019. Welcome to the Tidyverse. Journal of Open Source Software. https://doi.org/10.21105/joss.01686

Willcockson, M.A., Healton, S.E., Weiss, C.N., Bartholdy, B.A., Botbol, Y., Mishra, L.N., Sidhwani, D.S., Wilson, T.J., Pinto, H.B., Maron, M.I., Skalina, K.A., Toro, L.N., Zhao, J., Lee, C.-H., Hou, H., Yusufova, N., Meydan, C., Osunsade, A., David, Y., Cesarman, E., Melnick, A.M., Sidoli, S., Garcia, B.A., Edelmann, W., Macian, F., Skoultchi, A.I., 2021. H1 histones control the epigenetic landscape by local chromatin compaction. Nature 589, 293–298. https://doi.org/10.1038/s41586-020-3032-z

Woodcock, C.L., Skoultchi, A.I., Fan, Y., 2006. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome Res 14, 17–25. https://doi.org/10.1007/s10577-005-1024-3

Xhani, S., Lee, S., Kim, H.M., Wang, S., Esaki, S., Ha, V.L.T., Khanezarrin, M., Fernandez, G.L., Albrecht, A.V., Aramini, J.M., Germann, M.W., Poon, G.M.K., 2020. Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1. Science Advances 6, eaay3178. https://doi.org/10.1126/sciadv.aay3178

Xiao, L., Parolia, A., Qiao, Y., Bawa, P., Eyunni, S., Mannan, R., Carson, S.E., Chang, Y., Wang, X., Zhang, Y., Vo, J.N., Kregel, S., Simko, S.A., Delekta, A.D., Jaber, M., Zheng, H., Apel, I.J., McMurry, L., Su, F., Wang, R., Zelenka-Wang, S., Sasmal, S., Khare, L., Mukherjee, S., Abbineni, C., Aithal, K., Bhakta, M.S., Ghurye, J., Cao, X., Navone, N.M., Nesvizhskii, A.I., Mehra, R., Vaishampayan, U., Blanchette, M., Wang, Y., Samajdar, S., Ramachandra, M., Chinnaiyan, A.M., 2022. Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer. Nature 601, 434–439. https://doi.org/10.1038/s41586-021-04246-z

Xie, H., Ye, M., Feng, R., Graf, T., 2004a. Stepwise reprogramming of B cells into macrophages. Cell. https://doi.org/10.1016/S0092-8674(04)00419-2

Xie, H., Ye, M., Feng, R., Graf, T., 2004b. Stepwise reprogramming of B cells into macrophages. Cell. https://doi.org/10.1016/S0092-8674(04)00419-2

Zaret, K.S., Carroll, J.S., 2011. Pioneer transcription factors: Establishing competence for gene expression. Genes and Development. https://doi.org/10.1101/gad.176826.111

Zaret, K.S., Mango, S.E., 2016. Pioneer transcription factors, chromatin dynamics, and cell fate control. Current Opinion in Genetics and Development. https://doi.org/10.1016/j.gde.2015.12.003

Zaret, M.I.-D.K., 2014. Pioneer transcription factors in cell fate specification. Genes & Development. https://doi.org/10.1210/me.2014-1084

Zhang, Y., Cooke, M., Panjwani, S., Cao, K., Krauth, B., Ho, P.-Y., Medrzycki, M., Berhe, D.T., Pan, C., McDevitt, T.C., Fan, Y., 2012. Histone H1 Depletion Impairs Embryonic Stem Cell Differentiation. PLOS Genetics 8, e1002691. https://doi.org/10.1371/journal.pgen.1002691

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