Dissecting the Role of Nucleosomes in Transcription and the Application of CRISPR/Cas9 Technology

by

Lea Bengtson Witkowsky

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert Tjian, Chair
Professor Donald C Rio
Professor Carlos J Bustamante
Associate Professor Daniel Zilberman

Summer 2016

Dissecting the Role of Nucleosomes in Transcription and the Application of CRISPR/Cas9 Technology

Abstract

Dissecting the Role of Nucleosomes in Transcription and the Application of CRISPR/Cas9 Technology

by

Lea Bengtson Witkowsky

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Robert Tjian, Chair

In the 43 years since the first observation of nucleosome core particles as beads on a string, our understanding of the relationship between nucleosome structure and its function in living cells has made giant strides. Despite this progress, many questions remain unanswered regarding the role of nucleosomes in transcription initiation, and their interaction with the core transcriptional machinery. Here, I present my progress on addressing the role of TAF1 recognition of acetylated histone H4 tails, and its effect on transcription initiation and start site selection. Preliminary results suggest that H4 acetylation may play a role in start site selection.

One remarkable advancement in technology that will undoubtedly help answer remaining questions in chromatin biology is the repurposing of the prokaryotic CRISPR-Cas system to edit eukaryotic genomes with exquisite precision. Ironically, a tool that may shed light on complex eukaryotic chromatin interactions was evolved in a prokaryotic environment. My work with in vitro nucleosomes positioned me to address whether Cas9 is capable of engaging nucleosomal DNA. Both in vitro and in vivo analyses done in collaboration with Max Horlbeck and Jonathan Weissman at UCSF suggest that nucleosomes pose a strong barrier to Cas9.

An expert is a person who has
made all the mistakes that can
be made in a very narrow field.

Niels Bohr

For My Daughter, Georgia

Contents

List of Figures

List of Tables

Acknowledgments

I must first thank my committee for their advice and great patience with me, and most importantly, for allowing me to turn in this thesis. I also wish to thank Dr. Claudia Cattoglio for her mentorship and friendship, as well as Dr. Yick Fong for teaching me transcription and for his endless faith in me as a scientist. As mentioned in Chapter 3, I must also thank Dr. Karolin Luger, Pam Dyer, and Dr. Uma Muthurajan for teaching me to work with nucleosomes. Additionally, I owe a great debt to my husband, Dr. T. Blair Gainous, for holding me up through the hardest parts of my PhD, and more recently, for staying up all night to watch Georgia while I wrote this, and for proofing it as I went. I must also thank my parents, in particular my mother who has dropped her plans many times to keep me afloat, and my mother–in–law Vickie Gainous for flying across the country at a moment's notice to watch Georgia as I scrambled to finish this PhD.

And lastly, I must thank my mentor, Dr. Robert Tjian. I have been immensely lucky in my PhD to end up working for a man who is not only brilliant, but an incredibly compassionate and caring human being to boot. To him, I owe my entire PhD and career in science, and can never repay the kindness he has shown me.

Chapter 1

The Nucleosome

1.1 Introduction

In 1973 at Oak Ridge National labs in Tennessee, Don and Ada Olins peered through a magnifying glass at their electron micrographs of chicken erythrocyte chromatin spreads and saw the iconic structure of beads on a string [1, 2]. In that same year, nearly 1000 miles away at the University of Massachusetts, CLF Woodcock performed his own Miller spreads and identified the same architecture [1, 3]. Meanwhile, on the Pacific coast, KE Van Holde and his students had been performing circular dichroism and analytical sedimentation analysis on Micrococcal nuclease digested chromatin, the results of which pointed to a spherical particle in which proteins were highly compacting DNA [4–6]. And so, with at least three separate groups making the first observation and characterizations of the nucleosome core particle around the same time, so began decades of study of nucleosome structure and function. Some 43 years later, far more is known about the structure and function of this basic unit of chromatin, yet many questions still remain. In this chapter I provide a brief review of nucleosome biology and highlight persisting questions that are addressed in the following two chapters.

1.2 Nucleosome Structure

The eukaryotic nucleosome core particle is composed of two copies each of the four core histones, H2A, H2B, H3, and H4, with ∼147 bp of DNA wrapped 1.65 times around the histone octamer in a left–handed superhelix (Figure 1.1) [7, 8]. Histones bind to DNA primarily through hydrogen bonding between the histone fold domains and the phosphodiester backbone of DNA at every helical turn where the minor groove faces the histone octamer. These contacts are accompanied by conserved arginines that insert into the minor groove and help position the DNA. In addition to the histone fold interactions, the intrinsically disordered histone tails also make stabilizing contacts, both to other histones, and to the DNA. Together, over 120 direct interactions between the histone octamer and the DNA, as well as indirect water mediated H–bonding serve to overcome the 150bp persistence length of DNA and compact it by about 5–fold [9–11]. This wrapping serves as the first order of chromatin structure and is thought to have two main functions: (1) it serves to package the enormous amount of genetic material in eukaryotes into their relatively small nucleus, and (2) by physically obscuring the DNA, it can contribute to gene regulation.

Despite its extensive contacts with the histone octamer, nucleosomal DNA displays dynamic properties. One pivotal early study used restriction enzyme accessibility to calculate equilibrium constants for site exposure along the DNA [14]. In this study, Polach and Widom proposed a model in which the entry and exit sites of the DNA spontaneously peel away from the histone face before re–wrapping. Later studies confirm this model using single–molecule as well as bulk experiments [15–20]. This 'DNA breathing' occurs on the order of tens of milliseconds and can function to allow access to transcriptional machinery and regulators [17, 21–23].

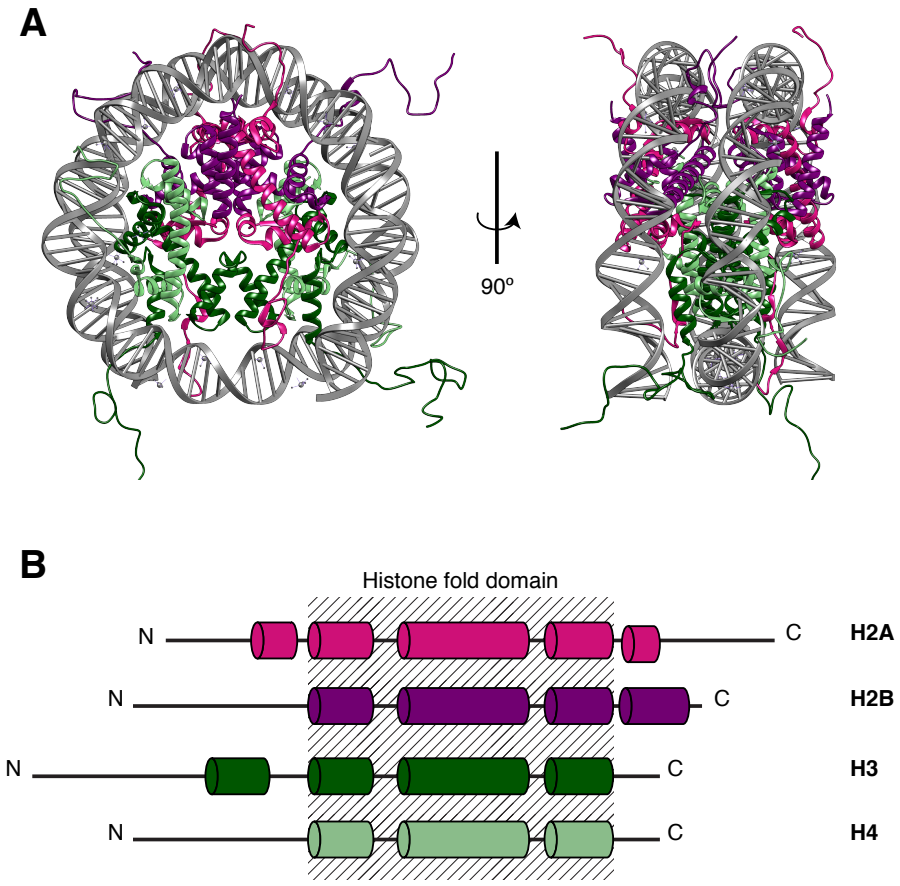


Figure 1.1: *Structure of the nucleosome core particle.* ***(A)*** *Crystal structure of the nucleosome core particle using Xenopus Laevis histones at 1.9 Angstrom Resolution ([12] PDBID 1KX5). DNA is shown as a grey ladder, while histones are represented as ribbons. Graphics were created using UCSF Chimera[13].* ***(B)*** *Color scheme for the histones in (A). The histone fold domains are aligned and highlighted with a striped background. Alpha helices are depicted as barrels.*

Adjacent nucleosomes are connected by 20–80 bp of linker DNA that associates with a linker histone (H1, H5) at the base of each core particle [11, 24] . H1 is expressed with the four core histones during S–phase, and is present in near stoichiometric amounts with each nucleosome core particle. Metazoan H1 has been observed to stabilize the nucleosome structure in vitro and contribute to higher order folding into a chromatin fiber [24]. Additionally, studies have connected H1 with global heterochromatin formation, including its recognition by the heterochromatin protein HP1 when methylated [25, 26]. These observations have lead to the assumption that H1 is a globally repressive factor in gene regulation, however knockdown experiments in multiple organisms have shown that rather than a global transcriptional effect, H1 influences only a subset of genes [26–28]. Thus, much remains to be learned about the role of linker histones in chromatin regulation.

1.3 Nucleosome Position

In all eukaryotic organisms surveyed, nucleosomes are observed to form a characteristic pattern around transcriptionally active genes (Figure 1.2). This pattern is distinguished by a nucleosome depleted region (NDR) of around 140 base pairs surrounding the transcription start site (TSS) with variation in length between species and promoters. This NDR is flanked with a well positioned nucleosome directly upstream and downstream, termed the -1 and +1 nucleosome respectively. Most organisms display a slightly higher occupancy and tighter phasing of the +1 nucleosome as compared to the -1, with a decaying periodic pattern of nucleosome occupancy within the body of active genes, ending in another NDR at the transcription termination site. This pattern has been highly documented and is reviewed exceptionally well in Hughes and Rando (2014) [29], and Jiang and Pugh (2009) [30].

Despite being a ubiquitous component of eukaryotic chromatin, and its ability to occupy over half of an organism's genomic DNA, nucleosomes have been observed to have some intrinsic sequence preference. As predicted by the extreme bending of DNA around the histone octamer, nucleosomes favor sequences that are more flexible and thus provide a lower energy barrier to wrapping [31]. Sequences that fall in the minor groove at the histone contact point every 10.2 bps favor the A/T dinucleotide, likely due to its propensity to form a narrower minor groove and thus tolerate a bent structure [29, 32]. Accordingly C/G dinucleotides are observed to be enriched in the opposite periodic phasic, allowing a wider major groove during bending. This preference for a periodic pattern of dinucleotides has been observed using genome–wide mapping techniques to identify the precise locations of endogenous nucleosomes in a number of different species [29, 30]. Jonathan Widom took a complementary approach to exploring intrisic sequence preferences by performing in vitro SELEX experiments to enrich a random pool of DNA fragments for tight binders of nucleosomes [33]. This experiment produced the highly popular 601 sequence that positions a stable and highly phased nucleosome in vitro, and confirmed dinucleotide rules for positioning.

Just as flexible or pre-bent structures might favor nucleosome binding, rigid sequences have a strong discouraging effect. Andrew Travers' group found early on that poly(dA:dT)

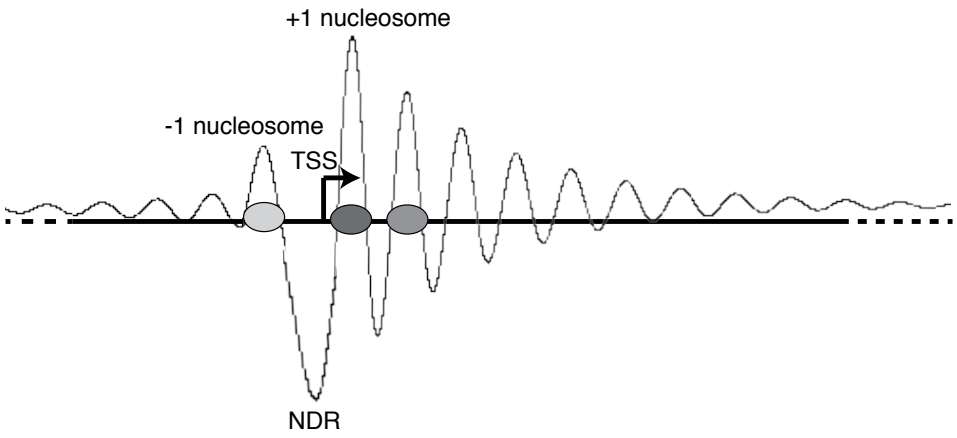


Figure 1.2: *Cartoon of nucleosome position surround transcription start sites. The nuclease digestion pattern of typical eukaryotic nucleosomes in vivo is illustrated with a periodic black line where peaks represent high nucleosome occupancy, and troughs represent low. Grey ovals depict the -1, +1, and +2 nucleosomes at typical active genes with the darkness of the color corresponding to their relative occupancy.*

tracts were intrinsically stiff and less likely to form a nucleosome [34]. Some of the strongest evidence in support of this model in the subsequent 30 years are experiments that reconstituted nucleosomes in vitro on genomic DNA followed by enzymatic digestion and sequencing of the protected nucleosomal DNA. These studies found a significant depletion of nucleosomes at poly(dA:dT) tracts, and point to a link between underlying sequence and intrinsic nucleosome positioning in vivo[29, 35–37]. The organism of choice for these genomic reconstitution studies was *S. cerevisiae* whose promoters are enriched in poly(dA:dT) tracts. Thus it was proposed that eukaryotes might encode their own nucleosome depletion at promoters.

As observed originally in yeast, promoter depletion has proven to hold true as a general principle of active genes in all eukaryotes surveyed. In seeming contradiction, however, is the observation that higher eukaryotes, such as vertebrates, have evolved much more GC rich promoters as compared to yeast, yet they still display nucleosome depletion. In fact over 70% of vertebrate genes are regulated by promoters with GC content higher than the surrounding sequence (CpG islands) [38, 39]. Speculation as to the mechanism of this depletion has led to studies testing the nucleosome occupancy of these vertebrate promoters in in vitro reconstitution assays. Surprisingly, the GC rich promoters were observed to self assemble nucleosome depleted regions in a similar manner to the AT rich promoters of yeast [40, 41]. Thus, it seems that while heteropolymeric GC rich sequences favor nucleosome occupancy, homopolymeric(dG/dC) sequences display stiff properties and are therefore likely to discourage nucleosome assembly [42]. While the mechanism of NDR formation at CpG islands in vertebrates may then be somewhat encoded at the genetic level, there are still conflicting results and many questions remain to be answered.

Despite a general agreement that nucleosomes exert some sequence preference, a rather

heated debate ensued in the years around 2009 as to what is primarily responsible for nucleosome positioning in vivo. Was it mostly based on intrinsic sequence preference, or were trans-acting factors such as chromatin remodelers the predominant driving force? This debate seems to have settled on a model in which the ground nucleosome organization state is encoded at the sequence level, while chromatin remodelers counteract preferred sites to create the more nuanced organization observed in vivo. A convincing experiment was performed in the Pugh lab in which yeast genomic DNA was first assembled in vitro to reveal intrinsic preferences, and then incubated with yeast extract and ATP to almost precisely recapitulate in vivo data and suggested that both intrinsic and trans–acting factors serve to create the nucleosome pattern observed in vivo [37]. Further work in metazoans revealed that chromatin remodelers actually exert their own sequence bias, and often function to counteract intrinsic sequence preferences of nucleosomes in vivo [43].

1.4 Histone Variants

In addition to the replication dependent expression of the five canonical histones, histone variants are also expressed in eukaryotes. These variants are expressed throughout the cell cycle and have evolved specialized functions. Here, I introduce the four most common variants, but a comprehensive review can be found at [44].

Two major H3 variants are found in most eukaryotic lineages; H3.3 and CenH3 (or CENP-A) with a variety of others specific to a subset of organisms. H3.3 is exchanged for canonical H3 at promoters associated with active transcription, and is thought to create a destabilized structure in order to facilitate assembly of the transcription machinery at the promoter and passage by the polymerase [45–47]. CenH3 on the other hand, is functional only in centromeres, and is critical for mitosis [48].

Besides variants of H3, most eukaryotes also encode H2A variants. Of these, the most studied and conserved are H2A.Z and H2A.X. H2A.X has been implicated in facilitating DNA repair and may also play a role in X inactivation, among other processes [49]. H2A.Z, like H3.3, replaces the canonical H2A at active promoters in many locations throughout the genome, and is though to contribute to a looser nucleosome structure [45, 50, 51].

1.5 Histone Modifications

Nucleosomes regulate gene expression by occluding gene regulatory sequences and by creating a physical impediment to RNA polymerase. While some histone variants are thought to relieve this physical hurdle (as mentioned above), histone modifications have been found to not only alter chromatin structure through direct changes in the physical properties of a nucleosome, but also to contribute to gene regulation through specific recognition by regulatory factors. A curated list of histone modifications, the proteins known to write

the modification, and the associated functions is provided below (Table 1.1). Here, I will briefly highlight one of the best studied examples.

1.5.1 Histone Acetylation

Histone lysines act as the receiving residue for many of the observed histone post translational modifications. As positively charged residues, their presence in histone tails is thought to contribute directly to nucleosome structure through electrostatic interactions with DNA [24]. Additionally, the positively charged lysines of the H4 tail in particular have been observed to contribute to higher order chromatin structure through internucleosomal interactions with an acidic patch on the H2A/H2B dimer of the adjacent nucleosome [10, 24]. Compared to other common lysine modifications, acetylation is the only process that results in charge neutralization. Thus, lysine acetylation might be expected to decondense chromatin and lead to looser nucleosomes. In agreement with this hypothesis, many examples have underscored the role of histone lysine acetylation in creating transcriptionally permissive chromatin. In general, genome–wide chromatin immunoprecipitation (ChIP) experiments have localized acetylated residues of H3 and hyper–acetylated H4 to the promoter proximal nucleosomes of active genes [52, 53]. In vitro biophysical studies have complemented this work by examining the impact of H4 acetylation on the nucleosome core particle's structure and dynamics, and demonstrated the creation of a destabilized nucleosome [9, 53]. Unlike other modifications such as methylation which can have activating or repressing roles depending on which residue is modified, all lysine acetylation events characterized to date display only transcriptionally activating phenotypes, likely due in part to the generalized mechanism of deionization. In addition to its structural effect, lysine acetylation is recognized by a specialized protein fold - the bromodomain. A number of proteins involved in active transcription, such as chromatin remodelers, harbor bromodomains, and are thought to be recruited to promoters to activate transcription through their interaction with specific acetylated lysines. However, a cause and effect relationship remains elusive. Despite a clear correlation with active transcription, no studies have examined the direct influence of this modification and its location on mechanisms of pre-initiation complex formation and transcription initiation at promoters. In Chapter two of this thesis, I describe my efforts to address this question.

1.6 Archaeal and eukaryotic nucleosomes and CRISPR systems

Contrary to what is written in almost every Biology 101 textbook, packaging of DNA by histone proteins is not a eukaryotic invention. Examples of histone proteins and nucleosome-like structures in prokaryotes are numerous [54], and conversely there is one notable example of the absence of histones in a eukaryotic organism - the dinoflagellate [55]. Due to their presences in certain branches of archaea, it seems that histones, and in fact, nucleosomes (though with altered composition), are an ancient evolutionary adaptation that arose prior

to the divergence of eukaryotes [54, 56, 57]. Recently, genome–wide mapping of euryarchaeal nucleosomes has demonstrated an even greater similarity in structure and function with eukaryotic nucleosomes than previously though [58, 59]. Further investigation into this evolutionary link may shed light on the function of the nucleosome in eukaryotes.

With the advent of prokaryotic CRISPR-Cas9 technology in the years surrounding 2012 and its repurposing for gene editing and regulation in eukaryotic cells, questions began to arise in the field about its ability to engage nucleosomal DNA. While bacteria have their own chromatin-like proteins that play a structural and functional role in gene regulation, they have no homology to eukaryotic histones, and are thought to have a much looser organization (though prokaryotic chromatin organization is still poorly understood) [60]. Additionally, since Cas9 evolved to target naked invading viral DNA, it was rather surprising that Cas9 could act on eukaryotic chromatin at all. We sought to address this question by taking a purified in vitro approach. In the summer of 2015 I presented my preliminary findings that nucleosomes impede Cas9 at the Ignite DNA Editing Supergroup (IDES) hosted by the Innovative Genomics Initiative at Berkeley. A graduate student from UCSF, Max Horlbeck, was also presenting at that meeting, and contacted me afterward to share some data he had collected but not presented on nucleosome position and CRISPR activity. When we realized we were both seeing the same result using two completely different approaches, we decided to collaborate and put our findings into a co-authored paper which was published in *eLife* the following year and is presented in Chapter 3 of this thesis.

While the current CRISPR system of choice is the RNA–containing holoenzyme Cas9 from the bacteria *Streptococcus pyogenes*, new CRISPR effector molecules are currently being investigated for repurposing in animals, and future candidates could include systems from archaea. Given that nearly an entire branch of archaea seems to utilize histones in much the same manner as eukaryotes, it will be interesting to see if and how CRISPR effectors from the phylum euryarchaeota can act in animal cells. In fact, it is tempting to hypothesize that inhibition by nucleosomes could be a mechanism of discrimination of self from non–self. For example, if by chance, CRISPR targets occur in the archaea's genome, the sequence would have a high probability of falling within an archaeal nucleosome, and could thus be protected from degradation, while the same sequence would be unprotected during a viral infection. This is a testable hypothesis and could lead to intriguing and useful findings if investigated in the future.

Table 1.1: *A curated list of histone modifications*

Histone	Residue	Modification	Writer	Function	Ref.
H2A	S1	P	S6K-alpha-5 *(Hs)*	Transcriptional Repression	[61]
	R3	Me	PRMT5 - Me2s *(Mm)*	Me2s - Transcriptional Repression	[62]
	K5	Ac	Tip60, p300/CBP *(Hs)*	Active Transcription	[63–65]
	K9	Ac		Active Transcription	[66, 67]
	K13	Ac	NuA4 *(Sc)*	Active Transcription	[68]
	K119	Ub	dRing *(Dm)*, RING1B *(Hs)*	Transcriptional Silencing, Polycomb	[69, 70]
	K120	P	BUB1 *(Hs)*	prevent chromosomal instability during mitosis	[71]
H2B	K5	Ac	p300, ATF2	Active Transcription	[65, 66, 72–74]
	K12	Ac	p300, ATF2	Active Transcription	[65, 66, 74]
	S14	P	Mst1 (vertebrates)	Apoptosis	[75]
	K15	Ac	p300, ATF2	Active Transcription	[65, 74]
	K120	Ub	BRE1A, BRE1B, RBX1 *(Hs)* RNF20/40 & UbcH6 *(Hs)*	Active Transcription	[76–78]
H3	R2	Me	CARM1/PRMT4 - Me2a *(Hs)* PRMT6 - Me2a *(Hs)*	CARM1: Transcriptional Activation (a few older studies) PRMT6: Transcriptional Repression, and silencing	Activation: [79, 80] Repression: [81, 82]
	T3	P	Haspin *(Hs)*	Centromere mitotic spindle function	[83, 84]
	K4	Me	Set1; me3 *(Sc)* Trx *(Ds)* Trr *(Ds)* SETD1A; me13 SETD1B *(Hs)* MLL; me13 MLL2 *(Hs)* MLL3; me1 3 *(Hs)* MLL4 *(Hs)* ASH1L; me1/3 *(Hs)* SETD7/SET9 *(Hs)*	rDNA/telomeric silencing (only shown in yeast) Active Enhancer Active Transcription	[85–88] [89–100]

Continued on next page

Table 1.1 – continued from previous page

Histone	Residue	Modification	Writer	Function	Ref.
	R8	Me	PRMT5 - Me2s *(Hs)*	examples of both Transcriptional Activation and Repression	Activation: [101, 102] Repression: [103, 104]
	K9	Ac	SAGA, GCN5 *(Sc)* SRC1 *(Hs)*	Active Transcription Histone deposition	[105–108]
		Me	Suv39h1, Suv39h2 *(Mm, Hs)* SETDB1, G9a *(Mm, Hs)* EHMT1/GLP, PRDM2/RIZ1 *(Hs)*	Traditionally considered correlated with Silenced Genes, but more recently has been shown at many active promoters in mammals	[109–113] [69, 114–118] [119, 120] [96, 121, 122]
	S10	P	Aurora-B kinase *(Sc, Ce)* Rsk2, Msk1, Msk2 *(Hs)* Snf1 *(Sc)* IKK-alpha *(Hs)*	Mitotic chromosome condensation IE gene activation Active Transcription Active Transcription	[123, 124] [125, 126] [127] [128, 129]
	T11	P	Dlk/Zip (mammals)	mitosis	[130]
	K14	Ac	TAF1 *(Dm, Hs)* GCN5 *(Mm, Sc)* p300, PCAF *(Hs)*	Active Transcription	[65, 131, 132]
	R17	Me	CARM1/PRMT4 - Me2a *(Hs, Mm)*	Active Transcription	[133]
	K18	Ac	SAGA, GCN5, Ada *(Sc)* p300/CBP *(Hs)*	Active Transcription Active Transcription, DNA replication	[25, 65, 105]
	K23	Ac	Sas3, SAGA *(Sc)* CBP *(Hs)*	Active Transcription	[25, 105, 134]
	R26	Me	CARM1/PRMT4 *(Hs)*	Active Transcription	[135, 136]
	K27	Ac	CBP, P300, GCN5 *(Dm, Sc)*	Active Transcription	[117, 137–139]
		Me	Ezh1/2, G9a	Transcriptional Silencing	[117, 140, 141]
	S28	P	Aurora-B kinase (mammals) MSK1, MSK2 (mammals)	mitosis immediate-early activation	[142] [125, 143]
	K36	Ac	GCN5 *(Sc)*	Active Transcription	[144]

Continued on next page

Table 1.1 – continued from previous page

Histone	Residue	Modification	Writer	Function	Ref.
		Me	Set2 *(Sc)* SETD2, NSD13 *(Hs)* Ash1 *(Hs)*	Active Transcription Anti-repression	[145] [146] [147] [148] [149]
H4	S1	P	CK2 *(Sc, Hs)*	DNA damage response Mitotic chromatin condensation	[150, 151] [152]
	R3	Me	PRMT1 - Me2a PRMT5/7 - Me2s	Me2a: Active Transcription Me2s: Transcriptional Repression	Active: [153–158] Repressive: [158–161]
	K5	Ac	Hat1 *(Tt, Mm, Hs)* Esa1/NuA4 *(Sc)* Tip60 ATF2, p300 *(Mm, Hs)*	Histone deposition Cell Cycle Progression, Active Transcription Active Transcription	[162–164] [165–167] [64] [65, 168]
	K8	Ac	Esa1/NuA4 *(Sc)* Tip60 ATF2, PCAF, p300 *(Mm, Hs)* Elp3	Active Transcription	[165] [64] [65, 74, 168] [169]
	K12	Ac	Hat1 *(Sc, Hs)* Esa1/NuA4 *(Sc)* Tip60 *(Hs)*	Histone deposition, Telomeric silencing Cell Cycle Progression, Active Transcription Active Transcription	[66, 162–164, 168, 170–173]
	K16	Ac	GCN5 *(Sc)* Esa1/NuA4 *(Sc)* Tip60 ATF2 *(Mm, Hs)* MOF *(Hs)*	Active Transcription	[64, 74, 131, 165, 168, 174, 175]
	K20	Me	Suv420h1/2 - me2, me3 *(Mm,Dm)* SETD8/Pr-SET7; me1 *(Hs,Dm)* Ash1 - me2 *(Dm)*	Gene silencing Gene silencing, Mitotic condensation Ash1: Trithorax activation	[176] [177–179] [180]

Chapter 2

Nucleosomes may provide an alternate transcription start site signal

2.1 Introduction

While a eukaryote's genome contains all the information necessary to direct cellular function and specialized processes throughout the organism's lifecycle, not every gene is expressed at the same level at the same time or location. In fact, precise control of gene expression is crucial in all domains of life. The first point of regulation occurs at the transcription level. In eukaryotes, transcriptional regulatory signals are integrated by the core machinery at the gene promoter and include information encoded directly in the DNA sequence as well as structurally through nucleosomes.

Core promoters of RNA polymerase II transcribed genes are characterized by around 100 or so base pairs flanking the transcription start site (TSS) and can contain sequence motifs that are bound by members of the core (i.e. basal) transcriptional machinery [39, 181]. These sequence motifs, termed core promoter elements (CPEs) encode start site information by orienting the pre–initiation complex and positioning the RNA polymerase at the correct location to initiate transcription. Core promoter elements also contribute to expression level and frequency through the strength of their interactions with the core machinery, and therefore their ability to recruit and maintain transcription complexes at the promoter [182, 183]. To date, only two components of the multi–protein pre–initiation complex are capable of recognizing DNA sequence specifically; TFIIB and TFIID [184]. TFIIB is known to recognize the two elements, the BREu and BREd, while all other known interactions between CPEs and the core machinery are through recognition by components of TFIID. A substantial number of identified metazoan CPEs and all the known interactions with the core machinery are depicted in Figure 2.1 and listed in Table 2.1.

Canonical TFIID is composed of the TATA binding protein (TBP) and ∼14 different

TBP–associated factors (TAFs) [185]. TBP and its core promoter element, the TATA box, represent one of the most ancient promoter interactions, as archaeal transcription is also driven by TBP and utilizes an archaeal TATA box [57]. TAFs, on the other hand, have evolved in conjunction with an increase in regulatory complexity among eukaryotes. In addition to recognizing a set of core promoter elements, they also harbor a number of specific interaction domains that are recognized by different activators and coactivators of transcription, as well as enzymatic properties such as acetyl transferase and kinase activity. Of these ∼14 TAFs, a few have been found to bind specific promoter sequences, and interestingly, in accordance with their evolution in eukaryotes, two have been identified to interact with specific histone post translational modifications associated with active transcription [185, 186]. The first of these two, TAF3 has been shown in vivo and in vitro to bind to H3K4me3, a mark highly correlated with transcriptionally active promoters through its PHD finger [187–190]. This interaction was shown to recruit TFIID and the pre–initiation complex and promote transcription at a number of genes [187, 190].

The second, and only other TAF known to interact with modified histone tails, is TAF1. Mammalian TAF1 contains a double bromodomain which is capable of binding hyperacetylated H4 tails in vitro [191]. Unlike TAF3, however, very little is known about this proposed interaction, with no evidence of direct involvement in transcription. Some data in yeast, however, do suggest a possible functional interaction in vivo. The double bromodomain is absent from TAF1 in yeast and is thought to instead reside in the protein, Bdf1 [192]. This protein has been shown to bind to acetylated H4 tails in vivo and exist in two complexes, one of which contains TFIID [193]. Additionally, a study at the PHO5 promoter found an interplay between the H4 acetylation binding activity of Bdf1 and its ability to drive transcription through TFIID in a manner that depended on promoter architecture [194]. Thus, it seems likely that the bromodomain of mammalian TAF1 may have an important functional role in cells. Furthermore, a vast majority of genes in mammals do not contain strong core promoter elements, and so pre–initiation complex positioning and TSS selection mechanisms must rely on other cues [39, 181, 195].

In this chapter, I present my efforts to address the role of TAF1 and H4 acetylation at promoter–flanking nucleosomes in transcription. I take a highly purified in vitro approach and present preliminary evidence of a role in start site selection, though further experiments are needed.

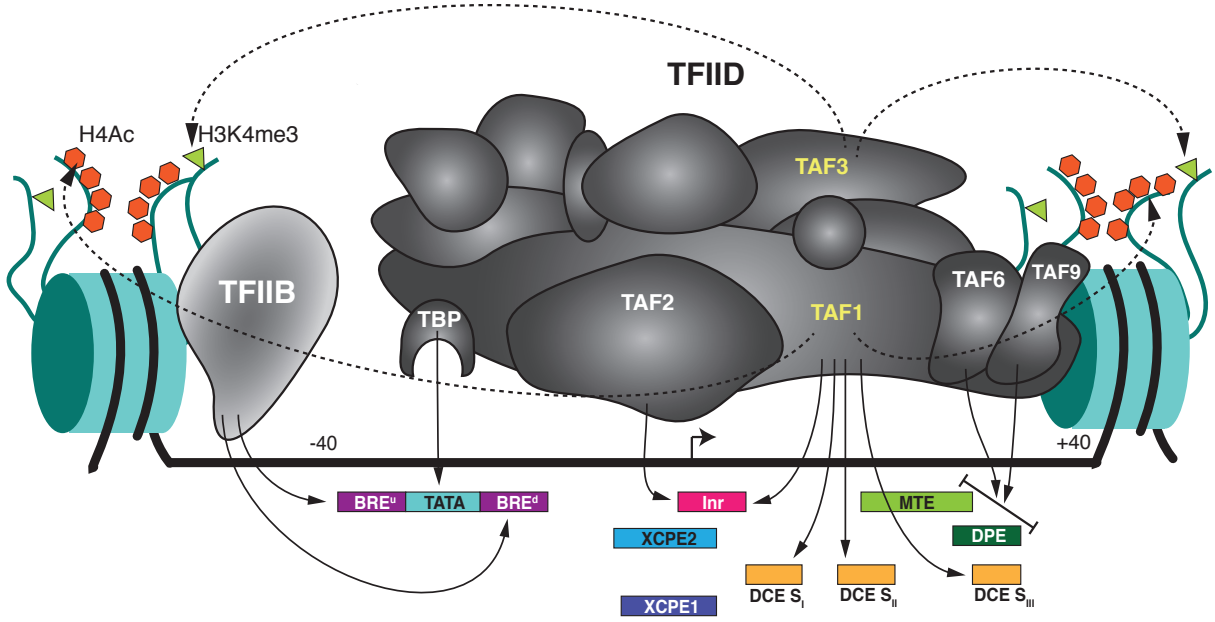


Figure 2.1: *Core Promoter Interactions. The locations of the majority of identified Metazoan core promoter elements are shown to scale with respect to the transcription start site (indicated by a forward arrow). The components of the basal transcription machinery previously shown to interact sequence specifically with the promoter are shown in gray - TFIID and TFIIB. Physical interactions between a core promoter element and a specific protein are indicated by solid, curved arrows. Of the basal transcription machinery, only TFIID has been found to recognize histone modifications found at the promoter. The two subunits of TFIID known to have histone modification recognition domains are labeled in yellow, and the modifications they recognize are indicated with dotted arrows. The promoter proximal nucleosomes are depicted in teal at the approximate location of the average mammalian -1 and +1 nucleosomes.*

2.2 Establishing an in vitro system

With the advent of ChIP-seq, a wealth of genome-wide information has informed our understanding of transcription initiation in eukaryotes and the role of chromatin in this process. However, these experiments can only provide a correlative relationship between histone modifications, promoter architecture, and transcriptional output. Even those experiments that have succeeded in creating mutations to test causality are still complicated by vast networks of uncontrollable and interdependent interactions. Thus, in order to separate out the relationship between nucleosome position, structural relaxation by histone acetylation,

Table 2.1: *Core Promoter Elements in RNA Polymerase II-Transcribed Genes. The information in this table was collected from a number of sources, namely a few noteworthy reviews: [***JuvenGershon:2010hg** *, 181, 195].*

Name	Position	Sequence	Interacting Protein	Types of genes	Frequency	Organisms
BREu	-38 to -32 directly upstream of TATA (if present)	SSRCGCC	TFIIB	identified in TATA containing promoters, but can also be found in TATA-less promoters		conserved from archaebacteria to humans
BREd	-23 to -17 directly downstream of TATA (if present)	RTDKKKK	TFIIB	identified in TATA containing promoters, but can also be found in TATA-less promoters		conserved from archaebacteria to humans
TATA	-30/-31 to -23/-24	TATAWAAR metazoans	TBP	- focused promoters - regulated genes	- strong TATA at ∼20% of yeast and ∼10% of human genome.[196, 197] - recent ChIP-exo experiments suggest that most promoters in yeast actually have a TATA with some deviation from consensus [198].	conserved from archaebacteria to humans
Inr	-2 to +4/5	$TCA_{+1}KTY$ in Drosophila $YYA_{+1}NWYY$ in humans	TAF1, TAF2	- present in focused or dispersed promoters, depending on how much of the sequence is present. - works cooperatively with MTE and DPE	most frequently found core promoter element	- Drosophila - human/mammalian
MTE	∼+18 to +29	CSARCSSAACGS	TAF6, TAF9	- TATA-less genes - functions in conjunction with Inr - can work with DPE		- Drosophila - human/mammalian
DPE	∼+28 to +34	RGWCGTG	TAF6, TAF9	- TATA-less genes - functions in conjunction with Inr - can work with MTE		- Drosophila - human/mammalian
DCE	S_I: +6 to +11 S_{II}: +16 to +21 S_{III}: +30 to +34	S_I: CTTC S_{II}: CTGT S_{III}: AGC	TAF1	- mutually exclusive with DPE - enriched in $TATA^+Inr^+$ promoters		human
XCPE1	-8 to +2	DSGYGGRASM in humans	unknown, but can use TAF-independent TBP	- TATA-less genes - dispersed promoters	1% of human genes (identified in humans)	human
XCPE2	-9 to +2	VCYCRTTRCMY	unknown, but can use TAF-independent TBP	- TATA-less genes - dispersed promoters		human

TAF1 recognition of acetylated histone tails, and transcription output, we turned to a highly purified and fully reconstituted mammalian system in vitro.

2.2.1 Creating a mammalian nucleosome

While many in vivo studies have examined mammalian nucleosome position and modification, the majority of the in vitro model systems have relied on *Xenopus laevis* histones – the histones used in the first crystal structure of the nucleosome, and the first protocol for reconstitution of a nucleosome using highly purified components [7]. *X. laevis* histones express well in bacteria, and because histones are highly conserved, they have become the standard for in vitro reconstitution. In support of their widespread use, structural studies of the nucleosome core particle reconstituted using histones from a number of different species show that the core structure is also highly conserved [9]. However, despite this conservation, there are sequence variations between metazoan histones, including a mammalian specific variant of histone H3 (Histone H3.1) (Figure 2.2). Upon sequence alignment, variable regions can be observed within the histone tails, where the majority of post translational modifications occur, and the only regions that are not visible in many structural studies. Thus, in order to rule out the possibility that mammalian specific interactions would be missed using the traditional *Xenopus laevis* system, I decided to use mammalian histones to create a model system.

Mouse histones, including the mammalian specific variant H3.1, were recombinantly expressed and purified from *E.coli* under the tutelage of Dr. Karolin Luger. At the time, there were few examples of recombinant mouse histone expression and purification, and as such, the best conditions were not standardized. Expression trials showed a wide range of efficiencies, but allowed identification of robust conditions. The final purification schemes for each mouse histone, and the subsequent refolding of the histone octamer are presented in Appendix B.

In order to precisely position a nucleosome for in vitro transcription, the Widom 601 sequence was chosen [33]. The ability of the 601 to assemble a stable and well positioned nucleosome using mouse histones was first confirmed using the minimal 147 bp sequence (Figure 2.3). Standard salt gradient dialysis of the purified mouse histone octamer with the DNA was used, and the concentration of the histone octamer was titrated relative to the DNA in order to find the optimal assembly ratio (Figure 2.3–A,B). Reactions were loaded onto a native PAGE gel and stained first with ethidium bromide (EtBr) to visualize the location of the DNA, and then with the protein stain, Page Blue to locate the histones. Nucleosome assembly is observed as a shift in the migration of the DNA to a higher position and its colocalization with histones (Figure 2.3–C). At undersaturating histone octamer concentrations, two species of nucleosomes are formed. This odd behavior is observed by the Luger lab, though not readily reported on. As the histone octamer is titrated to concentrations beyond a 1:1 molar ratio with DNA, a single position is favored, creating a fully assembled, homogeneous nucleosome. Thus, mouse histones are capable of forming a homogeneous nucleosome in vitro.

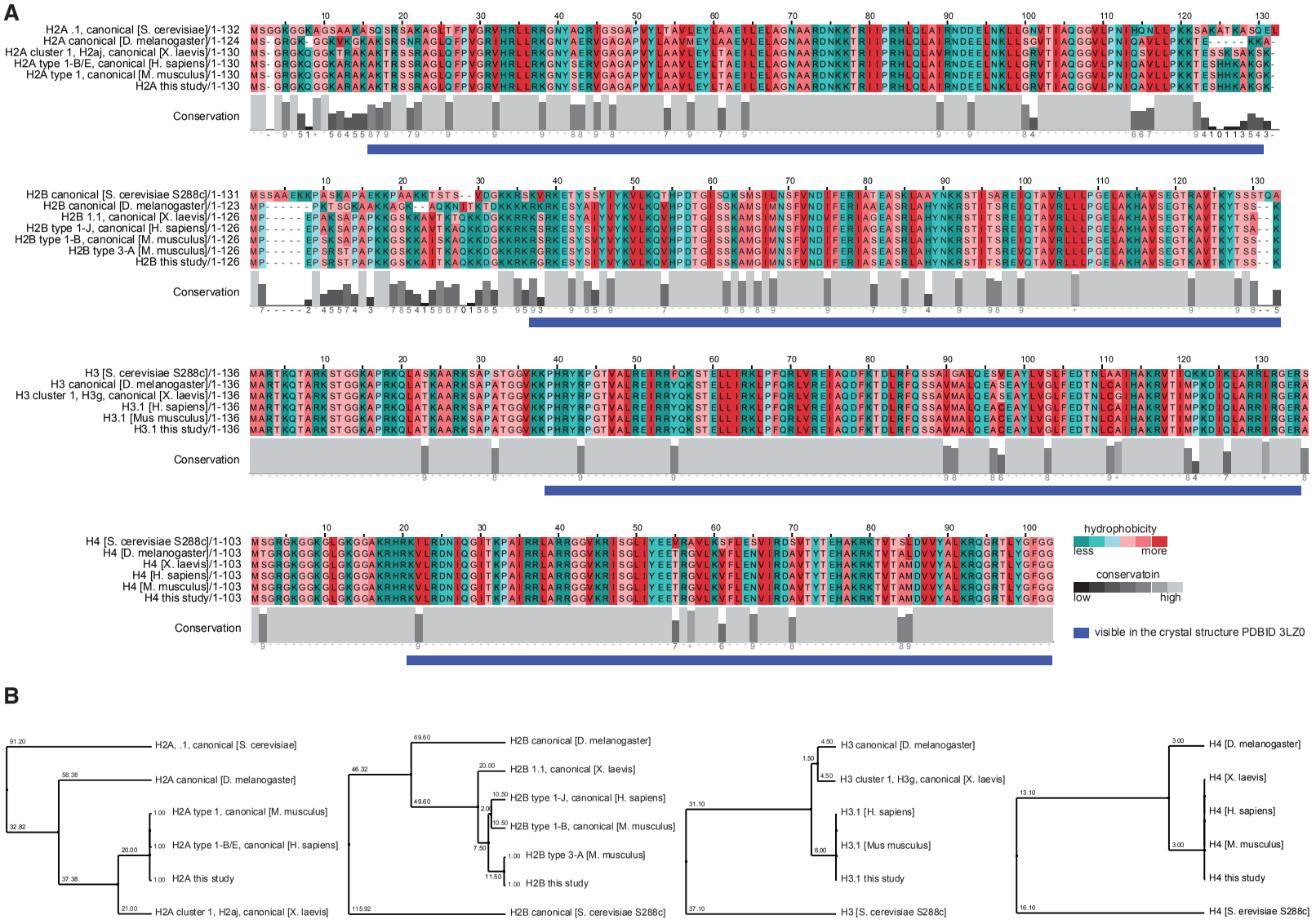


Figure 2.2: *Multiple sequence alignment of histones.* ***Continued on next page***

Continued Figure 2.2: ***(A)*** *Sequence alignment of canonical histones (as well as mammalian specific histone H3.1) with the histone sequence used in this study. The alignment was performed using Mafft with accuracy settings G-INS-i. Amino acids are colored according to their hydrophobicity, and conservation is shown in grey below. The regions appearing in the crystal structure, PDBID 3LZ0, are indicated with a blue bar.* ***(B)*** *Trees were calculated using BLOSUM62.*

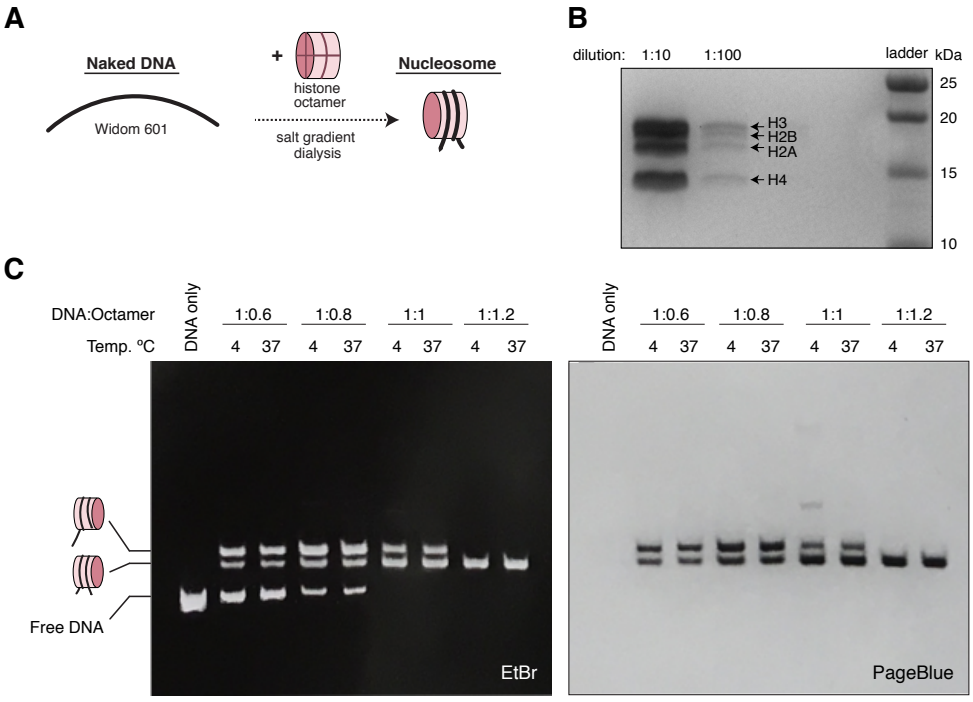


Figure 2.3: *Assembling a minimal mouse nucleosome.* ***(A)*** *Diagram of the method used to reconstitute a nucleosome using purified mouse histone octamer and the strong nucleosome positioning sequence, Widom 601. Salt gradient dialysis was used to deposit the octamer onto the DNA.* ***(B)*** *18% Bis-Tris-Mes PAGE gel of the purified mouse octamer used to reconstitute nucleosomes in this study. Correct stoichiometry can be observed by the relative staining of each histone.* ***(C)*** *Native PAGE gel showing the results of assembling a minimal mouse nucleosome using the scheme in (A). The gel was first stained with ethidium bromide (EtBr) and imaged using UV light, then stained with the protein stain, PageBlue, and imaged in transmitted light. Histone octamer was titrated to find the best condition, and assemblies were incubated at the indicated temperature post assembly at 4° C.*

Next, a transcription template was designed to position a nucleosome in the -1 and +1 position relative to the promoter (Figure 2.4–A). Most studies performing in vitro transcription on nucleosomal templates utilize chromatinized plasmids. This approach is able to faithfully recapitulate the dependence of transcription on chromatin remodeling [199], but does not allow precise control of nucleosome position, nor does it allow separation of the roles of each nucleosome in the transcriptional output. Thus, in order to fully dissect the relationship between TAF1 and acetylated H4 tails at the promoter, I chose to utilize a DNA fragment containing a single nucleosome in either the -1 or +1 position. With this strategy, the impact of acetylation at each nucleosome could be tested. Therefore, two constructs were designed, one with the 601 sequence inserted upstream of the promoter, and one with it inserted downstream. To control for sequence effects of the different surrounding sequences at each position, a short, previously characterized flanking sequence was included.

The well characterized and strong Super Core promoter was chosen as the promoter for the nucleosomal transcription templates [182]. However, since transcription in vivo often occurs on promoters with far fewer and weaker core elements than the Super Core, constructs containing mutations in either the TATA, initiator (Inr), DPE, or all three were made (Figure 2.4–B). By systematically mutating elements in the core promoter, their ability to compete with acetylated histone tails for binding by TAF1/TFIID and for orienting the pre–initiation complex on the promoter could be tested.

One possible outcome of weak or absent core promoter elements (CPEs) is an increased reliance on the position of nucleosomes to direct the transcription start site. I hypothesized that in the absence of orienting CPEs, positioning of the pre–initiation complex for start site selection could be compensated by the TAF1–acetylated H4 tail interaction. To test this, I created additional templates in which the nucleosomes were moved further away from the encoded TSS. If the TAF1–acetylated H4 tail interaction can direct start site selection, the resulting transcript should be shifted in accordance with the location of the nucleosome. However, longer templates are capable of assembling multiple nucleosomes and increase heterogeneity. Thus, in order to maintain only a single, homogenous nucleosome on each transcription template, I used a nucleosome discouraging sequence with which to space the the nucleosome away from the promoter (Figure 2.5). This sequence was taken from the Firefly luciferase gene from the vector pGL3-Basic in a region with low predicted occupancy as assayed by two separate nucleosome predicting algorithms. Importantly, because this vector is commonly used to test genomic sequences for promoter activity, it is known to have no detectable cryptic promoters, and thus should not direct transcription on its own.

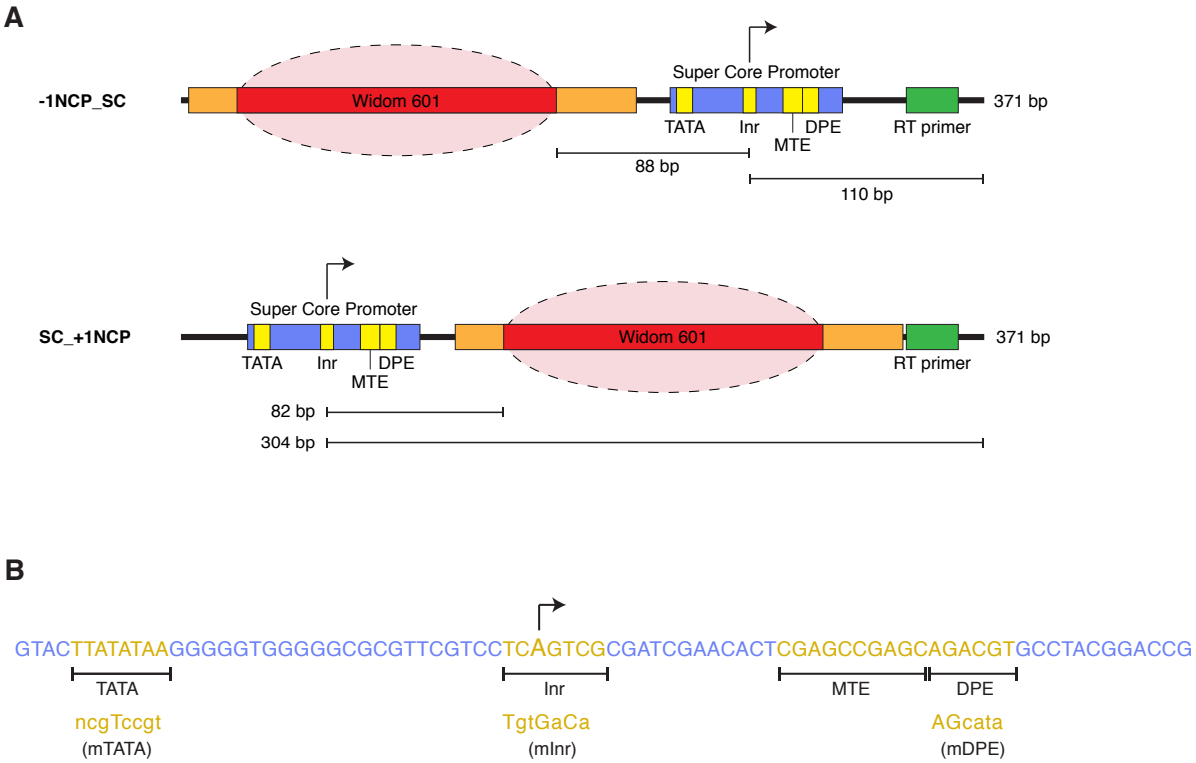


Figure 2.4: *DNA constructs for reconstituting a +1 and -1 nucleosome on a transcription template.* ***(A)*** *Transcription templates were designed to position a nucleosome either upstream (-1) or downstream (+1) of the transcription start site. A standard reverse transcription (RT) primer sequence was inserted at the end of each construct for visualizing in vitro transcription products using a* ^{32}P *primer. To control for effects of sequence context on the nucleosome, the Widom 601 sequence was flanked with a previously characterized sequence. The pink oval indicates the location of the histone octamer in the assembled nucleosome templates. The depicted constructs were created by cloning into a pBlueScript backbone.* ***(B)*** *The sequence of the Super Core promoter used in this study. The location of the first transcribed nucleotide is indicated by an arrow and a large 'A'. Mutations to the core promoter elements used are shown below and denoted with a lowercase m.*

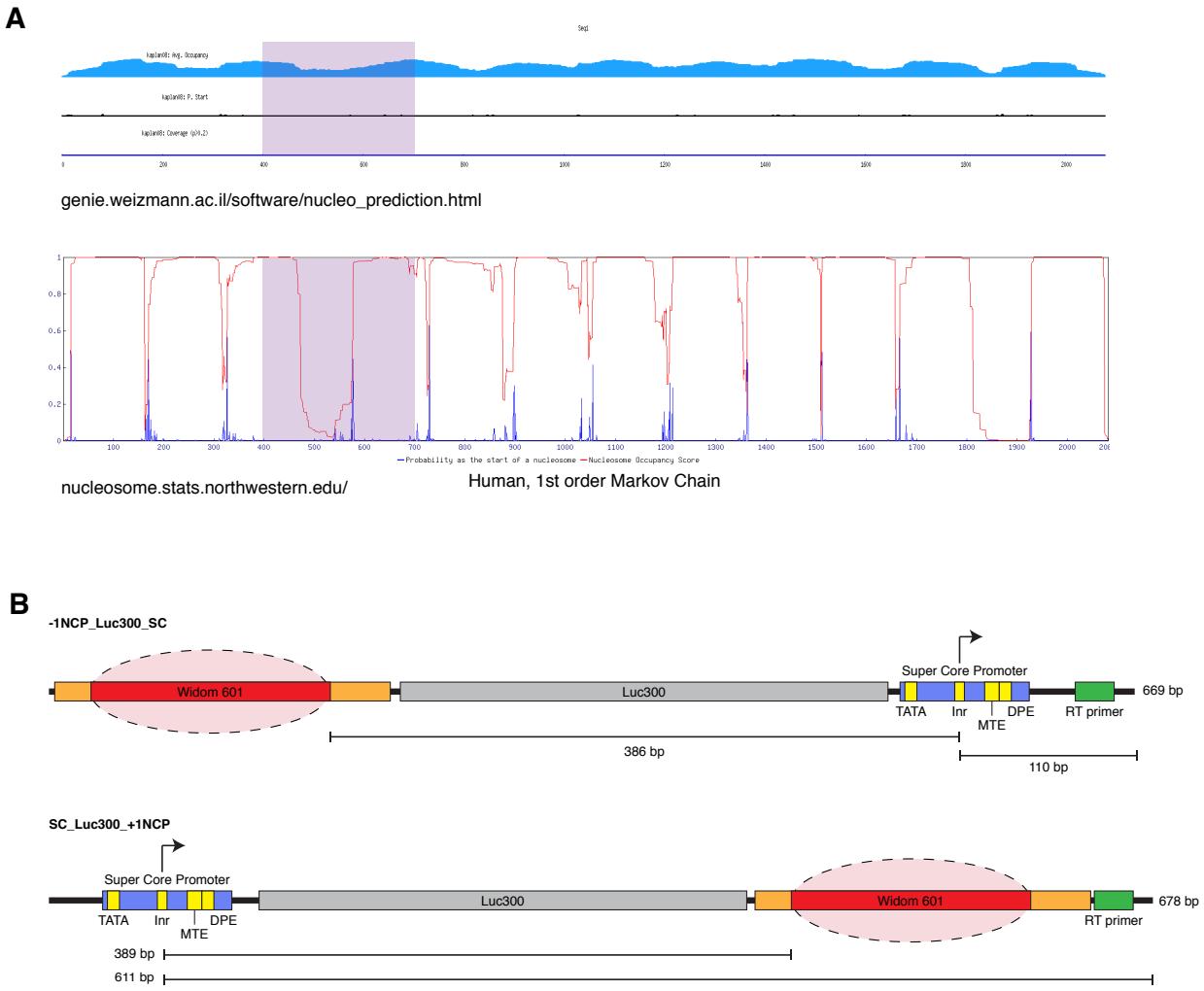


Figure 2.5: *Selection of a nucleosome spacing sequence.* ***(A)*** *The Firefly luciferase gene from the pGL3-Basic vector was analyzed using to two nucleosome occupancy predicting algorithms. The resulting analyses are shown. Top: predicted nucleosome occupancy is shown in blue along 1-2081bp of the luciferase gene. Bottom: The probability of that position as a start of a nucleosome is shown in blue, while the nucleosome occupancy score is shown in red. The region of particularly low predicted occupancy (highlighted in purple) in both algorithms was chosen as the nucleosome spacing sequence to use in this study.* ***(B)*** *Diagram of the transcription constructs resulting from inserting the chosen 300 bp section of the luciferase gene from (A) between the nucleosome positioning sequence and the promoter.*

Nucleosome assembly was subsequently tested on these fragments. During their creation, the templates were cloned into pBlueScript II SK+ backbones, and fragments for assembling nucleosomes were created by large scale PCR. Using the same salt gradient dialysis method as in Figure 2.3, templates were mixed with purified mouse histone octamer, and the resulting assemblies were assayed by native PAGE (Figure 2.6). Surprisingly, none of the DNA:octamer ratios tested yielded a positionally homogenous nucleosome. Although the shortest templates (represented by -1NCP_SC in Figure 2.6–A) do contain enough DNA to accommodate two tightly spaced nucleosomes, it was expected that undersaturating conditions, incubated at a warm temperature to favor the most thermodynamically stable position, would result in a single species, even if some of the DNA remained un–assembled. However, this was not observed. Additionally, despite the nucleosome discouraging sequence from the luciferase gene, the longer templates (represented by SC_Luc300_+1NCP in Figure 2.6–B), included even more nucleosome species at undersaturating octamer concentrations.

In an effort to shift the nucleosomes into a homogeneous population, assemblies were incubated at a variety of temperatures. Heat shifting has been used with endogenous nucleosome sequences to mobilize weak positions and populate the most thermodynamically stable state [200]. As seen in Figure **??**–A, even temperatures as high as 70° C resulted in multiple positions on the template. To determine if the inability to heat shift the templates was due to a property of the 601 sequence itself, the minimal 601 nucleosome was assembled at undersaturating conditions and subjected to the same heating regime. An additional incubation at 95° C was included as a control to strip the histones from the DNA (**??**–B). Surprisingly, even when incubated at 70° C, the minimal nucleosome occupied two positions on the 601. Thus, it seems that assembly of a homogeneous nucleosome on templates containing the 601 sequences cannot be achieved using either titration of histone octamer, or heat shifting. To circumvent this limitation, I set out to identify and purify the desired nucleosome species from this heterogeneous mixture.

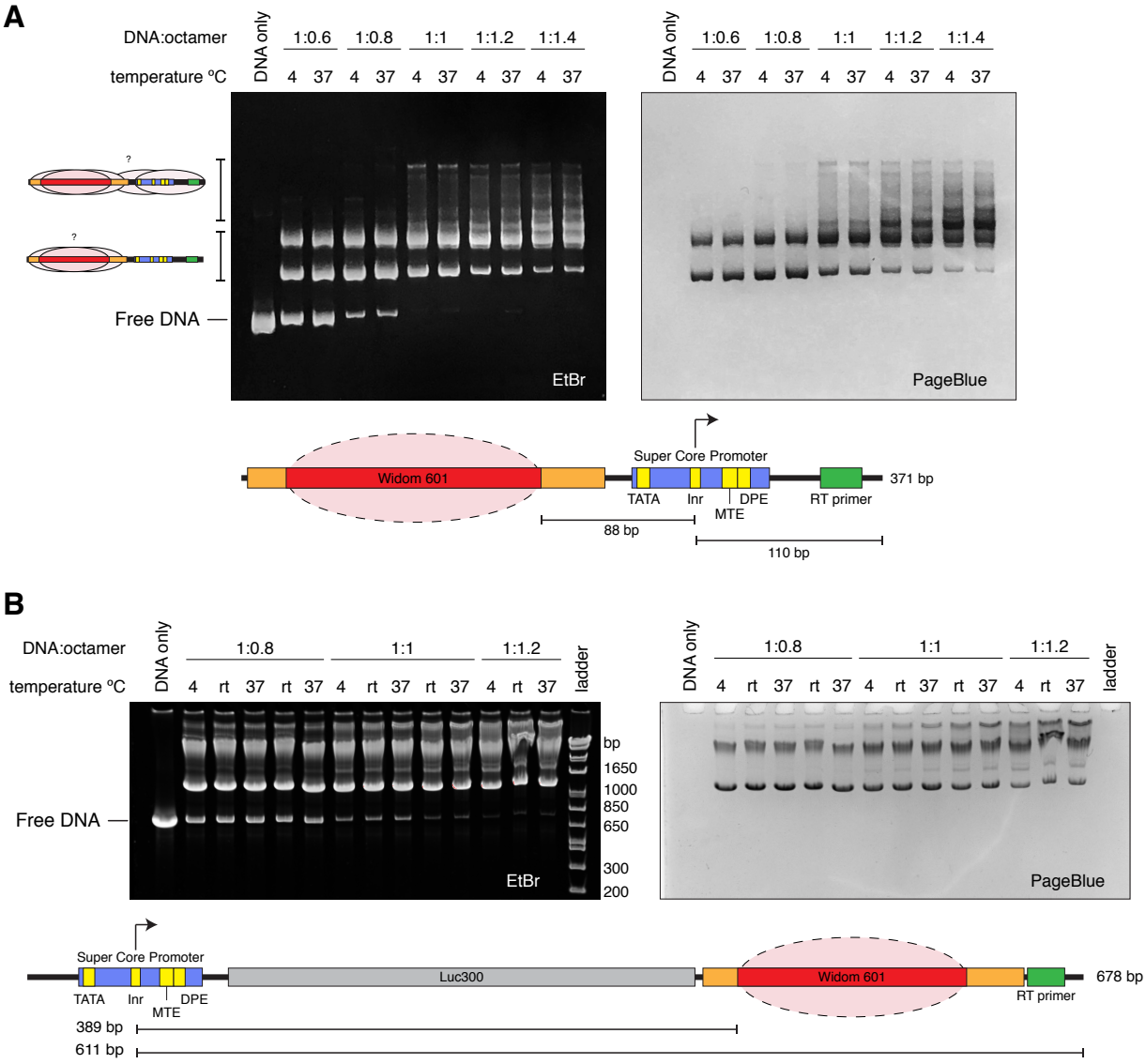


Figure 2.6: *Assembling nucleosomes on transcription templates leads to multiple species. Nucleosomes were assembled on the short -1NCP_SC* ***(A)*** *and the long SC_Luc300_+1NCP* ***(B)*** *transcription templates using the scheme described in Figure 2.3. Native PAGE gels show the results of nucleosome assemblies. The gels were first stained with EtBr and imaged using UV light, then stained with the protein stain, PageBlue, and imaged in transmitted light. Histone octamer was titrated, and assemblies were incubated at the indicated temperature for 1 hour post assembly at 4° C. Multiple species can be observed, with decreased mobility as octamer concentration increases.*

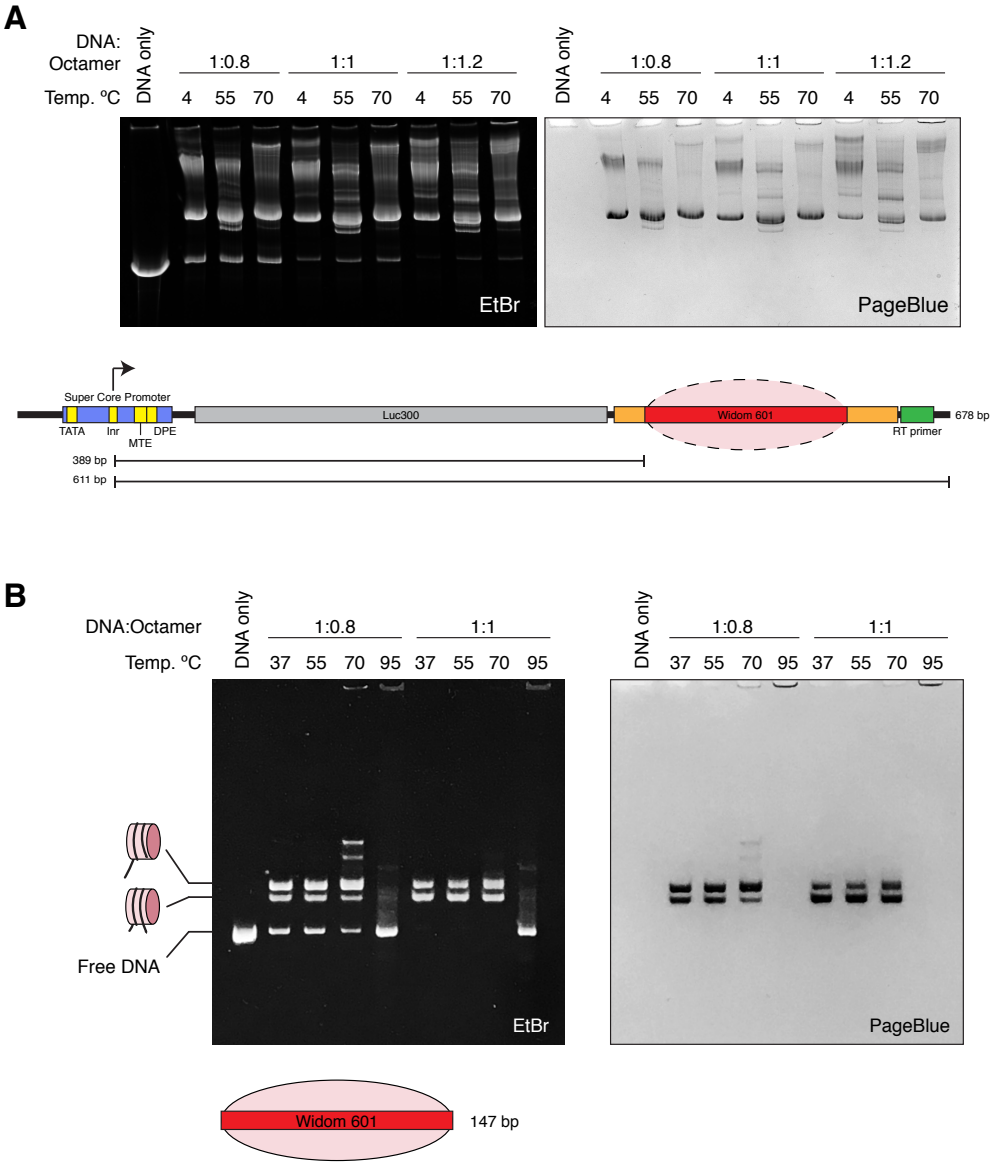


Figure 2.7: *Templates containing the Widom 601 sequence cannot be heat shifted. Native PAGE gels show the results of heating nucleosome assembly reactions to shift occupancy to the most thermodynamically stable position.* ***(A)*** *Nucleosome assemblies using the long SC_Luc300_+1NCP template were heated to the indicated temperatures for 1 hour post assembly at 4° C. Population of only one species could not be induced by heating as observed by multiple bands for each temperature and DNA:octamer ratio.* ***(B)*** *Same as in (A) but using only the minimal positioning sequence. Assemblies heated to 95° C were heated for 15 minutes rather than an hour. The 601 sequence cannot be heat shifted to populate a single nucleosome position as indicated by the prevalence of two (or more) histone–containing bands.*

Before the correct nucleosome species could be purified, it had to be identified. To achieve this, restriction enzyme accessibility was assayed at different sites along the template (Figure 2.8). First, nucleosomes were assembled on the short -1NCP_SC template at undersaturating octamer conditions (where there were the fewest nucleosome species), and each species was purified by extraction from native PAGE gel slices under non-denaturing conditions. Then, the purified species were incubated with XbaI and the resulting cleavage was assayed by native PAGE (Figure 2.8–A). Cleavage of the naked DNA template resulted in two fragments of different sizes – one containing the 601 sequence, and the other containing the promoter. If either of the two nucleosome species occupied the promoter half of the template, this cleavage product would remain shifted (bound by a nucleosome) in the native gel. Additionally, if either species spanned the XbaI site, the presence of the nucleosome would protect the DNA from cleavage, and neither of the cleavage products would be present in the native gel. This experiment shows that both species, termed A and B, are located away from the XbaI site, on the half of the template containing the 601 sequence.

To more precisely map the location of the nucleosome in species A and B, restriction enzyme accessibility was assayed at sites within and directly flanking the 601 sequence (2.8–B). Unpurified nucleosome assemblies were subjected to restriction digest, then loaded on a native PAGE gel. If the nucleosomes of either species were assembled over a restriction site, they would be expected to resist digestion and continue to migrate at the same rate as the undigested species. In contrast, if either species contained a nucleosome outside of a restriction site, the DNA would be digested and the resulting shorter nucleosome would be expected to run at a faster migration in a native gel. Additionally, when treated with proteinase K and extracted with phenol/chloroform/isoamyl alcohol to remove the histones, the resulting free DNA would be visible as digestion products. Surprisingly, assaying restriction sites within the 601 sequence resulted in full protection from digestion for both species A and B (Figure 2.8–B, left gel), while assaying sites flanking the 601 lead to full digestion of both species (Figure 2.8–B, right gel). This result suggested that there was only a small difference in nucleosome position between species A and B.

The piece of evidence that indicated which species was the correctly positioned nucleosome came as an accidental result. After gel purifying species A and B, each was dialyzed into a storage buffer over night, then run on a native gel (Figure 2.9). Purified species A remained a single species nucleosome after dialysis, while the purified species B repopulated the positional states of the unpurified assembly. Therefore, while both species are positioned over the 601 sequence, only species A forms a stable nucleosome. Thus, species A was chosen for transcription and acetylation experiments.

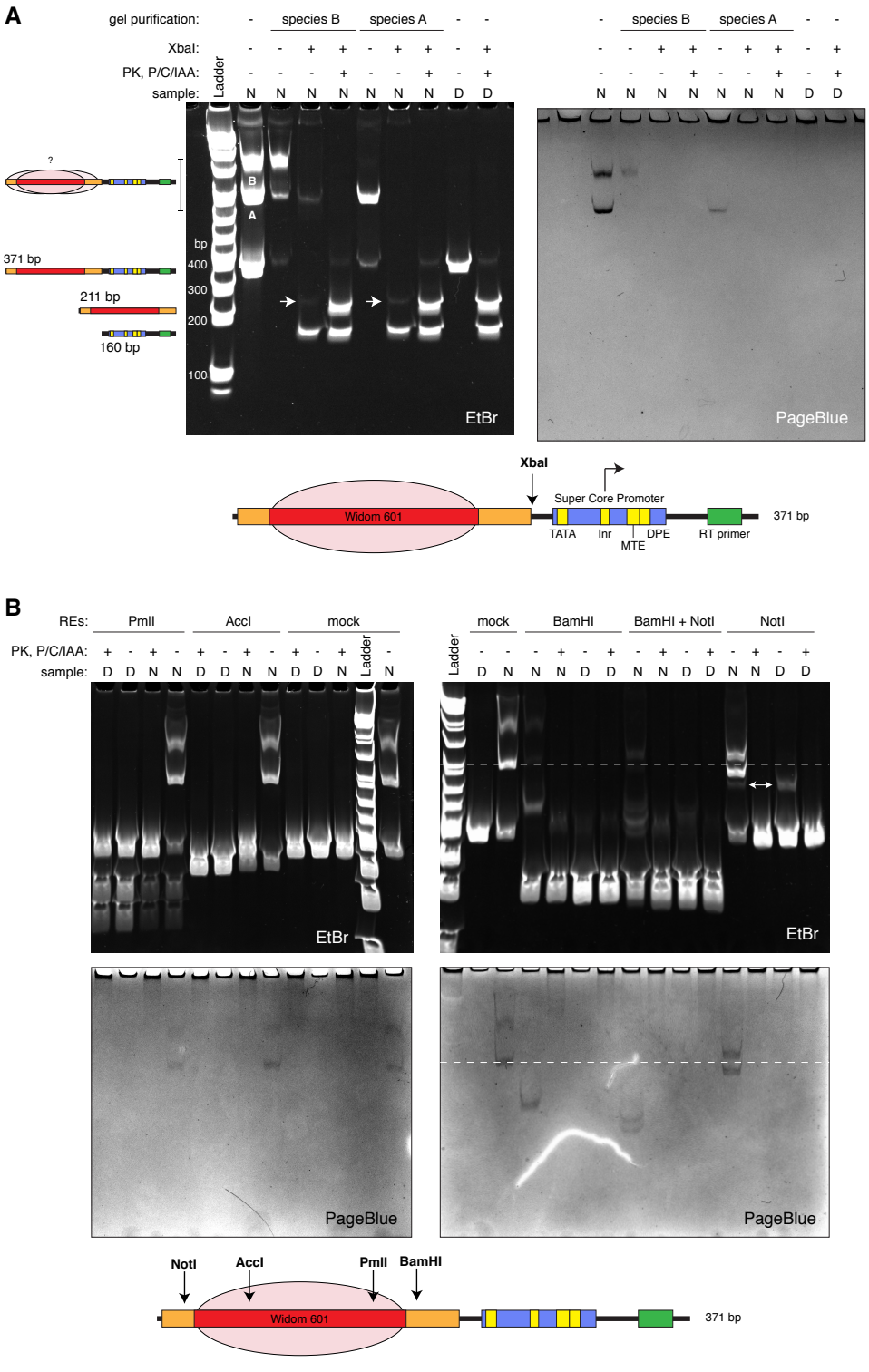


Figure 2.8: *Determining the correct nucleosome species by restriction enzyme accessibility assay.* *Continued on next page.*

Continued Figure 2.8: ***(A)*** *Above: Native PAGE gels showing the results of restriction enzyme accessibility assays on the template shown below either naked (D) or assembled into nucleosomes using undersaturating histone octamer concentrations (N). Prior to digestion by XbaI, the two major nucleosome species labeled as 'A' and 'B' on the gel were purified in their native form by gel extraction where indicated. The digestions were loaded onto a Native PAGE gel either directly, or after removal of histones from the sample using Proteinase K treatment followed by phenol/chloroform/isoamyl alcohol extraction. The location where the 601-containing restriction fragment would be if it were not bound by histones is indicated by arrows. The absence of this band, but the presence of free promoter fragment for both species A and B suggest that both species are composed of a nucleosome formed on the 601-containing half of the template, and not the promoter half. Below: diagram of the template used to assemble the nucleosomes with the location of the XbaI restriction site indicated with an arrow.* ***(B)*** *Same as in (A) except gel purification of nucleosome species was not performed prior to digestion with restriction enzymes (REs). The dotted line indicates the location species A runs in the gel. The double headed arrow points to the shift present in the DNA only digest, suggesting this is the restriction enzyme shifting the DNA. The ethidium bromide image brightness threshold was set to allow visualization of very weak bands.*

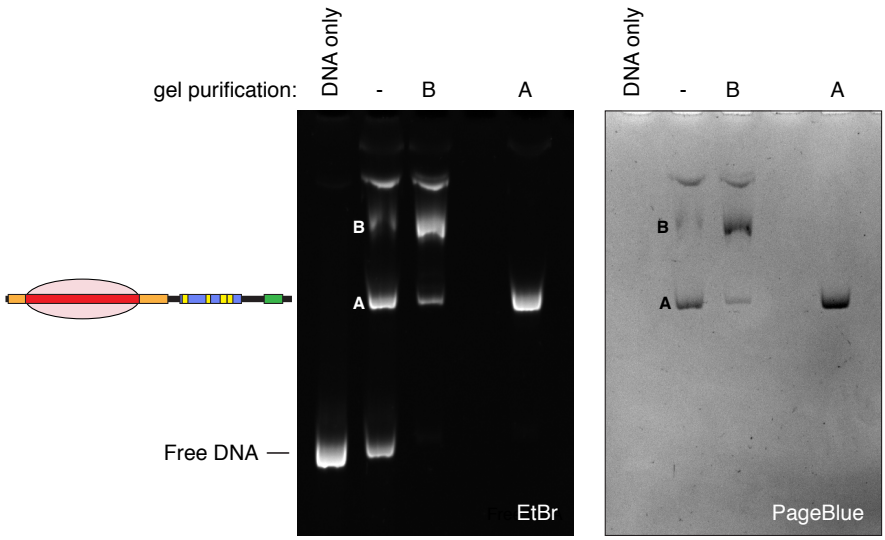


Figure 2.9: *Stability determines the correct nucleosome species. Species A and B were gel purified in their intact form and dialyzed at 4° C into storage buffer consisting of 20 mM Tris–HCl pH 7.5, 1 mM EDTA, and 1 mM DTT. The resulting purifications were loaded onto a native PAGE gel the next day. The purified species B can be seen to reform some of species A after gel purification, whereas when purified, species A does not repopulate species B, suggesting that it is more stable.*

2.2.2 Optimizing transcription

Before the nucleosome templates could be used in in vitro transcription assays, conditions for optimal transcription on a DNA fragment needed to be identified. Most in vitro transcription assays are performed on plasmids, and thus the precise number of base pairs required to form a landing pad large enough to allow proper assembly of the pre–initiation complex was not already apparent. The nucleosome templates modeling the downstream, +1 nucleosome, in particular needed to be tested for their ability to direct uninhibited transcription from a promoter adjacent to a DNA end. To address this question, templates with different numbers of base pairs upstream of the TATA box were created by PCR from the plasmid containing the construct SC_+1NCP (Figure 2.10–A). The resulting DNA fragments were transcribed using purified human basal transcription factors and purified human RNA polymerase II. To visualize the transcripts, a ^{32}P labeled primer was used in reverse transcription, and the reaction was run on a denaturing Urea-PAGE gel. Signal from the ^{32}P labeled reverse transcripts were imaged using a storage phosphor screen and imager. Additionally, since plasmids contain DNA that is not transcribed but is available to interact with the transcription machinery, they contain their own nonspecific competitor. The template fragments, on the other hand, do not. Thus, the presence of a nonspecific competitor DNA, dGdC, was tested.

The results suggest that a landing pad as small as 7 bp upstream of the TATA box is sufficient to drive robust transcription of a DNA fragment (Figure 2.10–B). Surprisingly, when regions past 28 bp upstream of the TATA box were included, the transcription output dropped dramatically and was comparable to transcription from the template in plasmid form (compare lanes 5–10 to 1–4 and 11–13 in Figure 2.10–B). Further investigation concluded that the observed decrease in transcription is due to a repressive element in the pBlueScript II SK+ backbone used to create the templates (Supplementary Figure A.2). Thus, a landing pad of 28 bp or less upstream of the TATA was chosen. Additionally, while the nonspecific competitor dGdC reduced overall transcription output, it reduced nonspecific transcripts to a greater degree, and was thus included in subsequent transcriptions where indicated.

Next, the dynamic range of transcription was tuned to allow observation of both increases and decreases in output (Figure 2.11). Using a template length found in Figure 2.10 to direct robust transcription, the concentration of RNA Polymerase II (Pol II) was titrated (Figure 2.10–A,B). No change in transcription output was observed, indicating that Pol II was not limiting at the concentrations tested. Because the goal was to test the effect of an interaction between TFIID and acetylated histone tails on transcription, it was important that the availability of Pol II in solution not be rate limiting. Thus, the lowest concentration of Pol II tested was chosen, and the dynamic range of the basal factors was tested (Figure 2.10–C). The results pointed to an intermediate level of basal factors that could be employed while still maintaining the dynamic range of the basal factor machinery and of transcript detection.

Once conditions were chosen to be within the dynamic range of transcription, the number of base pairs upstream of the TATA box was re-tested. Additionally, the number of base

pairs needed downstream of the RT primer annealing site to prevent underestimation of transcription output due to RNA degradation or incomplete transcription was tested (Figure 2.12). Based on the results, a final construct architecture of 35 bp upstream of the TATA and 13 bp downstream of the RT primer binding site was chosen.

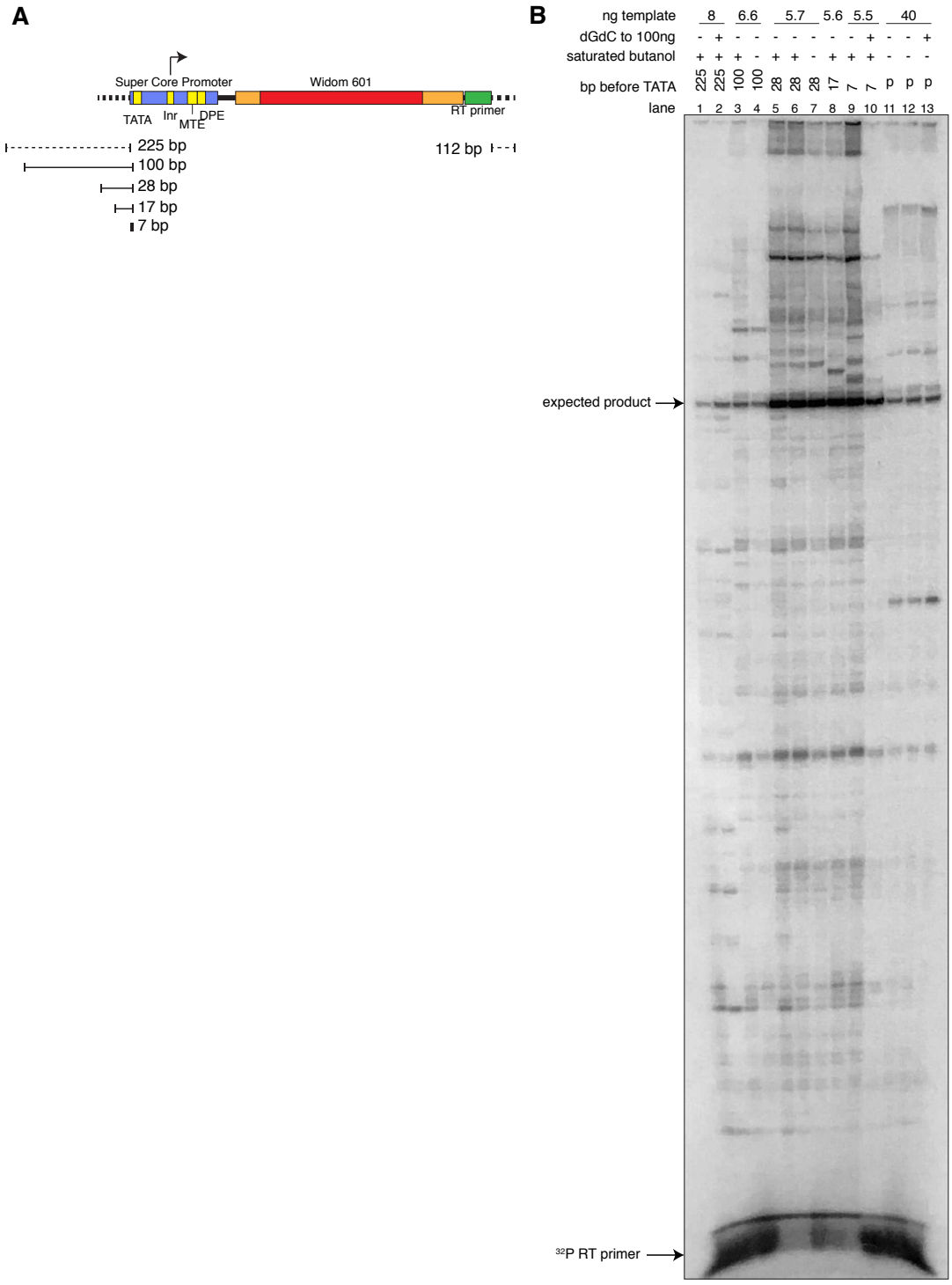


Figure 2.10: *Comparing transcription using a DNA fragment versus a plasmid reveals a repressive sequence element and the number of base pairs needed upstream of the promoter. Continued on next page.*

Continued Figure 2.10: ***(A)*** *Diagram of the transcription templates created using PCR from the construct p-1NCP_SC modeling a -1 nucleosome. PCR primers were chosen to produce fragments containing the indicated number of base pairs upstream of the TATA box and downstream of the RT primer binding site.* ***(B)*** *Phosphor image of a Urea-PAGE gel showing the results of transcription from the indicated templates. After incubation with purified human basal factors and RNA polymerase II at 30° C for 30 minutes, transcripts were reverse transcribed (RT) using a* ^{32}P *labeled primer and loaded on a denaturing Urea-PAGE gel. DNA purification conditions using saturated butanol are also tested, as well as the use of a DNA competitor, dGdC. The number of moles of template are the same in each reaction. Lanes 11–13 show transcription results using the construct diagramed in (A) in plasmid form.*

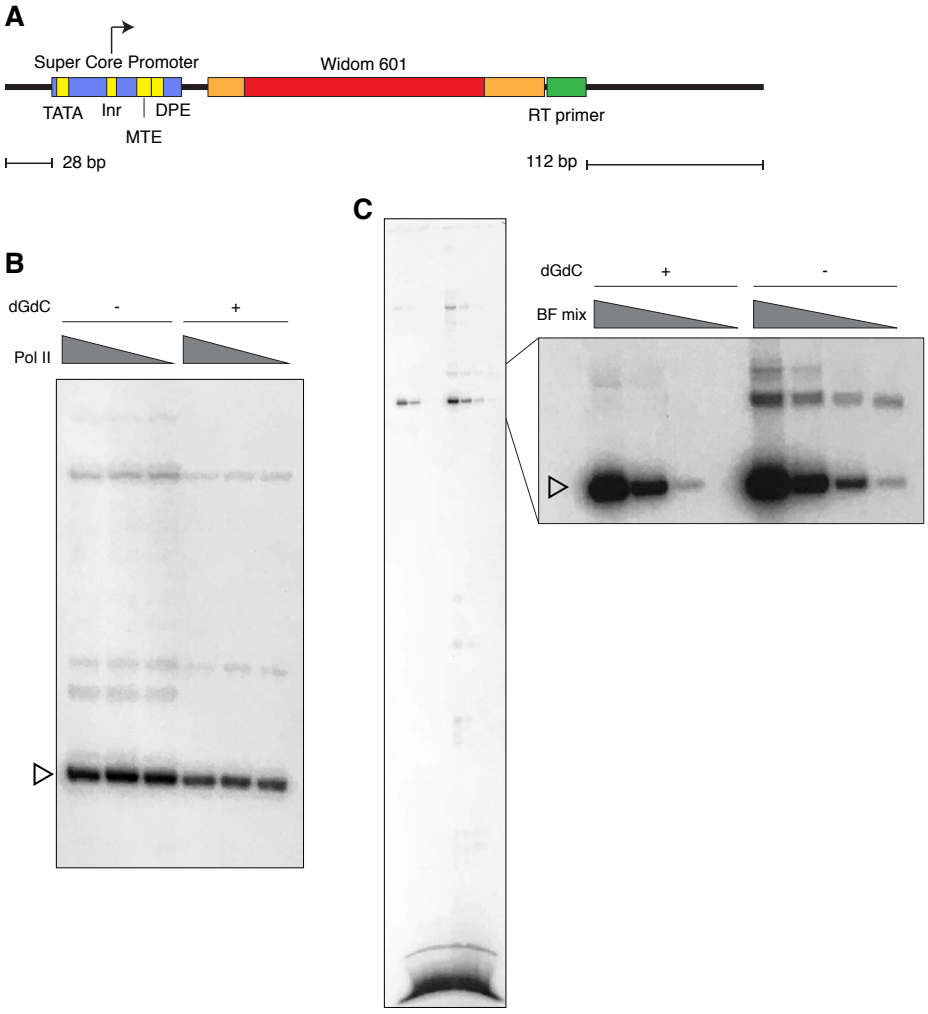


Figure 2.11: *Optimizing transcription on a fragment by titrating basal factors and RNA polymerase II.*

Continued on next page.

Continued Figure 2.11: ***(A)*** *Diagram of the DNA fragment (produced by PCR) used as a transcription template in (B) and (C).* ***(B)*** *Phosphor image of a Urea-PAGE gel showing the results of transcription from the template in (A). The amount of RNA polymerase II (Pol II) used in each reaction was titrated in two-fold steps to 1/4 the original amount. The effect of the DNA competitor, dGdC was also assessed. Each reaction contained 10 fmoles of template, and 50ng of dGdC where indicated.* ***(C)*** *Phosphor image of a Urea-PAGE gel showing the results of transcription from the template in (A) while titrating the amount of basal transcription factors used. The basal factor mix (BF mix) was titrated using a 3-fold serial dilution. The lowest concentration of RNA polymerase II used in (B) was selected for the reactions in (C). All other conditions were the same as in (B). Open triangles point to the expected transcript size. In order to be within the dynamic range, the basal factor concentration chosen for future transcription reactions was 1/3 the original amount.*

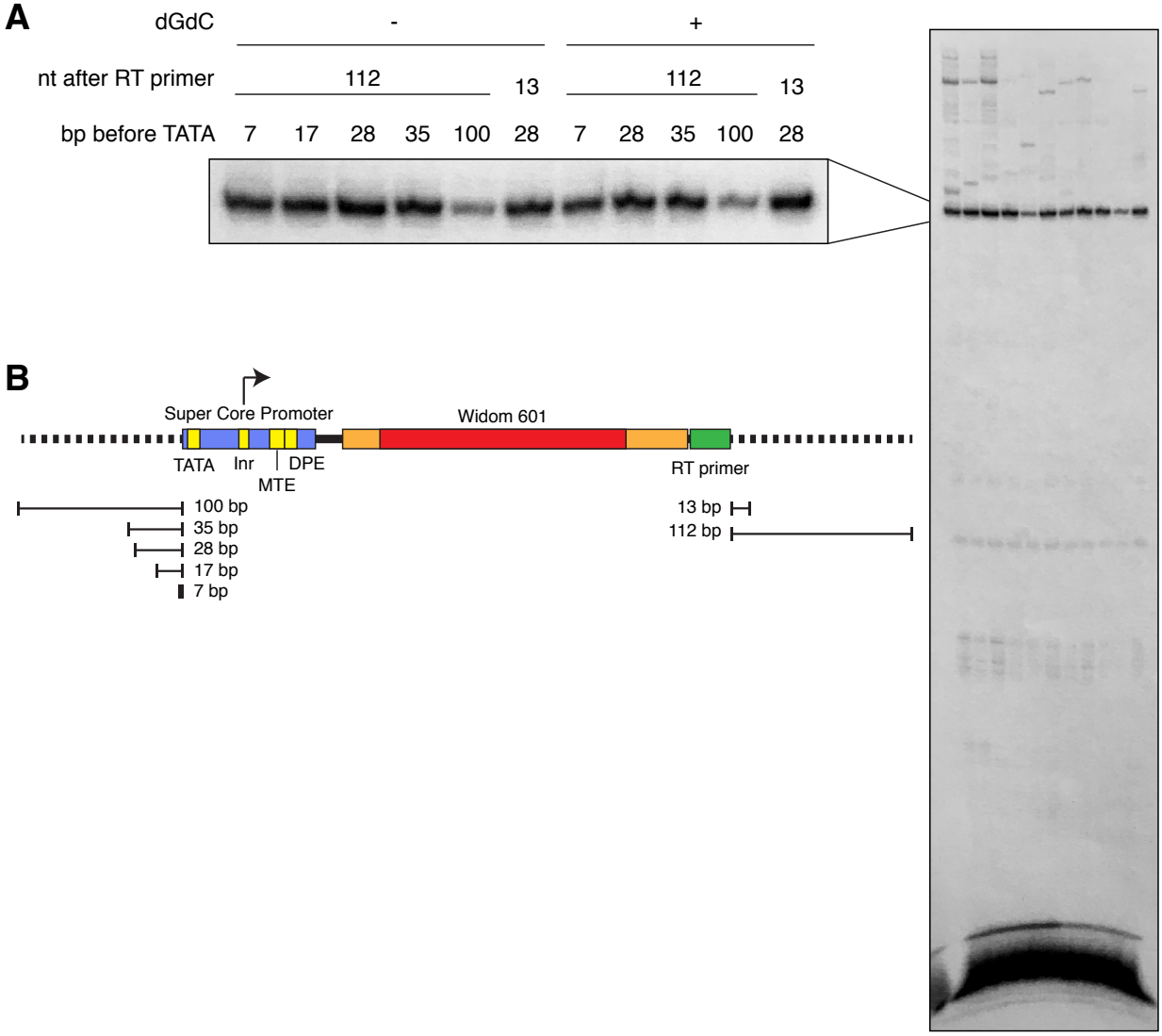


Figure 2.12: *Selecting final transcription conditions. Using the RNA polymerase II and Basal factor concentrations identified in Figure 2.11, the number of base pairs from the TATA box and the RT primer binding site required for efficient transcription were re-tested.* ***(A)*** *Phosphor image of a Urea-PAGE gel showing the results of transcription from the templates diagramed in* ***(B)****. As in Figure 2.10, when sequences beyond 28 bp upstream of the TATA box are included, transcription is repressed, suggesting the presence of a repressor at that site. Each reaction contained 10 fmoles of template, and 50ng of dGdC where indicated.*

2.2.3 Establishing an acetylation procedure

To test the effects of H4 acetylation on transcription initiation in vitro, a mechanism of site specific acetylation had to be chosen. For this purpose I recombinantly expressed and purified yeast Piccolo NuA4 following the method by Song Tan [201] (Figure A.1). NuA4 is the complex that acetylates H4 and H2A in vivo and is the only essential HAT in yeast [201, 202]. Its catalytic subunit, Esa1, is homologous to human Tip60 which performs the same function. To characterize the activity of the purified Piccolo NuA4, I first tested its ability to acetylate free mouse H4 histone using ^{14}C labeled acetyl coenzyme A (AcCoA) (Figure 2.13–A). Reactions were incubated from 40 seconds to one hour, before loading on an acid–urea gel to separate the different degrees of acetylation. The resulting gel was imaged using a storage phosphor screen and scanner, and then stained with PageBlue to visualize the unacetylated histone. Complete acetylation of histone H4 was complete soon after 1 minute, and controls demonstrated that Piccolo NuA4 had no detectable self modification activity that could complicate later experiments. Next, the short -1NCP_SC template was assembled into a nucleosome, gel purified, and assayed for site specific acetylation by Piccolo NuA4 (2.13–B,C). As previously reported using *Xenopus laevis* histones [203], Piccolo NuA4 demonstrated efficient and rapid acetylation of mouse H4 within the nucleosome, but also showed H2A acetylating activity (as determined by location in acid-urea gel and previous reports [203], Figure 2.13–D). This activity was present on a much longer timescale, and could therefore be disfavored by incubation at shorter timescales. In this section I have shown that using in vitro assembly of nucleosome using mouse histone octamers on transcriptional template, followed by gel purification, and site specific acetylation with a purified multi-subunit complex, I can create a highly pure, homogenous, and controlled system with which to test the interaction with TAF1 and its effect on transcription.

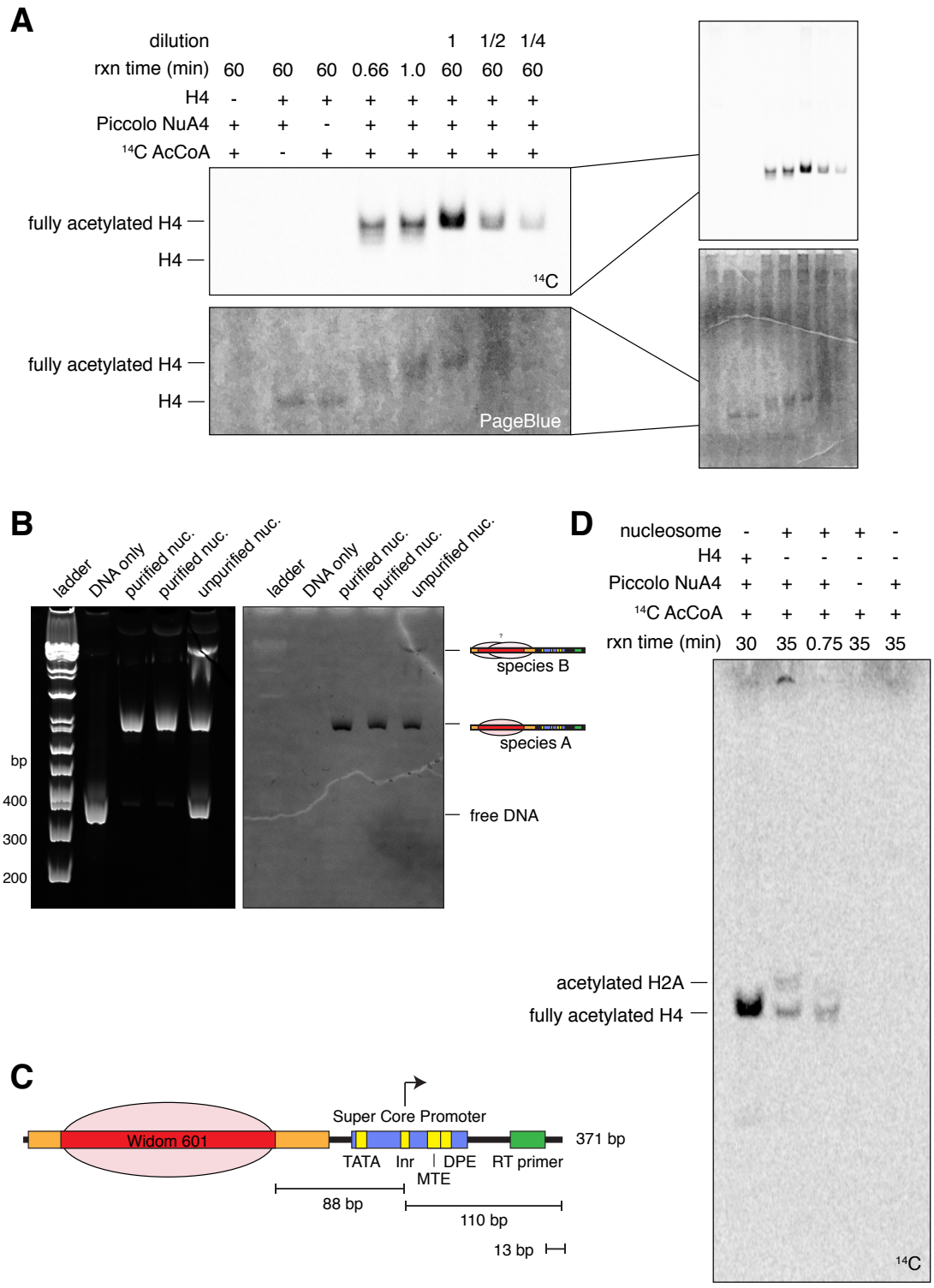


Figure 2.13: *Characterization of Piccolo NuA4 acetylation of mouse histone and nucleosomes. Continued on next page*

Continued Figure 2.13: ***(A)*** *Acid–urea PAGE gel showing the results of an acetylation trial on free mouse histone H4. Reactions were incubated at room temperature for the indicated time and then quenched using gel loading buffer containing 8M urea before being loaded on an acid–urea gel to resolve individual acetylation sites.* ^{14}C *labeled acetyl coenzyme A (AcCoA) was imaged using a storage phosphor screen and scanner, then stained with PageBlue to visualize the unacetylated histone. In the last lanes, the reactions were diluted by the indicated amount with loading buffer before being loaded on the gel.* ***(B)*** *Assembly and purification of the short -1NCP_SC template diagramed in* ***(C)****.* ***(D)*** *Acetylation assay performed as in (A), except using the assembled nucleosome template from (B). Free H4 was acetylated and loaded adjacent to the nucleosome to distinguish the positioin of acetylated H4 from acetylated H2A in the gel.*

2.2.4 Creating an acetylated H4 binding incompetent TAF1

One mode by which acetylation of histones is thought to facilitate transcription is by destabilizing nucleosome structure and thereby accommodating transcription factor binding at the promoter and passage through the nucleosome by RNA polymerase II [53, 204]. To separate the structural effects of histone acetylation from a possible TAF1–dependent effect, I made strategic point mutations in the doublebromodomain of TAF1. Previous work in yeast found that a point mutation in each of the bromodomains of Bdf1 could abolish its recognition of acetylated H4 tails while maintaining its stability in in vivo [193]. The selection of these mutations was based on the structural homology to Gcn5p [205]. Since Bdf1 is considered to be the yeast ortholog of the human TAF1 doublebromodomain, I reasoned that a similar point mutation would allow me to abolish acetyl–lysine recognition by TAF1 while maintaining its other roles in transcription. To identify these residues, I superimposed the crystal structure of the hTAF1 doublebromodomain with the structure of Gcn5p, as well as other bromodomains that were crystallized with acetylated lysine ligands (Figure 2.14). As expected, three of the residues identified in the structure of Gcn5p to be important for recognition of acetylated lysines aligned precisely with those of the other bromodomains. This allowed me to select tyrosine 1417, and 1540 in hTAF1 for mutation to alanine (Y1540 shown in Figure 2.14).

Because attempts to reconstitute TFIID from recombinant TAFs have proven unsuccessful, I designed a scheme to purify mutant–TAF1–containing TFIID by immunopurification through an introduced tag. A C–terminal HA tag was chosen for this purpose after testing both HA and FLAG tags at the N and C terminus of hTAF1 in HeLa cells (data not shown). Additionally, since TAF1 is an essential protein for cellular growth, expression of the tyrosine mutant could be toxic (*Saccharomyces* Genome Database). Expression under the control of a tetracycline inducible promoter was chosen to allow growth of tens of liters of HeLa cells before induction of the mutant TAF1 and purification through biochemical fractionation and HA immunopurification. Despite my progress on this strategy, the project was put on hold before I had the chance to test it. It is nonetheless presented here as it may be potentially

useful to somebody in the future.

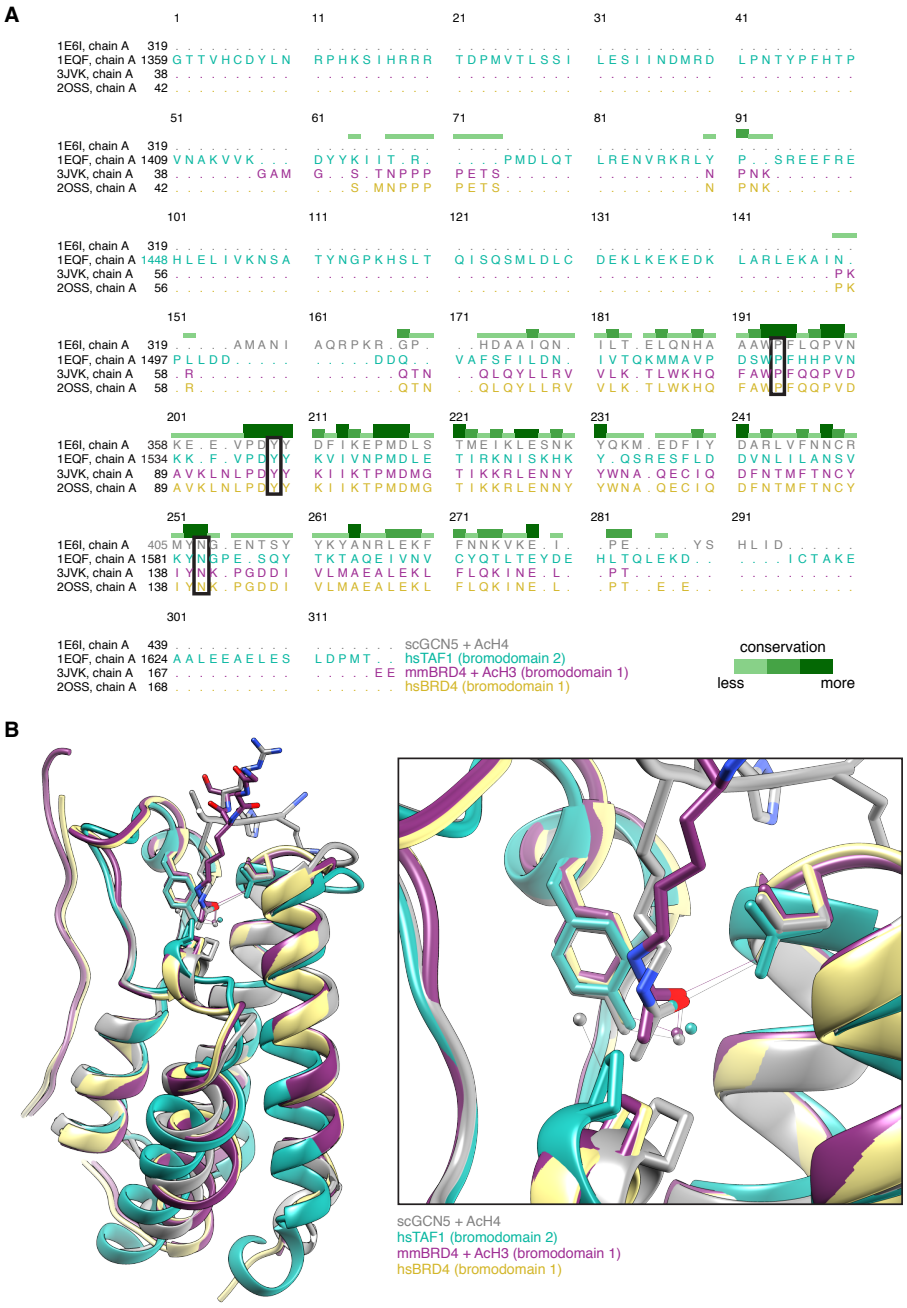


Figure 2.14: *Selection of TAF1 residues to mutate. Structural alignment of the crystal structures of three different bromodomains with bromodomain 1 from the human TAF1 structure [191, 205–207].* *Continued on next page*

Continued Figure 2.14: ***(A)*** *Sequence view of the structural alignment in (B). Boxed residues are depicted as sticks in the structure below and are known to be important for recognition of the acetylated form of lysine.* ***(B)*** *Structural alignment created using Chimera, MatchMaker [13]. The acetylated lysine is shown as a cartoon stick colored red at the oxygen atom, and blue at the nitrogen. Waters are shown as small spheres.*

2.3 Results and Discussion

2.3.1 Acetylation of H4 influences start site selection on a strong promoter

After establishing a well characterized mammalian nucleosome system for transcription in vitro, and setting up a precise acetylation procedure, I was ready to address the question posed at the beginning of this chapter: How does the interaction between TAF1 and acetylated H4 tails influence transcription? To address this, nucleosomal templates representing the -1 nucleosome at a strong promoter were tested first (Figure 2.15–A). Prior to transcription, nucleosome templates were incubated with Piccolo NuA4 and acetyl coenzyme A in order to hyper–acetylated the H4 tails. Transcription from these acetylated templates was then compared to unacetylated nucleosomal templates as well as naked DNA (Figure 2.15–B). To test for a dependence on TAF1, TFIID was titrated over four fold.

Two major results are observed. First, while assembly of a nucleosome at the -1 position decreased overall transcription amplitude at a strong promoter by about a third, acetylation of the nucleosome at H4 was not able to relieve this repression (Figure 2.15–C). This may be explained in part by the location of the nucleosome used here. While acetylation is known to induce a looser nucleosome structure and therefore may be expected to facilitate passage by the polymerase, transcription initiates away from the -1 nucleosome, and thus will not benefit from a weaker barrier to the polymerase. Conveniently, usage of a -1 nucleosomal template allowed me to circumvent the need for chromatin remodelers or elongation factors for these initial experiments.

Second, the transcription initiation site was influenced by H4 acetylation (Figure 2.15–D). The naked DNA template displayed three initiation sites approximately 1 bp apart. Assembly into a nucleosome focused initiation to TSS1, while acetylation of the nucleosome shifted the start site preference back toward that of the naked DNA. Importantly, this shift in start site selection was TFIID dependent, present at intermediate TFIID concentrations where TFIID is likely the limiting reagent for transcription, but not at higher concentrations where interactions between other basal factors might be rate limiting. This is in direct contrast to the influence of the -1 nucleosome on transcription amplitude. Repression by the presence of a -1 nucleosome is present at all concentrations of TFIID with no relief by acetylation of H4 observed. Thus, rather than having a major impact on recruitment of TFIID, acetylation of H4 at a strong promoter seems to have an effect on positioning the pre–

initiation complex for transcription. This role is in direct competition with the sequence cues of the core promoter elements. Thus, it would be informative to perform this experiment on nucleosomal transcription templates contain mutations at the CPEs. Additionally, despite a dependence on TFIID concentration, this experiment does not demonstrate the dependence on a direct interaction between TAF1 and acetylated H4 tails. To test for a direct causal effect of TAF1 interaction with H4 hyperacetylated tails, TFIID containing the mutations to the TAF1 doublebromodomain presented in Section 2.2.4 should be substituted for wtTFIID in transcription reactions. Unfortunately, I ran out of time to perform these experiments, and as such, this remains an open question.

While histone acetylation has been associated in vivo with active transcription, its precise role in initiating transcription has remained poorly understood. Moreover, genome–wide ChIP-seq studies demonstrate only correlation and cannot tease apart cause and effect. Here, I presented my work towards teasing apart the mechanism of influence of H4 acetylation on transcription initiation. In support of my preliminary findings, a couple papers have been published in the years since this project was put on pause. First, Vanja Haberle et al performed an elegant experiment in zebrafish published in 2014 in which they mapped TSSs using cap analysis of gene expression (CAGE) before and after the transition from maternal to zygotic transcription in the early embryo [208]. By comparing the initiation profiles to the locations of H3K4me3-marked nucleosomes, they found differential TSS usage between maternal and zygotic transcripts which correlated with a switch from a TATA directed start site to a seemingly nucleosome positioned start site as evidenced by no distance correlation with the zygotic TATA box, but instead a precise positioning of the +1 nucleosome with respect to the zygotic, but not maternal TSS. This positioning was punctuated by H3K4me3, a mark known to interact with TAF3. While these results are still correlative, they indicate a strong example of a possible link between TAF–histone interactions and start site selection (Figure 2.16).

Lastly, work in the Pugh lab using ChIP–exo has revealed a higher resolution picture of the pre-initiation complex at the promoter, and suggests a correlation with the position of the +1 nucleosome [198]. Additionally, the group previously observed a correlation between the binding of the TAF1 doublebromodomain homolog, Bdf1, H4 acetylation at the +1 nucleosome, and levels of transcription in yeast [209]. Taken together, these studies support the notion of a link between TFIID recognition of modified histones at the promoter and start site selection, but rather than a role for the -1 nucleosome as I observed, they point to a role for the +1 nucleosome. This hypothesis is easily testable using the system I built which includes a template representing the +1 nucleosome, but these questions may be reserved for another researcher to answer in the future.

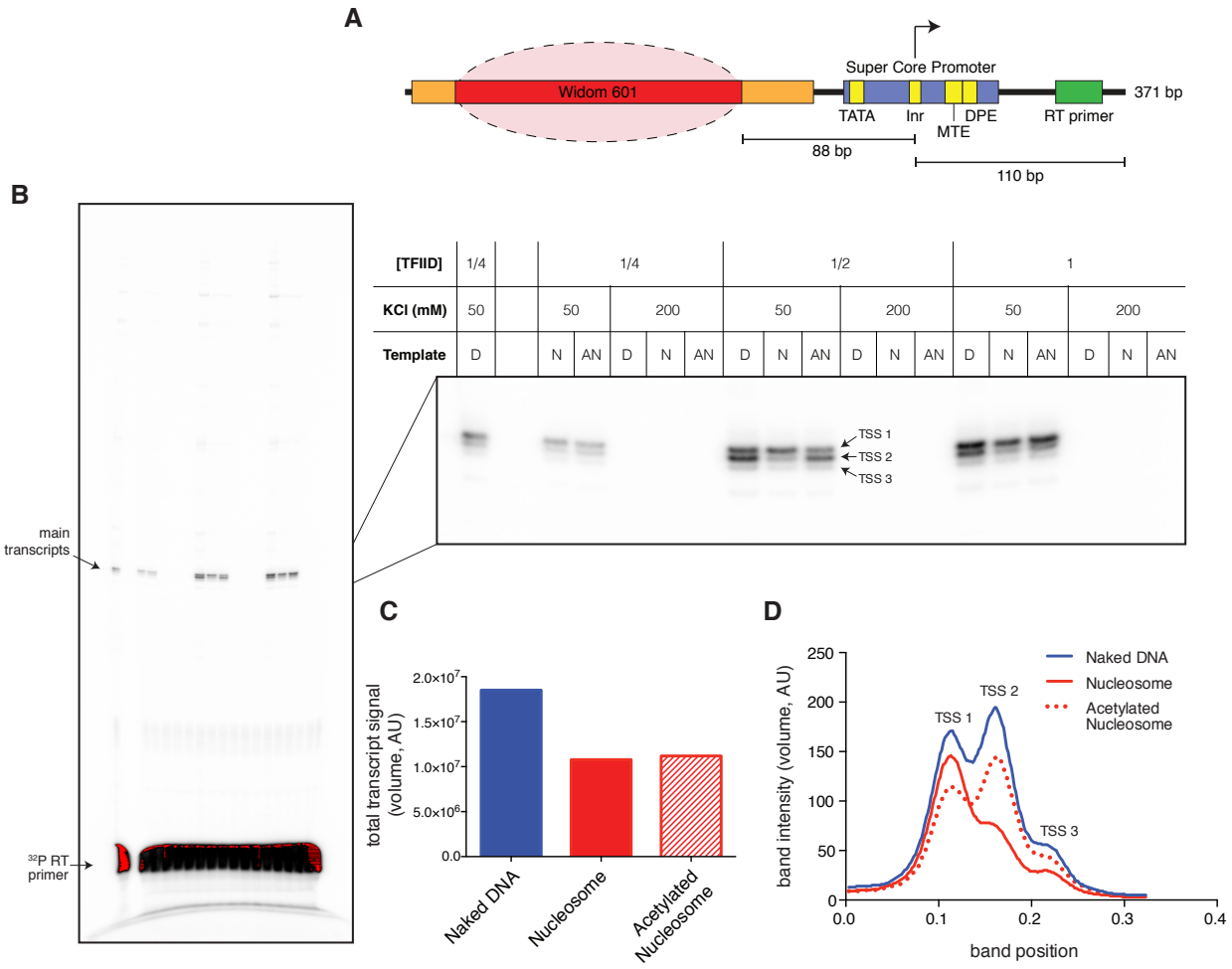


Figure 2.15: *Acetylation of H4 influences start site selection.* ***(A)*** *Nucleosome construct used in transcription in (B).* ***(B)*** *Phosphor image of a Urea-PAGE gel showing the results of transcription from the indicated templates. After incubation with purified human basal factors and RNA polymerase II at 30° C for 30 minutes, transcripts were reverse transcribed (RT) using a* ^{32}P *labeled primer and loaded on a denaturing Urea-PAGE gel. The templates used were either naked DNA 'D', nucleosomal template 'N', or H4 acetylated nucleosomal template 'AN'. TFIID concentrations were titrated and are labeled as a fraction of the most concentrated. Transcripts originating from three different start sites are labeled TSS1, TSS2, and TSS3.* ***(C)*** *Quantification of the integrated band signal as a sum of all transcripts at 1/2 TFIID concentration.* ***(D)*** *Quantification of the transcripts at 1/2 TFIID concentration. Band intensity is plotted as calculated from a line drawn directly vertically over the three main transcript bands.*

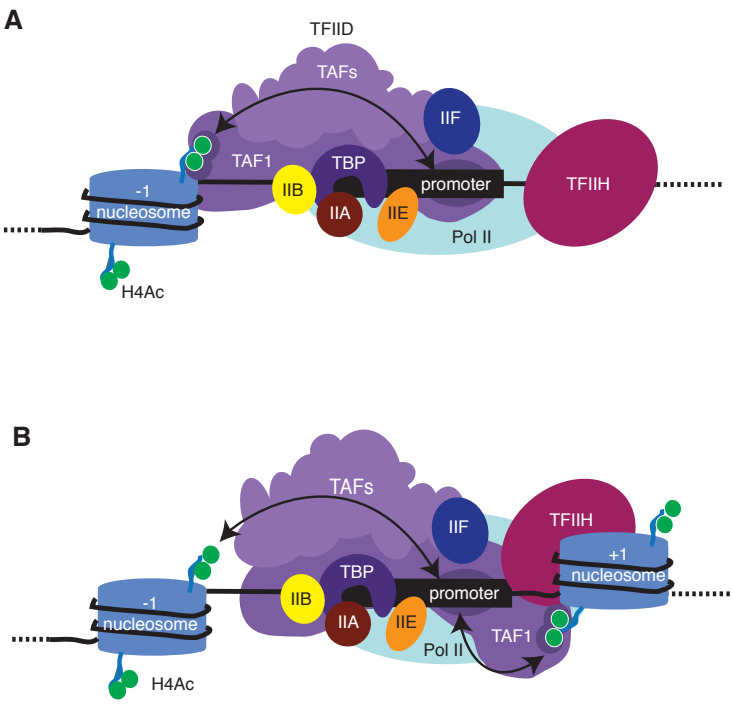


Figure 2.16: *Proposed model.* ***(A)*** *Cartoon demonstrating the proposed competition between TFIID recognition of core promoter elements and its interaction with acetylated H4 at the -1 nucleosome as observed in Figure 2.15.* ***(B)*** *A model including the potential role of the +1 nucleosome in interactions between the basal transcriptional machinery, core promoter elements, and acetylated histones flanking the promoter.*

2.4 Materials and Methods

2.4.1 Purification of Piccolo NuA4

Purification followed the procedure outlined in [201].

2.4.2 Restriction enzyme accessibility assays

Restriction enzyme accessibility assays were performed essentially as described in Chapter 3 of this thesis on page 64.

2.4.3 Transcription assays

In vitro transcription reactions were carried out essentially as described in [210]except immuno–affinity purified TFIIH and RNA polymerase II were used and purified in–house by Dr. Carla Inouye. Concentrations of the basal machinery are used as relative concentrations

based on standardized degree of activity in each preparations. The purification procedure for nucleosomal templates, as well as the amounts used are contained in the text and figure legends. When used, plasmids were phenol/chloroform/isoamyl alcohol extracted following Qiagen midi preparations.

Transcription reactions were performed in 0.1 M HEMG buffer and contained 2 mM Spermidine, 83 ng/μL, RNase IN, and 50 mM dGdC (unless otherwise specified). Basal factors included purified human Pol II, TFIID, TFIIB, TFIIF, TFIIE, and TFIIH. Briefly, basal factors were mixed prior to addition to the transcription templates, and reactions were incubated at 30° C for 30 minutes, to allow loading of the pre–initiation complex. Then, NTPs were added and the reactions were allowed to proceed for 30 minutes at 30° C. Transcription was stopped with a solution containing SDS and proteinase K (esentially as in the restriction enzyme accessibility reaction s described on page 64. After digestion at 37° C for 10 minutes, reactions were phenol/chloroform/isoamyl alcohol extracted and ethanol precipitated. Excess ^{32}P labeled primer was added, hybridized to the transcript, and reverse transcribed using SSIII at 37° C for 30 minutes. Reactions were ethanol precipitated and resuspended in formamide loading buffer and heated to 95° C for 6 minutes. Samples were then loaded on a 6% denaturing Urea-PAGE gel and analyzed using storage phosphor screens and a Bio–Rad Typhoon phosphor imager.

2.4.4 Histone and nucleosome acetylation

^{14}C Acetyl coenzyme A was used in preliminary assays to characterize acetylation, but for acetylation of templates used in transcription reactions, cold acetyl coenzyme A was used. Reactions were carried out at room temperature in 250 mM Tris–HCl pH 7.5, 5 mM DTT, and 0.1 mg/mL MSA. Piccolo NuA4 was added at 200 nM to the buffer containing nucleosomes or histones at 156 nM. Reactions were begun by adding acetyl coenzyme A to 40 μM. Reactions were incubated for the indicated time, and then quenched with loading buffer containing 8M urea, 10% glacial acetic acid, and a pinch of pyronin B for visualizing migration on the gel. Gels used to separate out acetylated histone species were acid–urea gels. For reactions using ^{14}C Acetyl coenzyme A, gels were analyzed using storage phosphor screens and a Bio–Rad Typhoon phosphor imager. Gels were subsequently stained with PageBlue to visualize un–modified histones.

2.4.5 Nucleosome species purification

Nucleosome assemblies were loaded on 6% native PAGE gels along with a small amount of sample in a lane adjacent to the preparative lane. Loading buffer included only sucrose and no dye as it interferes with migration in the gel. After a set running time determined empirically based on the template, the gel was sliced in half separating the preparative lane from the analytical lane. The analytical lane was stained with PageBlue and then repositioned next to the unstained half of the gel to direct the location to be excised. The identified bands were cut out using a razor blade and then inserted into small bags of dialysis tubing. Nucleosomes

were then electroeluted from the gel slice in 0.1X TBE with 0.1 mM DTT in a horizontal electrophoresis gel apparatus for about 2 hours at 100V, before reversing current for about a minute to dislodge the nucleosomes from the edge of the dialysis bag. The bags were then opened, and the solution containing the eluted nucleosomes was pipetted out.

2.4.6 Histone purification and nucleosome assembly

See pages 60, 63, and 74.

2.4.7 DNA constructs

Templates used a Widom 601 with a flanking sequence provided by the Luger lab as follows.

5' - A T C G G A C C C T A T A C G C G G C C G C C C T G G A G A A T C C C G G T G C C G A G G C C G C T C A A T T G G T C G T A G A C A G C T C T A G C A C C G C T T A A A C G C A C G T A C G C G C T G T C C C C C G C G T T T T A A C C G C C A A G G G G A T T A C T C C C T A G T C T C C A G G C A C G T G T C A G A T A T A T A C A T C C T G T G C A T G T G G A T C C G A A T T C A T A T T A A T T A A T A C T A G A T - 3'

The Super Core promoter was cloned out of pSCBCAT, and the nucleosome spacing sequence was cloned from the pGL3-Basic vector and is as follows.

5' - g c g a a c g a c a t t t a t a a t g a a c g t g a a t t g c t c a a c a g t a t g g g c a t t t c g c a g c c t a c c g t g g t g t t c g t t t c c a a a a a g g g g t t g c a a a a a a t t t t g a a c g t g c a a a a a a a a g c t c c c a a t c a t c c a a a a a a a t t a t t a t c a t g g a t t c t a a a a c g g a t t a c c a g g g a t t t c a g t c g a t g t a c a c g t t c g t c a c a t c t c a t c t a c c t c c c g g t t t t a a t g a a t a c g a t t t t g t g c c a g a g t c c t t c g a t a g g g a c a a g a c a a t t g c a c t g a t c a t g a a c t c c t c t g g a t c t a c t g g t c t g c - 3'

The reverse transcription primer was derived from the CAT gene in pSCBCAT and is: 5' - CCACCGTTGATATATCCCAATGGC - 3'.

Chapter 3

Nucleosomes impede Cas9 access to DNA in vivo and in vitro

Max A Horlbeck*, Lea B Witkowsky*, Benjamin Guglielmi, Joseph M Replogle, Luke A Gilbert, Jacqueline E Villalta, Sharon E Torigoe, Robert Tjian†, Jonathan S Weissman†

2016 **eLife** 5:e12677. doi: 10.7554/eLife.12677 [211]

3.1 Introduction

CRISPR (clustered regularly interspaced palindromic repeats) prokaryotic adaptive immune systems have yielded transformative tools for manipulating eukaryotic genomes. Most notably, the CRISPR-associated Cas9 protein from Streptococcus pyogenes, together with a single chimeric guide RNA (sgRNA), provides a programmable endonuclease that has revolutionized our ability to edit genomes [212]. Cas9 has been further modified to a nuclease-dead form (dCas9) to provide a programmable DNA-binding protein that can be fused to effector domains, making it possible to turn on or off targeted genes, mark specific genomic loci with fluorescent proteins, or alter epigenetic marks [212–219]. A central challenge in implementing these tools is identifying effective and specific sgRNAs. While much of the effort to define relevant rules has focused on the sequence of the target site and sgRNA [220–223], these only partially predict Cas9 activity and suggest that additional determinants likely exist.

Chromatin structure may represent a key parameter governing Cas9 efficacy in eukaryotic cells. CRISPR evolved in archea and bacteria [224] and is likely not optimized to explore and modify large, chromatin-bound eukaryotic genomes, a hypothesis supported by several studies that point to a correlation between rates of DNA binding and cleavage in regions of open chromatin as measured by DNase I hypersensitivity [220, 225, 226]. Additionally,

*these authors contributed equally to this work

†corresponding authors: jmlim@berkeley.edu (RT), Jonathan.Weissman@ucsf.edu (JSW)

recent single-molecule imaging studies have shown that dCas9 explores euchromatin more frequently than it does heterochromatin [227]. We hypothesized that nucleosomes, the basic unit of chromatin structure, are an important impediment to Cas9 recognition. Here, we addressed this hypothesis in vivo by leveraging large datasets collected from over 30 sgRNA tiling and genome-scale genetic screens [228]. We found that regions of high nucleosome occupancy in vivo, as determined by MNase-seq (micrococcal nuclease sequencing) [229, 230], were strongly depleted of highly active sgRNAs for CRISPR interference (CRISPRi) [214, 228] and nuclease-active Cas9. We complemented these results with in vitro experiments demonstrating that formation of a nucleosome provides a direct and profound block to dCas9 binding and Cas9 cleavage. Despite this strong barrier to Cas9 activity, we found that addition of a chromatin remodeling enzyme to chromatinized DNA in vitro can restore Cas9 access to nucleosomal DNA, highlighting one route by which CRISPR may still be able to modify chromatin in vivo. Our results reveal a fundamental aspect of the mechanism by which this transplanted bacterial enzyme interacts with eukaryotic chromatin, and provide a new dimension for selecting highly active sgRNAs.

3.2 Results

3.2.1 CRISPRi activity is periodic and out-of-phase with nucleosome positioning

In order to study the features of chromatin that affect CRISPRi activity, we integrated data from whole-genome screens testing for a wide range of phenotypes performed with a previously described CRISPRi library targeting each gene with ~10 sgRNAs [228]. We selected 30 screens performed in the cell line K562 expressing dCas9-KRAB (ML, BA, BB, CYP, MK, YC, JF, and JN, personal communication) and set a threshold for high-confidence hit genes, which allowed us to assess the relative strength of phenotypes for the sgRNAs targeting these genes. Specifically, we analyzed 18,380 sgRNAs targeting 1539 genes, and generated 'activity scores' by normalizing sgRNA phenotypes to the average of the 3 sgRNAs with the strongest phenotypes for each gene (Figure 3.1-A and Supplementary file 1 [211]). To assess how CRISPRi activity varies with respect to the transcription start site (TSS), we plotted the average sgRNA activity score as a function of distance to the FANTOM-annotated TSS [231] (Figure 3.1-B). This analysis revealed a robust, periodic pattern of activity, with peaks at ~190 bp intervals relative to the TSS. This periodicity was highly reminiscent of patterns previously described for nucleosomes [30]. Indeed, analysis of K562 MNase-seq data from the ENCODE consortium [230, 232] revealed that the average nucleosome signal was strongly anti-correlated with CRISPRi activity (Figure 3.1-B), suggesting that high nucleosome occupancy leads to low CRISPRi activity.

To further explore this inverse relationship between CRISPRi activity and nucleosome organization, we exploited the previous observation that nucleosome phasing is more pronounced in highly expressed genes [230]. We grouped genes by their expression in K562

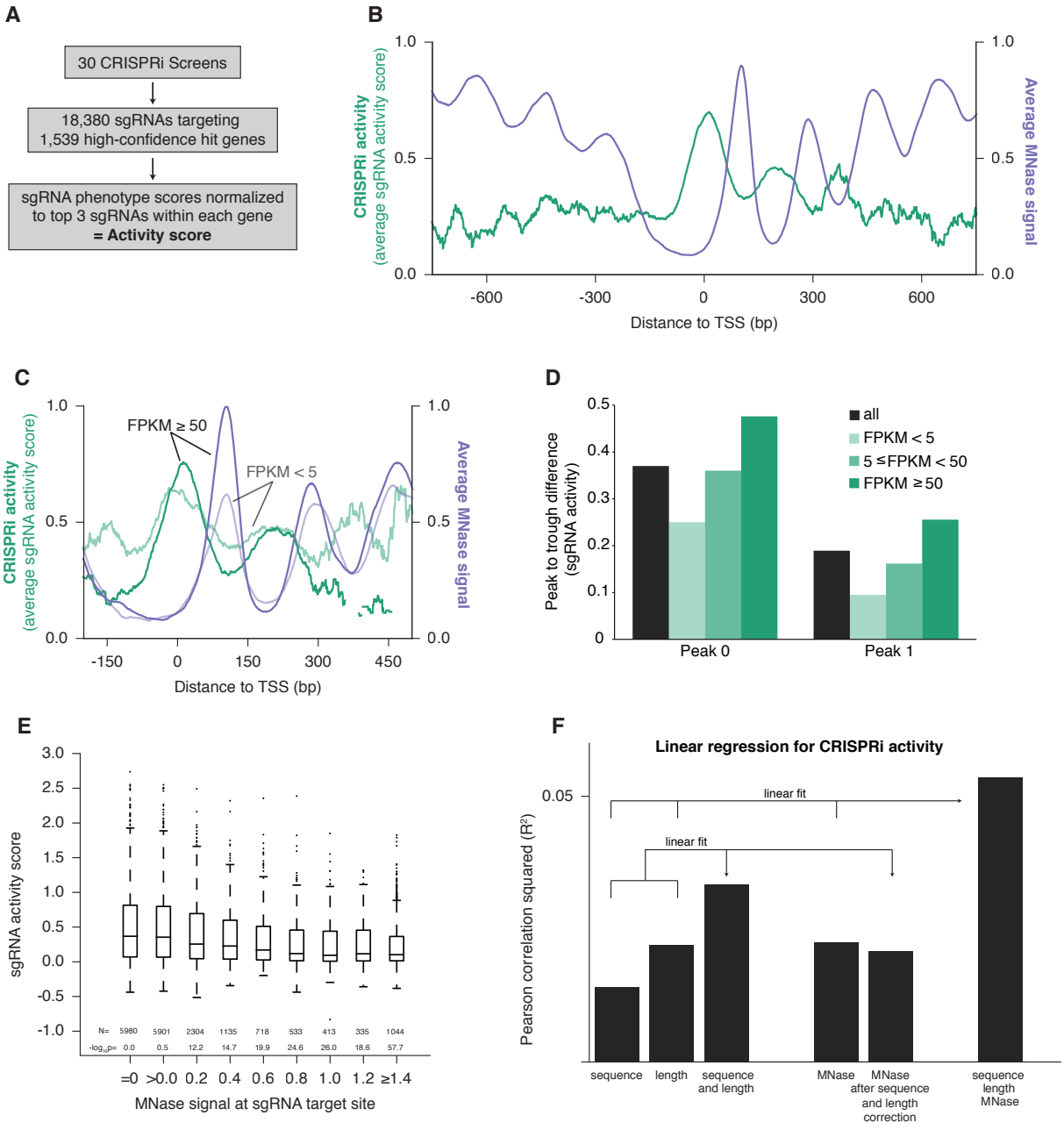


Figure 3.1: *CRISPRi activity anti-correlates with nucleosome occupancy.* ***(A)*** *Workflow for generating CRISPRi activity scores from pooled genetic screens. The resulting values are distributed around 0 for inactive sgRNAs and around 1 for highly active sgRNAs.* ***Continued on next page***

Continued Figure 3.1: ***(B)*** *Average CRISPRi activity and MNase-seq signal relative to the TSS. Green line represents average CRISPRi activity score of all sgRNAs within a 50bp window around each position. Purple line represents the K562 MNase-seq signal at each position averaged across all genes analyzed.* ***(C)*** *Average CRISPRi activity and MNase-seq signal for genes grouped by expression value. Genes were grouped into lower expression (light lines; $N = 240$), higher expression (heavy lines; $N = 368$), and medium expression (omitted for clarity; $N = 930$), and analyzed as in (B). Expression values were obtained as fragments per kilobase million (FPKM) from ENCODE K562 RNA-seq data. Average activity at positions with fewer than 10 sgRNAs within the 50bp window was not calculated.* ***(D)*** *Quantification of the amplitude of periodic CRISPRi activity. Peak and trough coordinates were obtained by calculating the local maxima and minima of the activity traces from analyses in (B) and (C). Peak 0 was defined as the local maximum closest to the TSS, peak 1 was defined as the next maximum downstream of peak 0, and troughs were defined as the minima immediately downstream of the respective peaks.* ***(E)*** *CRISPRi activity and target site nucleosome occupancy for individual sgRNAs. Target site nucleosome occupancy was calculated from the average MNase-seq signal at all genomic coordinates across the length of the sgRNA protospacer and the protospacer adjacent motif (PAM). sgRNAs were then binned by the target site nucleosome occupancy, displayed as box-and-whisker plots, and labeled according to the minimum value within the bin except where indicated. P-values were calculated by a two-tailed Mann-Whitney test comparing each bin to the =0.0 bin.* ***(F)*** *Linear regression for CRISPRi activity. The squared Pearson correlation was calculated for the sgRNA activity scores compared to the indicated individual parameters (bars 1, 2, and 4) or linear fits of multiple parameters (bars 3 and 6). sgRNA activity scores were corrected for sequence and length features (bar 5) by subtracting the linear fit of those two features.*

[232, 233] and analyzed CRISPRi activity and nucleosome occupancy within each group. In support of a connection, we found that peak to trough amplitudes of both features were larger for highly expressed genes (Figures 3.1-C,D).

We also analyzed the MNase signal at each sgRNA target site to determine whether nucleosome occupancy could explain variation in CRISPRi activity between individual sgRNAs (Figure 3.1-E). Consistent with the hypothesis that nucleosomes exclude dCas9, almost all of the highly active sgRNAs targeted sites with low MNase-seq signal. However, not all sgRNAs targeting sites with low nucleosome occupancy were highly active, matching previous findings that sgRNA and target DNA sequence features also influence efficiency of the CRISPR system [220–223]. To exclude the possibility that differences in sequence features alone between nucleosome-bound and -free regions could explain the periodicity of CRISPRi activity (rather than the presence of nucleosomes per se), we performed linear regression using MNase signal, sgRNA length [223, 228], and a validated sgRNA sequence scoring algorithm [221]. We found that each parameter individually correlated with sgRNA activity ($p < 10^{-58}$ for each; Figure 3.1-F). Importantly, correction for sequence and length features had minimal impact on the ability of the MNase signal to predict CRISPRi activity

($p < 10^{-85}$ after correction). Indeed, a linear fit of all three features provided a still stronger correlation ($p < 10^{-221}$; Figure 3.1-F, far right column), suggesting that incorporating nucleosome occupancy in future sgRNA design algorithms could significantly improve predictive value. We have recently developed a comprehensive algorithm for predicting highly active CRISPRi sgRNAs that, by accounting for nucleosome occupancy, higher order sequence features, and non-linear relationships in these parameters, shows even greater correlation with this dataset that was already enriched for active sgRNAs using our original CRISPRi library design principles [228] (cross-validation $R^2 = 0.32$; Horlbeck et al., manuscript in preparation).

3.2.2 Nuclease-active Cas9 activity anti-correlates with nucleosome occupancy

In order to generate a dataset for evaluating the effect of nucleosome positioning on nuclease-active Cas9 in K562 cells, we took advantage of our previously described library densely tiling sgRNAs in 10kb windows around the TSS of 49 genes known to modulate susceptibility to the toxin ricin [228, 234] and tested Cas9-expressing cells for resistance or sensitivity to ricin (Supplementary file 2 [211]). We observed phenotypes consistent with the expected knockdown phenotype primarily in coding sequences (CDS) but also in some promoter regions, consistent with recent results showing that modifications introduced by Cas9 can disrupt cis-regulatory regions [235]. Analysis of CDS-targeting sgRNAs revealed that strong phenotypes were found predominantly in regions of low MNase-seq signal (Figure 3.2-A), although this relationship was less pronounced than for CRISPRi. This may be due to decreased phasing of nucleosomes within the gene body [30]. When we analyzed all sgRNAs in the library, thus incorporating information from cis-regulatory regions where nucleosomes are well phased, we found the effect of nucleosome position to be even stronger (Supplementary Figure A.3). Although nucleosome occupancy was predictive of CRISPR activity independent of sgRNA sequence features, a much stronger correlation was obtained when all features were considered (Figure 3.2-B). Therefore, nucleosome organization likely represents an important feature for CRISPR sgRNA design and should be considered a key contributing factor in interpreting future tiling mutagenesis experiments of coding and non-coding regions.

3.2.3 Nucleosomes are sufficient to fully block cleavage by Cas9 in vitro

Our in vivo experiments reveal a strong anti-correlation between Cas9-mediated downstream phenotypic outputs and correlation in our data. For example, (d)Cas9 may bind or cut equally well in nucleosome-bound and un-bound regions, but may exert the observed modulation of gene expression through interference with transcriptional pausing, splicing, regulatory looping, or binding of important regulatory factors, processes which also correlate with nucleosome organization [236–242].

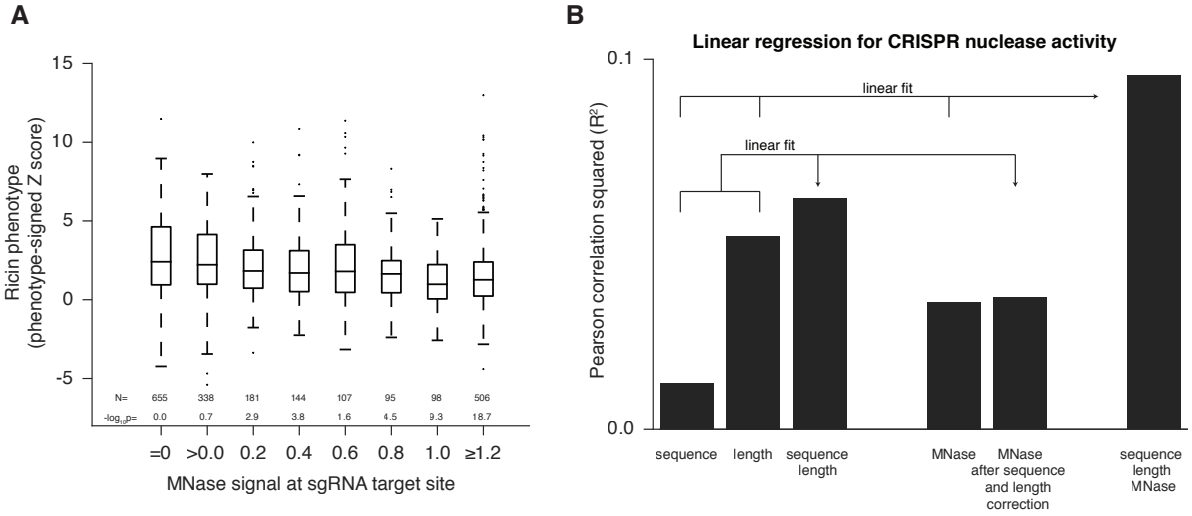


Figure 3.2: *Cas9 nuclease activity anti-correlates with nucleosome occupancy.* ***(A)*** *Cas9 nuclease phenotypes and target site nucleosome occupancy for individual sgRNAs targeting CDS regions. Ricin susceptibility phenotypes for each sgRNA are expressed as a Z score and are positive if the phenotype matches the expected knockdown phenotype. Target site nucleosome occupancy was calculated as in Figure 3.1-E. sgRNAs were then binned by the target site nucleosome occupancy, displayed as box-and-whisker plots, and labeled according to the minimum value within the bin except where indicated. P-values were calculated by a two-tailed Mann-Whitney test comparing each bin to the =0.0 bin.* ***(B)*** *Linear regression for Cas9 nuclease phenotypes. The squared Pearson correlation was calculated for the sgRNA activity scores compared to the indicated individual parameters (bars 1, 2, and 4) or linear fits of multiple parameters (bars 3 and 6). sgRNA activity scores were corrected for sequence and length features (bar 5) by subtracting the linear fit of those two features.*

To determine whether (d)Cas9 activity is indeed directly affected by the presence of nucleosomes, we turned to a purified in vitro reconstituted system. Mouse histones were recombinantly expressed, purified, and assembled into a nucleosome using 147bp of the Widom 601 positioning sequence (Figures 3.3-A,B) [33]. Conveniently, the 601 sequence contains numerous NGG protospacer adjacent motifs (PAMs) required for Cas9 recognition, spanning the full length of the DNA at different helical positions, allowing us to test the effects of position and solvent accesibility within the nucleosome (Figure 3.3-C). We first tested the ability of Cas9 to cleave nucleosomal DNA. We pre-loaded purified Cas9 (histidine tagged and HaloTagged) with in vitro transcribed sgRNA, then introduced either naked 601 DNA or that same DNA assembled into a nucleosome (Figures 3.3-A,D and Supplementary Figure A.4-A). Fluorescent labeling of the DNA allowed us to visualize the cleavage products on a denaturing Urea-PAGE gel. In agreement with our in vivo data, the nucleosome protected its DNA from cleavage by Cas9, and complete protection from cleavage was observed to be

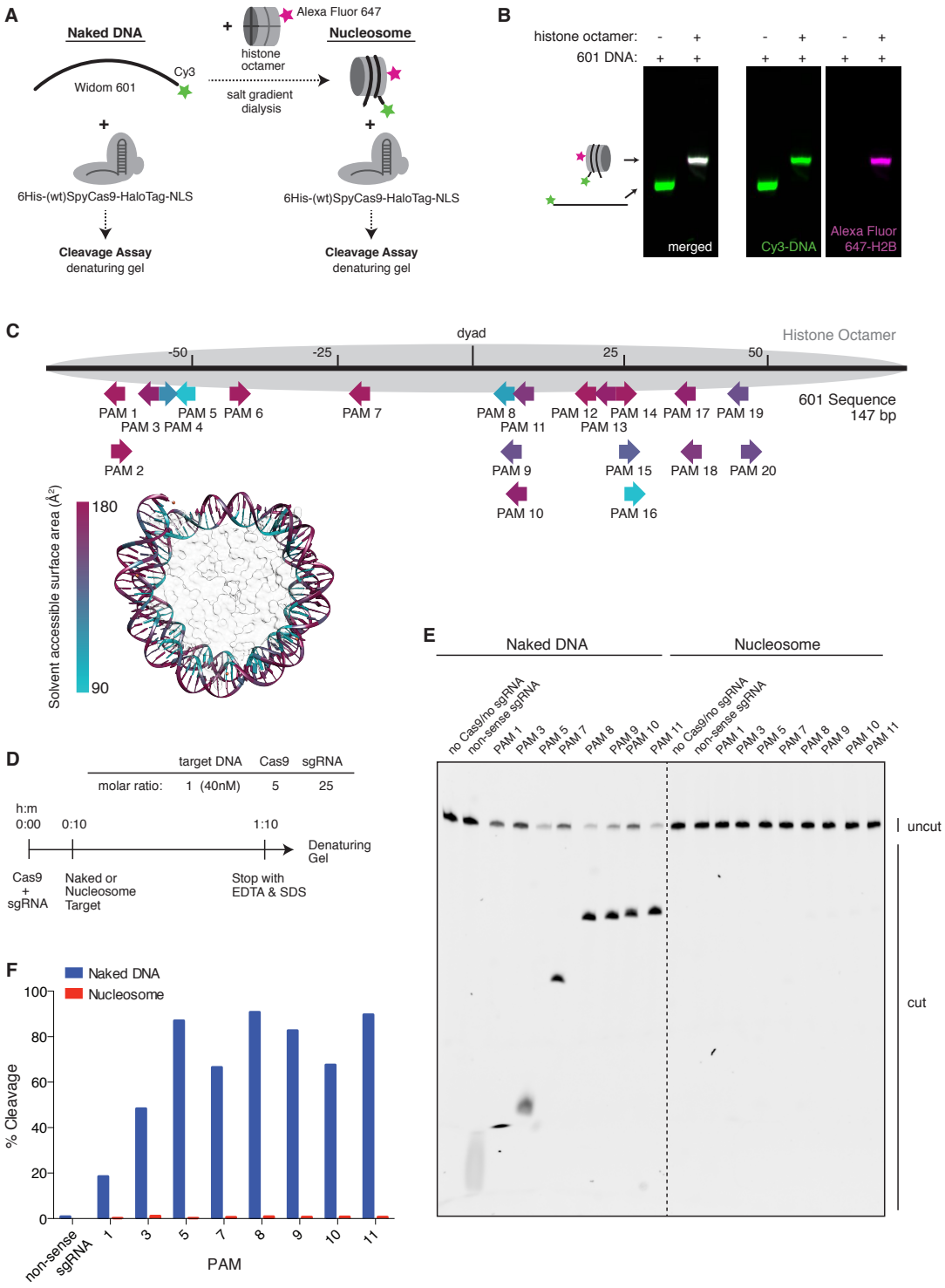


Figure 3.3: *Cas9 nuclease activity is blocked by the presence of a nucleosome in vitro.*
Continued on next page

Continued Figure 3.3: ***(A)*** *Schematic of the experimental setup for in vitro cleavage assays. Mononucleosomes were assembled by salt gradient dialysis of purified mouse histone octamer with the minimal nucleosome positioning sequence, Widom 601 (147 bp). Prior to assembly, DNA was 5'-end-labeled with Cy3 and histone H2B was fluorescently labeled using an introduced cysteine (T115C) coupled to Alexa Fluor 647. Purified His-tagged and HaloTagged Cas9 pre-loaded with in vitro transcribed sgRNAs were added to naked DNA or assembled nucleosomes, and DNA cleavage products were visualized using a denaturing gel imaged for Cy3-DNA fluorescence. See also Supplementary Figure A.4-A.* ***(B)*** *Confirmation that fully occupied, well-positioned nucleosomes were assembled. After assembly using salt gradient dialysis, the produced nucleosomes were visualized using a native PAGE gel imaged in the Cy3 and Alexa Fluor 647 channels. Full incorporation of the free DNA into a nucleosome occupying a single position on the DNA is indicated by the presence of a single-shifted band containing all of the detectable Cy3-DNA and Alexa Fluor 647-H2B signal.* ***(C)*** *Available PAMs and solvent accessibility of the 601 nucleosome positioning sequence. (Above) A schematic of the 601 sequence. The location of the histone octamer when assembled into a nucleosome is indicated by the gray oval. The location of PAMs within the double-stranded sequence are indicated with arrows spanning the 3 bp of the PAM, pointing in the 5' to 3' orientation of the NGG motif. The arrows are colored according to solvent accessibility at the center of the PAM as calculated from the crystal structure of the 601 nucleosome (PDBID 3LZ0, [243]). (Below) Crystal structure of the 601 nucleosome. For clarity, the surface area of the histones in the crystal structure has been made transparent. The DNA in the crystal structure is colored according to solvent accessibility using the same color scale as the PAMs above. Residues colored teal are less accessible, while residues colored fuchsia are more accessible by solvent.* ***(D)*** *Experimental conditions and timeline for cleavage assays.* ***(E)*** *Denaturing PAGE gel showing the results of a cleavage assay targeting the indicated PAMs. Cleavage reactions containing naked DNA were loaded on the left half of the gel, while reactions containing nucleosomes were loaded on the right. The DNA was imaged via a Cy3 fluorophore attached to the 5' end of the sgRNA-complimentary strand. A negative control was conducted with an sgRNA that had no sequence complementarity to the 601 sequence used (non-sense guide). See also Supplementary Figure A.4-B,C for additional controls.* ***(F)*** *Quantification of the gel in (D). For each lane, percent cleavage was determined by calculating the percent of the total band signal corresponding to cleaved DNA.*

independent of target position within the nucleosome (Figures 3.3-E,F and Supplementary Figure A.4-B,C). While our manuscript was under review, a study by Hinz and colleagues reported a similar in vitro finding that Cas9 nuclease activity is inhibited within the nucleosome but not at adjacent linker sequences [244]. Additionally, previous single-molecule and biochemical studies have established that nucleosomal DNA undergoes transient unwrapping or breathing at the entry and exit sites, creating a gradient of accessibility along the nucleosome [14–17, 24, 245]. This property is often credited for the observed position-dependent binding patterns of many transcription factors and DNA-binding proteins. Interestingly,

our data suggests that the cleavage activity of Cas9 in vitro is not detectably influenced by this effect, although less stable nucleosomes than the Widom 601 nucleosome or target sites closer to the nucleosome edge may exhibit more breathing and thus more accessibility than those tested here.

3.2.4 Cas9 is unable to bind nucleosomal DNA in vitro

Cleavage of DNA by Cas9 has been described as a stepwise process [226, 227, 246] in which Cas9 must first scan for PAMs, unzip the DNA duplex, and fully pair the guide RNA and target DNA prior to cleavage. While our in vitro data show that full pairing and cleavage is prevented by the presence of a nucleosome, we wondered if binding without cleavage, especially at the more accessible ends of the nucleosome, might still occur. Additionally, the first step in binding, PAM recognition, may be governed by helical location within the nucleosome as well as its proximity to the more dynamic ends. Histones make contacts with the DNA at every helical turn, thus a PAM may fall on the outside of the nucleosome, exposing it to solvent, or on the inside at the DNA-histone interface. To test the influence of target location within the nucleosome on Cas9 binding, we used an electrophoretic mobility shift assay to monitor binding of dCas9 to Cy3 end-labeled 601 DNA (Figures 3.4-A,B), either free or assembled into a nucleosome containing Alexa Fluor 647-labeled histone H2B (Supplementary Figure A.5-A). Consistent with our cleavage results, binding by dCas9 was abolished by the presence of the nucleosome, regardless of the targeted dCas9 binding site (Figure 3.4-CE and Supplementary Figure A.5-A,B).

To better understand how a nucleosome might impede Cas9 binding, we aligned the available crystal structures of DNA-bound Cas9 and the structure of the 601 nucleosome ([243] PDB ID 3LZ0, [247] PDB ID 4UN3). We superimposed the target DNA in the Cas9 crystal structure with the DNA in the nucleosome structure at a site where the two DNA paths gave the best fit. The resulting combined structure reveals that the Cas9 protein poses significant steric clashes with the histones (Figure 3.4-F). Given the extent of overlapping densities in the two structures, it seems unlikely that the histones and Cas9 could co-occupy the same piece of DNA. Additionally, it may be important to note that unlike other DNA binders such as transcription factors, binding by Cas9 constitutes melting of target DNA, which may pose an additional barrier to binding on a nucleosome. This hypothesis leaves two possible outcomes of targeting Cas9 to nucleosomal DNA: either Cas9 is capable of displacing histones in order to engage nucleosomal DNA, or it is excluded altogether. Our data support the latter conclusion.

3.2.5 The chromatin remodeling enzyme yChd1 can restore access to nucleosomal DNA in vitro

The nucleosome landscape in eukaryotic chromatin is dictated by both intrinsic DNA sequence preferences as well as extrinsic factors such as chromatin remodeling enzymes [30, 248, 249]. In order to model how these dynamics affect Cas9 access to nucleosomal DNA in

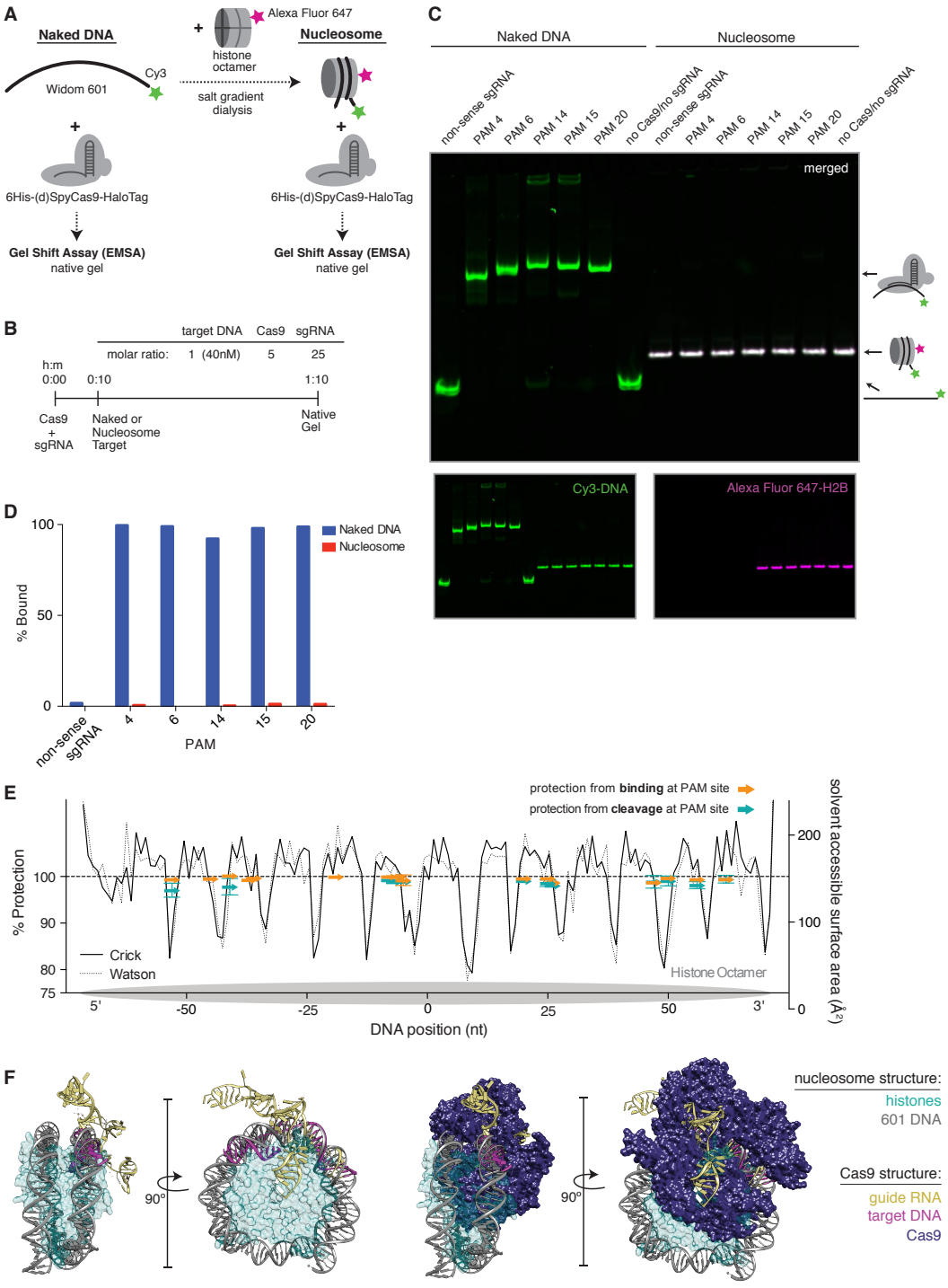


Figure 3.4: *dCas9 is unable to bind nucleosomal DNA in vitro.* ***Continued on next page***

Continued Figure 3.4: ***(A)*** *Schematic of the experimental setup for in vitro binding assays. Either naked DNA or assembled nucleosomes were incubated with catalytically dead Cas9 (dCas9, histidine tagged and HaloTagged), and binding was assessed by an electrophoretic mobility shift assay (EMSA).* ***(B)*** *Experimental conditions and timeline for binding assays.* ***(C)*** *A native PAGE gel showing the results of an EMSA in which dCas9 was targeted to the indicated PAMs on either naked or nucleosomal DNA. Gels were scanned for fluorescence from Cy3 on the DNA (green) and Alexa Fluor 647 on histone H2B (magenta). The two color channels were merged to identify the location of intact nucleosomes (white). See also Supplementary Figure A.5 and A.7 for reagent preparation and experimental conditions, and Supplementary Figure A.6 for comparison with wtCas9 binding.* ***(D)*** *Quantification of the gel in (C). Percent bound was determined by calculating the percent of the total band signal in each lane corresponding to Cas9-bound target as determined by a shift in mobility within the gel.* ***(E)*** *Summary of binding and cleavage results for each PAM tested. The ability of Cas9 to bind or cleave nucleosomal DNA at a targeted PAM is displayed as percent protection by the nucleosome, and was calculated by taking the ratio of binding or cleavage on nucleosomal DNA to that on naked DNA. While only the largest error bars are visible, replicates were performed for 15 of the 30 data points and are displayed with error bars showing standard deviation from the mean. The DNA positions plotted correspond to the three nucleotides of the targeted PAM. In order to compare percent protection from binding and cleavage with solvent accessibility, the PAMs are overlaid with a plot of the solvent accessible surface area for each strand (Watson or Crick) of DNA in the 601 nucleosome structure. The percent protection at each PAM, as well as the solvent accessibility were plotted so that the 5' end of each DNA strand begins at the left of the graph, where position 0 indicates the dyad.* ***(F)*** *Structural assessment of the ability of Cas9 to bind nucleosomal DNA. Superposition of the Cas9- guideRNA-DNA crystal structure (PDBID 4UN3, [247]) onto the 601 nucleosome crystal structure (PDBID 3LZ0, [243]) was achieved by alignment of the DNA path in both structures. (Left) To better view the alignment of the Cas9 target DNA with the nucleosomal DNA, the Cas9 protein density has been removed. (Right) After alignment of the DNA, inclusion of the Cas9 protein density reveals extensive steric clashes with the histones. Histone surface area was made partially transparent to better reveal the overlapping densities with Cas9.*

vitro, we turned to a chromatinized plasmid system. We dialyzed plasmid DNA containing a single 601 nucleosome positioning sequence with a sub-saturating quantity of purified histone octamers, and confirmed the quality of the resulting chromatin assemblies by MNase digestion (Figures 3.5-A,B). We tested whether a nucleosome was positioned at the 601 sequence using restriction enzyme accessibility mapping, and found that sites within the 601 sequence were protected from digestion while sites immediately adjacent were not, suggesting precise positioning of a high occupancy nucleosome (Figure 3.5-C). To recapitulate the effects of chromatin remodeling in vitro, we used a purified, truncated form of the Snf2-like chromatin remodeling enzyme Chd1 from Saccharomyces cerevisiae (yChd1), which had previously been shown to mediate nucleosome sliding in an ATP-dependent manner without

additional co-factors [250, 251]. To confirm yChd1 activity on our chromatinized plasmid, we used the frequent cutter, HaeIII, to digest the chromatin in the presence or absence of the remodeler. Upon addition of yChd1, we observed a shift toward lower molecular weight bands, indicative of dimished protection at HaeIII sites while still maintaining an overall chromatinized state (Figure 3.5-D).

We next sought to test whether yChd1 could affect Cas9's ability to access nucleosomal DNA. Before addition of the remodeler to our chromatinized plasmid, we found that sites within the 601 nucleosome were strongly protected from cleavage by Cas9, consistent with our mononucleosome results (Figure 3.5-E-G, PAM sites without remodeler). However, when the 601 nucleosome was remodeled by yChd1, as indicated by a loss of protection from restriction enzyme cleavage (approaching protection levels similar to those in the linker region), Cas9 cleavage efficiency was restored to around 80% of the corresponding naked plasmid control (Figures 3.5-EG). Notably, the percent protection at the EcoRI site adjacent to the positioned nucleosome did not decrease upon addition of yChd1, demonstrating that the decrease in protection along the 601 sequence was mediated by the nucleosome displacement activity of yChd1 rather than by a non-specific effect on cleavage efficiency. While our data with the chromatinized plasmid system confirm our findings that a well-positioned nucleosome provides a profound block to Cas9 cleavage, our further finding that chromatin remodeling restores access to nucleosomal DNA provides one potential mechanism by which Cas9 may efficiently modify broad portions of eukaryotic genomes. This plasmid model could be further exploited to assay the activity of Cas9 at nucleosome-free and boundary sites, and thus derive biophysical parameters governing Cas9-chromatin interactions.

3.3 Discussion

Despite its swift success as a repurposed tool for gene editing, imaging, and transcription modulation, the ability of the prokaryotic CRISPR/Cas9 system to effectively navigate eukaryotic chromatin has remained poorly understood. Here, we show that the nucleosome, the basic unit of chromatin, poses a strong barrier to Cas9, both in vitro and in vivo. By masking ~147 bp of DNA, the nucleosome effectively reduces the size of the eukaryotic genome available to Cas9. Previous studies using ChIP-seq to assay Cas9 binding have shown that off-target binding at PAM plus seed sequences more frequently occurs in regions of open chromatin [226, 252]. Our data expand upon these findings to show that the discrete pattern of nucleosome organization is able to modulate the efficiency of Cas9 binding and cleavage at on-target sites. The practical implications of these observations are underscored by our finding that accounting for nucleosome occupancy offers a significant improvement in predictive power for sgRNA design.

While our data show that nucleosomes strongly protect their DNA from Cas9 binding and cleavage in vitro, their organization in cells is not static. Transient displacement of nucleosomes occurs during replication, remodeling, and transcription. By adding the chromatin remodeling enzyme yChd1 to nucleosomes in vitro, we demonstrate that this displacement

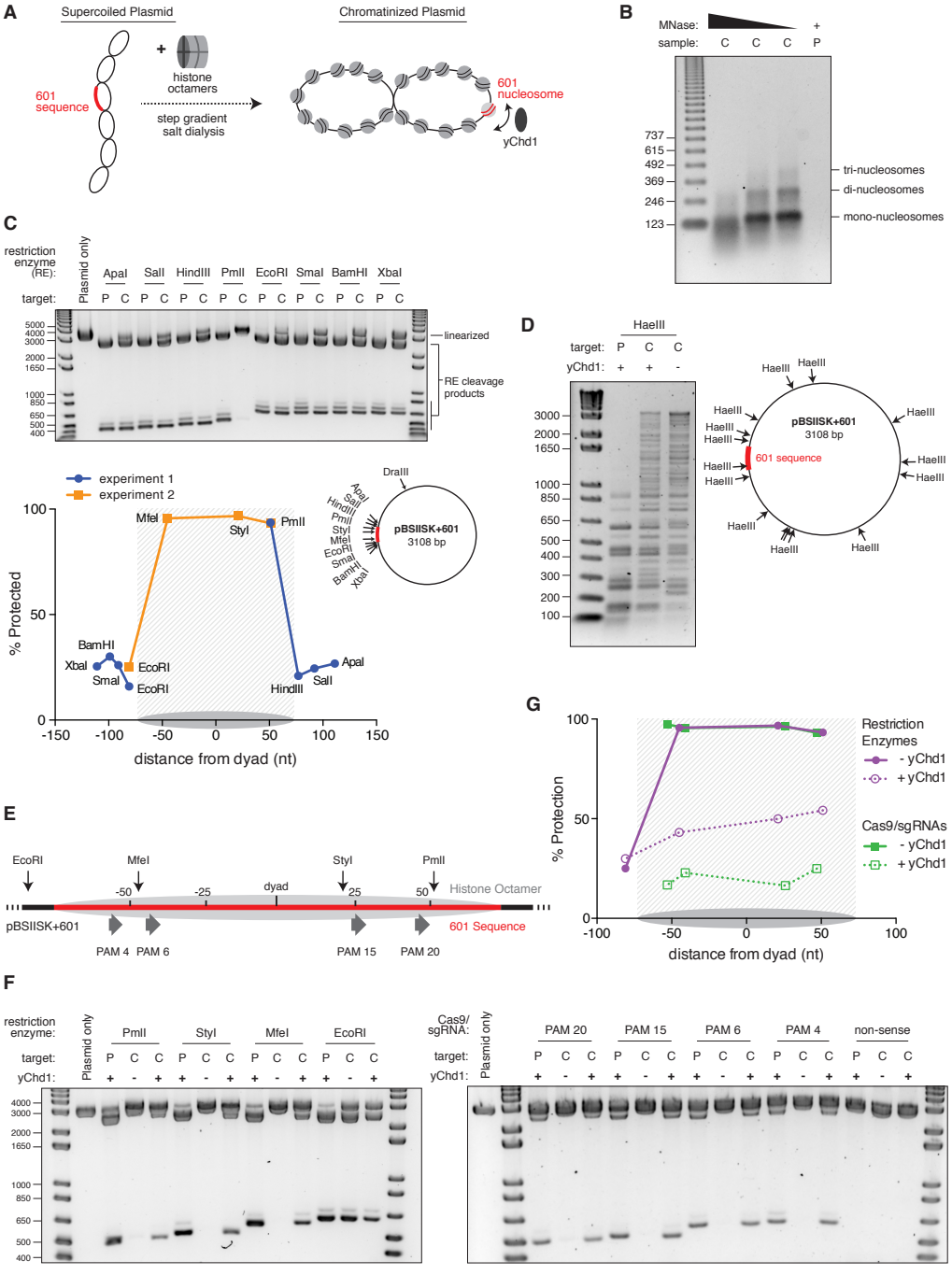


Figure 3.5: *Nucleosomes within chromatinized DNA can block cleavage by Cas9, but a chromatin remodeling factor can restore Cas9 access.* ***(A)*** *Schematic of the experimental setup. Supercoiled plasmid containing the 601 sequence inserted into a pBlueScript II SK (+) backbone (pBSIISK+601) was chromatinized by step gradient salt dialysis in the presence of histone octamer.* ***Continued on next page***

Continued Figure 3.5: *Purified yeast Chd1 (yChd1) remodeling factor was used to test the effect of ATP-dependent remodeling factors on Cas9 access to nucleosomal DNA.* ***(B)*** *Quality assessment of the chromatinized plasmid used in this study. Titrated amounts of Micrococcal Nuclease (MNase) were incubated with the chromatinized plasmid, and the resulting pattern of protection by assembled nucleosomes was visualized on a 1.3% agarose gel post-stained with ethidium bromide (EtBr). As a control, the supercoiled plasmid was also incubated with the lowest concentration of MNase.* ***(C)*** *A restriction enzyme accessibility assay (REAA) was used to assess the occupancy and position of the nucleosome assembled at the 601 sequence within the chromatinized plasmid. A panel of unique restriction enzyme sites spanning the 601 sequence were incubated with either the supercoiled plasmid, or the chromatinized plasmid. Cleavage was stopped, and protein was removed by incubation with proteinase K followed by Phenol:Chloroform:Isoamyl alcohol extraction and ethanol precipitation. (Top) The resulting DNA was then linearized using DraIII, and the level of cleavage by the restriction enzyme panel was visualized on a 1% agarose gel post-stained with EtBr. The label 'P' represents supercoiled plasmid, while 'C' represents chromatinized plasmid. (Bottom right) The location of the restriction sites used are indicated on a diagram of the plasmid. (Bottom left) After quantification of the gel, the percent protection from cleavage experienced in the chromatinized plasmid was plotted versus the location of the cleavage sites on the top strand of the 601 sequence. Experiment 1 refers to the REAA experiment shown in the gel above, while experiment 2 refers to the REAA experiment without remodeler shown in Figure 3.5-F. The grey shading indicates the borders of the 601 sequence, and the grey oval represents the corresponding nucleosome.* ***(D)*** *REAA experiment using the frequent cutter, HaeIII, to assess the remodeling activity around the chromatinized plasmid by the purified yChd1 chromatin remodeler. The resulting banding patterns were visualized on a 1.5% agarose gel post-stained with EtBr. Low molecular weight fragments indicate a high degree of HaeIII accessibility, while higher weight bands indicate protection from digestion.* ***(E)*** *Diagram showing the location of the restriction enzyme cleavage sites and the PAMs targeted by Cas9/sgRNA in the experiment shown in (F) and (G).* ***(F)*** *An accessibility assay was performed essentially as in C using either restriction enzymes or Cas9/sgRNAs in the presence or absence of the remodeler yChd1. The level of cleavage by the restriction enzyme panel (left) or Cas9/sgRNAs (right) was visualized on a 1.3% agarose gel post-stained with EtBr. A negative control was conducted with an sgRNA that had no sequence complementarity to the plasmid used (non-sense guide). The concentration of yChd1 used was the same as in panel (D)* ***(G)*** *Quantification of the gels shown in F. Percent protection from cleavage of the chromatinized plasmid in the presence or absence of the chromatin remodeler was calculated relative to the percent cleavage in the corresponding supercoiled plasmid control, and plotted at the location of the restriction enzyme cleavage sites or the center of the PAMs with respect to the 601 dyad.*

can in fact restore Cas9 access to DNA. However, despite brief exposure of nucleosomal DNA during remodeling and various other cellular processes, we still observe a clear anti-correlation between Cas9 activity and nucleosome occupancy in vivo, suggesting that the barrier to Cas9 target recognition exists even in a cellular environment. Indeed, the balance

between nucleosome disruption, turnover, and repositioning in the cell leads to the average level of occupancy and positioning at each site observed by MNase-seq [30, 230]. Thus, our data suggest that it is likely the overall effect of this average nucleosome positioning that leads to the observed anti-correlation with Cas9/sgRNA activity.

Additionally, it is important to note that there is likely a fundamental difference between applications that use dCas9 versus nuclease-active Cas9. Knock-down of transcription by CRISPRi likely requires persistent binding by dCas9 to continually block transcription, and would be largely ineffective during S-phase when transcription is globally shut down. In contrast, to make a genomic edit, Cas9 must succeed in cleaving DNA only once, and could potentially take advantage of nucleosome turnover during replication. The role these differences play in the dependence of Cas9 on nucleosome position is still not clear. We expect, however, that nucleosome position and occupancy will be of particular concern to applications that use the nuclease-dead Cas9 and require sustained binding. Future investigations into the role of cell cycle and nucleosome disruption may provide an additional piece to our understanding of the mechanism of Cas9 in eukaryotic cells. Furthermore, nucleosome organization represents only one aspect of eukaryotic chromatin, and thus, our results contribute a first step in understanding and exploiting how chromatin affects Cas9 activity to enable more sophisticated and precise rules for targeting Cas9.

3.4 Materials and methods

3.4.1 Analysis of CRISPRi sgRNA activity

The K562 dCas9-KRAB-BFP cell line was obtained from [228] and had been constructed from K562 cells obtained from ATCC. The resulting cell line tested negative for mycoplasma (MycoAlert Kit, Lonza, Basel, Switzerland) in regular screenings, and cytogenetic profiling by array comparative genomic hybridization (not shown) closely matched previous characterizations of the K562 cell line [253]. Data from 30 published [228] and unpublished screens (ML, BA, BB, CYP, MK, YC, JF, and JN, personal communication), conducted using the CRISPRi sgRNA library described in Gilbert et al., (2014) [228] in K562 cells constitutively expressing dCas9-KRAB-BFP, were processed through a standardized pipeline adapted from Bassik et al. (2013) [234], and Kampmann et al. (2013) [254]. Briefly, sgRNA phenotypes were calculated as the log2 enrichment of sequencing read counts between two conditions (e.g. initial and final timepoints for growth screens, untreated and treated for drug/toxin screens) and normalized to cell doubling differences where appropriate. Most screens were conducted in duplicate, and sgRNA phenotypes from the duplicates were averaged. To determine hit genes, each gene was given an effect size (average of strongest 3 sgRNA phenotypes by absolute value) and a confidence value (Mann-Whitney p-value of all ~10 sgRNAs compared to negative controls), and hits were selected using a score that integrates effect size and statistical confidence (3.1).

$$|[\text{effect } Z \text{ score}][\log_{10}(p\text{-value})]| \geqq 20 \text{ in any screen} \tag{3.1}$$

For genes with multiple TSS, each TSS was analyzed separately and the gene was assigned the highest score. Finally, sgRNA phenotypes were extracted for hit genes from the screen in which the gene scored as a hit and normalized to the average of the strongest 3 phenotypes to generate the 'sgRNA activity score'.

sgRNA positions were defined as the genomic coordinate of the 3' G of the NGG PAM (all genomic coordinates referenced in this text are from hg19). TSS positions were determined from the FANTOM5 project annotation (Riken) (http://fantom.gsc.riken.jp/5/datafiles/phase1.3/extra/TSS_ classifier/TSS_human.bed.gz; accessed March 2, 2015), using the downstream genomic coordinate of the corresponding 'p1@gene' BED file entry. All local averages were calculated in 50 bp windows centered around the indicated point. As sgRNA lengths including the PAM were ~24 bp and position was calculated relative to PAM, a window size of 50 bp captures all sgRNAs that directly neighbor the center point at either the 3' or 5' end. In order to quantify amplitude, the local averages were first smoothed using a low-pass Butterworth filter ($N = 4$, $Wn = 0.03$; SciPy signal processing module) and then peaks and troughs were calculated by determining the local maxima and minima, respectively. As described for Figure 3.1-D, Peak 0 was defined as the local maximum closest to the TSS, peak 1 was defined as the next maximum downstream of peak 0, and troughs were defined as the minima immediately downstream of the respective peaks.

3.4.2 MNase-seq data analysis

K562 MNase-seq data was obtained from the ENCODE consortium as processed continuous signal data (BigWig file format; accession number ENCFF000VNN, Michael Snyder lab, Stanford University). sgRNA target site signal was calculated as the average signal at all positions between the 5' end of the sgRNA and the 3' end of the PAM.

3.4.3 RNA-seq data analysis

K562 RNA-seq data were obtained from the ENCODE consortium as transcript quantifications (accession number ENCFF485YKK, Thomas Gingeras lab, Cold Spring Harbor Laboratories). Genes were assigned expression levels in units of FPKM according to their highest-expressed transcript.

3.4.4 Linear regression

Sequence score was calculated by passing the specified 30bp target site for each sgRNA to the on_target_score_calculator.py script provided by Doench and colleagues (accessed October 9, 2015) [221]. This sequence score, MNase signal as calculated above, and sgRNA length (base

pairs in protospacer and PAM) were each compared to sgRNA activity scores by Pearson correlation. Linear fits of the specified parameters were computed using multidimensional linear regression (Sci-kit learn linear_model package), and correction of the activity scores for sequence and length features was performed by subtracting the predicted scores based on the combined fit.

3.4.5 Ricin tiling screen

Screens for ricin susceptibility were performed essentially as previously described [228]. Briefly, K562 cells constitutively expressing Cas9-BFP from an SFFV (spleen focus-forming virus) promoter were transduced with our previously described pooled sgRNA tiling library packaged into lentivirus for a multiplicity of infection below 1. Duplicate screens were infected and subsequently treated independently. Infected cells were allowed to recover for 2 days, then selected with 0.75 mg/mL puromycin (Tocris) for 2 days, and finally allowed to recover from puromycin treatment for 2 days. Cells were then cultured for 19 days and were either treated with three pulses of 0.5 ng/ mL ricin administered for 24 hr and followed by re-suspension in fresh media, or passaged untreated. Genomic DNA was harvested from the endpoint untreated and treated samples and processed for high-throughput Illumina sequencing as previously described [228]. Screens were conducted at a minimum library coverage of 1000 cells per sgRNA, and sequenced to a median depth of ~500 reads per sgRNA. Phenotypes were calculated as $\log_2$ enrichments of read counts between untreated and treated conditions, normalized to cell doubling differences, and averaged between duplicates. Phenotype-signed Z scores were calculated by dividing all scores by the standard deviation of negative control phenotypes and then multiplying phenotypes by -1 for sgRNAs targeting genes shown to produce sensitizing phenotypes upon knockdown [234] such that positive values represent 'expected' phenotypes.

3.4.6 Protein purification

Mouse histones H2B(T115C) and H4 were recombinantly expressed in BL21(DE3)pLysS cells from expression plasmids gifted by Dr. Karolin Luger. Expression and purification of mH4 was conducted as previously described by the Luger lab [200, 255], while mH2B(T115C) was expressed and purified as described by the Cairns lab [256] with the following exception: after purification, histones were dialyzed against multiple changes of double distilled water and 1 mM b-mercaptoethanol (BME) before lyophilizing for storage. Purified recombinant mouse (Mus Musculus) histones H2A and H3 were gifts from Dr. Karolin Luger.

The labeling mutant, mH2B(T115C) was fluorescently labeled with ~five-fold molar excess Alexa Fluor 647 C2 maleimide dye (Thermo Fisher Scientific, Waltham, MA, USA) as follows. 2 mg of lyophilized mH2B(T115C) was dissolved at 2 mg/mL in labeling buffer (7 M GuHCl, 50 mM Tris-HCl pH 7.5, 1 mM TCEP) and nutated at room temperature for 30 min. 1 mg of the dye was then dissolved in 50 ml of anhydrous DMF under Argon gas, and approximately half of the dye solution was slowly mixed with the dissolved mH2B(T115C)

in the dark at room temperature to begin the labeling reaction. After nutating the reaction for 1 hr, the rest of the dissolved dye was slowly added, and the reaction was moved to 4° C overnight in the dark. In the morning, the reaction was quenched by adding over hundred-fold molar excess of BME.

Histone octamers were refolded and purified as previously described [200]. Specifically, 110 nmoles each mH3 and mH4 were refolded with 130 nmoles each mH2A and mH2B (T115C). The resulting octamer was concentrated using a 10,000 MWCO Spin-X UF concentrator (Corning, Tewksbury, MA, USA), then purified on a Superdex 200 HR (10/30) column (GE Life Sciences, Pittsburgh, PA, USA) using an Akta Explorer FPLC (GE Life Sciences, Pittsburgh, PA, USA) at 0.2 mL/min. Selected fractions were concentrated, flash frozen, and stored at -80° C until use.

A new purification scheme was conceived to achieve exceptionally high purity (d/wt)Cas9 (Supplementary Figure A.7). Nuclease active *S. pyogenes* 6His-Cas9-HaloTag-NLS was recombinantly expressed and purified from BL21(DE3)pLysS-Rosetta cells (Novagen/EMD Millipore, Darmstadt, Germany) using the expression plasmid pET302-6His-wtCas9-Halo-NLS, while the nuclease dead S. pyogenes 6His-dCas9(D10A,H840A)-HaloTag was recombinantly expressed and purified from BL21(DE3)pLysS cells using the expression plasmid pET302-6His-dCas9-Halo [227]. Bacterial cultures were grown in Terrific Broth II (MP Biomedicals, Santa Ana, CA, USA) at 37° C until an OD_{600} reached 0.4. Cultures were then transferred to an ice bath for ~15 min until an OD_{600} reached 0.5, at which point expression was induced with 0.2mM IPTG, and the cultures were moved to an 18° C shaker for 16 hr. Cells were harvested at 3000xg for 20 min, then resuspended in lysis buffer (500 mM NaCl, 50 mM Hepes pH 7.6, 5% glycerol, 1% Triton X-100, 10 mM imidazole, 1 mM benzamadine, 2.3 mg/mL aprotinin, 0.5 mM PMSF, and 1 tablet per 50 mL of Protease Inhibitor Cocktail (Roche, Basel, Switzerland). Cells were lysed using sonication on ice at 50% duty cycle, power 8, 30s on, 1 min off (Misonix/ Qsonica, LLC, Newtown, CT). The lysed cells were then ultracentrifuged at 4° C, 40K rpms, for 40 min to remove cell debris. The supernatant was then allowed to bind to Ni-NTA agarose resin (Qiagen, Hilden, Germany) by nutating at 4° C for 30 min. The resin containing bound (d/wt)Cas9 was poured into a mini column (Bio-Rad, Hercules, CA) and washed with 10 column volumes (CV) of lysis buffer, and 5 CVs of 20 mM Imidazole buffer (same as lysis buffer but with 20 mM imidazole). Elution of the (d/wt)Cas9 was achieved using 250 mM Imidazole buffer (same as lysis buffer but with 250 mM imidazole), and fractions were checked on an SDS-PAGE gel. Chosen fractions were pooled and diluted to a starting NaCl concentration of 200 mM using Buffer A (0 M NaCl, 50 mM Hepes pH 7.6, 5% glycerol, 1 mM DTT, 0.5 mM PMSF). A 5 mL HiTrap Q-HP (GE Life Sciences, Pittsburgh, PA, USA) and a 5 mL HiTrap SP-HP (GE Life Sciences, Pittsburgh, PA) column were attached in tandem (Q first in line) and equilibrated on an Akta FPLC at 10% Buffer B (same as Buffer A except with 2 M NaCl). The pooled (d/wt)Cas9 was filtered, then loaded onto the tandem columns at 2 mL/min. The columns were washed with 10% Buffer B until A_{280} and A_{260} returned to baseline, at which point the Q column was removed and (d/wt)Cas9 was eluted from the SP column using a gradient from 10% to 50% Buffer B over 10 CVs. Fractions were chosen using SDS-PAGE, pooled,

and dialyzed into storage buffer (200 mM NaCl, 50 mM Hepes pH 7.6, 5% glycerol, 1 mM DTT) (Supplementary Figure A.7). Aliquots were flash frozen in liquid nitrogen and stored at -80° C.

3.4.7 Target DNA purification

To make the fluorescent DNA used in this study, the 601 DNA sequence was amplified by PCR from plasmid pBSIISK+601 (parent 601 sequence plasmid gifted by K Luger) with the primers listed below. Two different PCR products were produced; one labeling the Watson strand of the 601 sequence with a 5' Cy3 dye (IDT, Coralville, IA, USA), and the other labeling the Crick strand. Large-scale (~2 mL) PCR reactions using in-house produced Pfu DNA polymerase were ethanol precipitated before loading onto a 20cm x 20cm 12% native TBE-PAGE gel for DNA purification. The PCR product was cut out of the gel, and the gel slice was crushed and soaked in 0.3 M sodium acetate pH 5.2 with multiple buffer changes. The pooled extract was ethanol precipitated, resuspended in 10 mM Tris-HCl pH 7.5, 10 mM NaCl, and stored at -20° C until use.

Watson primer pair
LW021: 5'– /5Cy3/ATCGGATGTATATATCTGACACGTGC –3'
LW022: 5'– ATCTGAGAATCCGGTGCCG –3'

Crick primer pair
LW025: 5'– /5Cy3/ATCTGAGAATCCGGTGCCGAG –3'
LW030: 5'– ATCGGATGTATATATCTGACACGTGC –3'

3.4.8 sgRNA production for in vitro experiments

sgRNA was produced by T7 transcription of a short DNA oligo template. To create this template, two oligos, one containing the T7 promoter and DNA target sequence, and the other encoding the invariable scaffolding of the sgRNA, were annealed and filled in using a single PCR cycle. The DNA template was ethanol precipitated, then transcribed using the T7 Quick High Yield RNA Synthesis Kit (New England Biolabs, Ipswich, MA) according to vendor instructions. The resulting sgRNAs were purified via 6% Urea-PAGE gel. The correct band was cut out and crushed and soaked in 0.3M sodium acetate pH 5.2. After ethanol precipitation, and resuspension in RNase-free water, aliquots were stored at -80° C until use. SgRNAs targeting the 601 sequence included 20 bp of complimentarity 5' to the targeted PAM, with the exception of PAM 4, which has room for only 19 bp of complimentarity. SgRNAs were used against PAMs 1, 315, and 1720. PAM 2 did not have a long enough target sequence available within the 601. The non-sense guides used in this study contained the target sequences 5'– ACATGTTGATTTCCTGAAA –3' or 5'– GATTTCACCTCTCAGCGCAT –3'.

Template oligonucleotides
5'– TTAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNNGTTTTAGAGC
TAGAAATAGC –3'
5'– AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTA
TTTTAACTTGCTATTTCTAGCTCTAAAAC –3'

3.4.9 Nucleosome assembly

Nucleosomes were assembled by salt gradient dialysis as previously described [200] using 50 mL home-made dialysis buttons and ~1 mM DNA. The Watson and the Crick labeled 601 DNA were assembled into separate nucleosomes. The best DNA:octamer molar ratio for optimal nucleosome assembly was selected by titrating octamer. The most homogenous assembly as judged by 6% Native TBE-PAGE gel was chosen for future assays.

3.4.10 Cleavage and binding assays

Cleavage and binding assays in Figures 3.3, 3.4, and their corresponding supplements were both conducted in the same manner, exchanging dCas9 for wtCas9 during binding reactions. First, the complete ribonucleoprotein complex was formed by incubating five-fold molar excess of the chosen sgRNA with (d/wt)Cas9 at 37° C for 10 min. Next, the DNA substrate (either naked DNA or nucleosome) was added to 40 nM (five-fold less than Cas9), and the reactions were returned to 37o for an hour. Reaction buffer contained 20 mM Hepes pH 7.5, 100 mM NaCl, 5 mM MgCl2, 1 mM EDTA, and 2.5 mg/mL insulin (Roche, Basel, Switzerland). Importantly, we found that including a nonspecific protein such as insulin prevents the nucleosomes from falling apart and getting lost to surfaces during binding and cleavage assays (Supplementary Figure A.5-B). Additionally, inclusion of insulin reduces nonspecific protein-protein interactions and aggregation. Binding reactions also contained 14.6% sucrose to allow direct loading onto a native gel. At 1 hr, binding reactions were loaded onto a pre-run 6% native TBE-PAGE gel at 4° C, while cleavage reactions were stopped using a 5X stop buffer (250 mM EDTA, 2% SDS), and prepared to load onto a 10% Urea-PAGE gel by adding 2X loading buffer (95% formamide, 20% DMSO, 5 mM EDTA, 0.025% Orange G) and heating to 95° C for 5 min before snap cooling on ice. Binding and cleavage reactions were repeated at least twice each for most of the PAMs targeted. Gels were imaged using a PharosFX Plus (Bio-Rad, Hercules, CA) and quantified using Image Lab (Bio-Rad, Hercules, CA). For binding and cleavage reactions, gels were scanned in the Cy3 fluorescence channel, while Alexa Fluor 647 fluorescence was also imaged for binding reactions. For cleavage reactions, labeling of either the Watson or Crick strand was chosen such that the fluorophore was always attached to the strand complimentary to the sgRNA used.

3.4.11 Structure alignment and solvent accessibility

Molecular graphics and analyses were produced using the UCSF Chimera package (supported by NIGMS P41-GM103311) [13]. The crystal structures for the 601 nucleosome ([243] PDBID 3LZ0) and the Cas9-sgRNA-DNA ternary complex ([247] PDBID 4UN3) were superimposed by aligning the DNA path in both structures using MatchMaker and manual manipulation. The solvent accessible surface area of the DNA in the 601 nucleosome structure was computed using residue areaSAS.

3.4.12 Reconstitution of nucleosomes on a plasmid (chromatin assemblies)

Supercoiled plasmid was produced using a Qiagen Maxi prep kit (Hilden, Germany). Histone octamers were prepared from purified histones as described above. DNA and histone octamer were mixed at a weight ratio of 1:0.8 (DNA to octamer) at a concentration of 0.4 mg/mL of DNA in 1X TE with 2 M NaCl. 50 mL homemade dialysis buttons were used to dialyse the solution by a step-wise gradient in 500 mL at room temperature. The gradient was as follows; 1.5 hr in 1 M NaCl in TE, 2 hr in 0.8 M NaCl in TE, 1.5 hr in 0.6 M NaCl in TE, and 2 hr in 0.05 M NaCl in TE. The resulting chromatin was stored at 4° C until use.

3.4.13 Micrococcal nuclease digestion assays of chromatin assembly reactions

MNase assays were performed essentially as in Torigoe et al., 2013 [257] and Alexiadis, 2002 [258]. MNase (Sigma-Aldrich, St. Louis) was resuspended at 200 units/mL in water and then disluted at 1:100, 1:500, or 1:1000 in a solution of 1X MNase reaction buffer (50 mM Tris-HCl pH7.9 and 5 mM CaCl2 dihydrate), 0.1 mg/mL insulin, and 10% glycerol. Chromatinized plasmid assemblies in 1X MNase buffer were added to each MNase dilution. As a control, an equal amount of unchromatinized super-coiled plasmid was added to the lowest MNase dilution. Each reaction was incubated for 11 min at room temperature, then stopped in a solution with final concentrations of 20 mM EGTA, 200 mM NaCl, 1% SDS, 20 mg/mL GlycoBlue (Thermo Fisher Scientific, Waltham, MA), and 13.4 mg/ mL Proteinase K (Roche, Basel, Switzerland) and incubated at 37° C for 30 min. Reactions wer Phenol/Chloroform/Isoamyl Alcohol extracted, ethanol precipitated, and run on a 1.3% agarose gel in 0.5X TBE. 3 mg of the 123 bp DNA ladder (Thermo Fisher Scientific, Waltham, MA) was used as a size standard.

3.4.14 Restriction enzyme accessibility assays

Either chromatin assemblies or super-coiled plasmid with equivalent amounts of DNA were added to restriction enzyme master mixes for final concentrations of 1X CutSmart buffer

(New England Biolabs, Ipswich, MA), 6 ng/mL DNA, and 0.4 U/mL of the indicated restriction enzyme (New England Biolabs). Reactions were incubated for 1 hr at NEB recommended temperatures. Reactions were stopped in final concentrations of 15 mM EDTA, 0.75% SDS, 150 mM NaCl, and 15 mg/mL Proteinase K and incubated at 37° C for 30 min. DNA was extracted with Phenol:Chloroform:Isoamyl Alcohol and ethanol precipitated, then resuspended in 1X CutSmart buffer and linearized by digesting with 1 U/mL DraIII-HF (New England Biolabs) at 37° C for 1 hr. The full reactions were run on an agarose gel and quantified as above.

3.4.15 Chromatin remodeling assays

Chromatin remodeling assays were performed simmilar to Torigoe et al., 2013 [257] and Alexiadis, 2002 [258]. Reactions contained 40 mM NaCl, 0.1 mg/mL BSA, 25 mM Tris-Acetate pH 7.5, 10 mM Mg-Acetate, 1 mM DTT, 1.2 mM ATP, and 6 ng/mL chromatinized or supercoiled plasmid. For restriction enzyme accessibility, the indicated restriction enzyme was added at 0.5 U/m L. For Cas9 accessiblity, Cas9/sgRNA ribonucleoproteins assembled as described above were added at 15.62 nM. A truncated version of the chromatin remodeling enzyme yChd1 was used spanning amino acids 1181274, also referred to as yChd1$_{\Delta NC}$ (gift of Dr. Ashok Patel and Dr. Gregory Bowman), and was added to the indicated reactions at a 0.2 mM final concentration. Reactions were incubated at 27° C for 1 hr, then processed as with the restriction enzyme accessibility assay above.

3.5 Acknowledgements

We would like to thank Dr. Karolin Luger, Pamela Dyer, and Dr. Uma Muthurajan for reagents and training in preparing and working with nucleosomes, and Dr. Ashok Patel and Dr. Gregory Bowman for the kind gift of yChd1. We would also like to thank Manuel Leonetti, Dr. Britt Adamson, Ben Barsi-Rhyne, Dr. Chong Y Park, Dr. Martin Kampmann, Yuwen Chen, Dr. Jonathan Friedman, and Dr. Jodi Nunnari for generously sharing unpublished screening data for determination of sgRNA activity. Additionally, we would like to thank members of the Tjian and Weissman labs, in particular, Dr. Elisa Zhang, Dr. Liangqi Xie, and Chiahao Tsui for their help with in vitro reagents, and Dr. Alex Fields and Joshua Dunn, for helpful discussions and assistance.

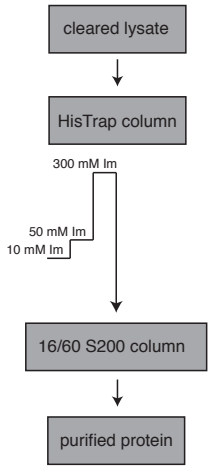
Appendix A

Supplementary figures

A

	Esa1	Piccolo NuA4	NuA4
organization:	catalytic subunit	subcomplex	complex
acetylation substrate:	histones	histones/ nucleosomes	histones/ nucleosomes
histone specificity:	all histones	H4 & H2A	H4 & H2A

B



C

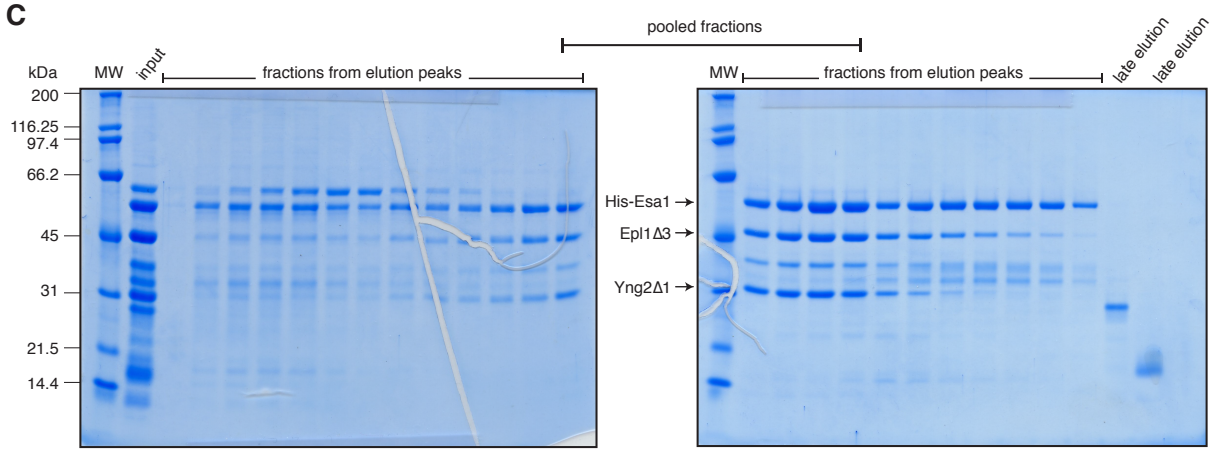


Figure A.1: *Purification of Piccolo NuA4.* ***(A)*** *Table listing the organization and specificity of the three levels of the NuA4 histone acetyltransferase complex used in this thesis.* ***(B)*** *Schematic of the procedure for purifying yeast Piccolo NuA4 from recombinant expression in E.coli.* ***(C)*** *SDS-PAGE gels showing the results of the final column in the purification of recombinant yeast Piccolo NuA4. The locations of the correct components are indicated with an arrow. Fractions pooled for use in this thesis are indicated.*

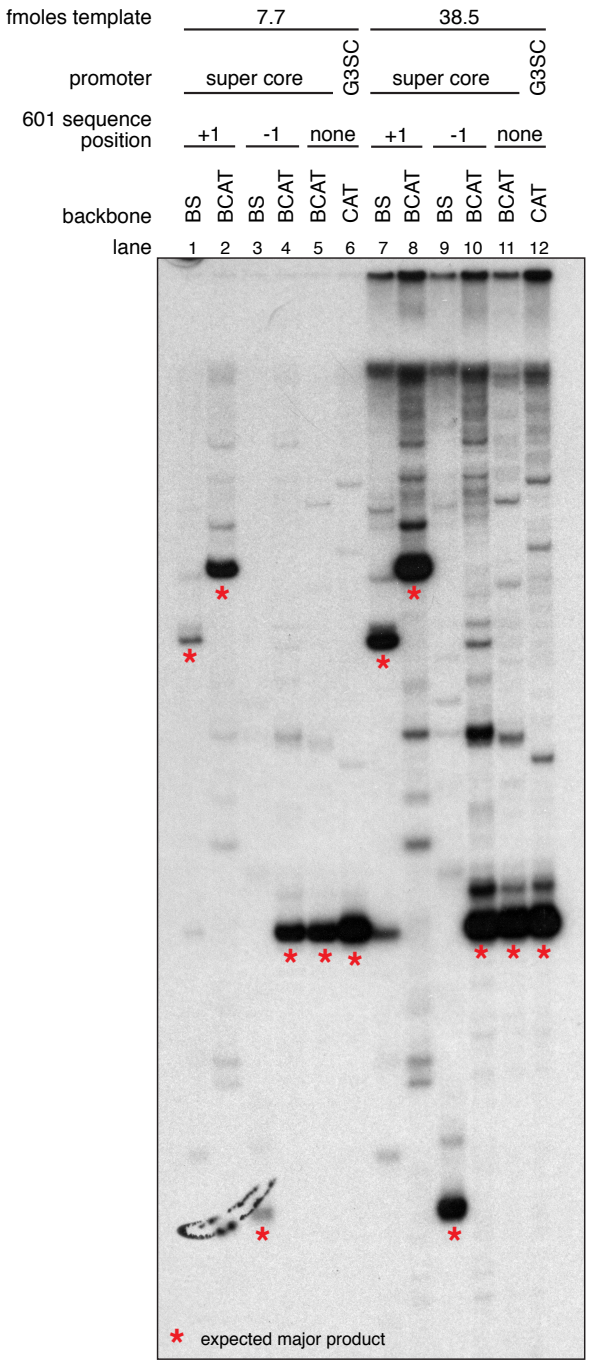


Figure A.2: *Insertion of the 601 sequence upstream or downstream of the super core promoter does not influence transcription, but the plasmid backbone does.*

Continued on the next page.

Continued Figure A.2: *Radiograph of a Urea-PAGE gel showing the results of in vitro transcription using purified basal factors and the indicated plasmids. After a 30 minute transcription at 37° C, transcripts were reverse transcribed (RT) using a ^{32}P labeled primer targeting the CAT gene present in constructs using the pBCAT and pCAT backbones. The constructs designed for modeling a -1 and +1 nucleosome for this study (shown in Figure 2.4) contain the same RT primer sequence inserted downstream of the Super Core promoter (SC). To assess the effects of the nucleosome positioning sequence on transcription, the 601 sequence was inserted into the pSC-BCAT plasmid. For comparison, a strong transcription template was used, pG3SC-CAT which contains 3 Sp1 binding sites ahead of a Super Core promoter. The location of the 601 sequence is labeled as +1 for downstream of the promoter, or -1 for upstream of the promoter. The expected products are indicated by red asterisks.*

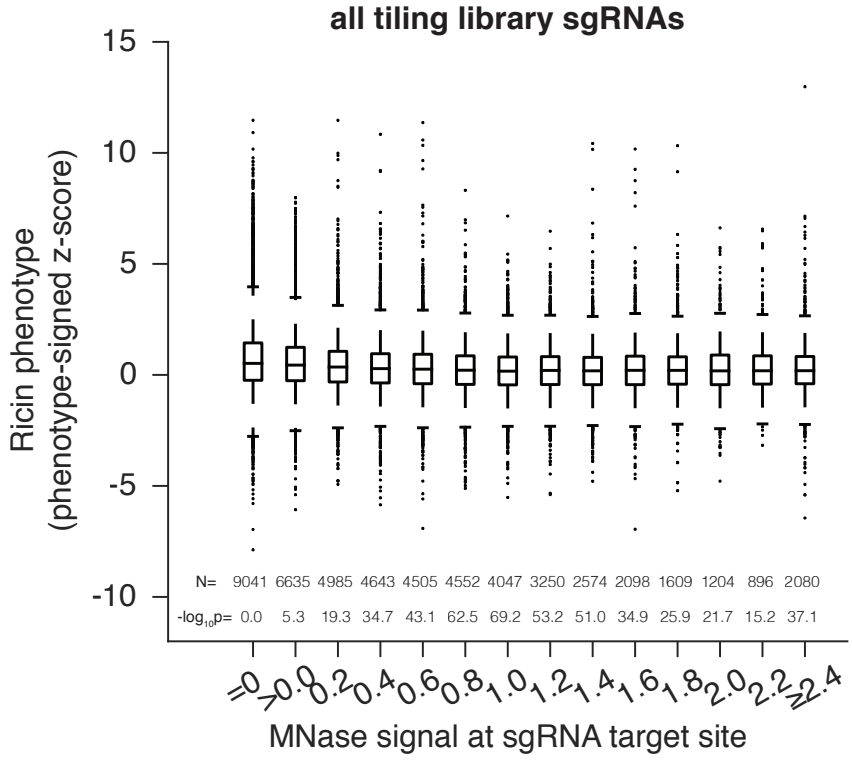


Figure A.3: *Cas9 nuclease activity anti-correlates with nucleosome occupancy at all target sites. Cas9 nuclease phenotypes and target site nucleosome occupancy for individual sgRNAs. As in Figure 3.2 A, but incorporating all CDS- and non-CDS-targeting sgRNAs in the ricin-susceptibility gene tiling library.*

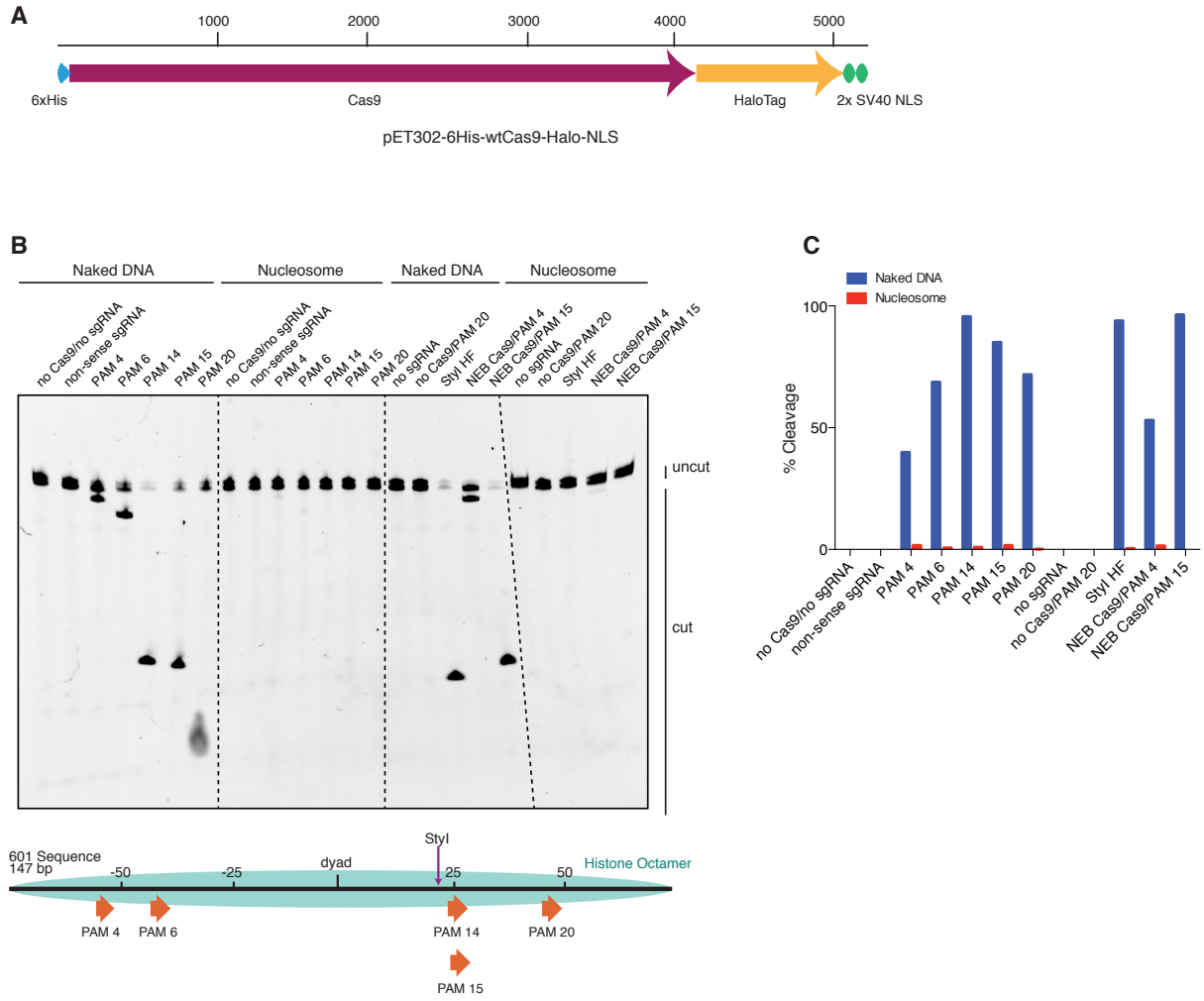


Figure A.4: *HaloTagged Cas9 activity is indistinguishable from untagged Cas9.* ***(A)*** *Diagram of the CDS region in the Cas9 expression plasmid used in this study.* ***(B)*** *Cleavage assay comparing the HaloTagged Cas9 construct used in this study with an untagged Cas9 commercially purchased from New England Biolabs (NEB, Ipswich, MA). Both forms of Cas9 were incubated with either naked DNA or the same DNA assembled into a nucleosome (see Figure 3.3 A). A positive control used the restriction enzyme, StyI-HF (from NEB), to target a sequence at a location within the DNA known to be fully protected upon assembly into a nucleosome. Unless explicitly labaled as NEB, all constructs of Cas9 (and dCas9) used were histidine tagged and HaloTagged.*

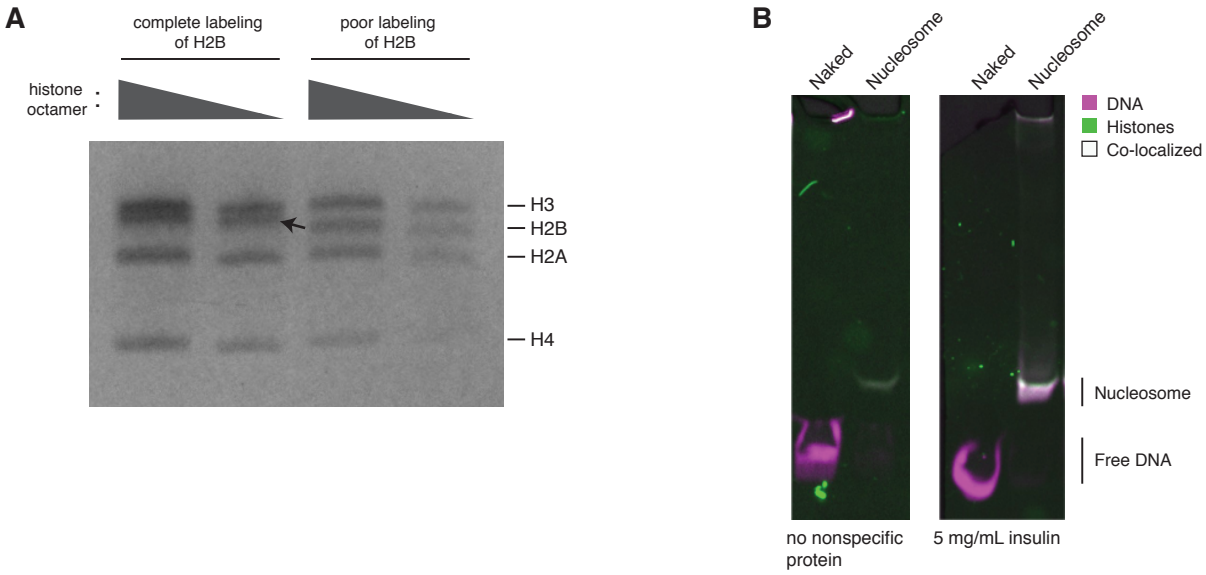


Figure A.5: *Quality controls.* ***(A)*** *H2B labeling is near complete. An SDS-PAGE gel after refolding and purifying the histone octamer containing fully labeled versus poorly labeled H2B. The arrow indicates a mobility shift of H2B corresponding to a fully labeled band. Each histone is present at equimolar ratios as indicated by PageBlue protein stain.* ***(B)*** *Nucleosome stability in Cas9 binding and cleavage assays is ensured by including a nonspecific protein in solution. On the left, a native PAGE gel showing naked DNA and nucleosomes under Cas9 binding and cleavage reaction conditions without a nonspecific protein in solution. On the right, the same conditions plus inclusion of a nonspecific competitor protein, insulin.*

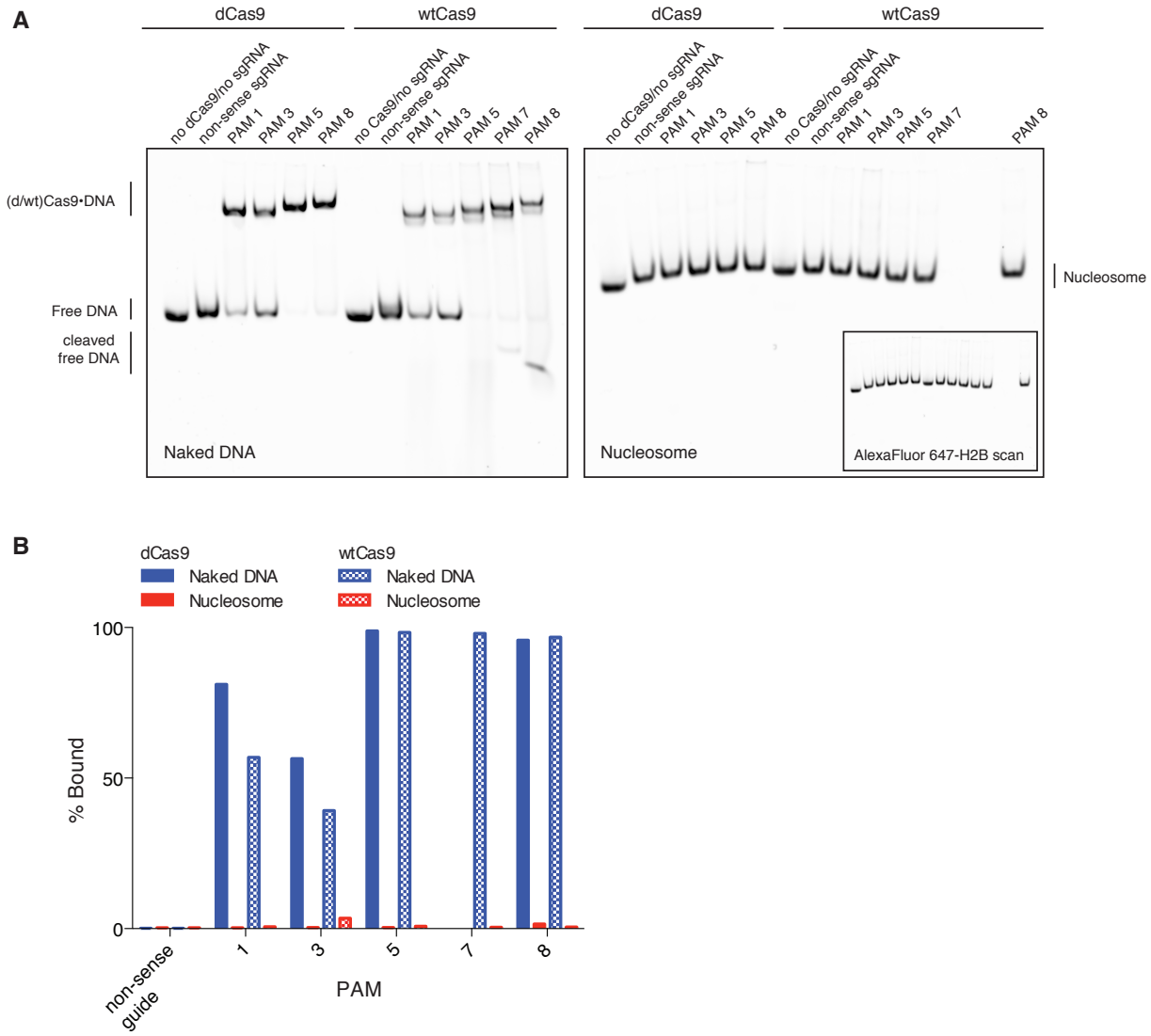


Figure A.6: *DNA binding by dCas9 is also representative of wtCas9 binding.* ***(A)*** *Two native PAGE gels showing the results of an EMSA binding assay comparing dCas9 and wtCas9. Binding to naked DNA is shown in the gel on the left, while binding to nucleosomes is shown in the gel on the right. Both gels were imaged in the Cy3-DNA channel, while the gel on the right was also imaged in the Alexa Fluor 647 H2B channel (inset image). Binding conditions for both dCas9 and wtCas9 were identical, and were as described in the methods.* ***(B)*** *Quantification of the Native PAGE gel in (A).*

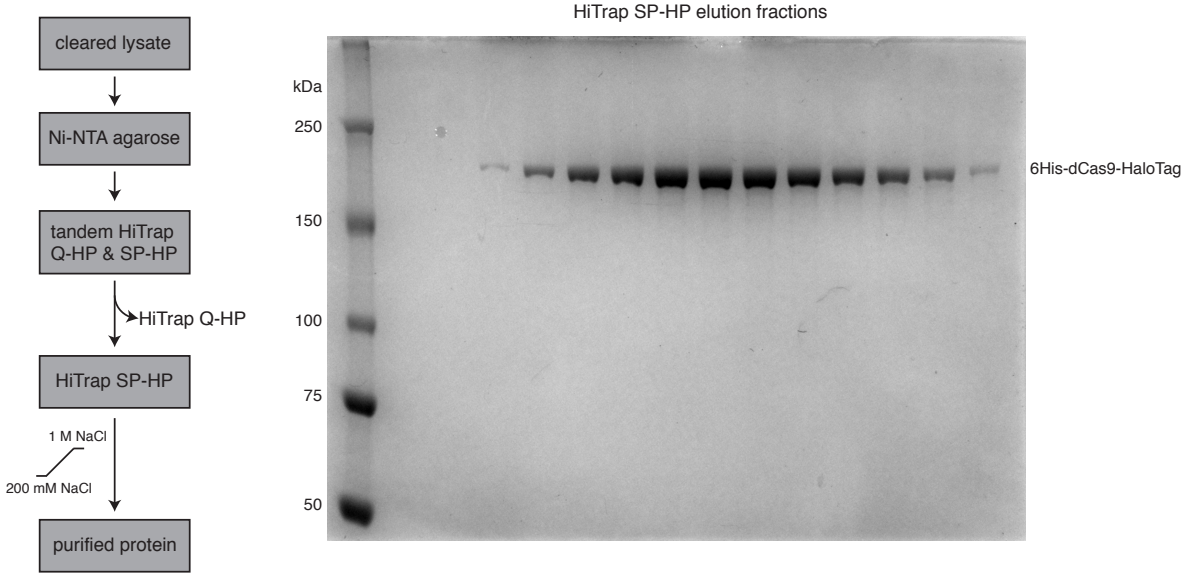


Figure A.7: *(wt/d)Cas9 purification strategy. Schematic showing the (wt/d)Cas9 purification strategy used in this study (left). SDS-PAGE gel stained with PageBlue (Life Technologies, Carlsbad, CA) showing the elution fractions from the HiTrap SP-HP column during purification of 6His-dCas9-HaloTag (right).*

Appendix B

Mouse Histone Expression and Purification

Mouse histones were purified using two strategies (Figure B.1). Histone H2A, H3, and H4 were purified while visiting the Luger lab in Colorado, and H2B was purified in the Tjian lab using a modified version of the Cairns lab procedure [256]. Both methods used in this thesis are detailed in the published *eLife* paper included in this thesis as Chapter 3, and can be found on page 60. Histone expression trials were carried out to identify optimal conditions for expression and purification and are detailed in Table B.1. Data for the purification of mouse histone H2B is found below (Figure B.2). Histone octamers were assembled as described in Chapter 3, page 60. The purification of the mouse histone octamer used in Chapter 2 of this thesis is shown in Figure B.3. Nucleosomes were reconstituted using salt gradient dialysis and a home made dialysis setup that incorporates techniques learned in the Luger lab, as well as from personal communication with Dr. Lacramioara Bintu and Dr. Toyotaka Ishibashi of the Bustamante lab. The home made dialysis buttons are created from the caps of small PCR tubes using dialysis tubing with a molecular weight cutoff at 6-8 kDa with unflavored floss attached in order to dangle the dialysis caps in buffer without being consumed by the stir bar (Figure B.4).

Table B.1: *Histone expression trials. Better expression is indicated with more (+). Those conditions that were chosen for large scale preps are indicated with •, while those conditions that failed to lyse well, despite good expression are indicated with §. Cell type abbreviations are as follows: BL21(DE3) (BD), BL21(DE3)pLysS (BDP), BL21-CodonPlus(DE3)-RIL (RILc+), BL21-CodonPlus(DE3)-RP (RPc+), BL21(DE3)-star (star). The media TBII abbreviates Terrific Broth II.*

histone	media	[IPTG]	induction time	cell type	induction OD_{600}	0.1% glucose	expression
H2B	2XYT	0.4 mM	2 hr	BD	0.4	+	
						-	
					0.6	+	+
						-	+
				BDP	0.4	+	
						-	+
					0.6	+	+
						-	+
				RILc+	0.4	+	
						-	
					0.6	+	+
						-	+
				RPc+	0.4	+	+
						-	+
					0.6	+	++
						-	++
			4 hr	BD	0.4	+	++++
				BDP	0.4	+	+++
				RPc+	0.4	+	++++ §
	TBII	0.2 mM	3.5 hr	BDP	0.5	-	+++++ •
				RP	0.5	-	+++++
				star	0.5	-	+++++
H2A	2XYT	0.4 mM	4 hr	BD	0.4	+	
				BDP	0.4	+	++++++ •
				RPc+	0.4	+	++
H3	2XYT	0.4 mM	2 hr	BDP	0.4	-	++++++ §

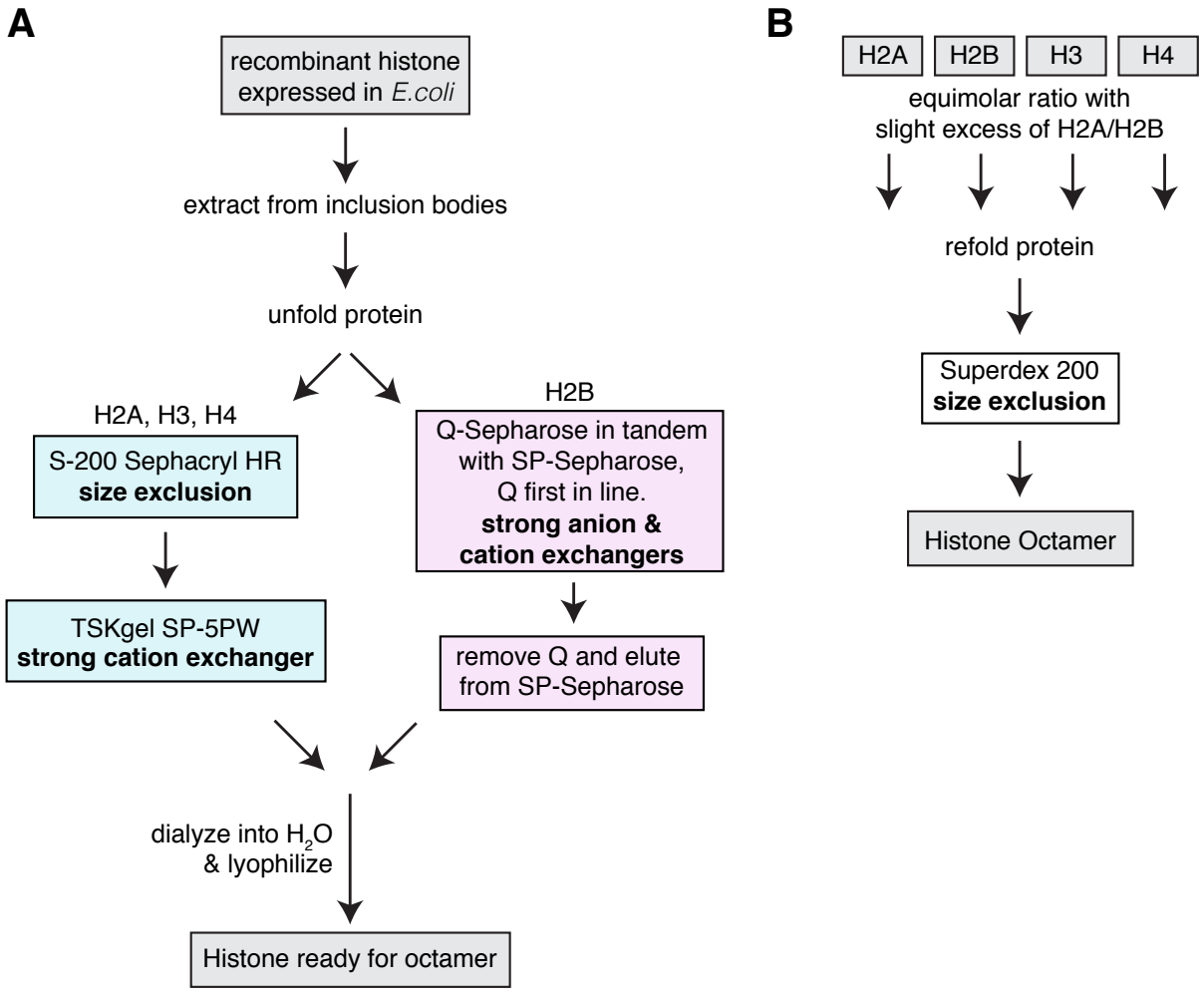


Figure B.1: *Histone purification scheme* ***(A)*** *Two purification schemes were employed to purify mouse histones. The scheme in blue was used to purify histones H2A, H3, and H4, while the scheme in pink was used to purify H2B.* ***(A)*** *Schematic of the purification strategy for creating a mouse histone octamer.*

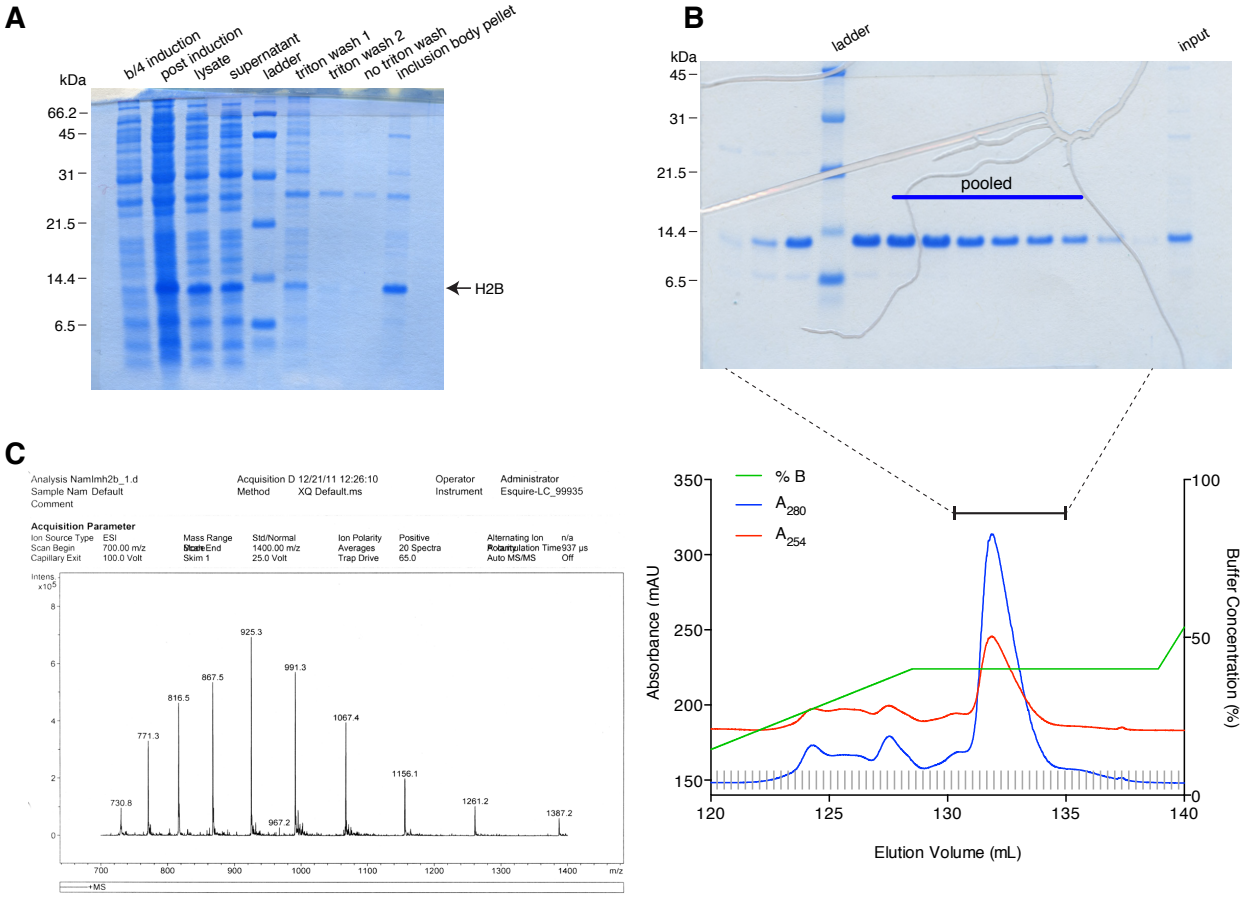


Figure B.2: *Purification of mouse H2B* ***(A)*** *SDS-PAGE gel showing the results of protein expression and inclusion body purification.* ***(B)*** *SDS-PAGE gel showing the results of H2B purification over a tandem Q and SP column. Fractions corresponding to the elution profile are indicated with a dotted line. Those fractions that were pooled for use in assembling the mouse histone octamers used in this thesis are indicated with a blue line.* ***(C*** *The results of mass–spec analysis performed by Dr. David King on the final purified mouse H2B showing a highly pure sample.*

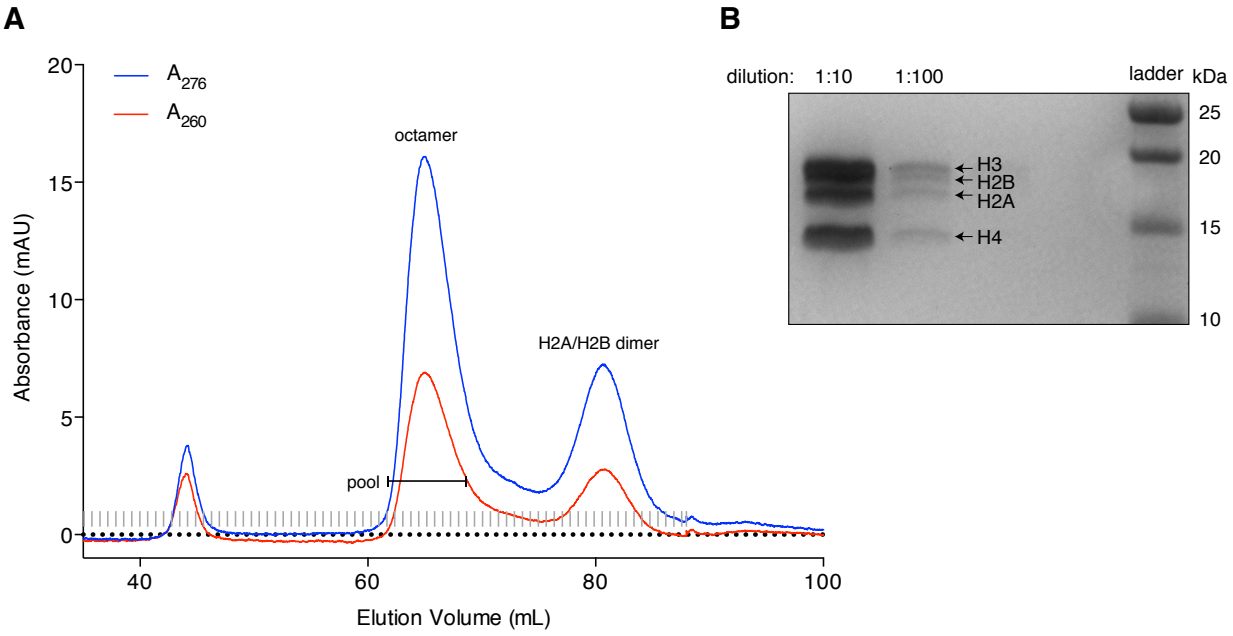


Figure B.3: *Refolding mouse histone octamer* ***(A)*** *Elution profile of the refolded mouse histone octamer separated on a Superdex 200 10/300 from GE Healthcare Life Sciences. The identity of the elution peaks are indicated, and the fractions pooled for use in nucleosome assemblies in this thesis are indicated.* ***(B)*** *18% Bis–Tris gel showing the final composition of the pooled octamer fractions from (A). Correct stoichiometry can be observed by the relatively even staining of all four histones.*

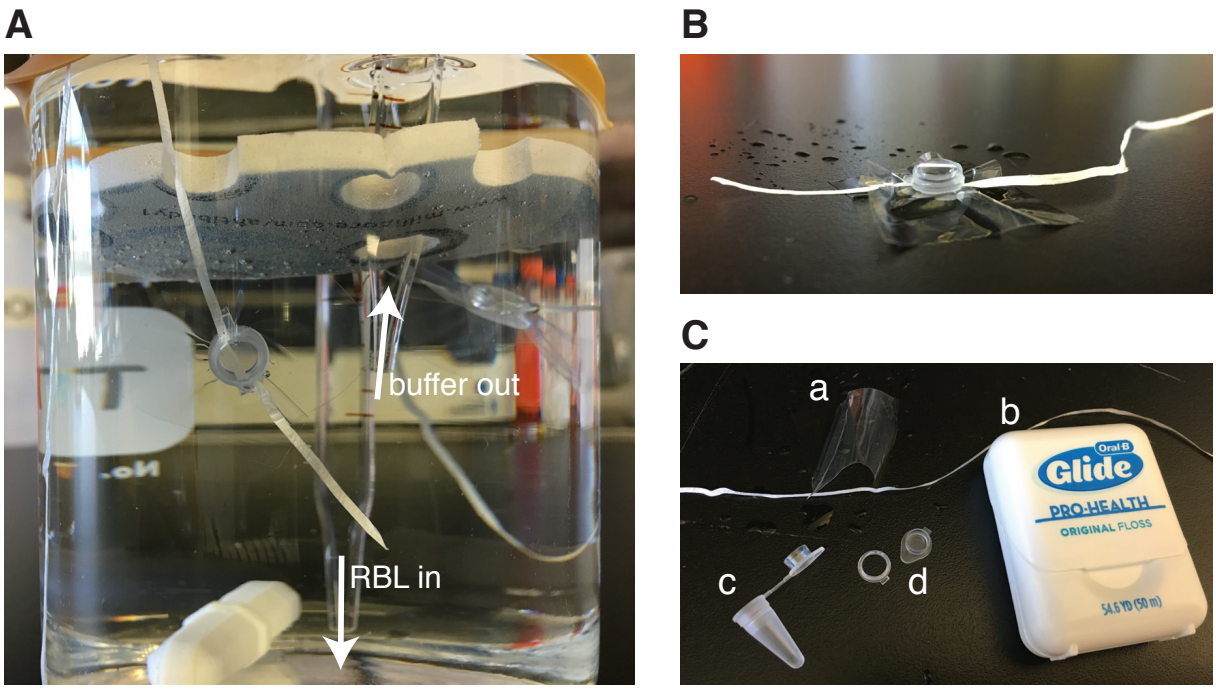


Figure B.4: *Apparatus for reconstituting nucleosomes.* ***(A)*** *Dialysis buttons are dangled in buffer using floss as high salt buffer is pumped out through a pipet as low salt buffer is pumped in to create the salt gradient.* ***(B)*** *Image of the home made dialysis buttons.* ***(C)*** *Components of the home made dialysis buttons. (a) 6-8 kDa MWCO dialysis tubing, (b) unflavored, waxed floss, (c) small PCR tube, (d) the result of cutting the top off the tube to make the bottom half of the dialysis tube, and the ring to hold on the tubing created by cutting a sliver of the PCR tube top.*

References

1. Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. *Nature Reviews Molecular Cell Biology* **4,** 809–814 (Oct. 2003).
2. Olins, A. L. & Olins, D. E. Spheroid Chromatin Units (ν Bodies). *Science* **183,** 330–332 (Jan. 1974).
3. Woodcock, C., Safer, J. P. & Stanchfield, J. E. Structural repeating units in chromatin: I. Evidence for their general occurrence. *Experimental cell research* **97,** 101–110 (1976).
4. Sahasrabuddhe, C. G. & Van Holde, K. E. The effect of trypsin on nuclease-resistant chromatin fragments. *The Journal of biological chemistry* **249,** 152–156 (Jan. 1974).
5. Rill, R. & Van Holde, K. E. Properties of Nuclease-Resistant Fragments of Calf Thymus Chromatin. *The Journal of biological chemistry* **248,** 1080–1083 (1973).
6. Van Holde, K. E. *Chromatin* (Springer Science & Business Media, 1989).
7. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–260 (Sept. 1997).
8. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311,** 532–537 (1984).
9. Luger, K. Structure and dynamic behavior of nucleosomes. *Current Opinion in Genetics & Development* **13,** 127–135 (Apr. 2003).
10. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS letters* **589,** 2914–2922 (2015).
11. McGinty, R. K. & Tan, S. Nucleosome Structure and Function. *Chemical reviews* **115,** 2255–2273 (Mar. 2015).
12. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *Journal of Molecular Biology* **319,** 1097–1113 (June 2002).
13. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25,** 1605–1612 (Oct. 2004).

14. Polach, K. J. & Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *Journal of Molecular Biology* **254,** 130–149 (Nov. 1995).

15. Choy, J. S. & Lee, T.-H. Structural dynamics of nucleosomes at single-molecule resolution. *Trends in Biochemical Sciences* **37,** 425–435 (Oct. 2012).

16. Li, G. & Widom, J. Nucleosomes facilitate their own invasion. *Nature Structural & Molecular Biology* **11,** 763–769 (Aug. 2004).

17. Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nature Structural & Molecular Biology* **12,** 46–53 (Jan. 2005).

18. Gansen, A. *et al.* Nucleosome disassembly intermediates characterized by single-molecule FRET. *Proceedings of the National Academy of Sciences* **106,** 15308–15313 (2009).

19. Boehm, V. *et al.* Nucleosome accessibility governed by the dimer/tetramer interface. *Nucleic acids research* **39,** 3093–3102 (Apr. 2011).

20. Brower-Toland, B. D. *et al.* Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proceedings of the National Academy of Sciences* **99,** 1960–1965 (Feb. 2002).

21. Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M. & Bustamante, C. Nucleosomal Fluctuations Govern the Transcription Dynamics of RNA Polymerase II. *Science* **325,** 626–628 (July 2009).

22. Alexandrov, B. S. *et al.* DNA breathing dynamics distinguish binding from nonbinding consensus sites for transcription factor YY1 in cells. *Nucleic acids research* **40,** 10116–10123 (Nov. 2012).

23. North, J. A. *et al.* Regulation of the nucleosome unwrapping rate controls DNA accessibility. *Nucleic acids research* **40,** 10215–10227 (Nov. 2012).

24. Luger, K. & Hansen, J. C. Nucleosome and chromatin fiber dynamics. *Current Opinion in Structural Biology* **15,** 188–196 (Apr. 2005).

25. Daujat, S., Zeissler, U., Waldmann, T., Happel, N. & Schneider, R. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *The Journal of biological chemistry* **280,** 38090–38095 (Nov. 2005).

26. Hergeth, S. P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO reports* **16,** 1439–1453 (Nov. 2015).

27. Sancho, M., Diani, E., Beato, M. & Jordan, A. Depletion of Human Histone H1 Variants Uncovers Specific Roles in Gene Expression and Cell Growth. *PLOS Genet* **4,** e1000227 (Oct. 2008).

28. Hellauer, K., Sirard, E. & Turcotte, B. Decreased expression of specific genes in yeast cells lacking histone H1. *The Journal of biological chemistry* **276,** 13587–13592 (2001).

29. Hughes, A. L. & Rando, O. J. Mechanisms underlying nucleosome positioning in vivo. *Annual review of biophysics* **43,** 41–63 (2014).

30. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics* **10,** 161–172 (Mar. 2009).

31. Drew, H. R. & Travers, A. A. DNA bending and its relation to nucleosome positioning. *Journal of Molecular Biology* **186,** 773–790 (Dec. 1985).

32. Olson, W. K. & Zhurkin, V. B. Working the kinks out of nucleosomal DNA. *Current Opinion in Structural Biology* **21,** 348–357 (June 2011).

33. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology* **276,** 19–42 (Feb. 1998).

34. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence Periodicities in Chicken Nucleosome Core Dna. *Journal of Molecular Biology* **191,** 659–675 (1986).

35. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458,** 362–U129 (2009).

36. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442,** 772–778 (2006).

37. Zhang, Z. *et al.* A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome. *Science* **332,** 977–980 (2011).

38. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences* **103,** 1412–1417 (Jan. 2006).

39. Juven-Gershon, T. & Kadonaga, J. T. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* **339,** 225–229 (Mar. 2010).

40. Ramirez-Carrozzi, V. R. *et al.* A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell* **138,** 114–128 (July 2009).

41. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474,** 516–520 (May 2011).

42. McCall, M., Brown, T. & Kennard, O. The crystal structure of d(G-G-G-G-C-C-C-C) a model for poly(dG) Δ poly(dC). *Journal of Molecular Biology* **183,** 385–396 (June 1985).

43. Moshkin, Y. M. *et al.* Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Molecular and cellular biology* **32,** 675–688 (Feb. 2012).

44. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nature Structural & Molecular Biology* **10,** 882–891 (Nov. 2003).

45. Thakar, A. *et al.* H2A.Z and H3.3 Histone Variants Affect Nucleosome Structure: Biochemical and Biophysical Studies. *Biochemistry* **48,** 10852–10857 (Oct. 2009).

46. Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3. 3 replacement patterns. *Nature genetics* **37,** 1090–1097 (2005).

47. Jin, C. & Felsenfeld, G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development* **21,** 1519–1529 (June 2007).

48. Quénet, D. & Dalal, Y. The CENP-A nucleosome: a dynamic structure and role at the centromere. *Chromosome Research* **20,** 465–479 (July 2012).

49. Turinetto, V. & Giachino, C. Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. *Nucleic acids research* **43,** 2489–2498 (Mar. 2015).

50. Luger, K., Suto, R. K., Clarkson, M. J. & Tremethick, D. J. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z - Nature Structural & Molecular Biology. *Nature Structural Biology* **7,** 1121–1124 (Dec. 2000).

51. Nekrasov, M. *et al.* Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. *Nature Structural & Molecular Biology* **19,** 1076–1083 (Nov. 2012).

52. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* **40,** 897–903 (July 2008).

53. Shahbazian, M. D. & Grunstein, M. Functions of Site-Specific Histone Acetylation and Deacetylation. *dx.doi.org* **76,** 75–100 (June 2007).

54. Sandman, K. & Reeve, J. N. Archaeal histones and the origin of the histone fold. *Current Opinion in Microbiology* **9,** 520–525 (Oct. 2006).

55. RIZZO, P. J. Those amazing dinoflagellate chromosomes. *Cell Research* **13,** 215–217 (Aug. 2003).

56. Ouzounis, C. A. & Kyrpides, N. C. Parallel origins of the nucleosome core and eukaryotic transcription from Archaea. *Journal of Molecular Evolution* **42,** 234–239 (Feb. 1996).

57. Koster, M. J. E., Snel, B. & Timmers, H. T. M. Genesis of chromatin and transcription dynamics in the origin of species. *Cell* **161,** 724–736 (May 2015).

58. Ammar, R. *et al.* Chromatin is an ancient innovation conserved between Archaea and Eukarya. *eLife* **1,** e00078 (Dec. 2012).

59. Peeters, E., Driessen, R. P. C., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nature reviews. Microbiology* **13,** 333–341 (May 2015).

60. Luijsterburg, M. S., Noom, M. C., Wuite, G. J. L. & Dame, R. T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *Journal of Structural Biology* **156,** 262–272 (Nov. 2006).

61. Zhang, Y., Griffin, K., Mondal, N. & Parvin, J. D. Phosphorylation of histone H2A inhibits transcription on chromatin templates. *The Journal of biological chemistry* **279,** 21866–21872 (May 2004).

62. Ancelin, K. *et al.* Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nature Cell Biology* **8,** 623–630 (June 2006).

63. Yamamoto, T. & Horikoshi, M. Novel Substrate Specificity of the Histone Acetyltransferase Activity of HIV-1-Tat Interactive Protein Tip60. *Journal of Biological Chemistry* **272,** 30595–30598 (Dec. 1997).

64. Kimura, A. & Horikoshi, M. Tip60 acetylates six lysines of a specific class in core histones in vitro. *Genes to cells : devoted to molecular & cellular mechanisms* **3,** 789–800 (Dec. 1998).

65. Schiltz, R. L. *et al.* Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. *The Journal of biological chemistry* **274,** 1189–1192 (Jan. 1999).

66. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* **40,** 897–903 (July 2008).

67. Zhang, L., Eugeni, E. E., Parthun, M. R. & Freitas, M. A. Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* **112,** 77–86 (Aug. 2003).

68. Altaf, M. *et al.* NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. *The Journal of biological chemistry* **285,** 15966–15977 (May 2010).

69. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431,** 873–878 (Oct. 2004).

70. Kapetanaki, M. G. *et al.* The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. *Proceedings of the National Academy of Sciences* **103,** 2588–2593 (Feb. 2006).

71. Kawashima, S. A., Yamagishi, Y., Honda, T., Ishiguro, K.-i. & Watanabe, Y. Phosphorylation of H2A by Bub1 Prevents Chromosomal Instability Through Localizing Shugoshin. *Science* **327,** 172–177 (Jan. 2010).

72. Puerta, C., Hernandez, F., Lopezalarcon, L. & Palacian, E. Acetylation of Histone H2A.H2B Dimers Facilitates Transcription. *Biochemical and Biophysical Research Communications* **210,** 409–416 (May 1995).

73. Galasinski, S. C., Louie, D. F., Gloor, K. K., Resing, K. A. & Ahn, N. G. Global regulation of post-translational modifications on core histones. *The Journal of biological chemistry* **277,** 2579–2588 (Jan. 2002).

74. Kawasaki, H., Taira, K. & Yokoyama, K. Histone acetyltransferase (HAT) activity of ATF-2 is necessary for the CRE-dependent transcription. *Nucleic Acids Symposium Series* **44,** 259–260 (Oct. 2000).

75. Cheung, W. L. *et al.* Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. *Cell* **113,** 507–517 (May 2003).

76. Pavri, R. *et al.* Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125,** 703–717 (May 2006).

77. Shema, E. *et al.* The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes & Development* **22,** 2664–2676 (Oct. 2008).

78. Zhu, B. *et al.* Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Molecular Cell* **20,** 601–611 (Nov. 2005).

79. Chen, D. *et al.* Regulation of Transcription by a Protein Methyltransferase. *Science* **284,** 2174–2177 (June 1999).

80. Koh, S. S., Chen, D., Lee, Y.-H. & Stallcup, M. R. Synergistic Enhancement of Nuclear Receptor Function by p160 Coactivators and Two Coactivators with Protein Methyltransferase Activities. *The Journal of biological chemistry* **276,** 1089–1098 (Jan. 2001).

81. Hyllus, D. *et al.* PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation. *Genes & Development* **21,** 3369–3380 (Dec. 2007).

82. Guccione, E. *et al.* Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* **449,** 933–937 (Oct. 2007).

83. Polioudaki, H. *et al.* Mitotic phosphorylation of histone H3 at threonine 3. *FEBS letters* **560,** 39–44 (Feb. 2004).

84. Dai, J., Sultan, S., Taylor, S. S. & Higgins, J. M. G. The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. *Genes & Development* **19,** 472–488 (Feb. 2005).

85. Briggs, S. D. *et al.* Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes & Development* **15,** 3286–3295 (Dec. 2001).

86. Roguev, A. *et al.* The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO journal* **20,** 7137–7148 (Dec. 2001).

87. Bryk, M. *et al.* Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in S. cerevisiae by a Sir2-independent mechanism. *Current biology : CB* **12,** 165–170 (Jan. 2002).

88. Nagy, P. L., Griesenbeck, J., Kornberg, R. D. & Cleary, M. L. A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. *Proceedings of the National Academy of Sciences* **99,** 90–94 (Jan. 2002).

89. Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences* **99,** 8695–8700 (June 2002).

90. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419,** 407–411 (Sept. 2002).

91. Lee, J.-H. & Skalnik, D. G. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *The Journal of biological chemistry* **280,** 41725–41731 (Dec. 2005).

92. Lee, J.-H., Tate, C. M., You, J.-S. & Skalnik, D. G. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *The Journal of biological chemistry* **282,** 13419–13428 (May 2007).

93. Milne, T. A. *et al.* MLL targets SET domain methyltransferase activity to Hox gene promoters. *Molecular Cell* **10,** 1107–1117 (Nov. 2002).

94. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics* **13,** 343–357 (May 2012).

95. Sierra, J., Yoshida, T., Joazeiro, C. A. & Jones, K. A. The APC tumor suppressor counteracts beta-catenin activation and H3K4 methylation at Wnt target genes. *Genes & Development* **20,** 586–600 (Mar. 2006).

96. Rybtsova, N. *et al.* Transcription-coupled deposition of histone modifications during MHC class II gene activation. *Nucleic acids research* **35,** 3431–3441 (2007).

97. Wang, H. *et al.* Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Molecular Cell* **8,** 1207–1217 (Dec. 2001).

98. Nishioka, K. *et al.* Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes & Development* **16,** 479–489 (Feb. 2002).

99. Wilson, J. R. *et al.* Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111,** 105–115 (Oct. 2002).

100. Zegerman, P., Canas, B., Pappin, D. & Kouzarides, T. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *The Journal of biological chemistry* **277,** 11621–11624 (Apr. 2002).

101. Richard, S., Morel, M. & Cléroux, P. Arginine methylation regulates IL-2 gene expression: a role for protein arginine methyltransferase 5 (PRMT5). *The Biochemical journal* **388,** 379–386 (May 2005).

102. Dacwag, C. S., Ohkawa, Y., Pal, S., Sif, S. & Imbalzano, A. N. The protein arginine methyltransferase Prmt5 is required for myogenesis because it facilitates ATP-dependent chromatin remodeling. *Molecular and Cellular Biology* **27,** 384–394 (Jan. 2007).

103. Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P. & Sif, S. Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Molecular and Cellular Biology* **24,** 9630–9645 (Nov. 2004).

104. Majumder, S. *et al.* Methylation of histone H3 and H4 by PRMT5 regulates ribosomal RNA gene transcription. *Journal of Cellular Biochemistry* **109,** 553–563 (Feb. 2010).

105. Grant, P. A. *et al.* Expanded lysine acetylation specificity of Gcn5 in native complexes. *The Journal of biological chemistry* **274,** 5895–5900 (Feb. 1999).

106. Schübeler, D. *et al.* Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes & Development* **14,** 940–950 (Apr. 2000).

107. Spencer, T. E. *et al.* Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature* **389,** 194–198 (Sept. 1997).

108. Guo, R., Chen, J., Mitchell, D. L. & Johnson, D. G. GCN5 and E2F1 stimulate nucleotide excision repair by promoting H3K9 acetylation at sites of damage. *Nucleic acids research* **39,** 1390–1397 (Mar. 2011).

109. O'Carroll, D. *et al.* Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. *Molecular and Cellular Biology* **20,** 9423–9433 (Dec. 2000).

110. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406,** 593–599 (Aug. 2000).

111. Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107,** 323–337 (Nov. 2001).

112. Nielsen, S. J. *et al.* Rb targets histone H3 methylation and HP1 to promoters. *Nature* **412,** 561–565 (Aug. 2001).

113. Vandel, L. *et al.* Transcriptional repression by the retinoblastoma protein through the recruitment of a histone methyltransferase. *Molecular and Cellular Biology* **21,** 6484–6494 (Oct. 2001).

114. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development* **16,** 919–932 (Apr. 2002).

115. Dodge, J. E., Kang, Y.-K., Beppu, H., Lei, H. & Li, E. Histone H3-K9 methyltransferase ESET is essential for early development. *Molecular and Cellular Biology* **24,** 2478–2486 (Mar. 2004).

116. Yang, L. *et al.* Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. *Oncogene* **21,** 148–152 (Jan. 2002).

117. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *The Journal of biological chemistry* **276,** 25309–25317 (July 2001).

118. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & Development* **16,** 1779–1791 (July 2002).

119. Kim, K.-C., Geng, L. & Huang, S. Inactivation of a Histone Methyltransferase by Mutations in Human Cancers. *Cancer Research* **63,** 7619–7623 (Nov. 2003).

120. Carling, T. *et al.* A histone methyltransferase is required for maximal response to female sex hormones. *Molecular and Cellular Biology* **24,** 7032–7042 (Aug. 2004).

121. Vakoc, C. R., Mandat, S. A., Olenchock, B. A. & Blobel, G. A. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Molecular Cell* **19,** 381–391 (Aug. 2005).

122. Wiencke, J. K., Zheng, S., Morrison, Z. & Yeh, R.-F. Differentially expressed genes are marked by histone 3 lysine 9 trimethylation in human cancer cells. *Oncogene* **27,** 2412–2421 (Apr. 2008).

123. Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106,** 348–360 (Nov. 1997).

124. Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102,** 279–291 (Aug. 2000).

125. Soloaga, A. *et al.* MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *The EMBO journal* **22,** 2788–2797 (June 2003).

126. Sassone-Corsi, P. *et al.* Requirement of Rsk-2 for Epidermal Growth Factor-Activated Phosphorylation of Histone H3. *Science* **285,** 886–891 (Aug. 1999).

127. Lo, W.-S. *et al.* Snf1–a Histone Kinase That Works in Concert with the Histone Acetyltransferase Gcn5 to Regulate Transcription. *Science* **293,** 1142–1146 (Aug. 2001).

128. Anest, V. *et al.* A nucleosomal function for IkappaB kinase-alpha in NF-kappaB-dependent gene expression. *Nature* **423,** 659–663 (June 2003).

129. Yamamoto, Y., Verma, U. N., Prajapati, S., Kwak, Y.-T. & Gaynor, R. B. Histone H3 phosphorylation by IKK-alpha is critical for cytokine-induced gene expression. *Nature* **423,** 655–659 (June 2003).

130. Preuss, U., Landsberg, G. & Scheidtmann, K. H. Novel mitosis-specific phosphorylation of histone H3 at Thr11 mediated by Dlk/ZIP kinase. *Nucleic acids research* **31,** 878–885 (Feb. 2003).

131. Kuo, M. H. *et al.* Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature* **383,** 269–272 (Sept. 1996).

132. Mizzen, C. A., Yang, X. J., Kokubo, T. & Brownell, J. E. The TAF II 250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87,** 1261–1270 (1996).

133. Daujat, S. *et al.* Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Current biology : CB* **12,** 2090–2097 (Dec. 2002).

134. Howe, L. *et al.* Histone H3 specific acetyltransferases are essential for cell cycle progression. *Genes & Development* **15,** 3144–3154 (Dec. 2001).

135. Schurter, B. T. *et al.* Methylation of Histone H3 by Coactivator-Associated Arginine Methyltransferase 1 †. *Biochemistry* **40,** 5747–5756 (May 2001).

136. An, W., Kim, J. & Roeder, R. G. Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell* **117,** 735–748 (June 2004).

137. Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development (Cambridge, England)* **136,** 3131–3141 (Sept. 2009).

138. Suka, N., Suka, Y., Carmen, A. A., Wu, J. & Grunstein, M. *Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin.* in *Molecular cell* (Aug. 2001), 473–479.

139. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 21931–21936 (Dec. 2010).

140. Cao, R. *et al.* Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298,** 1039–1043 (Nov. 2002).

141. Plath, K. *et al.* Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science* **300,** 131–135 (Apr. 2003).

142. Goto, H., Yasui, Y., Nigg, E. A. & Inagaki, M. Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. *Genes to cells : devoted to molecular & cellular mechanisms* **7,** 11–17 (Jan. 2002).

143. Zhong, S. *et al.* Ultraviolet B-induced phosphorylation of histone H3 at serine 28 is mediated by MSK1. *The Journal of biological chemistry* **276,** 33213–33219 (Aug. 2001).

144. Morris, S. A. *et al.* Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. *The Journal of biological chemistry* **282,** 7632–7640 (Mar. 2007).

145. Krogan, N. J. *et al.* Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Molecular and Cellular Biology* **23,** 4207–4218 (June 2003).

146. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO journal* **27,** 406–420 (Jan. 2008).

147. Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nature Cell Biology* **9,** 804–812 (July 2007).

148. Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *The Journal of biological chemistry* **286,** 7983–7989 (Mar. 2011).

149. Tanaka, Y., Katagiri, Z.-i., Kawahashi, K., Kioussis, D. & Kitajima, S. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* **397,** 161–168 (Aug. 2007).

150. Cheung, W. L. *et al.* Phosphorylation of histone H4 serine 1 during DNA damage requires casein kinase II in S. cerevisiae. *Current biology : CB* **15,** 656–660 (Apr. 2005).

151. Van Attikum, H. & Gasser, S. M. The histone code at DNA breaks: a guide to repair? *Nature Reviews Molecular Cell Biology* **6,** 757–765 (2005).

152. Barber, C. M. *et al.* The enhancement of histone H4 and H2A serine 1 phosphorylation during mitosis and S-phase is evolutionarily conserved. *Chromosoma* **112,** 360–371 (May 2004).

153. Wang, H. *et al.* Methylation of Histone H4 at Arginine 3 Facilitating Transcriptional Activation by Nuclear Hormone Receptor. *Science* **293,** 853–857 (Aug. 2001).

154. Strahl, B. D. *et al.* Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. *Current biology : CB* **11,** 996–1000 (June 2001).

155. Huang, S., Litt, M. & Felsenfeld, G. Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes & Development* **19,** 1885–1893 (Aug. 2005).

156. Balint, B. L. *et al.* Arginine methylation provides epigenetic transcription memory for retinoid-induced differentiation in myeloid cells. *Molecular and Cellular Biology* **25,** 5648–5663 (July 2005).

157. Li, X. *et al.* H4R3 methylation facilitates β-globin transcription by regulating histone acetyltransferase binding and H3 acetylation. *Blood* **115,** 2028–2037 (Mar. 2010).

158. Litt, M., Qiu, Y. & Huang, S. Histone arginine methylations: their roles in chromatin dynamics and transcriptional regulation. *Bioscience Reports* **29,** 131–141 (Apr. 2009).

159. Zhao, Q. *et al.* PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nature Structural & Molecular Biology* **16,** 304–311 (Mar. 2009).

160. Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P. & Sif, S. Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Molecular and Cellular Biology* **24,** 9630–9645 (Nov. 2004).

161. Girardot, M. *et al.* PRMT5-mediated histone H4 arginine-3 symmetrical dimethylation marks chromatin at G + C-rich regions of the mouse genome. *Nucleic acids research* **42,** 235–248 (Jan. 2014).

162. Sobel, R. E., Cook, R. G., Perry, C. A., Annunziato, A. T. & Allis, C. D. Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proceedings of the National Academy of Sciences* **92,** 1237–1241 (Feb. 1995).

163. Parthun, M. R., Widom, J. & Gottschling, D. E. The Major Cytoplasmic Histone Acetyltransferase in Yeast: Links to Chromatin Replication and Histone Metabolism. *Cell* **87,** 85–94 (1996).

164. Kruhlak, M. J. *et al.* Regulation of global acetylation in mitosis through loss of histone acetyltransferases and deacetylases from chromatin. *The Journal of biological chemistry* **276,** 38307–38319 (Oct. 2001).

165. Clarke, A. S., Lowell, J. E., Jacobson, S. J. & Pillus, L. Esa1p is an essential histone acetyltransferase required for cell cycle progression. *Molecular and Cellular Biology* **19,** 2515–2526 (Apr. 1999).

166. Smith, E. R. *et al.* ESA1 is a histone acetyltransferase that is essential for growth in yeast. *Proceedings of the National Academy of Sciences* **95,** 3561–3565 (Mar. 1998).

167. Allard, S. *et al.* NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *The EMBO journal* **18,** 5108–5119 (Sept. 1999).

168. Turner, B. M. & Fellows, G. Specific antibodies reveal ordered and cell-cycle-related use of histone-H4 acetylation sites in mammalian cells. *European journal of biochemistry / FEBS* **179,** 131–139 (Jan. 1989).

169. Winkler, G. S., Kristjuhan, A., Erdjument-Bromage, H., Tempst, P. & Svejstrup, J. Q. Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. *Proceedings of the National Academy of Sciences* **99,** 3517–3522 (Mar. 2002).

170. Jeppesen, P. & Turner, B. M. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell* **74,** 281–289 (July 1993).

171. Kleff, S., Andrulis, E. D., Anderson, C. W. & Sternglanz, R. Identification of a Gene Encoding a Yeast Histone H4 Acetyltransferase. *The Journal of biological chemistry* **270,** 24674–24677 (Oct. 1995).

172. Kelly, T. J., Qin, S., Gottschling, D. E. & Parthun, M. R. Type B histone acetyltransferase Hat1p participates in telomeric silencing. *Molecular and Cellular Biology* **20,** 7051–7058 (Oct. 2000).

173. Chang, L. *et al.* Histones in transit: cytosolic histone complexes and diacetylation of H4 during nucleosome assembly in human cells. *Biochemistry* **36,** 469–480 (Jan. 1997).

174. Akhtar, A. & Becker, P. B. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Molecular Cell* **5,** 367–375 (Feb. 2000).

175. Hilfiker, A., Kleiner, D. H., Pannuti, A. & Lucchesi, J. C. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. *The EMBO journal* **16,** 2054–2060 (Apr. 1997).

176. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development* **18,** 1251–1262 (June 2004).

177. Fang, J. *et al.* Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Current biology : CB* **12,** 1086–1099 (2002).

178. Nishioka, K. *et al.* PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Molecular Cell* **9,** 1201–1213 (June 2002).

179. Rice, J. C. *et al.* Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes & Development* **16,** 2225–2230 (Sept. 2002).

180. Beisel, C., Imhof, A., Greene, J., Kremmer, E. & Sauer, F. Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1 (Retracted article. See vol. 521, pg. 110, 2015). *Nature* **419,** 857–862 (2002).

181. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdisciplinary Reviews: Developmental Biology* **1,** 40–51 (Jan. 2012).

182. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nature Methods* **3,** 917–922 (Nov. 2006).

183. Tora, L. & Timmers, H. T. M. The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. *Trends in Biochemical Sciences* **35,** 309–314 (June 2010).

184. Danino, Y. M., Even, D., Ideses, D. & Juven-Gershon, T. The core promoter: At the heart of gene expression. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1849,** 1116–1131 (Aug. 2015).

185. Matangkasombut, O., Auty, R. & Buratowski, S. in *Proteins in Eukaryotic Transcription* 67–92 (Elsevier, 2004).

186. Malkowska, M., Kokoszynska, K., Rychlewski, L. & Wyrwicz, L. Structural bioinformatics of the general transcription factor TFIID. *Biochimie* **95,** 680–691 (Apr. 2013).

187. Lauberth, S. M. *et al.* H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. *Cell* **152,** 1021–1036 (2013).

188. Van Ingen, H. *et al.* Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3. *Structure (London, England : 1993)* **16,** 1245–1256 (Aug. 2008).

189. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131,** 58–69 (Oct. 2007).

190. Van Nuland, R. *et al.* Multivalent Engagement of TFIID to Nucleosomes. *PLOS ONE* **8,** e73495 (Sept. 2013).

191. Jacobson, R. H. Structure and Function of a Human TAFII250 Double Bromodomain Module. *Science* **288,** 1422–1425 (May 2000).

192. Matangkasombut, O., Buratowski, R. M., Swilling, N. W. & Buratowski, S. Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes & Development* **14,** 951–962 (Apr. 2000).

193. Ladurner, A. G., Inouye, C., Jain, R. & Tjian, R. Bromodomains Mediate an Acetyl-Histone Encoded Antisilencing Function at Heterochromatin Boundaries. *Molecular Cell* **11,** 365–376 (Feb. 2003).

194. Martinez-Campa, C. *et al.* Precise Nucleosome Positioning and the TATA Box Dictate Requirements for the Histone H4 Tail and the Bromodomain Factor Bdf1. *Molecular Cell* **15,** 69–81 (July 2004).

195. Danino, Y. M., Even, D., Ideses, D. & Juven-Gershon, T. The core promoter: At the heart of gene expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849,** 1116–1131 (Aug. 2015).

196. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116,** 699–709 (2004).

197. Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389,** 52–65 (Mar. 2007).

198. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483,** 295–301 (Mar. 2012).

199. Lemon, B., Inouye, C., King, D. S. & Tjian, R. Selectivity of chromatin-remodelling cofactors for ligand-activated transcription. *Nature* **414,** 924–928 (Dec. 2001).

200. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* **375,** 23–44 (2004).

201. Barrios, A. *et al.* Expression and purification of recombinant yeast Ada2/Ada3/Gcn5 and Piccolo NuA4 histone acetyltransferase complexes. *Methods* **41,** 271–277 (Mar. 2007).

202. Doyon, Y., SELLECK, W., Lane, W. S., TAN, S. & Cote, J. Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans. *Molecular and Cellular Biology* **24,** 1884–1896 (Feb. 2004).

203. Berndsen, C. E. *et al.* Nucleosome recognition by the Piccolo NuA4 histone acetyltransferase complex. *Biochemistry* **46,** 2091–2099 (Feb. 2007).

204. Anton Eberharter, P. B. B. Histone acetylation: a switch between repressive and permissive chromatin: Second in review series on chromatin dynamics. *EMBO reports* **3,** 224–229 (Mar. 2002).

205. Owen, D. J. *et al.* The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *The EMBO journal* **19,** 6141–6149 (2000).

206. Vollmuth, F., Blankenfeldt, W. & Geyer, M. Structures of the dual bromodomains of the P-TEFb-activating protein Brd4 at atomic resolution. *The Journal of biological chemistry* **284,** 36547–36556 (Dec. 2009).

207. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–231 (Mar. 2012).

208. Haberle, V. *et al.* Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507,** 381–385 (Mar. 2014).

209. Koerber, R. T., Rhee, H. S., Jiang, C. & Pugh, B. F. Interaction of Transcriptional Regulators with Specific Nucleosomes across the Saccharomyces Genome. *Molecular Cell* **35,** 889–902 (Sept. 2009).

210. Ryu, S. & Tjian, R. Purification of transcription cofactor complex CRSP. *Proceedings of the National Academy of Sciences* **96,** 7137–7142 (June 1999).

211. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5,** e12677 (Mar. 2016).

212. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346,** 1258096–1258096 (Nov. 2014).

213. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (Feb. 2013).

214. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (July 2013).

215. Maeder, M. L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nature Methods* **10,** 977–979 (July 2013).

216. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155,** 1479–1491 (Dec. 2013).

217. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 3002–3007 (Mar. 2015).

218. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33,** 510–517 (May 2015).

219. Kearns, N. A. *et al.* Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nature Methods* **12,** 401–403 (Mar. 2015).

220. Chari, R., Mali, P., Moosburner, M. & Church, G. M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods* **12,** 823–826 (Sept. 2015).

221. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology* **32,** 1262–1267 (Dec. 2014).

222. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343,** 80–84 (2014).

223. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Research* **25,** 1147–1157 (Aug. 2015).

224. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology* **13,** 722–736 (Nov. 2015).

225. Singh, R., Kuscu, C., Quinlan, A., Qi, Y. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research* **43,** e118–e118 (Oct. 2015).

226. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32,** 670–676 (July 2014).

227. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350,** 823–826 (2015).

228. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159,** 647–661 (Oct. 2014).

229. Johnson, S. M., Tan, F. J., McCullough, H. L., Riordan, D. P. & Fire, A. Z. Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. *Genome Research* **16,** 1505–1516 (Dec. 2006).

230. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474,** 516–520 (June 2011).

231. FANTOM Consortium and the RIKEN PMI and CLST (DGT) *et al.* A promoter-level mammalian expression atlas. *Nature* **507,** 462–470 (Mar. 2014).

232. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (Sept. 2012).

233. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (Sept. 2012).

234. Bassik, M. C. *et al.* A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **152,** 909–922 (Feb. 2013).

235. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527,** 192–197 (Nov. 2015).

236. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* **16,** 167–177 (Mar. 2015).

237. Denisov, D. A., Shpigelman, E. S. & Trifonov, E. N. Protective nucleosome centering at splice sites as suggested by sequence-directed mapping of the nucleosomes. *Gene* **205,** 145–149 (Dec. 1997).

238. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annual Review of Biochemistry* **84,** 165–198 (2015).

239. Segal, E. & Widom, J. What controls nucleosome positions? *Trends in Genetics* **25,** 335–343 (Aug. 2009).

240. Tilgner, H. & Guigo, R. From chromatin to splicing: RNA-processing as a total artwork. *Epigenetics* **5,** 180–184 (Oct. 2014).

241. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (Sept. 2012).

242. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162,** 108–119 (July 2015).

243. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal Structures of Nucleosome Core Particles Containing the '601′ Strong Positioning Sequence. *Journal of Molecular Biology* **403,** 1–10 (Oct. 2010).

244. Hinz, J. M., Laughery, M. F. & Wyrick, J. J. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry* **54,** acs.biochem.5b01108–7066 (Nov. 2015).

245. Tomschik, M., van Holde, K. & Zlatanova, J. Nucleosome dynamics as studied by single-pair fluorescence resonance energy transfer: a reevaluation. *Journal of fluorescence* **19,** 53–62 (Jan. 2009).

246. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (Mar. 2014).

247. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (Sept. 2014).

248. Flaus, A. & Owen-Hughes, T. Mechanisms for ATP-dependent chromatin remodelling: the means to the end. *The FEBS journal* **278,** 3579–3595 (Oct. 2011).

249. Hargreaves, D. C. & Crabtree, G. R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Research* **21,** 396–420 (Mar. 2011).

250. Patel, A. *et al.* Decoupling nucleosome recognition from DNA binding dramatically alters the properties of the Chd1 chromatin remodeler. *Nucleic acids research* **41,** 1637–1648 (Feb. 2013).

251. Patel, A., McKnight, J. N., Genzor, P. & Bowman, G. D. Identification of residues in chromodomain helicase DNA-binding protein 1 (Chd1) required for coupling ATP hydrolysis to nucleosome sliding. *The Journal of biological chemistry* **286,** 43984–43993 (Dec. 2011).

252. O'Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. & Segal, D. J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic acids research* **43,** 3389–3404 (Mar. 2015).

253. Naumann, S., Reutzel, D., Speicher, M. & Decker, H. J. Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization. *Leukemia research* **25,** 313–322 (Apr. 2001).

254. Kampmann, M., Bassik, M. C. & Weissman, J. S. Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **110,** E2317–26 (June 2013).

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

256. Wittmeyer, J., Saha, A. & Cairns, B. DNA translocation and nucleosome remodeling assays by the RSC chromatin remodeling complex. *Methods Enzymol* **377,** 322–343 (2004).

257. Torigoe, S. E., Patel, A., Khuong, M. T., Bowman, G. D. & Kadonaga, J. T. ATP-dependent chromatin assembly is functionally distinct from chromatin remodeling. *eLife* **2,** e00863 (2013).

258. Alexiadis, V. & Kadonaga, J. T. Strand pairing by Rad54 and Rad51 is enhanced by chromatin. *Genes & Development* **16,** 2767–2771 (Nov. 2002).