

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Single-molecule sequencing to uncover high-resolution epigenome patterns

Permalink

<https://escholarship.org/uc/item/530075f6>

Author

Abdulhay, Nour J

Publication Date

2022

Peer reviewed|Thesis/dissertation

Single-molecule sequencing to uncover high-resolution epigenome patterns

by
Nour Abdulhay

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in
Biomedical Sciences

in the
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

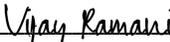
FDD44359FCC6487... Hani Goodarzi
Chair

DocuSigned by:

43B... Geeta Narlikar

DocuSigned by:

417... Barbara Panning

DocuSigned by:

12BA217F83A048B... Vijay Ramani

Committee Members

Copyright 2022

by

Nour Abdulhay

Dedicated to my parents, Suzanne and Gazi

Acknowledgments

I'd like to acknowledge everyone who supported me during this pursuit of knowledge, particularly during periods when I did not believe in myself. I am incredibly grateful for this experience, which challenged me significantly but ultimately allowed me to grow as a scientist. Pursuing a PhD has challenged my perspective on many things – but especially on how to engage in, conduct, and communicate scientific research in a positive manner. Graduate school has not just made me a better scientist but to my surprise, I believe a better friend, partner, and family member. This experience forced me to prioritize my mental health, which has allowed me to show up as a better version of myself and to adopt a growth mindset in the face of difficulties and setbacks (in science and in life).

I'd like to first acknowledge those who encouraged me to apply to graduate school in the first place. As a lab manager/research assistant in Vijay Sankaran's lab at Boston Children's Hospital, I had the privilege of learning from incredible scientists who encouraged me to pursue graduate school from the beginning – a special shout out to Ah Ram Kim, Anindita Basak, Aoi Wakabayashi, Claudia Fiorini, Jacob Ulirsch, and many others I worked with during my time there. My training at BCH laid a strong foundation of resilience upon which I relied heavily during graduate school.

At UCSF, I'd like to acknowledge my thesis committee members – Barbara Panning, Geeta Narlikar, and Hani Goodarzi - thank you for your genuine kindness and guidance throughout this experience. I'd also like to thank Geeta Narlikar and her entire lab for being such incredible research collaborators, and especially Laura Hsieh for being an amazing mentor in biochemistry as I navigated a field I was unfamiliar with. Thank you to Hani and Barbara for always encouraging me to stay positive in my research and life. Further, thank you to everyone in the Goodarzi lab, from RAs to my fellow BMS graduate students Bahar Zirak and Brian Woo, to post docs – you were my second lab home and I definitely couldn't have done this without all of you!

Next, I'd like to acknowledge my main advisor Vijay Ramani whose excitement and enthusiasm for the basic sciences knows no bounds. You've challenged me as a scientist and helped widen my perspective on how to approach scientific questions. Thank you for your patience, support, and encouragement throughout my PhD journey! It was really an honor to be a part of the early stages of your PI career as a Sandler fellow and to help the lab mature into its current form. Lastly, thank you to everyone in the Ramani lab from its earliest members to the most recent – Aidan Keith, Colin McNally, Siva Kasinathan, Dan Xu, Scott Nanda, Camille Moore, Coco Wu, and Megan Ostrowski – it has been inspiring to work with such dedicated and talented individuals. I'm excited to see the amazing work everyone does in the future!

Contributions

All published and unpublished work in this dissertation was conducted under Dr. Vijay Ramani.

Chapter 1 partially includes previously published material as seen in: NJ Abdulhay and V Ramani, At least two to tango: Choreographing chromatin through cooperative footprints, *Mol Cell*, 2021;81:8.

Chapter 2 of this dissertation includes previously published material as seen in: NJ Abdulhay, CP McNally et al., Massively multiplex single-molecule oligonucleosome footprinting, *eLife* 2020;9:e59404.

Chapter 3 of this dissertation contains material from a manuscript currently under review. It is currently available in open access journal: NJ Abdulhay, CP McNally, LJ Hsieh et al., Single-fiber nucleosome density shapes the regulatory output of a mammalian chromatin remodeling enzyme, bioRxiv, <https://doi.org/10.1101/2021.12.10.472156>

Nour Abdulhay

Abstract

Nearly all essential nuclear processes act on DNA packaged into series of nucleosomes termed chromatin fibers. However, our understanding of how these processes (e.g. DNA replication, RNA transcription, chromatin extrusion, nucleosome remodeling) actually occur on such fibers remains unresolved. Our current understanding of the beads-on-a-string arrangement of nucleosomes has been built largely on high-resolution sequence-agnostic imaging methods and sequence-resolved bulk biochemical techniques. To bridge the divide between these approaches, we present the single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA). SAMOSA is a high-throughput single-molecule sequencing method that combines adenine methyltransferase footprinting and single-molecule real-time DNA sequencing to natively and nondestructively measure nucleosome positions on individual chromatin fibers. SAMOSA data allows unbiased classification of single-molecular 'states' of nucleosome occupancy on individual chromatin fibers. We leverage this to estimate nucleosome regularity and spacing on single chromatin fibers genome-wide, at predicted transcription factor binding motifs, and across human epigenomic domains. Our analyses suggest that chromatin is comprised of both regular and irregular single-molecular oligonucleosome patterns that differ subtly in their relative abundance across epigenomic domains. This irregularity is particularly striking in constitutive heterochromatin, which has typically been viewed as a conformationally static entity. Our proof-of-concept study provides a powerful new methodology for studying nucleosome organization at a previously intractable resolution and offers up new avenues for modeling and visualizing higher order chromatin structure.

ATP-dependent chromatin remodelers are one of the major regulators of patterns within the epigenome. As a follow-up from our proof-of-concept study, we developed SAMOSA-ChAAT, a massively multiplex single-molecule footprinting platform to map the primary structure of individual,

precisely-reconstituted chromatin templates subjected to virtually any chromatin-associated reaction. As proof-of-concept, we apply SAMOSA-ChAAT to study ATP-dependent chromatin remodeling by the essential imitation switch (ISWI) ATPase SNF2h, whose mechanism-of-action remains contentious. Using our approach, we discover that SNF2h operates as a density-dependent, length-sensing nucleosome sliding enzyme, whose ability to decrease or increase DNA accessibility depends on single-fiber nucleosome density. We validate our *in vitro* findings with single-fiber accessibility measurements *in vivo*, finding that the regulatory ‘mode’ of SNF2h-containing complexes (*i.e.* ‘opening’ vs. ‘closing’ chromatin) is dictated by the underlying nucleosome-density of individual chromatin fibers: at canonically-defined heterochromatin, SNF2h generates evenly-spaced nucleosome arrays of multiple nucleosome repeat lengths; at SNF2h-dependent accessible sites, the enzyme slides nucleosomes to increase accessibility of motifs for the essential transcription factor CTCF. Our approach and data demonstrate, for the first time, how chromatin remodelers can effectively sense nucleosome density to induce diametrically-opposed regulatory effects within the nucleus. More generally, our novel approach promises molecularly-precise views of any of the essential processes shaping nuclear physiology.

Table of Contents

Chapter 1: Chromatin patterns dictate gene expression	1
1.1 Introduction	2
1.2 Anatomy of the nucleosome, the fundamental repeating subunit of chromatin.....	2
1.3 Nucleosome properties distinctly decorate chromatin	5
<i>Contributions of non-canonical histone variants</i>	<i>6</i>
<i>Mechanisms of histone turnover during DNA replication.....</i>	<i>7</i>
<i>Mechanisms of histone turnover throughout the cell cycle</i>	<i>8</i>
<i>Histone patterns organize the genome and dictate gene regulation</i>	<i>10</i>
1.4 ATP-dependent chromatin remodelers modulate chromatin to dictate gene expression.....	12
<i>Major ATP-dependent chromatin remodelers</i>	<i>12</i>
<i>Mammalian SWI/SNF functions.....</i>	<i>13</i>
<i>ISWI generates nucleosome arrays.....</i>	<i>14</i>
<i>ACF has dual roles to tune transcription</i>	<i>16</i>
1.5 Transcription factor cooperativity promotes dynamic binding at regulatory motifs.....	17
1.6 Epigenetics and chromatin are inextricably linked to disease.....	20
<i>Brief review: epigenetics in cancer</i>	<i>20</i>
<i>ATP-dependent chromatin remodelers in cancer</i>	<i>21</i>
<i>Oncohistones within histone H3 disrupt Polycomb-repressed chromatin.....</i>	<i>23</i>
1.7 Current toolkit to map the epigenome	25
<i>The development of 2nd generation NGS tools</i>	<i>25</i>
<i>Common methods to map chromatin</i>	<i>26</i>
1.8 Harnessing third-generation sequencing tools to study chromatin	27

<i>Harnessing PacBio SMRT sequencing to study chromatin at high resolution</i>	29
<i>Harnessing SAMOSA to determine how SNF2h remodels single-molecule, long templates</i>	31
Chapter 2: Massively multiplex single-molecule oligonucleosome footprinting	56
2.1 Abstract	57
2.2 Current methods to study nucleosome positioning and their limitations	57
2.3 Applying long-read sequencing to detect genome patterns at high resolution	59
2.4 Single-molecule real-time sequencing of adenine-methylated chromatin captures nucleosome footprints	60
2.5 SAMOSA captures regular nucleosome-DNA interactions in vivo through nuclease-cleavage and adenine-methylation simultaneously	62
2.6 SAMOSA enables unbiased classification of chromatin fibers on the basis of regularity and nucleosome repeat length	63
2.7 SAMOSA captures the transient nucleosome occupancy of transcription factor binding motifs	64
2.8 Heterogeneous oligonucleosome patterns comprise human epigenomic domains	66
2.9 Discussion	69
2.10 Figures	72
2.11 Supplementary Figures	77
2.12 Materials and Methods	86
2.13 Acknowledgments and Additional Information	98

Chapter 3: Single-fiber nucleosome density dictates the regulatory mode of a chromatin

remodeler.....	125
3.1 Abstract.....	126
3.2 Bulk average sequencing techniques are unable to measure single chromatin reaction states	127
3.3 Overview: Applying SAMOSA to study single-molecule reactions of ISWI remodeler reactions.....	127
3.4 Single-molecule footprinting of intact chromatin fibers reconstituted on genomic sequences.....	129
3.5 SAMOSA-ChAAT captures single-fiber chromatin remodeling reaction outcomes.....	131
3.6 SNF2h does not preferentially clamp trinucleosomes on individual fibers	133
3.7 Density-dependent SNF2h remodeling generates chromatin fibers with a range of regular spacings.....	134
3.8 SNF2h remodeling modulates motif site exposure frequency in a nucleosome density and sequence-dependent manner.....	136
3.9 SNF2h loss in vivo leads to bidirectional, domain-specific shifts in chromatin fiber structure.....	137
3.10 SNF2h tunes heterogeneous fiber usage patterns to specify bulk chromatin accessibility..	138
3.11 Discussion	140
3.13 Supplementary Figures	153
3.14 Supplementary Tables.....	165

3.15 Materials & Methods.....	175
Chapter 4: Future Directions	219
4.1 Multi-omic techniques paired with SAMOSA.....	220
<i>Targeted genomic capture with SAMOSA</i>	<i>221</i>
<i>Chromosome conformation capture with SAMOSA</i>	<i>222</i>
<i>Mapping nascent chromatin with SAMOSA</i>	<i>224</i>

List of Figures

Chapter 2 Figures

Figure 2. 1: Overview of the single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA).....	72
Figure 2. 2: In vivo SAMOSA captures oligonucleosome structure by combining MNase digestion of chromatin with adenine methylation footprinting.....	73
Figure 2. 3: SAMOSA reveals distribution of oligonucleosome patterns genome-wide.....	74
Figure 2. 4: SAMOSA captures bulk and single-molecule evidence of transcription factor-DNA interaction simultaneously via two orthogonal molecular signals.....	75
Figure 2. 5: Human epigenomic states are punctuated by specific oligonucleosome patterns.....	76

Chapter 2 Supplementary Figures

Supplementary Figure 2. 1: Quality control of in vitro nucleosome arrays assembled through salt-gradient dialysis.....	77
Supplementary Figure 2. 2: Mean raw and quantile normalized interpulse durations for in vitro SAMOSA experiments.....	78
Supplementary Figure 2. 3: Adenine methylation by the EcoGII enzyme is specific to accessible adenines and is protected against by the nucleosome.....	79
Supplementary Figure 2. 4: Average linker methylation and individually called dyad positions are qualitatively similar across the length of the nonanucleosomal array molecule.....	80
Supplementary Figure 2. 5: Unmethylated and fully methylated array DNA does not display the same periodic patterning of modified bases seen in methylated chromatin.....	81
Supplementary Figure 2. 6: Coverage enrichment of SAMOSA versus other data types at ChromHMM annotated genomic regions.....	82

Supplementary Figure 2. 7: Further characterization of clustered footprinted molecules.	83
Supplementary Figure 2. 8: Cluster sizes and numbers of motif-containing molecules for each transcription factor chosen for study.	84
Supplementary Figure 2. 9: Cluster enrichment for each transcription factor studied.....	85

Chapter 3 Figures

Figure 3. 1: SAMOSA-ChAAT enables massively multiplex dissection of single-fiber nucleosome positioning on in vitro reconstituted genomic chromatin fibers.....	145
Figure 3. 2: SAMOSA-ChAAT reveals chromatin remodeling outcomes at single-fiber resolution.	146
Figure 3. 3: SNF2h does not fix trinucleosomal spacings on individual chromatin fibers.	147
Figure 3. 4: Autocorrelation analyses reveal that chromatin density influences the heterogeneous reaction outcomes of SNF2h remodeling.	148
Figure 3. 5: SNF2h remodeling influences motif site exposure in a nucleosome-density and sequence-dependent manner.....	149
Figure 3. 6: Mapping the <i>in vivo</i> consequences of SNF2h remodeling in murine embryonic stem cells at single fiber resolution.	150
Figure 3. 7: Exploring SNF2h-mediated chromatin closing and opening by integrating ATAC-seq and SAMOSA data.	151
Figure 3. 8: A model of SNF2h-mediated chromatin regulation based on results of this study.	152

Chapter 3 Supplementary Figures

Supplementary Figure 3. 1: Computational pipeline for inferring DNA accessibility from measured inter-pulse distance (IPD).....	153
Supplementary Figure 3. 2: SAMOSA-ChAAT generalizes to other genomic sequences of interest.	154

Supplementary Figure 3. 3: SAMOSA-ChAAT is accurate and reproducible.....	155
Supplementary Figure 3. 4: Reproducibility of SAMOSA-ChAAT remodeling experiments and horizon plots for all catalytic conditions tested.....	156
Supplementary Figure 3. 5: Nucleosome spacing correlation is slightly impacted by chromatin remodeling, and dinucleosome distances are not fixed by SNF2h remodeling.....	157
Supplementary Figure 3. 6: SNF2h remodeling alters the probability of observing specific chromatin fiber structures.	158
Supplementary Figure 3. 7: Autocorrelation clustering analysis for fiber S2.....	159
Supplementary Figure 3. 8: An improved, in situ SAMOSA assay for profiling single-fiber chromatin structure in vivo.....	160
Supplementary Figure 3. 9: Reproducibility of <i>in vivo</i> SAMOSA data from knockout, addback, and E14 mESCs.....	161
Supplementary Figure 3. 10: Relative abundance of each chromatin fiber type at differentially accessible ATAC-seq peaks.	162
Supplementary Figure 3. 11: Single-fiber nucleosome occupancy patterns around predicted Ctf binding sites in vivo.....	163
Supplementary Figure 3. 12: Cutoffs for counting nucleosomes in inaccessible regions based on length.	164

Lists of Tables

Supplementary Table 3. 1: Summary of sequencing depths for all experiments performed in this study. 165

Supplementary Table 3. 2: Summary of average nucleosomes and standard deviation for all in vitro experiments..... 172

List of Abbreviations

5mC = 5-methylcytosine (epigenetic cytosine methyl modification)

Bp = basepair

CpG = addition of methyl group to C5 position of cytosines in a CpG dinucleotide context

CCS = circular consensus sequence

CLR = Continuous Long Reads

CRC = chromatin remodeling complex

hESC / hiPSC = human embryonic stem cell / human induced pluripotent stem cell

ENCODE = The Encyclopedia of DNA Elements

GRF = gene regulatory factor

HAND/SANT/SLIDE = HSS (domain of ISWI)

HAT = histone acetyl transferase

HDAC = histone deacetylase

ID = internucleosomal distance

IPD = interpulse duration

ISWI = Imitation Switch (a CRC)

m6dA = N-6-methyldeoxyadenosine (epigenetic adenine methyl modification)

mESC = mouse embryonic stem cell

MNase = micrococcal nuclease

NCP = nucleosome core particle

NGS = Next-generation sequencing

NRL = Nucleosome Repeat Length

PacBio = Pacific Biosciences

PTM = post-translational modification

RC = replication-coupled (nucleosome assembly)

RI = replication-independent (nucleosome assembly)

RSC = Remodeling the structure of chromatin (a CRC)

SAMOSa = Single-molecule Adenine Methylated Oligonucleosome Sequencing Assay

SAMOSa-ChAAT = Single-molecule Adenine Methylated Oligonucleosome Sequencing Assay,
Chromatin Accessibility of Assembled Templates

SHL = super-helix location

SMRT = Single-molecule real time

SWI/SNF = SWItch/Sucrose Non-Fermentable (a CRC)

TF = transcription factor

ZMW = zero-mode waveguide

Chapter 1: Chromatin patterns dictate gene expression

1.1 Introduction

At the core of the epigenome, chromatin provides a blueprint to regulate all of gene expression. Inside a single eukaryotic cell, nuclear DNA must undergo a series of hierarchical steps to compact nearly 2m of DNA into the nucleus. This compaction is made possible through nucleosomes, the fundamental repeating unit of chromatin, where ~147 DNA base pairs are wound ~1.65 times around the histone octamer core in a left-handed superhelix. Through nucleosome-assisted compaction, nuclear DNA is compacted inside the nucleus by many thousands of nucleosomes, which are separated by varying amounts of linker DNA, approximately 20-80 base pairs (Cutter and Hayes, 2015). Nucleosomes both passively and actively template the majority of nuclear interactions essential to life, including bookmarking active and repressed chromosomal compartments via post-translational modifications (Zhou et al., 2010), determining target site access for transcription factors (Spitz and Furlong, 2012), and safeguarding the genome from mutational agents (Papamichos-Chronakis and Peterson, 2012).

1.2 Anatomy of the nucleosome, the fundamental repeating subunit of chromatin

Early Electron Microscopy first visualized the typical ‘beads-on-a-string’ view of uncondensed chromatin, where nucleosomes are separated by linker DNA (Kornberg, 1974; Olins and Olins, 1974). This was followed by low-resolution X-Ray crystallization studies (Finch et al., 1977; Rhodes et al., 1989; Richmond et al., 1984) which were essential in revealing the structure of a single nucleosome core particle (NCP), which consists of an octamer of highly conserved histone proteins complexed together. More recently, higher resolution crystallography (Luger et al., 1997), single-particle Cryo-EM (Armache et al., 2019; Nakane et al., 2018) , and FRET (Deindl et al., 2013; Yang et al., 2006a) have provided unprecedented and detailed views into NCP structure and dynamics. Since then, high-resolution crystal structures of NCPs from many species and those containing non-canonical histone variant structure have been resolved.

Through decades of now improved imaging and biochemical structural assays, we now understand that formation of a single NCP occurs when two individual H3-H4 histone heterodimers dimerize to form a tetramer, which interacts with the central part of the DNA superhelix. Next, two H2A-H2B dimers complex with the exposed portions of the (H3-H4)₂ tetramer to bind DNA at the last half turn of the superhelix to form a more stable, canonical NCP (Richmond et al., 1984). Core histones have flexible and unstructured tails which protrude out of the NCP. All four core histones have N-terminal histone tails in addition to two C-terminal histone tails of H2A. These histone tails are highly basic and frequent subject to a wide array of reversible, post-translational modifications (PTMs), including but not limited to phosphorylation, acetylation, and methylation of serine, lysine, and arginine residues (Bannister and Kouzarides, 2011; Rothbart and Strahl, 2014; Zhou et al., 2010). Many *in vitro* biochemical studies of tail-less NCPs demonstrate their importance for maintaining nucleosome stability and structure, and even in the formation of higher-order chromatin states (Iwasaki et al., 2013). Similar biochemical studies also demonstrate that histone tails may bind to nucleosomal DNA, linker DNA, and even the acidic patches of neighboring nucleosomes (Angelov et al., 2001; Mutskov et al., 1998; Yang et al., 2007).

How exactly do these proteins complex on extremely bent DNA, and what are the determinants of nucleosome positioning? Core histones consist of a high percentage of small, positively charged proteins, particularly Lys⁺ and Arg⁺ amino acids in the histone-fold domain, which promote histone dimerization formation and allow negatively charged DNA to effectively wrap around the histone octamer (Arents and Moudrianakis, 1995). There are fourteen superhelix locations (SHLs) where the major groove of the DNA backbone interacts with histones, in which SHL 0 resides at the nucleosomal dyad. These cumulative interactions comprise up to >100 direct atomic interactions (Luger et al., 1997).

Despite these relatively stable atomic interactions, the nucleosome-DNA relationship itself is very dynamic, particularly through the assistance of ATP-dependent remodeling machinery (covered in section 1.4). More recent studies therefore suggest a dynamic model of non-canonical nucleosome states existing *in vivo*. ‘Fragile’ nucleosomes were initially proposed as nucleosomes with higher enzymatic digestion

sensitivity, often at the -1 nucleosome position adjacent to nucleosome-free region (NFR) and upstream of the transcription start site (TSS) (Lai and Pugh, 2017). The $(H3-H4)_2$ tetramer alone is capable of creating a nucleosome-like structure, protecting ~ 80 bp of DNA from micrococcal nuclease (MNase) digestion (Richmond et al., 1984). These non-canonical nucleosomes may be intermediates related to transcription or replication, or perhaps subnucleosomes dynamically existing throughout the cell-cycle to support all DNA-dependent processes. Further studies are needed to understand the specific role of subnucleosomes and/or nucleosome intermediates *in vivo*.

Sequence-dependent determinants of nucleosome positioning are also important to understand, as nucleosomes dictate access to underlying gene regulatory sites. Beyond the inherent amino acid properties which promote NCP formation, underlying DNA structural and sequence properties dictate non-random nucleosome-positioning both *in vitro* and *in vivo* (Kaplan et al., 2009). Biochemical reconstitution studies demonstrate that histone octamers have a preference for DNA sequences with more inherent flexibility (Luger et al., 1997). Early studies show that nucleosomal sequences contain quasi-periodic nucleotide distribution, reflecting the anisotropic (direction-dependent) bending preferences of DNA (Lowary and Widom, 1998a). For instance, the A/T-rich minor grooves of the DNA helix faces inward toward the histone-fold, while G/C-rich minor grooves preferentially face outward (Drew and Travers, 1985). Sequences conferring DNA flexibility promote histone octamer positioning and allow for the phenomenon of nucleosomal phasing of arrays to occur. One of the major deterrents of nucleosome-positioning is poly(deoxyadenylic-deoxythymidylic) (poly(dA:dT)) tracts, likely due to their inherent bending properties, which are resistant to the sharp bends conducive to NCP formation (Segal and Widom, 2009). However, poly(dA:dT) are highly abundant in eukaryotes, and may play an important functional role in nucleosome-free regions (NFRs) prior to transcription start sites (TSS) (Suter et al., 2000).

The most common DNA sequence utilized to study nucleosome dynamics *in vitro* for biochemical reconstitution studies is the ‘Widom-601’ sequence. The 601 sequence was originally isolated by SELEX (Systematic Evolution of Ligands by Exponential Enrichment) via a pool of 10^{12} chemically randomized

artificial 220bp fragments competitively salt dialyzed with chicken histone octamer. This yielded a rank of the highest affinity positions, with 601 having the highest affinity for nucleosome positioning (Lowary and Widom, 1998a). The 601 sequence along with other highest ranked sequences were all AT-rich at the DNA helical periods. Other sequences commonly used for nucleosome positioning include variations of the 601 sequence, 5S ribosomal DNA, and MMTV 3' LTR (Donehower et al., 1981; Flaus, 2012; Simpson and Stafford, 1983).

To facilitate compaction in the nucleus, nucleosomes are often accompanied by linker histone H1, which associates with the superhelix both at the entrance or exit of the NCP, likely in close proximity to H2A molecules on the outside of the NCP. This is termed a chromatosome (Simpson, 1978). Interestingly, phosphorylation of histone H1 increases as cells enter mitosis, demonstrating a cell-cycle dependent role for H1 in chromatin condensation or decondensation (Harshman et al., 2013). It was originally thought that nucleosomal arrays undergo further compaction, via chromatosomes, in mitotic and interphase nuclei into a 30-nm chromatin fiber, known as the solenoid model (Finch and Klug, 1976). However, the existence of the 30-nm chromatin fiber *in vivo* was originally contested, - cryo-EM of vitreous sections (CEMOVIS) performed on mitotic cells showed no evidence of this higher-order chromatin structure (Dubochet et al., 1988). More recent, higher-resolution methods provide further evidence contesting the static, 30-nm chromatin fibers in both mitotic and interphase chromatin (Cai et al., 2018; Maeshima et al., 2010; Ou et al., 2017). These recent studies demonstrate various *in vivo* models for higher-order chromatin formation, none of which include regular 30-nm fibers, and instead suggest a heterogenous and dynamical compaction model of chromatin compaction.

1.3 Nucleosome properties distinctly decorate chromatin

During various stages of nucleosome-mediated DNA compaction into chromatin, underlying DNA must be made available via histone-dependent and histone-independent mechanisms. There are several key factors that assist gene regulatory factors (GRFs) in overcoming chromatin's inherently gene repressive

environment: non-canonical histone deposition, post-translational modification (PTM) of histones, ATP-dependent chromatin remodelers, and transcription factor cooperativity.

Contributions of non-canonical histone variants

What are the major mechanisms that ensure nucleosomes are disassembled and reassembled in the proper orientation and space? Since a large quantity of histones are required for assembly onto newly synthesized DNA, a majority of histone synthesis is coupled to DNA replication in the S Phase, typically forming the canonical NCP. There are, however, several non-canonical histone variants which contribute to the nucleosome population in either a replication-coupled (RC) or replication-independent (RI) manner. These histone variants vary in their sequence homology percentages and many are conserved throughout eukaryotes. They typically have highly specialized function, including establishing genomic boundaries and cell-type specific differentiation (Henikoff and Ahmad, 2005; Henikoff and Smith, 2015).

Histone H3 variants have been widely studied for their roles as markers of specialized chromatin. For instance, H3.3 is an essential and highly conserved replication-independent variant, which is deposited throughout the cell-cycle (H3 and H3.3 only differ by four amino acids). It is assembled at actively transcribed genes, promoters, and gene regulatory elements (Ahmad and Henikoff, 2002). CENP-A shares ~60% homology to H3 and is deposited specifically at centromeres. CENP-A containing centromeres are essential for kinetochore assembly, the area of which microtubules of the mitotic spindle fibers attach to chromosomes during cell division (Henikoff and Smith, 2015).

There are many H2A variants which have diverse and often conflicting roles. H2A.Z has been linked to both activating and repressive genomic regions (Weber and Henikoff, 2014). H2A.X assembly and phosphorylation is established early at DNA double-stranded breaks, making it an essential player for the DNA damage response, and other repair-associated remodeling and transcriptional pathways. Other H2A variants macroH2A and H2A.B carry out conflicting roles in transcriptional repression and activation,

respectively (Henikoff and Smith, 2015). While H2A and H3 variants have been the most widely studied, the remaining histone subunits do demonstrate chromatin specialization. H2B variants are understudied, but TH2B and others seem to play a key role in male germ cell differentiation (Montellier et al., 2013). Lastly, it was originally thought that no H4 histone variants existed, but recently a rare variant H4G was discovered which regulates ribosomal DNA transcription in breast cancer (Long et al., 2019).

Mechanisms of histone turnover during DNA replication

Chromatin regulates underlying accessibility of DNA for all essential processes in the nucleus including transcription, replication, DNA repair, and recombination. During many of these processes, nucleosomes must be constantly reshuffled through dynamic disassembly, reassembly, as well as histone exchange. Histone turnover is the process of complete or partial removal of the nucleosome, followed by deposition of newly synthesized histones (Henikoff and Smith, 2015).

During DNA replication for instance, chromatin must be faithfully duplicated into daughter cells. First, nucleosomes must be disassembled ahead of the replication fork to provide access to the replisome replication machinery, and then promptly reassembled behind the fork onto new daughter strands. Unsurprisingly then, the process of RC nucleosome assembly during DNA replication involves the highly orchestrated assembly of both parental and newly synthesized histones, in concert with the replisome and its associated factors. Specialized histone chaperones, including but not limited to FACT (facilitates chromatin transaction) and CAF-1 (chromatin assembly factor-1) are essential to this process. As the replicative helicase CMG unwinds DNA, FACT interacts with the MCM2 (minichromosome maintenance 2) subunit of the helicase to cooperatively bind and evict histones (Smith and Stillman, 1991; Zhang et al., 2020).

The next key aspect of RC nucleosome assembly is deposition of both recycled parental and newly synthesized histones. Further studies are needed fully understand the mechanisms of how histones are

faithfully evicted, passed to segregated daughter strands, and replaced with either parental or naive histones. We do know that PCNA, an essential DNA replication sliding clamp, recruits CAF-1 to the replication fork to mark nascent chromatin that is ready for nucleosome assembly. Parental (H3-H4)₂ tetramer are more frequently recycled than H2A-H2B dimers and preserve their PTMs, which is a key method for epigenetic inheritance. As for newly synthesized histones, they often contain H4K5 and H4K12 di-acetylation, which may be important for genome stability (Serra-Cardona and Zhang, 2018; Stewart-Morgan et al., 2020). Lastly, it is thought that INO80 (covered below) promotes replication fork stability through the recovery of stalled replication forks, as well as evicts misplaced H2A.Z. Lastly, the α -thalassemia/mental retardation X-linked (ATR-X) chromatin remodeling factor, implicated in the DNA Damage response (DDR) pathway, is also essential for replication stress-tolerance. ATR-X has been shown to deposit H3.3 variant, perhaps through a maintenance mechanism at newly replicated DNA (Khurana and Oberdoerffer, 2015).

It is now well known that the DNA replication leading and lagging (Okazaki) strands are synthesized differently due to strand orientation and the ability to synthesize continuously or discontinuously, respectively. How then is histone deposition, particularly from the primary PTM-carrier (H3-H4)₂ tetramer, influenced by leading and lagging strand synthesis? While this mechanism is not well understood, it is thought that parental histones contribute equally to daughter strands. While the mechanism has been debated, studies suggest that H3-H4 tetramers are inherited intact in order to pass on PTMs, while parental H2A-H2B dimers segregate randomly to daughter cells (Stewart-Morgan et al., 2020; Xu et al., 2010). Further studies are needed to understand the extent and mechanism of parental contribution to daughter strands in DNA replication.

Mechanisms of histone turnover throughout the cell cycle

While the bulk of histone synthesis occurs in S phase, nucleosome dynamics are still necessary outside of DNA replication. How then does this occur throughout the cell cycle, such as when gene regulatory factors require underlying access for transcription, or during cell growth and differentiation?

During eukaryotic RNA Polymerase II (Pol II)-mediated transcription, the polymerase must overcome nucleosome roadblocks through careful nucleosome disassembly followed by prompt re-assembly (Li et al., 2007a). As a brief review, faithful RNA Pol II transcription of DNA into stable RNA for downstream protein synthesis involves initiation, elongation, and termination. Initiation involves the binding of RNA Pol II and other transcription factors to form the pre-initiation complex (PIC), where promoter sequence (and distal enhancer) binding occurs, DNA is denatured, and the open DNA bubble engages with the RNA Pol II active site for synthesis to begin. Pausing at promoter-proximal sites acts as an additional signal for the active elongation step, where RNA synthesis occurs. Lastly, termination occurs when RNA is released from both DNA and RNA Pol II.

How do RNA Pol II and associated factors access underlying DNA while maintaining local nucleosome density to prevent aberrant transcription? There are many players that contribute to histone displacement in transcription including histone chaperones, histone PTMs and variants, ATP-dependent nucleosome remodeling, and transcription factor cooperativity (Workman, 2006). Despite its role as the key RNA polymerase for mRNA synthesis, RNA Pol II cannot transcribe through nucleosomes *in vitro* except in hexasomes, a standard nucleosome missing one H2A/H2B dimer (Arimura et al., 2012). There has been much debate over the extent to which histones are displaced in transcription. In the context of RNA Pol II transcription, rapid histone exchange of the H2A-H2B dimer occurs (Kireeva et al., 2002; Scott and Campos, 2020). This may be due to stable hexasome association with DNA, which allows for the maintenance of local nucleosome density.

To further assist in these processes, the chromatin remodeler complex (CRC) SWI/SNF is initially recruited by transcriptional activators and is thought to evict histones, allowing formation into PIC. SWI/SNF contains a bromodomain that can bind to acetylated histone Lysines modified by SAGA, a histone acetyltransferase (HAT) (Carey et al., 2006). During elongation, the histone chaperon FACT, possibly through recruitment of the CRC RSC, simultaneously dissociates and redeposits H2A-H2B dimer while maintaining (H3-H4)₂ tetramer stability (Carey et al., 2006). Behind RNA Pol II, histone chaperones

FACT, HIRA, and SUPT6H maintain nucleosome reassembly (Scott and Campos, 2020). Overall, the manner and extent to which histones are displaced during RNA pol II transcription, particularly ATP-dependent chromatin remodelers, requires further investigation.

Histone patterns organize the genome and dictate gene regulation

As mentioned before, nucleosome positioning is influenced by underlying DNA structure and sequence. Nucleosome positioning and density is non-random, where promoters and enhancers are typically nucleosome-sparse, while transcribed genes are nucleosome-dense. Other mechanisms which dictate nucleosome positioning include histone variants and post-translational modification (PTMs). As previously mentioned, histone tails and even nucleosome globular domains are subject to frequent and reversible PTMs which is regulated by histone ‘writers’ and ‘erasers’ (Venkatesh and Workman, 2015). These processes are essential to marking genomic regions, contributing to epigenetic diversity and inheritance, and regulating cellular development and differentiation.

Histone variants may act as a marker for particular genomic regions. Most RNA Pol II genes contain a nucleosome-free region over the promoter, flanked by two well-positioned nucleosomes, termed the +1 and -1 nucleosome. These nucleosome-free regions contain TF binding occupancy sites and may be marked by histone variants. The flanking nucleosomes often contain H2A.Z variant, a marker for transcriptional activation (as well as silent gene promoters). As a mechanism for maintaining local nucleosome density, H3.3 variant is deposited over actively transcribed genes (Ahmad and Henikoff, 2002; Venkatesh and Workman, 2015).

While chromatin structure is constantly evolving in a cell-type and cell-cycle specific manner, there are several known demarcations between heterochromatin and euchromatin. Many histone PTMs are correlated with typically transcriptionally active (euchromatin, loose packaging of nucleosomes) or transcriptionally silent (heterochromatin, tight packaging of nucleosomes) genomic regions. For instance,

an active gene typically contains H3K4me2/3 at promoters and TSSs followed by H3K36me2/3 at downstream gene bodies. A majority of PTMs, with the exception of methylation, induce a change to the net charge of the nucleosome. Acetylation of histones may loosen histone-DNA binding properties, and thus contribute to the higher rates of histone exchange associated with active transcription (Kenzaki and Takada, 2015; Li et al., 2007a).

PTMs are incredibly diverse and many are associated with heterochromatic regions, - H3K9me2/3 and H3k27me2/3 are typically associated with repressed genes. While Lysine methylation is not typically associated with active transcription, there are several exceptions and a particular PTM may contribute to multiple chromatin outputs. Heterochromatin is further divided into constitutive and facultative heterochromatin, many regions of which contain satellite DNA. Satellite DNA, arrays of tandem repeats, are mainly found in heterochromatin or tightly packed regions of chromosomes. These regions are typically non-coding but may undergo transcription. The periodic A/T distribution and inherent DNA curvature of satellite DNA makes it conducive to heterochromatic packing.

Constitutive heterochromatin is majorly found within telomeres and centromeres, particularly pericentromeric heterochromatin. These regions are typically gene-poor and contain repetitive satellite DNA, hypoacetylated nucleosomes, and high levels of H3K9me3 mark. It is accompanied by heterochromatin protein 1 (HP1), an essential component to packaging heterochromatin (Sanulli et al., 2019). These regions remain condensed throughout the cell cycle, and are thought to be important for genome stability and chromosomal segregation (Barski et al., 2007). Facultative heterochromatin includes silenced genes, which may be expressed under certain conditions in cellular development or differentiation. These genes are typically hypoacetylated and marked by H3K27me2/3. Examples include inactive X chromosome, imprinted genes, or other transcriptionally repressed genes.

1.4 ATP-dependent chromatin remodelers modulate chromatin to dictate gene expression

Major ATP-dependent chromatin remodelers

Chromatin-remodeling complexes (CRCs) utilize ATP hydrolysis to alter histone-DNA contacts, through eviction, sliding, or assembly of nucleosomes. There are four major families of CRCs: switching defective/sucrose nonfermenting (SWI/SNF), imitation SWI (ISWI), inositol requiring 80 (INO80), and the nucleosome remodeling and deacetylation (NuRD)/Mi-2/chromodomain helicase DNA-binding (CHD). These CRCs are evolutionarily conserved, multi-subunit complexes which perform specialized DNA-dependent processes to modulate local regulatory DNA. While the ATPase confers ATP hydrolysis required activity, accessory subunits with distinct domains confer biological specificity, such as chromatin reader domains (SANT, bromodomains, chromodomains). While many of the remodelers have been studied with biochemical assays on simplified mononucleosomes *in vitro*, how remodelers affect nucleosomes within the context of higher order chromatin structure has yet to be fully elucidated.

The following sections will primarily focus on the two most-well studied CRCs: SWI/SNF and ISWI. However, I will give a brief overview of INO80 and CHD. The INO80 CRC plays a major role in DNA replication, transcription, and DNA repair pathways. Through its ATPase Ino80 and 14 other subunits, it has specialized nucleosome-modulating abilities, including histone exchange and sliding. INO80 can catalyze histone exchange H2A with variant H2A.Z, which is enriched at TSSs. Further, INO80 slides nucleosomes *in vitro* to position the +1 nucleosome at promoters, supporting its role in gene activation (Krietenstein et al., 2016). INO80 slides nucleosomes continuously up to 20bp and requires a minimum of ~33bp of linker DNA (Udugama et al., 2011). It senses flanking DNA length independently of its ATPase, unlike the ISWI remodeler spacing activity discussed below. The accessory subunits of INO80 are likely to have an important role in sliding at specific target sites (Zhou et al., 2018). For instance, INO80 is recruited to double-strand breaks, where the Nhp10 subunit interacts with phosphorylated gamma H2A.X (Udugama et al., 2011).

The CHD/NuRD complex has ATP-dependent remodeling activity similar to that of SWI/SNF, but also has a histone deacetylase. NuRD ATPase activity is conferred through the SWI/SNF family subclass Mi-2 proteins (or chromodomain-helicase-DNA-binding protein CHD3 or CHD4). NuRD contains histone deacetylase proteins HDAC1 and HDAC2, as well as the methyl-CpG-binding domain protein MBD2 of MBD3. Lastly, NuRD contains the histone-binding proteins RbAp46 and RbAp4 and the metastasis-associated proteins MTA (MTA1, 2, or 3); the functional roles of these particular subunits remain largely unknown. NuRD is thought to be involved in thyroid hormone receptor mediated transcriptional repression, but its mechanisms of action are not well understood (Denslow and Wade, 2007; Xue et al., 1998; Zhang and Li, 2011). It has been also suggested that NuRD facilitates Polycomb repression, acting in opposition to SWI/SNF (Bracken et al., 2019).

Mammalian SWI/SNF functions

Both SWI/SNF and ISWI share homology for a catalytic ATPase subunit, both of which are members of the SF2 superfamily of helicases. Despite this shared homology to carry out DNA translocation events, both CRCs carry out highly distinct DNA modulating mechanisms (Racki and Narlikar, 2008). SWI/SNF CRCs are highly-conserved, large multi-subunit complexes, all sharing homology to the yeast SWI2/SNF2 ATPase of yeast Remodeling the Structure of Chromatin (RSC) (Racki and Narlikar, 2008). There are three major classes of mammalian SWI/SNF (BAF), including BRG1/BRM-associated Factor (BAF), non-canonical BAF (ncBAF), and polybromo-associated BAF (PBAF). Brg1 (SMARCA4) is the common ATPase subunit and is the critical component needed to carry out ATP-dependent translocation of nucleosomes. Each BAF complex identified to date contains up to 15 subunits encoded from 29 different genes, performing highly specialized functions, including roles in DNA repair, RNA Polymerase II transcription, and chromosomal functions.

BAF complexes carry out many actions on nucleosomes including but not limited to altering positions of nucleosomes on DNA, generating nucleosomes containing loops, and partial or full eviction of nucleosomes. BAF is thought to play an important role in histone eviction and exchange at promoters and enhancers by opening chromatin for gene activation. Certain BAF complexes may be recruited to specific loci via their bromodomain which recognizes acetylated histones. This supports evidence showing that BAF opposes H3K27me2/3 Polycomb-repression (Bracken et al., 2019; Weber et al., 2021; Wilson et al., 2010).

ISWI generates nucleosome arrays

The ISWI family of remodelers, along with CHD1 and INO80, generates regularly spaced nucleosomes to facilitate formation of chromatin structure and proper gene expression. Similar to many CRCs, ISWI members play an essential role in DNA repair, transcription, and replication. All ISWI members consist of two ATPases, SNF2h or SNF2l, both of which contain the conserved DEAD/H helicase ATPase domain in their N-terminal domains for ATP hydrolysis, as well as a HAND/SANT/SLIDE (HSS) domain for substrate recognition and catalysis in their C-terminal domains to mediate DNA and histone binding. Overall, the ISWI family has various functions and target recognition abilities, dictated by accessory proteins within a given complex (Narlikar et al., 2013; Zhou et al., 2016). Several ISWI complexes are involved in nucleosome spacing, including ACF, CHRAC and RSF; other SNF2h-containing complexes include NoRC and WICH (Längst and Becker, 2001). Human ACF (ATP-dependent chromatin-assembly factor) is formed by SNF2h and the noncatalytic subunit ACF1/BAZ1A (Ito et al., 1997, 1999). hACF has long been used as a model to study the mechanisms of ISWI based nucleosome spacing, along with SNF2h alone to determine if the ATPase motor is sufficient to recapitulate similar modes of action, due to their small size and specific role in nucleosome spacing. Through a highly debated mechanism, SNF2h and ACF can both slide end positioned mononucleosomes to the center and evenly space nucleosome arrays, sensing up to 40bp and 60bp of linker DNA, respectively (Racki et al., 2009; Yang et al., 2006a).

A major question in the field of ATP-dependent remodelers has been whether SNF2h has a remodeler-specific mechanism to create nucleosome regularity. There are two conflicting models on how SNF2h may mechanistically achieve nucleosome regularity. Previous assays from the Korber lab examined bulk ISWI-remodeled, MNase digested arrays with varying histone density. Their findings suggested that SNF2h slides nucleosomes via a clamping mechanism or ‘density-independent’ model. In this model, SNF2h may create fixed internucleosomal distances (IDs), much like ruler, regardless of the underlying density of nucleosomes at its target regions (Lieleg et al., 2015a). An alternative model suggests that SNF2h instead uses its HSS domain to sense flanking DNA length on either side of a nucleosome and pushes nucleosomes towards the side with longer flanking DNA (Yang et al., 2006b). This indicates that instead SNF2h stochastically slides nucleosomes, depending on underlying nucleosome density at its target regions. Regardless of the mechanisms, it is unclear whether or not equal spacing is mediated directly through nucleosome-nucleosome contacts or indirectly through remodelers. Further, studies done thus far mainly rely on bulk techniques using micrococcal nuclease (MNase) digestion to measure to measure NRLs, which does not allow for resolution of these density-related models. A technique which allows resolution of single molecules on remodeler reactions would be highly beneficial for clarifying conflicting ISWI-spacing models, as well as for understanding other remodeler-mediated processes.

Despite its relatively well-established role, ISWI and all remodelers are sequence agnostic. How then, do ATP-dependent chromatin remodelers carry out their function at known target sites to achieve specialized remodeling outcomes? All CRCs sense well-established molecular cues on and around the NCP, including DNA modification, histone PTMs, flanking DNA length, and the H2A/H2B acidic patch. Several studies have been performed to understand the mechanism by which these ISWI-remodelers space nucleosomes, and in turn how this function contributes to the formation of chromatin at specific loci. Since SNF2h recapitulates the biochemical function of sliding nucleosomes for the ISWI remodelers, mechanistic studies have been performed with this enzyme alone to understand ISWI spacing. Two conserved regulatory regions from *Drosophila* ISWI were identified, which negatively regulate ATPase activity (AutoN) or the

coupling of hydrolysis to slide nucleosomes (NegC) (Clapier and Cairns, 2012). In this study, they discovered that in their respective ground states, AutoN inhibits ATPase activity while NegC domain inhibits ATPase coupling to translocate DNA in a manner highly sensitive to flanking DNA length (Clapier and Cairns, 2012).

Since these autoinhibitory regions act as ‘break modules’, in order to initiate remodeling, ISWI therefore requires positive regulation through activating epitopes within the nucleosome itself. It is well established that the ‘basic patch’ of the Histone H4 tail and extranucleosomal (linker) DNA, both positively regulate ISWI binding (Clapier et al., 2001). The HSS domain of SNF2h requires linker DNA to initiate this remodeling activity, and was suggested to regulate DNA translocation activity by acting antagonistically to NegC (Clapier and Cairns, 2012). Histone H4 is recognized by the second RecA lobe within the ATPase domain of SNF2h. The H4 tail is also important for compaction of nucleosome arrays, and associates with the acidic patch on neighboring nucleosomes to promote condensation of nucleosomes into higher chromatin fibers (Dorigo et al., 2004; Schalch et al., 2005). The acidic patch is a conserved, negatively charged surface formed by histones H2A/H2B and recently, studies have shown that its interaction with the AutoN and NegC modules relieve their autoinhibition. Acidic patch mutants reduce the ability of SNF2h to expose DNA to restriction enzymes as well as reduce the rate at which the mutant nucleosomes are centered on end-positioned nucleosomes compared to WT nucleosomes (Gamarra et al., 2018). Overall, it is not yet understood how these negative and positive elements contribute to nucleosome spacing on longer chromatin templates.

ACF has dual roles to tune transcription

In addition to its role in nucleosome spacing, ISWI members have specialized biological role in chromatin formation and may be associated with both transcriptional repression and activation *in vivo* (Chioda et al., 2010). ISWI can slide nucleosomes toward nucleosome depleted regions (NDRs) of promoters, preventing accessibility to transcription factors. ACF is thought to be essential for the formation

of compact heterochromatin, wherein regularly spaced arrays are a hallmark (Cuaycong and Elgin, 2001). Earlier studies in Acf embryos demonstrated its role in formation of repressive pericentric heterochromatin as well as polycomb-mediated silencing. The ACF1 subunit has been suggested to target pericentromeric heterochromatin, and was found to be required for efficient DNA replication through condensed chromatin (Collins et al., 2002). Conversely, ISWI has also been shown to promote transcription by mediating SNF2h-dependent accessibility of CTCF loci (Barisic et al., 2019a). This suggests that ISWI remodelers cooperate with CTCF at their binding motifs to promote both insulation at topologically associated domains and chromosomal folding.

1.5 Transcription factor cooperativity promotes dynamic binding at regulatory motifs

Through strict, yet flexible genomic compartmentalization, nucleosome positioning coordinates the accessibility of underlying DNA to *trans*-acting gene regulatory factors, enabling essential DNA transactions (Jiang and Pugh, 2009; Klemm et al., 2019; Luger et al., 1997). How, then, do TFs gain specific access to correct binding motifs buried deep within nucleosomal DNA? As nucleosomes serve as barriers to most chromatin-templated activity, TFs must gain access to *cis*-regulatory elements (CREs) such as promoters and enhancers, often over incredibly long genomic distances.

Collaboration amongst TFs, or binding cooperativity, is considered an established mechanism to promote gene modulation and increase binding affinity at specific motifs. There are several distinct models to explain TF cooperativity: a “billboard” model wherein the co-occurrence of TF binding sites (but not, necessarily, their ordering and spacing) defines cell-type-specific CRE activity; an “enhanceosome” model, wherein motif placement and spacing promotes co-binding of multiple factors to drive element activity; and a “collective” model, wherein a few sequence specific factors (potentially bridged by many other proteins) nucleate the assembly of an active CRE (Spitz and Furlong, 2012). The accuracy of any of these models depends on the extent to which TF cooperativity occurs *in vivo*. Classical studies of TFs have

demonstrated many examples of obligate cooperative TF interactions, as with basic leucine zipper (bZIP; e.g., Jun-Fos) and basic helix-loop-helix (bHLH; e.g., Myc-Max) TFs.

Whether such cooperative interactions exist beyond classic multimerization, however, has remained an open question due to technical limitations; for example, nucleolytic footprinting (Hesselberth et al., 2009) inherently destroys the chromatin fiber and is typically averaged over many chromatin templates, while *in vitro* reconstitution approaches such as CAP-SELEX (Jolma et al., 2015) cannot recapitulate the complexity of dynamic trans environments *in vivo*. Two recent studies applied integrative genomic analyses and cutting-edge genomic footprinting techniques to reveal extensive *in vivo* cooperativity between TFs in both fly and mouse (Rao et al., 2021; Sönmezer et al., 2021). Sönmezer et al. (2021) and Rao et al. (2021) both explore the role of multiple TF co-occupancy by quantifying binding events and patterns on single DNA molecules.

Rao et al. (2021) analyze TF co-binding events at *Drosophila* S2 enhancer elements by overlaying a combination of new and existing genomic datasets. First, they identified putative TF binding sites by analyzing short micrococcal nuclease (MNase) digested fragments overlapping enhancers defined using STARR-seq (Arnold et al., 2013). The authors demonstrate that predicted S2 enhancers harbor on average 4 discrete TF footprints, though they note that this estimate is simply a bulk average. To move past this average, the authors apply integrative analysis of MNase-seq data by fragment length versus midpoint “V-plots” (Henikoff et al., 2011). Intriguingly, these analyses revealed multiple examples of long MNase footprints that span multiple TF binding sites, suggestive of cases where co-occupancy of adjacent TFs protects a stretch of DNA from MNase digestion. The authors then leverage previously published data from S2 cells (Krebs et al., 2017a) to confirm that these MNase-derived footprints match co-occupancy patterns seen by whole genome bisulfite sequencing of methyltransferase footprints. They note that co-occupancy is most common at regulatory elements with high nucleosome turnover, hinting at a conflict between nucleosomes and cooperatively-binding TFs as they compete for the same regulatory DNA.

Sönmezer et al. (2021) improved upon their previously published dual enzyme single-molecule footprinting protocol to scale analyses to mammalian-sized genomes. Taking advantage of genetically-engineered murine embryonic stem cell (mESC) lines devoid of CpG methylation (Domcke et al., 2015) and hybrid capture-based targeted sequencing, the authors were able to achieve high, single-molecule coverage of 78,807 footprinted mESC regulatory elements. The resulting data are striking. The authors observe diverse examples of singular and cooperative TF binding to DNA and are able to estimate TF occupancy rates by explicitly quantifying nucleosome-bound and TF-bound footprint states at individual CREs for myriad factors, including Myc-Max, Nrfl, Ctf, and Rest. As in fly, the mESC epigenome harbors many cooperative TF interactions, and these cooperative interactions occur most frequently in regions of high nucleosome occupancy. Using the TF Nrfl as an example, the authors directly test the biophysical importance of these cooperative interactions; after knocking-down Nrfl levels via siRNA, the authors observe clear shifts in TF occupancy and nucleosome loading in the vicinity of Nrfl sites on single molecules *in vivo*—a powerful marker of the significance of these findings.

Broadly speaking, these two studies highlight the amazing complexity of cell-type-specific regulation of CREs. CRE activity is the product of an intricate dance between nucleosomes and TFs: while TF binding in nucleosome-poor regions may happen in isolation, accessing nucleosome-rich DNA requires multiple TFs to “tango,” as protein-DNA interactions are passed from core histones to sequence-specific DNA binding factors. Indeed, a common theme of both of these studies is their support for a model of “nucleosome-mediated cooperativity.” Co-binding of TFs has long been postulated as a mechanism to overcome the energetic barrier posed by nucleosome-wound DNA (Mirny, 2010). The clear and conserved patterns of intranucleosomally spaced co-binding events shown here are highly suggestive of aspects of this model, though as Sönmezer et al. find through simulation, the model likely requires an ATP-dependent process (such as chromatin remodeling) to best recapitulate *in vivo* observations.

These studies raise an exciting slew of questions regarding the logic of *cis* regulation: what are the biophysical mechanisms by which TF pairs cooperate at these length scales? Many TFs are known to

physically reshape DNA upon binding, and distinguishing between the interaction modes underlying this cooperativity will be an exciting direction of study. What is the significance of the accessible nucleosome- and TF-free states seen in both studies? Most TFs turn over their cognate sites rapidly *in vivo*, and it is tempting to speculate a regulatory role for the “un-bound” states observed in both studies. Finally, how do the activities of ATP-dependent chromatin remodelers impact TF co-occupancy to effect physiologically relevant gene expression changes? As genomic footprinting methods continue evolving at a rapid pace, answering these and myriad other questions will become possible, a tantalizing prospect for the chromatin field at large.

1.6 Epigenetics and chromatin are inextricably linked to disease

Brief review: epigenetics in cancer

Oncogenesis is a multistep process leading to cancerous cells characterized by abnormal growth and replication, evasion of normal apoptotic signaling pathways, and tissue invasion and metastasis (Hanahan and Weinberg, 2000). Beyond genetic mechanisms of cancer, it is now widely accepted that epigenetics plays a larger role in oncogenesis than previously determined. DNA modifications and histone PTMs, as discussed previously, regulate DNA-templated processes including transcription, replication, and DNA repair. There are several established epigenomic characteristics of tumors. Tumors are often characterized by alterations in widespread 5mC (5-methylcytosine), including hypermethylation of 5'-cytosine-phosphate-guanine-3' (CpG) islands and global hypomethylation at repetitive regions. Dysregulation of methylation may be due to driver mutations in key chromatin regulators, such as histone tail regulators (histone acetyltransferases and demethylases), DNA methyltransferases (DNMTs) and demethylases, as well as ATP-dependent chromatin remodelers. For example, the histone methyltransferase MLL2 contains mutations in over 90% of follicular lymphoma cases (Gonzalez-Perez et al., 2013). Interestingly, alterations of methylation in cancer often occur at poised genes essential to cellular

development. These regions are often marked with bivalent histone modifications, including both active (H3K4me3) and repressive (H3K27me2) marks, although the mechanisms of how methylation alterations at bivalent regions contribute to cancer is not well-understood. Since these mutated epigenomic regulators play a key role in oncogenesis, investigations into targetable pathways for therapeutic development have been an active area of interest. Several approved drug therapies exist which target epigenetic regulators and their pathways, including inhibitors of DNA methyltransferase (DNMT), histone deacetylase (HDAC), and the non-receptor tyrosine kinase JAK2 (Dawson and Kouzarides, 2012; Golemis et al., 2018; Gonzalez-Perez et al., 2013).

ATP-dependent chromatin remodelers in cancer

It is widely reported that mutations in regulators of the genome lead to a broad array of human cancers, particularly in mutations within subunits of ATP-dependent remodelers and within the histone core subunits themselves (Ferraro, 2016; Narlikar et al., 2013). The subunits of the mammalian SWI/SNF complex, or BAF (Brg/Brahma-associated factors), are the most commonly mutated chromatin regulators in human cancer, occurring in a startling 20% of all sequenced human tumors. Specific BAF subunit mutations lead to particular cancers, suggesting each subunit may have a cell and tissue-specific tumor suppressor role. Mutations are also more frequently heterozygous; dosage sensitivity mechanisms are likely at play contributing to formation of disease (Kadoch and Crabtree, 2015). As for mutations themselves, the highest frequency occurs in SMARCA4, which encodes the BAF ATPase subunit Brg1. SMARCA4 mutations leading to SWI/SNF mutant cancers are more aggressive and associated with poor prognosis. We still have a limited understanding of the mechanisms of BAF-mutated oncogenesis. Despite their widespread prevalence in many different human cancers, there are no existing therapies to target BAF related diseases. This is challenging due to the complex, combinatorial nature of BAF. There are three main complexes identified to date containing up to 15 subunits encoded from 29 different genes. To further complicate, BAF subunit mutations are typically complete loss of function (LOF) mutations, and thus

impossible targets. Remaining BAF subunits may compensate for the LOF and enable tumor growth (Alfert et al., 2019; Wanior et al., 2021).

BAF does not have an isolated role in evicting nucleosomes and is instead a part of a larger regulatory network with other chromatin modifiers and transcription factors. BAF and PRC2 are thought to act antagonistically via EZH2 of the Polycomb Complex, which facilitates H3K27me deposition and transcriptional silencing (Alfert et al., 2019; Francis et al., 2001; Poepsel et al., 2018; Wilson et al., 2010). It is thought that SMARCA4 is a mutational hotspot leading to impaired ATPase activity and remodeling, as well as impaired competition with the PRC2 complex. For example, knockdown of Brg1 increases transcription of several polycomb subunits of PRC1 and PRC2, which were also found to be directly bound by Brg1 from ChIP-seq experiments (Alfert et al., 2019). Further, Jarid2, a specific PRC2 component required for differentiation of ESCs, interacts with BAF and counteracts PRC2 methyltransferase activity (Kadoch et al., 2017). Overall, we do not have a clear understanding of how impaired BAF remodeling at these genomic regions, due to SMARCA4 mutations, contributes to disease.

Future studies are necessary to better understand how mutated BAF targets nucleosomes, and how they cooperate (or compete) with enzymatic factors and transcription factors that regulate underlying motifs, particularly those involved in cellular differentiation and self-renewal. While we have good mechanistic insight into SWI/SNF ability to evict and replace histones on mononucleosome templates, we still do not have a very clear idea of how SWI/SNF and its individual subunit function on longer chromatin templates. Previous studies have deciphered the mechanisms of various SWI/SNF complex activities on mononucleosomal templates using FRET, restriction enzyme accessibility, gel shifts, and, most recently, Cryo-EM) (Blosser et al., 2009; Ganji et al., 2016; Han et al., 2020; Kassabov et al., 2003; Mashtalir et al., 2020). However, no *in vitro* studies to date have studied how BAF remodels oligonucleosomal fibers. A future study may include endogenous purification of both wildtype and mutant BAF complexes to determine how common SMARCA4 and ARID1A subunit mutations alter chromatin dynamics in a simplified system (Kadoch and Crabtree, 2015). This brings to mind several exciting questions: How do

wildtype and mutant BAF complexes remodel long chromatin fibers comprised of unmodified and modified nucleosome core particles? ii.) How do BAF complexes collaborate with other *trans* regulatory factors (*e.g.*, the polycomb complex) to compartmentalize chromatin and regulate gene expression? iii.) How do these differential activities impact global gene regulatory networks *in vivo*? Future studies comprising combinatorial enzymatic reactions, single-molecule, and long-read sequencing could clarify how the BAF complex operates at the resolution of single chromatin fibers, how mutations in BAF modulate single-fiber activity, and how these activities impact disease-specific gene regulation.

It should be noted that many other chromatin-remodeler related factors contain mutations at a lower occurrence, including ATRX, Chd1, and Chd4 (Alver et al., 2017; Kadoch et al., 2013). There is also an emerging role for ISWI remodelers in cancer. Several cancers have identified ISWI aberrant expression and mutated subunits which contribute to oncogenic pathways and tumorigenesis. SMARCA1 (SNF2L) is thought to have a tumor-specific suppressor role in several cancers. SMARCA4 (SNF2H) is overexpressed in ovarian and breast cancer, hepatocellular carcinoma, and acute myeloid leukemia (Li et al., 2021). Future studies are essential to understand the underlying mechanism of ISWI in various cancers, to determine how they interplay with transcription factors and promote oncogenesis.

Oncohistones within histone H3 disrupt Polycomb-repressed chromatin

Mutations in epigenetic modifiers are the most common mutations in childhood cancers (Panditharatna and Filbin, 2020). Histone core mutations or ‘oncohistones’ have more recently been identified and implicated in a broad range of cancers, including over 60% of pediatric high grade gliomas (pHGG). The landscape of oncohistone mutations ranges from singular driver mutations with high genetic penetrance in rare tumors to low frequency mutations in more common cancers, suggesting a broader role for histones and thus, altered chromatin structure, in oncogenesis (Behjati et al., 2013; Nacev et al., 2019; Schwartzenruber et al., 2012). One of the most well-characterized oncohistones is a dominant mutation within histone H3 encoding p.Lys27Met (H3K27M), a driver of diffuse intrinsic pontine glioma (DIPG), a

pediatric and young adult HGG restricted to the midline structures of the brain (Schwartzentruber et al., 2012). Oncohistones within histone H3 preferentially occur within the non-canonical, replication-independent variant histone H3.3. H3 or variant H3.3 mutations lead to clinically distinct phenotypes, with H3.3K27M-positive DIPG tumors being associated with worse prognosis (Castel et al., 2015; Mohammad and Helin, 2017). Histone H3.3 is enriched at specific genomic locations such as Transcription Start Sites (TSSs), pericentric heterochromatin, active genes, and telomeres (Goldberg et al., 2010). However, the mechanisms of H3K27M-driven (both canonical and variant) DIPG are not well understood.

H3K27M mutations are thought to disrupt EZH2-mediated methylation at H3K27, an essential repressive mark. EZH2 is a key enzymatic component of Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2), which tune higher order chromatin compaction through H3K27 methylation. PRC2 acts with its writer enzyme EZH2 to catalyze H3K27-di and -tri methylation, and PRC1 mediates gene repression and chromatin compaction (Comet et al., 2016). Interestingly, aggressive H3.3K27M in DIPG acts as a dominant-negative inhibitor of EZH2, which leads to a characteristic global reduction of H3K27me_{2/3} and increase in H3K27ac (Chan et al., 2013; Lewis et al., 2013). Several cellular and tumor models have been used to characterize H3.3K27M-positive DIPG, leading to conflicting theories on how H3.3K27M acts dominantly to drive tumorigenesis in such a cell-type specific manner. One hypothesis suggests H3K27M-containing nucleosomes sequester and inactivate EZH2 to disrupt the normal distribution of H3K27me₃ (Lewis et al., 2013). Another study demonstrated that despite a global H3K27me₃ reduction, deposition of this mark is retained at specific loci, possibly due to preferential recruitment of PRC2 to strong affinity sites that in turn activate oncogenes (Mohammad et al., 2017). Another study suggests H3.3K27M prevents spread of H3K27me₃ from PRC2 binding sites to larger silencing domains, and that the mutation is required for tumorigenesis (Harutyunyan et al., 2019). It has not yet been investigated whether these related epigenomic changes lead to altered nucleosome positioning at typical Polycomb-repressed chromatin and how this may contribute to tumorigenesis within DIPG, and more broadly in other oncohistone-related cancers.

1.7 Current toolkit to map the epigenome

The development of 2nd generation NGS tools

Over forty years ago, the advent of chemical chain termination sequencing followed by Sanger's dideoxy method revolutionized biology as we know it (van Dijk et al., 2018). Sanger sequencing technology led to the Human Genome Project starting in 2001 (Lander et al., 2001) with the goal of fully sequencing the 3B bases of DNA that make up the human haploid genome. Sanger sequencing a single human genome via capillary array electrophoresis (CAE) cost approximately \$10m. This was soon followed by second or next generation sequencing (NGS) technology, which has since allowed for faster and cheaper genome sequencing far exceeding Moore's Law (Hayden, 2014). Efforts to make sequencing cheaper, more efficient and accurate were initially made possible through massive amounts of grant funding through the US National Human Genome Research Institute (NHGRI). Funding academic labs to create these technologies ultimately contributed to the formation of current industry leaders in the field, including the NGS market-leader Illumina, which harnesses reversible terminator chemistry and sequencing-by-synthesis to sequence genomes at a now unprecedented pace.

Large-scale studies to identify functional portions of the genome, particularly The Encyclopedia of DNA Elements (ENCODE Project Consortium) were greatly accelerated through NGS technology (Dunham et al., 2012a). We have since uncovered unprecedented detail of the whole genome beyond its linear DNA bases, including but not limited to histone modification, chromatin structure, mapped transcriptional regions and transcription factor (TF) association. These studies were largely made possible with the switch to high-throughput, 2nd generation NGS. Despite incredible progress in NGS technology, there is still much to learn about the unmappable regions of the human genome. As previously mentioned, NGS technology, while extremely powerful, is limited by short-reads and the inability to reliably map repetitive regions. The updated Genome Reference Consortium (GRC) (hg38) contains approximately 151 mega-base pairs (Mbp) of unknown, unmappable sequence, mostly consisting of heterochromatic regions (pericentromeric, subtelomeric, ribosomal DNA (rDNA), and human satellite gaps). What efforts have been

made to resolve these large gaps? As of 2022, the first fully resolved human genome was made available (T2T-CHM13), owing to the international and collaborative work of the Telomere-to-Telomere (T2T) consortium. In a series of studies, the complete genomic and epigenetic maps of previously unmappable human centromeres (Altemose et al., 2022; Gershman et al., 2022; Hoyt et al., 2022), made possible by the advances in 3rd generation, long-read sequencing (covered in the following section).

Common methods to map chromatin

The existing methods to study nucleosome positioning historically use cleavage reagents (e.g. dimethyl sulphate (Becker et al., 1986), hydroxyl radicals (Tullius, 1988), and nucleases (Hewish and Burgoyne, 1973)) followed by gel electrophoresis and/or Southern blotting to map the abundance, accessibility, and nucleosome repeat lengths (NRLs) of chromatin fibers (Richard-Foy and Hager, 1987). More recently, these methods have been coupled to high-throughput short-read sequencing (Zentner and Henikoff, 2014), enabling genome-wide measurement of average nucleosome positions. Common chromatin accessibility methods include DNase-seq, FAIRE-seq, MNase-seq, and ATAC-seq (Klein and Hainer, 2020). Other methods to map chromatin-bound proteins include ChIP-seq, DamID, and CUT&RUN. ChIP-seq is often used to avoid enzyme-associated sequence preference and is an alternative chromatin mapping method, which utilizes immunoprecipitation of histone marks associated with gene activation or repression. Chromatin digestion via sonication is followed by immunoprecipitation with a histone antibody of interest and high-throughput sequencing. Recent efforts have been made to improve the resolution of ChIP-seq, including ChIP-exo, where immunoprecipitated and crosslinked fragments are digested with an exonuclease, allowing for both enrichment and resolution of the crosslink within the DNA fragment. Lastly, CUT&RUN harnesses MNase-based immunoprecipitation by using an MNase-Antibody to bind histone or epitope of interest (Klein and Hainer, 2020). While powerful, these current chromatin mapping methods share key limitations: measurement of individual protein-DNA interactions inherently requires destruction of the chromatin fiber and averaging of signal across many short molecules.

1.8 Harnessing third-generation sequencing tools to study chromatin

The advent of third-generation (i.e. high-throughput, long-read) shotgun sequencing offers a potential solution to many sequence resolution issues mentioned above (Shema et al., 2018). There are two major existing third-generation sequencing platforms, Oxford Nanopore and Pacific Biosciences (PacBio), the latter of which will be primarily discussed. To briefly review, Nanopore sequencing technology does not require a sequencing polymerase, and instead detects the electrical current corresponding to individual DNA bases in real-time. Nanopore allows for the detection of incredibly long DNA sequences, but can only sequence the forward/reverse strand once, limiting its sequence accuracy. For the purpose of our studies, we chose PacBio SMRT sequencing to obtain highly accurate consensus reads, as will be discussed below (Oberbeckmann et al., 2019; Shipony et al., 2020; Wang et al., 2019)

PacBio's SMRT (Single Molecule Real Time) is a novel NGS platform that allows for highly accurate sequencing and obtaining longer read lengths up to 10-20kb. SMRT sequencing produces high fidelity reads through reads of a closed, single molecule DNA template prepared through ligation of hairpin adaptors or 'SMRTbells' onto both ends of an end-repaired dsDNA template of interest. These SMRTbell templates are sequenced by a high-fidelity processive DNA polymerase (traditionally Phi29) affixed to the bottom of a 100nm-diameter nanophotonic structure called a Zero-mode waveguide (ZMW), which can hold an astonishingly miniscule volume up to 20 zeptoliters (10^{-21} L). One single sequencing cell, which can contain multiplexed population of libraries (SMRTbells containing unique barcoded adapters), contains 8 million ZMWs, allowing for theoretically up to 8 million single-molecule reads in a sequencing run (Eid et al., 2009).

How does PacBio SMRT Sequencing reaction occur? First, a high-fidelity polymerase affixed to the ZMW uses a single molecule of DNA as a template for synthesis, creating a DNA-template-polymerase complex. The polymerase in each ZMW incorporates fluorescently-labeled nucleotides when illuminated from below by a laser light, which then emits a distinct emission spectra detected and recorded by a camera in real-time (Eid et al., 2009). The phospholinked nucleotides that are introduced into the ZMW chamber

are unique compared to Illumina nucleotides in that they carry a fluorophore on the terminal phosphate rather than base, so the enzyme cleaves the label when incorporated and leaves behind the natural strand of DNA. These signals are converted to long-read sequences called Continuous Long Reads (CLR) where each pass is a subread that can be combined into a highly accurate set of sequences allowing for highly accurate circular consensus sequencing (CCS). This allows the advantage of combining subreads and combining overlapping lengths to improve accuracy. Overall, the polymerase can make multiple passes around the single molecule template, generating many subreads within a single ZMW to yield highly accurate (and easily error-corrected) CCS reads. The read length has improved with the advent of PacBio Sequel I and Sequel II systems which can read up to 20kb length sequences, compared to the maximum read length of Paired End 250bp for Illumina HiSeq 2500. There are many advantages of SMRT sequencing: sequencing through and obtaining accurate reads at repetitive regions, differentiating haplotypes, and avoiding bias and loss of coverage associated with high/low GC content areas.

Another key advantage of SMRT sequencing is the native detection of epigenomic modifications such as methylation (Flusberg et al., 2010a). First, the processive polymerase incorporates a phospholinked dNTP, leading to the raw data output of the bases corresponding fluorescent pulse intensity. As the incorporated base is held for milliseconds it produces a distinct light emission and pulse time, called the interpulse duration (IPD). The phosphodiester bond formation liberates the dye-linker-pyrophosphate product, diffusing out of the ZMW and ending the associated pulse. Lastly, the polymerase translocates to the next position allowing the next nucleotide to bind the active site and for a new fluorescent pulse to begin. Interestingly and much to an experimentalists advantage, when the polymerase encounters an epigenetic DNA modification, the polymerase stalls leading to a distinct IPD compared to its unmodified counterpart. This polymerase pausing can detect different base modifications with varying reliability. While its ability to detect 5-methylcytosine (5mC) modifications are weak, its ability to detect adenine methylation (N-6-methyldeoxyadenosine, m6dA) is very prominent (~5x longer IPD than an unmodified adenine) (Flusberg et al., 2010b). More recently, analyses methods have allowed for better detection of endogenous

5-methylcytosine (5mC) which will allow scientist to avoid the harsh side-effects associated with bisulfite conversion typically used in short-read sequencing (Liu et al., 2019; Tse et al., 2021)

It should be noted that the kinetic rate of DNA polymerase is known to be sensitive to sequence context, where roughly 80% of the IPD variation can be explained by a 10-basepair sequence context (7 bases upstream and 2 bases downstream from the incorporation site). The upstream bases which have bigger impact on the enzyme kinetic rate can be accounted for in your downstream analyses. Lastly, since SMRT sequencing reads are strand specific with respect to the detection of these kinetic variation events, modifications can be inferred in a strand specific manner (Feng et al., 2013).

Harnessing PacBio SMRT sequencing to study chromatin at high resolution

Nucleosome-assisted DNA compaction is essential for mitotic segregation of chromosomes, and yet nucleosomes must act dynamically to promote associated replication and transcriptional processes. There are therefore several key questions that must be asked to understand these contradictory roles. How, when, and where do nucleosomes form on arrays *in vivo*? What are the mechanisms, including ATP-dependent chromatin remodeling, that promote this highly dynamic process of nucleosomal reassembly and disassembly? How do these basic nucleosomal processes determine overall chromatin structure, including compaction into fibers and 3D nuclear organization? Lastly, how do associated epigenomic mutations contribute or lead to human disease? The following work aims to address many of these questions by employing higher-resolution sequencing method to map nucleosomes on longer arrays genome-wide.

High-throughput sequencing technologies paired with enzymatic footprinting assays have revolutionized our understanding of genome-wide epigenomic and chromatin patterns. However, short-read sequencing technologies capture ensemble measurements of DNA molecules. In Dr. Vijay Ramani's lab at UCSF, we sought to seek novel methods to solve this 'sequence resolution' issue, by developing an assay that captures chromatin patterns on single, long molecules of DNA. We hypothesized that methylating

accessible chromatin with a non-specific bacterial adenine methyltransferase and sequencing this modified DNA with PacBio's SMRT epigenomic sequencing, would enable detection of nucleosome and transcription factor footprints on multi-kilobase DNA sequences.

Why choose adenine methylation for mapping nucleosomes, and potentially transcription factor footprints on long DNA sequences? DNA methyltransferases have historically been used as an alternative to Chromatin Immunoprecipitation sequencing (ChIP-seq) in order to probe for DNA-protein interactions, but ultimately have been limited in motif specificity across the genome, leading to bias and inaccuracy of defined binding sites (Hoose Scott A. and Kladde, 2006; Steensel BV. and Henikoff S., 2000). A more recent study demonstrated that EcoGII paired with a proximity labeling technique could successfully map protein-DNA interactions *in vivo* to any DNA motif of interest (methyl adenine identification, MadID) (Sobecki et al., 2018). When targeted to both GATC-poor and GATC-rich genomic regions, EcoGII adenine methylation was widespread and provided high coverage throughout the genome (Sobecki et al., 2018). Importantly,

While adenine methylation of RNA has exhibited an increasingly important role in bacterial and mammalian RNA (Li et al., 2019) it is extremely sparse across the entire human (at 0.05% of total adenines) and mouse genome on DNA, making adenine methylation an ideal non-specific experimental marker for nucleosome footprinting (Xiao et al., 2018). The DNA methyltransferase EcoGII was identified and cloned from the pathogenic C227-11 *Escherichia coli* genome (Fang et al., 2012) and found to methylate adenines non-specifically in a GATC sequence context *in vitro* and *in vivo* on either DNA or RNA substrates up to 99% (Murray et al., 2018a). Pertinent to our goal of developing a tool which allows mapping of nucleosome positioning on long chromatin templates, we decided to harness EcoGII paired with long-read sequencing and the inherent epigenomic detection made possible with SMRT sequencing.

In the following chapters, we demonstrate our novel method: Single-molecule Adenine Methylated Oligonucleosome Sequencing Assay (SAMOSA), which combines adenine methyltransferase footprinting of nucleosomes with base modification detection on the PacBio single-molecule real time sequencer to

measure nucleosome positions on single chromatin templates. SAMOSA is a high-throughput single-molecule sequencing method that combines adenine methyltransferase footprinting and single-molecule real-time DNA sequencing to natively and nondestructively measure nucleosome positions on individual chromatin fibers. SAMOSA data allows unbiased classification of single-molecular 'states' of nucleosome occupancy on individual chromatin fibers. We leverage this to estimate nucleosome regularity and spacing on single chromatin fibers genome-wide, at predicted transcription factor binding motifs, and across human epigenomic domains. Our analyses suggest that chromatin is comprised of both regular and irregular single-molecular oligonucleosome patterns that differ subtly in their relative abundance across epigenomic domains. This irregularity is particularly striking in constitutive heterochromatin, which has typically been viewed as a conformationally static entity. Our proof-of-concept study provides a powerful new methodology for studying nucleosome organization at a previously intractable resolution and offers up new avenues for modeling and visualizing higher order chromatin structure.

Harnessing SAMOSA to determine how SNF2h remodels single-molecule, long templates

Current methods to footprint nucleosomes rely on Illumina short-read sequencing, which generate bulk average measurements of nucleosome positioning and regularity. Many processes essential to gene regulation occur on individual chromatin fibers, including transcription factor binding and ATP-dependent nucleosome remodeling. In order to better understand these processes and how they may lead to dysregulated gene expression in disease, we require single molecule measurements. SAMOSA allows for measurement of nucleosomes (occupancy and regularity) on single-molecule fibers using PacBio's long-read Single Molecule Real Time (SMRT) sequencing, which can natively detect epigenetic modifications. As a prior proof-of-concept (Abdulhay et al., 2020), we demonstrated that methylation with a non-specific adenine methyltransferase footprints nucleosomes at previously intractable resolution when applied both to assembled Widom-601 arrays *in vitro* and within a cellular context *in vivo* from K562 cells.

We hypothesized that we could further apply SAMOSA to visualize discrete patterns of remodeled nucleosomes on *in vitro* assembled chromatin. We first aimed to assemble histones at varying densities (as inspired by work from the Korber and Pugh labs) via salt gradient dialysis onto any mammalian sequence of interest (where underlying nucleosome sequence preference is unknown). We chose two previously studied murine sites (~3kb in size) which were shown to have low or high dependency on SNF2h for CTCF binding (Barisic et al., 2019a). The assembled chromatin arrays were validated with imaging and MNase digestion laddering. We performed m6a footprinting and subjected the libraries to PacBio sequencing as previously performed and determined that we can indeed assemble nucleosomes onto any site of interest and detect nucleosome footprints with our generated methylation and accessibility computational predictions. We call this newer technique: SAMOSA-ChAAT (Chromatin Accessibility of Assembled Templates).

We applied remodeling reactions to the assembled templates with SNF2h, the ATPase of the Imitation Switch (ISWI) family of remodelers, which has been shown to create regularly spaced nucleosome arrays. A major question in the field of ATP-dependent remodelers has been whether SNF2h has a remodeler-specific mechanism to create nucleosome regularity. There are two conflicting models on how SNF2h may mechanistically achieve nucleosome regularity. Previous assays from the Korber lab examined bulk ISWI-remodeled, MNase digested arrays with varying histone density. Their findings suggested that SNF2h slides nucleosomes via a clamping mechanism where, regardless of underlying nucleosome density, like a ruler, SNF2h will create fixed internucleosomal distances (IDs) on its target region. An alternative model, more in line with findings in the Narlikar lab which has extensively studied ISWI mechanism, suggest that SNF2h instead uses its HSS domain to sense flanking DNA length on either side of a nucleosome and pushes nucleosomes towards the side with longer flanking DNA. This indicates that instead SNF2h stochastically slides nucleosomes, and this is dependent on underlying nucleosome density. This is a long-winded way of saying: we can distinguish these two models now that we have a higher resolution technique to determine single-fiber nucleosome positions. When we compared

unremodeled with remodeled arrays, we find that the average spacing on remodeled single fibers decreases as the underlying nucleosome density increases, demonstrating that SNF2h remodels in a stochastic, length-sensing manner.

Next, we wanted to determine whether SNF2h follows this same behavior as a length-sensing remodeler *in vivo*. We applied a modified *in vivo* adaptation of SAMOSA to murine embryonic stem cells (mESCs) devoid of SNF2h and SNF2h WT re-expressed, both previously generated and gifted by Dirk Schübeler's lab. Similar to their findings, we found that SNF2h KO leads to decreased NRLs and an enrichment of irregular fibers. When comparing the KO and WT re-expression cell lines, we found that initial nucleosome location and their density dictates SNF2h-mediated accessibility at target sites. SNF2h has been shown to play a role both in the formation of heterochromatin (repressed chromatin) and enabling TF accessibility (opening chromatin). We found that at heterochromatin, SNF2h generated regularly spaced nucleosomes with multiple NRLs. At SNF2h-dependent CTCF sites, SNF2h slides nucleosomes to increase motif accessibility. How SNF2h operates is therefore dependent on underlying nucleosome density, suggesting that SNF2h senses density to assist in opposing actions essential for gene regulation (open/active vs closed/inactive chromatin).

REFERENCES

- Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020a). Massively multiplex single-molecule oligonucleosome footprinting. *Elife* 9.
- Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020b). Massively multiplex single-molecule oligonucleosome footprinting. *Elife* 9, 1–23.
- Ahmad, K., and Henikoff, S. (2001). Modulation of a Transcription Factor Counteracts Heterochromatic Gene Silencing in *Drosophila*. *Cell* 104, 839–847.
- Ahmad, K., and Henikoff, S. (2002). The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Mol. Cell* 9, 1191–1200.
- Alfert, A., Moreno, N., and Kerl, K. (2019). The BAF complex in development and disease. *Epigenetics Chromatin* 2019 121 12, 1–15.
- Altemose, N., Logsdon, G.A., Bzikadze, A. V., Sidhwani, P., Langley, S.A., Caldas, G. V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., et al. (2022). Complete genomic and epigenetic maps of human centromeres. *Science* (80-.). 376.
- Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., and Roberts, C.W.M. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* 2017 81 8, 1–10.
- Angelov, D., Vitolo, J.M., Mutskov, V., Dimitrov, S., and Hayes, J.J. (2001). Preferential interaction of the core histone tail domains with linker DNA. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6599.
- Arents, G., and Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11170.
- Arimura, Y., Tachiwana, H., Oda, T., Sato, M., and Kurumizaka, H. (2012). Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry* 51,

3302–3309.

Armache, J.P., Gamarra, N., Johnson, S.L., Leonard, J.D., Wu, S., Narlikar, G.J., and Cheng, Y. (2019). Cryo-EM structures of remodeler-nucleosome intermediates suggest allosteric control through the nucleosome. *Elife* 8.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* (80-.). 339, 1074–1077.

Badenhorst, P., Voas, M., Rebay, I., and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.* 16, 3186–3198.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.

Baldi, S., Krebs, S., Blum, H., and Becker, P.B. (2018). Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. *Nat. Struct. Mol. Biol.* 2018 259 25, 894–901.

Baldi, S., Korber, P., and Becker, P.B. (2020). Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. *Nat. Struct. Mol. Biol.* 2020 272 27, 109–118.

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

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019a). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* 569, 136–140.

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019b). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* 569, 136–140.

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

Barutcu, A.R., Lajoie, B.R., Fritz, A.J., McCord, R.P., Nickerson, J.A., Van Wijnen, A.J., Lian, J.B., Stein, J.L., Dekker, J., Stein, G.S., et al. (2016). SMARCA4 regulates gene expression and higher-order chromatin structure in proliferating mammary epithelial cells. *Genome Res.* 26, 1188–1201.

Becker, P.B., Gloss, B., Schmid, W., Strähle, U., and Schütz, G. (1986). In vivo protein–DNA interactions

in a glucocorticoid response element require the presence of the hormone. *Nat.* 1986 3246098 324, 686–688.

Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., Davies, H., et al. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat. Genet.* 45, 1479–1482.

Blosser, T.R., Yang, J.G., Stone, M.D., Narlikar, G.J., and Zhuang, X. (2009). Dynamics of nucleosome remodelling by individual ACF complexes. *Nature* 462, 1022–1027.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422.

Bracken, A.P., Brien, G.L., and Verrijzer, C.P. (2019). Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev.* 33, 936–959.

Brahma, S., and Henikoff, S. (2019). RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. *Mol. Cell* 73, 238-249.e3.

Brahma, S., and Henikoff, S. (2020). Epigenome Regulation by Dynamic Nucleosome Unwrapping. *Trends Biochem. Sci.* 45, 13–26.

Cai, S., Song, Y., Chen, C., Shi, J., and Gan, L. (2018). Natural chromatin is heterogeneous and self-associates in vitro. *Mol. Biol. Cell* 29, 1652–1663.

Carey, M., Li, B., and Workman, J.L. (2006). RSC Exploits Histone Acetylation to Abrogate the Nucleosomal Block to RNA Polymerase II Elongation. *Mol. Cell* 24, 481–487.

Castel, D., Philippe, C., Calmon, R., Le Dret, L., Truffaux, N., Boddaert, N., Pagès, M., Taylor, K.R., Saulnier, P., Lacroix, L., et al. (2015). Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. *Acta Neuropathol.* 130, 815–827.

Chan, K.M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., David James, C., Jenkins, R., et al. (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27

methylation and gene expression. *Genes Dev.* *27*, 985–990.

Chioda, M., Vengadasalam, S., Kremmer, E., Eberharter, A., and Becker, P.B. (2010). Developmental role for ACF1-containing nucleosome remodellers in chromatin organisation. *Development* *137*, 3513–3522.

Clapier, C.R., and Cairns, B.R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* *492*, 280–284.

Clapier, C.R., Längst, G., Corona, D.F. V., Becker, P.B., and Nightingale, K.P. (2001). Critical Role for the Histone H4 N Terminus in Nucleosome Remodeling by ISWI. *Mol. Cell. Biol.* *21*, 875–883.

Collins, N., Poot, R.A., Kukimoto, I., García-Jiménez, C., Delleire, G., and Varga-Weisz, P.D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat. Genet.* *32*, 627–632.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat. Rev. Cancer* *16*, 803–810.

Correll, S.J., Schubert, M.H., and Grigoryev, S.A. (2012). Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J.* *31*, 2416–2426.

Cuaycong, M.H., and Elgin, S.C.R. (2001). Long-Range Nucleosome Ordering Is Associated with Gene Silencing in. *Mol. Cell. Biol.* *21*, 2867–2879.

Cutter, A.R., and Hayes, J.J. (2015). A Brief Review of Nucleosome Structure. *FEBS Lett.* *589*, 2914.

Dann, G.P., Liszczak, G.P., Bagert, J.D., Müller, M.M., Nguyen, U.T.T., Wojcik, F., Brown, Z.Z., Bos, J., Panchenko, T., Pihl, R., et al. (2017). ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature* *548*, 607–611.

Davidson, I.F., and Peters, J.M. (2021). Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* *2021* 227 22, 445–464.

Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* *150*, 12–27.

Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladde, M.P., and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent

nucleosomes. *Mol. Cell* 38, 590–602.

Deindl, S., Hwang, W.L., Hota, S.K., Blosser, T.R., Prasad, P., Bartholomew, B., and Zhuang, X. (2013). ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell* 152, 442–452.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26, 5433–5438.

De Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Dargham, D.B., Jounier, S., Humbertclaude, H., Ribierre, F., Baulard, C., et al. (2016). Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* 530, 113–116.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet.* 34, 666–681.

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

Donehower, L.A., Huang, A.L., and Hager, G.L. (1981). Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. *J. Virol.* 37, 226–238.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* (80-.). 306, 1571–1573.

Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* 186, 773–790.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21, 129–228.

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

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

Nat. 2012 4897414 489, 57–74.

Ebbert, M.T.W., Farrugia, S.L., Sens, J.P., Jansen-West, K., Gendron, T.F., Prudencio, M., McLaughlin, I.J., Bowman, B., Seetin, M., Dejesus-Hernandez, M., et al. (2018). Long-read sequencing across the C9orf72 “GGGGCC” repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Mol. Neurodegener.* 13, 1–17.

Eberharter, A., Ferrari, S., Längst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., and Becker, P.B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO J.* 20, 3781–3788.

Ehrensberger, A.H., Franchini, D.M., East, P., George, R., Matthews, N., Maslen, S.L., and Svejstrup, J.Q. (2015). Retention of the Native Epigenome in Purified Mammalian Chromatin. *PLoS One* 10, e0133246.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138.

Erdel, F., and Rippe, K. (2011). Chromatin remodelling in mammalian cells by ISWI-type complexes—where, when and why? *Authors J. Compil.* ^a 278, 3608–3618.

Eustermann, S., Schall, K., Kostrewa, Di., Lakomek, K., Strauss, M., Moldt, M., and Hopfner, K.P. (2018). Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature* 556, 386–390.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* 30, 1232–1239.

Feng, Z., Fang, G., Korlach, J., Clark, T., Luong, K., Zhang, X., Wong, W., and Schadt, E. (2013). Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic. *PLOS Comput. Biol.* 9, e1002935.

Ferraro, A. (2016). Altered primary chromatin structures and their implications in cancer development. *Cell. Oncol.* 39, 195–210.

Filion, G.J., van Bommel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping

Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell* 143, 212–224.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 73, 1897–1901.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. *Nature* 269, 29–36.

Flaus, A. (2012). Principles and practice of nucleosome positioning in vitro.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010a). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010b). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 2010 76 7, 461–465.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol. Cell* 8, 545–556.

Fyodorov, D. V., Blower, M.D., Karpen, G.H., and Kadonaga, J.T. (2004). Acfl confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes Dev.* 18, 170–183.

Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of Nucleosome Positioning in the Human Genome. *PLOS Genet.* 8, e1003036.

Gamarra, N., Johnson, S.L., Trnka, M.J., Burlingame, A.L., and Narlikar, G.J. (2018). The nucleosomal acidic patch relieves auto-inhibition by the ISWI remodeler SNF2h. *Elife* 7.

Garcia, J.F., Dumesic, P.A., Hartley, P.D., El-Samad, H., and Madhani, H.D. (2010). Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. *Genes Dev.* 24, 1758–1771.

Gershman, A., Sauria, M.E.G., Guitart, X., Vollger, M.R., Hook, P.W., Hoyt, S.J., Jain, M., Shumate, A.,

Razaghi, R., Koren, S., et al. (2022). Epigenetic patterns in a complete human genome. *Science* (80-). 376.

Ghandi, M., Mohammad-Noori, M., Ghareghani, N., Lee, D., Garraway, L., and Beer, M.A. (2016). gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* 32, 2205–2207.

Gibson, B.A., Doolittle, L.K., Schneider, M.W.G., Jensen, L.E., Gamarra, N., Henry, L., Gerlich, D.W., Redding, S., and Rosen, M.K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* 179, 470-484.e21.

Gilbert, N., and Allan, J. (2001). Distinctive higher-order chromatin structure at mammalian centromeres. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11949–11954.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin Architecture of the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. *Cell* 118, 555–566.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* 140, 678–691.

Golemis, E.A., Scheet, P., Beck, T.N., Scolnick, E.M., Hunter, D.J., Hawk, E., and Hopkins, N. (2018). Molecular mechanisms of the preventable causes of cancer in the United States. *Genes Dev.* 32, 868–902.

Gonzalez-Perez, A., Jene-Sanz, A., and Lopez-Bigas, N. (2013). The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* 14, 1.

Grüne, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F.V., Becker, P.B., and Müller, C.W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol. Cell* 12, 449–460.

Hamiche, A., Sandaltzopoulos, R., Gdula, D.A., and Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* 97, 833–842.

Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. *Cell* 100, 57–70.

Harshman, S.W., Young, N.L., Parthun, M.R., and Freitas, M.A. (2013). H1 histones: current perspectives and challenges. *Nucleic Acids Res.* 41, 9593.

Harutyunyan, A.S., Krug, B., Chen, H., Papillon-Cavanagh, S., Zeinieh, M., De Jay, N., Deshmukh, S., Chen, C.C.L., Belle, J., Mikael, L.G., et al. (2019). H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me₂/me₃ and is essential for glioma tumorigenesis. *Nat. Commun.* *10*.

Hayden, E.C. (2014). The \$ 1,000 genome. *Nature* *507*, 294–295.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* *21*, 133–153.

Henikoff, S., and Smith, M.M. (2015). Histone Variants and Epigenetics. *Cold Spring Harb. Perspect. Biol.* *7*.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 18318–18323.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* *2009* *6*, 283–289.

Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* *52*, 504–510.

Hoose Scott A. and Kladde, M.P. (2006). DNA Methyltransferase Probing of DNA-Protein Interactions. In *Gene Mapping, Discovery, and Expression: Methods and Protocols*, M. Bina, ed. (Totowa, NJ: Humana Press), pp. 225–244.

Hota, S.K., Bhardwaj, S.K., Deindl, S., Lin, Y.C., Zhuang, X., and Bartholomew, B. (2013). Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat. Struct. Mol. Biol.* *20*, 222–229.

Hoyt, S.J., Storer, J.M., Hartley, G.A., Grady, P.G.S., Gershman, A., Lima, L.G. de, Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., et al. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* (80-.). 376.

Huang, B., Babcock, H., and Zhuang, X. (2010). Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *Cell* *143*, 1047–1058.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., and Wheeler, T.J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Res.* *44*, D81–D89.

Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* *90*, 145–155.

Ito, T., Levenstein, M.E., Fyodorov, D. V., Kutach, A.K., Kobayashi, R., and Kadonaga, J.T. (1999). ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.* *13*, 1529–1539.

Iwasaki, W., Miya, Y., Horikoshi, N., Osakabe, A., Taguchi, H., Tachiwana, H., Shibata, T., Kagawa, W., and Kurumizaka, H. (2013). Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* *3*, 363.

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

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nat.* *2015* *527*, 384–388.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* *1*.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* *45*, 592–601.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M.G., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* *49*, 213–222.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* *458*, 362–366.

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

Kenzaki, H., and Takada, S. (2015). Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLOS Comput. Biol.* *11*, e1004443.

Khurana, S., and Oberdoerffer, P. (2015). Replication Stress: A Lifetime of Epigenetic Change. *Genes* 2015, Vol. 6, Pages 858-877 *6*, 858–877.

Kim, J.H., Ebersole, T., Kouprina, N., Noskov, V.N., Ohzeki, J.I., Masumoto, H., Mravinac, B., Sullivan, B.A., Pavlicek, A., Dovat, S., et al. (2009). Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res.* *19*, 533–544.

Kim, J.M., Visanpattanasin, P., Jou, V., Liu, S., Tang, X., Zheng, Q., Li, K.Y., Snedeker, J., Lavis, L.D., Lionnet, T., et al. (2021). Single-molecule imaging of chromatin remodelers reveals role of atpase in promoting fast kinetics of target search and dissociation from chromatin. *Elife* *10*.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome Remodeling Induced by RNA Polymerase II: Loss of the H2A/H2B Dimer during Transcription. *Mol. Cell* *9*, 541–552.

Klein, D.C., and Hainer, S.J. (2020). Genomic methods in profiling DNA accessibility and factor localization. *Chromosom. Res.* *28*, 69.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 2018 204 *20*, 207–220.

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

Kornberg, R.D., and Stryer, L. (1988). Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* *16*, 6677–6690.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., and Burger, L. (2017a). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol. Cell* *67*, 411–422.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017b). Genome-wide

Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol. Cell* 67, 411-422.e4.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* 167, 709-721.e12.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* 18, 548.

Lai, B., Gao, W., Cui, K., Xie, W., Tang, Q., Jin, W., Hu, G., Ni, B., and Zhao, K. (2018). Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nat.* 2018 5627726 562, 281–285.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Längst, G., and Becker, P.B. (2001). Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci.* 114, 2561–2568.

Lee, K.M., and Narlikar, G. (2001). Assembly of Nucleosomal Templates by Salt Dialysis. *Curr. Protoc. Mol. Biol.* 54, 21.6.1-21.6.16.

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Sadowski, N., Simpson, J.T., Sedlazeck, F.J., and Timp, W. (2019). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *BioRxiv* 504993.

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Gershman, A., Sadowski, N., Sedlazeck, F.J., Hansen, K.D., Simpson, J.T., and Timp, W. (2020). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* 2020 1712 17, 1191–1199.

Leonard, J.D., and Narlikar, G.J. (2015). A Nucleotide-Driven Switch Regulates Flanking DNA Length Sensing by a Dimeric Chromatin Remodeler. *Mol. Cell* 57, 850–859.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation

found in pediatric glioblastoma. *Science* (80-.). *340*, 857–861.

Li, B., Carey, M., and Workman, J.L. (2007a). The Role of Chromatin during Transcription. *Cell* *128*, 707–719.

Li, B., Carey, M., and Workman, J.L. (2007b). The role of chromatin during transcription. *Cell* *128*, 707–719.

Li, Y., Gong, H., Wang, P., Zhu, Y., Peng, H., Cui, Y., Li, H., Liu, J., and Wang, Z. (2021). The emerging role of ISWI chromatin remodeling complexes in cancer. *J. Exp. Clin. Cancer Res.* 2021 401 *40*, 1–27.

Li, Z., Zhao, P., and Xia, Q. (2019). Epigenetic Methylations on N6-Adenine and N6-Adenosine with the same Input but Different Output. *Int. J. Mol. Sci.* *20*.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015a). Nucleosome Spacing Generated by ISWI and CHD1 Remodelers Is Constant Regardless of Nucleosome Density. *Mol. Cell. Biol.* *35*, 1588–1605.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015b). Nucleosome Spacing Generated by ISWI and CHD1 Remodelers Is Constant Regardless of Nucleosome Density. *Mol. Cell. Biol.* *35*, 1588–1605.

Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., and Song, C.X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* *37*, 424–429.

Long, M., Sun, X., Shi, W., Yanru, A., Leung, S.T.C., Ding, D., Cheema, M.S., MacPherson, N., Nelson, C.J., Ausio, J., et al. (2019). A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res.* *47*, 8399–8409.

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

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

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of

the nucleosome core particle at 2.8 Å resolution. *Nat.* 1997 389:6648–6658, 251–260.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* 304, 3–19.

Maeshima, K., Hihara, S., and Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo? *Curr. Opin. Cell Biol.* 22, 291–297.

Mashtalir, N., Dao, H.T., Sankar, A., Liu, H., Corin, A.J., Bagert, J.D., Ge, E.J., D’Avino, A.R., Filipovski, M., Michel, B.C., et al. (2021). Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes. *Science* (80-.). 373, 306–315.

McBride, M.J., Mashtalir, N., Winter, E.B., Dao, H.T., Filipovski, M., D’Avino, A.R., Seo, H.S., Umbreit, N.T., St. Pierre, R., Valencia, A.M., et al. (2020). The nucleosome acidic patch and H2A ubiquitination underlie mSWI/SNF recruitment in synovial sarcoma. *Nat. Struct. Mol. Biol.* 2020 279 27, 836–845.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 22534–22539.

Mivelaz, M., Cao, A.M., Kubik, S., Zencir, S., Hovius, R., Boichenko, I., Stachowicz, A.M., Kurat, C.F., Shore, D., and Fierz, B. (2020). Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. *Mol. Cell* 77, 488–500.e9.

Mohammad, F., and Helin, K. (2017). Oncohistones: Drivers of pediatric cancers. *Genes Dev.* 31, 2313–2324.

Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D.P., Højfeldt, J.W., Comet, I., Zheng, C., Johansen, J.V., Rapin, N., Porse, B.T., et al. (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.* 23, 483–492.

Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., Shiota, H., Debernardi, A., Héry, P., Curtet, S., et al. (2013). Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev.* 27, 1680–1692.

Murray, I.A., Morgan, R.D., Luyten, Y., Fomenkov, A., Corrêa, I.R., Dai, N., Allaw, M.B., Zhang, X., Cheng, X., and Roberts, R.J. (2018a). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* *46*, 840–848.

Murray, I.A., Morgan, R.D., Luyten, Y., Fomenkov, A., Corrêa, I.R., Dai, N., Allaw, M.B., Zhang, X., Cheng, X., and Roberts, R.J. (2018b). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* *46*, 840–848.

Mutskov, V., Gerber, D., Angelov, D., Ausio, J., Workman, J., and Dimitrov, S. (1998). Persistent Interactions of Core Histone Tails with Nucleosomal DNA following Acetylation and Transcription Factor Binding. *Mol. Cell. Biol.* *18*, 6293.

Nabilsli, N.H., Deleyrolle, L.P., Darst, R.P., Riva, A., Reynolds, B.A., and Kladde, M.P. (2014). Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res.* *24*, 329–339.

Nacev, B.A., Feng, L., Bagert, J.D., Lemiesz, A.E., Gao, J.J., Soshnev, A.A., Kundra, R., Schultz, N., Muir, T.W., and Allis, C.D. (2019). The expanding landscape of ‘oncohistone’ mutations in human cancers. *Nature* *567*, 473–478.

Nakane, T., Kimanius, D., Lindahl, E., and Scheres, S.H.W. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* *7*.

Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* *154*, 490–503.

Oberbeckmann, E., Wolff, M., Krietenstein, N., Heron, M., Ellins, J.L., Schmid, A., Krebs, S., Blum, H., Gerland, U., and Korber, P. (2019). Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* *29*, 1996–2009.

Oberbeckmann, E., Niebauer, V., Watanabe, S., Farnung, L., Moldt, M., Schmid, A., Cramer, P., Peterson, C.L., Eustermann, S., Hopfner, K.P., et al. (2021). Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. *Nat. Commun.* *2021* *121* *12*, 1–17.

Olins, A.L., and Olins, D.E. (1974). Spheroid Chromatin Units (*v* Bodies). *Science* (80-.). *183*, 330–332.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* (80-.). 357.

Panditharatna, E., and Filbin, M.G. (2020). The growing role of epigenetics in childhood cancers. *Curr. Opin. Pediatr.* 32, 67–75.

Papamichos-Chronakis, M., and Peterson, C.L. (2012). Chromatin and the genome integrity network. *Nat. Rev. Genet.* 2013 141 14, 62–75.

Partensky, P.D., and Narlikar, G.J. (2009). Chromatin Remodelers Act Globally, Sequence Positions Nucleosomes Locally. *J. Mol. Biol.* 391, 12–25.

Patel, A.B., Moore, C.M., Greber, B.J., Luo, J., Zukin, S., Ranish, J., and Nogales, E. (2019). Architecture of the chromatin remodeler RSC and insights into its nucleosome engagement. *Elife* 8.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* 25, 154.

Polach, K.J., and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J. Mol. Biol.* 254, 130–149.

Pott, S. (2017). Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *Elife* 6.

Racki, L.R., and Narlikar, G.J. (2008). ATP-dependent Chromatin Remodeling Enzymes: Two Heads are not Better, Just Different. *Curr Opin Genet Dev* 18, 137–144.

Racki, L.R., Yang, J.G., Naber, N., Partensky, P.D., Acevedo, A., Purcell, T.J., Cooke, R., Cheng, Y., and Narlikar, G.J. (2009). The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nat.* 2009 4627276 462, 1016–1021.

Ramani, V., Qiu, R., and Shendure, J. (2019). High Sensitivity Profiling of Chromatin Structure by MNase-SSP. *Cell Rep.* 26, 2465-2476.e4.

Rao, S., Ahmad, K., and Ramachandran, S. (2021). Cooperative Binding between Distant Transcription Factors is a Hallmark of Active Enhancers. *Mol. Cell* 81, 255-267.e6.

Rhodes, D., Brown, R.S., and Klug, A. (1989). Crystallization of nucleosome core particles. *Methods*

Enzymol. *170*, 420–428.

Richard-Foy, H., and Hager, G.L. (1987). Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. *EMBO J.* *6*, 2321–2328.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. *Nature* *311*, 532–537.

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

Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta* *1839*, 627.

Sanulli, S., Trnka, M.J., Dharmarajan, V., Tibble, R.W., Pascal, B.D., Burlingame, A.L., Griffin, P.R., Gross, J.D., and Narlikar, G.J. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nat.* 2019 5757782 *575*, 390–394.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* *436*, 138–141.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T.W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A.K., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* *482*, 226–231.

Scott, W.A., and Campos, E.I. (2020). Interactions With Histone H3 & Tools to Study Them. *Front. Cell Dev. Biol.* *8*.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* *19*, 65–71.

Serra-Cardona, A., and Zhang, Z. (2018). Replication-Coupled Nucleosome Assembly in the Passage of Epigenetic Information and Cell Identity. *Trends Biochem. Sci.* *43*, 136–148.

Shema, E., Bernstein, B.E., and Buenrostro, J.D. (2018). Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat. Genet.* 2018 511 *51*, 19–25.

Shipony, Z., Marinov, G.K., Swaffer, M.P., Sinnott-Armstrong, N.A., Skotheim, J.M., Kundaje, A., and

Greenleaf, W.J. (2020). Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* 2020 173 17, 319–327.

Simpson, R.T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* 17, 5524–5531.

Simpson, R.T., and Stafford, D.W. (1983). Structural features of a phased nucleosome core particle. *Proc. Natl. Acad. Sci. U. S. A.* 80, 51.

Smith, S., and Stillman, B. (1991). Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J.* 10, 971–980.

Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., and Shendure, J. (2016). Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* 164, 57–68.

Sobecki, M., Souaid, C., Boulay, J., Guerineau, V., Noordermeer, D., and Crabbe, L. (2018). MadID, a Versatile Approach to Map Protein-DNA Interactions, Highlights Telomere-Nuclear Envelope Contact Sites in Human Cells. *Cell Rep.* 25, 2891-2903.e5.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* (80-.). 344, 376–380.

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

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 2012 139 13, 613–626.

Steensel BV., and Henikoff S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* 18, 424–428.

Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* (80-.). 368.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory.

Nat. Cell Biol. 2020 224 22, 361–371.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J. Biol. Chem.* 281, 16279–16288.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–9445.

Suter, B., Schnappauf, G., and Thoma, F. (2000). Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.* 28, 4083.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Reports* 2019 91 9, 1–12.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res.* 25, 1491–1498.

Tse, O.Y.O., Jiang, P., Cheng, S.H., Peng, W., Shang, H., Wong, J., Chan, S.L., Poon, L.C.Y., Leung, T.Y., Chan, K.C.A., et al. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 118.

Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 83, 1021–1026.

Tullius, T.D. (1988). DNA footprinting with hydroxyl radical. *Nat.* 1988 3326165 332, 663–664.

Udugama, M., Sabri, A., and Bartholomew, B. (2011). The INO80 ATP-Dependent Chromatin Remodeling Complex Is a Nucleosome Spacing Factor. *Mol. Cell. Biol.* 31, 662.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–522.

Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nat.* 1997 3886642 388, 598–602.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 2015 163 16, 178–189.

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen,

E., et al. (2020). Global reference mapping of human transcription factor footprints. *Nat.* 2020 5837818 583, 729–736.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 2020 173 17, 261–272.

Wang, Y., Wang, A., Liu, Z., Thurman, A.L., Powers, L.S., Zou, M., Zhao, Y., Hefe, A., Li, Y., Zabner, J., et al. (2019). Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* 29, 1329–1342.

Wanior, M., Krämer, A., Knapp, S., and Joerger, A.C. (2021). Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy. *Oncogene* 2021 4021 40, 3637–3654.

Weber, C.M., and Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. *Genes Dev.* 28, 672.

Weber, C.M., Hafner, A., Kirkland, J.G., Braun, S.M.G., Stanton, B.Z., Boettiger, A.N., and Crabtree, G.R. (2021). mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol.* 2021 286 28, 501–511.

Weintraub, H., and Groudine, M. (1976). Chromosomal Subunits in Active Genes Have an Altered Conformation. *Science* (80-). 193, 848–856.

Wiechens, N., Singh, V., Gkikopoulos, T., Schofield, P., Rocha, S., and Owen-Hughes, T. (2016). The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLOS Genet.* 12, e1005940.

Wilson, B.G., Wang, X., Shen, X., McKenna, E.S., Lemieux, M.E., Cho, Y.J., Koellhoffer, E.C., Pomeroy, S.L., Orkin, S.H., and Roberts, C.W.M. (2010). Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* 18, 316–328.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 1–5.

Workman, J.L. (2006). Nucleosome displacement in transcription. *Genes Dev.* 20, 2009–2017.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nat.* 2016 5327599 532, 329–333.

Xiao, C. Le, Zhu, S., He, M., Chen, D., Zhang, Q., Chen, Y., Yu, G., Liu, J., Xie, S.Q., Luo, F., et al. (2018). N 6 -Methyladenine DNA Modification in the Human Genome. *Mol. Cell* 71, 306-318.e7.

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., and Wu, C. (2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* 8, 531–543.

Xu, M., Long, C., Chen, X., Huang, C., Chen, S., and Zhu, B. (2010). Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. *Science* (80-.). 328, 94–98.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. *Mol. Cell* 2, 851–861.

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

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

Yang, Z., Zheng, C., and Hayes, J.J. (2007). The core histone tail domains contribute to sequence-dependent nucleosome positioning. *J. Biol. Chem.* 282, 7930–7938.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* 149, 1461–1473.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364.

- Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* *25*, 2227–2241.
- Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* *37*, 76–81.
- Zentner, G.E., and Henikoff, S. (2014). High-resolution digital profiling of the epigenome. *Nat. Rev. Genet.* *15*, 814–827.
- Zhang, Y., and Li, Y. (2011). The Expanding Mi-2/NuRD Complexes: A Schematic Glance: <https://doi.org/10.4137/PRI.S6329> *3*, 79–109.
- Zhang, W., Feng, J., and Li, Q. (2020). The replisome guides nucleosome assembly during DNA replication. *Cell Biosci.* *2020* *101* *10*, 1–14.
- Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* (80-.). *332*, 977–980.
- Zhou, C.Y., Johnson, S.L., Gamarra, N.I., and Narlikar, G.J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors. *Annu. Rev. Biophys.* *45*, 153–181.
- Zhou, C.Y., Johnson, S.L., Lee, L.J., Longhurst, A.D., Beckwith, S.L., Johnson, M.J., Morrison, A.J., and Narlikar, G.J. (2018). The yeast INO80 complex operates as a tunable DNA length-sensitive switch to regulate nucleosome sliding. *Mol. Cell* *69*, 677.
- Zhou, V.W., Goren, A., and Bernstein, B.E. (2010). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* *2011* *121* *12*, 7–18.
- Zofall, M., Persinger, J., and Bartholomew, B. (2004). Functional Role of Extranucleosomal DNA and the Entry Site of the Nucleosome in Chromatin Remodeling by ISW2. *Mol. Cell. Biol.* *24*, 10047–10057.

**Chapter 2: Massively multiplex single-molecule
oligonucleosome footprinting**

2.1 Abstract

Our understanding of the beads-on-a-string arrangement of nucleosomes has been built largely on high-resolution sequence-agnostic imaging methods and sequence-resolved bulk biochemical techniques. To bridge the divide between these approaches, we present the single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA). SAMOSA is a high-throughput single-molecule sequencing method that combines adenine methyltransferase footprinting and single-molecule real-time DNA sequencing to natively and nondestructively measure nucleosome positions on individual chromatin fibers. SAMOSA data allows unbiased classification of single-molecular 'states' of nucleosome occupancy on individual chromatin fibers. We leverage this to estimate nucleosome regularity and spacing on single chromatin fibers genome-wide, at predicted transcription factor binding motifs, and across human epigenomic domains. Our analyses suggest that chromatin is comprised of both regular and irregular single-molecular oligonucleosome patterns that differ subtly in their relative abundance across epigenomic domains. This irregularity is particularly striking in constitutive heterochromatin, which has typically been viewed as a conformationally static entity. Our proof-of-concept study provides a powerful new methodology for studying nucleosome organization at a previously intractable resolution and offers up new avenues for modeling and visualizing higher order chromatin structure.

2.2 Current methods to study nucleosome positioning and their limitations

The nucleosome is the atomic unit of chromatin. Nucleosomes passively and actively template the majority of nuclear interactions essential to life by determining target site access for transcription factors (Spitz and Furlong, 2012), bookmarking active and repressed chromosomal compartments via post-translational modifications (Zhou et al., 2010), and safeguarding the genome from mutational agents (Papamichos-Chronakis and Peterson, 2012). Our earliest views of the beads-on-a-string arrangement of chromatin derived from classical electron micrographs of chromatin fibers (Olins and Olins, 1974), which have since been followed by both light (Huang et al., 2010) and electron microscopy (Ou et al., 2017; Song

et al., 2014) studies of *in vitro*-assembled and *in vivo* chromatin. In parallel, complementary biochemical methods using nucleolytic cleavage have successfully mapped the subunit architecture of chromatin structure at high resolution. These cleavage-based approaches can be stratified into those that focus primarily on chromatin accessibility (Klemm et al., 2019) (i.e. measuring ‘competent’ active chromatin (Weintraub and Groudine, 1976)), and those that survey nucleosomal structure uniformly across active and inactive genomic compartments. Understanding links between chromatin and gene regulation requires sensitive methods in all three of these broad categories: in this study, we advance our capabilities in the third, focusing on a novel method to map oligonucleosomal structures genome-wide.

The existing methods to study nucleosome positioning historically use cleavage reagents (e.g. dimethyl sulphate (Becker et al., 1986), hydroxyl radicals (Tullius, 1988), nucleases (Hewish and Burgoyne, 1973)) followed by gel electrophoresis and/or Southern blotting to map the abundance, accessibility, and nucleosome repeat lengths (NRLs) of chromatin fibers (Richard-Foy and Hager, 1987). More recently, these methods have been coupled to high-throughput short-read sequencing (Zentner and Henikoff, 2014), enabling genome-wide measurement of average nucleosome positions. While powerful, all these methods share key limitations: measurement of individual protein-DNA interactions inherently requires destruction of the chromatin fiber and averaging of signal across many short molecules. These limitations extend even to single-molecule methyltransferase-based approaches (Kelly et al., 2012; Krebs et al., 2017b; Nabils et al., 2014), which have their own biases (e.g. CpG/GpC bias; presence of endogenous m⁵dC in mammals; DNA damage due to bisulfite conversion), and are still subject to the short-length biases of Illumina sequencers. While single-cell (Lai et al., 2018; Pott, 2017) and long-read single-molecule (Baldi et al., 2018) genomic approaches have captured some of this lost contextual information, single-cell data are generally sparse and single-molecule Array-seq data must be averaged over multiple molecules. Ultimately, these limitations have hindered our understanding of how combinations of ‘oligonucleosomal patterns’ (i.e., discrete states of nucleosome positioning and regularity on single DNA molecules) give rise to active and silent chromosomal domains.

2.3 Applying long-read sequencing to detect genome patterns at high resolution

Massively parallel sequencing utilized in conjunction with tools that interrogate the epigenome at base-pair resolution have revolutionized our understanding of nucleosome positions at a genome-wide scale. As mentioned above, existing techniques to map nucleosomes at base-pair resolution (MNase-seq, DNA methylation footprinting) are limited by Illumina short-read sequencing, harshness of bisulfite conversion which destroys chromatin fibers, and bulk nucleosome measurements. We therefore require a quantitative and single-molecule tool which limits these biases and precisely maps long oligonucleosomes on a genome-wide scale.

The advent of third-generation (i.e. high-throughput, long-read) sequencing discussed in Chapter 1 offers a potential solution to many of these issues (Shema et al., 2018). Here, we demonstrate Single-molecule Adenine Methylated Oligonucleosome Sequencing Assay (SAMOSA), a method that combines adenine methyltransferase footprinting of nucleosomes with base modification detection on the PacBio single-molecule real-time sequencer (Flusberg et al., 2010a) to measure nucleosome positions on single chromatin templates. We first present proof-of-concept of SAMOSA using gold-standard *in vitro* assembled chromatin fibers, demonstrating that our approach captures single-molecule nucleosome positioning at high-resolution. We next apply SAMOSA to oligonucleosomes derived from K562 cells to profile single-molecule nucleosome positioning genome-wide. Our data enables unbiased classification of oligonucleosomal patterns across both euchromatic and heterochromatic domains. These patterns are influenced by multiple epigenomic phenomena, including the presence of predicted transcription factor binding motifs and post-translational histone modifications. Consistent with estimates from previous studies, our approach reveals enrichment for long, regular chromatin arrays in actively elongating chromatin, and highly accessible, disordered arrays at active promoters and enhancers. Surprisingly, we also observe a large amount of heterogeneity within constitutive heterochromatin domains, with both mappable H3K9me3-decorated regions and human major satellite sequences harboring a mixture of

irregular and short-repeat-length oliognucleosome types. Our study provides a proof-of-concept framework for studying chromatin at single-molecule resolution while suggesting a highly dynamic nucleosome-DNA interface across chromatin sub-compartments.

2.4 Single-molecule real-time sequencing of adenine-methylated chromatin captures nucleosome footprints

Existing methyltransferase accessibility assays either rely on bisulfite conversion (Kelly et al., 2012; Krebs et al., 2017b; Nabils et al., 2014) or use the Oxford Nanopore platform to detect DNA modifications (Oberbeckmann et al., 2019; Shipony et al., 2020; Wang et al., 2019). We hypothesized that high-accuracy PacBio single-molecule real-time sequencing could detect m⁶dA deposited on chromatin templates to natively measure nucleosome positioning. To test this hypothesis, we used the nonspecific adenine methyltransferase EcoGII (Murray et al., 2018b) to footprint nonanucleosomal chromatin arrays generated through salt-gradient dialysis (**Figure 2.1-figure supplement 2.1**), using template DNA containing nine tandem repetitive copies of the Widom 601 nucleosome positioning sequence (Lowary and Widom, 1998b) separated by ~46 basepairs (bp) of linker sequence followed by ~450 bp of sequence without any known intrinsic affinity for nucleosomes. After purifying DNA, polishing resulting ends, and ligating on barcoded SMRTBell adaptors, we subjected libraries to sequencing on PacBio Sequel or Sequel II flow cells, using unmethylated DNA and methylated naked DNA as controls (**Figure 2.1A**). After filtering low quality reads, we analyzed a total of 33,594 single molecules across all three conditions. Across both platforms, we observed higher average interpulse duration (IPD) in samples exposed to methyltransferase, consistent with a rolling circle polymerase ‘pausing’ at methylated adenine residues in template DNA (**Figure 2.1—figure supplement 2.2**). Further inspection of footprinted chromatin samples sequenced on either platform revealed strong specificity for altered IPD values only at thymines falling outside Widom 601 repeat sequences, in contrast with fully methylated naked template and unmethylated controls (**Figure 2.1—figure supplement 2.3A,B**). These patterns were subtly influenced by the associated

10-mer context of sequenced bases, consistent with possible enzymatic biases, but also previous observations of sequence-influenced shifts in polymerase kinetics (**Figure 2.1—figure supplement 2.4**; (Feng et al., 2013)). These results suggest that the PacBio platform can natively detect ectopic m⁶dA added to chromatinized templates.

We next developed a computational approach to assign a posterior probability describing the likelihood that an A/T basepair is methylated given IPD signals found within the same molecule (i.e., ‘modification probability’). We then paired this approach with a simple peak-calling strategy to approximate nucleosomal dyad positions. To benchmark this pipeline, we first calculated the distance between called nucleosome dyads and expected 601 dyad positions (**Figure 2.1B**). Observed dyads were highly concordant with expected positions (median \pm median absolute deviation [MAD] = 4 ± 2.97 bp), consistent with our data accurately capturing the expected 601 dyad. We next calculated the expected distances between nucleosomes given our dyad callset (i.e. a computationally defined nucleosome repeat length [NRL]; **Figure 2.1C**). Compared with the expected repeat length of 193 bp, our calculated results were similarly accurate at both two-dyad resolution (pairwise distance between adjacent dyads; median \pm MAD = 193 ± 7.40 bp) and averaged single-molecule resolution (median \pm MAD = 192 ± 1.30 bp). Both these measurements were qualitatively uniform across all molecules, independent of the positions of individual nucleosomes along individual array molecules (**Figure 2.1—figure supplement 2.5**). Finally, we directly visualized the modification probabilities of individual sequenced chromatin molecules and observed that modification patterns occurred in expected linker sequences (**Figure 2.1D**), and not in unmethylated or fully methylated control samples (**Figure 2.1—figure supplement 2.6A,B**). These results demonstrate that EcoGII footprinting is specific for unprotected DNA and that kinetic deviations observed in the data are not simply the result of primary sequence biases in the template itself. We hereafter refer to this approach as SAMOSA.

2.5 SAMOSA captures regular nucleosome-DNA interactions in vivo through nuclease-cleavage and adenine-methylation simultaneously

Having shown that SAMOSA can footprint *in vitro* assembled chromatin fibers, we sought to apply our approach to oligonucleosomal fragments from living cells. Multiple prior studies have suggested that a light micrococcal nuclease (MNase) digest followed by disruption of the nuclear envelope and overnight dialysis can be used to gently liberate oligonucleosomes into solution without dramatically perturbing nucleosomal structure (Ehrensberger et al., 2015; Gilbert and Allan, 2001; Gilbert et al., 2004). After lightly digesting and solubilizing oligonucleosomes from human K562 nuclei, we methylated chromatin with EcoGII and sequenced methylated molecules on the Sequel II platform (n = 1,855,316 molecules total; **Figure 2.2A**). As controls, we also shallowly sequenced deproteinated K562 oligonucleosomal DNA, and deproteinated oligonucleosomal DNA methylated with the EcoGII enzyme.

In vivo SAMOSA has several advantages compared to existing MNase- or methyltransferase-based genomic approaches. Our approach combines MNase-derived cuts flanking each fragment with methyltransferase footprinting of nucleosomes. MNase cuts mark the boundary of genomic ‘barrier’ elements like nucleosomes and can be tuned by modifying digestion conditions; accordingly, fragment length distributions from *in vivo* SAMOSA data display patterns emblematic of bulk nucleosomal array regularity (**Figure 2.2B**; **Figure 2.2—figure supplement 2.1**). Modification patterns of sequenced molecules then capture nucleosome-positioning information at single-molecule resolution; this is evident in single-molecule averages of modification probability in chromatin samples with respect to fully methylated and unmethylated controls (**Figure 2.2C**). While previous approaches for studying nucleosome regularity may capture each of the former information types, this method is, to our knowledge, the first that simultaneously captures the positioning of protein-DNA interactions through nucleolytic cleavage, and (through DNA methylation) the positioning of proximal protein-DNA interactions on the same single-molecule.

2.6 SAMOSA enables unbiased classification of chromatin fibers on the basis of regularity and nucleosome repeat length

The relative abundance and diversity of oligonucleosome patterns across the human genome remains unknown. Given the single-molecule nature of SAMOSA, we speculated that our data could be paired with a state-of-the-art community detection algorithm to systematically cluster footprinted molecules on the basis of single-molecule nucleosome regularity and NRL (i.e. ‘oligonucleosome patterns’). To ease detection of signal regularity on single molecules, we computed autocorrelograms for each molecule in our dataset ≥ 500 bp in length, and subjected resulting values to unsupervised Leiden clustering (Traag et al., 2019). Cluster sizes varied considerably, but were consistent across both replicates, with each cluster containing 6.54% (Cluster 4)–20.1% (Cluster 1) of all molecules (**Figure 3A**). The resulting seven clusters (Figure 3—figure supplement 1A) capture the spectrum of oligonucleosome patterning genome-wide, stratifying the genome by both NRL and array regularity. Accounting for the coverage biases presented above, the measurements shown in **Figure 3A** provide a rough estimate of the equilibrium composition of the genome with respect to these patterns.

The diversity in nucleosome regularity and repeat length across these clusters is visually apparent when inspecting average modification probabilities of the 5' 1000 bp of each cluster (**Figure 2.3B**). To better annotate each of these clusters, we characterized each with respect to methylation extent and distribution of computed single-molecule NRLs. We first inspected the average modification probabilities of each molecule across clusters, finding that these averages were largely invariant (**Figure 2.3—figure supplement 2.1B**). This suggests that our clustering approach does not simply classify oligonucleosomes based on the amount of methylation on each molecule. We next estimated within-cluster heterogeneity in single-molecule NRLs using a simple peak-calling approach. We scanned each autocorrelogram for secondary peaks, and annotated the location of each peak to compute an estimated NRL. We then visualized these distributions as violin plots for each cluster (Figure 3C). Our data broadly fall into two categories: irregular clusters made up of molecules spanning multiple NRLs and lacking a strong regular periodicity,

and highly regular clusters with defined single-molecule NRLs ranging from ~172 bp (i.e. chromosome plus 5 bp DNA) to >200 bp. Based on the median NRLs and regularities inferred from these analyses, we named these clusters irregular-short (IRS), irregular-long (IRL), irregular-170 (IR170), regular repeat length 172 (NRL172), regular repeat length 187A and B (NRL187A/B), and regular repeat length 192 (NRL192). The difference between irregular and regular clusters is clear when closely inspecting histograms of NRL calls from selected clusters (**Figure 2.3D**; **Figure 2.3—figure supplement 2.1C**), as well as the modification patterns on individual molecules (**Figure 2.3E**). Our analyses also varied with respect to the fraction of molecules per cluster where a secondary peak could be detected (0.50%–38.2% of molecules across specific clusters; **Figure 2.3—figure supplement 2.1D**). Failure to detect a peak within a single-molecule autocorrelogram could be due to multiple factors, including technical biases (e.g. random undermethylated molecules). We observed, however, that more ‘missing’ NRL estimates occurred in irregular clusters, suggesting that at least a fraction of failed peak calls occurred due to lack of intrinsic regularity along individual footprinted molecules. These analyses together demonstrate that SAMOSA data can be clustered in an unbiased manner, thus enabling estimates of the equilibrium composition of the genome with respect to oligonucleosome regularity and repeat length.

2.7 SAMOSA captures the transient nucleosome occupancy of transcription factor binding motifs

We next explored the extent to which our data captures chromatin structure at predicted K562 transcription factor (TF)-binding sites (ENCODE Project Consortium, 2012 (Dunham et al., 2012b)). Both endo- and exo-nucleolytic MNase cleavage activities are obstructed by genomic protein-DNA contacts; resulting fragment-ends thus capture both nucleosomal- and TF-DNA interactions (Henikoff et al., 2011; Ramani et al., 2019). Inspection of cleavage patterns about six different TF-binding sites (CTCF, NRF1, NRSF/REST, PU.1, c-MYC, GATA1) (**Figure 2.4A–F**) revealed signal resembling traditional MNase-seq data, with fragment ends accumulating immediately proximal to predicted TF-binding motifs, and, in the case of some TFs (i.e. CTCF, REST, PU.1), showed characteristic patterns of phased nucleosomes. Further

analysis of m⁶ dA signal in sequenced molecules harboring motifs with at least 500 nucleotides of flanking DNA revealed examples of methyltransferase accessibility coincident with TF motifs (e.g. CTCF, NRF1, c-MYC), but also cases where single-molecule averages demonstrated weak or no differential signal when compared to equal numbers of molecules drawn from random genomic regions matched for GC-percentage and repeat content (e.g. GATA1; **Figure 2.4G–L**). Importantly, our methylation data do not appear to capture TF ‘footprints’ as seen in DNase I, hydroxyl radical, or MNase cleavage data—this could be due to turnover of transcription factors during our solubilization process, or owed to sterics, as EcoGII is roughly twice the molecular weight of *S. aureus* micrococcal nuclease (Murray et al., 2018b).

In theory, single-molecule footprinting data should distinguish nucleosome-bound and nucleosome-free states for molecules containing TF-binding sites. These accessibility patterns should be specific to TF-binding motifs (i.e. not present in control molecules matched for GC/repeat content). To test whether our assay captured such signal, we clustered all molecules shown in **Figure 2.4G–L** (including control molecules) using Leiden clustering, using modification probabilities extracted in a 500 bp window surrounding the predicted motif site/control site. In total, we defined 13 discrete states of template accessibility across all surveyed molecules (**Figure 2.4M**; cluster sizes shown in **Figure 2.4—figure supplement 2.1**). We interpreted these states on the basis of methyltransferase accessibility as: methyltransferase-resistant motifs (MR); nucleosome-occluded motifs (NO1-8); stochastically accessible motifs (wherein motif accessibility is slightly elevated near the DNA entry/ exit point of a footprinted nucleosome; SA1-2); accessible motifs (A); and hyper-accessible motifs (HA). Notably, the patterns within these clusters were evident at single-molecule resolution (**Figure 2.4N**). Most transcription factors (excepting PU.1 and GATA1—the latter of which may productively bind nucleosomal DNA (Zaret and Carroll, 2011)) were significantly enriched for specific states as defined above, and all control regions were markedly depleted for molecules harboring the accessible ‘A’ and ‘HA’ states, hinting at the biological relevance of these patterns (**Figure 2.5A**). We speculate that the broad distribution of these states across both TF-binding sites and controls represent distributions of nucleosome ‘registers’ surrounding typical

transcription factor binding motifs (i.e. states MR; NO-1–8). A fraction of these registers (i.e. states SA1/2) may stochastically permit transcription factor binding (perhaps through transient unwrapping of the nucleosome (Polach and Widom, 1995)), enabling formation of a new nucleosome register (i.e. state ‘A’), and subsequent generation of a highly accessible state (‘HA’; model illustrated in **Figure 2.5B**). The relative fraction of molecules in an ‘SA’ state could conceivably be modulated by TF intrinsic properties (e.g. ability to bind partially nucleosome-wrapped DNA (Zaret and Mango, 2016)), or extrinsic factors (e.g. local concentration of ATP-dependent chromatin remodeling enzymes (Narlikar et al., 2013)). While correlation of our replicates demonstrates the reproducibility and robustness of these findings (**Figure 2.5—figure supplement 2.1**), future experimental follow-up coupling our protocol with perturbed biological systems and deeper sequencing are necessary to quantitatively interrogate this model.

2.8 Heterogeneous oligonucleosome patterns comprise human epigenomic domains

Short-read and long-read sequencing of nucleolytic fragments in mammals have suggested that NRLs vary across epigenomic domains, with euchromatin harboring shorter NRLs on average and heterochromatic domains harboring longer NRLs (Gaffney et al., 2012; Snyder et al., 2016; Valouev et al., 2011), but the relative heterogeneity of these domains remains unknown. We speculated that SAMOSA data could be used to estimate single-molecule oligonucleosome pattern heterogeneity across epigenomic domains. We revisited the seven oligonucleosome patterns defined above and examined the distribution of patterns across collections of single molecules falling within ENCODE-defined H3K4me3, H3K4me1, H3K36me3, H3K27me3, and H3K9me3-decorated chromatin domains. To control for the impact of GC-content on these analyses, we also included GC-/ repeat content matched control molecules for each epigenomic mark surveyed. Furthermore, to take advantage of the long-read and relatively unbiased nature of our data, we also incorporated molecules deriving from typically unmappable human alpha, beta, and gamma satellite DNA sampled directly from raw CCS reads.

We visualized the relative heterogeneity of these domains and controls in two ways: using histograms of computed single-molecule NRL estimates (**Figure 2.6A**), and by using stacked bar graphs to visualize cluster membership (**Figure 2.6B**). A striking finding of our analyses was that each epigenomic domain surveyed was comprised of a highly heterogeneous mixture of oligonucleosome patterns. In most cases, these patterns differed only subtly from control molecules with respect to regularity and NRL. In specific cases, we observed small effect shifts in the estimated median NRLs for specific domains—for example, a shift of ~5 bp (180 bp vs. 185 bp) in H3K9me3 chromatin with respect to random molecules, and a shift of ~4 bp (182 bp vs 186 bp) for H3K36me3. These shifts were also evident in the fraction of molecules with successful peak calls: H3K4me3 decorated chromatin, for example, had markedly fewer (78.0% vs 88.6%) successful calls compared to control molecules, a finding consistent with the expected irregularity of active promoter oligonucleosomes. We note that all these measured parameters would be unattainable using any existing biochemical method and that these preliminary findings argue against the abundance of homogeneous oligonucleosome structures in either heterochromatic or euchromatic nuclear regions.

On first glance, our data appear to run counter to previous observations demonstrating that epigenomic domains can be delineated by differences in bulk nucleosome positioning as measured by nuclease digestion. One possible explanation for this is that epigenomic domains subtly, but significantly, vary in their relative composition of distinct oligonucleosome patterns, and the resulting average of these differences is the signal captured in MNase-Southern and other cleavage-based measurements. We tested this hypothesis by constructing a series of statistical tests to determine whether each of the seven defined oligonucleosome patterns were significantly enriched or depleted across chromatin domains and matched control regions (**Figure 2.6C; reproducibility analyses summarized in Figure 2.6—figure supplement 2.1**). Our results suggest that chromatin domains are demarcated by their relative usage of specific oligonucleosome patterns. Consistent with expectations, active chromatin marked by H3K4me3 and H3K4me1 are punctuated by a mixture of irregular oligonucleosome patterns (namely, clusters IRL and IR170). For transcription elongation associated H3K36me3 decorated chromatin, both short-read mapping

in human and long-read bulk array regularity mapping in *D. melanogaster* have suggested relatively short, regular nucleosome repeat lengths (Baldi et al., 2018; Valouev et al., 2011). Our data partially corroborate this finding in human K562 cells: H3K36me3-domains are punctuated by irregular IRS oligonucleosome patterns (Fisher's Exact Odds Ratio [O.R.]=1.13; $q = 1.71E-50$) and regular, short NRL172 patterns (O.R. = 1.39; $q = 3.69E-170$).

Our assay also allows us to assess compositional biases in heterochromatic domains. Short-read-based human studies and classical MNase mapping of constitutive heterochromatin have suggested that H3K9me3-decorated chromatin harbor (i) long nucleosome repeat lengths on average, and (ii) are highly regular. These estimates are susceptible to artifacts, as heterochromatic nucleosomes are expected to be both strongly phased and weakly positioned. Our data partially disagree with prior estimates—across both H3K9me3 and Satellite molecules we observe enrichment for irregular IRS nucleosome conformers (Satellite O.R. = 1.13; $q = 5.71E-11$; H3K9me3 O.R. = 1.35; $q = 3.95E-23$). Still, these enriched conformers were accompanied by enrichment for regular NRL172 oligonucleosome patterns for both states (Satellite O.R. = 1.61; $q = 5.25E-80$; H3K9me3 O.R. = 1.23; $q = 3.86E-6$). These analyses demonstrate that prior NRL estimates by short-read sequencing may have been confounded by *in vivo* heterogeneity in nucleosome positions, that heterochromatic nucleosome conformations can be both irregular and diverse, and finally, highlight the value of SAMOSA for accurately studying nucleosome structure in heterochromatin.

Taken as a whole, our data suggest two fundamental properties of human epigenomic domains: first, epigenomic domains are comprised of a diverse array of oligonucleosome patterns varying substantially in intrinsic regularity and average distance between regularly spaced nucleosomes; second: epigenomic domains are demarcated by their usage of these oligonucleosome patterns. We find that all epigenomic states are characterized by a diverse mixture of oligonucleosomal conformers - many conformational states are neither significantly depleted nor enriched with respect to all molecules surveyed, further hinting at the diverse composition of chromatin domains genome-wide.

2.9 Discussion

Here, we present the SAMOSA, a method for resolving nucleosome-DNA interactions using the EcoGII adenine methyltransferase and PacBio single-molecule real-time sequencing. Our approach has multiple advantages over existing methyltransferase-based sequencing approaches: first, by using a relatively nonspecific methyltransferase, we avoid the primary sequence biases associated with GpC/ CpG methyltransferase footprinting methods; second, by natively detecting modifications using the single-molecule real-time sequencer, we reduce enzymatic sequence bias and avoid sample damage associated with sodium bisulfite conversion; finally, and most importantly, our approach unlocks the study of protein-DNA interactions at length-scales previously unallowed by Illumina sequencing.

Our study does have limitations. While the current SAMOSA protocol enriches fragments ranging from ~500 bp to ~ 2 kb in size, high-quality PacBio CCS sequencing is compatible with fragments ranging from 10 to 15 kbp. We anticipate that with further optimization (e.g. optimization of digestion conditions), SAMOSA will be applicable to longer arrays, enabling kilobase-domain-scale study of single-molecule oligonucleosome patterning. Indeed, our preliminary SAMOSA experiments varying digestion conditions demonstrate the feasibility of such variations (**Figure 2.2—figure supplement 2.1**). Second, our approach involves methylating fibers, following solubilization of oligonucleosomal fragments, and is thus unlikely to capture protein-DNA interactions weaker or more transient than the stable nucleosome-DNA interaction. Such transient interactions could be captured in future work by modifying the protocol to footprint nuclei prior to MNase-solubilization. Third, our proof-of-concept was performed in unsynchronized K562 cells, and thus we cannot yet address the contribution of a biological process like the cell cycle to the observed heterogeneity. Finally, as a proof-of-concept our approach falls short of generating a high-coverage reference map of the K562 epigenome; as sequencing costs for PacBio decrease and sequence-enrichment technologies (e.g. CRISPR-based enrichment (Ebbert et al., 2018); SMRT-ChIP (Wu et al., 2016)) for the platform mature, SAMOSA may routinely be used to generate reference datasets with hundred-to-thousand-fold single-molecular coverage of genomic sites of interest.

Our data confirms that the human epigenome is made up of a diverse array of oligonucleosome patterns, including highly regular arrays of varying nucleosome repeat lengths, and irregular arrays where nucleosomes are positioned without a detectable periodic signature (Baldi et al., 2020). Our results broadly agree with a recent approach employing electron tomography to map the in situ structure of mammalian nuclei, which found chromatin to be highly heterogeneous at the length scale of multi-nucleosome interactions, and failed to detect evidence of a 30 nm fiber or other homogeneous higher order compaction states (Ou et al., 2017). At the sequencing depth presented here, these oligonucleosome patterns significantly, if subtly, vary across different epigenomic domains. Surprisingly, we find that both mappable (H3K9me3 ChIP-seq peaks) and unmappable (human satellite sequence) constitutive heterochromatin are enriched for irregular oligonucleosome patterns in addition to expected regular arrays—the presence of these irregular fibers may have been previously missed due to an understandable reliance on bulk averaged methods (e.g. MNase- Southern) for studying constitutive heterochromatin. This is strongly supported by orthogonal analysis of heterochromatin-spanning K562 reads generated using the recently published, conceptually similar Fiber-seq method (Stergachis et al., 2020), which also reveal that H3K9me3 domains are enriched for irregular chromatin fibers (**Figure 2.6—figure supplement 2.2**). Given the robustness of this finding, it is tempting to speculate that this irregularity may be linked to the dynamic restructuring of heterochromatic nucleosomes by factors like HP1 (Sanulli et al., 2019), which may promote phase-separation of heterochromatin. While stratification of analyzed satellite sequences into H3K9me3-decorated alpha/beta, and H3K9me3-free gamma satellite (Kim et al., 2009) provides correlative support for this notion (**Figure 2.6—figure supplement 2.3**), future studies combining SAMOSA with cellular perturbation of heterochromatin-associated factors are necessary to directly address this possibility.

More generally, future work employing our technique must focus on questioning the biological significance of this global heterogeneity: for example, is the fraction of stochastically accessible transcription factor binding sites (i.e. motif ‘site exposure’ frequency (Ahmad and Henikoff, 2001; Polach and Widom, 1995)) important for TF-DNA binding in nucleosome-occluded genomic regions? What is the

interplay between transcription factor ‘pioneering’ and stochastic site accessibility? What are the global roles of ATP-dependent chromatin remodeling enzymes (i.e. SWI/SNF; ISWI; INO80; CHD) in maintaining these patterns genome-wide (Brahma and Henikoff, 2020)? Our approach also unlocks a set of conceptual questions regarding the nature of chromatin secondary structure. Significant genome-wide efforts have revealed that metazoan epigenomes are punctuated by regions of concerted histone modification and subnuclear positioning (ENCODE Project Consortium, 2012 ((Dunham et al., 2012b), (Filion et al., 2010)), but approaches for studying the distribution of oligonucleosomal patterns associated within these same regions are lacking. Given recent work suggesting that NRLs can specify the ability of nucleosomal arrays to phase separate (Gibson et al., 2019), it is likely that SAMOSA and similar assays may provide an important bridge between *in vitro* biochemical observations of chromatin and *in vivo* genome-wide ‘catalogs’ of oligonucleosome patterning.

SAMOSA adds to the growing list of technologies that use high-throughput single-molecule sequencing to explore the epigenome (Baldi et al., 2018; Lee et al., 2019; Shipony et al., 2020; Stergachis et al., 2020; Wang et al., 2019). We foresee the broad applicability of this and similar approaches to dissect gene regulatory processes at previously intractable length-scales. Our approach and associated analytical pipelines demonstrate the versatility of high-throughput single-molecule sequencing—namely the ability to cluster single-molecules in an unsupervised manner to uncover molecular states previously missed by short-read approaches. Our analytical approach bears many similarities to methods used in single-cell analysis, and indeed many of the technologies and concepts typically used for single-cell genomics (Trapnell, 2015) (*e.g.* clustering; trajectory analysis) will likely have value when applied to single-molecule epigenomic assays. Our approach also follows in the footsteps of multi-omic Illumina assays like NoME-seq and MapIT, representing the first of what we anticipate will be many ‘multi-omic’ third-generation sequencing assays. As third-generation sequencing technologies advance, it will likely become possible to encode multiple biochemical signals on the same single-molecules, thus enabling causal inference of the logic and ordering of biochemical modifications on single chromatin templates.

2.10 Figures

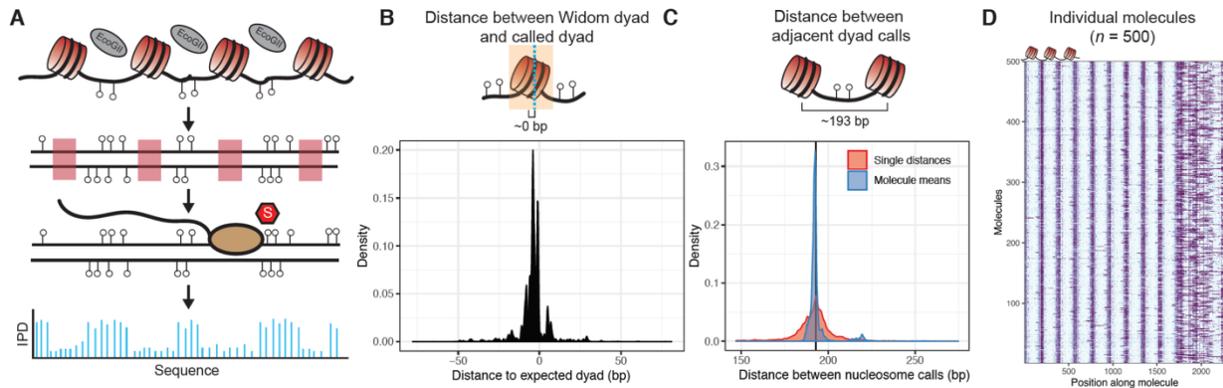


Figure 2. 1: Overview of the single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA).

A.) In the SAMOSA assay, chromatin is methylated using the nonspecific EcoGII methyltransferase, DNA is purified, and then subjected to sequencing on the PacBio platform. Modified adenine residues are natively detected during SMRT sequencing due to polymerase pausing, leading to an altered interpulse duration at modified residues.

B.) SAMOSA data can be used to accurately infer nucleosome dyad positions given a strong positioning sequence. Shown are the distributions of called dyad positions with respect to the known Widom 601 dyad. Called dyads fall within a few nucleotides of the expected dyad position (median \pm median absolute deviation [MAD] = 4 ± 2.97 bp).

C.) SAMOSA data accurately recapitulates the known nucleosome repeat lengths (NRL) of in vitro assembled chromatin fibers. Called NRLs are strongly concordant with the expected 193 repeat length (pairwise distance between adjacent dyads median \pm MAD = 193 ± 7.40 bp; single-molecule averaged repeat length median \pm MAD = 192 ± 1.30 bp).

D.) Expected nucleosome footprints in SAMOSA data can be visually detected with single-molecule resolution (n = 500 sampled footprinted chromatin molecules).

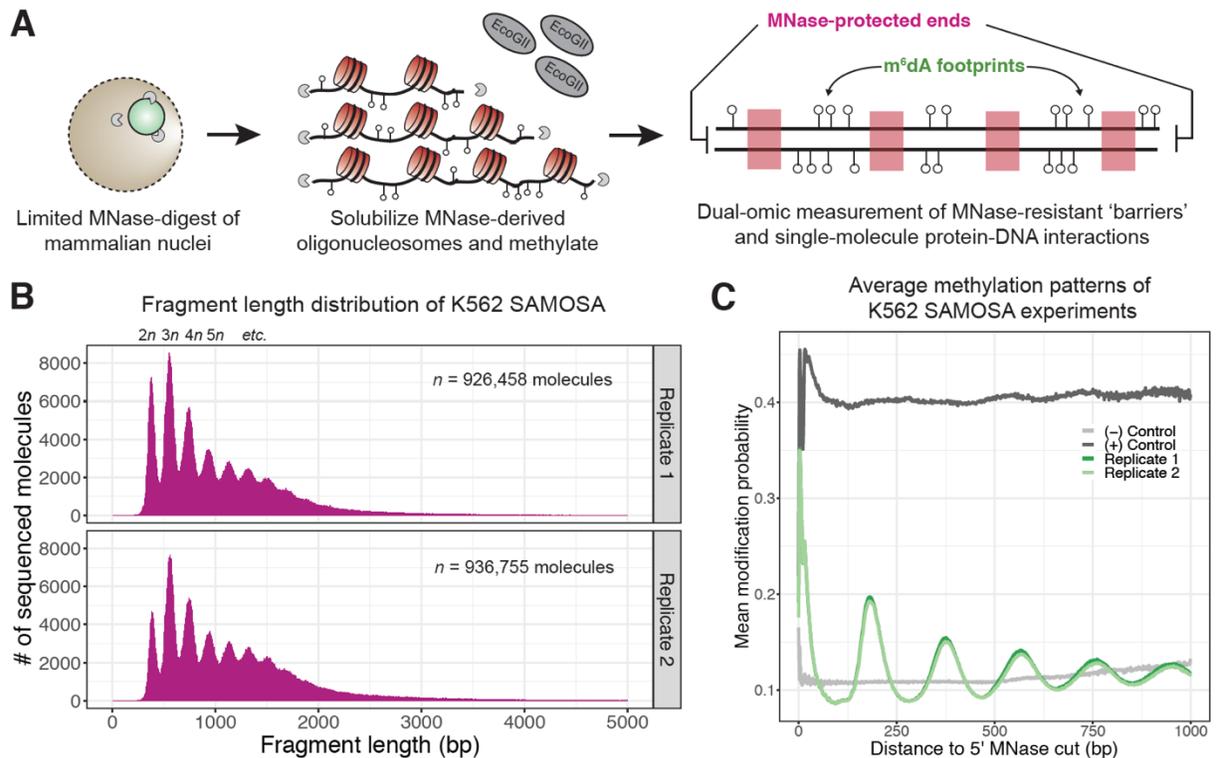


Figure 2. 2: In vivo SAMOSA captures oligonucleosome structure by combining MNase digestion of chromatin with adenine methylation footprinting.

A.) An overview of the in vivo SAMOSA protocol: oligonucleosomes are gently solubilized from nuclei using micrococcal nuclease and fusogenic lipid treatment. Resulting oligonucleosomes are footprinted using the EcoGII enzyme and sequencing on the PacBio platform. Each sequencing molecules captures two orthogonal biological signals: MNase cuts that capture 'barrier' protein-DNA interactions, and m^6dA methylation protein-DNA footprints.

B.) Fragment length distributions for in vivo SAMOSA data reveal expected oligonucleosomal laddering (bin size = 5 bp).

C.) Averaged modification probabilities from SAMOSA experiments demonstrate the ability to mark nucleosome-DNA interactions directly via methylation. Modification patterns seen in the chromatin sample are not seen in unmethylated oligonucleosomal DNA or fully methylated K562 oligonucleosomal DNA.

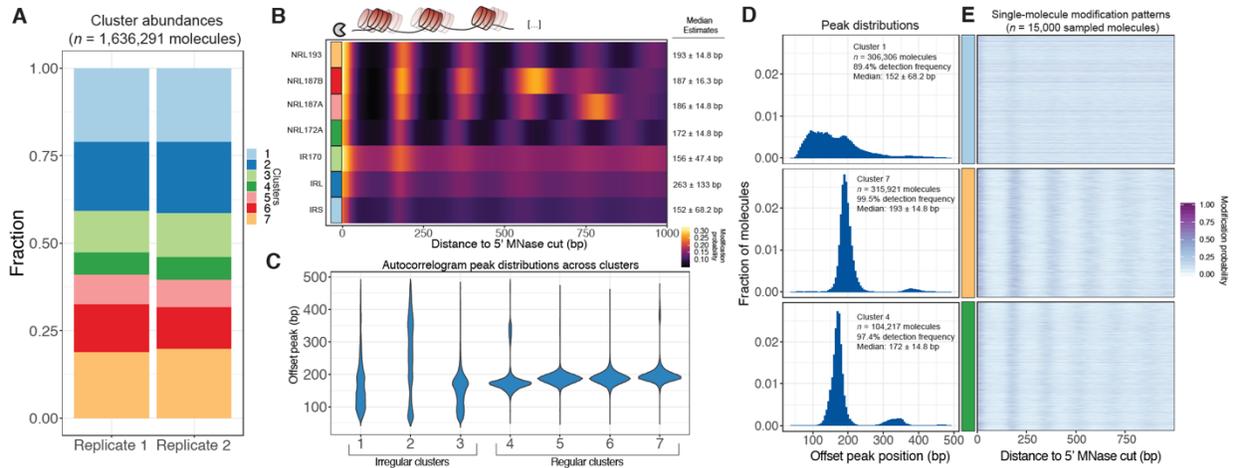


Figure 2. 3: SAMOSA reveals distribution of oligonucleosome patterns genome-wide.

A.) Stacked bar chart representation of the contribution of each cluster to overall signal across two replicate experiments in K562 cells.

B.) Average modification probability as a function of sequence for each of the seven defined clusters. Left: Manually annotated cluster names based on NRL estimates computed by calling peaks on single-molecule autocorrelograms; Right: Median and median absolute deviation for single-molecule NRL estimates determined for each cluster.

C.) Violin plot representation of the distributions of single-molecule NRL estimates for each cluster. Clusters can be separated into three 'irregular' and four 'regular' groups of oligonucleosomes.

D.) Histogram of single-molecule NRL estimates for Clusters 1, 4, and 7, along with

E.) 5,000 randomly sampled molecules from each cluster.

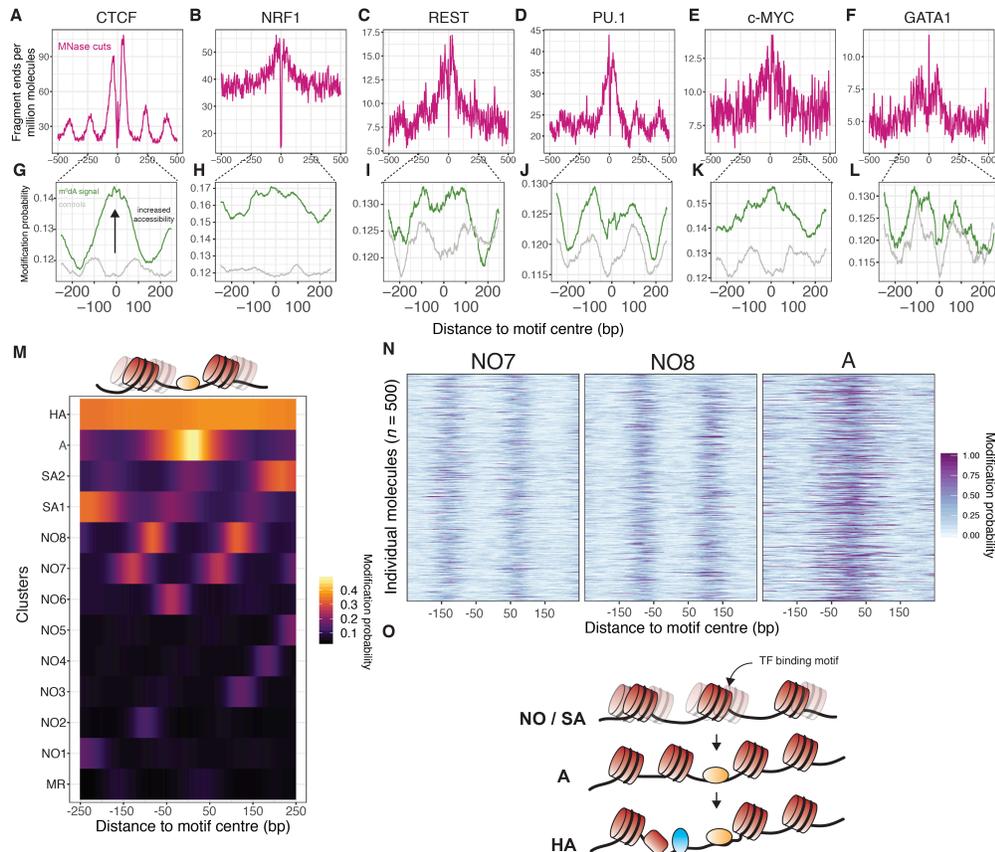


Figure 2. 4: SAMOSA captures bulk and single-molecule evidence of transcription factor-DNA interaction simultaneously via two orthogonal molecular signals.

A.-F.) SAMOSA MNase-cut signal averaged over predicted CTCF, NRF1, REST, PU.1, c-MYC, and GATA1 binding motifs in the K562 epigenome. All binding sites were predicted from ENCODE ChIP-seq data.

G.-L.) m⁶A signal for the same transcription factors, averaged over molecules containing predicted binding sites and at least 250 bases flanking DNA on either side of the predicted motif. Methylation patterns at predicted sites were compared against average profiles taken from randomly drawn molecules from GC%- and repeat-content-matched regions of the genome (calculated for each ENCODE ChIP-seq peak set).

M.) Results of clustering motif-containing molecules using the Leiden community detection algorithm. Clusters were manually annotated as containing molecules that were: ‘methylation resistant’ (MR), nucleosome occupied (NO1-8), stochastically accessible (SA1-2), accessible (A), or hyper-accessible (HA).

N.) Heatmap representation of single-molecule accessibility profiles for clusters NO7, NO8 , and A (500 randomly sampled molecules per cluster).

O.) Our data may be explained by the Widom ‘site exposure’ model in vivo. Transcription factor binding motifs are stochastically exposed as nucleosomes toggle between multiple ‘registers’ as seen in Figure 3M (states NO and SA). Transcription factor binding perhaps enforces a favorable nucleosome register (state A), which can then seed hyper-accessible states / further TF-DNA interactions (state HA).

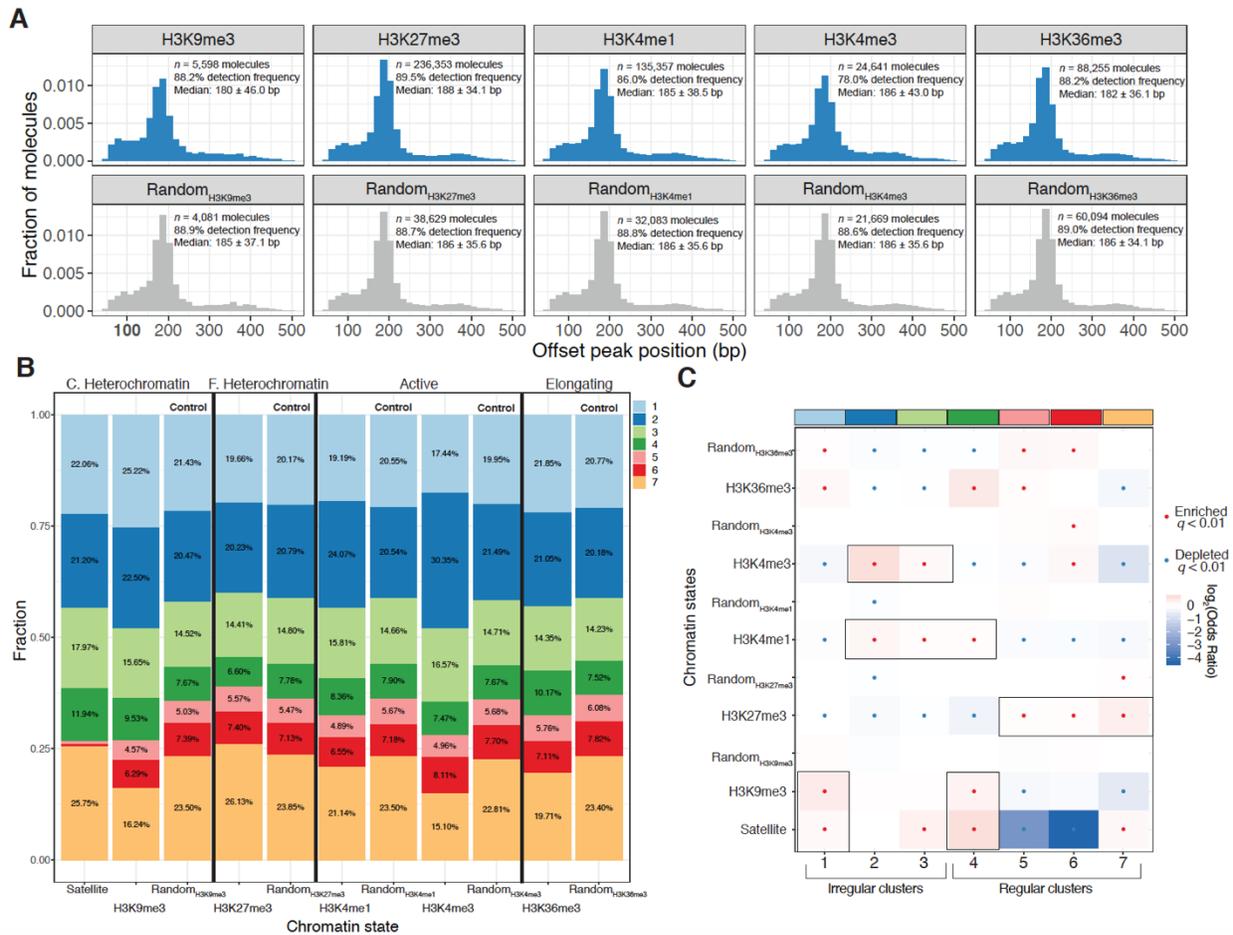


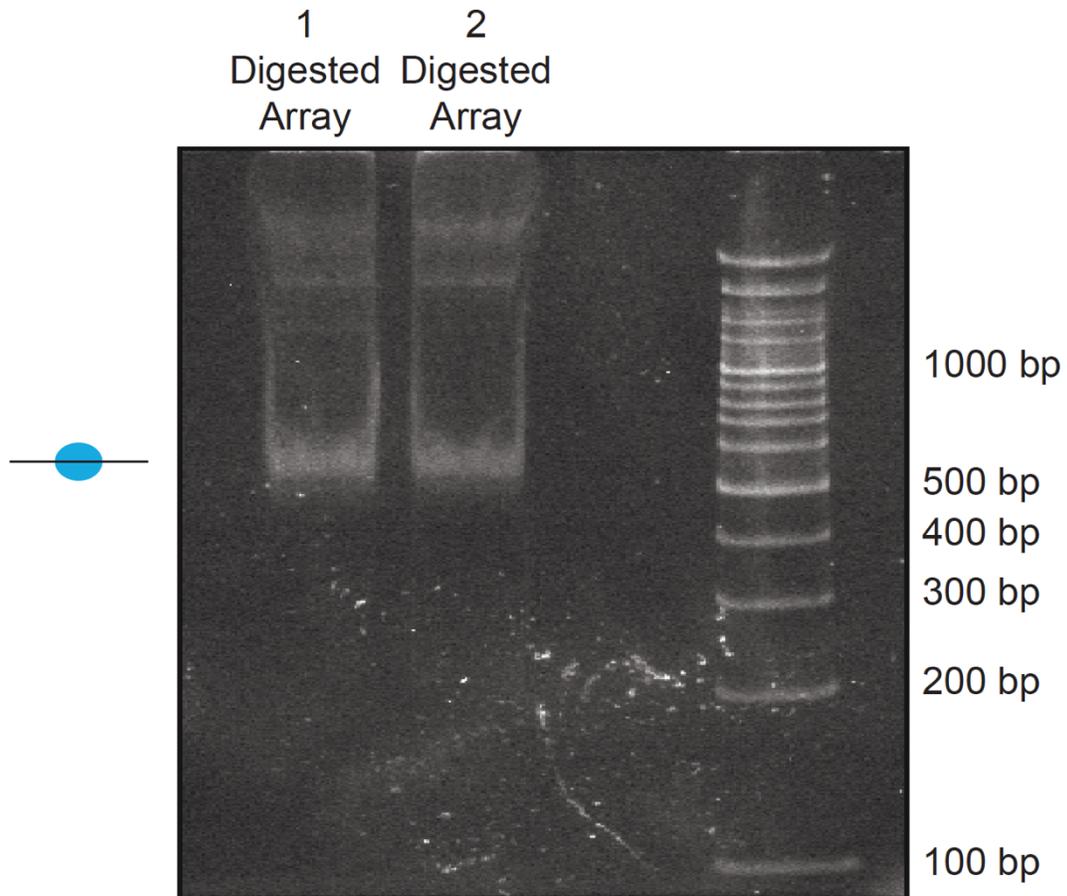
Figure 2. 5: Human epigenomic states are punctuated by specific oligonucleosome patterns.

A.) Histogram representations of the estimated single-molecule NRLs for five different epigenomic domains compared to control sets of molecules matched for GC and repeat content. Inset: Numbers of molecules plotted, median NRL estimates with associated median absolute deviations, and the percent of molecules where a peak could not be detected.

B.) Stacked bar chart representation of the relative composition of each epigenomic domain with respect to the seven clusters defined in Figure 3. C. Heterochromatin: constitutive heterochromatin; F. Heterochromatin: facultative heterochromatin.

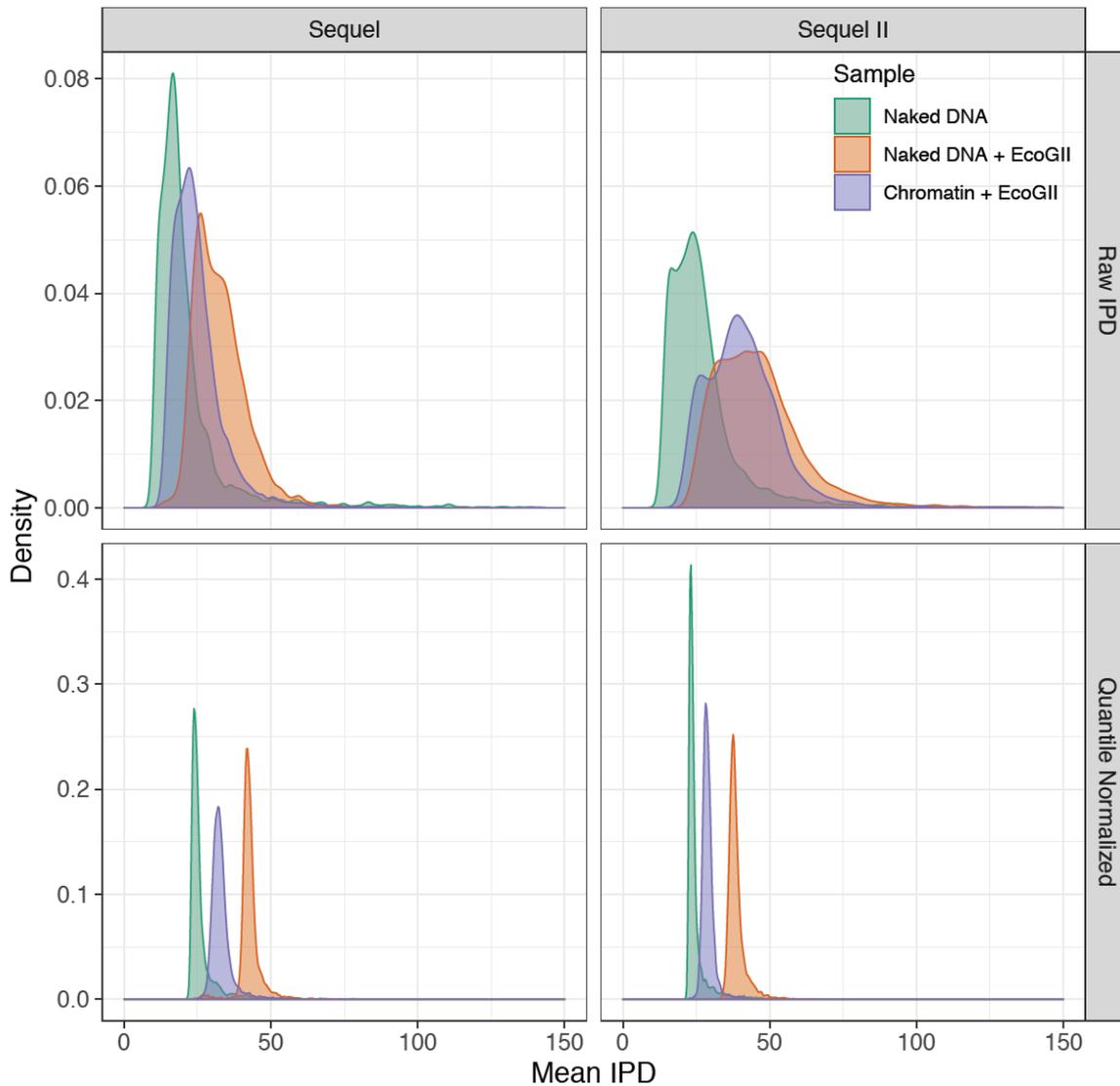
C.) Heatmap of enrichment test results to determine nucleosome conformers that are enriched or depleted for each chromatin state. Tests qualitatively appearing to be chromatin-state specific are highlighted with a black box. Significant tests following multiple hypothesis correction marked with a black dot. Fisher's Exact Test was used for all comparisons.

2.11 Supplementary Figures



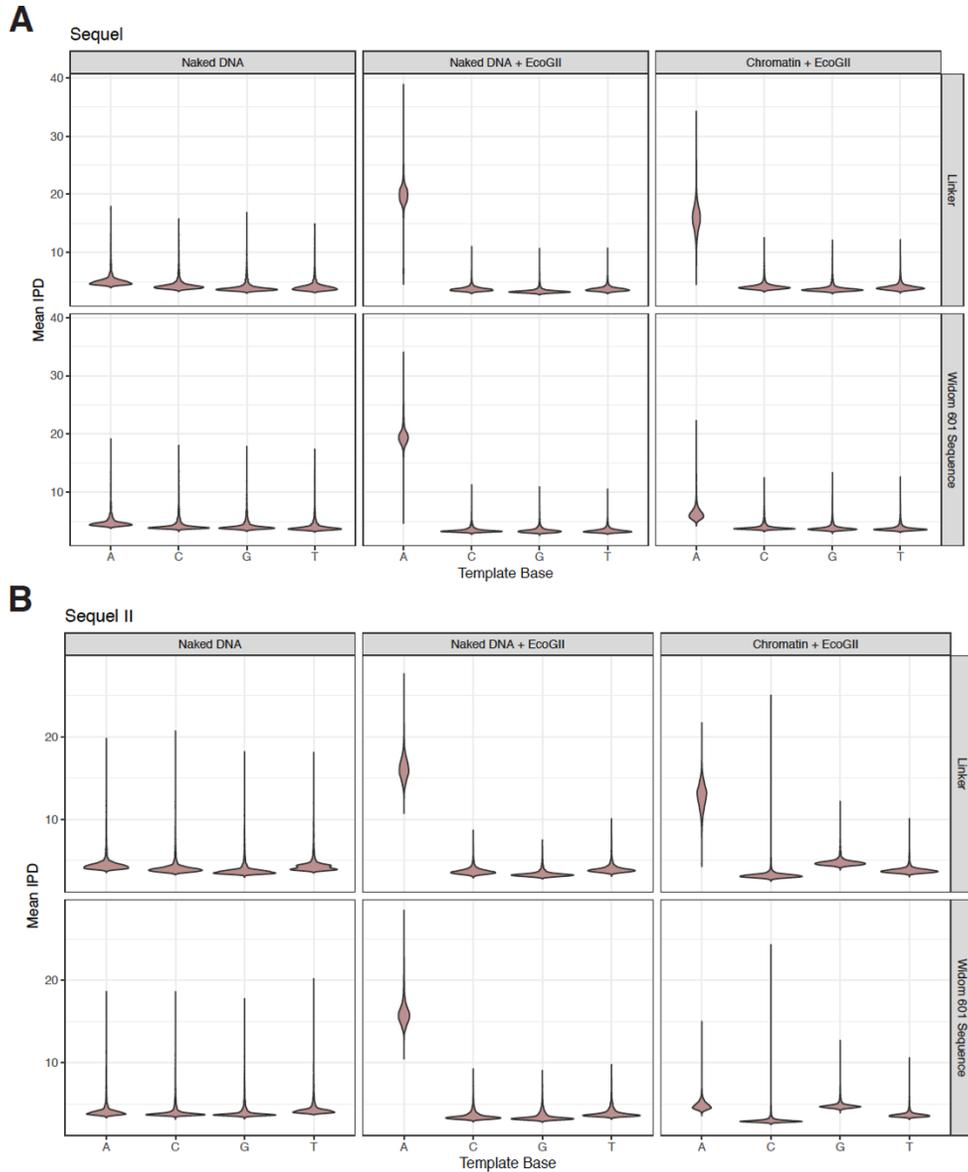
Supplementary Figure 2. 1: Quality control of in vitro nucleosome arrays assembled through salt-gradient dialysis.

Nonnucleosomal arrays were assembled in duplicate as previously and checked for assembly extent via restriction enzyme digest. In both cases, the smallest digestion products corresponded to a protected mononucleosomal fragment, suggesting that there is minimal underassembly of the resulting arrays.



Supplementary Figure 2. 2: Mean raw and quantile normalized interpulse durations for in vitro SAMOSA experiments.

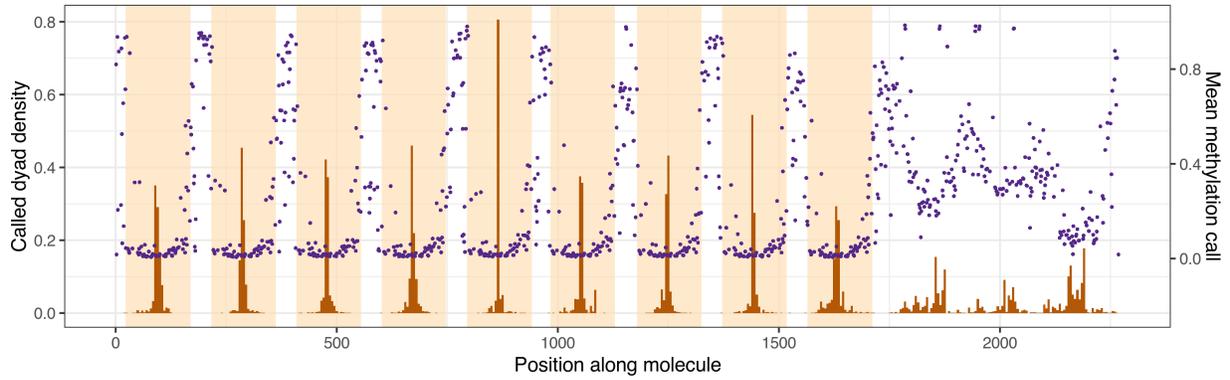
In vitro SAMOSA experiments demonstrate intermediate single-molecule average interpulse durations compared to unmethylated array DNA and fully methylated deproteinated array DNA. Data are similarly separated on the Sequel and Sequel II platforms and quantile normalization further aids in separating chromatin from control samples, particularly on the Sequel II platform.



Supplementary Figure 2. 3: Adenine methylation by the EcoGII enzyme is specific to accessible adenines and is protected against by the nucleosome.

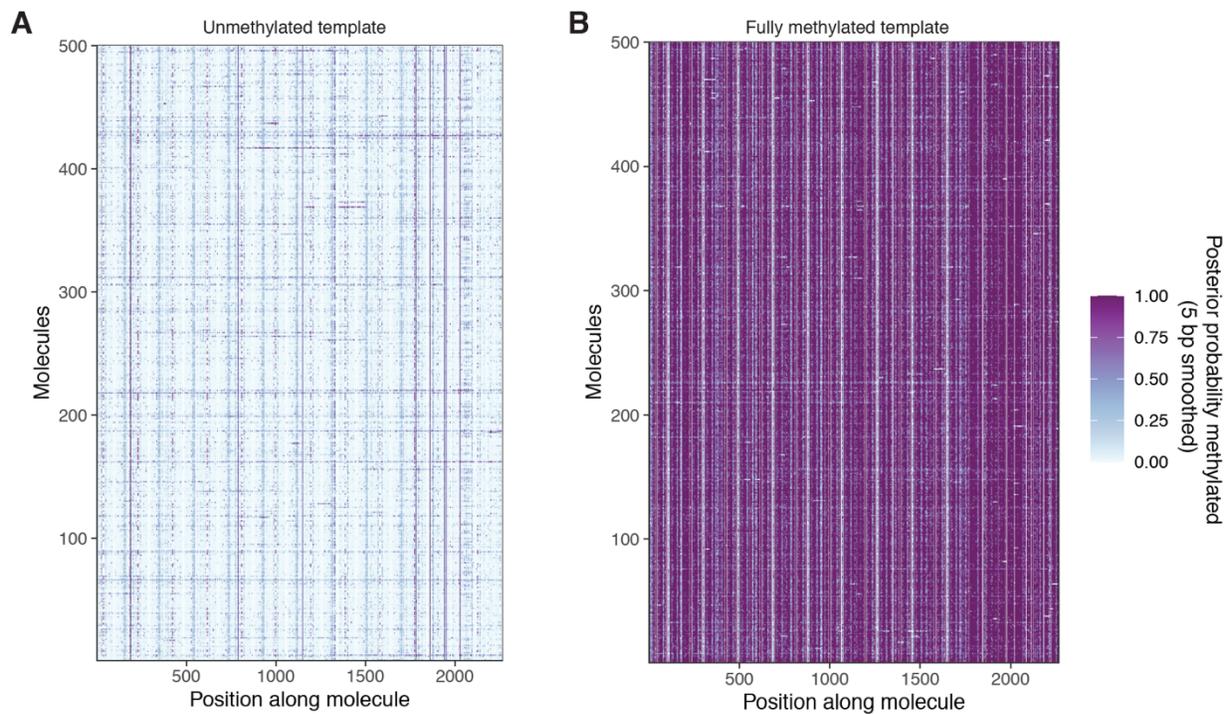
A.) Violin plots of the distribution of average, quantile normalized IPD values for each nucleotide on controls (unmethylated / fully methylated naked DNA) and chromatin separated by nucleotides falling within the Widom 601 nucleosome positioning sequence or linker DNA. In the chromatin context, only adenine nucleotides falling within the linker are modified, consistent with protection of bases by nucleosomes positioned by the Widom 601 sequence.

B.) As in A, but for data from the Sequel II platform.



Supplementary Figure 2. 4: Average linker methylation and individually called dyad positions are qualitatively similar across the length of the nonnucleosomal array molecule.

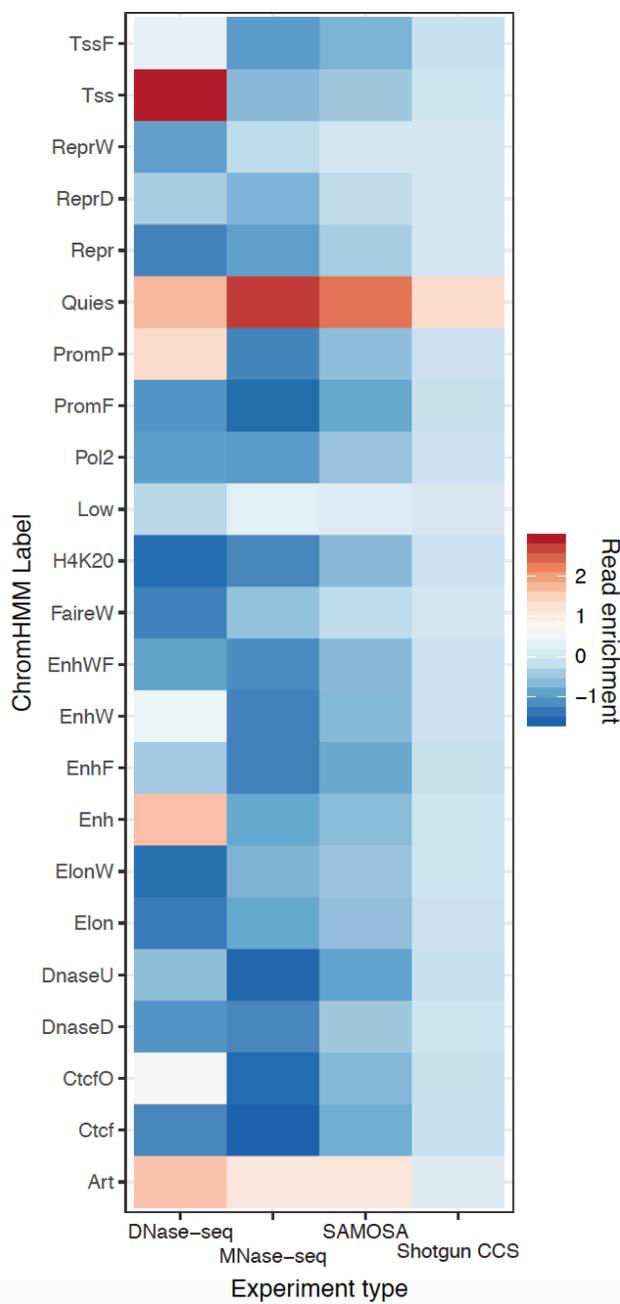
Histograms of called dyad positions for each occurrence of a Widom 601 repeat unit (orange shading), averaged over all sequenced chromatin molecules are shown in brown. Mean methylation calls for each linker sequence (sequence outside orange shading) are shown in purple.



Supplementary Figure 2. 5: Unmethylated and fully methylated array DNA does not display the same periodic patterning of modified bases seen in methylated chromatin.

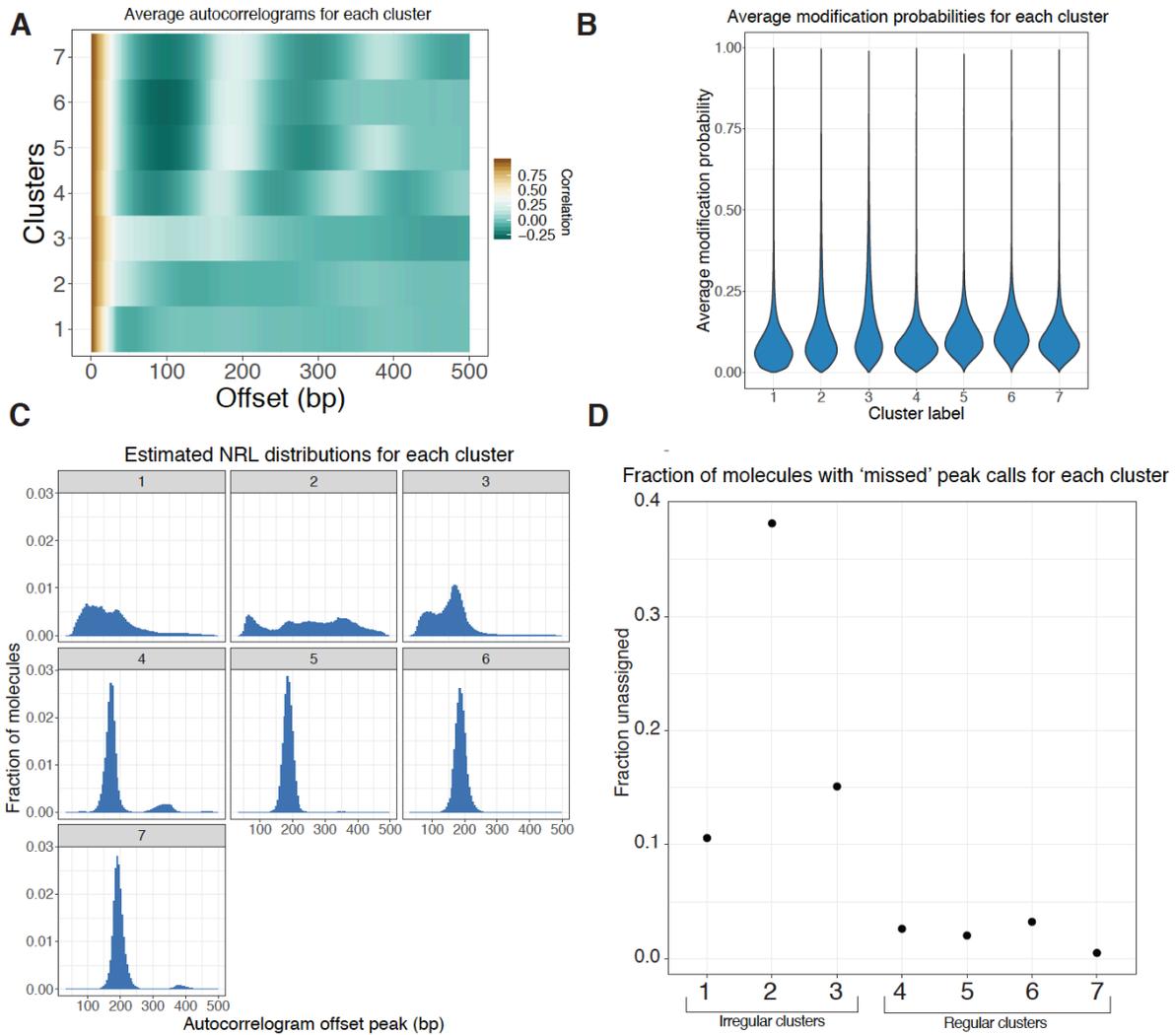
A.) Smoothed modification probabilities for 500 molecules of unmethylated array DNA.

B.) Smoothed modification probabilities for 500 molecules of fully methylated naked array DNA. In both cases, data are smoothed using a 5 bp rolling mean on the calculated modification probabilities at template A nucleotides.



Supplementary Figure 2. 6: Coverage enrichment of SAMOSA versus other data types at ChromHMM annotated genomic regions.

SAMOSA coverage is less biased for open / active chromatin than short-read assays of chromatin accessibility, and is more comparable to short-read MNase-seq assays. SAMOSA coverage is more biased than shotgun PacBio sequencing of the human genome, likely due to the combined use of MNase as a cleavage reagent and the chromatin solubilization protocol used.



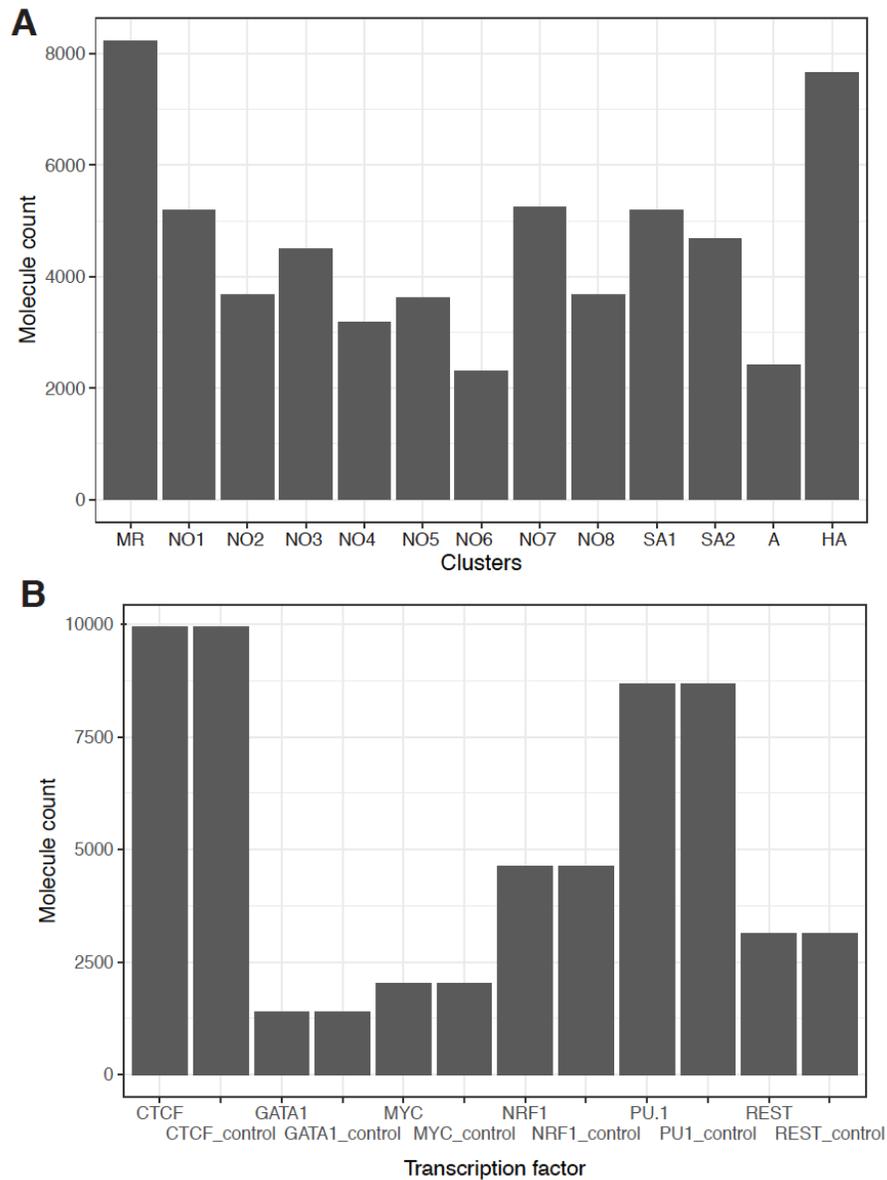
Supplementary Figure 2. 7: Further characterization of clustered footprinted molecules.

A.) Average autocorrelograms for the seven Leiden clusters.

B.) Violin plots of the single-molecule average modification probabilities for each cluster. Clusters do not substantially differ with respect to modification probability, suggesting that clustering is not simply driven by methylation extent.

C.) As in **Figure 3C**; NRL distribution estimates for each of the seven clusters.

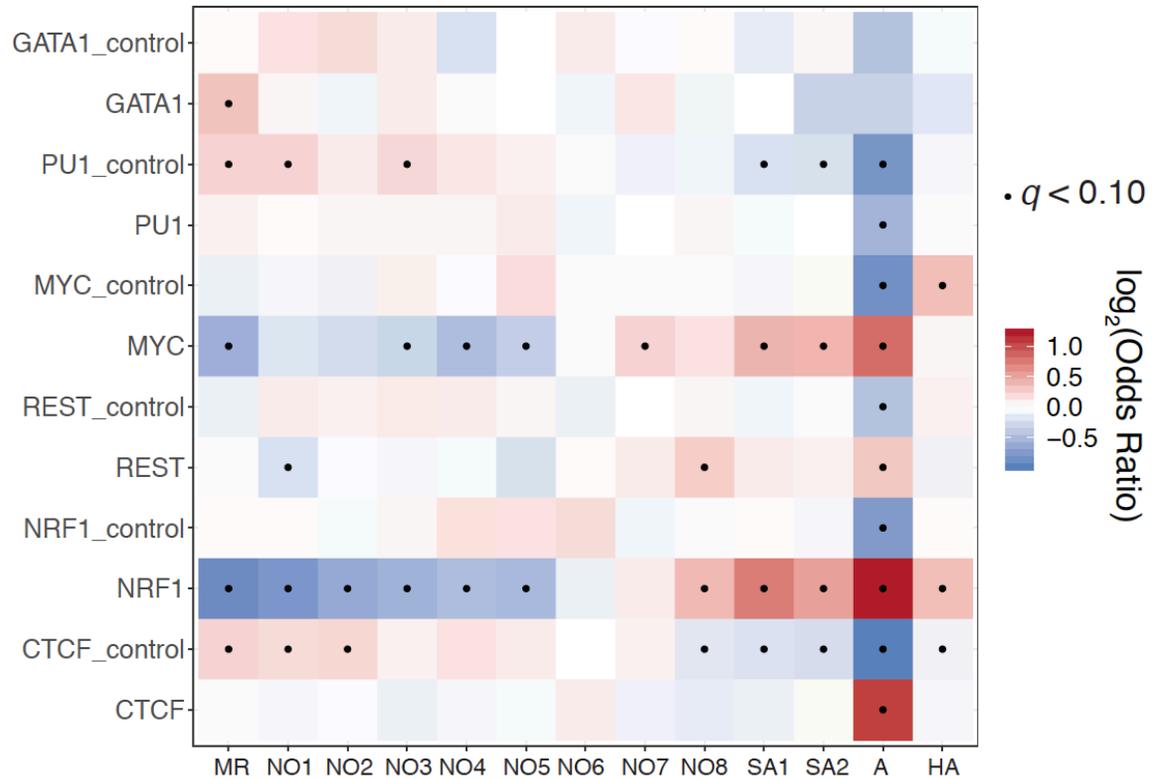
D.) Autocorrelogram peak-calling fails in a fraction of reads in each cluster; this fraction appears to be negatively associated with the 'regularity' of the cluster.



Supplementary Figure 2. 8: Cluster sizes and numbers of motif-containing molecules for each transcription factor chosen for study.

A.) Leiden cluster sizes for cluster shown in Figure 4.

B.) Counts of molecules harboring respective transcription factor binding sites. For each transcription factor, we sampled an equal number of randomly drawn molecules taken from regions GC- / repeat-content matched against TF ChIP-seq peaks.



Supplementary Figure 2. 9: Cluster enrichment for each transcription factor studied.

We performed Fisher's exact tests to determine relative enrichment and depletion of each cluster for each transcription factor surveyed in **Figure 4**. Cluster 'A' is consistently depleted across control molecules but enriched across molecules containing bona fide transcription factor binding motifs, suggesting that the clusters identified in this study are biologically relevant. Fishers Exact test odds ratios are plotted in heatmap form and all enrichment tests that are statistically significant under a false discovery rate of 10% ($q < 0.1$) are marked with a black dot.

2.12 Materials and Methods

Preparation of nonanucleosome arrays via salt-gradient dialysis

The nonanucleosome DNA in a plasmid was purified by Gigaprep (Qiagen) and the insert was digested out with EcoRV, ApaLI, XhoI and StuI. The insert was subsequently purified using a Sephacryl S1000 super fine gel filtration (GE Healthcare). Histones were purified and octamer was assembled as previously described (Luger et al., 1999). To assemble the arrays, the nonanucleosome DNA was mixed with octamer and supplementing dimer, then dialyzed from high salt to low salt (Lee and Narlikar, 2001). EcoRI sites engineered in the linker DNA between the nucleosomes, and digestion by EcoRI was used to assess the quality of nucleosome assembly.

SAMOSa on nonanucleosomal chromatin arrays

For the chromatin arrays, 1.5 mg of assembled array was utilized as input for methylation reactions with the non-specific adenine EcoGII methyltransferase (New England Biolabs, high concentration stock; 2.5E4U/mL). For the naked DNA controls, 2 mg of DNA was utilized as input for methylation reactions. Methylation reactions were performed in a 100 μ L reaction with Methylation Reaction buffer (1X CutSmart Buffer, 1 mM S-adenosyl-methionine (SAM, New England Biolabs)) and incubated with 2.5 μ L EcoGII at 37°C for 30 min. SAM was replenished to 6.25 mM after 15 min. Unmethylated controls were similarly supplemented with Methylation Reaction buffer, minus EcoGII and replenishing SAM, and the following purification conditions. To purify DNA, the samples were all subsequently incubated with 10 μ L Proteinase K (20 mg/mL) and 10 μ L 10% SDS at 65°C for a minimum of 2 hr up to overnight. To extract the DNA, equal parts volume of Phenol-Chloroform was added and mixed vigorously by shaking, spun (max speed, 2 min). The aqueous portion was carefully removed and 0.1x volumes of 3M NaOAc, 3 μ L of GlycoBlue and 3x volumes of 100% EtOH were added, mixed gently by inversion, and incubated overnight

at -20 °C. Samples were then spun (max speed, 4 °C, 30 min), washed with 500 mL 70% EtOH, air dried and resuspended in 50 µL EB. Sample concentration was measured by Qubit High Sensitivity DNA Assay.

Preparation of in vitro SAMOSA SMRT libraries

The purified DNA from nonanucleosome array and DNA samples were used in entirety as input for PacBio SMRTbell library preparation (~1.5–2 mg). Preparation of libraries included DNA damage repair, end repair, SMRTbell ligation, and Exonuclease according to manufacturer's instruction. After Exonuclease Cleanup and a double 0.8x Ampure PB Cleanup, sample concentration was measured by Qubit High Sensitivity DNA Assay (1 mL each). To assess for library quality, samples (1 mL each) were run on an Agilent Bioanalyzer DNA chip. Libraries were sequenced on either Sequel I or Sequel II flow cells (UC Berkeley QB3 Genomics). Sequel II runs were performed using v2.0 sequencing chemistry and 30 hr movies.

Cell lines and cell culture

K562 cells (ATCC) were grown in standard media containing RPMI 1640 (Gibco) supplemented with 10% Fetal Bovine Serum (Gemini, Lot#A98G00K) and 1% Penicillin-Streptomycin (Gibco). Cell lines were regularly tested for mycoplasma contamination and confirmed negative with PCR (NEB Neb-Next Q5 High Fidelity 2X Master Mix).

Isolation of nuclei, MNase digest, and overnight dialysis

100E6 K562 cells were collected by centrifugation (300xg, 5 min), washed in ice cold 1X PBS, and resuspended in 1 mL Nuclear Isolation Buffer (20 mM HEPES, 10 mM KCl, 1 mM MgCl₂, 0.1% Triton X-100, 20% Glycerol, and 1X Protease Inhibitor (Roche)) per 5–10 e6 cells by gently pipetting 5x with a wide-bore tip to release nuclei. The suspension was incubated on ice for 5 min, and nuclei were pelleted (600xg, 4 °C, 5 min), washed with Buffer M (15 mM Tris-HCl pH 8.0, 15 mM NaCl, 60 mM KCl, 0.5 mM

Spermidine), and spun once again. Nuclei were resuspended in 37°C pre-warmed Buffer M supplemented with 1 mM CaCl₂ and distributed into two 1 mL aliquots. For digestion, micrococcal nuclease from *Staphylococcus aureus* (Sigma, reconstituted in ddH₂O, stock at 0.2 U/uL) was added at 1U per 50E6 nuclei, and nuclei were digested for 1 min. at 37°C. EGTA was added to 2 mM immediately after 1 min to stop the digestion and incubated on ice. For nuclear lysis and liberation of chromatin fibers, MNase-digested nuclei were collected (600xg, 4°C, 5 min) and resuspended in 1 mL per 50E6 nuclei of Tep20 Buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EGTA, 20 mM NaCl, and 1X Protease Inhibitor (Roche) added immediately before use) supplemented with 300 mg/ mL of Lysolethicin (L-a-Lysophosphatidylcholine from bovine brain, Sigma, stock at 5 mg/mL) and incubated at 4°C overnight. To remove nuclear debris the next day, dialyzed samples were spun (12,000xg, 4°C, 5 min) and the soluble chromatin fibers present in the supernatant were collected. Sample concentration was measured by Nanodrop. SAMOSA experiments with variable digestion conditions were performed as above, except temperature (37°C vs. 4°C) and time (1 min vs. 10 min vs. 60 min) were varied, starting cell counts were increased to 200E6 for prepared nuclei for varied condition experiments, and gTube spins were omitted.

SAMOSA on K562-derived oligonucleosomes

Dialyzed chromatin was utilized as input (1.5 mg) for methylation reactions with the non-specific adenine EcoGII methyltransferase (New England Biolabs, high concentration stock 2.5e4U/mL). Reactions were performed in a 200 mL reaction with 1X CutSmart Buffer and 1 mM S-adenosyl- methionine (SAM, New England Biolabs) and incubated with 2.5 mL enzyme at 37°C for 30 min. SAM was replenished to 6.25 mM after 15 min. Non-methylation controls were similarly supplemented with Methylation Reaction buffer, minus EcoGII and replenishing SAM, and purified by the following conditions. To purify all DNA samples, reactions were incubated with 10 mL of RNaseA at room temperature for 10 min, followed by 20 uL Proteinase K (20 mg/mL) and 20 uL 10% SDS at 65°C for a minimum of 2 hr up to overnight. To extract the DNA, equal parts volume of Phenol-Chloroform was added and mixed vigorously by shaking, spun

(max speed, 2 min). The aqueous portion was carefully removed and 0.1x volumes of 3M NaOAc, 3 mL of GlycoBlue and 3x volumes of 100% EtOH were added, mixed gently by inversion, and incubated overnight at -20 °C. Samples were then spun (max speed, 4 °C, 30 min), washed with 500 mL 70% EtOH, air dried and resuspended in 50 mL EB. Sample concentration was measured by Qubit High Sensitivity DNA Assay. Naked DNA Positive methylation controls were collected from aforementioned non-methylated controls post-purification (25 mL, ~500 ng), methylated with EcoGII as previously stated, and purified again by the following conditions.

Preparation of in vivo SAMOSA SMRT libraries

Purified DNA from MNase-digested K562 chromatin oligonucleosomes (methylated, non-methylated control, purified then methylated) were briefly spun in a Covaris G-Tube (3380xg, 1 min) in efforts to shear gDNA uniformly to 10 kB prior PacBio library preparation. The input concentration was approximately 575 ng for methylated and non-methylated samples, and approximately 320 ng for purified then methylated samples. Samples were concentrated with 0.45x of AMPure PB beads according to manufacturer's instructions. The entire sample volume was utilized as input for subsequent steps in library preparation, which included DNA damage repair, end repair, SMRTbell ligation, and Exonuclease cleanup according to manufacturer's instructions. For SMRTbell ligations, unique PacBio SMRT-bell adaptors (100 mM stock) were annealed to a 20 mM working stock in 10 mM Tris-HCl pH 7.5 and 100 mM NaCl in a thermocycler (85 °C 5 min, RT 30 s, 4 °C hold) and stored at -20 °C for long-term storage. After exonuclease cleanup and double Ampure PB cleanups (0.45X), the sample concentrations were measured by Qubit High Sensitivity DNA Assay (1 mL each). To assess for size distribution and library quality, samples (1 uL each) were run on an Agilent Bioanalyzer DNA chip. Libraries were sequenced on Sequel II flow cells (UC Berkeley QB3 Genomics Core). In vivo data were collected over three 30 hr Sequel II movie runs; the first with a 2 hr pre-extension time and the second two with a 0.7 hr pre-extension time.

Data analysis

All raw data will be made available at GEO Accession GSE162410; processed data is available at Zenodo (<https://doi.org/10.5281/zenodo.3834705>). All scripts and notebooks for reproducing analyses in the paper are available at <https://github.com/RamaniLab/SAMOSA> (Abdulhay, 2020; copy archived at [swh:1:rev:208027064183d042adede691b935cad9e79106a3](https://swh.io/rev/208027064183d042adede691b935cad9e79106a3)).

We apply our method to two use cases in the paper, and they differ in the computational workflow to analyze them. The first is for sequencing samples where every DNA molecule should have the same sequence, which is the case for our in vitro validation experiments presented in Figure 1. The second use case is for samples from cells containing varied sequences of DNA molecules. We will refer to the first as homogeneous samples, and the second as genomic samples. The workflow for genomic samples will be presented first in each section, and the deviations for homogeneous samples detailed at the end.

500U hia5 K562 Fiber-seq data from Stergachis et al., 2020 were downloaded using Google Cloud Services via SRA accession SRP252718 and processed as below.

Sequencing read processing

Sequencing reads were processed using software from Pacific Biosciences. The following describes the workflow for genomic samples:

Demultiplex reads

Reads were demultiplexed using lima. The flag ‘-same’ was passed as libraries were generated with the same barcode on both ends. This produces a BAM file for the subreads of each sample.

Generate circular consensus sequences (CCS)

CCS were generated for each sample using `ccs` (Travers et al., 2010). Default parameters were used other than setting the number of threads with `-j`. This produces a BAM file of CCS.

Align CCS to the reference genome

Alignment was done using `pbmm2` (Li, 2016), and run on each CCS file, resulting in BAM files containing the CCS and alignment information.

Generate missing indices

Our analysis code requires PacBio index files (`.pbi`) for each BAM file. `pbmm2` does not generate index files, so missing indices were generated using `pbindex`.

For homogeneous samples, replace step three with this alternate step 3.

Align subreads to the reference genome

`pbmm2` was run on each subreads BAM file (the output of step 1) to align subreads to the reference sequence, producing a BAM file of aligned subreads.

Sample reference preparation

Our script for analyzing samples relies on a CSV file input that contains information about each sample, including the locations of the relevant BAM files and a path to the reference genome. The CSV needs a header with the following columns: `index`: Integer indices for each sample. We write the table using `pandas` `.to_csv` function, with parameters `index = True, index_label='index'` `cell`: A unique name for the SMRT cell on which the sample was sequenced `sampleName`: The name of the sample

unalignedSubreadsFile: This will be the file produced by step one above. This should be an absolute path to the file.

ccsFile

This is the file produced by step two above alignedSubreadsFile: This is the file produced by the alternate step three above. It is required for homogeneous samples but can be left blank for genomic samples.

alignedCcsFile

This is the file produced by step three above. It is required for genomic samples but can be left blank for homogeneous samples.

Reference

The file of the reference genome or reference sequence for the sample.

Extracting IPD measurements and calling methylation

The script extractIPD.py accesses the BAM files, reads the IPD values at each base and uses a gaussian mixture model to generate posterior probabilities of each adenine being methylated. extractIPD takes two positional arguments. The first is a path to the above sample reference CSV file. The second is a specification for which sample to run on. This can be either an integer index value, in which case extractIPD will run on the corresponding row. Alternatively, it can be a string containing the cell and sampleName, separated by a period. Either way extractIPD will run on the specified sample using the paths to the BAM files contained within the CSV.

extractIPD produces the following three output files when run on genomic samples: processed/onlyT/{cell}_{sampleName}_onlyT_zmwinfo.pickle: This file is a 'pandas' dataframe stored as a pickle,

and can be read with the ‘pandas.read_pickle’ function. This dataframe contains various information about each individual ZMW.

Processed/onlyT/{cell}_{sampleName}_onlyT.pickle

This file contains the normalized IPD value at every thymine. The data is stored as a dictionary object. The keys are the ZMW hole numbers (stored in the column ‘zwm’ in the zmwinfo dataframe), and the values are numpy arrays. The arrays are 1D with length equal to the length of the CCS for that molecule. At bases that are A/T, there will be a normalized IPD value. Each G/C base and a few A/T bases for which an IPD value couldn’t be measured will contain NaN.

Processed/binarized/{cell}_{sampleName}_bingmm.pickle

This file contains the posterior probability of each adenine being methylated. The data format is identical to the _onlyT.pickle file above, except the numpy array contains values between 0 and 1, where the higher values indicate a higher confidence that the adenine is methylated.

When run on homogeneous samples the following output files are alternately produced: processed/onlyT/{cell}_{sampleName}_onlyT.npy: This numpy array has a column for every base in the reference sequence, and a row for each DNA molecule that passes the filtering threshold. A normalized IPD value is stored for each adenine that could be measured at A/T bases, other bases are NaN.

Processed/binarized/{cell}_{sampleName}_bingmm.npy

This numpy array is the same shape as the _onlyT.npy file above. The values are posterior probabilities for an adenine being methylated, ranging from 0 to 1.

Dyad calling on in vitro methylated chromatin arrays

Nucleosome positions were predicted in nonnucleosomal array data by taking a 133 bp wide rolling mean across the molecule and finding each local minimum peak at least 147 bp apart from each other.

k-mer analyses of negative and positive control experiments

To investigate the role of sequence context in our methylation calls, we examined the distribution of normalized IPD values for our in vitro negative and positive controls. We binned the adenines by sequence context using two base pairs on the 5' side of the template base and five base pairs on the 3' side. These bases were previously found to have the strongest influence on IPD value (Feng et al., 2013). We combined both replicates for negative and positive controls and plotted a heatmap where each row is a sequence context, and the color intensity is the histogram counts of molecules with a normalized IPD value in that bin. Negative control, positive control, and both combined were each plotted. K-mer contexts were sorted by their mean normalized IPD in the combined set. The sequence contexts were separately plotted.

In vivo analyses

We smooth the posterior probabilities calculated in the paper to account for regions with low local A/T content and generally denoise the single-molecule signal. For *in vitro* analyses, we smooth the calculated posterior probabilities using a 5 bp rolling mean. For all *in vivo* analyses in the paper that involve calculation of single-molecule autocorrelograms, averaging over multiple templates, and visualizing individual molecules, we smooth posteriors with a 33 bp rolling mean. For all autocorrelation calculations we ignore regions where compared lengths would be unequal; this has the effect of rendering the returned autocorrelogram exactly $0.5 * \text{the input length}$.

Averages of the modification signal across the first 1 kb of K562 oligonucleosomes

We took all molecules at least 500 nt in length and concatenated all of the resulting matrices from each of

the four separate samples/runs, and then plotted the NaN-sensitive mean over the matrix as a function of distance along the molecule.

Clustering analysis of all chromatin molecules ≥ 500 bp in length

We used Leiden clustering cluster all molecules in our dataset passing our lower length cutoff. Resolution and `n_neighbors` were manually adjusted to avoid generating large numbers of very small clusters (i.e. <100 molecules). All parameters used for plotting figures in the paper are recapitulated in the Jupyter notebook. Our clustering strategy was as follows: first, we smoothed raw signal matrices with a 33 bp NaN-sensitive running mean. We next computed the autocorrelation function for each molecule in the matrix, using the full length of the molecule up to 1000 bp. We then used Scanpy (Wolf et al., 2018) to perform Leiden clustering on the resulting matrix. We visualized the resulting cluster averages with respect to the average autocorrelation function, and with respect to averaged modification probabilities for each cluster. For a subset of clusters we also randomly sampled 500– 5000 molecules to directly visualize in the paper.

Computing single-molecule autocorrelograms and estimating NRLs on single molecules

We computed single-molecule autocorrelograms and discovered peaks on these autocorrelograms as follows: for each molecule, we used the `scipy` (Virtanen et al., 2020) `find_peaks` function to in the computed autocorrelogram and annotated the location of that peak. We also kept track of the molecules where `find_peaks` could not detect a peak using the given parameters, which we optimized manually by modifying peak height/width to detect peaks on the averaged autocorrelograms. In our hands, these parameters robustly detect peaks between 180 and 190 bp in auto-correlogram averages, consistent with the expected bulk NRL in K562 cells (analyses by A Rendeiro; [zenodo.org/ record/3820875](https://zenodo.org/record/3820875)). For each collection of single-molecule autocorrelogram peaks we computed the median, the median absolute deviation, and visualized the distribution of peak locations as a histogram.

TF-binding motif analyses and enrichment tests

K562 TF-binding sites were predicted as in Ramani et al., 2019. Briefly, we downloaded IDR-filtered ENCODE ChIP-seq peaks for CTCF, NRF1, REST, c-MYC, PU.1, and GATA1, and then used FIMO (Bailey et al., 2009) to predict TF binding sites within these peaks using CISTRROME PWM definitions for each transcription factor. For MNase-cleavage analyses, we plotted the abundance of MNase cuts (two per molecule) with respect to TF binding sites and plotted these as number of cleavages per molecules sequenced. To examine modification probabilities around TF-binding sites, we wrote a custom script (`zmw_selector.py`) to find the ZMWs that overlap with features of interest (*e.g.* transcription factor binding sites). We extracted all ZMWs where a portion of the read alignment falls within 1 kb of a given feature, and annotated the position of the alignment starts, ends, and strand with respect to the feature. We then used these coordinates and strand information to extract all modification signal falling within a 500 bp window centered at each TF binding site. For control sites, we used the `gkmSVM` package (Ghandi et al., 2016) to find GC-/repeat content matched genomic regions for each peakset. We constructed a series of enrichment tests (Fisher's Exact) to determine odds ratios/p values to find specific cluster label–transcription factor pairs that were enriched with respect to the total set of all labeled molecules. Finally, we used the Storey q-value package (Storey and Tibshirani, 2003) to correct for the number of Fisher's exact tests performed.

Enrichment tests for chromatin states

We used a custom python script (`zmw_selector_bed.py`) or directly scanned for satellite-containing CCS reads (see below) to extract molecules that fall within ENCODE-defined chromatin states/pertain to human major satellite sequences. We then used a Python dictionary linking ZMW IDs to indices along the total matrix of molecules to link Cluster IDs and chromatin states. Finally, we constructed a series of enrichment tests (Fisher's Exact) to determine odds ratios/p values to find specific cluster label-chromatin state pairs that were enriched with respect to the total set of all labeled molecules. We then used the Storey q-value

package to correct for the number of Fisher's exact tests performed. Control molecules were drawn as above, using the gkmSVM package to find GC/repeat content matched genomic regions for each peakset.

Selection of satellite-containing reads

Circular consensus reads with minimum length of 1 kb bearing satellites were identified using BLAST searching against a database containing DFAM (Hubley et al., 2016) consensus sequences for alpha (DF0000014.4, DF0000015.4, DF0000029.4), beta (DF0000075.4, DF0000076.4, DF0000077.4, DF0000078.4, DF0000079.4), and gamma (DF0000148.4, DF0000150.4, DF0000152.4) satellites using blastn with default parameters. Satellite containing reads were further filtered such that they contained at minimum two hits to satellite consensus sequences and matches spanned at least 50% of the consensus sequence. These labels were then used to separate out sequences for the analyses presented in Figure 6—figure supplement 3.

Data Availability

All raw data are available at GEO Accession GSE162410; processed data is available at Zenodo (<https://doi.org/10.5281/zenodo.3834705>). All scripts and notebooks for reproducing analyses in the paper are available at <https://github.com/RamaniLab/SAMOSA> (copy archived at <https://archive.softwareheritage.org/swh:1:rev:208027064183d042adede691b935cad9e79106a3/>).

2.13 Acknowledgments and Additional Information

The authors thank Daniele Canzio (UCSF), Hiten Madhani (UCSF), Srinivas Ramachandran (CU Denver), and Christopher Weber (Stanford) for helpful discussions and comments on the manuscript. The authors thank Shana McDevitt, Robert Munch, and the UC Berkeley Vincent J Coates Genomics Sequencing Laboratory for assisting with PacBio sequencing.

Competing interests

Geeta J Narlikar: Reviewing editor, eLife. Jason G Underwood: JGU is an employee of Pacific Biosciences, Inc and holds stock in this company. The other authors declare that no competing interests exist.

REFERENCES

Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020a). Massively multiplex single-molecule oligonucleosome footprinting. *Elife* 9.

Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020b). Massively multiplex single-molecule oligonucleosome footprinting. *Elife* 9, 1–23.

Ahmad, K., and Henikoff, S. (2001). Modulation of a Transcription Factor Counteracts Heterochromatic Gene Silencing in *Drosophila*. *Cell* 104, 839–847.

Ahmad, K., and Henikoff, S. (2002). The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Mol. Cell* 9, 1191–1200.

Alfert, A., Moreno, N., and Kerl, K. (2019). The BAF complex in development and disease. *Epigenetics Chromatin* 2019 121 12, 1–15.

Altemose, N., Logsdon, G.A., Bzikadze, A. V., Sidhwani, P., Langley, S.A., Caldas, G. V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., et al. (2022). Complete genomic and epigenetic maps of human centromeres. *Science* (80-.). 376.

Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., and Roberts, C.W.M. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* 2017 81 8, 1–10.

Angelov, D., Vitolo, J.M., Mutskov, V., Dimitrov, S., and Hayes, J.J. (2001). Preferential interaction of the core histone tail domains with linker DNA. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6599.

Arents, G., and Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11170.

Arimura, Y., Tachiwana, H., Oda, T., Sato, M., and Kurumizaka, H. (2012). Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry* *51*, 3302–3309.

Armache, J.P., Gamarra, N., Johnson, S.L., Leonard, J.D., Wu, S., Narlikar, G.J., and Cheng, Y. (2019). Cryo-EM structures of remodeler-nucleosome intermediates suggest allosteric control through the nucleosome. *Elife* *8*.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* (80-.). *339*, 1074–1077.

Badenhorst, P., Voas, M., Rebay, I., and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.* *16*, 3186–3198.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* *37*, W202–W208.

Baldi, S., Krebs, S., Blum, H., and Becker, P.B. (2018). Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. *Nat. Struct. Mol. Biol.* *2018 259 25*, 894–901.

Baldi, S., Korber, P., and Becker, P.B. (2020). Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. *Nat. Struct. Mol. Biol.* *2020 272 27*, 109–118.

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

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019a). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* *569*, 136–140.

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019b). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* *569*, 136–140.

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

Barutcu, A.R., Lajoie, B.R., Fritz, A.J., McCord, R.P., Nickerson, J.A., Van Wijnen, A.J., Lian, J.B., Stein, J.L., Dekker, J., Stein, G.S., et al. (2016). SMARCA4 regulates gene expression and higher-order chromatin structure in proliferating mammary epithelial cells. *Genome Res.* 26, 1188–1201.

Becker, P.B., Gloss, B., Schmid, W., Strähle, U., and Schütz, G. (1986). In vivo protein–DNA interactions in a glucocorticoid response element require the presence of the hormone. *Nat.* 1986 3246098 324, 686–688.

Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., Davies, H., et al. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat. Genet.* 45, 1479–1482.

Blosser, T.R., Yang, J.G., Stone, M.D., Narlikar, G.J., and Zhuang, X. (2009). Dynamics of nucleosome remodelling by individual ACF complexes. *Nature* 462, 1022–1027.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422.

Bracken, A.P., Brien, G.L., and Verrijzer, C.P. (2019). Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev.* 33, 936–959.

Brahma, S., and Henikoff, S. (2019). RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. *Mol. Cell* 73, 238-249.e3.

Brahma, S., and Henikoff, S. (2020). Epigenome Regulation by Dynamic Nucleosome Unwrapping. *Trends Biochem. Sci.* 45, 13–26.

Cai, S., Song, Y., Chen, C., Shi, J., and Gan, L. (2018). Natural chromatin is heterogeneous and self-

associates in vitro. *Mol. Biol. Cell* 29, 1652–1663.

Carey, M., Li, B., and Workman, J.L. (2006). RSC Exploits Histone Acetylation to Abrogate the Nucleosomal Block to RNA Polymerase II Elongation. *Mol. Cell* 24, 481–487.

Castel, D., Philippe, C., Calmon, R., Le Dret, L., Truffaux, N., Boddaert, N., Pagès, M., Taylor, K.R., Saulnier, P., Lacroix, L., et al. (2015). Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. *Acta Neuropathol.* 130, 815–827.

Chan, K.M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., David James, C., Jenkins, R., et al. (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev.* 27, 985–990.

Chioda, M., Vengadasalam, S., Kremmer, E., Eberharter, A., and Becker, P.B. (2010). Developmental role for ACF1-containing nucleosome remodellers in chromatin organisation. *Development* 137, 3513–3522.

Clapier, C.R., and Cairns, B.R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* 492, 280–284.

Clapier, C.R., Längst, G., Corona, D.F. V., Becker, P.B., and Nightingale, K.P. (2001). Critical Role for the Histone H4 N Terminus in Nucleosome Remodeling by ISWI. *Mol. Cell. Biol.* 21, 875–883.

Collins, N., Poot, R.A., Kukimoto, I., García-Jiménez, C., Delleire, G., and Varga-Weisz, P.D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat. Genet.* 32, 627–632.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat. Rev. Cancer* 16, 803–810.

Correll, S.J., Schubert, M.H., and Grigoryev, S.A. (2012). Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J.* 31, 2416–2426.

Cuaycong, M.H., and Elgin, S.C.R. (2001). Long-Range Nucleosome Ordering Is Associated with Gene Silencing in. *Mol. Cell. Biol.* 21, 2867–2879.

Cutter, A.R., and Hayes, J.J. (2015). A Brief Review of Nucleosome Structure. *FEBS Lett.* 589, 2914.

Dann, G.P., Liszczak, G.P., Bagert, J.D., Müller, M.M., Nguyen, U.T.T., Wojcik, F., Brown, Z.Z., Bos, J., Panchenko, T., Pihl, R., et al. (2017). ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature* 548, 607–611.

Davidson, I.F., and Peters, J.M. (2021). Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* 2021 227 22, 445–464.

Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150, 12–27.

Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladde, M.P., and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. *Mol. Cell* 38, 590–602.

Deindl, S., Hwang, W.L., Hota, S.K., Blosser, T.R., Prasad, P., Bartholomew, B., and Zhuang, X. (2013). ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell* 152, 442–452.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26, 5433–5438.

De Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Dargham, D.B., Jounier, S., Humbertclaude, H., Ribierre, F., Baulard, C., et al. (2016). Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* 530, 113–116.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet.* 34, 666–681.

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

Donehower, L.A., Huang, A.L., and Hager, G.L. (1981). Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. *J. Virol.* 37, 226–238.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* (80-.). 306, 1571–1573.

Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* 186, 773–790.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21, 129–228.

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

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

Ebbert, M.T.W., Farrugia, S.L., Sens, J.P., Jansen-West, K., Gendron, T.F., Prudencio, M., McLaughlin, I.J., Bowman, B., Seetin, M., Dejesus-Hernandez, M., et al. (2018). Long-read sequencing across the C9orf72 “GGGGCC” repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Mol. Neurodegener.* 13, 1–17.

Eberharter, A., Ferrari, S., Längst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., and Becker, P.B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO J.* 20, 3781–3788.

Ehrensberger, A.H., Franchini, D.M., East, P., George, R., Matthews, N., Maslen, S.L., and Svejstrup, J.Q. (2015). Retention of the Native Epigenome in Purified Mammalian Chromatin. *PLoS One* *10*, e0133246.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* *323*, 133–138.

Erdel, F., and Rippe, K. (2011). Chromatin remodelling in mammalian cells by ISWI-type complexes—where, when and why? *Authors J. Compil.* *a* *278*, 3608–3618.

Eustermann, S., Schall, K., Kostrewa, Di., Lakomek, K., Strauss, M., Moldt, M., and Hopfner, K.P. (2018). Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature* *556*, 386–390.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* *30*, 1232–1239.

Feng, Z., Fang, G., Korlach, J., Clark, T., Luong, K., Zhang, X., Wong, W., and Schadt, E. (2013). Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic. *PLOS Comput. Biol.* *9*, e1002935.

Ferraro, A. (2016). Altered primary chromatin structures and their implications in cancer development. *Cell. Oncol.* *39*, 195–210.

Filion, G.J., van Bommel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell* *143*, 212–224.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* *73*, 1897–1901.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. *Nature* *269*, 29–36.

Flaus, A. (2012). Principles and practice of nucleosome positioning in vitro.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010a). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010b). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 2010 76 7, 461–465.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol. Cell* 8, 545–556.

Fyodorov, D. V., Blower, M.D., Karpen, G.H., and Kadonaga, J.T. (2004). Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes Dev.* 18, 170–183.

Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of Nucleosome Positioning in the Human Genome. *PLOS Genet.* 8, e1003036.

Gamarra, N., Johnson, S.L., Trnka, M.J., Burlingame, A.L., and Narlikar, G.J. (2018). The nucleosomal acidic patch relieves auto-inhibition by the ISWI remodeler SNF2h. *Elife* 7.

Garcia, J.F., Dumesic, P.A., Hartley, P.D., El-Samad, H., and Madhani, H.D. (2010). Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. *Genes Dev.* 24, 1758–1771.

Gershman, A., Sauria, M.E.G., Guitart, X., Vollger, M.R., Hook, P.W., Hoyt, S.J., Jain, M., Shumate, A., Razaghi, R., Koren, S., et al. (2022). Epigenetic patterns in a complete human genome. *Science* (80-.). 376.

Ghandi, M., Mohammad-Noori, M., Ghareghani, N., Lee, D., Garraway, L., and Beer, M.A. (2016).

gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* 32, 2205–2207.

Gibson, B.A., Doolittle, L.K., Schneider, M.W.G., Jensen, L.E., Gamarra, N., Henry, L., Gerlich, D.W., Redding, S., and Rosen, M.K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* 179, 470–484.e21.

Gilbert, N., and Allan, J. (2001). Distinctive higher-order chromatin structure at mammalian centromeres. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11949–11954.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin Architecture of the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. *Cell* 118, 555–566.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* 140, 678–691.

Golemis, E.A., Scheet, P., Beck, T.N., Scolnick, E.M., Hunter, D.J., Hawk, E., and Hopkins, N. (2018). Molecular mechanisms of the preventable causes of cancer in the United States. *Genes Dev.* 32, 868–902.

Gonzalez-Perez, A., Jene-Sanz, A., and Lopez-Bigas, N. (2013). The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* 14, 1.

Grüne, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F.V., Becker, P.B., and Müller, C.W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol. Cell* 12, 449–460.

Hamiche, A., Sandaltzopoulos, R., Gdula, D.A., and Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* 97, 833–842.

Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. *Cell* 100, 57–70.

Harshman, S.W., Young, N.L., Parthun, M.R., and Freitas, M.A. (2013). H1 histones: current perspectives and challenges. *Nucleic Acids Res.* *41*, 9593.

Harutyunyan, A.S., Krug, B., Chen, H., Papillon-Cavanagh, S., Zeinieh, M., De Jay, N., Deshmukh, S., Chen, C.C.L., Belle, J., Mikael, L.G., et al. (2019). H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nat. Commun.* *10*.

Hayden, E.C. (2014). The \$ 1,000 genome. *Nature* *507*, 294–295.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* *21*, 133–153.

Henikoff, S., and Smith, M.M. (2015). Histone Variants and Epigenetics. *Cold Spring Harb. Perspect. Biol.* *7*.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 18318–18323.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* *2009 6* *6*, 283–289.

Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* *52*, 504–510.

Hoose Scott A. and Kladde, M.P. (2006). DNA Methyltransferase Probing of DNA-Protein Interactions. In *Gene Mapping, Discovery, and Expression: Methods and Protocols*, M. Bina, ed. (Totowa, NJ: Humana Press), pp. 225–244.

Hota, S.K., Bhardwaj, S.K., Deindl, S., Lin, Y.C., Zhuang, X., and Bartholomew, B. (2013). Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat. Struct. Mol. Biol.* *20*, 222–229.

Hoyt, S.J., Storer, J.M., Hartley, G.A., Grady, P.G.S., Gershman, A., Lima, L.G. de, Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., et al. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* (80-.). 376.

Huang, B., Babcock, H., and Zhuang, X. (2010). Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *Cell* 143, 1047–1058.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., and Wheeler, T.J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Res.* 44, D81–D89.

Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* 90, 145–155.

Ito, T., Levenstein, M.E., Fyodorov, D. V., Kutach, A.K., Kobayashi, R., and Kadonaga, J.T. (1999). ACF consists of two subunits, Acfl and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.* 13, 1529–1539.

Iwasaki, W., Miya, Y., Horikoshi, N., Osakabe, A., Taguchi, H., Tachiwana, H., Shibata, T., Kagawa, W., and Kurumizaka, H. (2013). Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* 3, 363.

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

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nat.* 2015 5277578 527, 384–388.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* 1.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013).

Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* *45*, 592–601.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M.G., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* *49*, 213–222.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* *458*, 362–366.

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

Kenzaki, H., and Takada, S. (2015). Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLOS Comput. Biol.* *11*, e1004443.

Khurana, S., and Oberdoerffer, P. (2015). Replication Stress: A Lifetime of Epigenetic Change. *Genes* *2015*, Vol. 6, Pages 858-877 *6*, 858–877.

Kim, J.H., Ebersole, T., Kouprina, N., Noskov, V.N., Ohzeki, J.I., Masumoto, H., Mravinac, B., Sullivan, B.A., Pavlicek, A., Dovat, S., et al. (2009). Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res.* *19*, 533–544.

Kim, J.M., Visanpattanasin, P., Jou, V., Liu, S., Tang, X., Zheng, Q., Li, K.Y., Snedeker, J., Lavis, L.D., Lionnet, T., et al. (2021). Single-molecule imaging of chromatin remodelers reveals role of atpase in promoting fast kinetics of target search and dissociation from chromatin. *Elife* *10*.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome Remodeling Induced by RNA Polymerase II: Loss of the H2A/H2B Dimer during

Transcription. *Mol. Cell* 9, 541–552.

Klein, D.C., and Hainer, S.J. (2020). Genomic methods in profiling DNA accessibility and factor localization. *Chromosom. Res.* 28, 69.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220.

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

Kornberg, R.D., and Stryer, L. (1988). Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* 16, 6677–6690.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., and Burger, L. (2017a). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol. Cell* 67, 411–422.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017b). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol. Cell* 67, 411-422.e4.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* 167, 709-721.e12.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* 18, 548.

Lai, B., Gao, W., Cui, K., Xie, W., Tang, Q., Jin, W., Hu, G., Ni, B., and Zhao, K. (2018). Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nat.* 562, 281–285.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409,

860–921.

Längst, G., and Becker, P.B. (2001). Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci.* *114*, 2561–2568.

Lee, K.M., and Narlikar, G. (2001). Assembly of Nucleosomal Templates by Salt Dialysis. *Curr. Protoc. Mol. Biol.* *54*, 21.6.1-21.6.16.

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Sadowski, N., Simpson, J.T., Sedlazeck, F.J., and Timp, W. (2019). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *BioRxiv* 504993.

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Gershman, A., Sadowski, N., Sedlazeck, F.J., Hansen, K.D., Simpson, J.T., and Timp, W. (2020). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* *2020* 1712 *17*, 1191–1199.

Leonard, J.D., and Narlikar, G.J. (2015). A Nucleotide-Driven Switch Regulates Flanking DNA Length Sensing by a Dimeric Chromatin Remodeler. *Mol. Cell* *57*, 850–859.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* (80-.). *340*, 857–861.

Li, B., Carey, M., and Workman, J.L. (2007a). The Role of Chromatin during Transcription. *Cell* *128*, 707–719.

Li, B., Carey, M., and Workman, J.L. (2007b). The role of chromatin during transcription. *Cell* *128*, 707–719.

Li, Y., Gong, H., Wang, P., Zhu, Y., Peng, H., Cui, Y., Li, H., Liu, J., and Wang, Z. (2021). The emerging role of ISWI chromatin remodeling complexes in cancer. *J. Exp. Clin. Cancer Res.* *2021* 401 *40*, 1–27.

Li, Z., Zhao, P., and Xia, Q. (2019). Epigenetic Methylations on N6-Adenine and N6-Adenosine with the same Input but Different Output. *Int. J. Mol. Sci.* *20*.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015a). Nucleosome Spacing Generated by ISWI and CHD1 Remodelers Is Constant Regardless of Nucleosome Density. *Mol. Cell. Biol.* *35*, 1588–1605.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015b). Nucleosome Spacing Generated by ISWI and CHD1 Remodelers Is Constant Regardless of Nucleosome Density. *Mol. Cell. Biol.* *35*, 1588–1605.

Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., and Song, C.X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* *37*, 424–429.

Long, M., Sun, X., Shi, W., Yanru, A., Leung, S.T.C., Ding, D., Cheema, M.S., MacPherson, N., Nelson, C.J., Ausio, J., et al. (2019). A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res.* *47*, 8399–8409.

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

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

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nat.* *1997* 3896648 *389*, 251–260.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* *304*, 3–19.

Maeshima, K., Hihara, S., and Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo?

Curr. Opin. Cell Biol. 22, 291–297.

Mashtalir, N., Dao, H.T., Sankar, A., Liu, H., Corin, A.J., Bagert, J.D., Ge, E.J., D’Avino, A.R., Filipovski, M., Michel, B.C., et al. (2021). Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes. *Science* (80-). 373, 306–315.

McBride, M.J., Mashtalir, N., Winter, E.B., Dao, H.T., Filipovski, M., D’Avino, A.R., Seo, H.S., Umbreit, N.T., St. Pierre, R., Valencia, A.M., et al. (2020). The nucleosome acidic patch and H2A ubiquitination underlie mSWI/SNF recruitment in synovial sarcoma. *Nat. Struct. Mol. Biol.* 2020 279 27, 836–845.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 22534–22539.

Mivelaz, M., Cao, A.M., Kubik, S., Zencir, S., Hovius, R., Boichenko, I., Stachowicz, A.M., Kurat, C.F., Shore, D., and Fierz, B. (2020). Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. *Mol. Cell* 77, 488-500.e9.

Mohammad, F., and Helin, K. (2017). Oncohistones: Drivers of pediatric cancers. *Genes Dev.* 31, 2313–2324.

Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D.P., Højfeldt, J.W., Comet, I., Zheng, C., Johansen, J.V., Rapin, N., Porse, B.T., et al. (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.* 23, 483–492.

Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., Shiota, H., Debernardi, A., Héry, P., Curtet, S., et al. (2013). Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev.* 27, 1680–1692.

Murray, I.A., Morgan, R.D., Luyten, Y., Fomenkov, A., Corrêa, I.R., Dai, N., Allaw, M.B., Zhang, X.,

Cheng, X., and Roberts, R.J. (2018a). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* *46*, 840–848.

Murray, I.A., Morgan, R.D., Luyten, Y., Fomenkov, A., Corrêa, I.R., Dai, N., Allaw, M.B., Zhang, X., Cheng, X., and Roberts, R.J. (2018b). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* *46*, 840–848.

Mutskov, V., Gerber, D., Angelov, D., Ausio, J., Workman, J., and Dimitrov, S. (1998). Persistent Interactions of Core Histone Tails with Nucleosomal DNA following Acetylation and Transcription Factor Binding. *Mol. Cell. Biol.* *18*, 6293.

Nabils, N.H., Deleyrolle, L.P., Darst, R.P., Riva, A., Reynolds, B.A., and Kladde, M.P. (2014). Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res.* *24*, 329–339.

Nacev, B.A., Feng, L., Bagert, J.D., Lemiesz, A.E., Gao, J.J., Soshnev, A.A., Kundra, R., Schultz, N., Muir, T.W., and Allis, C.D. (2019). The expanding landscape of ‘oncohistone’ mutations in human cancers. *Nature* *567*, 473–478.

Nakane, T., Kimanius, D., Lindahl, E., and Scheres, S.H.W. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* *7*.

Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* *154*, 490–503.

Oberbeckmann, E., Wolff, M., Krietenstein, N., Heron, M., Ellins, J.L., Schmid, A., Krebs, S., Blum, H., Gerland, U., and Korber, P. (2019). Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* *29*, 1996–2009.

Oberbeckmann, E., Niebauer, V., Watanabe, S., Farnung, L., Moldt, M., Schmid, A., Cramer, P., Peterson, C.L., Eustermann, S., Hopfner, K.P., et al. (2021). Ruler elements in chromatin remodelers set nucleosome

array spacing and phasing. *Nat. Commun.* 2021 121 12, 1–17.

Olins, A.L., and Olins, D.E. (1974). Spheroid Chromatin Units (v Bodies). *Science* (80-). 183, 330–332.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O’Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* (80-). 357.

Panditharatna, E., and Filbin, M.G. (2020). The growing role of epigenetics in childhood cancers. *Curr. Opin. Pediatr.* 32, 67–75.

Papamichos-Chronakis, M., and Peterson, C.L. (2012). Chromatin and the genome integrity network. *Nat. Rev. Genet.* 2013 141 14, 62–75.

Partensky, P.D., and Narlikar, G.J. (2009). Chromatin Remodelers Act Globally, Sequence Positions Nucleosomes Locally. *J. Mol. Biol.* 391, 12–25.

Patel, A.B., Moore, C.M., Greber, B.J., Luo, J., Zukin, S., Ranish, J., and Nogales, E. (2019). Architecture of the chromatin remodeler RSC and insights into its nucleosome engagement. *Elife* 8.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* 25, 154.

Polach, K.J., and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J. Mol. Biol.* 254, 130–149.

Pott, S. (2017). Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *Elife* 6.

Racki, L.R., and Narlikar, G.J. (2008). ATP-dependent Chromatin Remodeling Enzymes: Two Heads are not Better, Just Different. *Curr Opin Genet Dev* 18, 137–144.

Racki, L.R., Yang, J.G., Naber, N., Partensky, P.D., Acevedo, A., Purcell, T.J., Cooke, R., Cheng, Y., and Narlikar, G.J. (2009). The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nat.*

2009 4627276 462, 1016–1021.

Ramani, V., Qiu, R., and Shendure, J. (2019). High Sensitivity Profiling of Chromatin Structure by MNase-SSP. *Cell Rep.* 26, 2465-2476.e4.

Rao, S., Ahmad, K., and Ramachandran, S. (2021). Cooperative Binding between Distant Transcription Factors is a Hallmark of Active Enhancers. *Mol. Cell* 81, 255-267.e6.

Rhodes, D., Brown, R.S., and Klug, A. (1989). Crystallization of nucleosome core particles. *Methods Enzymol.* 170, 420–428.

Richard-Foy, H., and Hager, G.L. (1987). Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. *EMBO J.* 6, 2321–2328.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. *Nature* 311, 532–537.

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

Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta* 1839, 627.

Sanulli, S., Trnka, M.J., Dharmarajan, V., Tibble, R.W., Pascal, B.D., Burlingame, A.L., Griffin, P.R., Gross, J.D., and Narlikar, G.J. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nat.* 2019 5757782 575, 390–394.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436, 138–141.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T.W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A.K., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin

remodelling genes in paediatric glioblastoma. *Nature* 482, 226–231.

Scott, W.A., and Campos, E.I. (2020). Interactions With Histone H3 & Tools to Study Them. *Front. Cell Dev. Biol.* 8.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* 19, 65–71.

Serra-Cardona, A., and Zhang, Z. (2018). Replication-Coupled Nucleosome Assembly in the Passage of Epigenetic Information and Cell Identity. *Trends Biochem. Sci.* 43, 136–148.

Shema, E., Bernstein, B.E., and Buenrostro, J.D. (2018). Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat. Genet.* 2018 511 51, 19–25.

Shipony, Z., Marinov, G.K., Swaffer, M.P., Sinnott-Armstrong, N.A., Skotheim, J.M., Kundaje, A., and Greenleaf, W.J. (2020). Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* 2020 173 17, 319–327.

Simpson, R.T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* 17, 5524–5531.

Simpson, R.T., and Stafford, D.W. (1983). Structural features of a phased nucleosome core particle. *Proc. Natl. Acad. Sci. U. S. A.* 80, 51.

Smith, S., and Stillman, B. (1991). Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J.* 10, 971–980.

Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., and Shendure, J. (2016). Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* 164, 57–68.

Sobecki, M., Souaid, C., Boulay, J., Guerineau, V., Noordermeer, D., and Crabbe, L. (2018). MadID, a Versatile Approach to Map Protein-DNA Interactions, Highlights Telomere-Nuclear Envelope Contact

Sites in Human Cells. *Cell Rep.* 25, 2891-2903.e5.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* (80-). 344, 376–380.

Sönmez, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., and Krebs, A.R. (2021). Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. *Mol. Cell* 81, 255-267.e6.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 2012 139 13, 613–626.

Steensel BV., and Henikoff S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* 18, 424–428.

Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* (80-). 368.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* 2020 224 22, 361–371.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J. Biol. Chem.* 281, 16279–16288.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–9445.

Suter, B., Schnappauf, G., and Thoma, F. (2000). Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.* 28, 4083.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected

communities. *Sci. Reports* 2019 9, 1–12.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res.* 25, 1491–1498.

Tse, O.Y.O., Jiang, P., Cheng, S.H., Peng, W., Shang, H., Wong, J., Chan, S.L., Poon, L.C.Y., Leung, T.Y., Chan, K.C.A., et al. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 118.

Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 83, 1021–1026.

Tullius, T.D. (1988). DNA footprinting with hydroxyl radical. *Nat.* 1988 3326165 332, 663–664.

Udugama, M., Sabri, A., and Bartholomew, B. (2011). The INO80 ATP-Dependent Chromatin Remodeling Complex Is a Nucleosome Spacing Factor. *Mol. Cell. Biol.* 31, 662.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–522.

Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nat.* 1997 3886642 388, 598–602.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 2015 163 16, 178–189.

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. (2020). Global reference mapping of human transcription factor footprints. *Nat.* 2020 5837818 583, 729–736.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific

computing in Python. *Nat. Methods* 2020 173 17, 261–272.

Wang, Y., Wang, A., Liu, Z., Thurman, A.L., Powers, L.S., Zou, M., Zhao, Y., Hefe, A., Li, Y., Zabner, J., et al. (2019). Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* 29, 1329–1342.

Wanior, M., Krämer, A., Knapp, S., and Joerger, A.C. (2021). Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy. *Oncogene* 2021 4021 40, 3637–3654.

Weber, C.M., and Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. *Genes Dev.* 28, 672.

Weber, C.M., Hafner, A., Kirkland, J.G., Braun, S.M.G., Stanton, B.Z., Boettiger, A.N., and Crabtree, G.R. (2021). mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol.* 2021 286 28, 501–511.

Weintraub, H., and Groudine, M. (1976). Chromosomal Subunits in Active Genes Have an Altered Conformation. *Science* (80-.). 193, 848–856.

Wiechens, N., Singh, V., Gkikopoulos, T., Schofield, P., Rocha, S., and Owen-Hughes, T. (2016). The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLOS Genet.* 12, e1005940.

Wilson, B.G., Wang, X., Shen, X., McKenna, E.S., Lemieux, M.E., Cho, Y.J., Koellhoffer, E.C., Pomeroy, S.L., Orkin, S.H., and Roberts, C.W.M. (2010). Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* 18, 316–328.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 1–5.

Workman, J.L. (2006). Nucleosome displacement in transcription. *Genes Dev.* 20, 2009–2017.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nat.* 2016 5327599 532, 329–333.

Xiao, C. Le, Zhu, S., He, M., Chen, D., Zhang, Q., Chen, Y., Yu, G., Liu, J., Xie, S.Q., Luo, F., et al. (2018). N 6 -Methyladenine DNA Modification in the Human Genome. *Mol. Cell* 71, 306-318.e7.

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., and Wu, C. (2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* 8, 531–543.

Xu, M., Long, C., Chen, X., Huang, C., Chen, S., and Zhu, B. (2010). Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. *Science* (80-.). 328, 94–98.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. *Mol. Cell* 2, 851–861.

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

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

Yang, Z., Zheng, C., and Hayes, J.J. (2007). The core histone tail domains contribute to sequence-dependent nucleosome positioning. *J. Biol. Chem.* 282, 7930–7938.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* 149, 1461–1473.

- Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* *515*, 355–364.
- Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* *25*, 2227–2241.
- Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* *37*, 76–81.
- Zentner, G.E., and Henikoff, S. (2014). High-resolution digital profiling of the epigenome. *Nat. Rev. Genet.* *15*, 814–827.
- Zhang, Y., and Li, Y. (2011). The Expanding Mi-2/NuRD Complexes: A Schematic Glance: <https://doi.org/10.4137/PRI.S63293>, 79–109.
- Zhang, W., Feng, J., and Li, Q. (2020). The replisome guides nucleosome assembly during DNA replication. *Cell Biosci.* *2020* *101* *10*, 1–14.
- Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* (80-.). *332*, 977–980.
- Zhou, C.Y., Johnson, S.L., Gamarra, N.I., and Narlikar, G.J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors. *Annu. Rev. Biophys.* *45*, 153–181.
- Zhou, C.Y., Johnson, S.L., Lee, L.J., Longhurst, A.D., Beckwith, S.L., Johnson, M.J., Morrison, A.J., and Narlikar, G.J. (2018). The yeast INO80 complex operates as a tunable DNA length-sensitive switch to regulate nucleosome sliding. *Mol. Cell* *69*, 677.
- Zhou, V.W., Goren, A., and Bernstein, B.E. (2010). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* *2011* *121* *12*, 7–18.

Zofall, M., Persinger, J., and Bartholomew, B. (2004). Functional Role of Extranucleosomal DNA and the Entry Site of the Nucleosome in Chromatin Remodeling by ISW2. *Mol. Cell. Biol.* 24, 10047–10057.

**Chapter 3: Single-fiber nucleosome density dictates
the regulatory mode of a chromatin remodeler**

3.1 Abstract

Nearly all essential nuclear processes act on DNA packaged into series of nucleosomes termed chromatin fibers. However, our understanding of how these processes (e.g. DNA replication (Stewart-Morgan et al., 2020), RNA transcription (Li et al., 2007b), chromatin extrusion (Davidson and Peters, 2021), nucleosome remodeling (Narlikar et al., 2013)) actually occur on such fibers remains unresolved. To address this deficit, we present SAMOSA-ChAAT, a massively multiplex single-molecule footprinting platform to map the primary structure of individual, precisely reconstituted chromatin templates subjected to virtually any chromatin-associated reaction. As proof-of-concept, we apply SAMOSA-ChAAT to study ATP-dependent chromatin remodeling by the essential imitation switch (ISWI) ATPase SNF2h, whose mechanism-of-action remains contentious. Using our approach, we discover that SNF2h operates as a density-dependent, length-sensing nucleosome sliding enzyme, whose ability to decrease or increase DNA accessibility depends on single-fiber nucleosome density. We validate our in vitro findings with single-fiber accessibility measurements in vivo, finding that the regulatory ‘mode’ of SNF2h-containing complexes (i.e. ‘opening’ vs. ‘closing’ chromatin) is dictated by the underlying nucleosome-density of individual chromatin fibers: at canonically-defined heterochromatin, SNF2h generates evenly-spaced nucleosome arrays of multiple nucleosome repeat lengths; at SNF2h-dependent accessible sites, the enzyme slides nucleosomes to increase accessibility of motifs for the essential transcription factor CTCF. Our approach and data demonstrate, for the first time, how chromatin remodelers can effectively sense nucleosome density to induce diametrically-opposed regulatory effects within the nucleus. More generally, our novel approach promises molecularly-precise views of any of the essential processes shaping nuclear physiology.

3.2 Bulk average sequencing techniques are unable to measure single chromatin reaction states

Nucleosomes regulate DNA transactions essential to life. Nuclear regulatory factors, such as sequence-specific transcription factors (TFs), polymerases, DNA repair machinery, extrusive condensin and cohesin complexes, and ATP-dependent chromatin remodeling complexes (i.e. ‘chromatin remodelers’), all must navigate long stretches of nucleosomes (i.e. ‘chromatin fibers’) to enable cell-type-specific gene regulation. However, studying how such regulatory factors act on individual chromatin fibers has been challenging, as methods capable of resolving such interactions are fundamentally lacking. Existing biochemical approaches for studying chromatin in bulk (e.g. FRET; gel remodeling) (Yang et al., 2006a), or at single-molecule resolution (e.g. smFRET (Deindl et al., 2013); cryo-EM (Armache et al., 2019)), provide high-resolution views of mononucleosomes, but are generally incapable of capturing the state of individual fibers. Classical footprinting-based approaches for studying chromatin interactions are powerful, but rely on bulk averaging of nucleolytic products over many templates (Becker et al., 1986; Hewish and Burgoyne, 1973; Richard-Foy and Hager, 1987; Tullius, 1988). Averaging such signal is problematic, as both nucleosome positions, and the average nucleosome spacing on individual chromatin fibers, can vary substantially across a population of even identical DNA templates (Baldi et al., 2020). To address this clear methodological gap, we developed a generalizable, high-throughput platform to footprint and follow how chromatin regulatory factors remodel nucleosome positions on reconstituted chromatin fibers. We term this platform SAMOSA-ChAAT, a novel approach that combines our previously published single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA) with precise, *in vitro* chromatin reconstitution on mammalian genomic sequences (Chromatin Accessibility of Assembled Templates).

3.3 Overview: Applying SAMOSA to study single-molecule reactions of ISWI remodeler reactions

Here we apply SAMOSA-ChAAT to study chromatin remodeling by the essential imitation switch (ISWI) ATPase SNF2h and address a major unsolved area of SNF2h mechanism (Tsukiyama et al., 1995). Chromatin remodelers are essential macromolecular complexes that harness the energy of ATP to slide,

evict, load, or transfer core histones (Narlikar et al., 2013). Mammalian ISWI complexes catalyze nucleosome sliding via the ATPase motors sucrose nonfermenting 2 homologue / homologue-like (SNF2h / SNF2I), to facilitate DNA replication, repair, transcriptional activation, and repression (Erdel and Rippe, 2011). A key activity of ISWI complexes is to organize nucleosomes into equally-spaced arrays in the context of heterochromatin. Yet how ISWI complexes equalize spacing remains debated: some studies have proposed a ‘clamping’ or ‘ruler’ model of nucleosome density-independent ISWI remodeling, while others suggest a ‘length-sensing’ model of nucleosome density-dependent ISWI remodeling. Importantly, studies performed to-date harbor unique limitations that confound resolution of either model. Bulk and single-molecule studies, for instance, have been performed in the context of mononucleosomes (Blosser et al., 2009; Racki et al., 2009; Yang et al., 2006a) while *in vitro* activity measurements on chromatin fibers have relied on bulk nuclease digestion (Krietenstein et al., 2016; Lieleg et al., 2015b; Oberbeckmann et al., 2021; Zhang et al., 2011). Delineation between these models is further complicated by the facts that i.) equally-spaced nucleosome arrays can randomly emerge downstream of a barrier without invoking nucleosome remodeling (i.e. ‘statistical’ positioning) (Kornberg and Stryer, 1988), and ii.) primary sequence can influence initial nucleosome positions (Lowary and Widom, 1998b). Distinguishing between these models, and ultimately understanding how ISWI can both repress chromatin accessibility, and enable TF binding for factors like CTCF (Barisic et al., 2019b; Wiechens et al., 2016) thus requires single-fiber-resolved measurements.

Using our approach, we rule out the clamping model for SNF2h. We find that instead SNF2h activity is consistent with the length-sensing model, as the average SNF2h-catalyzed nucleosome spacing on single fibers decreases as underlying nucleosome density increases. We explore the physiological significance of this model using binding motifs for the TF CTCF, demonstrating that the initial nucleosome-location and density ‘states’ of a fiber ultimately dictate the fold-change in relative accessibility of sites following remodeling. Finally, to understand the implications of this length-sensing model *in vivo*, we footprint individual chromatin fibers in living murine embryonic stem cells (mESCs) devoid of SNF2h.

Our results demonstrate that at heterochromatic sites known to be targeted by ISWI complexes, SNF2h acts to generate a population of evenly-spaced nucleosome arrays with short, but variant, nucleosome repeat lengths; conversely, at promoters and CTCF binding sites, ISWI complexes increase DNA accessibility by sliding nucleosomes to create irregular fibers. Taken as a whole, our study offers a new paradigm for single-fiber chromatin remodeling, wherein nucleosome sliding in the context of varying nucleosome density programs DNA accessibility outcomes.

3.4 Single-molecule footprinting of intact chromatin fibers reconstituted on genomic sequences

Prior biochemical studies on chromatin fibers have used arrays composed of a nucleosome positioning sequence such as Widom 601 (Dechassa et al., 2010; Mivelaz et al., 2020), or fibers reconstituted from yeast genomic DNA (Kaplan et al., 2009; Krietenstein et al., 2016; Zhang et al., 2011). We previously demonstrated that our SAMOSA protocol could accurately resolve single-fiber nucleosome footprints on 601-based chromatin fibers (Abdulhay et al., 2020b). To enable study of more native-like fibers, we extended the SAMOSA approach to footprint individual fibers reconstituted on mammalian genomic sequences through salt gradient dialysis (SGD). To this end, we devised a general workflow we term SAMOSA-ChAAT (**Chromatin Accessibility of Assembled Templates; Figure 3.1A**), in which chromatin fibers with desired biochemical properties (e.g. nucleosome density) are assembled from genomic templates, subjected to the SAMOSA m⁶dA footprinting protocol, and sequenced on the PacBio Sequel II to natively detect m⁶dA modifications reflective of accessible DNA bases.

As proof-of-concept for this approach, we cloned two ~3 kilobase sequences from the *M. musculus* genome (hereafter, sequences ‘S1’ and ‘S2’), carried out the SAMOSA-ChAAT workflow across four specified histone octamer:DNA ratios, and sequenced resulting molecules and controls to high depth (**samples and sequencing depths summarized in Supplementary Table 1**). We then analyzed the resulting sequencing data using a custom pipeline that combines a neural network with a hidden markov model (NN-HMM; Methods) to detect stretches of unmethylated DNA directly from PacBio data, while

accounting for primary sequence biases for the EcoGII m6dAase and PacBio sequencing polymerase (**Supplementary Figure 3.1**). Consistent with the assembly of histones into nucleosome core particles with varying degrees of ‘breathability,’ we were able to call stretches of unmethylated DNA on sequenced molecules (i.e. ‘footprints’) ranging from ~120 – 160 nucleotides (nts) in size (**Figure 3.1B**), in addition to short (< 30 nt) footprints suggestive of nonspecific histone-DNA interactions (e.g. H2A/H2B-DNA). Footprint sizes increased along with chromatin density, suggesting i.) that higher nucleosome densities inhibit DNA breathing along assembled nucleosome core particles (footprint sizes for S1 predicted mononucleosomes: 130 ± 10.4 bp to 141 ± 19.3 bp for 5:1 to 20:1 fibers; data reported as median \pm median absolute deviation), and ii.) that higher nucleosome densities promote formation of closely spaced di- and tri-nucleosome structures. Our data also enable estimates of the number of nucleosomes per individual template (i.e. ‘nucleosome density’); accordingly, inferred nucleosome counts on single molecules closely matched targeted assembly extents (**Figure 3.1C**; **values reported as mean \pm standard deviation in Supplementary Table 3.2**).

Nucleosome assembly can be influenced by the underlying shape and rigidity of template DNA, which varies strongly as a function of DNA sequence (Rohs et al., 2009). To ascertain patterns of favored nucleosome positioning in bulk, we generated footprint length versus footprint midpoint ‘horizon plots’ (analogous to fragment length vs midpoint ‘V-plots’ (Henikoff et al., 2011)) for each assembly condition and sequence (**Figure 3.1D**). Our approach allows for explicit mapping and classification of footprints of all sizes as a function of target sequence, clearly revealing both sequence-directed nucleosome positioning, and regions that favor formation of closely-packed primary structures (e.g. dinucleosomes with virtually no intervening linker DNA).

To move beyond these bulk averages, we next explored our data at single-molecule resolution (**Figure 3.1E,F**) using UMAP dimensionality reduction (McInnes et al., 2018) and Leiden community detection (Traag et al., 2019). We found i.) that UMAP projections natively capture the biochemical parameter of chromatin density (**Figure 3.1E**), and ii.) that unbiased clustering enables detection of

mutually exclusive nucleosome positions for molecules from identical SGD preparations (**see purple and green clusters in Figure 3.1F**). Importantly, our data satisfy a wide set of controls. First, our footprint-size analyses, nucleosome-density measurements, horizon plot visualizations, UMAP reductions, and cluster profiles were all consistent for the completely different sequence S2 (**Supplementary Figure 3.2**). Second, our analytical pipeline accurately detected expected footprint sizes and positions from Widom 601 chromatin fibers with known dyad positions (**Supplementary Figure 3.33A-B**), albeit with a longer mononucleosome footprint size consistent with less DNA breathing on Widom 601 nucleosomes (Lowary and Widom, 1998b). Finally, our nucleosome occupancy measurements were highly quantitatively reproducible across replicates (**Supplementary Figure 3.3C-F**). Together, these data demonstrate the sensitivity, reproducibility, and generalizability of the SAMOSA-ChAAT approach.

3.5 SAMOSA-ChAAT captures single-fiber chromatin remodeling reaction outcomes

We next used SAMOSA-ChAAT to study ATP-dependent chromatin remodeling by purified SNF2h at single-fiber resolution. Our multiplexing strategy allowed us to study multiple reaction conditions in parallel, including a pre-catalytic state (SNF2h(-)ATP), a remodeled steady state (SNF2h(+)ATP), and an uncatalyzed state where ADP is added instead of ATP (SNF2h(+)ADP), which we collectively compared against unremodeled chromatin fibers (i.e. ‘native’ fibers). The endpoint of 15 minutes was chosen to enable reaction completion (> 3 half-times) based on previously measured rate-constants (Yang et al., 2006a). The vast majority of data we collected were under single-turnover and saturating reaction conditions ($[SNF2h] \gg [template]$), though we also performed a small subset of control experiments under multiple turnover conditions where $[SNF2h] < [template]$. Including all replicate and control experiments, and after filtering out molecules that failed quality-control, we analyzed 3.17E6 footprinted molecules across multiple sequencing runs, amounting to a single-molecule fold-coverage of 1.72E6X and 1.45E6X for templates S1 and S2, respectively (**Supplementary Table 3.2**). As with our initial SGD preps, bulk repositioning of

nucleosomes was highly quantitatively reproducible by our footprinting assay (**Supplementary Figure 3.4A-H**).

We focused on exploring the effects of remodeling on ‘10:1’ density S1 and S2 chromatin fibers. First, we visualized the bulk consequences of remodeling fibers through horizon plots (**Figure 3.2A-B; Supplementary Figure 3.4I-J**). Prior studies with mononucleosomes have shown that ATP-dependent chromatin remodelers move nucleosomes off from strong and weak positioning sequences with similar rates (Partensky and Narlikar, 2009). We found that SNF2h remodeling also decreases sequence-dependent nucleosome positioning on fibers—nucleosome-sized footprint midpoints occupied virtually all possible positions along the sequences, overriding observed sequence-dependencies on native fibers. Next, we performed visual inspection of individual sampled fibers before and after remodeling (**Figure 3.2C-D**). Across both fiber types, we observed expected hallmarks of SNF2h remodeling, including the formation of what appear to be evenly-spaced nucleosomal arrays. Importantly, several aspects of our remodeling data recapitulate existing knowledge of how SNF2h binds and remodels mononucleosomes: for instance, remodeling did not substantially impact the estimated numbers of nucleosomes per template, consistent with ISWI remodelers predominantly sliding, not evicting or loading nucleosomes (**Supplementary Table 3.2**), and the precatalytic condition (SNF2h(-)ATP) yielded larger footprints on average but little change in preferred nucleosome positions on templates, consistent with the HSS domain of SNF2h interrogating DNA flanking the nucleosome (Grüne et al., 2003; Hota et al., 2013; Leonard and Narlikar, 2015).

We next hypothesized that single-molecule data might offer insight into SNF2h-catalyzed ‘transitions’ between thermodynamically-favorable versus translocated nucleosome positions. To address this, we formalized the UMAP and Leiden clustering methods into a pipeline for visualizing and clustering SAMOSA remodeling data across multiple catalytic conditions (**Figure 3.2E**), and applied this to native and remodeled 10:1 S1 fibers. UMAP visualization and Leiden clustering (**Figure 3.2F**) of 10:1 S1 fibers across all catalytic conditions resulted in seven individual clusters. These seven clusters were nonuniformly distributed across each catalytic condition, suggesting that our unsupervised clustering approach can

distinguish fiber types on the basis of remodeling condition (**Figure 3.2G**). Visualizing these distributions as alluvial plots (**Figure 3.2H**) demonstrates the potential of our approach; by quantifying the relative changes in cluster usage over different conditions (e.g. increase in cluster 2 representation under saturating SNF2h remodeling conditions), our data enable a comprehensive view of heterogeneous state spaces. These data and analyses demonstrate the ability of our platform to study and potentially model dynamic changes in chromatin fiber structure at high-throughput and at single-molecule resolution.

3.6 SNF2h does not preferentially clamp trinucleosomes on individual fibers

Our data provide a unique opportunity to distinguish between the ‘clamping’ and ‘length-sensing’ models of SNF2h remodeling (**Figure 3.3A**). To begin testing these two models, we queried how remodeled fibers were structured at the level of trinucleosomes. Specifically, we computed the following: for every series of three successive footprints on each molecule, we determined the distance between the second and third footprint midpoints (‘d2’) and plotted this as a function of the distance between the first and second footprint midpoints (‘d1’; **Figure 3.3B**). The simplest prediction of the clamping model is that the distances between three successive remodeled nucleosomes should cluster around a fixed length, with unequal distances resulting from trinucleosomes where nucleosome n3 is positioned distantly enough to evade clamping (**Figure 3.3C**). Conversely, if SNF2h moves nucleosomes via length-sensing, we expect to observe weak correlation between d1 and d2, with remodeling i.) increasing the correlation between these distances as a function of increasing nucleosome density, and ii.) decreasing the average distance between nucleosomes as a function of density (**Figure 3.3D**).

Trinucleosomal distance scatterplots for unremodeled and remodeled S1 and S2 fibers are visualized in Figure 3.3E-H. In all cases, SNF2h remodeling had a clear visual effect on pairwise distance distributions, converting sequence-programmed distance-dependencies (shown as ‘patches’ of enrichment in unremodeled fibers) into a random configuration of successive distances. The degree to which remodeling impacted the observed correlation between distances varied across different density

preparations and different sequences (**Supplementary Figure 3.5A-B**). In some cases, remodeling altered the sign of the correlation, though this was not necessarily dependent on starting chromatin density. On first glance, our results seem evocative of the clamping model, as fixed distances appear more frequently at higher densities. Comparisons of remodeled scatter plots (**Figure 3.3F,H**) with unremodeled scatter plots (**Figure 3.3E,G**), however, reveal that much of this signal is in fact encoded prior to remodeling, and is reliant on high nucleosome density. In reality, SNF2h remodeling products do not display density-independent trinucleosomal spacings, nor do they display favored dinucleosomal spacings (**Supplementary Figure 3.5C-D**). Instead, our data are consistent with SNF2h translocating nucleosomes towards the longer linker DNA without clamping di- and tri-nucleosomes on individual fibers.

3.7 Density-dependent SNF2h remodeling generates chromatin fibers with a range of regular spacings

To further discriminate between existing models of SNF2h activity, we carried out a second single-molecule analysis—directly detecting array regularity on individual fibers through autocorrelation (**Figure 3.4A**). We previously demonstrated that single-molecule autocorrelograms and unbiased clustering could be used to simultaneously ascertain regularity and NRL directly from single-molecule data (Abdulhay et al., 2020b). We applied this analysis to all footprinted native and remodeled fibers, and then clustered molecules on the basis of similar autocorrelograms, arriving at 8 distinct S1 fiber clusters (**average signals shown in Figure 3.4B**). These clusters classified footprinted molecules by increasing average distance between nucleosomes across entire single DNA templates, simultaneously capturing molecules with consistent NRLs (e.g. clusters 1 - 3), and molecules where a regular pattern was not detected (clusters 7-8; likely owing to hyper- or hypomethylation of the fiber). To ascertain how cluster usage differed as a function of nucleosome density, we visualized cluster enrichment as a heatmap of effect sizes stratified by density (**Figure 3.4C**), as well as stacked bar graphs capturing the absolute abundance of each cluster as functions of density and remodeling (**Figure 3.4D**).

Importantly, our method allowed us to visualize results for both native and remodeled fibers to account for the random formation of nucleosome arrays by statistical positioning downstream of free DNA template ends (Kornberg and Stryer, 1988). Most prior biochemical reactions have been studied at higher nucleosome densities (equivalent of 18 nucleosomes per S1 template) used in our studies. The products of SNF2h remodeling at these densities appear less heterogeneous than the products at lower densities (**Figure 3.4C,D**). However, at these densities the starting architecture of fibers is also less heterogeneous than at lower densities due to the effects of statistical positioning. More broadly, these results illustrate how nucleosome density influences the state distribution of remodeling outcomes. Even at relatively low fiber densities (e.g. 5 – 7 nucleosomes per template), SNF2h remodeling generates a distribution of regular fibers of various predicted NRLs.

To better quantify this distribution before and after remodeling, we computed the probability of observing fibers in each cluster at each nucleosome density, and calculated a log-odds ratio comparing probability distributions for remodeled versus native fibers (**Supplementary Figure 3.6**). These plots capture the complexity of the ‘state space’ of SNF2h remodeling: remodeling increases the representation (shown in red) of multiple fiber types at multiple densities, with underlying nucleosome density shaping the steady-state distribution of fiber structures. These results, which extend to a completely different primary DNA sequence (**Supplementary Figure 3.7**), provide further evidence for the length-sensing model of SNF2h activity, and suggest two fundamental properties of remodeling by the SNF2h enzyme: array formation and spacing is influenced by underlying nucleosome density, and reaction outcomes are highly heterogeneous at all densities when the starting nucleosome architectures are explicitly considered. Importantly, these *in vitro* results are inconsistent with the clamping model.

3.8 SNF2h remodeling modulates motif site exposure frequency in a nucleosome density and sequence-dependent manner

How ATP-dependent chromatin remodelers and TFs collaborate in trans to access the chromatinized genome is an area of ongoing investigation. We reasoned that our dataset would allow us to ask how nucleosome density, primary sequence, and nucleosome sliding activities synergize to regulate the accessibility of specific sequence motifs. Sequences S1 and S2 collectively harbor three predicted murine 17 bp Ctf binding motifs (referred to here as S1.1, S2.1, and S2.2), and the accessibility of these motifs *in vivo* depends to various extents on SNF2h activity (Barisic et al., 2019b). We aimed to define how SNF2h activity may increase or decrease the nucleosome occupancy likelihood of these motifs (**Figure 3.5A**). To do so, we examined the relative methyltransferase accessibility of every possible 17-mer along our sequences before and after remodeling and computed a log-odds ratio that captures relative increases (in red) and decreases (in blue) in motif accessibility as a function of sequence position, chromatin density and SNF2h remodeling (**Figure 3.5B**; S1 fold changes from -10.5-fold decrease to 63.6-fold increase; S2 fold changes from 25.8-fold decrease to 60.1-fold increase). We highlight two features of SNF2h activity resulting from this analysis: first, remodeler-dependent motif exposure *in vitro* is influenced by both starting nucleosome positions and nucleosome density; second, the ability of SNF2h to create accessibility at particular Ctf sites (**Figure 3.5C-E**; specific fold changes noted inline) varies substantially, as remodeling differentially impacts the relative accessibility of sites in a sequence- and density-dependent manner (e.g. 3.76-fold increase after remodeling of 8:1 density templates for S1.1; 1.07-fold decrease after remodeling 3:1 density templates for S2.2). The maps presented here provide an essential starting point for modeling how the coupled influences of sequence, nucleosome density, and nucleosome sliding activities quantitatively increase or decrease the probability of site exposure on a population of templates.

3.9 SNF2h loss *in vivo* leads to bidirectional, domain-specific shifts in chromatin fiber structure

Substantial prior work has demonstrated that SNF2h-containing complexes can open and repress chromatin accessibility (Barisic et al., 2019b; Fyodorov et al., 2004; Wiechens et al., 2016; Xiao et al., 2001), but how these regulatory modes are deployed *in vivo* remains unclear. We speculated that length-sensing by SNF2h may enable these regulatory modes, and performed *in vivo* experiments to test this hypothesis. In mammalian cells, SNF2h (encoded by SMARCA5 / Smarca5) is one of two possible catalytic subunits for several ISWI remodeling complexes, each with specific genome-wide localization and activity (Badenhorst et al., 2002; Ito et al., 1997; Varga-Weisz et al., 1997). Importantly, SNF2h is nonessential in murine embryonic stem cells (mESCs), offering a unique opportunity to study how steady-state chromatin fiber structure *in vivo* is impacted by the addition of SNF2h *in trans* (Barisic et al., 2019b). We applied an improved, *in situ* version of the SAMOSA protocol (**Supplementary Figure 3.8**) to footprint feeder-cultured mESCs devoid of SNF2h (Smarca5^{-/-} mESCs; i.e. ‘knockout’), knockout cells expressing a wildtype copy of the SNF2h protein (i.e. ‘addback’) (Barisic et al., 2019b), and wildtype feeder-free E14 mESCs (i.e. ‘E14’). Across all cell lines and including biological replicates, we sequenced 1.66E7 individual fibers, the equivalent of ~9X haploid coverage of the mouse genome. We used these data to ask (**Figure 3.6A**): i.) how does SNF2h-loss impact the distribution of fiber structures genome-wide; and ii.) where do SNF2-mediated structural changes occur across the mESC epigenome?

We first inspected the average methylation profile of footprinted fibers from knockout and addback cells (**Figure 3.6B**); consistent with previous results, we found that knockout cells had globally longer NRLs compared to addback cells (Barisic et al., 2019b). We then clustered single-molecule autocorrelograms for all data falling within one of eleven different genomic domains. As previously observed in K562 cells (Abdulhay et al., 2020b), this analysis resulted in clusters that stratify the genome on the basis of array regularity and NRL (**Figure 3.6C**). Our unsupervised approach yielded 8 clusters—six regular clusters ranging in NRL from ~172 bp to ~198 bp, and two irregular clusters with weak nucleosome phasing (IRL200) or undetected phasing (IR). These clusters were tightly associated with the

underlying nucleosome density of clustered fibers (**Figure 3.6D**). Knockout cells were globally enriched for irregular IRL200 / IR clusters and low-density NRL198 fibers, while addback cells were conversely enriched for higher density, short NRL fiber types (**Figure 3.6E**), consistent with the pattern observed on bulk averages. These patterns were highly reproducible, both visually (**Supplementary Figure 3.9A**) and quantitatively (**Supplementary Figure 3.9B**).

In previous work, we observed that epigenomic domains were highly heterogeneous with respect to chromatin fiber usage (Abdulhay et al., 2020). In light of this heterogeneity, we sought to determine whether SNF2h loss would consistently impact fiber usage patterns across the epigenome, in line with ‘clamping’ activity *in vivo*. To test this, we examined fibers falling into one of nine different epigenomic domains (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3, telomeric sequence, major satellite, and minor satellite), and computed the relative enrichment and depletion of fiber types within each domain in knockout and addback cells (**Figure 3.6F**; reproducibility shown in **Supplementary Figure 3.9C-D**). Intriguingly, we found that the reintroduction of SNF2h in addback cells had domain-specific effects: at predicted active promoters, for instance, the addition of SNF2h leads to increased representation of ‘irregular’ and long NRL fibers; at predicted H3K36me3 regions, SNF2h increased the representation of intermediate-length NRL fibers; finally, at typically unmappable heterochromatic major and minor satellite sequences, the addition of SNF2h led to increased representation of short NRL fiber types, consistent with SNF2h condensing chromatin fibers in this context. These results demonstrate that SNF2h remodeling outcomes are domain-specific, suggest that domain-level nucleosome density impacts SNF2h remodeling, and provide further evidence against ISWI clamping *in vivo*.

3.10 SNF2h tunes heterogeneous fiber usage patterns to specify bulk chromatin accessibility

To further explore how SNF2h might differentially regulate specific sites *in vivo*, we re-analyzed paired ATAC-seq datasets from knockout and wildtype mESCs (**Supplementary Figure 3.10A**) (Barisic et al., 2019b). This allowed us to define genomic sites that (directly or indirectly) depend on SNF2h to

either open or close chromatin in cell populations. We then extended our enrichment analyses to examine how single-fiber usage patterns are modulated by SNF2h for these sets of loci. Comparing knockout and addback fibers across these sites, we find that sites dependent on SNF2h to remain closed demonstrate a subtle increase in the relative abundance of regular, long NRL arrays, while sites that depend on SNF2h to create accessibility have an increased representation of irregular fibers (**Supplementary Figure 3.10B**). These fiber usage distributions are highly heterogeneous (**Supplementary Figure 3.10C**), hinting that SNF2h tunes, rather than specifies, fiber usage patterns to change bulk chromatin accessibility.

We next focused specifically on the transcription factor Ctf, whose occupancy in mammalian cells depends on SNF2h activity (Barisic et al., 2019b; Wiechens et al., 2016). Specifically we examined, across both cell lines, footprinted chromatin fibers containing known gold-standard Ctf bound sites in the mESC epigenome (Yue et al., 2014), as well as fibers harboring motif matches drawn randomly from the approximately 2.9E6 Ctf sites found across the murine genome (i.e. unoccupied Ctf sites) (Vierstra et al., 2020). Averaging signal over all of these fibers centered at predicted Ctf motifs, we recapitulated previous findings, observing that Ctf site accessibility is dramatically impacted by the loss of SNF2h in vivo, although SNF2h-loss does not lead to complete occlusion of factor binding sites (**Figure 3.7A**; compare with unbound Ctf sites). These results were orthogonally corroborated by ATAC fragment analyses at these same sites.

We next quantified the impact of SNF2h remodeling on Ctf occupancy patterns on single-molecules. To do this, we aggregated all single-molecule observations within a 500 nucleotide window surrounding bound sites and clustered all molecules to determine nucleosome occupancy patterns immediately surrounding Ctf binding sites. Our analyses yielded 6 primary patterns of nucleosome occupancy over these motifs (**Figure 3.7B**; **Supplementary Figure 3.11**), including two irregular, or ‘offset’ nucleosome occupancy patterns (clusters O1 and O2), three nucleosome occluded patterns harboring well-phased nucleosomes in various registers (clusters R1 – R3), and a TF accessible cluster of molecules, a subset of which display unmethylated footprints potentially capturing direct Ctf-DNA

interactions (cluster A) (**Supplementary Figure 3.11**). Cluster usage was highly heterogeneous across knockout and addback cells (**Figure 3.7C**), a finding that was underscored by analysis of a single SNF2h-dependent Ctf site (12X coverage in knockout; 10X coverage in addback) in the murine genome (**Figure 3.7D**). At this site, fibers of each type were observed, demonstrating extensive heterogeneity in fiber patterns at a single, SNF2h-regulated locus. In line with this, quantification of differential cluster usage across knockout and addback cells revealed a modest depletion (1.43-fold for cluster O2) for offset clusters and modest enrichment (2.04-fold for cluster A) for the accessible cluster in addback cells. These distributions suggest that SNF2h again tunes chromatin accessibility at these sites through continuous nucleosome sliding, and not through programmed clamping of nucleosomes; we speculate that this sliding is in turn regulated by local nucleosome density, which controls the relative amount of extranucleosomal flanking DNA available to allow SNF2h catalysis (Leonard and Narlikar, 2015).

3.11 Discussion

Dissecting chromatin remodeling outcomes at single-fiber resolution using SAMOSA-ChAAT

Modern chromatin biology sits amid a ‘resolution revolution.’ Advances in cryogenic electron microscopy (‘cryo-EM’) have provided us with near-atomic views of macromolecular chromatin-interacting complexes (Armache et al., 2019; Eustermann et al., 2018; Patel et al., 2019). Complementarily, advances in single-molecule and high-resolution microscopic approaches *in vitro* and *in vivo* have provided new views of dynamic and often heterogeneous chromatin conformations (Blosser et al., 2009; Boettiger et al., 2016; Kim et al., 2021). Finally, advances in high-throughput short-read sequencing have offered near nucleotide-resolution maps of where and how these complexes engage with chromatin genome-wide, across myriad substrates *in vitro*, and even at the resolution of single-cells (De Dieuleveult et al., 2016; Henikoff et al., 2011; Krietenstein et al., 2016; Lai et al., 2018). SAMOSA-ChAAT provides a fourth advance in chromatin resolution—datasets describing the molecularly-resolved activity of chromatin regulators on individual chromatin fibers. Our data and associated computational pipelines offer a new frontier for

quantifying dynamic chromatin-associated processes that complements existing high-resolution approaches. In future work, we anticipate using SAMOSA-ChAAT to study post-translationally modified chromatin fibers, as well as fibers undergoing additional dynamic nuclear processes (e.g. transcription, replication, loop extrusion).

SNF2h is a length-sensing, nucleosome-density dependent chromatin remodeler

Chromatin remodelers regulate nucleosome spacing *in vitro* and *in vivo*, but the question of how chromatin remodelers like SNF2h space nucleosomes on individual fibers remains open. Using SAMOSA-ChAAT, we performed the first (to our knowledge) single-fiber-resolution footprinting experiments on reconstituted, remodeled, murine genomic templates of varying nucleosome density. Our *in vitro* results highlight three key properties of SNF2h remodeling: first, remodeling outcomes are highly heterogeneous and largely ablate sequence-programmed nucleosome positions, consistent with prior findings that SNF2h remodeling rates are insensitive to nucleosome stability, and implying that remodelers can override intrinsic DNA driven nucleosome positioning (Partensky and Narlikar, 2009; Zhang et al., 2011); second, remodeling products bear little evidence of so-called ‘clamping,’ as the final distributions of internucleosomal distances and single-fiber nucleosome arrangements catalyzed by SNF2h remodeling vary as a function of underlying chromatin density; third, both primary sequence and nucleosome density contribute to whether SNF2h increases or decreases DNA site accessibility, which we demonstrate using the cognate binding motif for the CTCF transcription factor (**Figure 3.8A**). Given our results, we propose an updated model for SNF2h action on chromatin fibers. *In vitro*, SNF2h does not program fixed IDs; instead, length equalization is a steady-state outcome of moving nucleosomes in the direction of the longer linker DNA, consistent with a length-sensing model for SNF2h activity.

SNF2h employs distinct regulatory modes to repress and open chromatin in vivo

How do the length-sensing properties of SNF2h manifest *in vivo*? All of the SNF2h activities discussed here, including length-dependent sliding (Stockdale et al., 2006; Yang et al., 2006a; Zofall et al., 2004), active positioning of nucleosomes downstream of barriers (Krietenstein et al., 2016; Yen et al., 2012), and the formation of well-spaced nucleosome arrays (Fyodorov et al., 2004; Ito et al., 1997), have been noted in previous work, but how these sometimes disparate activities harmonize to impact gene regulation *in vivo* has remained elusive. Our data from living mESCs provide evidence that linker-length-sensing can account for all of these classically-defined activities (**Figure 3.8B**). At regions where SNF2h is required to maintain heterochromatic structure (i.e. regions of relatively high nucleosome density), the remodeler converts irregular and long NRL fibers into well-spaced nucleosome arrays with multiple short NRLs. How well-ordered arrays repress chromatin remains unknown, but it is tempting to speculate that this process either facilitates ‘elimination’ of nucleosome-free regions (NFRs) by preventing cryptic NFR formation (Garcia et al., 2010), by promoting chromatin compaction (Correll et al., 2012), or by generating NRLs particularly suited for phase separation (Gibson et al., 2019). At euchromatic regions where SNF2h is required to create chromatin accessibility (i.e. regions with relatively low nucleosome density), sliding is now used by SNF2h to generate ‘disordered’ fibers to increase the site exposure frequency of cis-regulatory elements like CTCF binding sites. While our data explain the multiple observations of coupled ISWI activity and TF binding (Badenhorst et al., 2002; Barisic et al., 2019b; Wiechens et al., 2016; Xiao et al., 2001), how SNF2h remodeling facilitates other dynamic nuclear processes (e.g. transcription, replication, repair, higher-order chromatin folding) remains an exciting question.

Nucleosome density as an additional substrate cue for targeting chromatin remodeling activity

Our understanding of how sequence-non-specific chromatin remodeling complexes achieve specificity at genomic loci is still developing. Prior work has uncovered myriad remodeler-targeting ‘cues,’

including post-translational histone modifications (Clapier and Cairns, 2012; Dann et al., 2017; Mashtalir et al., 2021), TFs (Barisic et al., 2019b; Brahma and Henikoff, 2019; De Dieuleveult et al., 2016), three-dimensional (3D) chromosomal architecture (Barisic et al., 2019b; Barutcu et al., 2016; Weber et al., 2021), composition of the nucleosome core particle (Dann et al., 2017; Gamarra et al., 2018; McBride et al., 2020), and, as discussed here, availability of extranucleosomal flanking DNA (Stockdale et al., 2006; Yang et al., 2006a; Zofall et al., 2004). Here, we elaborate on the length-sensing model by connecting DNA length-sensing on mononucleosomes to nucleosome density of individual chromatin fibers. At high nucleosome densities, flanking DNA is occluded and ISWI remodeling outcomes are constrained to create populations of evenly-spaced arrays capable of repressing chromatin. These fiber-type distributions are likely further regulated by ISWI complex composition (Eberharther et al., 2001; Hamiche et al., 1999; Ito et al., 1997; Varga-Weisz et al., 1997). At low nucleosome densities, extranucleosomal DNA is abundant and ISWI catalysis is unconstrained. This enables continuous nucleosome sliding, allowing trans acting factors to overcome nucleosomal repression of regulatory DNA.

Study limitations & Concluding Remarks

Our study does have limitations. First, we have restricted our analyses to the ISWI ATPase SNF2h—as discussed above, SNF2h operates in the context of complexes whose additional non-catalytic subunits can modulate SNF2h activity. In future work, we anticipate broadly applying SAMOSA-ChAAT to study how ATP-dependent chromatin remodeling complexes and other chromatin-associated factors alter chromatin fiber structure. Second, our assay inherently captures the accessibility of primary sequence in the context of chromatin, which can be influenced by both primary structure (e.g. histone-DNA interactions) and higher-order solution chromatin structures (e.g. fiber folding). As such, our accessibility patterns must not be overinterpreted as solely representing nucleosome occupancy, and our assay does not capture the same real-time dynamics as *in vitro* and *in vivo* single-molecule experiments. Finally, like other *in vivo* single-molecule footprinting studies to date (Abdulhay et al., 2020; Lee et al., 2020; Sönmezer et al., 2021;

Stergachis et al., 2020), our mESC data are necessarily averaged over a multitude of discrete cellular states, including the cell cycle. As chromatin structure changes dramatically as a function of cell cycle progression, a fraction of the single-molecule patterns we observe *in vivo* must represent known ‘bespoke’ chromatin structures (e.g. mitotic chromatin). Future work in synchronized or post-mitotic cells will help discriminate the specificity of observed single-molecule patterns at distinct cell cycle stages.

Many questions remain: how is SNF2h activity constrained? SNF2h-containing complexes require specific flanking DNA lengths to enable catalysis—do other factors act *in trans* to effectively gate flanking DNA lengths and inhibit SNF2h activity? How do SNF2h-containing complexes collaborate with other remodeling families? At actively transcribed gene bodies, for instance, INO80, CHD, and ISWI complexes regulate nucleosome positioning in an evolutionarily-conserved manner, and dissecting the contribution of each remodeler remains important. How might nucleosome density (and, by extension, SNF2h activity) be controlled *in vivo*? In mammals, nucleosome density is likely regulated on a diversity of length scales, ranging from local (e.g. ATP-dependent chromatin remodeling; histone chaperones; histone modification; replication, transcription, and repair), to global (e.g. genome compartmentalization / phase separation; loop extrusion; subnuclear localization). We envision a regulatory circuit wherein the concentration of core histones can be tuned within large chromatin domains by specific *trans* and *cis*-regulatory elements. This circuit could influence the regulatory outputs of remodeling complexes over long genomic distances, allowing higher-order genome conformation to instruct local interpretation of regulatory DNA.

3.12 Figures

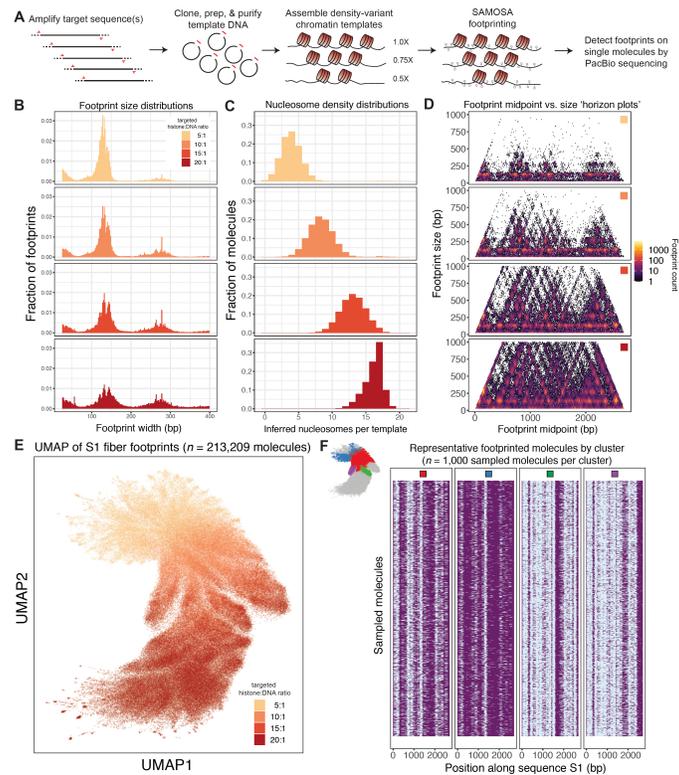


Figure 3. 1: SAMOSA-ChAAT enables massively multiplex dissection of single-fiber nucleosome positioning on *in vitro* reconstituted genomic chromatin fibers.

A.) Schematic overview of the SAMOSA-ChAAT protocol, wherein genomic sequences are cloned, purified, and assembled into chromatin fibers with desired biochemical properties (e.g. nucleosome density) through salt gradient dialysis. Fibers are then footprinted with a nonspecific adenine methyltransferase and sequenced on the PacBio platform to assess single-molecule nucleosome positioning.

B.) A custom analytical pipeline enables detection of methyltransferase footprints on sequenced fibers. Footprint sizes from SAMOSA-ChAAT experiments carried out at varying nucleosome densities follow closely with expected nucleosome sizes, plus expected ‘breathing’ of DNA around the histone octamer, with the extent of breathing decreasing as nucleosome density increases.

C.) SAMOSA-ChAAT data enables direct estimation of the absolute number of nucleosomes per footprinted fiber, which track well with expected nucleosome densities based on targeted octamer : DNA ratios during salt gradient dialysis.

D.) Footprint length vs. midpoint ‘horizon’ plots for footprinted fibers. Average nucleosome positions display sequence dependencies.

E.) UMAP dimensionality reduction of fiber accessibility data. UMAP patterns recapitulate known differences in nucleosome density in footprinted fibers.

F.) Visualization of a subset of sampled molecules following Leiden clustering of single molecule data. Individual Leiden clusters (cluster positions inset) capture mutually exclusive nucleosome positions consequent of chromatin fiber assembly.

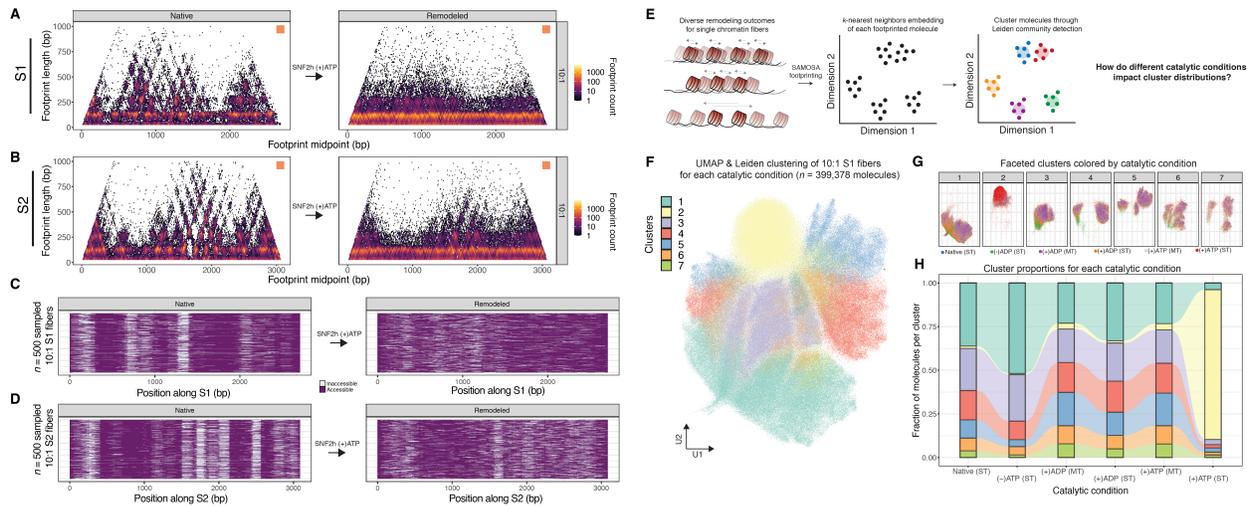


Figure 3. 2: SAMOSA-ChAAT reveals chromatin remodeling outcomes at single-fiber resolution.

A-B.) Footprint length versus footprint midpoint horizon plots capture bulk outcomes of SNF2h remodeling of 10:1 targeted density S1 fibers (A) and S2 fibers (B). In both cases, remodeling relaxes bulk sequence preferences for nucleosome assembly.

C-D.) Sampled single-molecule data for 10:1 S1 (C) and S2 (D) fibers before and after single-turnover SNF2h remodeling. Remodeling translocates nucleosomes from preferred starting positions, leading to the formation of nucleosome arrays.

E.) Schematic representation of a computational approach to infer how remodeling alters fiber structure, by using a combination of k-nearest neighbors visualization (UMAP) and clustering (Leiden) to define the ‘state-space’ of possible chromatin accessibility patterns.

F.) UMAP visualization and Leiden clustering (colors) of all 10:1 fibers from six different catalytic condition experiments.

G.) UMAP visualization faceted by cluster type, colored by catalytic condition.

H.) Alluvial plots visualize shifts in cluster distribution across each catalytic condition.

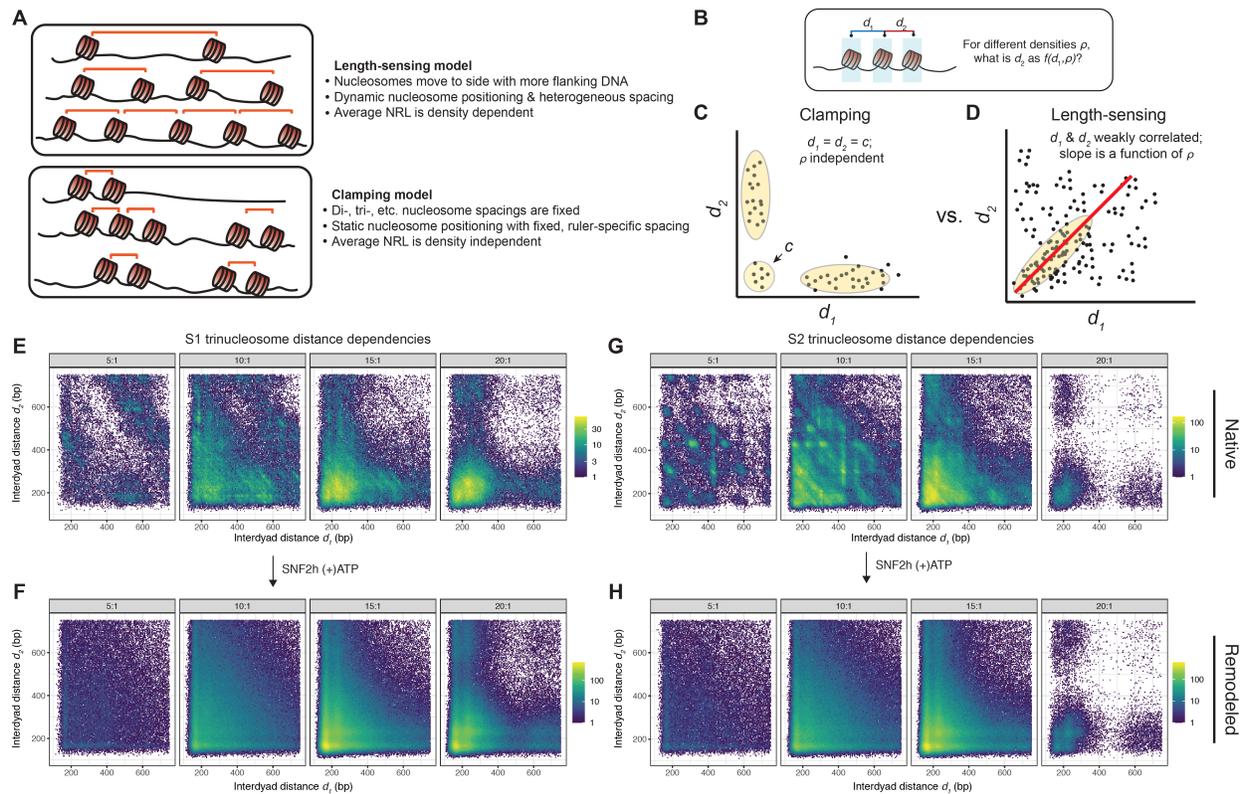


Figure 3. 3: SNF2h does not fix trinucleosomal spacings on individual chromatin fibers.

A.) Schematic of two competing models for how SNF2h creates regular nucleosome arrays; the ‘length-sensing’ model suggests that array regularity is a steady-state product of flanking DNA sampling by the remodeler and subsequent translocation of nucleosomes towards sides with longer flanking DNA length. The ‘clamping’ model suggests that SNF2h clamps nucleosomes to fixed distances.

B.) We constructed an analysis to determine how SNF2h remodels trinucleosomes on individual fibers by testing for correlation between the distances between successive nucleosomes.

C.) Schematic of expected results of the trinucleosome distance analysis if SNF2h acts according to the ‘clamping’ model.

D.) Schematic of expected results of the trinucleosome distance analysis if SNF2h acts according to the ‘length-sensing’ model.

E-F.) Scatter plots of the resulting correlations, for unremodeled S1 (E) and remodeled S1 fibers. Remodeling diversifies possible internucleosomal distances, as evidenced by the loss of ‘patchy’ sequence-dependent paired nucleosome distances upon remodeling.

G-H.) As in E-F.), but for unremodeled S2 fibers (G) and remodeled S2 fibers.

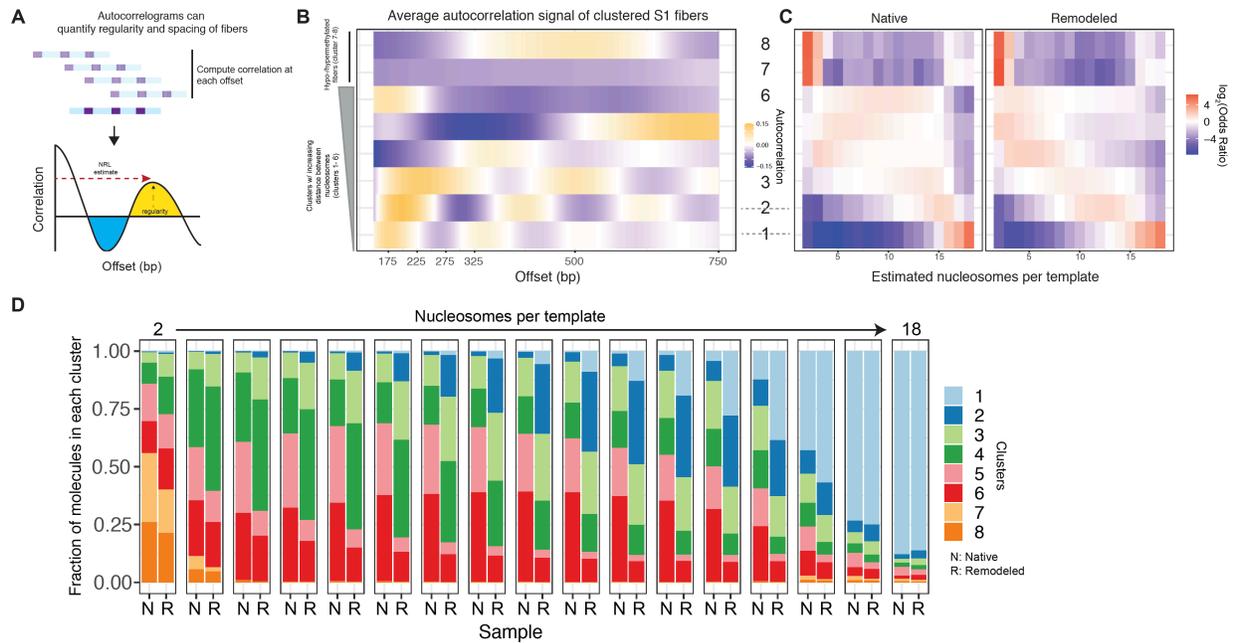


Figure 3. 4: Autocorrelation analyses reveal that chromatin density influences the heterogeneous reaction outcomes of SNF2h remodeling.

A.) Single-molecule autocorrelation can be used to estimate relative spacing and regularity of single, footprinted chromatin fibers. A schematic of how the autocorrelation function is calculated is illustrated here.

B.) Results of Leiden clustering of all native and remodeled fiber autocorrelograms. Leiden clustering results in 8 clusters, the first six of which have been classified on the basis of increasing average distance between nucleosomes, and the latter two of which appear to harbor highly hypo- or hyper-methylated fibers.

C.) Fisher’s exact test enrichment results (log-transformed odds ratios) surveying the relative enrichment and depletion of clusters at each chromatin density (# of nucleosomes per template). Separation of results between ‘Native’ and ‘Remodeled’ classes demonstrates the contribution of statistical nucleosome positioning in unremodeled samples, while illustrating how density can affect fiber state distributions upon remodeling.

D.) A stacked bar-chart representation of the relative abundance of each fiber-type, stratified by estimated chromatin fiber density; N: native, R: remodeled.

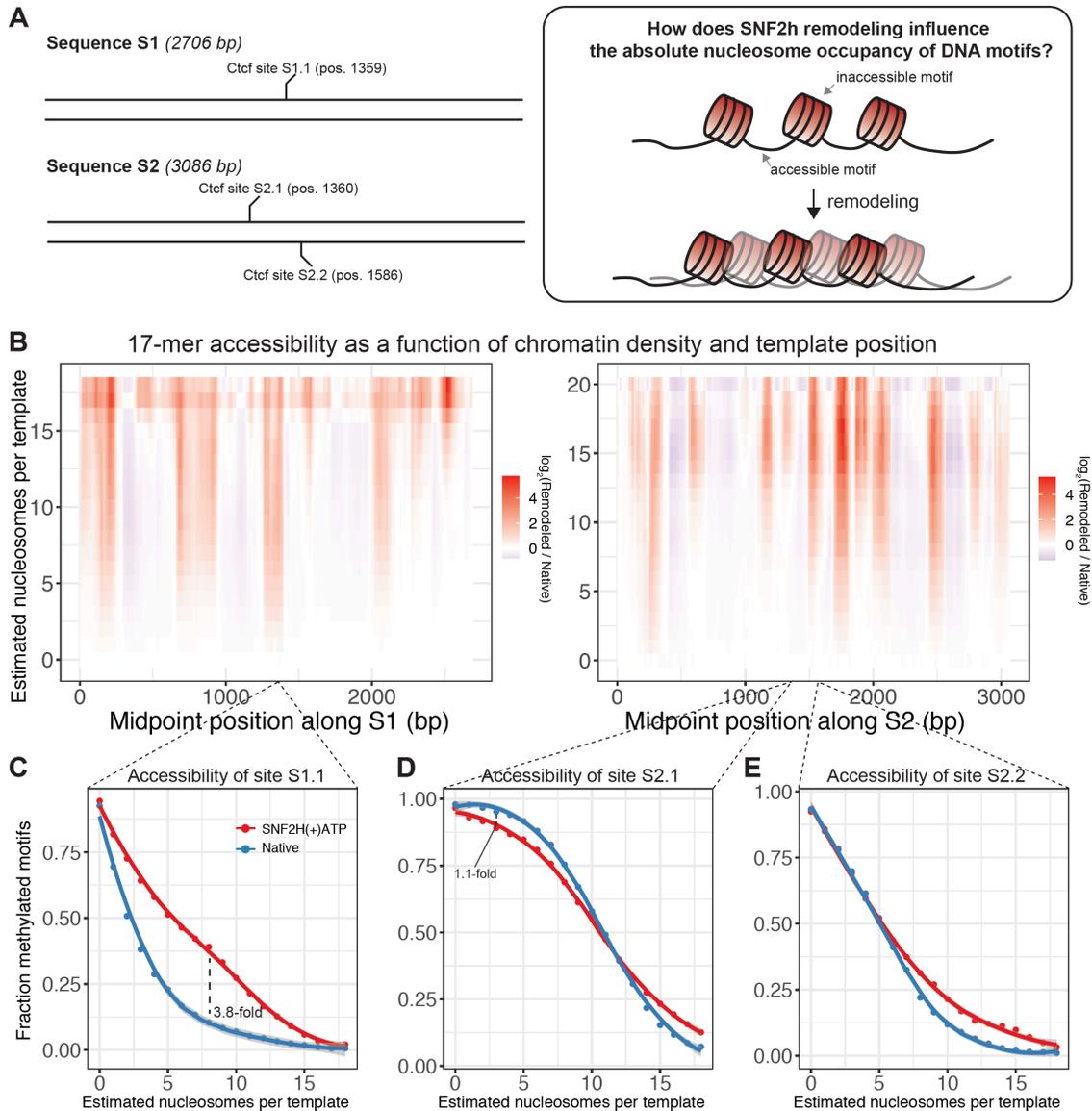


Figure 3. 5: SNF2h remodeling influences motif site exposure in a nucleosome-density and sequence-dependent manner.

A.) Sequences S1 and S2 collectively contain 3 matches to the canonical 17 bp CTCF / Ctcf binding site (referred to here as S1.1, S2.1, and S2.2). We sought to determine how SNF2h remodeling could influence the probability that these 17 bp sites might be accessible using methylation as a proxy.

B.) Density vs. sequence heatmaps relating the \log_2 fold change in probability of accessibility of every possible 17 bp motif along sequences S1 and S1. Regions in red represent the 17 bp sites where remodeling increases the probability of methylation; conversely, regions in blue are sites where remodeling decreases the probability of methylation. Accessibility was calculated as a binary variable gated on whether > 90% of the 17 bp motif was methylated on an individual molecule.

C.-E.) Fractional accessibility as a function of nucleosome density for CTCF / Ctcf site S1.1 (C), site S2.1 (D), and site 2.2 (E) for native fibers (blue) and remodeled fibers (red). Specific values for fold change increases (C) or decreases (D) highlighted inline.

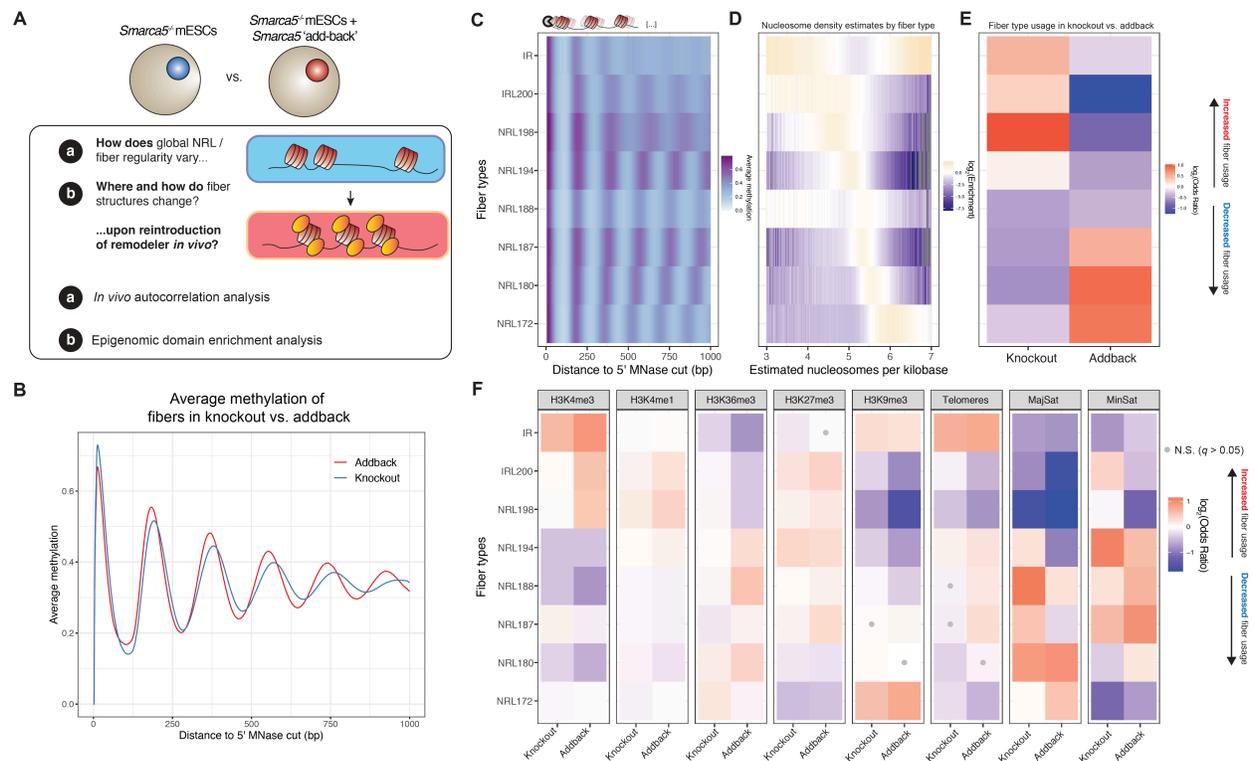


Figure 3. 6: Mapping the *in vivo* consequences of SNF2h remodeling in murine embryonic stem cells at single fiber resolution.

A.) Overview of *in vivo* experimental design. We performed *in vivo* SAMOSA footprinting in mESCs devoid of SNF2h (*Smarca5*^{-/-}; ‘knockout’ cells) and mutant mESCs overexpressing *Smarca5* cDNA (addback cells). We then used the resulting single-molecule data to determine how the addition of SNF2h impacts chromatin fiber architecture globally and at specific epigenomic domains.

B.) Average single-molecule methylation patterns in knockout cells (blue) and addback cells (red). Reintroduction of SNF2h decreases the overall nucleosome repeat length (NRL) on individual fibers.

C.) Leiden clustered single-molecule autocorrelograms stratify the mESC epigenome into eight distinct clusters (i.e. ‘fiber types’), two ‘irregular’ clusters and six regular clusters arranged here in decreasing NRL.

D.) Single-molecule nucleosome density estimates for molecules in each of the six clusters, colored by increasing mean nucleosome density. Calculations were made with respect to the background frequency of fiber density estimates.

E.) Global enrichment of fiber usage plotted as a log-transformed Fisher’s Exact odds ratio of fiber types in knockout versus addback cell lines. Knockout cells are enriched for irregular and long NRL fibers, and the addition of SNF2h leads to the depletion of long NRL / irregular fiber types and enrichment of short NRL fibers in addback cells.

F.) Differences in fiber type composition across varying epigenomic domains in knockout versus addback cells. Actively transcribed regions are enriched for lower-density fibers, and heterochromatic regions are enriched for higher-density fibers. Enrichments are calculated as log-transformed Fisher’s Exact odds ratios. Tests that are not significant (*q*-value threshold > 0.05) are marked with a grey dot.

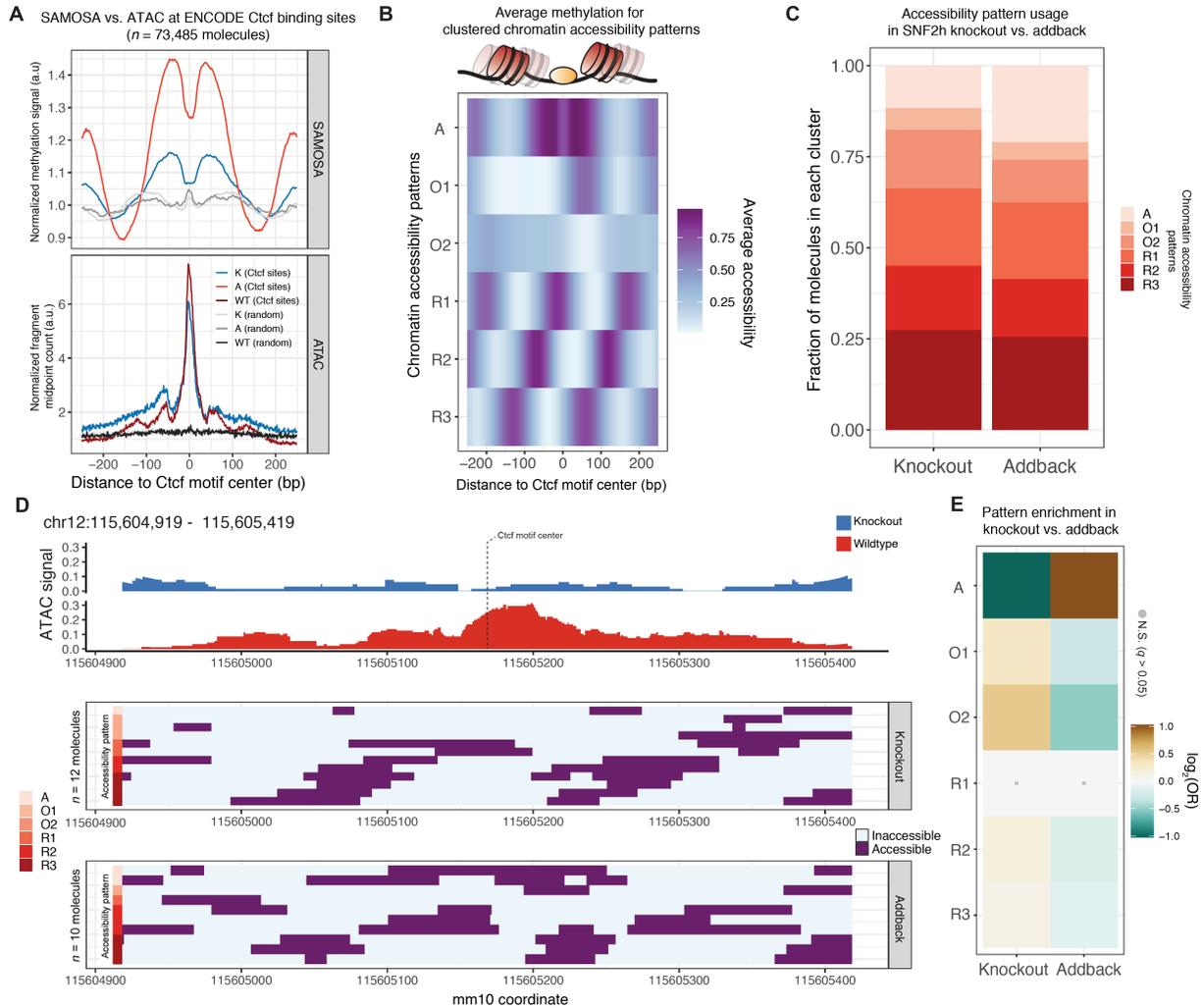


Figure 3. 7: Exploring SNF2h-mediated chromatin closing and opening by integrating ATAC-seq and SAMOSA data.

A.) Line plots for normalized SAMOSA (top) and ATAC-seq (bottom) signal at ENCODE-defined Ctf sites and unbound Ctf motif matches in the mESC epigenome.

B.) Average accessibility patterns for clustered Ctf site occupancy patterns. Clusters can be broken down into three regular nucleosome occupancy patterns (R1 – R3), two irregular occupancy patterns (O1-O2), and one accessible cluster (A).

C.) Stacked bar chart representation of occupancy pattern cluster distribution in knockout vs. addback cells.

D.) Single-site representation of a Ctf site where bulk chromatin accessibility is lost in the knockout genome. Single-molecules covering this site are shown below and are labeled with respective cluster labels. **G.)** Fisher’s exact test results for enrichment (gold) or depletion (green) of occupancy patterns in knockout vs. addback cells. Tests that are not significant following q-value correction ($q < 0.05$) are marked with a grey dot.

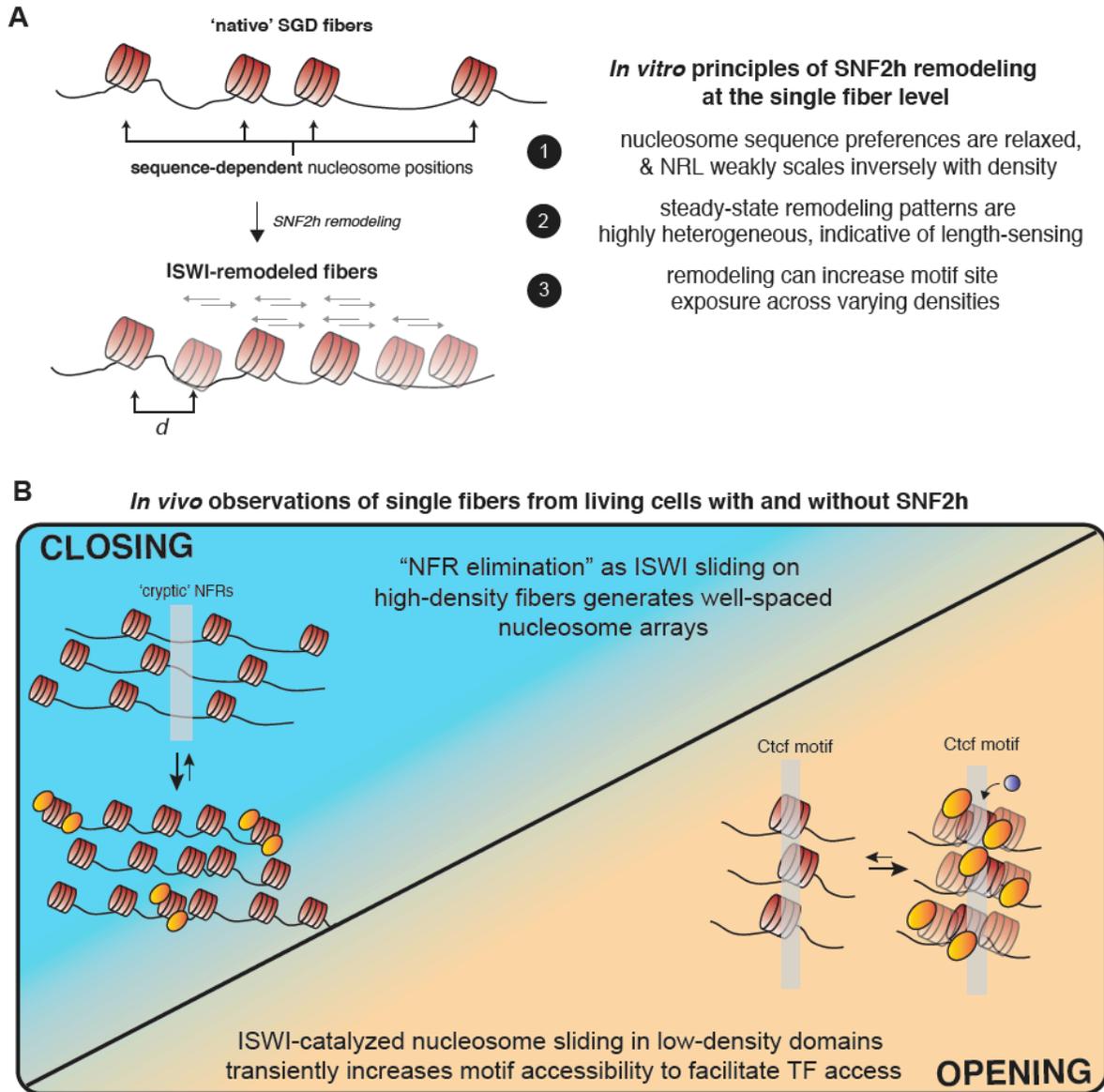
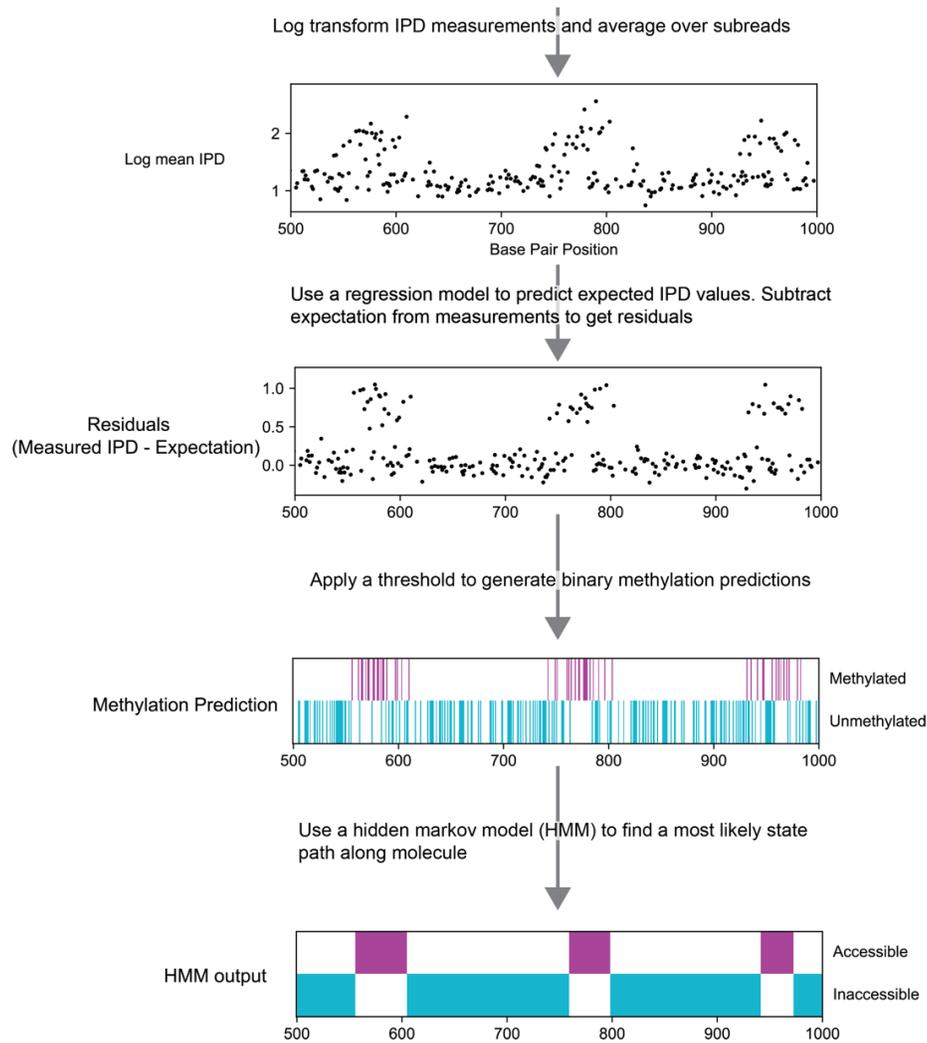


Figure 3. 8: A model of SNF2h-mediated chromatin regulation based on results of this study.

A.) In vitro, SNF2h acts stochastically, sliding nucleosomes that are pre-positioned by primary DNA sequence to new positions. Reaction outcomes are highly heterogeneous and NRLs created by SNF2h remodeling are weakly inversely proportional to the underlying nucleosome density of fiber substrates—evidence that SNF2h operates through length-sensing of extranucleosomal DNA. Finally, remodeling can increase (or decrease) motif site exposure in a nucleosome-density- and sequence-dependent manner.

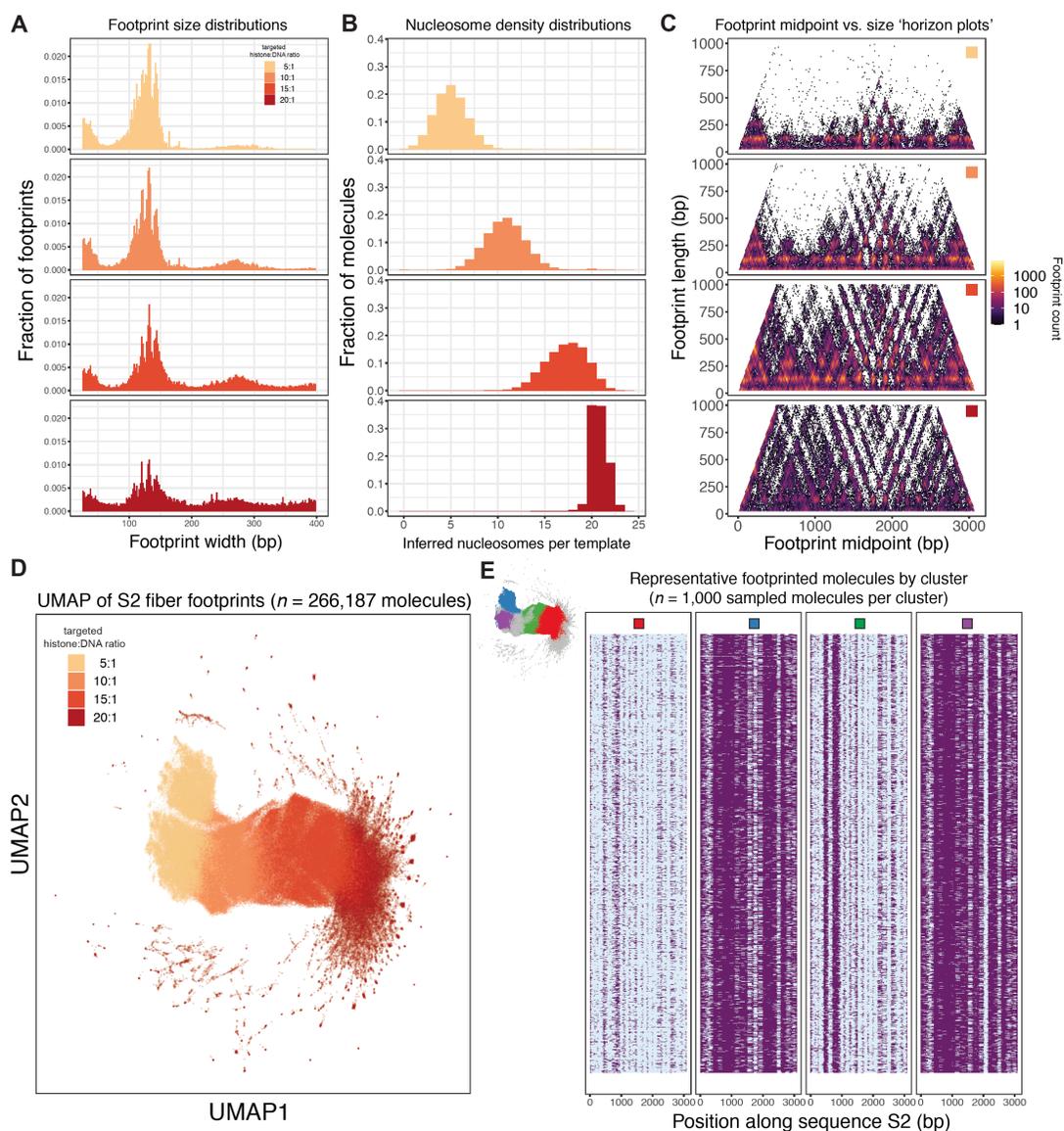
B.) In vivo, SNF2h length-sensing can explain two diametrically-opposed regulatory functions. At high-nucleosome-density repressed regions, SNF2h increases the representation of multiple types of regular, short NRL fibers, presumably to facilitate the elimination of cryptic NFRs. At lower-nucleosome-density, accessible cis-regulatory elements (e.g. CTCF / Ctfc binding sites), SNF2h slides nucleosomes to increase the site exposure frequency of these sites, enabling TF access.

3.13 Supplementary Figures



Supplementary Figure 3. 1: Computational pipeline for inferring DNA accessibility from measured inter-pulse distance (IPD).

Shown is example data for a portion of a methylated molecule containing nucleosomes assembled onto regularly spaced Widom 601 sequences. The pipeline starts with \log_{10} transforming the IPD measurements and averaging over all subreads. Next, to reduce noise from DNA sequence effects and inter-molecular variation, a neural network regression model that was trained on unmethylated DNA is used to regress out the expected IPD at each adenine. The regression model takes into account the DNA sequence context as well as molecule level IPD distribution measurements. The residuals show greater signal, and a threshold is then applied to the residuals to get binary methylation predictions. A hidden Markov model (HMM) is then used to synthesize the information from all adenines across the molecule into a single trace of accessible and inaccessible regions. The HMM model uses the frequency at which adenines in different sequence contexts were methylated in unmethylated and fully methylated control molecules to set expectations for observing methylation in accessible and inaccessible regions of chromatin. This HMM output was used for all downstream analyses.



Supplementary Figure 3. 2: SAMOSA-ChAAT generalizes to other genomic sequences of interest.

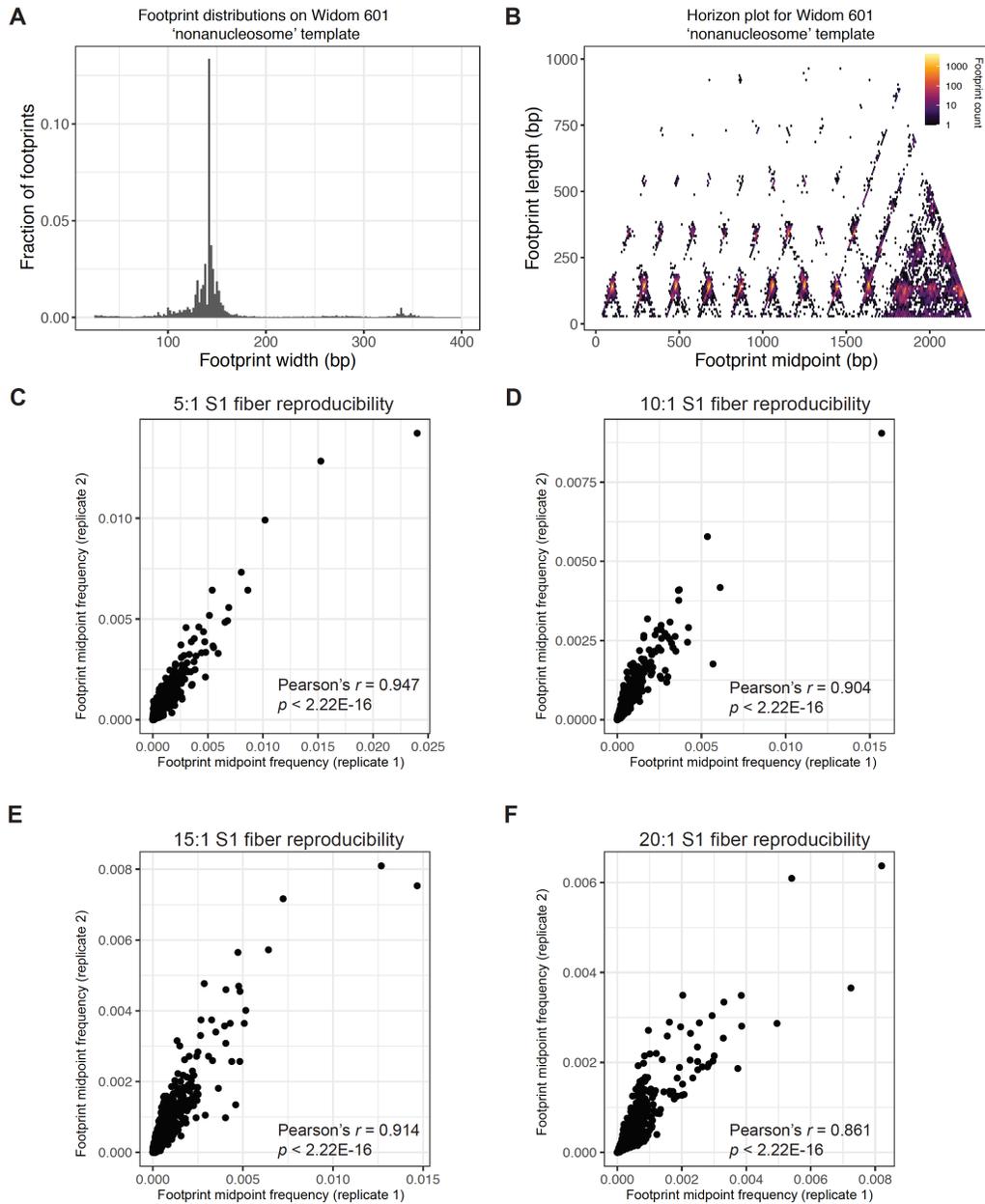
A.) As in Figure 1B, but for a completely different murine sequence ('S2'). Footprint sizes from SAMOSA-ChAAT experiments carried out at varying nucleosome densities follow closely with expected nucleosome sizes, plus expected 'breathing' of DNA around the histone octamer, with the extent of breathing decreasing as nucleosome density increases.

B.) SAMOSA-ChAAT data enables direct estimation of the absolute number of nucleosomes per footprinted S2 fiber.

C.) Footprint length vs. midpoint 'horizon' plots for footprinted S2 fibers.

D.) UMAP dimensionality reduction of S2 fiber accessibility data.

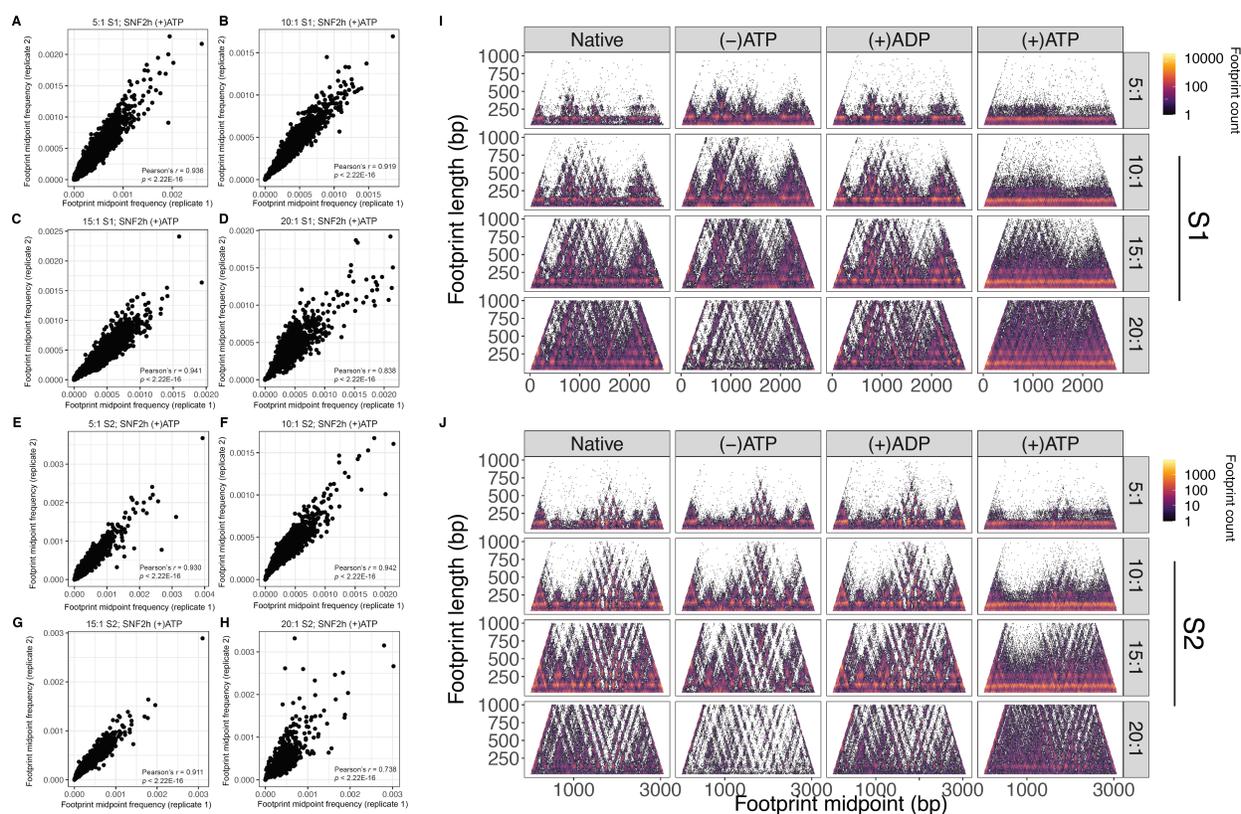
E.) Visualization of a subset of sampled molecules following Leiden clustering of single molecule data.



Supplementary Figure 3.3: SAMOSA-ChAAT is accurate and reproducible.

A.-B.) Widom 601 nonanucleosomal fiber data from 28 was reprocessed using the NN-HMM. Called footprints are the expected length of 601-assembled nucleosomes (A), and horizon plots reveal positioned nucleosomes at expected positions (B).

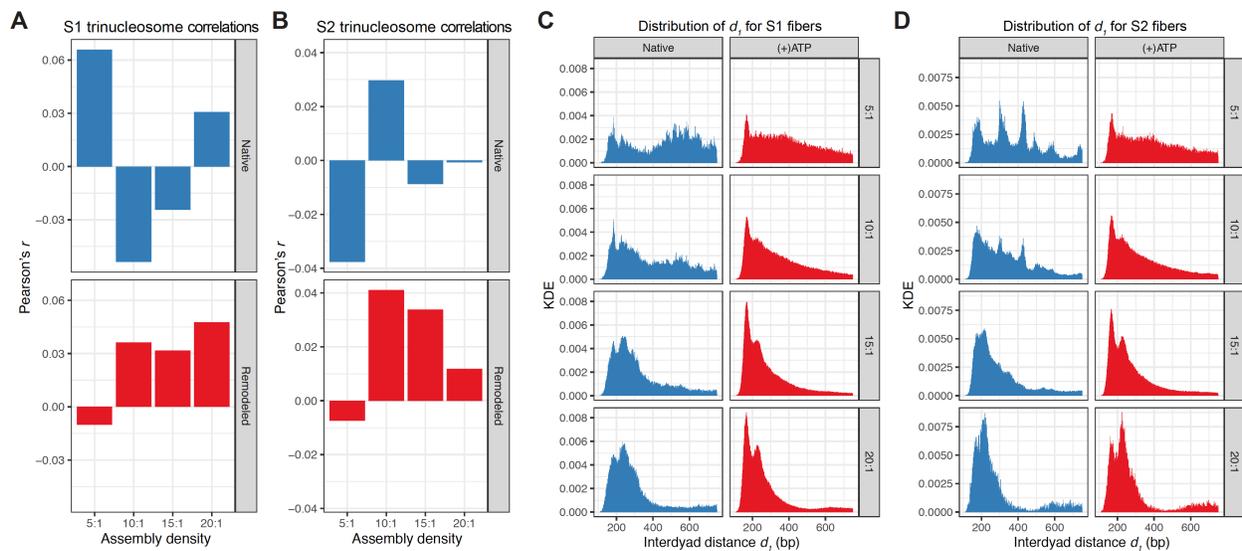
C.-F.) Correlation of footprint abundances for S1 fibers of each density across two replicates (different remodeling reactions on fibers generated from different salt gradient dialysis preps). SAMOSA-ChAAT measurements are highly reproducible.



Supplementary Figure 3. 4: Reproducibility of SAMOSA-ChAAT remodeling experiments and horizon plots for all catalytic conditions tested.

A.-H.) Scatter plots and associated Pearson's r values for correlations between two biological replicate remodeling experiments, for each density tested, for both S1 (A-D) and S2 (E-H) fibers.

I.-J.) Horizon plots for S1 (I) and S2 (J) fibers, for native, pre-catalytic, (+)ADP, and remodeled fibers (all averages are over single-turnover experiments; multi-turnover data is omitted for this visualization).

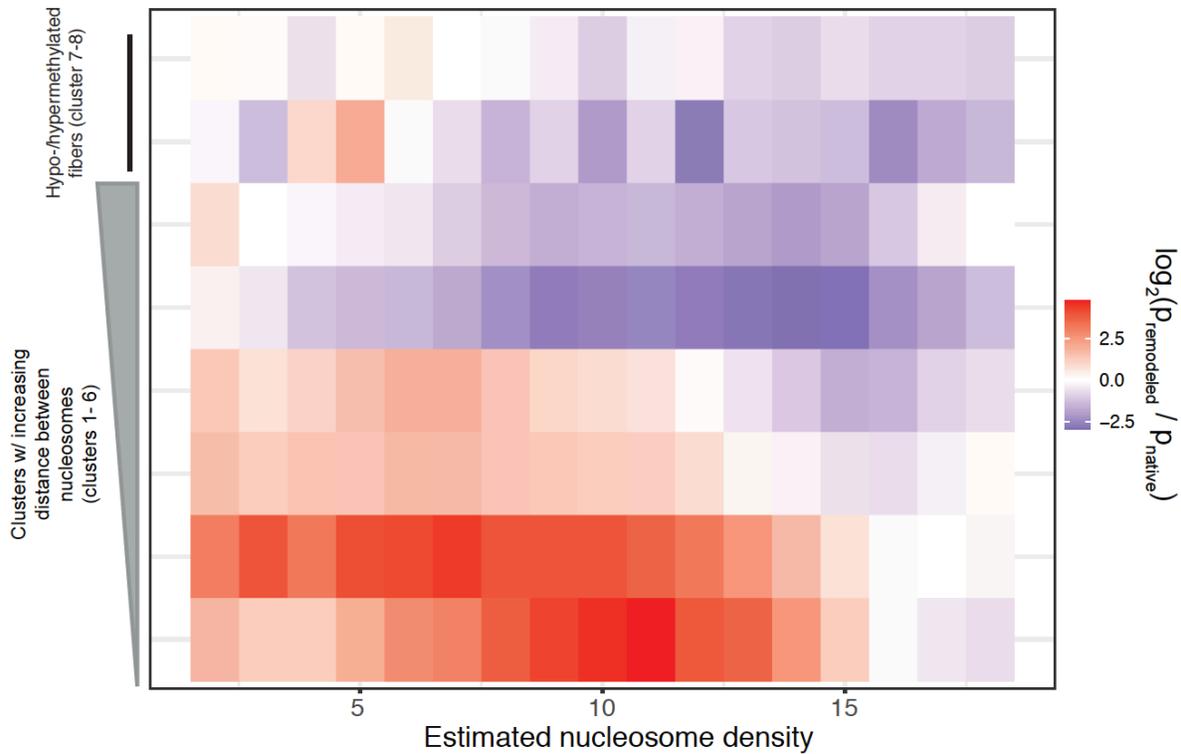


Supplementary Figure 3.5: Nucleosome spacing correlation is slightly impacted by chromatin remodeling, and dinucleosome distances are not fixed by SNF2h remodeling.

A.-B.) Bar chart representation of Pearson's r values for each assembly density, for unremodeled (blue) and remodeled (red) S1 (A) and S2 (B) fibers. Correlations are globally very low (indicating little coupling between the two distances), and is slightly impacted in both directions upon remodeling, though the effects are small.

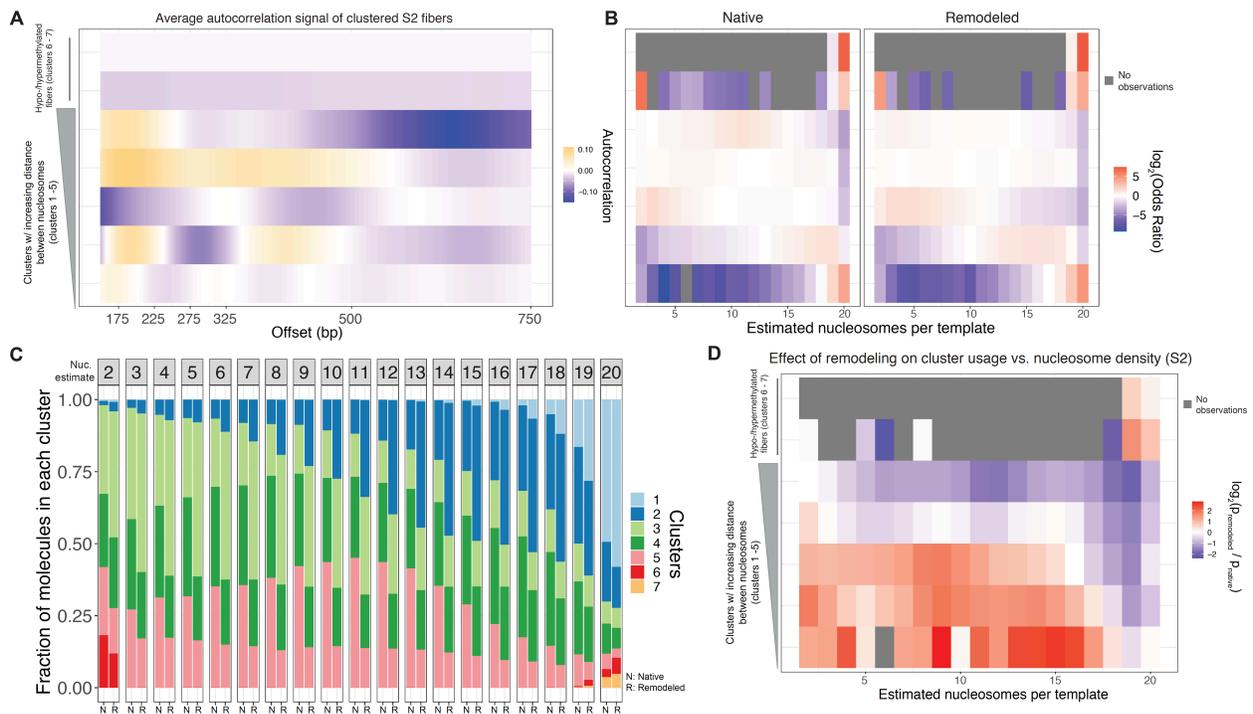
C.-D.) Histograms (binsize = 1 bp) for unremodeled (blue) and remodeled (red) interdyad distances (i.e. d_1) for S1 (C) and S2 (D) fibers.

Effect of remodeling on cluster usage vs. nucleosome density



Supplementary Figure 3. 6: SNF2h remodeling alters the probability of observing specific chromatin fiber structures.

SNF2h remodeling alters the probability of observing specific chromatin fiber structures and remodeling outcomes are highly heterogeneous. Log-odds visualization capturing the likelihood that chromatin fiber states are increased (red) or decreased (blue) in representation with respect to native, unremodeled fibers. p_{remodeled}: fraction of molecules of a given cluster for each nucleosome density estimate after SNF2h remodeling; p_{native}: fraction of molecules of a given cluster for each nucleosome density estimate before SNF2h remodeling.



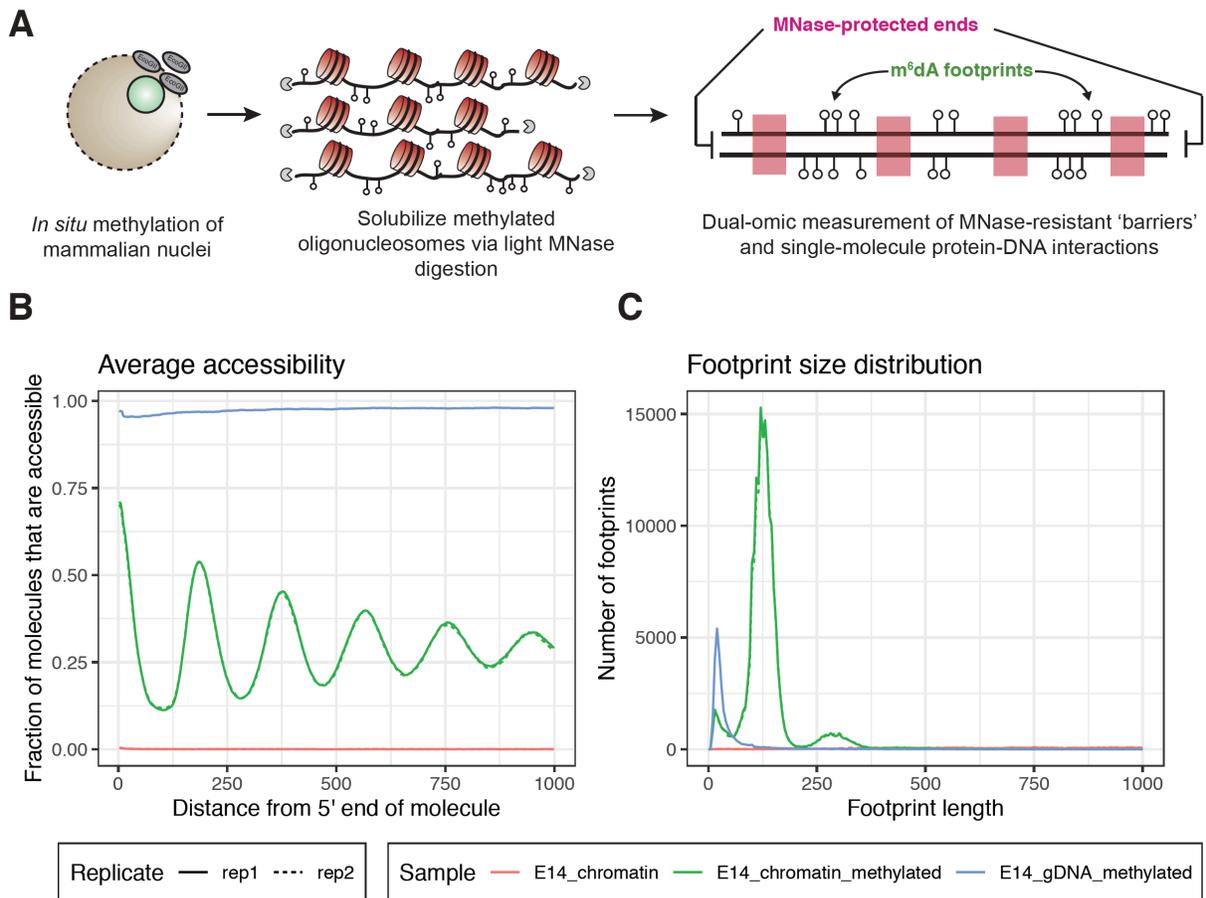
Supplementary Figure 3. 7: Autocorrelation clustering analysis for fiber S2.

A.) As in Figure 4B, but for fiber S2; unsupervised Leiden clustering of single-molecule autocorrelations yields 7 clusters.

B.) As in Figure 4C, Fishers exact effect sizes for cluster enrichment as a function of estimated chromatin density.

C.) As in Figure 4D, a stacked bar-chart representation of the relative abundance of each fiber-type, stratified by estimated chromatin fiber density, for S2 fibers.

D.) Log-odds visualization capturing the likelihood that chromatin fiber states are increased (red) or decreased (blue) in representation with respect to native, unremodeled S2 fibers. premodeled: fraction of molecules of a given cluster for each nucleosome density estimate after SNF2h remodeling; pnative: fraction of molecules of a given cluster for each nucleosome density estimate before SNF2h remodeling.

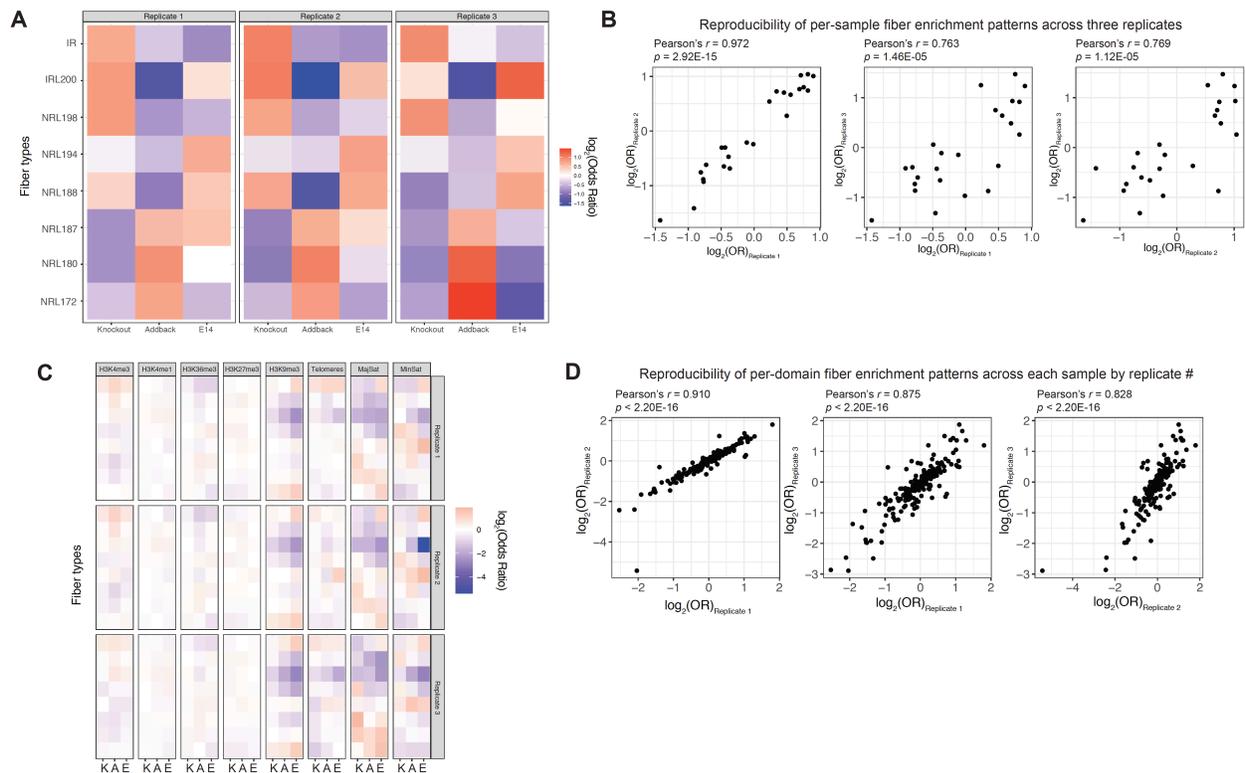


Supplementary Figure 3. 8: An improved, *in situ* SAMOSA assay for profiling single-fiber chromatin structure *in vivo*.

A.) We improved on our previously published SAMOSA protocol by performing EcoGII methylation in intact nuclei, which we then digest with a limited MNase digestion to liberate oligonucleosomes. These molecules are then sequenced on the PacBio Sequel II platform and harbor two information types: MNase cuts that mark the position of 'barriers' along the genome, and m⁶dA footprints that capture protein-DNA interactions.

B.) Our NN-HMM can be applied to estimate chromatin accessibility on individual molecules. Shown here is data from E14 mESCs. Nucleosome periodicity is seen in footprinted chromatin, but not in positive (methylated naked DNA) and negative (unmethylated E14 gDNA) controls. The 5' and 3' ends of molecules are massively enriched for MNase-defined 'barriers' (generally, the edge of nucleosome core particles).

C.) The NN-HMM can predict footprint sizes, which range from nucleosome length, to subnucleosomal protections indicative of transcription factor-DNA interactions.



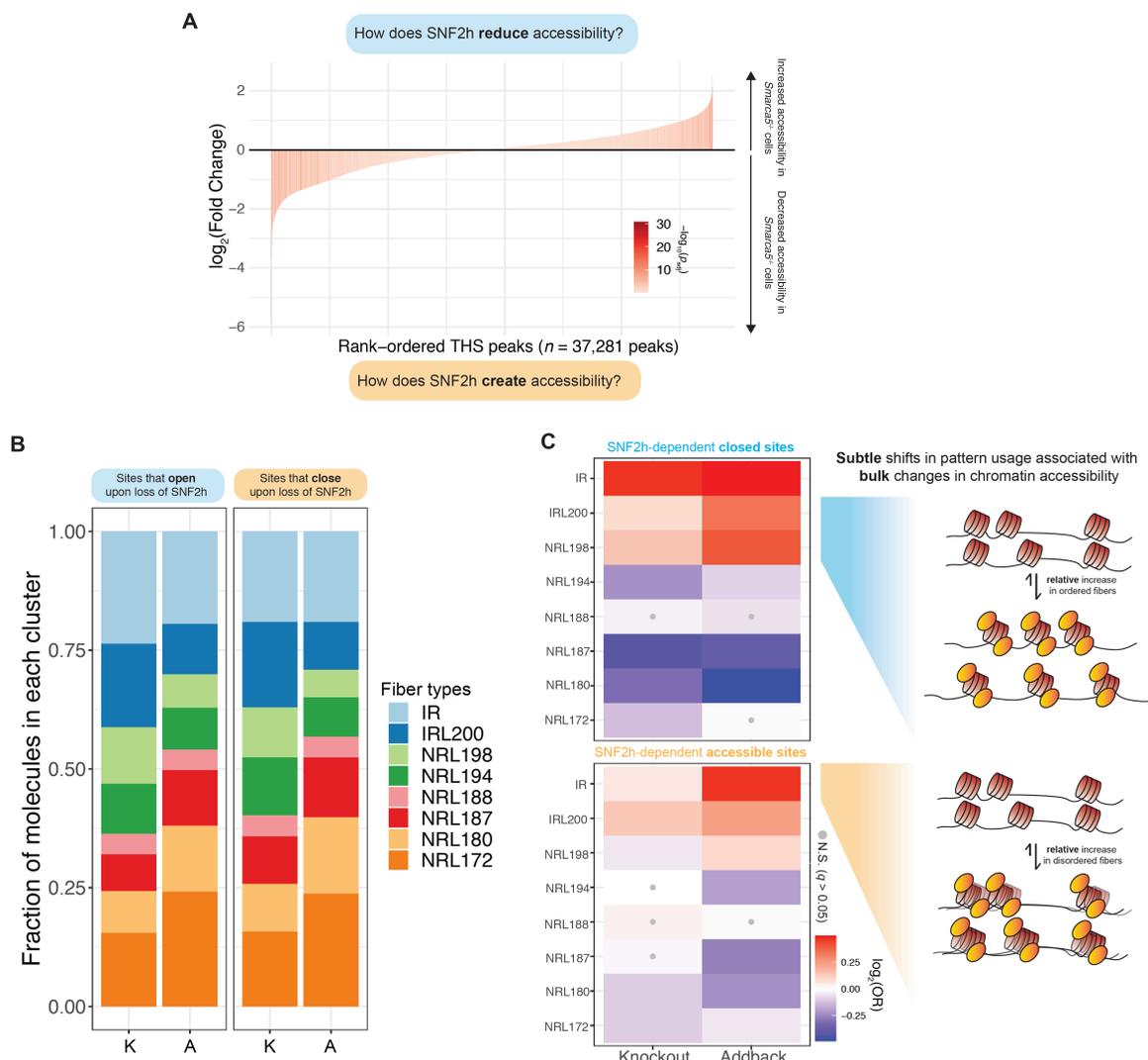
Supplementary Figure 3. 9: Reproducibility of *in vivo* SAMOSA data from knockout, addback, and E14 mESCs.

A.) Fisher's exact test results for sample-level fiber enrichment stratified by biological replicate.

B.) Pearson's r correlations, p -values, and associated scatter plots of effect sizes for three biological replicates.

C.) Fisher's exact test results for domain-level fiber enrichment stratified by biological replicate.

D.) Pearson's r correlations, p -values, and associated scatter plots of effect sizes for domain-level analyses, across three biological replicates.

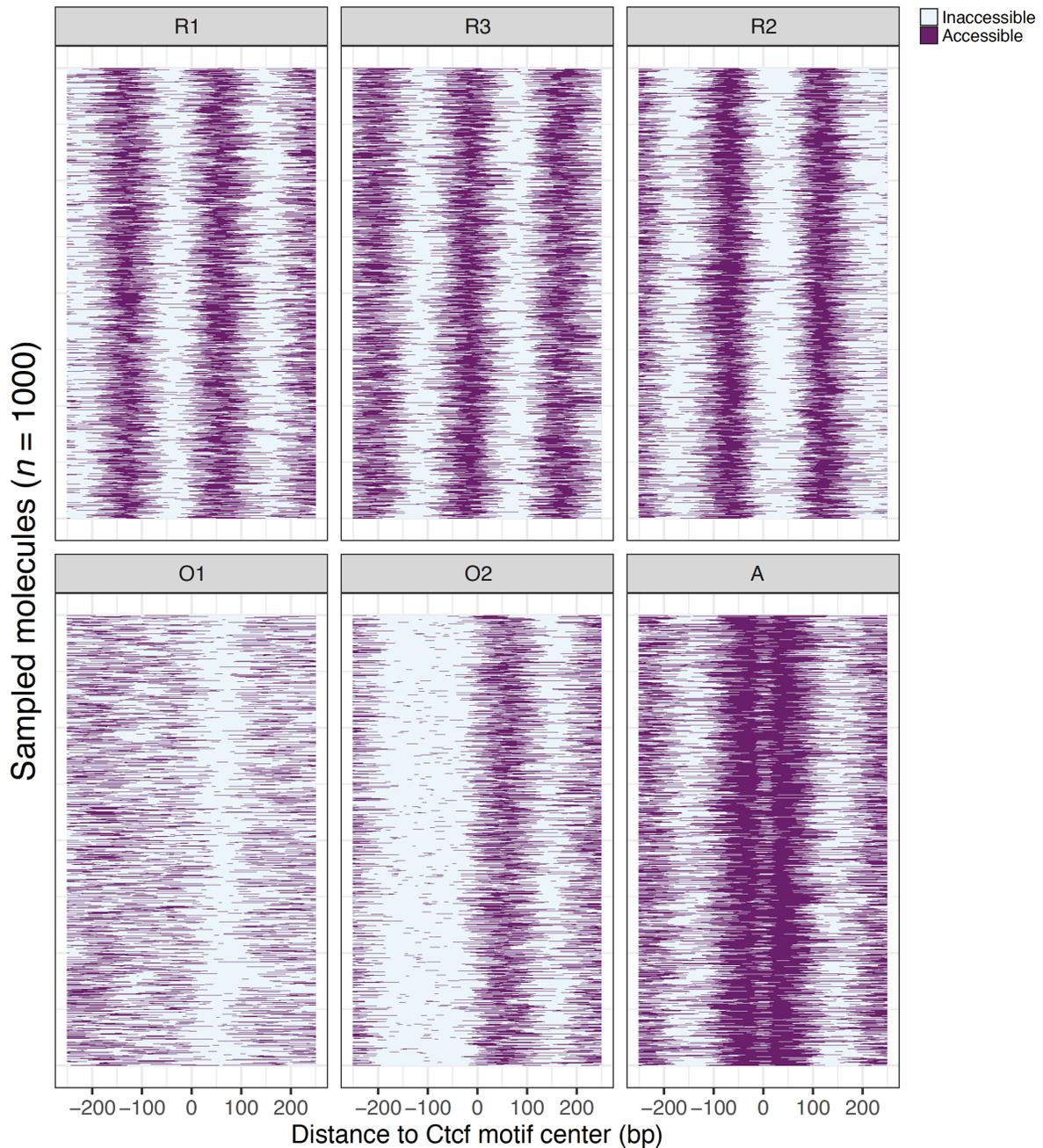


Supplementary Figure 3. 10: Relative abundance of each chromatin fiber type at differentially accessible ATAC-seq peaks.

A.) We re-analyzed published ATAC-seq data to determine statistically significant Tn5-hypersensitive sites (THSs) that are genetically dependent on SNF2h for maintaining chromatin accessibility patterns. Regions that open upon SNF2h-loss can be used to study how SNF2h reduces bulk chromatin accessibility, while regions that close upon SNF2h-loss can be used to study how SNF2h maintains bulk chromatin accessibility.

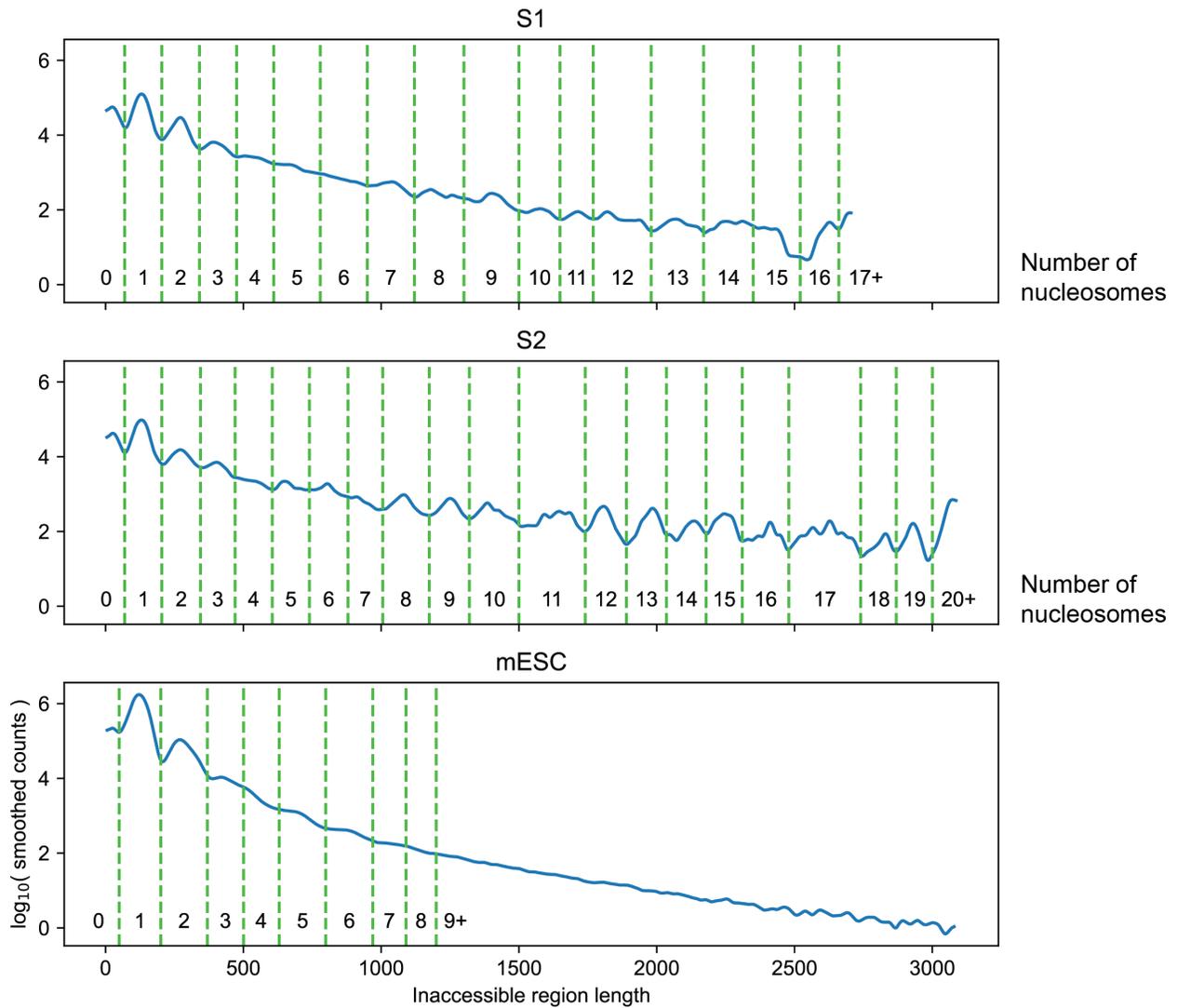
B.) A stacked bar chart representation of the relative usage of each fiber type in (left) regions that are significantly more accessible in knockout cells versus wildtype cells, and (right) regions that are significantly less accessible in knockout cells versus wildtype cells.

C.) Fisher's exact test results for fibers falling in regions that open (top) and regions that close (bottom) in knockout cells by ATAC-seq. Upon reintroduction of SNF2h in addback cells, regions that open increase representation of evenly-spaced chromatin fibers, while regions that close increase the relative representation of irregular fibers.



Supplementary Figure 3. 11: Single-fiber nucleosome occupancy patterns around predicted Ctf binding sites in vivo.

1000 randomly sampled chromatin fibers, with accessibility signal centered around the predicted Ctf binding motif, for each of six different nucleosome occupancy patterns obtained through unsupervised Leiden clustering. R1 – R3 are regularly phased nucleosomes, O1 – O2 are ‘irregular’ nucleosome occupancy patterns that appear to occlude the cognate binding motif for Ctf, and A represents an ‘accessible’ state. A fraction of A molecules display a ‘footprint’ of unmethylated DNA precisely over the Ctf binding site, indicative of a molecule where Ctf was bound during the footprinting reaction.



Supplementary Figure 3. 12: Cutoffs for counting nucleosomes in inaccessible regions based on length.

To estimate the number of nucleosomes on each DNA molecule, cutoffs were defined to delineate between the number of estimated nucleosomes within an inaccessible region. Green dashed lines show the cutoffs, and the numbers below indicate the number of nucleosomes that sized region is counted as. Different cutoffs were used for S1, S2, and mESC molecules, based on the distributions and peaks in region length for each.

3.14 Supplementary Tables

Supplementary Table 3. 1: Summary of sequencing depths for all experiments performed in this study.

For in vivo samples, ‘biorep’ refers to different biological replicate experiments, while ‘techrep’ refers to ‘technical replicates’ where the same sequencing library was sequenced on multiple PacBio Sequel II runs.

Sample Name	in vitro / in vivo	# of molecules
S1_5:1_Native_Rep1	In vitro	27699
S1_10:1_Native_Rep1	In vitro	34876
S1_15:1_Native_Rep1	In vitro	35456
S1_20:1_Native_Rep1	In vitro	41820
S1_5:1_(+)ATP_ST_Rep1	In vitro	46908
S1_10:1_(+)ATP_ST_Rep1	In vitro	49183
S1_15:1_(+)ATP_ST_Rep1	In vitro	54314
S1_20:1_(+)ATP_ST_Rep1	In vitro	66019
S1_5:1_(+)ATP_MT_Rep1	In vitro	7207
S1_10:1_(+)ATP_MT_Rep1	In vitro	7749
S1_15:1_(+)ATP_MT_Rep1	In vitro	14689
S1_20:1_(+)ATP_MT_Rep1	In vitro	18073
S1_5:1_(+)ADP_ST_Rep1	In vitro	51269
S1_10:1_(+)ADP_ST_Rep1	In vitro	59983
S1_15:1_(+)ADP_ST_Rep1	In vitro	58869

Sample Name	in vitro / in vivo	# of molecules
S1_20:1_(+)ADP_ST_Rep1	In vitro	50736
S1_5:1_(+)ADP_MT_Rep1	In vitro	8202
S1_10:1_(+)ADP_MT_Rep1	In vitro	10467
S1_15:1_(+)ADP_MT_Rep1	In vitro	12220
S1_20:1_(+)ADP_MT_Rep1	In vitro	17212
S1_5:1_(-)ATP_ST_Rep1	In vitro	57492
S1_10:1_(-)ATP_ST_Rep1	In vitro	55734
S1_15:1_(-)ATP_ST_Rep1	In vitro	50702
S1_20:1_(-)ATP_ST_Rep1	In vitro	50249
S2_5:1_Native_Rep1	In vitro	27699
S2_10:1_Native_Rep1	In vitro	34876
S2_15:1_Native_Rep1	In vitro	35456
S2_20:1_Native_Rep1	In vitro	41820
S2_5:1_(+)ATP_ST_Rep1	In vitro	34479
S2_10:1_(+)ATP_ST_Rep1	In vitro	38640
S2_15:1_(+)ATP_ST_Rep1	In vitro	49336
S2_20:1_(+)ATP_ST_Rep1	In vitro	43342
S2_5:1_(+)ATP_MT_Rep1	In vitro	4260

Sample Name	in vitro / in vivo	# of molecules
S2_10:1_(+)ATP_MT_Rep1	In vitro	8971
S2_15:1_(+)ATP_MT_Rep1	In vitro	15395
S2_20:1_(+)ATP_MT_Rep1	In vitro	25431
S2_5:1_(+)ADP_ST_Rep1	In vitro	45597
S2_10:1_(+)ADP_ST_Rep1	In vitro	41348
S2_15:1_(+)ADP_ST_Rep1	In vitro	51470
S2_20:1_(+)ADP_ST_Rep1	In vitro	45351
S2_5:1_(+)ADP_MT_Rep1	In vitro	7527
S2_10:1_(+)ADP_MT_Rep1	In vitro	11525
S2_15:1_(+)ADP_MT_Rep1	In vitro	21467
S2_20:1_(+)ADP_MT_Rep1	In vitro	21186
S2_5:1_(-)ATP_ST_Rep1	In vitro	35102
S2_10:1_(-)ATP_ST_Rep1	In vitro	41964
S2_15:1_(-)ATP_ST_Rep1	In vitro	37744
S2_20:1_(-)ATP_ST_Rep1	In vitro	38859
S1_5:1_Native_Rep2	In vitro	8686
S1_10:1_Native_Rep2	In vitro	5773
S1_15:1_Native_Rep2	In vitro	11352

Sample Name	in vitro / in vivo	# of molecules
S1_20:1_Native_Rep2	In vitro	47547
S1_5:1_(+)ATP_ST_Rep2	In vitro	27463
S1_10:1_(+)ATP_ST_Rep2	In vitro	44435
S1_15:1_(+)ATP_ST_Rep2	In vitro	45480
S1_20:1_(+)ATP_ST_Rep2	In vitro	54941
S1_5:1_(+)ATP_MT_Rep2	In vitro	33475
S1_10:1_(+)ATP_MT_Rep2	In vitro	40585
S1_15:1_(+)ATP_MT_Rep2	In vitro	61531
S1_20:1_(+)ATP_MT_Rep2	In vitro	56409
S1_5:1_(+)ADP_ST_Rep2	In vitro	45214
S1_10:1_(+)ADP_ST_Rep2	In vitro	56101
S1_15:1_(+)ADP_ST_Rep2	In vitro	50308
S1_20:1_(+)ADP_ST_Rep2	In vitro	45507
S1_5:1_(+)ADP_MT_Rep2	In vitro	35135
S1_10:1_(+)ADP_MT_Rep2	In vitro	35686
S1_15:1_(+)ADP_MT_Rep2	In vitro	58548
S1_20:1_(+)ADP_MT_Rep2	In vitro	69814
S2_5:1_Native_Rep2	In vitro	24433

Sample Name	in vitro / in vivo	# of molecules
S2_10:1_Native_Rep2	In vitro	34336
S2_15:1_Native_Rep2	In vitro	46634
S2_20:1_Native_Rep2	In vitro	34544
S2_5:1_(+)ATP_ST_Rep2	In vitro	19691
S2_10:1_(+)ATP_ST_Rep2	In vitro	28590
S2_15:1_(+)ATP_ST_Rep2	In vitro	32631
S2_20:1_(+)ATP_ST_Rep2	In vitro	42131
S2_5:1_(+)ATP_MT_Rep2	In vitro	25323
S2_10:1_(+)ATP_MT_Rep2	In vitro	33649
S2_15:1_(+)ATP_MT_Rep2	In vitro	44732
S2_20:1_(+)ATP_MT_Rep2	In vitro	48249
S2_5:1_(+)ADP_ST_Rep2	In vitro	26139
S2_10:1_(+)ADP_ST_Rep2	In vitro	29922
S2_15:1_(+)ADP_ST_Rep2	In vitro	46692
S2_20:1_(+)ADP_ST_Rep2	In vitro	45585
S2_5:1_(+)ADP_MT_Rep2	In vitro	23294
S2_10:1_(+)ADP_MT_Rep2	In vitro	38394
S2_15:1_(+)ADP_MT_Rep2	In vitro	31501

Sample Name	in vitro / in vivo	# of molecules
S2_20:1_(+)ADP_MT_Rep2	In vitro	48805
E14_biorep1_techrep1	In vivo	210215
E14_biorep2_techrep1	In vivo	189956
SNF2hKO_biorep1_techrep1	In vivo	767470
SNF2hKO_biorep2_techrep1	In vivo	550580
SNF2hWTAB_biorep1_techrep1	In vivo	438531
SNF2hWTAB_biorep2_techrep1	In vivo	623798
E14_biorep1_techrep2	In vivo	173438
E14_biorep2_techrep2	In vivo	151407
E14_biorep3_techrep1	In vivo	506067
SNF2hKO_biorep3_techrep1	In vivo	113986
SNF2hWTAB_biorep3_techrep1	In vivo	439218
E14_biorep3_techrep2	In vivo	416576
SNF2hKO_biorep3_techrep2	In vivo	93892
SNF2hWTAB_biorep3_techrep2	In vivo	361727
E14_biorep3_techrep3	In vivo	765720
SNF2hKO_biorep3_techrep3	In vivo	163927
SNF2hWTAB_biorep3_techrep3	In vivo	650995

Sample Name	in vitro / in vivo	# of molecules
SNF2hKO_biorep1_techrep2	In vivo	185098
SNF2hKO_biorep2_techrep2	In vivo	130861
SNF2hWTAB_biorep1_techrep2	In vivo	93524
SNF2hWTAB_biorep2_techrep2	In vivo	99783
E14_biorep1_techrep3	In vivo	1597410
SNF2hKO_biorep2_techrep3	In vivo	6958
SNF2hKO_biorep1_techrep3	In vivo	2666318
SNF2hKO_biorep1_techrep4	In vivo	2869820
SNF2hWTAB_biorep1_techrep3	In vivo	2356244

Supplementary Table 3. 2: Summary of average nucleosomes and standard deviation for all in vitro experiments.

ST refers to single-turnover remodeling reaction conditions, while MT refers to multi-turnover reaction conditions. Data are shown as mean \pm standard deviation.

Fiber Type	Density	Nucleosomes / template	Remodeling status
S1	5:1	4.00 \pm 1.80	Native
S1	10:1	7.93 \pm 2.22	Native
S1	15:1	12.6 \pm 2.16	Native
S1	20:1	15.8 \pm 1.60	Native
S2	5:1	5.17 \pm 2.00	Native
S2	10:1	11.0 \pm 2.16	Native
S2	15:1	17.3 \pm 2.19	Native
S2	20:1	20.8 \pm 0.918	Native
S1	5:1	3.77 \pm 1.93	(+) ATP (ST)
S1	10:1	7.28 \pm 2.04	(+) ATP (ST)
S1	15:1	11.8 \pm 2.01	(+) ATP (ST)
S1	20:1	15.5 \pm 1.80	(+) ATP (ST)
S2	5:1	4.68 \pm 1.93	(+) ATP (ST)
S2	10:1	9.92 \pm 2.11	(+) ATP (ST)
S2	15:1	15.6 \pm 2.10	(+) ATP (ST)
S2	20:1	20.9 \pm 0.987	(+) ATP (ST)

Fiber Type	Density	Nucleosomes / template	Remodeling status
S1	5:1	4.62 ± 1.88	(-) ATP (ST)
S1	10:1	10.2 ± 2.22	(-) ATP (ST)
S1	15:1	15.4 ± 1.65	(-) ATP (ST)
S1	20:1	17.0 ± 0.918	(-) ATP (ST)
S2	5:1	5.22 ± 2.00	(-) ATP (ST)
S2	10:1	11.4 ± 2.10	(-) ATP (ST)
S2	15:1	18.5 ± 1.72	(-) ATP (ST)
S2	20:1	20.8 ± 0.909	(-) ATP (ST)
S1	5:1	3.60 ± 1.45	(+) ATP (MT)
S1	10:1	6.96 ± 1.88	(+) ATP (MT)
S1	15:1	11.9 ± 1.87	(+) ATP (MT)
S1	20:1	15.2 ± 1.50	(+) ATP (MT)
S2	5:1	5.37 ± 1.72	(+) ATP (MT)
S2	10:1	11.2 ± 1.84	(+) ATP (MT)
S2	15:1	18.0 ± 1.81	(+) ATP (MT)
S2	20:1	20.7 ± 0.818	(+) ATP (MT)
S1	5:1	4.07 ± 1.78	(+) ADP (ST)
S1	10:1	8.02 ± 1.90	(+) ADP (ST)

Fiber Type	Density	Nucleosomes / template	Remodeling status
S1	15:1	12.8 ± 1.81	(+) ADP (ST)
S1	20:1	16.1 ± 1.39	(+) ADP (ST)
S2	5:1	5.39 ± 2.13	(+) ADP (ST)
S2	10:1	11.3 ± 2.22	(+) ADP (ST)
S2	15:1	17.4 ± 1.99	(+) ADP (ST)
S2	20:1	20.8 ± 0.965	(+) ADP (ST)
S1	5:1	3.66 ± 1.65	(+) ADP (MT)
S1	10:1	6.91 ± 1.77	(+) ADP (MT)
S1	15:1	11.9 ± 1.88	(+) ADP (MT)
S1	20:1	15.3 ± 1.47	(+) ADP (MT)
S2	5:1	5.46 ± 2.01	(+) ADP (MT)
S2	10:1	11.3 ± 1.84	(+) ADP (MT)
S2	15:1	18.2 ± 1.79	(+) ADP (MT)
S2	20:1	20.7 ± 0.797	(+) ADP (MT)

3.15 Materials & Methods

Cloning M. musculus genomic sites for nucleosome array assembly

Two separate sites within the *M. musculus* reference genome containing CTCF sites were chosen for histone assembly. The CTCF genomic sites will be referred to as Sequence 1 “S1” (chr1:156,887,669-156,890,368, 2712bp) and Sequence 2 “S2” (chr1:156,890,410-156,893,258, 2861 bp). S1 and S2 were PCR amplified (NEBNext® Q5 2X Master Mix) from purified E14 mESC genomic DNA with primers containing homology to a Zeocin-resistance multicutter plasmid backbone as well as dual EcoRV sites for downstream separation of insert from backbone. The plasmid backbone sequence of interest containing homology was prepared with PCR amplification and the remaining parental plasmid was digested away (1 uL DpnI in 1X CutSmart at 37°C for 1 hour). All PCR products were subsequently run out on a 1% Agarose gel and gel purified. After gel purification, standard Gibson Cloning for S1 or S2 inserts plus PCR-amplified/DpnI-digested backbone was performed using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) at 3:1 insert to vector ratio. Transformation was performed with Stellar Competent Cells (Takara) which were thawed on ice. 2 uL of assembly reaction was added to 50 uL competent cells and flicked to mix 4-5 times. The mixture was incubated on ice for 30 minutes, heat shocked at 42°C for 30 seconds, and placed on ice for 2 minutes. 950 uL of SOC media was added to the mixture, and an outgrowth step was performed at 37°C for 1 hour shaking at 1000RPM. The entire mixture was added to pre-warmed Zeocin plates and incubated overnight. Colony PCR was performed to test for insert presence – 8 colonies were selected per site and ran on a 1% Agarose gel. Four colonies containing the insert were selected per sequence and minipreped overnight in Low Salt LB Broth containing Zeocin (25ug/mL). Plasmids were subsequently Sanger Sequenced (Genewiz) to confirm insert sequence and one clone was selected per site for downstream experiments.

Preparation of S1 and S2 arrays via Salt Gradient Dialysis

To assemble nucleosomes onto the sequences of interest, the S1 and S2 plasmids were purified using a GigaPrep kit (Qiagen). To isolate the insert, purified plasmids were restriction enzyme digested (S1: EcoRV, ApaLI, XhoI, BsrBI and S2: EcoRV, BsrBI, BssSal/BssSi-v2, FseI, BstXI, PflFI). Each insert was purified by size exclusion chromatography. Plasmid gigapreps were performed with a dam⁺ E. coli strain, - GATC sequences were ignored for downstream analysis of in vitro experiments. Initial restriction enzyme tests were performed with the plasmids to confirm proper digestion of the backbone, so that the insert could be purified. Xenopus histones were purified according to previously described methods (Luger et al., 1999), and chromatin was assembled using salt gradient dialysis (Lee and Narlikar, 2001), with varying ratios of histone:DNA.

Enzyme remodeling on in vitro oligonucleosome chromatin arrays

S1 or S2 arrays assembled at varying histone:DNA concentrations (50 nM arrays) were remodeled under single-turnover, saturating enzyme conditions (9 μ M SNF2h) or under multiple turnover conditions (10 nM SNF2h). All remodeling reactions were performed in 12.5 mM HEPES pH 7.5, 3 mM MgCl₂, 70 mM KCl, and 0.02% NP-40. Reactions were started with the addition of saturating ATP, ADP (2mM) or no nucleotide and incubated for 15 minutes at room temperature. All reactions were quenched immediately with an equal volume of ADP (34mM) in 1X TE resulting in 25 nM arrays.

SAMOSAs on in vitro oligonucleosome chromatin arrays.

SAMOSAs were performed on remodeled arrays as well as unremodeled arrays and unassembled DNA controls using the non-specific adenine EcoGII methyltransferase (New England Biolabs, high concentration stock 2.5e4 U/mL) as previously described (Abdulhay et al., 2020). For the remodeled arrays,

entire reaction volume was methylated with 31.25 U (1.25uL) of EcoGII. For unremodeled arrays, 1000 nM of input was methylated with 2.5uL EcoGII. For the unassembled, naked S1 and S2 DNA, 3ug input DNA was methylated with 5ul of EcoGII. Methylation reactions were performed in a 100uL reaction containing 1X CutSmart Buffer and 1mM S-adenosyl-methionine (SAM, New England Biolabs) and incubated at 37°C for 30 minutes. SAM was replenished to 3.15 mM after 15 minutes. Unmethylated S1 and S2 naked DNA controls were similarly supplemented with Methylation Reaction buffer, minus EcoGII and replenishing SAM, and the following purification conditions. To purify the remodeled and unremodeled DNA, the samples were subsequently incubated with 10uL Proteinase K (20mg/mL) and 10uL 10% SDS at 65°C for a minimum of 2 hours up to overnight. To extract the DNA, equal parts volume of Phenol-Chloroform-Isoamyl was added and mixed vigorously by shaking and then spun (max speed, 2 min). The aqueous portion was carefully removed and 0.1x volume 3M NaOAc, 3uL of GlycoBlue, and 3x volume of 100% EtOH were added, mixed gently by inversion, and incubated either at -80°C for four hours or overnight at -20°C. Samples were spun (max speed, 4°C, 30 min), washed with 500uL of 70% EtOH, air dried, and resuspended in 50uL EB Buffer. Sample concentration was measured by Qubit High Sensitivity DNA Assay.

Preparation of in vitro SAMOSA PacBio SMRT Libraries

The Purified DNA from array and DNA samples was used in entirety as input for PacBio SMRTbell library preparation as previously described 28. Briefly, preparation of libraries included DNA damage repair, end repair, SMRTbell ligation, and Exonuclease cleanup according to manufacturer's instruction. After Exonuclease cleanup and a double 0.45x Ampure PB Cleanup, sample concentration was measured by Qubit High Sensitivity DNA Assay (1uL each). To assess for library quality, samples (1uL each) were run on the Agilent TapeStation D5000 Assay. Libraries were sequenced on Sequel II 8M SMRTcells in-house. In vitro experiment data were collected over several pooled 30 h Sequel II movie runs with either a 0.6 h or 2 h pre-extension time and either a 2 h or 4 h immobilization time.

Cell lines and cell culture

Published SNF2h knockout and re-expression mouse embryonic stem cells were provided under MTA by the Dirk Schübeler Laboratory at FMI 24. Cells were thawed and grown for at least two passages onto CF-1 Irradiated Mouse Embryonic Feeder cells (Gibco A34181). Feeder cells were depleted from mESCs for at least two passages prior to collection for SAMOSA experiments. E14 mESCs were gifted from Elphege Nora Laboratory at UCSF. All cell lines were mycoplasma tested upon arrival, routinely tested, and confirmed negative with PCR (NEBNext® Q5 2X Master Mix). All feeder and mESC cultures were grown on 0.2% gelatin. mESCs were maintained in KnockOut DMEM 1X (Gibco) supplemented with 10% Fetal Bovine Serum (Phoenix Scientific, Lot# BW-067C18), 1% 100X GlutaMAX (Gibco), 1% 100X MEM Non-Essential Amino Acids (Gibco), 0.128mM 2-Mercaptoethanol (BioRad), and 1X Leukemia Inhibitory Factor (purified and gifted by Barbara Panning Lab at UCSF).

SAMOSA on mESC-derived oligonucleosomes

Isolation of nuclei

Nuclei were collected for the in vivo SAMOSA protocol as previously described (Abdulhay et al., 2020). Briefly, all nuclei were collected per cell line by centrifugation (300xg, 5 min), washed in ice cold 1X PBS, and resuspended in 1 mL Nuclear Isolation Buffer ((20mM HEPES, 10mM KCl, 1mM MgCl₂, 0.1% Triton X-100, 20% Glycerol, and 1X Protease Inhibitor (Roche)) per 5-10e6 cells by gently pipetting 5x with a wide-bore tip to release nuclei. The suspension was incubated on ice for 5 minutes, and nuclei were pelleted (600xg, 4°C, 5 min), washed with Buffer M (15mM Tris-HCl pH 8.0, 15 mM NaCl, 60mM KCl, 0.5mM Spermidine), and spun once again. Nuclei were counted via hemocytometer and either

slow frozen or split for each experimental condition (plus or minus EcoGII methylation). To slow freeze nuclei, nuclei were resuspended in Freeze Buffer (20mM HEPES pH 7.5, 150mM NaCl, 0.5mM spermidine (Sigma), 1X Protease Inhibitor (Roche), 10% DMSO) and stored at -80°C.

Adenine methylation, MNase digest, and overnight dialysis

To proceed to the modified in vivo SAMOSA protocol for direct methylation of nuclei, fresh nuclei were resuspended in Methylation Reaction Buffer (Buffer M containing 1mM SAM). 200uL methylation reactions were performed (10uL EcoGII per 1e6 nuclei) and incubated at 37°C for 30 minutes. SAM was replenished to 6.25mM after 15 minutes. Unmethylated controls were similarly supplemented with Buffer M + SAM, minus EcoGII and replenishing SAM. Samples were spun (600xg, 4°C, 5 min) and resuspended in cold MNase digestion Buffer (Buffer M containing 1mM CaCl₂). MNase digestion of nuclei was performed in 200uL reactions and 0.02 Units of MNase was added per 1e6 nuclei (Sigma, micrococcal nuclease from *Staphylococcus aureus*) at 4°C for either 45 minutes or 1 hour. EGTA was added to 2mM to stop the digestion and incubated on ice. For nuclear lysis and liberation of chromatin fibers, MNase-digested nuclei were collected (600xg, 4°C, 5 min) and resuspended in ~250uL of Tep20 Buffer (10mM Tris-HCl pH 7.5, 0.1mM EGTA, 20mM NaCl and 1X Protease Inhibitor (Roche) added immediately before use) supplemented with 300ug/mL of Lysolethicin (Sigma, L- α -Lysophosphatidylcholine from bovine brain) and rotated overnight at 4°C. Dialyzed samples were spun to remove nuclear debris (12,000xg, 4°C, 5 minutes) and soluble chromatin fibers in the supernatant were collected. Sample concentration was measured by Nanodrop and chromatin fibers were analyzed by standard agarose gel electrophoresis.

To generate a naked DNA positive control for downstream analysis, gDNA was extracted from E14 mESCs with Lysis Buffer (10mM Tris-Cl pH 8.0, 100mM NaCl, 25mM EDTA pH 8.0, 0.5% SDS,

0.1mg/mL Proteinase K) and purified with the following conditions. Methylation reactions were performed as previously stated, with 3ug DNA as input and 5uL EcoGII (125U), followed by a second purification as follows. To purify all DNA samples, reactions were incubated with 10uL of RNase A at room temperature for 10 minutes, followed by 10uL Proteinase K (20mg/mL) and 10uL 10% SDS at 65°C for a minimum of 2 hours up to overnight. To extract the DNA, equal parts volume of Phenol-Chloroform was added and mixed vigorously by shaking, and spun (max speed, 2 min). The aqueous portion was carefully removed and 0.1x volumes of 3M NaOAc, 3uL of GlycoBlue and 3x volumes of 100% EtOH were added, mixed gently by inversion, and incubated overnight at -20°C. Samples were then spun (max speed, 4°C, 30 min), washed with 500uL 70% EtOH, air dried and resuspended in 50uL EB. Sample concentration was measured by Qubit High Sensitivity DNA Assay.

Preparation of in vivo SAMOSA PacBio SMRT Libraries

Purified DNA from mESCs (methylated, unmethylated, naked DNA positive controls) was used to prepare PacBio SMRT libraries using either the SMRTbell Express Template Prep Kit 1.0 (blunt end ligation) or 2.0 (A/T overhang ligation). For the SNF2h KO and SNF2h WT AB mESC purified SAMOSA samples, a minimum of 500ng up to 1.5ug was utilized as input with SMRTbell Express Template Prep Kit 1.0. For the E14 mESCs, a minimum of ~400ng up to 1.7ug was utilized as input with the SMRTbell Express Template Prep Kit 2.0. The naked DNA E14 positive control was sheared with a Covaris G-Tube (5424 Rotor, 3381xg for 1 minute) and sheared to approximately 10,000 bp. Sample size distribution was checked with the Agilent Bioanalyzer DNA chip. The entire sample was utilized as input for library preparation with the PacBio SMRTbell Express Template Prep Kit 2.0. Briefly, all library preparations included DNA damage repair, end repair, SMRTbell ligation with either blunt or overhang unique adapters, and Exonuclease cleanup according to manufacturer's instructions. Unique PacBio SMRTbell adapters (100uM stock) were annealed to 20uM in annealing buffer (10mM Tris-HCl pH 7.5 and 100mM NaCl) in a thermocycler (95°C 5 min, RT 30 mins, 4°C hold) and stored at -20°C for long-term storage. After

exonuclease cleanup and Ampure PB cleanups (0.45X for 1.0 preparation or 1X for 2.0 preparation), the sample concentrations were measured by Qubit High Sensitivity DNA Assay (1uL each). To assess for size distribution and library quality, samples (1 uL each) were run on an Agilent Bioanalyzer DNA chip. Libraries were sequenced in house on Sequel II 8M SMRTcells. In vivo data were collected over several pooled 30 h Sequel II movie runs with either a 0.6 h or 2 h pre-extension time and either a 2 h or 4 h immobilization time.

SMRT data processing

We applied our method to two use cases in the paper, and they differ in the computational workflow to analyze them. The first is for sequencing samples where every DNA molecule has the same sequence, which is the case for our remodeling experiments on the S1 and S2 sequences, presented in Figures 1-5. The second use case is for samples from cells containing varied sequences of DNA molecules, such as the murine in vivo follow up experiments presented in Figures 6-7. The first will be referred to as homogeneous samples, and the second as genomic samples. The workflow for homogenous samples will be presented first in each section, and the deviations for genomic samples detailed at the end.

Sequencing read processing

Sequencing reads were processed using software from Pacific Biosciences. The following describes the workflow for homogenous samples:

Demultiplex reads

Reads were demultiplexed using lima. The flag `--same` was passed as libraries were generated with the same barcode on both ends. This produces a BAM file for the subreads of each sample.

Generate Circular Consensus Sequences (CCS)

CCS were generated for each sample using `ccs`. Default parameters were used other than setting the number of threads with `-j``. This produces a BAM file of CCS.

Align subreads to the reference genome

`pbmm2`, the pacbio wrapper for `minimap2` 71, was run on each subreads BAM file (the output of step 1) to align subreads to the reference sequence, producing a BAM file of aligned subreads.

Generate missing indices

Our analysis code requires pacbio index files (`.pbi`) for each BAM file. `pbmm2`` does not generate index files, so missing indices were generated using `pbindex``.

For genomic samples, replace step 3 with this alternate step 3

Align CCS to the reference genome

Alignment was done using `pbmm2`, and run on each CCS file, resulting in BAM files containing the CCS and alignment information.

Extracting IPD measurements

The IPD values were accessed from the BAM files and \log_{10} transformed after setting any IPD measurements of 0 frames to 1 frame. Then, for each ZMW, at each base in the CCS (for genomic samples) or amplicon reference (for homogenous samples), for both strands, the log transformed IPD values in all subreads were averaged. These mean log IPD values for the molecule were then exported along with the percentiles of log IPD values across subreads within that molecule.

Predicting methylation status of individual adenines

Predicting methylation in homogenous samples

For homogenous samples dimensionality reduction was used to capture variation in IPD measurements between molecules, and then the reduced representations and IPD measurements were used to predict methylation. For each of S1 and S2, the non-adenine mean log IPD measurements from one unmethylated control sample were used to train a truncated singular value decomposition model. The input measurements had the mean of each base subtracted before training. The Truncated SVD class of scikit-learn was used and trained in 20 iterations to produce 40 components. The trained model was then used to transform all molecules in all samples into their reduced representations. Each resulting component had its mean subtracted and was divided by its standard deviation.

Next, a neural network model was trained to predict the mean log IPD at each base in unmethylated control molecules. The dimension reduced representation of the molecules were provided as input to the model, and the output was a value for each adenine on both strands of the amplicon molecule. The neural network was composed of four dense layers with 600 units each, with relu activation and he uniform initialization. A 50% dropout layer was placed after each of these four layers. A final dense layer produced an output for each adenine in the amplicon reference. The model was trained on a negative control sample using Keras, Adam optimizer, mean square error loss, 100 epochs and a batch size of 128. The trained model was then used to predict the mean log IPD value at all adenines in all molecules in all samples. This prediction was subtracted from the measured mean log IPD to get residuals.

A large positive residual represents slower polymerase kinetics at that adenine than would be expected given the sequence context and molecule and is thus evidence of methylation. To find a cutoff of how large the residual should be to be called as methylated, we assembled a dataset of residuals from an equal proportion of molecules from a fully methylated naked DNA control and an unmethylated control.

For each individual adenine a student's t-distribution mixture model was fit to the residuals using the python package *smm* (Peel and McLachlan, 2000). A two-component model was fit with a tolerance of $1e-6$, and a cutoff was found where that residual value was equally likely to originate from either of the two components. Adenines were then filtered by whether a sufficiently informative cutoff had been found. The three criteria for using the methylation predictions at that adenine in further analysis were: 1) The mean of at least one t-distribution had to be above zero, 2) The difference between the means of the two t-distributions had to be at least X , where X was chosen separately for each amplicon reference but varied from 0.1 to 0.3, and 3) At least 2% of the training data was over the cutoff. These were lenient cutoffs that allowed the methylation predictions at 90+% of adenines to be included in downstream analysis. This was done because the next HMM step accounts for the frequency of methylation predictions in unmethylated and fully methylated control samples, and thus adenine bases where methylation prediction was poor will be less informative of DNA accessibility.

Predicting methylation in genomic samples

Methylation prediction was made in a similar fashion for genomic samples, with deviations necessitated by the differences in the data. Unlike in homogenous samples, dimensionality reduction could not be used to capture inter-molecular variation due to varying DNA sequences. Instead IPD percentiles were used as neural network inputs. As described above in Extracting IPD measurements, log IPD percentiles were calculated across all subreads in each molecule separately for each template base. Every 10th percentile from 10th to 90th inclusive, for template bases C, G, and T, were used as neural network input. The other input was the DNA sequence context around the measured base, given for three bases 5' of the template adenine and ten bases 3' of the template adenine, one-hot encoded. The neural network was a regression model predicting the measured mean log IPD at that template adenine. The neural network consisted of four dense layers with 200 units each, relu activation, and he uniform initialization. The training data was 5,000,000 adenines each from six different unmethylated control samples. The validation data for early

stopping was 5,000,000 adenines from each of two more unmethylated control samples. The model was trained using Keras, Adam optimizer, 20 epochs with early stopping (patience of 2 epochs), and a batch size of 128.

To determine at which adenines the methylation prediction was usefully informative and accurate, we used a second neural network model to predict the IPD residual in a positive control sample from sequence context. Sequence contexts that consistently produced residuals near zero in a positive control would be likely never methylated by EcoGII, or always methylated endogenously. The input to this network was the one-hot encoded sequence context as described above. The output was the measured log IPD with predicted log IPD subtracted. The training data was a fully methylated naked DNA sample of E14. Mean log IPD residuals were calculated using the above trained model. 20,000,000 adenines were used as training data and 10,000,000 as validation data. The neural network consisted of three dense layers of 100 units, relu activation, and he uniform initialization. The model was trained using Adam optimizer for two epochs with a batch size of 128. After examining the output of the trained model on negative and positive controls and chromatin, we settled on a cutoff of 0.6 for the predicted residual in positive control for calling a sequence context as usable for downstream analysis, and a cutoff of 0.42 for the mean log IPD residual for calling an adenine as methylated.

Predicting molecule-wide DNA accessibility using Hidden Markov Models

Predicting DNA accessibility in homogeneous samples

To go beyond individual methylation predictions and predict DNA accessibility along each molecule we applied a Hidden Markov Model (HMM). An HMM model was constructed for each amplicon reference, with two states for every adenine at which methylation was predicted: one state representing that adenine being inaccessible to the methyltransferase, and another representing it being accessible. The emission

probabilities were all Bernoulli distributions, with the probability of observing a methylation in an inaccessible state being the fraction of unmethylated control molecules predicted to be methylated at that adenine, and the probability of observing a methylation in an accessible state being the fraction of fully methylated naked DNA control molecules predicted to be methylated at that adenine. 0.5 was added to the numerator and denominator of all fractions to avoid any probabilities of zero. An initial state was created with an equal probability of transitioning into either accessible or inaccessible states. Transition probabilities between adenines were set using the logic that for an expected average duration in a single state of L , by the geometric distribution at each base the probability of switching states at the next base will be $\frac{1}{L}$. The probability of staying in the same state from one adenine to the next is thus $(1 - \frac{1}{L})^B$, where B is the distance in bases between adenines. The probability of switching to the other state at the next adenine is then 1 minus that value. Different values of the average duration L were tested, and ultimately a value of 1000 bp was used. This is much higher than expected, but has the beneficial result of requiring a higher burden of evidence to motivate switching states and thus minimizes spurious switching.

With the HMM model constructed, the most likely state path was found using the Viterbi algorithm for all molecules in all samples, with the predicted methylation at each adenine provided as the input. Models were constructed and solved using pomegranate (Schreiber, 2018). The solved path was output as an array with accessible adenines as 1, inaccessible as 0, and non-adenine and uncalled bases interpolated.

Predicting DNA accessibility in genomic samples

In genomic samples DNA accessibility was predicted in a similar fashion to homogenous, except that the HMM model had to be individually constructed for each molecule due to varying DNA sequences, and rather than empirically measuring the fraction of methylation in positive and control samples at each position, neural networks were trained to predict the fraction of methylation in each from sequence context.

A neural network model was trained to predict the predicted methylation status of adenines in the positive control sample based on sequence context. The output from this model was used to approximate the probability of an adenine in that sequence context getting predicted as methylated if it was accessible to EcoGII. The sample used for training was the same naked DNA E14 methylated sample used to train the positive residual prediction model. Approximately 27,600,000 adenines were used as the training set and 7,000,000 as the validation set. The input was the one-hot encoded sequence context. The neural network consisted of three dense layers of 200 units, relu activation and he uniform initialization. The training output were binary methylation predictions, so the final output of the network had a sigmoid activation and binary cross-entropy was used as the loss. The model was trained with Adam optimizer for seven epochs with the batch size increasing each epoch from 256 to a max of 131,072.

An identical network was trained to predict the predicted methylation status of adenines in the unmethylated negative control samples. The output from this model was used to approximate the probability of an adenine in that sequence context getting predicted as methylated if it was not accessible to EcoGII. This one was trained using adenines combined from four different unmethylated samples, and approximately 28,100,000 adenines were used as the training set and 7,100,000 as the validation set.

The HMM models were constructed in an identical manner to that described above for homogenous samples, except for genomic data an HMM model was constructed for each sequenced molecule individually. States and transition probabilities and observed output were the same. The emission probability of observing methylation at each accessible state was the output of the trained positive control methylation prediction model, and for inaccessible states was the output of the trained negative control methylation prediction model. As with homogenous samples, the HMM was solved using the observed methylation and the Viterbi algorithm.

Defining inaccessible regions and counting nucleosomes

Inaccessible regions were defined from the HMM output data as continuous stretches with accessibility ≤ 0.5 . To estimate the number of nucleosomes contained within each inaccessible region, a histogram of inaccessible region lengths was generated for each data type (sequence S1, S2, and murine in vivo). Periodic peaks in these histograms were observed that approximated expected sizes for stretches containing one, two, three, etc. nucleosomes. Cutoffs for the different categories were manually defined using the histogram, including a lower cutoff for sub-nucleosomal regions (Supplementary Figure 12).

Processed data analysis

All processed data analyses and associated scripts will be made available at <https://github.com/RamaniLab/SAMOS-CHAAAT>. Most processed data analyses proceeded from data tables generated using custom python scripts. Resulting data tables were then used to compute all statistics reported in the paper and perform all visualizations (using tidyverse and ggplot2 in R). Below, we describe each analysis in text form, while noting that all code is freely available at the above link.

UMAP and Leiden clustering analyses

All UMAP and Leiden clustering analyses were performed using the scanpy package (Wolf et al., 2018). All UMAP visualizations 31 were made using default parameters in scanpy. Leiden clustering 32 was performed using resolution = 0.4; clusters were then filtered on the basis of size such that all clusters that collectively summed up to $< 5\%$ of the total dataset were removed. In practice, this served to remove long tails of very small clusters defined by the Leiden algorithm.

Signal correlation analyses

We converted footprint data files into a vector of footprint midpoint abundance for sequences S1 and S2 by summing footprint midpoint occurrences and normalizing against the total number of footprints. We then correlated these vectors across replicate experiments using R for both correlation calculations and plotting associated scatterplots.

Trinucleosome analyses

Using processed footprint midpoint data files, we examined, for each footprinted fiber, the distances between all consecutive footprints sized between 100 and 200 bp, and plotted these distances against each other. All calculations were made on processed data tables generated using scripts described in the associated Jupyter notebook.

Autocorrelation analyses

Autocorrelations for in vitro and in vivo data were calculated using python, and then clustered as described above. All scripts for computing autocorrelation are available at the above link.

CTCF motif analyses

We examined the relative accessibility of 17 nucleotide windows tiling sequences S1 and S2 for each footprinted molecule before (native) and after (remodeled) remodeling, summarizing accessibility as a binary value thresholded on whether $> 0.9 * \text{the window length}$ was accessible on a single molecule. We then stored these values in a data frame, and plotted the relative fractions of accessible windows against each other as log-odds values.

In vivo chromatin fiber analyses

All autocorrelation and clustering analyses were done as previously performed (Abdulhay et al., 2020b). Autocorrelation and clustering were performed above. Nucleosome density enrichment plots were generated by estimating probability distributions for background (all molecules) and cluster-specific (clustered molecules) molecules, and computing log-odds from these distributions. All per-fiber nucleosome density measurements were calculated as above. Fisher's Exact enrichment tests were carried out using *scipy* in Python as in Abdulhay et al (2020). All p-values calculated were then corrected using a Storey q-value correction, using the *qvalue* package in R (Storey and Tibshirani, 2003). Multiple hypothesis correction was performed for all domain-level Fisher's tests (including ATAC peak analyses) and cutoffs were made at $q < 0.05$.

Molecules falling within ENCODE-defined epigenomic domains were extracted using scripts published in Abdulhay et al (2020).

ATAC data reanalysis

SNF2hKO and WT ATAC-seq data (Barisic et al., 2019b) were downloaded, remapped to mm10 using *bwa*, converted to sorted, deduplicated BAM files, and then processed using *macs2* to define accessibility peaks. Peaks were then filtered for reproducibility using the ENCODE IDR framework, and reproducible peaks were preserved for downstream analyses. Reproducible peaks for SNF2hKO and WT samples were pooled and merged using *bedtools merge*, and then used to generate count matrices using *bedtools bamcoverage*. Resulting count matrices for replicate experiments were then fed into *DESeq2* to define statistically significant differentially accessible peaks with an adjusted p-value cutoff of 0.05.

In vivo TF binding analyses

Molecules containing previously-defined (Ramani et al., 2019) ENCODE-backed Ctf bound motifs were extracted using scripts previously published (Abdulhay et al., 2020b). Control molecules were obtained by randomly sampling a resource of Ctf motif matches (provided by: (Vierstra et al., 2020)). Single-molecule signals centered at the predicted motif center were stored as an array and clustered as above to obtain the clusters shown in the text. Enrichment analyses and associated multiple-hypothesis correction were performed as above for all enrichment tests performed for this array of Ctf sites.

Satellite sequence analyses

Detecting mouse minor (centromeric) and major (pericentromeric) satellite is challenging because of the similarity of these two sequences (including internal / self-similarity). The latter is also an issue with the telomere repeat. To use BLAST to find matches to these sequences, the output must be processed to remove overlapping matches, which is done here heuristically using an implementation of the weighted interval scheduling dynamic programming algorithm that seeks to optimize the summed bitscores for non-overlapping matches to all three sequences (minor satellite, major satellite, and telomeres). This is not a perfect solution to the problem, in part because it treats the alignment for the three different repeats as effectively equivalent and we do not believe the alignments produced by BLAST are optimal compared to e.g. Smith-Waterman alignment, and the attendant fuzziness introduced may lead to removal of a small fraction of bona fide matches.

Given the similarity of major and minor satellite sequences in particular, using the DFAM minor (SYNREP_MM, accession DF0004122.1) and major (GSAT_MM, accession: DF0003028.1) satellite consensus sequences, which both exceed well-established monomer lengths of ~120 bp (minor) and ~234 bp (major), produces too many overlapping hits. Thus, we used more representative sequences from

Genbank, specifically M32564.1 for major satellite, and X14462.1 for minor satellite. The telomere repeat sequence was constructed by pentamerizing the telomere repeat (i.e. [TTAGGG] x 5). All code used for these analyses is deposited at the above GitHub link.

Data Availability

All processed data will be made available at Zenodo (<https://doi.org/10.5281/zenodo.5770727>). Raw data and a portion of the processed data will be uploaded to GEO at GEO accession GSE197979. All scripts and notebooks used for data analysis in this study will be made available at <https://github.com/RamaniLab/SAMOSA-ChAAT>.

REFERENCES

Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020a). Massively multiplex single-molecule oligonucleosome footprinting. *Elife* 9.

Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020b). Massively multiplex single-molecule oligonucleosome footprinting. *Elife* 9, 1–23.

Ahmad, K., and Henikoff, S. (2001). Modulation of a Transcription Factor Counteracts Heterochromatic Gene Silencing in *Drosophila*. *Cell* 104, 839–847.

Ahmad, K., and Henikoff, S. (2002). The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Mol. Cell* 9, 1191–1200.

Alfert, A., Moreno, N., and Kerl, K. (2019). The BAF complex in development and disease. *Epigenetics Chromatin* 2019 121 12, 1–15.

Altemose, N., Logsdon, G.A., Bzikadze, A. V., Sidhwani, P., Langley, S.A., Caldas, G. V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., et al. (2022). Complete genomic and epigenetic maps of human centromeres. *Science* (80-.). 376.

Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., and Roberts, C.W.M. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* 2017 81 8, 1–10.

Angelov, D., Vitolo, J.M., Mutskov, V., Dimitrov, S., and Hayes, J.J. (2001). Preferential interaction of the core histone tail domains with linker DNA. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6599.

Arents, G., and Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11170.

Arimura, Y., Tachiwana, H., Oda, T., Sato, M., and Kurumizaka, H. (2012). Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry* *51*, 3302–3309.

Armache, J.P., Gamarra, N., Johnson, S.L., Leonard, J.D., Wu, S., Narlikar, G.J., and Cheng, Y. (2019). Cryo-EM structures of remodeler-nucleosome intermediates suggest allosteric control through the nucleosome. *Elife* *8*.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* (80-.). *339*, 1074–1077.

Badenhorst, P., Voas, M., Rebay, I., and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.* *16*, 3186–3198.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* *37*, W202–W208.

Baldi, S., Krebs, S., Blum, H., and Becker, P.B. (2018). Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. *Nat. Struct. Mol. Biol.* *2018 259 25*, 894–901.

Baldi, S., Korber, P., and Becker, P.B. (2020). Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. *Nat. Struct. Mol. Biol.* *2020 272 27*, 109–118.

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

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019a). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* *569*, 136–140.

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019b). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* *569*, 136–140.

- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129, 823–837.
- Barutcu, A.R., Lajoie, B.R., Fritz, A.J., McCord, R.P., Nickerson, J.A., Van Wijnen, A.J., Lian, J.B., Stein, J.L., Dekker, J., Stein, G.S., et al. (2016). SMARCA4 regulates gene expression and higher-order chromatin structure in proliferating mammary epithelial cells. *Genome Res.* 26, 1188–1201.
- Becker, P.B., Gloss, B., Schmid, W., Strähle, U., and Schütz, G. (1986). In vivo protein–DNA interactions in a glucocorticoid response element require the presence of the hormone. *Nat.* 1986 3246098 324, 686–688.
- Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., Davies, H., et al. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat. Genet.* 45, 1479–1482.
- Blosser, T.R., Yang, J.G., Stone, M.D., Narlikar, G.J., and Zhuang, X. (2009). Dynamics of nucleosome remodelling by individual ACF complexes. *Nature* 462, 1022–1027.
- Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422.
- Bracken, A.P., Brien, G.L., and Verrijzer, C.P. (2019). Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev.* 33, 936–959.
- Brahma, S., and Henikoff, S. (2019). RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. *Mol. Cell* 73, 238-249.e3.
- Brahma, S., and Henikoff, S. (2020). Epigenome Regulation by Dynamic Nucleosome Unwrapping. *Trends Biochem. Sci.* 45, 13–26.
- Cai, S., Song, Y., Chen, C., Shi, J., and Gan, L. (2018). Natural chromatin is heterogeneous and self-

associates in vitro. *Mol. Biol. Cell* 29, 1652–1663.

Carey, M., Li, B., and Workman, J.L. (2006). RSC Exploits Histone Acetylation to Abrogate the Nucleosomal Block to RNA Polymerase II Elongation. *Mol. Cell* 24, 481–487.

Castel, D., Philippe, C., Calmon, R., Le Dret, L., Truffaux, N., Boddaert, N., Pagès, M., Taylor, K.R., Saulnier, P., Lacroix, L., et al. (2015). Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. *Acta Neuropathol.* 130, 815–827.

Chan, K.M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., David James, C., Jenkins, R., et al. (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev.* 27, 985–990.

Chioda, M., Vengadasalam, S., Kremmer, E., Eberharter, A., and Becker, P.B. (2010). Developmental role for ACF1-containing nucleosome remodellers in chromatin organisation. *Development* 137, 3513–3522.

Clapier, C.R., and Cairns, B.R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* 492, 280–284.

Clapier, C.R., Längst, G., Corona, D.F. V., Becker, P.B., and Nightingale, K.P. (2001). Critical Role for the Histone H4 N Terminus in Nucleosome Remodeling by ISWI. *Mol. Cell. Biol.* 21, 875–883.

Collins, N., Poot, R.A., Kukimoto, I., García-Jiménez, C., Delleire, G., and Varga-Weisz, P.D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat. Genet.* 32, 627–632.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat. Rev. Cancer* 16, 803–810.

Correll, S.J., Schubert, M.H., and Grigoryev, S.A. (2012). Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J.* 31, 2416–2426.

Cuaycong, M.H., and Elgin, S.C.R. (2001). Long-Range Nucleosome Ordering Is Associated with Gene Silencing in. *Mol. Cell. Biol.* 21, 2867–2879.

Cutter, A.R., and Hayes, J.J. (2015). A Brief Review of Nucleosome Structure. *FEBS Lett.* 589, 2914.

Dann, G.P., Liszczak, G.P., Bagert, J.D., Müller, M.M., Nguyen, U.T.T., Wojcik, F., Brown, Z.Z., Bos, J., Panchenko, T., Pihl, R., et al. (2017). ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature* 548, 607–611.

Davidson, I.F., and Peters, J.M. (2021). Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* 2021 227 22, 445–464.

Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150, 12–27.

Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladde, M.P., and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. *Mol. Cell* 38, 590–602.

Deindl, S., Hwang, W.L., Hota, S.K., Blosser, T.R., Prasad, P., Bartholomew, B., and Zhuang, X. (2013). ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell* 152, 442–452.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26, 5433–5438.

De Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Dargham, D.B., Jounier, S., Humbertclaude, H., Ribierre, F., Baulard, C., et al. (2016). Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* 530, 113–116.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet.* 34, 666–681.

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

Donehower, L.A., Huang, A.L., and Hager, G.L. (1981). Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. *J. Virol.* 37, 226–238.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* (80-.). 306, 1571–1573.

Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* 186, 773–790.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21, 129–228.

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

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

Ebbert, M.T.W., Farrugia, S.L., Sens, J.P., Jansen-West, K., Gendron, T.F., Prudencio, M., McLaughlin, I.J., Bowman, B., Seetin, M., Dejesus-Hernandez, M., et al. (2018). Long-read sequencing across the C9orf72 “GGGGCC” repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Mol. Neurodegener.* 13, 1–17.

Eberharter, A., Ferrari, S., Längst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., and Becker, P.B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO J.* 20, 3781–3788.

Ehrensberger, A.H., Franchini, D.M., East, P., George, R., Matthews, N., Maslen, S.L., and Svejstrup, J.Q. (2015). Retention of the Native Epigenome in Purified Mammalian Chromatin. *PLoS One* *10*, e0133246.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* *323*, 133–138.

Erdel, F., and Rippe, K. (2011). Chromatin remodelling in mammalian cells by ISWI-type complexes—where, when and why? *Authors J. Compil.* *a* *278*, 3608–3618.

Eustermann, S., Schall, K., Kostrewa, Di., Lakomek, K., Strauss, M., Moldt, M., and Hopfner, K.P. (2018). Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature* *556*, 386–390.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* *30*, 1232–1239.

Feng, Z., Fang, G., Korfach, J., Clark, T., Luong, K., Zhang, X., Wong, W., and Schadt, E. (2013). Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic. *PLOS Comput. Biol.* *9*, e1002935.

Ferraro, A. (2016). Altered primary chromatin structures and their implications in cancer development. *Cell. Oncol.* *39*, 195–210.

Filion, G.J., van Bommel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell* *143*, 212–224.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* *73*, 1897–1901.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. *Nature* *269*, 29–36.

Flaus, A. (2012). Principles and practice of nucleosome positioning in vitro.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010a). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010b). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 2010 76 7, 461–465.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol. Cell* 8, 545–556.

Fyodorov, D. V., Blower, M.D., Karpen, G.H., and Kadonaga, J.T. (2004). Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes Dev.* 18, 170–183.

Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of Nucleosome Positioning in the Human Genome. *PLOS Genet.* 8, e1003036.

Gamarra, N., Johnson, S.L., Trnka, M.J., Burlingame, A.L., and Narlikar, G.J. (2018). The nucleosomal acidic patch relieves auto-inhibition by the ISWI remodeler SNF2h. *Elife* 7.

Garcia, J.F., Dumesic, P.A., Hartley, P.D., El-Samad, H., and Madhani, H.D. (2010). Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. *Genes Dev.* 24, 1758–1771.

Gershman, A., Sauria, M.E.G., Guitart, X., Vollger, M.R., Hook, P.W., Hoyt, S.J., Jain, M., Shumate, A., Razaghi, R., Koren, S., et al. (2022). Epigenetic patterns in a complete human genome. *Science* (80-.). 376.

Ghandi, M., Mohammad-Noori, M., Ghareghani, N., Lee, D., Garraway, L., and Beer, M.A. (2016).

gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* 32, 2205–2207.

Gibson, B.A., Doolittle, L.K., Schneider, M.W.G., Jensen, L.E., Gamarra, N., Henry, L., Gerlich, D.W., Redding, S., and Rosen, M.K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* 179, 470–484.e21.

Gilbert, N., and Allan, J. (2001). Distinctive higher-order chromatin structure at mammalian centromeres. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11949–11954.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin Architecture of the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. *Cell* 118, 555–566.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* 140, 678–691.

Golemis, E.A., Scheet, P., Beck, T.N., Scolnick, E.M., Hunter, D.J., Hawk, E., and Hopkins, N. (2018). Molecular mechanisms of the preventable causes of cancer in the United States. *Genes Dev.* 32, 868–902.

Gonzalez-Perez, A., Jene-Sanz, A., and Lopez-Bigas, N. (2013). The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* 14, 1.

Grüne, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F.V., Becker, P.B., and Müller, C.W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol. Cell* 12, 449–460.

Hamiche, A., Sandaltzopoulos, R., Gdula, D.A., and Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* 97, 833–842.

Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. *Cell* 100, 57–70.

Harshman, S.W., Young, N.L., Parthun, M.R., and Freitas, M.A. (2013). H1 histones: current perspectives and challenges. *Nucleic Acids Res.* *41*, 9593.

Harutyunyan, A.S., Krug, B., Chen, H., Papillon-Cavanagh, S., Zeinieh, M., De Jay, N., Deshmukh, S., Chen, C.C.L., Belle, J., Mikael, L.G., et al. (2019). H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nat. Commun.* *10*.

Hayden, E.C. (2014). The \$ 1,000 genome. *Nature* *507*, 294–295.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* *21*, 133–153.

Henikoff, S., and Smith, M.M. (2015). Histone Variants and Epigenetics. *Cold Spring Harb. Perspect. Biol.* *7*.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 18318–18323.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* *2009* *6*, 283–289.

Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* *52*, 504–510.

Hoose Scott A. and Kladdde, M.P. (2006). DNA Methyltransferase Probing of DNA-Protein Interactions. In *Gene Mapping, Discovery, and Expression: Methods and Protocols*, M. Bina, ed. (Totowa, NJ: Humana Press), pp. 225–244.

Hota, S.K., Bhardwaj, S.K., Deindl, S., Lin, Y.C., Zhuang, X., and Bartholomew, B. (2013). Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat. Struct. Mol. Biol.* *20*, 222–229.

Hoyt, S.J., Storer, J.M., Hartley, G.A., Grady, P.G.S., Gershman, A., Lima, L.G. de, Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., et al. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* (80-.). 376.

Huang, B., Babcock, H., and Zhuang, X. (2010). Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *Cell* 143, 1047–1058.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., and Wheeler, T.J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Res.* 44, D81–D89.

Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* 90, 145–155.

Ito, T., Levenstein, M.E., Fyodorov, D. V., Kutach, A.K., Kobayashi, R., and Kadonaga, J.T. (1999). ACF consists of two subunits, Acfl and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.* 13, 1529–1539.

Iwasaki, W., Miya, Y., Horikoshi, N., Osakabe, A., Taguchi, H., Tachiwana, H., Shibata, T., Kagawa, W., and Kurumizaka, H. (2013). Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* 3, 363.

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

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nat.* 2015 5277578 527, 384–388.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* 1.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013).

Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* *45*, 592–601.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M.G., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* *49*, 213–222.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* *458*, 362–366.

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

Kenzaki, H., and Takada, S. (2015). Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLOS Comput. Biol.* *11*, e1004443.

Khurana, S., and Oberdoerffer, P. (2015). Replication Stress: A Lifetime of Epigenetic Change. *Genes* *2015*, Vol. 6, Pages 858-877 *6*, 858–877.

Kim, J.H., Ebersole, T., Kouprina, N., Noskov, V.N., Ohzeki, J.I., Masumoto, H., Mravinac, B., Sullivan, B.A., Pavlicek, A., Dovat, S., et al. (2009). Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res.* *19*, 533–544.

Kim, J.M., Visanpattanasin, P., Jou, V., Liu, S., Tang, X., Zheng, Q., Li, K.Y., Snedeker, J., Lavis, L.D., Lionnet, T., et al. (2021). Single-molecule imaging of chromatin remodelers reveals role of atpase in promoting fast kinetics of target search and dissociation from chromatin. *Elife* *10*.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome Remodeling Induced by RNA Polymerase II: Loss of the H2A/H2B Dimer during

Transcription. *Mol. Cell* 9, 541–552.

Klein, D.C., and Hainer, S.J. (2020). Genomic methods in profiling DNA accessibility and factor localization. *Chromosom. Res.* 28, 69.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220.

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

Kornberg, R.D., and Stryer, L. (1988). Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* 16, 6677–6690.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., and Burger, L. (2017a). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol. Cell* 67, 411–422.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017b). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol. Cell* 67, 411–422.e4.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* 167, 709–721.e12.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* 18, 548.

Lai, B., Gao, W., Cui, K., Xie, W., Tang, Q., Jin, W., Hu, G., Ni, B., and Zhao, K. (2018). Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nat.* 562, 281–285.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409,

860–921.

Längst, G., and Becker, P.B. (2001). Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci.* *114*, 2561–2568.

Lee, K.M., and Narlikar, G. (2001). Assembly of Nucleosomal Templates by Salt Dialysis. *Curr. Protoc. Mol. Biol.* *54*, 21.6.1-21.6.16.

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Sadowski, N., Simpson, J.T., Sedlazeck, F.J., and Timp, W. (2019). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *BioRxiv* 504993.

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Gershman, A., Sadowski, N., Sedlazeck, F.J., Hansen, K.D., Simpson, J.T., and Timp, W. (2020). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* *2020* 1712 *17*, 1191–1199.

Leonard, J.D., and Narlikar, G.J. (2015). A Nucleotide-Driven Switch Regulates Flanking DNA Length Sensing by a Dimeric Chromatin Remodeler. *Mol. Cell* *57*, 850–859.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* (80-.). *340*, 857–861.

Li, B., Carey, M., and Workman, J.L. (2007a). The Role of Chromatin during Transcription. *Cell* *128*, 707–719.

Li, B., Carey, M., and Workman, J.L. (2007b). The role of chromatin during transcription. *Cell* *128*, 707–719.

Li, Y., Gong, H., Wang, P., Zhu, Y., Peng, H., Cui, Y., Li, H., Liu, J., and Wang, Z. (2021). The emerging role of ISWI chromatin remodeling complexes in cancer. *J. Exp. Clin. Cancer Res.* *2021* 401 *40*, 1–27.

Li, Z., Zhao, P., and Xia, Q. (2019). Epigenetic Methylations on N6-Adenine and N6-Adenosine with the same Input but Different Output. *Int. J. Mol. Sci.* *20*.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015a). Nucleosome Spacing Generated by ISWI and CHD1 Remodelers Is Constant Regardless of Nucleosome Density. *Mol. Cell. Biol.* *35*, 1588–1605.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015b). Nucleosome Spacing Generated by ISWI and CHD1 Remodelers Is Constant Regardless of Nucleosome Density. *Mol. Cell. Biol.* *35*, 1588–1605.

Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., and Song, C.X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* *37*, 424–429.

Long, M., Sun, X., Shi, W., Yanru, A., Leung, S.T.C., Ding, D., Cheema, M.S., MacPherson, N., Nelson, C.J., Ausio, J., et al. (2019). A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res.* *47*, 8399–8409.

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

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

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nat.* *1997* 3896648 *389*, 251–260.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* *304*, 3–19.

Maeshima, K., Hihara, S., and Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo?

Curr. Opin. Cell Biol. 22, 291–297.

Mashtalir, N., Dao, H.T., Sankar, A., Liu, H., Corin, A.J., Bagert, J.D., Ge, E.J., D’Avino, A.R., Filipovski, M., Michel, B.C., et al. (2021). Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes. *Science* (80-.). 373, 306–315.

McBride, M.J., Mashtalir, N., Winter, E.B., Dao, H.T., Filipovski, M., D’Avino, A.R., Seo, H.S., Umbreit, N.T., St. Pierre, R., Valencia, A.M., et al. (2020). The nucleosome acidic patch and H2A ubiquitination underlie mSWI/SNF recruitment in synovial sarcoma. *Nat. Struct. Mol. Biol.* 2020 279 27, 836–845.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 22534–22539.

Mivelaz, M., Cao, A.M., Kubik, S., Zencir, S., Hovius, R., Boichenko, I., Stachowicz, A.M., Kurat, C.F., Shore, D., and Fierz, B. (2020). Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. *Mol. Cell* 77, 488-500.e9.

Mohammad, F., and Helin, K. (2017). Oncohistones: Drivers of pediatric cancers. *Genes Dev.* 31, 2313–2324.

Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D.P., Højfeldt, J.W., Comet, I., Zheng, C., Johansen, J.V., Rapin, N., Porse, B.T., et al. (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.* 23, 483–492.

Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., Shiota, H., Debernardi, A., Héry, P., Curtet, S., et al. (2013). Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev.* 27, 1680–1692.

Murray, I.A., Morgan, R.D., Luyten, Y., Fomenkov, A., Corrêa, I.R., Dai, N., Allaw, M.B., Zhang, X.,

Cheng, X., and Roberts, R.J. (2018a). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* *46*, 840–848.

Murray, I.A., Morgan, R.D., Luyten, Y., Fomenkov, A., Corrêa, I.R., Dai, N., Allaw, M.B., Zhang, X., Cheng, X., and Roberts, R.J. (2018b). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* *46*, 840–848.

Mutskov, V., Gerber, D., Angelov, D., Ausio, J., Workman, J., and Dimitrov, S. (1998). Persistent Interactions of Core Histone Tails with Nucleosomal DNA following Acetylation and Transcription Factor Binding. *Mol. Cell. Biol.* *18*, 6293.

Nabils, N.H., Deleyrolle, L.P., Darst, R.P., Riva, A., Reynolds, B.A., and Kladde, M.P. (2014). Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res.* *24*, 329–339.

Nacev, B.A., Feng, L., Bagert, J.D., Lemiesz, A.E., Gao, J.J., Soshnev, A.A., Kundra, R., Schultz, N., Muir, T.W., and Allis, C.D. (2019). The expanding landscape of ‘oncohistone’ mutations in human cancers. *Nature* *567*, 473–478.

Nakane, T., Kimanius, D., Lindahl, E., and Scheres, S.H.W. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* *7*.

Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* *154*, 490–503.

Oberbeckmann, E., Wolff, M., Krietenstein, N., Heron, M., Ellins, J.L., Schmid, A., Krebs, S., Blum, H., Gerland, U., and Korber, P. (2019). Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* *29*, 1996–2009.

Oberbeckmann, E., Niebauer, V., Watanabe, S., Farnung, L., Moldt, M., Schmid, A., Cramer, P., Peterson, C.L., Eustermann, S., Hopfner, K.P., et al. (2021). Ruler elements in chromatin remodelers set nucleosome

array spacing and phasing. *Nat. Commun.* 2021 121 12, 1–17.

Olins, A.L., and Olins, D.E. (1974). Spheroid Chromatin Units (v Bodies). *Science* (80-). 183, 330–332.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O’Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* (80-). 357.

Panditharatna, E., and Filbin, M.G. (2020). The growing role of epigenetics in childhood cancers. *Curr. Opin. Pediatr.* 32, 67–75.

Papamichos-Chronakis, M., and Peterson, C.L. (2012). Chromatin and the genome integrity network. *Nat. Rev. Genet.* 2013 141 14, 62–75.

Partensky, P.D., and Narlikar, G.J. (2009). Chromatin Remodelers Act Globally, Sequence Positions Nucleosomes Locally. *J. Mol. Biol.* 391, 12–25.

Patel, A.B., Moore, C.M., Greber, B.J., Luo, J., Zukin, S., Ranish, J., and Nogales, E. (2019). Architecture of the chromatin remodeler RSC and insights into its nucleosome engagement. *Elife* 8.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* 25, 154.

Polach, K.J., and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J. Mol. Biol.* 254, 130–149.

Pott, S. (2017). Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *Elife* 6.

Racki, L.R., and Narlikar, G.J. (2008). ATP-dependent Chromatin Remodeling Enzymes: Two Heads are not Better, Just Different. *Curr Opin Genet Dev* 18, 137–144.

Racki, L.R., Yang, J.G., Naber, N., Partensky, P.D., Acevedo, A., Purcell, T.J., Cooke, R., Cheng, Y., and Narlikar, G.J. (2009). The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nat.*

2009 4627276 462, 1016–1021.

Ramani, V., Qiu, R., and Shendure, J. (2019). High Sensitivity Profiling of Chromatin Structure by MNase-SSP. *Cell Rep.* 26, 2465-2476.e4.

Rao, S., Ahmad, K., and Ramachandran, S. (2021). Cooperative Binding between Distant Transcription Factors is a Hallmark of Active Enhancers. *Mol. Cell* 81, 255-267.e6.

Rhodes, D., Brown, R.S., and Klug, A. (1989). Crystallization of nucleosome core particles. *Methods Enzymol.* 170, 420–428.

Richard-Foy, H., and Hager, G.L. (1987). Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. *EMBO J.* 6, 2321–2328.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. *Nature* 311, 532–537.

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

Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta* 1839, 627.

Sanulli, S., Trnka, M.J., Dharmarajan, V., Tibble, R.W., Pascal, B.D., Burlingame, A.L., Griffin, P.R., Gross, J.D., and Narlikar, G.J. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nat.* 2019 5757782 575, 390–394.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436, 138–141.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T.W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A.K., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin

remodelling genes in paediatric glioblastoma. *Nature* 482, 226–231.

Scott, W.A., and Campos, E.I. (2020). Interactions With Histone H3 & Tools to Study Them. *Front. Cell Dev. Biol.* 8.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* 19, 65–71.

Serra-Cardona, A., and Zhang, Z. (2018). Replication-Coupled Nucleosome Assembly in the Passage of Epigenetic Information and Cell Identity. *Trends Biochem. Sci.* 43, 136–148.

Shema, E., Bernstein, B.E., and Buenrostro, J.D. (2018). Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat. Genet.* 2018 511 51, 19–25.

Shipony, Z., Marinov, G.K., Swaffer, M.P., Sinnott-Armstrong, N.A., Skotheim, J.M., Kundaje, A., and Greenleaf, W.J. (2020). Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* 2020 173 17, 319–327.

Simpson, R.T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* 17, 5524–5531.

Simpson, R.T., and Stafford, D.W. (1983). Structural features of a phased nucleosome core particle. *Proc. Natl. Acad. Sci. U. S. A.* 80, 51.

Smith, S., and Stillman, B. (1991). Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J.* 10, 971–980.

Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., and Shendure, J. (2016). Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* 164, 57–68.

Sobecki, M., Souaid, C., Boulay, J., Guerineau, V., Noordermeer, D., and Crabbe, L. (2018). MadID, a Versatile Approach to Map Protein-DNA Interactions, Highlights Telomere-Nuclear Envelope Contact

Sites in Human Cells. *Cell Rep.* 25, 2891-2903.e5.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* (80-). 344, 376–380.

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

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 2012 139 13, 613–626.

Steensel BV., and Henikoff S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* 18, 424–428.

Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* (80-). 368.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* 2020 224 22, 361–371.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J. Biol. Chem.* 281, 16279–16288.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–9445.

Suter, B., Schnappauf, G., and Thoma, F. (2000). Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.* 28, 4083.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected

communities. *Sci. Reports* 2019 9, 1–12.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res.* 25, 1491–1498.

Tse, O.Y.O., Jiang, P., Cheng, S.H., Peng, W., Shang, H., Wong, J., Chan, S.L., Poon, L.C.Y., Leung, T.Y., Chan, K.C.A., et al. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 118.

Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 83, 1021–1026.

Tullius, T.D. (1988). DNA footprinting with hydroxyl radical. *Nat.* 1988 3326165 332, 663–664.

Udugama, M., Sabri, A., and Bartholomew, B. (2011). The INO80 ATP-Dependent Chromatin Remodeling Complex Is a Nucleosome Spacing Factor. *Mol. Cell. Biol.* 31, 662.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–522.

Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nat.* 1997 3886642 388, 598–602.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 2015 163 16, 178–189.

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. (2020). Global reference mapping of human transcription factor footprints. *Nat.* 2020 5837818 583, 729–736.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific

computing in Python. *Nat. Methods* 2020 173 17, 261–272.

Wang, Y., Wang, A., Liu, Z., Thurman, A.L., Powers, L.S., Zou, M., Zhao, Y., Hefe, A., Li, Y., Zabner, J., et al. (2019). Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* 29, 1329–1342.

Wanior, M., Krämer, A., Knapp, S., and Joerger, A.C. (2021). Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy. *Oncogene* 2021 4021 40, 3637–3654.

Weber, C.M., and Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. *Genes Dev.* 28, 672.

Weber, C.M., Hafner, A., Kirkland, J.G., Braun, S.M.G., Stanton, B.Z., Boettiger, A.N., and Crabtree, G.R. (2021). mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol.* 2021 286 28, 501–511.

Weintraub, H., and Groudine, M. (1976). Chromosomal Subunits in Active Genes Have an Altered Conformation. *Science* (80-.). 193, 848–856.

Wiechens, N., Singh, V., Gkikopoulos, T., Schofield, P., Rocha, S., and Owen-Hughes, T. (2016). The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLOS Genet.* 12, e1005940.

Wilson, B.G., Wang, X., Shen, X., McKenna, E.S., Lemieux, M.E., Cho, Y.J., Koellhoffer, E.C., Pomeroy, S.L., Orkin, S.H., and Roberts, C.W.M. (2010). Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* 18, 316–328.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 1–5.

Workman, J.L. (2006). Nucleosome displacement in transcription. *Genes Dev.* 20, 2009–2017.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nat.* 2016 5327599 532, 329–333.

Xiao, C. Le, Zhu, S., He, M., Chen, D., Zhang, Q., Chen, Y., Yu, G., Liu, J., Xie, S.Q., Luo, F., et al. (2018). N 6 -Methyladenine DNA Modification in the Human Genome. *Mol. Cell* 71, 306-318.e7.

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., and Wu, C. (2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* 8, 531–543.

Xu, M., Long, C., Chen, X., Huang, C., Chen, S., and Zhu, B. (2010). Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. *Science* (80-.). 328, 94–98.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. *Mol. Cell* 2, 851–861.

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

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

Yang, Z., Zheng, C., and Hayes, J.J. (2007). The core histone tail domains contribute to sequence-dependent nucleosome positioning. *J. Biol. Chem.* 282, 7930–7938.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* 149, 1461–1473.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* *515*, 355–364.

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

Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* *37*, 76–81.

Zentner, G.E., and Henikoff, S. (2014). High-resolution digital profiling of the epigenome. *Nat. Rev. Genet.* *15*, 814–827.

Zhang, Y., and Li, Y. (2011). The Expanding Mi-2/NuRD Complexes: A Schematic Glance: <https://doi.org/10.4137/PRI.S63293>, 79–109.

Zhang, W., Feng, J., and Li, Q. (2020). The replisome guides nucleosome assembly during DNA replication. *Cell Biosci.* *2020* *101* *10*, 1–14.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* (80-.). *332*, 977–980.

Zhou, C.Y., Johnson, S.L., Gamarra, N.I., and Narlikar, G.J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors. *Annu. Rev. Biophys.* *45*, 153–181.

Zhou, C.Y., Johnson, S.L., Lee, L.J., Longhurst, A.D., Beckwith, S.L., Johnson, M.J., Morrison, A.J., and Narlikar, G.J. (2018). The yeast INO80 complex operates as a tunable DNA length-sensitive switch to regulate nucleosome sliding. *Mol. Cell* *69*, 677.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2010). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* *2011* *121* *12*, 7–18.

Zofall, M., Persinger, J., and Bartholomew, B. (2004). Functional Role of Extranucleosomal DNA and the Entry Site of the Nucleosome in Chromatin Remodeling by ISW2. *Mol. Cell. Biol.* 24, 10047–10057.

Chapter 4: Future Directions

4.1 Multi-omic techniques paired with SAMOSA

The generation of multiple types of data from a single biological sample promises to provide more comprehensive epigenomic maps to better understand human health and disease. Through streamlined integration of multiple data sets derived from a single experiment, we can gather richer biological information beyond singular genomics data, allowing us to better understand complex patterns in biology and ultimately uncover relevant therapeutic pathways. For instance, chromatin is not the static beads-on-a-string structure as traditional images would depict. Instead, it is a constantly shifting, dynamic structure that potentiates all cellular functions, - we just need the correct tools to truly visualize it. My dissertation has demonstrated the promise of multi-omics technology to map chromatin dynamics, through the development of a novel technique which harnesses non-destructive methyltransferase footprinting to map genome-wide nucleosome patterns, as well as 3rd generation PacBio sequencing to generate single-molecule and long-reads.

Through development of the single-molecule adenine methyltransferase oligonucleosome sequencing assay (SAMOSA), we were empowered to survey known epigenomic domains at higher-resolution. We next harnessed this technology to produce a more well-defined role of the essential Imitation Switch (ISWI) ATPase SNF2h, which slides nucleosomes depending on the local underlying nucleosomal density of its target loci, while working in concert with other transcriptional effectors to promote distinct transcriptional outcomes. Despite these exciting findings, we have demonstrated a small piece of what is possible with this technique. Beyond SAMOSA, scientists will continue to improve upon existing high-throughput sequencing techniques and iteratively build new technologies to generate rich biological data sets. For instance, multi-omic single cell sequencing promises to provide higher resolution genomic, transcriptomic, and epigenomic information beyond bulk-cell population measurements, which have been demonstrated here. There are many exciting applications of these multi-omic techniques, from studying the basic mechanisms of chromatin regulators to understanding how chromatin patterns are altered in a disease state. The following sections are dedicated to preliminary data and future directions which expand upon the

SAMOSAs technique to generate even more biologically targeted and relevant sequencing data. This includes pairing SAMOSA with the following: targeted genomic capture, chromosome conformation, and mapping nascent chromatin.

Targeted genomic capture with SAMOSA

While extremely informative and powerful, many current chromatin mapping methods including ChIP-seq, DamID, and CUT&RUN share key limitations: measurement of individual protein-DNA interactions inherently requires destruction of the chromatin fiber and averaging of signal across many short molecules. ChIP-enriched DNA for specific histone modifications without a PCR amplification step is compatible with SMRT sequencing (Wu et al., 2016). We therefore hypothesized that utilizing SAMOSA chromatin followed by immunoprecipitation (IP) and pulldown for a histone PTM of interest may allow for targeted enrichment, long-range transcription factor binding accessibility measurements, and higher coverage for regions of interest. Similar recent efforts cleverly combine antibody tethered to a protein A-m6A methyltransferase for enrichment of genomic regions using formaldehyde-fixed nuclei followed by nanopore sequencing (Altemose et al., 2022). However, formaldehyde fixation may inconsistently crosslink nuclear proteins, potentially leading to misinterpretation of downstream data analysis (Gavrilov et al., 2015). Further, while nanopore sequencing can obtain incredibly long reads, it is still limited by reduced read accuracy. We therefore sought to incorporate native chromatin dynamics while generating highly accurate reads with PacBio's CCS reads.

I initially sought out to combine a classic Native ChIP protocol with our pre-established SAMOSA protocol using live K562 nuclei. As is typical of existing ChIP-seq protocols, IP efficiency is highly dependent on both wash buffer composition and quality antibodies, in order to eliminate non-specific binding while obtaining robust pulldown. It is essential to determine the quality of even ChIP-grade antibodies and to include important controls, such as chromatin starting input and histone H3. I observed that the amount of solubilized chromatin generated via short MNase digests to obtain long fragments is

lower than highly digested samples. This does not however seem to influence IP efficiency, indicating that obtaining extra long-reads for IP is possible. After troubleshooting MNase digestion, wash buffer composition, and antibody quality, I demonstrate that we can enrich >3-8 fold at active genomic loci for H3K27ac and H3K4me3 compared to input. Further optimization of enrichment and probing other genomic regions including heterochromatin are important next steps. Lastly, developing methods that decrease required chromatin input and reduce sequencing costs will be essential. This may include alternative methods of target enrichment for low abundance proteins prior to PacBio Sequencing. For instance, other no-amplification, NGS targeted enrichment systems implementing CRISPR-Cas9 guide RNAs to cleave regions of interest may be extended to long-read PacBio sequencing and are an exciting prospect (Ebbert et al., 2018; Quan et al., 2019).

Chromosome conformation capture with SAMOSA

Chromosome conformation capture techniques (3C, 4C, Hi-C, ChIA-PET, DNase-HiC) are widely used to interrogate spatial organization of chromatin by quantifying the interactions between loci that are nearby in 3D space, and yet linearly separated in the genome (enhancer-promoter contacts, for instance). These protocols typically involve the fixation of protein-DNA with formaldehyde, followed by restriction enzyme (RE) DNA cleavage, solubilizing and sonicating to release complexes containing linked interactions, and subjecting them to bulk, short-read HTS. Despite many exciting results from these methods, they are limited by their ensemble, short-read sequences. Further, it has been suggested that these methods may miss important DNA regulatory interactions. For instance, detecting bridges via formaldehyde protein-DNA crosslinking may be too rare, or even produce technically inaccurate results (Gavrilov et al., 2015). The latter may be due to crosslinking disparities between genomic subcompartments (i.e. heterochromatin versus euchromatin), as well as cell-cycle dependent differences (condensed mitotic chromosomes).

It is essential to improve upon existing 3C techniques to accurately capture 3D genome dynamics. Upon thinking about ways to improve upon existing techniques, we thought it beneficial to obtain a multi-omic dataset which includes nucleosome positioning and chromatin spatial organization information. I conducted several pilot experiments where I considered alternatives to formaldehyde fixation. Using isolated K562 nuclei, I performed SAMOSA as previously done, followed by either native or several fixed conditions (1% formaldehyde, 1mM EGS, 1mM DSG). EGS (ethylene glycol bis(succinimidyl succinate)) and DSG (disuccinimidyl glutarate) are protein-protein crosslinkers containing amine-reactive NHS-ester ends. They have often been used in combination with formaldehyde to perform dual-crosslinking X-ChIP. Following the nuclear fixation (or native condition), chromatin was digested with a general restriction enzyme (HindIII or BamHI), end repaired with Klenow, and proximity ligated with T4 DNA ligase. Next, crosslinks were reversed and DNA was lightly sonicated and prepared for PacBio sequencing as usual. Upon assessing sequencing run quality metrics, all fixation methods decrease read-length quality but do not decrease the estimated number of passes around the molecule, when compared to natively sequenced samples. EGS-fixed samples have the worst performance out of all fixation methods, with DSG-fixed samples performing more similarly to FA-fixed. Upon analysis of the data, nucleosome footprinting patterns are reliably detected in both native and fixed samples, generating similar accessibility plots to both controls (-HiC where digestion and proximity ligation were excluded), as well as to previous *in vivo* SAMOSA data. Upon further analysis, few supplementary alignments were generated, indicating lack of enrichment for 3D contacts with these alternative fixation methods. Further optimization of fixation, potentially through dual-fixation, may be required to determine if alternative protein-protein fixation methods are sufficient to capture 3D contacts. Exciting, relevant future experiments may include pairing chromosome conformation plus SAMOSA to study long-range enhancer-promoter using degron-tagged cells to study essential chromatin regulators, especially BAF where dosage sensitivity is important in a disease context.

Mapping nascent chromatin with SAMOSA

Classic cell proliferation assays monitor cellular growth over time, cell division rate, metabolic activity, and DNA synthesis. As replicating cells undergo DNA synthesis in the S phase of the cell cycle, pyrimidine deoxynucleoside thymidine analogues may be inserted into replicating DNA to tag and characterize dividing cells over time. 5-bromo-2-deoxyuridine (BrdU) is the most common analogue used; BrdU-incorporated DNA is often quantified with harsh denaturing methods (HCl, heat) to expose BrdU, followed by labeling with anti-BrdU antibodies for downstream experiments (including Immunohisto/cytochemistry (IHC/ICC, enzyme linked immunosorbent assay (ELISA), Fluorescence-activated cell sorting (FACS)). Other halogenated thymidine analogues include 5-chlorodeoxyuridine (CldU) and 5-iododeoxyuridine (IdU). BrdU, CldU, and IdU are nearly identical to thymidine, except the methyl group on the 5-carbon is substituted with the bulky halogen molecule. Alternative DNA synthesis tagging methods harness the non-halogenated thymidine analogue 5-ethynyl-2'-deoxyuridine (Edu), which contains an alkyne group readily detected via 'Click Chemistry' (fluorescent azide probe and copper catalysis). Edu Click Chemistry allows for preservation of the cellular structure without harsh denaturing reagents (Cavanagh et al., 2011). While valuable, it should be noted that while Edu does not require a denaturing reagent for detection, it is cytotoxic to cells at high concentrations, inducing the DNA damage response (phosphorylation of H2A.X) and leading to cell cycle perturbation and apoptosis (Zhao et al., 2013).

During DNA replication, chromatin structure must be faithfully duplicated into newly replicating daughter strands. This involves the prompt replacement of displaced histones in a highly orchestrated manner, as covered extensively in chapter 1.3. Recent studies have been performed to better understand the mechanisms of chromatin replication, namely through labeling of nascent chromatin with thymidine analogues, typically used to assay cell proliferation as mentioned above. SCAR-seq (sister chromatids after replication by DNA sequencing) was developed to distinguish the distribution of recycled (parental) histones and naïve (new) histones onto the replicated daughter strands in mESCs. With this technique,

newly replicated histone-DNA was labeled with biotinylated EdU and isolated from parental histone-DNA via biotin click chemistry and streptavidin pulldown. After separating and sequencing new and parental DNA strands, they found that parental histones segregate with a slight preference for the leading strand, likely through the replicative helicase subunit MCM2 (Petryk et al., 2018). Another study performed thymidine-analog pulsing with BrdU in yeast strains with mutant Mcm2, which is an essential component of the replicative helicase that binds the H3-H4 tetramers. Their technique, eSPAN (Enrichment and sequencing of protein-associated nascent DNA), demonstrated Mcm2-Ctf4-Pol α are critical for parental histone recycling onto the lagging strand (Gan et al., 2018; Stewart-Morgan et al., 2020; Zhang et al., 2020)

We initially hypothesized that by incorporating BrdU during the exponential growth phase of cell culture prior to SAMOSA, during sequencing the SMRT high-fidelity polymerase may generate a distinct pause during incorporation of the halogenated (or non-halogenated) thymidine analogue compared to a normal, unmodified thymidine. This would allow us to distinguish nascent chromatin and nucleosome footprints at single-molecules with long-reads, which has never been done before. Initial pilot experiments included BrdU pulsing at multiple timepoints (50mM BrdU on 5e6 cells at 0, 30, 60, and 120 minutes, followed by 16 hours, up to 24 hours) prior to normal *in vivo* SAMOSA. I initially aimed to verify BrdU incorporation by intracellular flow and observed only a small population (<5%) of BrdU positive cells in the bulk population, possibly due to cell cycle kinetics and/or antibody quality. We later demonstrated that BrdU incorporation can be verified via simple slot blot (denature via HCl and incubate with anti-BrdU antibody), which shows an increasing gradient of BrdU incorporation in E14 mESCs incubated at 10 minutes, 1 hour, 6 hour and 24 hours.

Upon PacBio sequencing of our initial experiments, we observed a marked decrease in the local base rate (LBR) of the polymerase, indicating excessive pausing. To determine if this signal is real in sequenced molecules, we used a similar machine learning method used for adenine methylation to predict which whole molecules contain BrdU. For the training model, the molecules derived from the 24-hour pulse (without footprinting) which contain BrdU were used to predict IPDs. Our initial analyses demonstrate

evidence of BrdU incorporation, although the IPD is not nearly as distinct as m6a, indicating that a true-positive template is necessary for neural network predictions. As a proper positive control, I PCR-generated gDNA amplicons containing BrdUTP, in addition to normal dNTPs and half dTTP/BrdUTP. Out of ~100,000 molecules passing quality filtering, we found that only a subset of molecules (~10,000) were required for training, indicating that low input samples may work efficiently with this method.

We have since generated more control amplicons or pulsed with alternative thymidine analogs to determine if they generate distinct IPDs on the SMRT sequencer, including halogenated IdU and CIdU, as well as non-halogenated EdU, 4-thio-thymidine (S4dT, or 4sT), and even dUTP (Mitter et al., 2020). We initially pulsed cells with EdU, and prepared control gDNA amplicons with EdUTP as previously performed. For these alternative analogues, we observed that EdU pulsing produces a cytotoxic effect at longer pulse conditions, and interestingly, significantly disrupts the sequencer. IdU and CIdU do not have a cytotoxic effect on cells and do not disrupt the sequencer. Lastly, we were unable to produce 4sT or dUTP amplicons, and chose not to continue with these analogues. As an alternative true positive control, we generated amplicons from MNase-digested chromatin (with and without footprinting) and performed Illumina-style PCR using either EdUTP or BrdUTP, followed by SMRTbell library preparation for PacBio sequencing. Further computational training to determine distinct IPDs on BrdU, IdU, and CIdU incorporated molecules is the next step. Once this is reliably in place, we can confidently call analogue-incorporated molecules to study nucleosome patterns on nascent chromatin, which is a very exciting prospect.

REFERENCES

- Altomose, N., Maslan, A., Smith, O.K., Sundararajan, K., Brown, R.R., Mishra, R., Detweiler, A.M., Neff, N., Miga, K.H., Straight, A.F., et al. (2022). DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. *Nat. Methods* 2022 1–13.
- Blosser, T.R., Yang, J.G., Stone, M.D., Narlikar, G.J., and Zhuang, X. (2009). Dynamics of nucleosome remodelling by individual ACF complexes. *Nature* 462, 1022.
- Cavanagh, B.L., Walker, T., Norazit, A., and Meedeniya, A.C.B. (2011). Thymidine analogues for tracking DNA synthesis. *Molecules* 16, 7980–7993.
- Ebbert, M.T.W., Farrugia, S.L., Sens, J.P., Jansen-West, K., Gendron, T.F., Prudencio, M., McLaughlin, I.J., Bowman, B., Seetin, M., Dejesus-Hernandez, M., et al. (2018). Long-read sequencing across the C9orf72 “GGGGCC” repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Mol. Neurodegener.* 13, 1–17.
- Gan, H., Serra-Cardona, A., Hua, X., Zhou, H., Labib, K., Yu, C., and Zhang, Z. (2018). The Mcm2-Ctf4-Pol α Axis Facilitates Parental Histone H3-H4 Transfer to Lagging Strands. *Mol. Cell* 72, 140-151.e3.
- Ganji, M., Docter, M., Le Grice, S.F.J., Abbondanzieri, E., Harada, B.T., Hwang, W.L., Deindl, S., Bartholomew, B., Zhuang, X., Mit, H., et al. (2016). Stepwise Nucleosome Translocation by RSC Remodeling Complexes. *Biophys. J.* 110, 515a.
- Gavrilov, A., Razin, S. V., and Cavalli, G. (2015). In vivo formaldehyde cross-linking: it is time for black box analysis. *Brief. Funct. Genomics* 14, 163–165.
- Han, Y., Reyes, A.A., Malik, S., and He, Y. (2020). Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nat.* 2020 5797799 579, 452–455.
- Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* 1.
- Kassabov, S.R., Zhang, B., Persinger, J., and Bartholomew, B. (2003). SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol. Cell* 11, 391–403.

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., et al. (2020). A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. *Cell* 183, 802-817.e24.

Mitter, M., Gasser, C., Takacs, Z., H Langer, C.C., Tang, W., Jessberger, G., Beales, C.T., Neuner, E., Ameres, S.L., Peters, J.-M., et al. (2020). Conformation of sister chromatids in the replicated human genome. *Nature* 586, 139.

Petryk, N., Dalby, M., Wenger, A., Stromme, C.B., Strandsby, A., Andersson, R., and Groth, A. (2018). MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science* (80-.). 361, 1389–1392.

Quan, J., Langelier, C., Kuchta, A., Batson, J., Teyssier, N., Lyden, A., Caldera, S., McGeever, A., Dimitrov, B., King, R., et al. (2019). FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *Nucleic Acids Res.* 47, e83–e83.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* 2020 224 22, 361–371.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nat.* 2016 5327599 532, 329–333.

Zhang, W., Feng, J., and Li, Q. (2020). The replisome guides nucleosome assembly during DNA replication. *Cell Biosci.* 2020 101 10, 1–14.

Zhao, H., Halicka, H.D., Li, J., Biela, E., Berniak, K., Dobrucki, J., and Darzynkiewicz, Z. (2013). DNA Damage Signaling, Impairment of Cell Cycle Progression, and Apoptosis Triggered by 5-Ethynyl-2'-deoxyuridine Incorporated into DNA. *Cytometry. A* 83, 979–988.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Nour Abdulkay

0F401D5929D74F1...

Author Signature

5/23/2022

Date