Title
The Evolutionary Roles of H2A and H2B in Genome Compaction in Eukaryotes

Author
Macadangdang, Benjamin

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Evolutionary Roles of H2A and H2B in Genome Compaction in Eukaryotes

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Molecular Biology

by

Benjamin Macadangdang

ABSTRACT OF THE DISSERTATION

The Evolutionary Roles of H2A and H2B in Genome Compaction in Eukaryotes

by

Benjamin Macadangdang

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2014

Professor Siavash Kurdistani, Chair

As eukaryotes have evolved from simple, unicellular organisms to more complex multicellular species, both the genome size and the nuclear volume have increased. However, the rates of increase of these two parameters have not been equal with genome size increasing much faster than the volume of the nucleus. This disproportionality has necessitated an increase in chromatin compaction for larger genome organisms. Although histones, through formation of the nucleosome, are the basic building blocks of chromatin, it is unknown whether histones have evolved to facilitate these higher compaction ratios. Analysis of histone sequences from 160 organisms representing all eukaryotic kingdoms revealed that there are significant changes in both the H2A N-terminus and H2B C-terminus sequences that have systematically evolved as genomes expanded. In the H2A N-terminus, larger genome species have acquired more arginines, and in the H2B C-terminus, lysines. In parallel with acquisition of positively charged residues, both H2A and H2B have lost polar residues. To determine whether these evolutionary

sequence changes contribute to increased genome compaction, a series of *in vivo* and *in vitro* molecular biological and biochemical experiments were carried out using both budding yeast and human cell lines as model systems. Insertion of precisely positioned arginines into the H2A N-terminus of a small-genome organism substantially increased chromatin compaction while their absence markedly diminished compaction in cells with large genomes. This effect was recapitulated using *in vitro* assembled nucleosomal arrays with unmodified histones, indicating that H2A N-terminus directly modulates the chromatin fiber likely through intra- and inter-nucleosomal arginine-DNA contacts to enable tighter compaction. Insertions of lysines into the H2B C-terminus of a small genome organism also increased compaction. Combinations of H2A-H2B sequence changes revealed that chromatin compaction is enhanced as compared to single histone mediated compaction and that the H2A N-terminus and H2B C-terminus work additively. An area physically near the H2B C-terminus in the nucleosomal structure, which I have termed the ABC Domain (H2A/H2B Compaction Domain), accounts for the majority of the chromatin compaction mediated by H2A and H2B. These findings reveal a simple evolutionary mechanism for regulation of chromatin compaction that has enabled organisms with larger genome achieve higher compaction ratios.

The dissertation of Benjamin Macadangdang is approved.

Michael F. Carey

Michael Grunstein

Amy Rowat

Stephen T. Smale

Siavash K. Kurdistani, Committee Chair

University of California, Los Angeles

2014

TABLE OF CONTENTS

FIGURES

TABLES

ACKNOWLEDGEMENTS

Submitted manuscript contributors (a reproduction of which is Chapter 2)

Evolution of Histone 2A for Chromatin Compaction in Eukaryotes

Benjamin R. Macadangdang[1, 2]†, Amit Oberai[1]†, Tanya Spektor[1]†, Oscar A. Campos[1,2], Fang Sheng[1], Michael F. Carey[1,2], Maria Vogelauer[1], and Siavash K. Kurdistani[1, 2, 3, 4]*

[1]Department of Biological Chemistry, [2]Molecular Biology Institute, [3]Department of Pathology and Laboratory Medicine, and [4]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.

†These authors contributed equally.

*Correspondence to: Siavash Kurdistani
Tel: 310.794.5194
skurdistani@mednet.ucla.edu

I undoubtedly need to thank Siavash for his guidance during my academic journey. He took me into his laboratory as a medical student still unsure about my professional course in life and gave me an appreciation for research that I will take with me in the future. I admire his open-ended thinking and curiosity towards biology and science which has allowed him to ponder extraordinary questions while still remaining grounded in sound scientific principles. His teaching has facilitated my studies and has laid the groundwork for my upcoming endeavors.

I also need to thank Maria Vogelauer for her guidance and support during my entire time in Siavash's laboratory. She not only taught me many of the techniques that I know, but she also helped me grow as a scientist. Her strong knowledge base provided a foundation for my learning and her academic input into these studies cannot be overstated. Additionally, she performed several of the experiments presented in Chapter 2. She performed the RNA expression analysis in yeast, the cell cycle experiments, the spot tests on different drugs, and provided the foundation for the MNase digestion experiments as well provide gels that we used for the submitted paper.

I would like to thank both Amit Oberai and Tanya Spektor for their contributions to this research. They were both equal contributors to the submitted paper, a portion of which is reproduced in Chapter 2. Amit started this project by gathering the data for the histone sequences and was first to establish the correlations to genome size. He also found the SNP associated with autism located in H2A from online databases, made several of the yeast strains, and did the growth curves. His computational work has been invaluable to our progress and his data was the basis of our work computational work in Chapter 3. Tanya's work includes all of our *in vitro* nucleosome assembly experiments as well as our nuclear area measurements in human cells. She also performed the IF part of the combined IF-FISH protocol in human cells.

I need to thank Oscar Campos for helping me with the competition assay experiments as well as for his discussions relating to our research and, in general, science as a whole. Our progress through this project was greatly assisted by his input. Beena Biju created several of our yeast strains used in Chapter 3, as well as performed some FISH experiments in H2B yeast mutants. Thanks to all the other past and present members of the Kurdistani Lab for help along the way.

My committee members, Michael Carey, Michael Grunstein, Amy Rowat, and Stephen Smale, have been very supportive and provided valuable advice to the project. I appreciate all the time and effort that they have put in to help me grow. I would not be where I am without their encouragement and assistance.

I would like to thank my family for their love and support throughout my lifetime. They have taught me the values that I carry with me. I am fortunate to have parents that are as encouraging and caring in all of my pursuits as they have been.

I would like to thank all the mentoring I received as an undergraduate by Chris Dwyer, as a post-baccalaureate by Chava Kimchi-Sarfaty and Michael Gottesman, and as a medical student by Philip Goodman. They have helped me tremendously during those stages of my life.

Finally, thanks to my buddy, David Fraley, whom I met when I moved to Los Angeles. Although he has Down syndrome, he has taught me more about life and happiness than most. I am lucky for the time we were able to spend together.

Funding

My funding sources have been the Howard Hughes Medical Fellows Program (2010-2011), the Phillip Whitcome Training Grant (2011-2013), and the UCLA Graduate Division Dissertation Year Fellowship (2013-2014).

VITA

Education

2003-2007	Electrical and Computer Engineering, B.S.E. Computer Science, B.S. Duke University, Durham, NC
2008-Present	Doctor of Medicine candidate Duke University School of Medicine, Durham, NC
2011-Present	Doctor of Philosophy, Molecular Biology candidate University of California, Los Angeles, CA

Publications and Presentations

Macadangdang B*, Zhang N*, Lund PE, Marple AH, Okabe M, Gottesman MM, Appella DH, & Kimchi-Sarfaty C. Inhibition of multidrug resistance by SV40 pseudovirion delivery of an antigene peptide nucleic acid (PNA) in cultured cells. PloS one 6, e17981 (2011). *=Equal contribution

Gillet, JP, **Macadangdang B**, Fathke RL, Gottesman M M & Kimchi-Sarfaty C. in Gene Therapy of Cancer 5-54 (Springer, 2009).

MBI Departmental Retreat, UCLA, Los Angeles, CA, January 2014; poster presentation

Biological Chemistry Floor Meeting, UCLA, Los Angeles, CA, January 2014; oral presentation

JCCC Gene Regulation monthly meeting, UCLA, Los Angeles, CA, December 2013; oral presentation

MBI Departmental Retreat, UCLA, Lake Arrowhead, CA; October 2012; poster presentation

Biological Chemistry Floor Meeting, UCLA, Los Angeles, CA, October 2012; oral presentation

Biological Chemistry Departmental Retreat, UCLA, Calmingos Ranch, CA, May 2012; poster presentation

Duke University AOA retreat, Duke University School of Medicine, Durham, NC, August 2011; poster presentation

JCCC Gene Regulation monthly meeting, UCLA, Los Angeles, CA, June 2011; oral presentation

Howard Hughes Medical Fellows Retreat, HHMI, Chevy Chase, MD, May 2011; poster presentation

CHAPTER 1

Introduction: Chromatin compaction in eukaryotes

1.1 DNA packaging within the domains of life

DNA molecules present inside all cells are not freely-floating entities but rather are organized in some manner, reducing their radius of gyration which is a measure of occupied volume. This reduction is accomplished through a variety of forces acting on DNA, such as supercoiling[1], macromolecule crowding[2], and DNA binding proteins[3-6]. The most basic of the DNA binding proteins are known as "DNA wrappers" and some form of these proteins exist in all domains of life, although the actual structure and protein sequences are not conserved. In bacteria, two types of proteins, Lrp (leucine-responsive regulatory protein) and HU (heat unstable) proteins, participate in wrapping DNA and, along with other DNA binding proteins, such as *Fis* and H-NS, maintain the nucleoid structure[7-10]. Archaea, which are more closely related to eukaryotes than are bacteria, contain histone-like proteins[11-13]. These proteins exist as dimers in solution with sequence similarities to two of the eukaryotic histones, H3 and H4, such as three alpha helices separated by β-strand loops[14]. The dimer of histone-like proteins then form a tetramer and can wrap approximately 60-90 basepairs (bps) of DNA within 1.5 positive toroidal (left-handed) supercoils[15,16]. However, unlike their eukaryotic histone counterparts, archaeal histone-like proteins lack N-terminal and C-terminal tails[17].

In eukaryotes, DNA resides in the nucleus and is organized into chromatin, which consists of the DNA, histone proteins, ncRNA, and the many different non-histone DNA binding proteins. The basic building block of chromatin, the "DNA wrapper" in eukaryotes, are the histone proteins, which form an octamer comprised of two copies of each of the four core histones – H2A, H2B, H3, and H4[18,19]. The octamer and the DNA wrapped around it are collectively known as the nucleosome. Besides dinoflagellates, which do not express histone genes but rather histone-like proteins[20], all eukaryotes contain histones with long N-terminal tails that are post-translationally

modified by a variety of marks, including acetylation, methylation, phosphorylation, ubiquitination, and sumoylation[21-26]. These marks can usually be precisely read by other proteins through domains like bromodomains, which recognize acetylated lysines, and chromodomains and Tudor domains, which recognize methylated lysines[27,28]. In this way, histones serve to organize the genome for DNA-based functions such as transcription, DNA replication, and DNA damage repair[29-31].

1.2 The hierarchy of chromatin compaction

1.2.1 The eukaryotic nucleosome

In the eukaryotic nucleus, chromatin is condensed through a hierarchy of processes which are able to compact DNA by many orders of magnitude. This hierarchy starts with the histone proteins themselves, which, as an octamer, directly interact with the double stranded DNA molecule through the intercalation of fourteen arginines into the DNA minor groove[19,32-34]. Doing so, the octamer is able to wrap 145-147 bps of DNA in 1.65 turns around the core[19,35]. The nucleosomal structure repeats with regular periodicity every 160-220 bps throughout the entire genome. This conformation is sometimes called the 10 nm fiber or "beads on a string," referring to the way chromatin appeared on electron micrographs when it was first visualized in 1973[36-38].

Because it has been crystalized several times, the above view of the nucleosome often gives the impression that the 146 bps of DNA and octamer of histones are static. However, it is now beginning to be appreciated that the nucleosome is dynamic and other conformations of it may exist *in vivo*. For instance, nucleosomal "breathing" was shown using Förster Resonance Energy Transfer (FRET) experiments where DNA becomes unwrapped from the nucleosome for 10-50

ms and stays wrapped for ~250 ms, possibly allowing recognition of the underlying DNA sequence[39]. Computational modeling has shown that a "gaping" nucleosome, where the H2A/H2B dimers detach from each other, can more efficiently stack adjacent nucleosomes and form a more compact fiber[40-42]. Other proposed conformations include the hexasome[43,44], tetrasome[45], reversome (R-octasome)[46], and lexosome (split nucleosome)[47], although the *in vivo* prevalence and biological implications of each are still unknown.

1.2.2 The role of histone tails

Histones also function beyond this first, basic level of chromatin compaction mediated directly by DNA interactions to the nucleosome core. Each of the four core histones has an N-terminal tail (often referred to as the N-terminal domain [NTD]) that protrudes out from the globular core of the nucleosome to mediate interactions with proteins and DNA. H2A has a long C-terminal domain (CTD) that protrudes from the nucleosome as well. The position of these tails within the nucleosome structure is relevant to how they can function[19,35]. The DNA molecule normally enters and exits the nucleosome on the same side (Figure 2-1A,C) where three out of the five tails are located. The H3 NTD is the longest of the tails and leaves the core between the DNA gyres near the entry/exit (Figure 2-1A-C). The H4 NTD and the H2A CTD both protrude from the surface of the nucleosome but the H2A CTD is in a unique position to directly interact with both the entering and exiting DNA with its docking domain (Figure 2-1A,C). On the opposite side of the nucleosome, the H2A NTD and H2B NTD protrude out, with the H2B NTD jutting between the DNA gyres and the H2A NTD from the surface (Figure 2-1A,B,D).

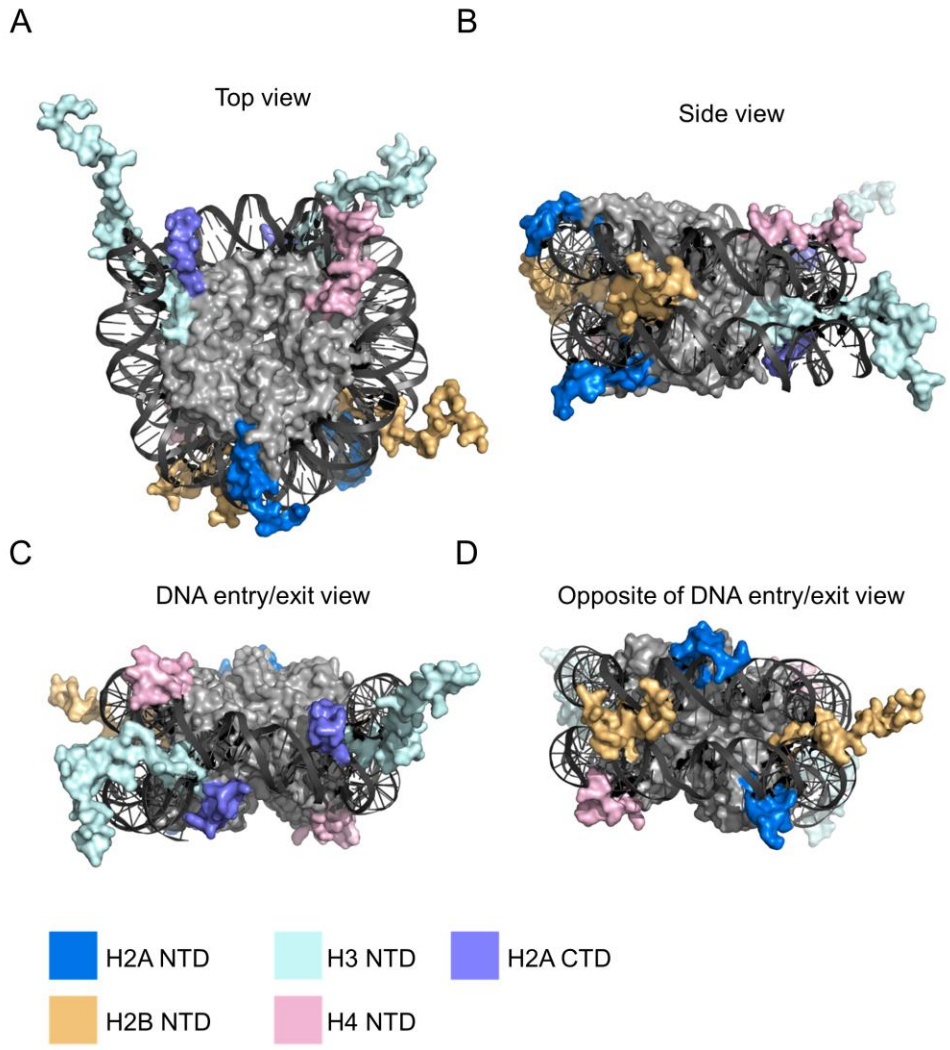

Figure 2-1 – The histone tails protrude from the nucleosome in a symmetrical manner. Nucleosomal structure (1KX5)[35] with each of the NTDs and the H2A CTD colored as indicated. **(A-D)** The orientation of the view is given above the picture. The histone cores are colored gray. DNA is colored black.

The functions of these tails in chromatin compaction can be divided into two main categories: 1) direct interactions of the NTDs and 2) modulations of those interactions by post-translational modifications. To probe the direct contributions of histone N-terminal tails, *in vitro* nucleosomal array reconstitution experiments were used with wildtype (WT) and mutant octamers. Normally, nucleosomal arrays are assembled by incubating histone octamers with a DNA template that contains a strong positioning sequence (e.g., the 601 sequence). Under low salt, these assembled

arrays form an open structure that corresponds to the "beads on a string" conformation[48]. This open structure progressively folds into higher order, more compact structures with the addition of higher concentrations of salt, mainly divalent cations[49]. Along with these maximally folded arrays, arrays can also oligomerize together under even higher divalent salt concentrations, which are thought to mimic long range fiber-fiber interactions[50].

When compared to WT histones, octamers containing tailless histones could still assemble onto the DNA template. However, tailless histones were unable to form maximally compacted structures and were unable to oligomerize[50-52]. In addition, aggregation of mononucleosomes was prevented when tailless histones, but not WT histones, were incubated with various cations and polyamines[53]. Interestingly, when hybrid octamers containing full H2A/H2B proteins but H3/H4 without tails or octamers containing full H3/H4 but tailless H2A/H2B were used, oligomerization could still occur but needed higher $MgCl_2$ concentrations than normal[54]. Although these hybrid arrays could oligomerize, they were unable to form maximally folded structures. These results indicate that all four tails are needed for fully compacted intra-array associations (array folding), but that the N-terminal tails may have a redundant function for inter-array associations (oligomerization). Additionally, these data suggest that the mechanism for array folding and oligomerization is distinct.

When the crystal structure of the nucleosome was finally solved by Luger *et al.* in 1997, she noted that during the crystallization, the H4 N-terminal tail crystalized near the H2A/H2B acidic patch of an adjacent nucleosome, effectively making an inter-nucleosomal contact[19,32,55]. Nucleosomal array reconstitution with individually deleted histone tails found that the H4 N-terminal tail, and specifically residues 14-19, are needed to form fully compacted arrays[56]. The

role of the interaction between H4 and the H2A/H2B acidic patch became clearer when homogenous acetylation of H4K16 prevented nucleosomal arrays from fully folding[57]. Subsequent studies have now confirmed this interaction through the ability to crosslink these structures *in vitro*[58], H4K16ac disrupting long array reconstitution[59], and cation-induced chromatin folding[60]. The *in vivo* relevance of H4K16ac has been recognized for some time as this modification is found at promoters of some genes[61] and is absent from yeast heterochromatin[62-64].

In addition to H4K16ac, another lysine on the H4 N-terminal tail, H4K20, is involved with chromatin compaction. *In vitro* nucleosomal arrays assembled with a trimethylated H4K20 (H4K20me3) exhibited larger sedimentation coefficients but did not oligomerize more easily[65]. This indicates that H4K20me3 is involved in local intra-array associations to maximally fold the chromatin fiber.

The H2B CTD is involved with chromatin compaction and heterochromatin formation. H2B K121 is a lysine that is present across eukaryotes within the highly conserved motif of AVTKY. It is known to be post-translationally modified by monoubiquitination[66,67]. Using *in vitro* nucleosomal arrays with chemically modified tails, ubiquitination of H2B (uH2B) was found to inhibit both maximally folded array formation as well as oligomerization[68]. The authors found that uH2B prevents the transition of intermediate folded arrays to maximally folded arrays and that effect is specific for ubiquitin because Hub1, a similar sized protein to ubiquitin, did not prevent maximal folding. These results indicate that uH2B causes chromatin to adopt an open structure which increases accessibility and correlates well with the *in vivo* observations that uH2B is present at some active genes with RNA Pol II[69-71]. Another residue within this

conserved motif, T119 (T122 in yeast), is needed for proper gene silencing at yeast telomeres[72]. This effect is independent of the ubiquitinated state of H2BK121.

1.2.3 The role of histone variants

Histone variants are isoforms of canonical histones that can substitute into the nucleosome structure. The primary amino acid sequence of each variant differs from the canonical form which provides them with unique functions. Unlike canonical histones that are expressed only during S phase, variant histones are expressed throughout the cell cycle[73]. H4 is the only histone for which no variant has been discovered in humans, although some variants have been discovered in other organisms[74-76]. H2B has a variant that is expressed in germ cells such as the testis and only differs slightly from the canonical form[77,78]. In contrast, both H3 and H2A have variants that are more ubiquitously expressed and the functions and effects of which on genome compaction are better understood.

Most eukaryotes contain two variants of H3. At the centromeres, nucleosomes contain a special H3 variant called CENP-A or CenH3 (Cse4p in yeast). This variant is needed for proper kinetochore assembly and attachment during mitosis[79,80]. CenH3 affects chromatin architecture though several different mechanisms. First, unlike the canonical nucleosome that wraps DNA in a left-handed wrap, nucleosomes with CenH3 induce positive supercoils implying a right-handed wrap[81]. The consequence of this difference is proposed to protect the centromeric chromatin from fully condensing during mitosis and thus allowing assembly of the kinetochore[82]. Secondly, nucleosomes containing CenH3 protect fewer basepairs of DNA (120 bps) than nucleosomes containing canonical histones (145-147 bps)[19,83,84]. CenH3 is proposed to exist *in vivo* in a tetrameric form[83], but it was crystalized in an octamer[84].

Higher eukaryotes contain an additional H3 variant, H3.3 that differs in humans by only four amino acids. Although H3.3 is found at transcriptionally active genes[85], its effect on chromatin architecture is unknown. Interestingly, H3.3's N-terminal tail is modified by post-translational modifications more easily than the canonical H3 NTD, but this observation may be a consequence of its genomic positioning rather than structural differences[86].

In comparison to the other canonical histones, H2A is the most diverse and contains the greatest number of variants. To date, there are four main H2A variants in higher eukaryotes – H2A.Z, H2A.X, H2A.Bbd, and macroH2A[87]. Plants also have an additional H2A variant, H2A.W, that has an extended CTD that interacts with the linker region to protect an additional 16 bps of DNA[88]. Similar to H2A.W, most of the H2A variants differ from the canonical H2A in the CTD as well as L1 region where it interacts with the other H2A in the nucleosome structure.

H2A.Z was first described in 1980 by West and Bonner[89] and it seems to have only evolved in eukaryotes once[90]. In higher eukaryotes such as mammals and flies, H2A.Z is an essential protein[91,92] but in the yeasts *S. cerevisiae* and *S. pombe*, strains with the gene knocked out are still viable[93-96]. Within the genome, H2A.Z is usually found near transcription start sites, such as the nucleosomes immediately surrounding the nucleosome-free region at promoters[97,98]. Current literature shows no consensus for the function of H2A.Z, as some reports indicate that it is gene activating while others say it represses transcription[99-102]. In terms of chromatin architecture, H2A.Z has an extended acidic patch and *in vitro* nucleosomal array reconstitution experiments demonstrated that arrays containing H2A.Z more readily fold due to this increased area[103,104].

H2A.X is another H2A variant that was described by West and Bonner but, unlike H2A.Z, it seems to have arisen multiple times throughout evolution[89]. H2A.X has a CTD motif SQ[E/D]Φ

where Φ represents a hydrophobic residue. In higher eukaryotes, H2A.X exists as a separate isoform, but in yeast, such as *S. cerevisiae*, this unique CTD motif is incorporated into the canonical H2A protein[96]. Distinct from H2A.Z, H2A.X is a nonessential variant because organisms such as *Caenorhabditis elegans* do not encode for an H2A.X isoform[105,106]. The serine within the CTD motif becomes phosphorylated upon DNA damage (γH2A.X) and spreads over long distances within the genome[107]. In terms of chromatin architecture, γH2A.X is thought to recruit both chromatin remodeling factors and DNA damage repair enzymes, both of which disrupt the local chromatin structure.

The other two H2A variants, H2A.Bbd and macroH2A, have been more recently discovered but their effects on chromatin structure are more well known. H2A.Bbd (Barr Body Deficient) is named because it is excluded from silenced X-chromosome[108]. It is also found at active genes and with acetylated H4. Concerning the primary protein sequence, H2A.Bbd has a CTD that is much shorter than the canonical H2A and thus is missing key residues within the docking domain. In mice, H2A.Bbd is also known as H2A.Lap1 (lack of acidic patch)[109]. Due to these two structural differences, the nucleosomes containing H2A.Bbd only protect 118-130 bps of DNA, far fewer than the canonical nucleosome; and *in vitro* chromatin folding is impaired with nucleosomal arrays reconstituted with H2A.Bbd[110-112]. In contrast to H2A.Bbd, macroH2A has a very long CTD consisting of approximately 250 residues within the CTD (compared with the entire canonical H2A which contains 130 residue in total)[113]. macroH2A is enriched on the silent X-chromosome and is thought to repress transcription as it overlaps frequently with developmental genes silenced by Polycomb[114,115]. The extremely long CTD of macroH2A1.1 (an isoform of macroH2A), interacts with NAD metabolites and enzymes such as PARP1 and promotes *in vitro* array condensation[116,117].

1.2.4 The role of linker length

The next piece in the hierarchy of chromatin structure beyond the nucleosome is the linker DNA, or the DNA that connects successive nucleosomes together. The term nucleosome repeat length (NRL) refers to the amount of DNA that wraps around the nucleosome (147 bps) plus the amount of DNA within the linker region. NRLs are extremely variable, not only between organisms but also between cell types and transcription states. For instance, the shortest known NRL is in the fungi, *S. pombe* and *Aspergillus nidulans*, both of which have an NRL of approximately 154 bps (or 7 bp linker length)[118,119]. On the other end of the spectrum, echinoderm sperm have NRLs that are approximately 245 bps[120]. Within the rat brain, the NRL length differs significantly between neurons (165 bps) and glial cells (201 bps)[121]. Even within the same cell type, $CD4^+$ T cells, promoters and enhancers were associated with NRLs of 178-187 bps while heterochromatin was associated with NRLs of 205 bps[122].

The effect of different linker lengths on the ability of a chromatin fiber to fold was measured *in vitro* through several different studies[59,123,124]. These studies used electron microscopy measurements to evaluate fiber diameters of long nucleosomal arrays (>50 nucleosomes/template) of varying NRLs, both with and without linker histones. Interestingly, it was found that arrays with short NRLs (167 bps), could only form fibers that were approximately 21 nm wide with a packing ratio of ~6 nucleosomes per 11 nm. Arrays with intermediate NRLs (197 bps), could form fibers approximately 34 nm wide with packing ratios of ~11 nucleosomes per 11 nm. Finally, arrays with long NRLs (217 bps) had fiber diameters approximately 42 nm wide with packing ratios of ~15-17 nucleosomes per 11 nm. Fiber diameter did not linearly increase between short NRLs to long NRLs but rather jumped discretely from 20 to 30 to 40 nm, indicating that the NRL most likely affects the shape of the fiber that can form (one start vs two

start helix, see Section 1.3.1 for more information). Longer NRLs may allow the fiber to position the nucleosomes in favorable conformations to allow better stacking, which increases the width of the fiber but reduces the linear length.

As the DNA molecule traverses between nucleosomes in the linker region, it twists upon itself in the usual B-form DNA. The consequence of this twisting is that the linker length may affect the rotational angle between adjacent nucleosomes. For instance, the crystal structure of the tetranucleosome with a short NRL was crystalized with adjacent nucleosomes orthogonal to each other[125]. The effect of short differences between NRLs was probed *in vitro* using nucleosomal arrays and EM measurements[126]. It was found that nucleosomal arrays formed the most compacted structures only with the correct NRL and that multiples of 10 bps from this length again allowed compact structure formation. For instance, linker lengths of 22 bps and 32 bps formed completely folded arrays, while linker lengths of 24 bps and 27 bps prevented this maximally folded formation presumably because the rotational angle between adjacent nucleosomes disfavored compact stacking.

1.2.5 The role of linker histones

Another consequence of different lengths of linker DNA is the ability of linker histones to bind to the stretch of DNA between nucleosomes. Short linker lengths prevent binding of linker histones while long linker lengths facilitate binding. Indeed, the stoichiometry of linker histones per nucleosomes is vastly different in species and cell types with varying NRLs[127]. Species or cell types with short NRLs, such as yeast or neuronal cells, have very few (<0.5) linker histones/nucleosome. However, as NRL increased linearly, so too did the number of linker histones/nucleosome, up to 1.3 in chicken erythrocytes[128].

Linker histones are comprised of the histone H1/H5 family of genes. Structurally, they have a well conserved globular domain, a short NTD, and a long positively charged CTD[129]. The globular domain consists of three helices that can bind DNA close to the dyad while the CTD is thought to help to condense the linker DNA by cooperative folding[130,131]. Binding of H1 to the chromatin fiber helps to protect an additional 20 bps of DNA from *Micrococcal* nuclease digestion[132]. Thus, it is believed that the role of linker histones is to bind and condense chromatin, leading to transcriptional silencing. Therefore, it was surprising when knock out of linker histones is yeast and *Tetrahymena* revealed very little phenotypic affect[133,134]. Also, depletion of embryonic H1 in a *Xenopus* system did not interfere with embryogenesis[135].

In vitro nucleosomal reconstitution studies have demonstrated that linker histones do not alter the folding pathway of arrays but may help to stabilize more condensed forms of the chromatin[136]. Along the same lines, optical tweezer experiments that measured the force needed to pull apart long nucleosomal arrays found that linker histones do not affect the stiffness or the length of the fiber[137]. This is in contrast with EM studies that showed that the most condensed fibers could only be formed in the presence of linker histones, thus implying that fiber folding is different in their presence[124]. Linker histones were deemed to be important for nucleosome arrays of long linker lengths, where linker lengths of 43-64 bps required linker histones to fold completely[126]. Additionally, it was found that linker histones, along with divalent cations such as magnesium, can bend the linker DNA in order to form different compact structures[138].

1.2.6 The role of nucleosome binding proteins and molecular mimcry

Because of the strong association between the acidic patch of H2A/H2B and the H4 N-terminal tail, it is not surprising that several different proteins have mimicked this interaction in nature

with consequences for chromatin architecture. Crystal structures and biochemical analyses have demonstrated that the Kaposi's sarcoma herpevirus LANA protein[139], Interleukin-33 (IL-33)[140], the Ran GTPase RCC1[141], Sir3[142], and high-mobility group nucleosomal protein 2 (HMGN2)[143] all bind to the acidic patch in the nucleosome. This binding either disrupts (HMGN2) or promotes (Sir3) chromatin condensation, affecting the local genome architecture.

1.2.7 The role of structural proteins

Apart from the nucleosomal binding proteins to the acidic patch, many different proteins are known to affect chromosome condensation *in vivo*. The heterochromatin protein 1 (HP1) recognizes H3K9 methylation and binds to chromatin[144,145]. This mark is associated with gene silencing in eukaryotes ranging from fission yeast[146] to flies[147] to humans[148] and is found at heterochromatin, or chromatin that is highly condensed[149]. This may be due to the ability of HP1 to promote associates between each other on adjacent nucleosomes[150]. In addition, HP1 may help to recruit other structural proteins, such as cohesin, to condense the surrounding chromatin[151].

Another group of proteins that strongly affect chromatin structure are the Polycomb group proteins (PcGs). PcG proteins silence developmental proteins such as the homeotic (Hox) genes in flies[152]. There are three major complexes comprised of PcG proteins: PhoRC, PRC1, and PRC2. The PRC2 complex contains proteins such as EZH, ESC, SUZ12, and NURF-55 and will trimethylate H3K27 (H3K27me3) on nucleosomes of developmental genes that need to be silenced. PRC1 will recognize the H3K27me3 mark and contains the RING1B E3 ubiquitin ligase that will ubiquitinate H2A K119[153]. These marks will spread throughout the gene and the

close association of the PRCs, especially PRC1, will cause chromatin condensation and heterochromatin formation[154].

The structural maintenance of chromosome (SMC) proteins comprise two larg complexes that regulate genome architecture throughout the different phases of the cell cycle – condensin and cohesin. These complexes are formed from a heterodimer of SMC proteins whose tails associate together and whose heads can open and close, effectively forming a hinge[155]. Both condensin and cohesin are strictly regulated through post-translational modifications and play a role in the 3D structure of the genome in interphase and mitosis[156,157].

1.3 The structure of chromatin in interphase

1.3.1 Models of the 30 nm fiber

The most basic level of chromatin compaction, the nucleosome or "beads on a string" conformation, is known to exist within nuclei of eukaryotic cells. It is commonly believed that these nucleosomes stack and fold into higher order structures throughout the phases of the cell cycle, including interphase. This next level of compaction, or secondary structure, is commonly referred to as the 30 nm fiber. *In vitro* nucleosomal array reconstitution lends credence to this idea because the arrays can form progressively more compact structures with increasing salt and these fibers have been visualized by techniques such as electron microscopy and atomic force microscopy with fiber diameters close to 30 nm[49,123,124].

There are several different theoretical models for the conformation about the 30 nm fiber that differ in the path of the DNA and stacking of the nucleosomes. The two most common models are the one-start solenoid model[158,159] and the two-start zigzag model[160,161]. In the one-start solenoid model, consecutive nucleosomes stack adjacently to each other, forming a larger

superhelix. Due to the physical dimensions of the nucleosome particle, stacking in the one-start solenoid model results in approximately 6-8 nucleosomes per 11 nm. The ensuing structure should behave similar to a Hookian spring whereby the force needed to stretch the fiber is proportional to the length of the fiber and a spring constant. This was found to be the case when nucleosomes were assembled onto arrays with a NRL of 197 bps and magnetic tweezer experiments measured the force needed to stretch the array[137].

In the two-start helical model, consecutive nucleosomes are not physically adjacent to each other but rather zigzag across the diameter of the fiber. This arrangement results in nucleosomes stacking with other nucleosomes located two apart from each other (j±2). For instance, nucleosome 3 will stack between nucleosomes 1 and 5 and further down the fiber, nucleosome 16 will stack between nucleosomes 14 and 18. The two-start helical model predicts that two helicies will wrap together to form the fiber and, depending on linker lengths, compact between 6-12 nucleosomes per 11 nm. Several studies have now shown that chromatin can adopt a zigzag pattern. The crystal structure of the tetranucleome using a NRL of 177 bps revealed that the chromatin was stabilized in the zigzag pattern with straight linkers between nucleosomes[125]. Additionally, cross-linking studies found that fibers with NRLs of 167, 177, and 208 bps all were compatible with the two-start helix[162]. Finally, molecular tweezer experiments to probe the force it took to pull apart a chromatin fiber with a NRL of 167 bps found a stiffer fiber than with a NRL of 197 bps, indicating that a NRL of 167 bps favors a two-start organization[137].

Interestingly, a study that used Monte Carlo simulations coupled with EM-assisted nucleosome interaction capture (EMANIC) was able to quantify the types of higher order structures formed by nucleosomal arrays at physiological salt concentrations[138]. The authors found that most

arrays adopted a zigzag pattern, consistent with the two-start helix. However, when the divalent cation concentration was increased, arrays then favored a one-start solenoid organization, but some still retained the zigzag formation. It was proposed that the divalent cations help to bend the linker DNA to promote the solenoid structure. Thus, it was concluded that both structures may exist simultaneously and no one structure is ubiquitous.

Recently, a newer model was proposed which takes into account many of the observations from literature related to the structure of the 30 nm fiber[163]. This model, called the ribbon model, was designed to account for the differences in fiber diameters observed with different NRLs by Rhodes and colleagues[123,124]. Theoretically, the fiber diameter of the solenoid model as well as the two-start helical model should both be invariant to linker length, but a variant of the two-start helical model, the two-start crossed-linker fiber, should depend on linker length. However, what has been observed in EM measurements is that the fiber diameter adopts discrete diameters, with short, intermediate, and long NRLs adopting fibers that are ~20 nm, ~30 nm, and ~40 nm wide, respectively[123,124]. The ribbon model accounts for these discrete changes in fiber diameter by proposing that the NRL dictates the number of ribbons that twist upon each other to form the fiber. The number of ribbons then dictates the diameter. Short NRLs would only allow twisting of three ribbons upon each other with a diameter of ~23 nm. Twisting of five ribbons allows for a diameter of 33 nm and twisting of seven ribbons allows for a diameter of 44 nm[164]. The number of ribbons also dictates the wedge angle of the stacking nucleosomes and the path of the linker DNA.

1.3.2 Does the 30 nm fiber exist *in vivo*?

The existence of chromatin forming fibers that are 30 nm in diameter *in vivo* initially came from EM studies that looked at the 3D architecture of the sperm of a starfish, *P. miniata*[165]. In the EM images, individual nucleosomes could be seen that adopted a zigzag pattern with fibers approximately 30 nm in width. However, subsequent studies using cryo-EM with *Xenopus* mitotic chromosomes found that the 10 nm chromatin was not further condensed into 30 nm fibers, even during mitosis[166]. Electron spectroscopic imaging, a technique that allows visualization of the DNA phosphate backbone, demonstrated that chromatin was organized into a meshwork within the nucleus with no higher order structures beyond the 10 nm fiber seen[167].

In yeast, there have been conflicting results on the compaction of chromatin *in vivo*. Bystricky *et al.* used fluorescence *in situ* hybridization (FISH) to show that interphase chromatin has a persistence length ranging from 170-220 nm and a mass density of 110-150 nm[168]. These values correspond to a stiff 30 nm fiber with 7-10 nucleosomes per 11 nm. In contrast, Dekker used chromosome conformation capture (3C) to estimate the contact frequencies of chromosomes and concluded that yeast interphase chromatin exists in an open conformation with only 1.2-2.4 nucleosomes per 11 nm, values that are inconsistent with the presence of a 30 nm fiber[169].

Because of the lack of evidence for the 30 nm fiber *in vivo*, its existence in interphase chromatin has been called into question[170-172]. Instead, hypotheses have proposed that chromatin exists in the 10 nm form and is condensed into local regions of higher or lower density based on the transcription state and epigenetic marks[172]. These hypotheses would be consistent with other chromatin capture techniques, such as Hi-C[173,174].

1.4 The eukaryotic need for chromatin compaction

1.4.1 Genome size expands more rapidly than nuclear size

It is generally well accepted that eukaryotes with larger genomes also have correspondingly larger cell volumes and nuclear volumes[175]. However, the rates at which genome size and nuclear size have expanded are not the same, and, in fact, genome size has expanded much more rapidly. For instance, with an average volume of a yeast nucleus[176] of 3.0 μm^3 and a genome size of ~12 Mbp, yeast have approximately 4 Mbp/μm^3. Humans have an average nuclear volume of 200 μm^3 and a genome size of ~3100 Mbp, requiring nuclei to pack 15.5 Mbp/μm^3. Therefore, in terms of strict DNA density per unit volume, human nuclei are approximately four times denser than yeast nuclei, and this does not even take into account the amount of RNA and proteins within the nucleus.

The disproportionate expansion of genome size relative to nuclear volume occurs across the eukaryotic domain in species with genome sizes ranging from a few Mbp to several thousand Mbp[177]. This trend is also apparent when considering different classes and taxa as well, such as vertebrates[178,179] (Figure 2-2). In Figure 2-2, simultaneous measurements were made of nuclear volume and genome size in erythrocytes of varies species. The DNA density is plotted against the genome size. If genome size expanded proportionally to nuclear size, a flat trendline would be expected. Instead, the trendline slopes positively, indicating that genome sizes expand more rapidly than nuclear volume in this dataset.

The consequence of faster expansion of genome size relative to nuclear volume is that organisms with larger genomes need to compact their genomes to a greater degree. In fact, Vinogradov measured the native compaction levels by dye incorporation in cells with intact membranes

(compacted chromatin) vs damaged membranes (de-compacted chromatin) in several vertebrates ranging in genome size[180]. The author found that vertebrates with larger genomes contained chromatin that was more compact than the chromatin of species with smaller genomes.

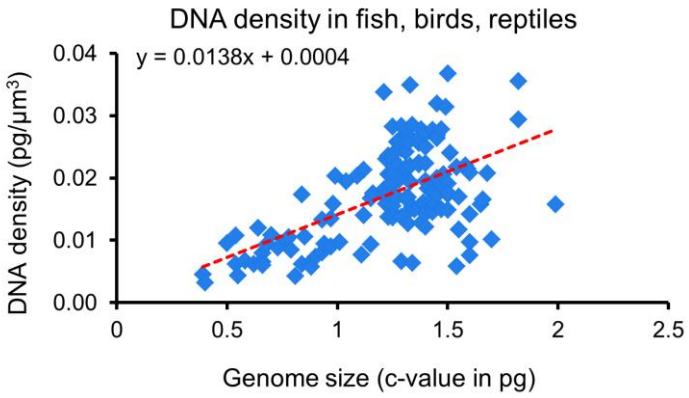

Figure 2-2 – DNA density increases with genome size. The genome size of several fish, birds, and reptiles was plotted against the density of DNA. The trendline is shown in a red dashed line. Note the equation of the trendline in the top left. Data was obtained from www.genomesize.com[179].

Although higher eukaryotes possess several different methods for genome compaction (as outlined in Section 1.2), their intrinsic need for greater compaction levels opens up several interesting questions. How do higher eukaryotes natively deal with more DNA per unit volume? What mechanisms promote basal levels of compaction? Which came first, genome size expansion or nuclear size expansion? Some eukaryotes have small genome sizes but extremely small cellular/nuclear sizes. How do they cope with the increased DNA density?

Histones are thought to be very conserved proteins but there have been sequence divergence among portions of the histones. Interestingly, it has been unknown whether histones, themselves, have evolved to aid eukaryotes in genome compaction which is the subject of this dissertation.

CHAPTER 2

The evolution of H2A to mediate chromatin compaction in eukaryotes

2.1 Summary

During eukaryotic evolution, genome size has increased disproportionately to nuclear volume, necessitating greater degrees of chromatin compaction in higher eukaryotes which have evolved several mechanisms for genome compaction. However, it has been unknown whether histones themselves have evolved to regulate chromatin compaction. Analysis of histone sequences from 160 eukaryotes revealed that the H2A NTD has systematically acquired arginines as genomes expanded. Insertion of arginines into their evolutionary conserved position in H2A of a small-genome organism increased linear compaction by as much as 40%, while their absence markedly diminished compaction in cells with large genomes. This effect was recapitulated *in vitro* with nucleosomal arrays using unmodified histones, indicating that H2A N-terminus directly modulates the chromatin fiber likely through intra- and inter-nucleosomal arginine-DNA contacts to enable tighter nucleosomal packing. Our findings reveal a novel evolutionary mechanism for regulation of chromatin compaction and may explain the frequent mutations of the H2A N-terminus in cancer.

2.2 Introduction

2.2.1 The *in vivo* role of the H2A NTD

As outlined in Section 1.2.2, the N-terminal tails of histones play important roles in the ability of chromatin to mediate local interactions (array folding) and long range interactions (oligomerization). Both the H4 NTD and the H2B CTD have specifically been shown to be involved in these processes but the roles of the NTDs of the other histones is less well understood. Functionally, the H3 NTD, the longest of the N-terminal tails, can be extensively epigenetically modified and it is thought that this is the main purpose of the H3 tail[24].

In a series of experiments in the 1980s, Grunstein and colleagues showed that the H2A NTD and the H2B NTD were both vital but redundant[181]. Removal of either the H2A NTD or the H2B NTD from yeast did not affect cell viability but removal of both tails simultaneously was lethal. In addition, the very N-terminal portions of the tails could be switched without large effects on viability. Besides the H2B NTD, the H2A NTD is redundant with the H4 NTD as well. In humans, the first nine residues of H2A and H4 are identical with the exception of a glutamine insertion at position 6 of H2A. Simultaneous deletions of both the H2A and H4 NTDs was lethal for yeast cells and deletion of both the H2A NTD and H3 NTD resulted in a severe growth phenotype and sensitivity to DNA damage [182].

Within the yeast cell, the H2A NTD has been implicated in global transcription repression. Two domains in H2A, residues 4-20 and residues 30-45, were identified in repression assays[183] and later through microarray gene expression[184] to be important for preventing transcription of a number of gene products. The region within H2A 4-20 was narrowed down to residues 16-20, which is known as the H2A Repression Domain (HAR Domain). Later, it was found that the

HAR domain can regulate the amount of H2B K123 ubiquitylation which, in turn, helps to regulate H3K4me3, an epigenetic mark associated with active promoters[185]. Besides the crosstalk between H2B ubiquitylation and H3 methylation, the HAR domain serves as a recruitment location for the klesin subunit of condensin[156].

The H2A NTD is known to have a few post-translational modifications. Two of the lysines, K7[61,186] and K4[187] can be acetylated, though H2AK4ac was found to be of very low abundance from mass spectrometry analysis. Although H2AK7ac can be found at gene promoters, mutation of both K4 and K7 to glycines had relatively little effect on transcription[184]. In higher eukaryotes, H2A contains an arginine at position 3, and due to the similarity of the H2A and H4 tails, R3 can be methylated in both tails by protein arginine methyltransferases (PRMT) 1, 5, 6, and 7[188]. Recently, H2A R11 was also found to be methylated by PRMT 1 and 6 but not 5 and 7[189]. The biological function of R3 and R11 methylation is still unknown.

The role of the H2A NTD in chromatin compaction is not well understood. *In vitro* nucleosome reconstitution experiments with single histone NTD deletions found that only the H4 NTD affected chromatin folding, as arrays containing histones without either the H2A NTD, H2B NTD, or H3 NTD sedimented similarly to WT[56]. Conversely, crosslinking experiments with dinucleosomes demonstrated that both the H2A and H2B NTDs, but not the H3 or H4 NTDs, were able to make histone-DNA contacts independent of salt concentration.[190]. The H2A NTD does play a role in longer range interactions, as it was found that each of the NTDs function both independently and additively in the oligomerization process of nucleosomal arrays[191]. The experiments just outlined were all done *in vitro*, but the *in vivo* role of the H2A NTD in compaction of chromatin is still largely unknown and is the subject of Chapter 2.

2.3 Results

2.3.1 H2A NTD protein sequence correlates to genome size

2.3.1.1 Arginines positively correlate to genome size while serines and threonines negatively correlate to genome size

As demonstrated in Section 1.4, eukaryotes with larger genomes need to compact their chromatin to a greater extent than organisms with small genomes. Histone proteins play a vital role in chromatin architecture, but it is unknown whether they have co-evolved with genome size to aid in this increased compaction seen in higher eukaryotes. To determine whether this is the case, we gathered the protein sequence of the four canonical histones – H2A, H2B, H3, and H4 – from 160 fully sequenced eukaryotic organisms. This group of eukaryotes contained species comprising of fungi, protozoa, plants, and animals and ranged in genome size from 8 Mbp (*B. bovis*) to 5600 Mbp (*B. gargarizans*). In order to discover trends across the entire dataset, we first grouped each organism into one of three categories based on genome size – small (<100 Mbp), medium (100-1000 Mbp), or large (>1000 Mbp). Next, we split the four core histones into their known protein domains[192]. H2A and H2B were split into the N-terminal Domain (NTD), Histone-Fold region (HF), and the C-terminal Domain (CTD) while H3 and H4 were split into the NTD and HF. We then analyzed the protein sequences of each canonical histone in its entirety or individual domains within each of the genome size categories to uncover potential patterns in relation to genome size.

Residue composition analysis demonstrated that the strongest trends within any of the canonical histones occurred in the H2A NTD whereby the number of arginine (Arg or R) residues increased with expanding genome size. Figure 2-1A shows this strong correlation.

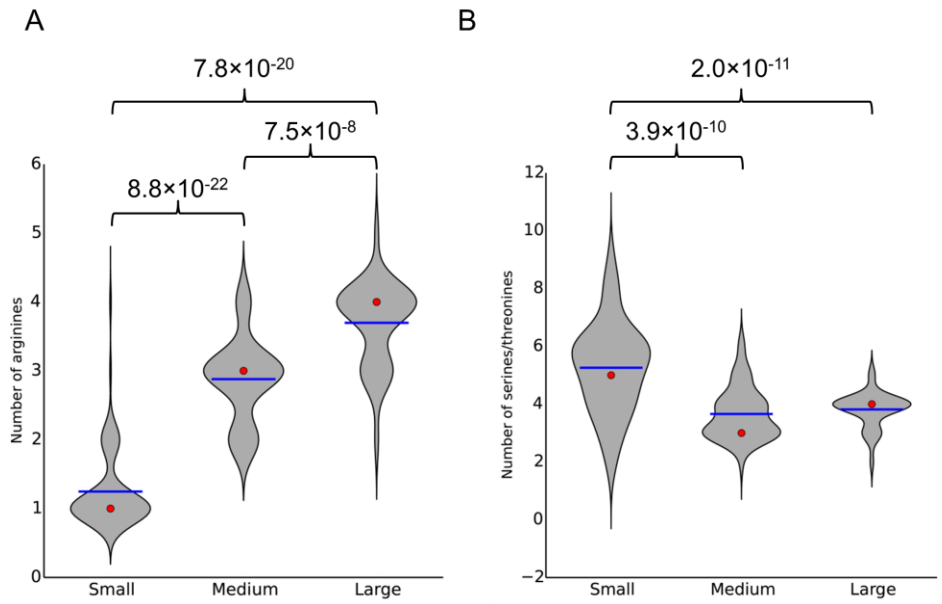

Figure 2-1 – Arginines in the H2A NTD positively correlate to genome size while serines/threonines negatively correlate. Violin plot showing the distribution of the number of arginines **(A)** or serines/threonines **(B)** in the H2A NTD within each genome size category. Small: < 100 Mbp, Medium: 100-1000 Mbp, Large: >1000 Mbp. The width of the violin plot gives the kernel density function of the distribution. The red dot shows the median value. The blue line shows the mean value. P-values comparing means of categories are shown above the plots.

In addition to the pattern of arginines within the H2A NTD, the number of serines/threonines (Ser or S and Thr or T) shows the inverse relationship through a negative correlation between the frequency of serines/threonines and genome size (Figure 2-1B). Other residues, such as lysines (Lys or K), did not correlate with genome size.

2.3.1.2 Arginines and serines/threonines are precisely positioned within the H2A NTD

To determine whether the acquisition of arginines within the H2A NTD with increasing genome size occurs randomly or at specific regions, we first compared the H2A protein sequence between *S. cerevisiae* (yeast), which has a small genome size of 12 Mbp, and *H. sapiens*, which has a large genome size of 3108 Mbp (Figure 2-2A). Yeast, like most small genome sized organisms, contain only one arginine, located at position 17 (R17). In contrast, humans contain four arginines located at positions 3, 11, 17, 20. The arginine at position 17 is conserved

between yeast and humans, while R3 and R11 are absent in yeast and R20 is correspondingly a lysine (K20) in yeast (Figure 2-2A). Yeast contain two serines in the middle of the H2A NTD (S10 and S15) that are absent in the human sequence.

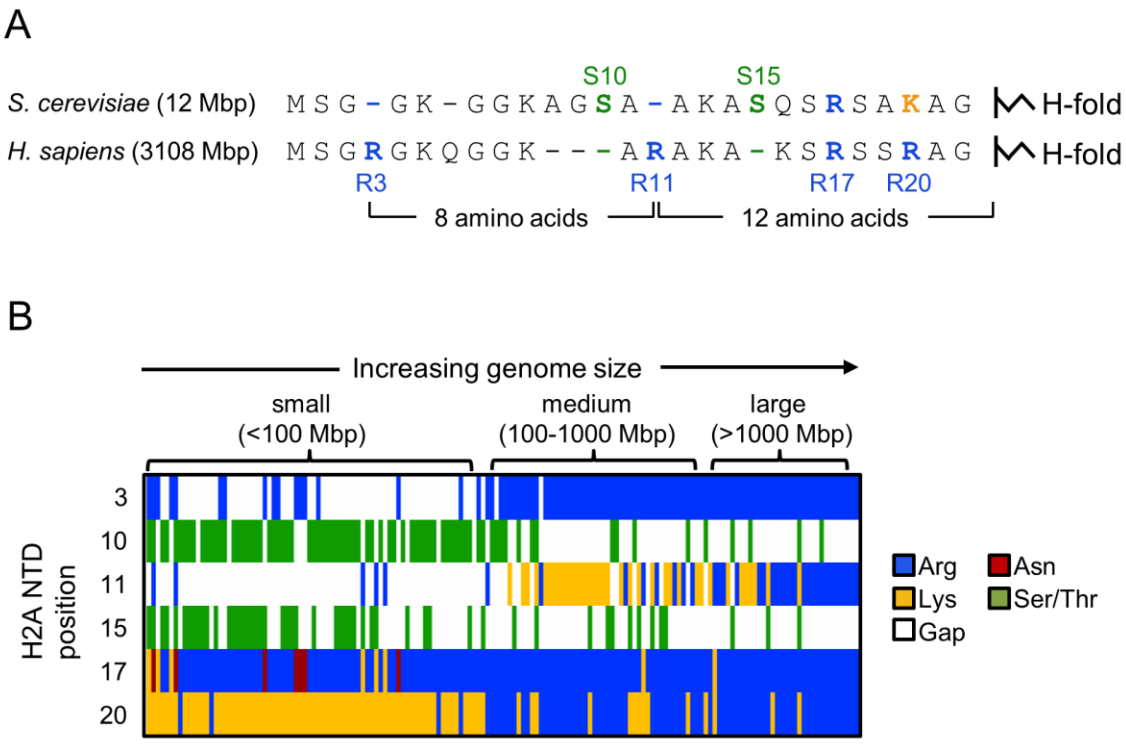

Figure 2-2 – Arginines are precisely positioned within the H2A NTD across eukaryotes. (A) The H2A NTD protein sequence of *S. cerevisiae* and *H. sapiens*. Note the spacing between R11 and the histone fold and between R3 and R11. **(B)** Heat map of H2A NTD residue composition at the indicated positions ordered by genome size.

The presence of a conserved arginine (R17) within both yeast and humans suggest that arginines may be precisely positioned within the H2A NTD of eukaryotes. To determine whether this is the case, we aligned the H2A NTD protein sequences of all 160 species in our dataset and created a heatmap of amino acid composition by position in order to visually compare residues across organisms (Figure 2-2B). Interestingly, this analysis revealed similar trends as was observed between yeast and humans. R17 is present in almost all organisms examined, suggesting a conserved function of this residue[185]. R3 is predominantly absent in organisms with

small genomes but present in medium- and large-sized genome species. At position 20, most small genome sized species contain a lysine, which is mutated to an arginine in organisms with medium-sized genomes and larger. Position 11 shows the most variability, being absent in most small genome species, a lysine in most medium genome species, and an arginine in large genome species. Analysis of positions 10 and 15 revealed that serines/threonines are predominantly present only in small genome species and not in medium or large genome species.

In order to further investigate the physical positioning of the evolutionary variable arginines and serines/threonines in the H2A NTD, we plotted their distance from both the N-terminus and the histone fold region (Figure 2-3). Although the N-terminus is traditionally used to number residues, we surprisingly found that the distance of all arginines, except R3, was very variable relative to the N-terminus. R11 varied from 9-13 residues from the N-terminus, R17 varied from 16-25 amino acids from the N-terminus, and R20 varied from 17-28 from the N-terminus. In contrast, when counted relative to the histone fold, R11 is mostly found 12 amino acids to the fold with a few instances of being positioned 11 amino acids away and R17 and R20 are always positioned 6 and 3 amino acids from the fold. All serines/threonines in the H2A NTD are also better conserved relative to the histone fold than the N-terminus.

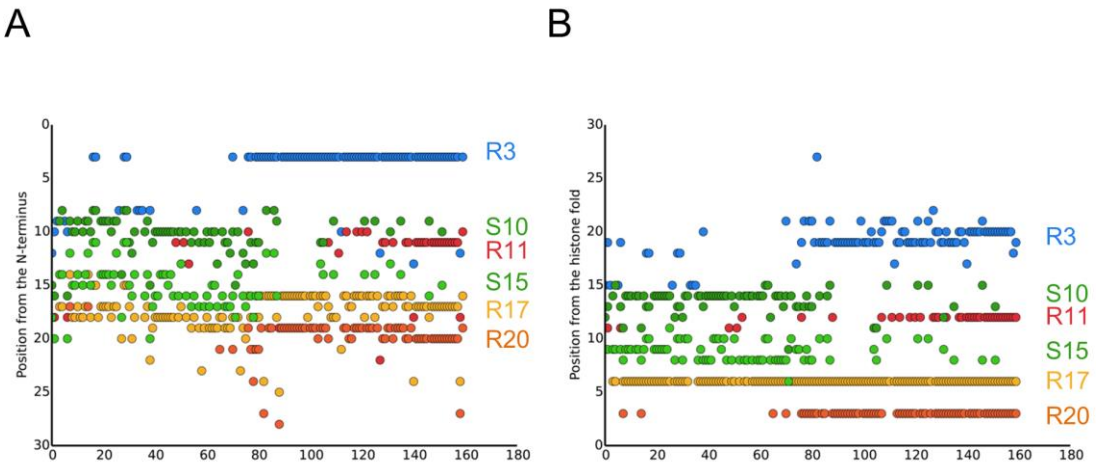

Figure 2-3 – H2A NTD arginines are better conserved relative to the histone fold than the N-terminus. The position of the evolutionarily variable residues is plotted relative to the H2A N-terminus **(A)** or histone fold **(B)**. Note the y-axis in both graphs.

2.3.1.3 H2A NTD arginines are surrounded by conserved motifs

Because surrounding residues on histones are often important for recognition by chromatin readers[193-195], we examined the motifs near the four conserved arginines in the H2A NTD (Figure 2-4). R3 and R17 both had very well conserved motifs of GRG and SRS, respectively. AKA is present at R20 in organisms with small genomes and SRA is present in organisms with large genomes. The motif at R11 also varied with genome size, with medium sized genome species containing VKG and large genome species containing ARA.

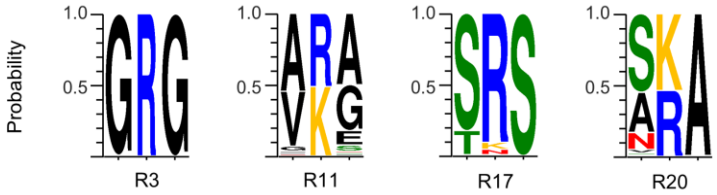

Figure 2-4 – H2A NTD arginines are surrounded by conserved motifs. The proportion of each surrounding residue of the four conserved arginines in the H2A NTD is plotted.

2.3.1.4 Summary

Together, the results presented here in Section 2.3.1, which comprises our computational analysis concerning the protein sequences in the H2A NTD, suggest several patterns emerge with respect to increasing genome size. First, the H2A NTD acquires arginines and loses serines/threonines as genome size expands (Figure 2-1). Second, these changes occur in sequential order, with small genomes containing S10, S15, and R17, medium genomes containing R3, K11, R17, and R20, and large genomes containing R3, R11, R17, and R20 (Figure 2-2). Third, arginines, in the H2A NTD are precisely positioned relative to the histone fold, suggesting that the spacing and physical distance from the fold matters (Figure 2-3). Lastly, H2A NTD arginines are surrounded by conserved motifs (Figure 2-4).

2.3.2 H2A NTD arginines within the crystal structure

2.3.2.1 H2A R3 and R11 are located on the surface of the nucleosome

Because the positioning of the H2A NTD arginines is so well conserved from the histone fold, we hypothesized that this position related to the physical location of the residues as the tail protrudes out from the core particle (Figure 2-3). To examine the three dimensional positions of the H2A NTD arginines, we used a crystal structure of the nucleosome in which all the evolutionarily changed arginines had been fully crystalized in at least one of the two H2A tails[35]. Inspection of this crystal structure revealed that two of the arginines, R3 and R11, are located on the surface of the nucleosome, while the other two arginines, R17 and R20, are more buried in the structure (Figure 2-5).

The orientation of the positive charge of R3 and R11 are pointing outward, suggesting that they could lead to possible *trans* interactions. R17 and R20 both have their positive charges pointing

inward towards the DNA backbone as the gyre wraps around the core particle, suggesting that they may aid in intranucleosomal positioning of the tail.

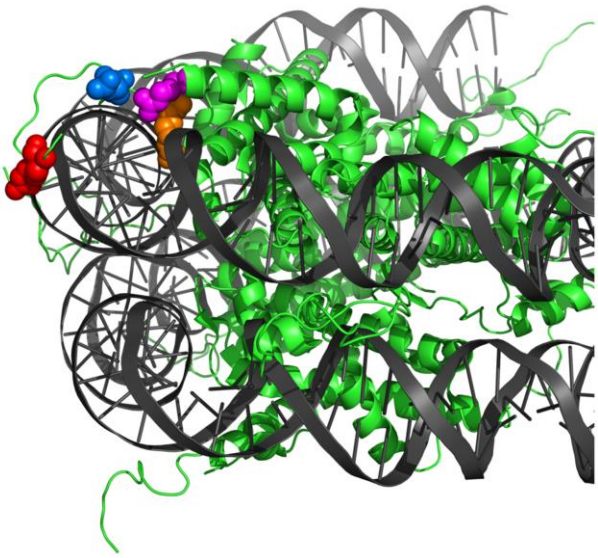

Figure 2-5 – R3 and R11 are located on the surface of the nucleosome. Crystal structure of the nucleosome particle with the four evolutionarily changed arginines of Chain C colored. R3 is red. R11 is blue. R17 is orange. R20 is magenta.

2.3.2.2 H2A R11 makes an internucleosomal connection

Further analysis of the crystal structure revealed that R11 makes a close internucleosomal contact with the DNA phosphate backbone that wraps around a different particle in the crystal lattice (Figure 2-6). R11 is 2.90 Å from the neighboring DNA backbone while 4.09 Å away from the self DNA backbone, making it closer to the DNA of the neighboring nucleosome. R3, on the other hand, is located 2.87 Å from the self DNA backbone.

2.3.2.3 Summary

Together, these analyses show that at least two of the H2A NTD arginines, R3 and R11, are positioned on the nucleosomal surface in such a way as to possibly aid in tighter nucleosomal stacking by shielding the negative charge of either self or neighboring DNA phosphate

backbones. This may lead to more compact chromatin. The other two evolutionary H2A arginines, R17 and R20, are more buried from the surface and thus may not directly contribute to nucleosomal stacking, but rather may help to position the H2A NTD in an optimal physical location.

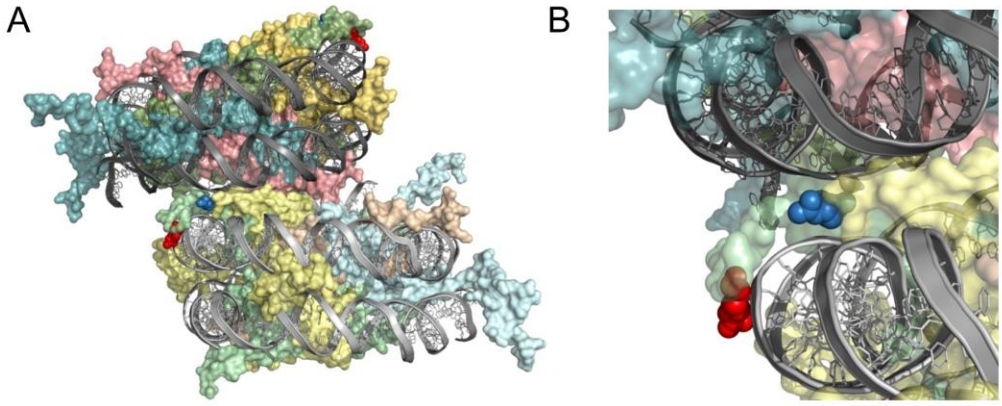

Figure 2-6 – R11 makes an internucleosomal interaction. Structure of a dinucleosome pair from (A) the side or (B) close up highlighting possible internucleosomal interactions between arginines and DNA backbone. Green is H2A, yellow is H2B, cyan is H3, and salmon is H4. The red and blue spheres are R3 and R11, respectively, both of which are in chain C[35].

2.3.3 H2A NTD arginines compact chromatin in yeast

2.3.3.1 H2A arginines increase chromatin compaction along a chromosomal arm

Both DNA density per unit volume (Section 1.4) and the number of arginines in the H2A NTD (Figure 2-1) correlate positively with genome size. Because the physical location of the H2A NTD arginines in the nucleosome core particle allows for possible interactions (Figure 2-6), we hypothesized the H2A arginines may facilitate the greater chromatin compaction seen in organisms with larger genomes. To test this hypothesis, we took advantage of a strain of *S. cerevisiae* that had both of its genomic copies of H2A deleted and instead expressed H2A on a plasmid (TSY107)[181,196]. We can then use this strain to introduce mutations into the H2A gene

to mimic single or multiple insertions of arginines, deletions of serines, or combinations of the two (see Table 5-1 for all strains tested). To test the effect of single arginine insertions, we created strains containing R3, R11, R17K (this is the only arginine natively found in yeast), and K20R. To test serine deletions, we created strains with ΔS15 and ΔG9ΔS10 (ΔGS10). Because the spacing between the arginines and the histone fold is highly conserved (Figure 2-3), R11ΔS15 was constructed to move R11 into a position where it is 12 residues from the fold, much like what is observed in large genome species. Finally, we also created double arginine mutants, R3R11 and R3(ΔGS10)R11, with the latter mutant created in order to preserve the spacing between R3 and R11 (Figure 2-2A – note the spacing in the human sequence between R3 and R11). To control for positive charge, we created mutants with lysines (K3, K11, K11ΔS15) and to control for any arginine insertion, we created an R6 mutant.

To measure chromatin compaction, we differentially labeled two parts of the yeast Chr XVI that are spaced 275 kbp apart using fluorescent *in situ* hybridization (FISH) (Figure 2-7A)[168,197]. These probes were visualized by confocal microscopy and the physical distance between them was measured in a single field of view where both probes were simultaneously present within each nucleus, as was previously done[168]. When compared to an isogenic wild type (WT), addition of a single arginine R3 or R11 compacted chromatin, with a mean interprobe distance that was decreased by 18% and 23%, respectively (Figure 2-7B-C). However, the mutation of K20 to an arginine (K20R) or R17 to lysine (R17K) did not statistically change these distances, indicating that not every H2A arginine affects compaction. The lysine controls, K3 and K11, also had little effect on the measured interprobe distances, indicating a specific function through arginines.

The combination mutant, R3R11, further enhanced the compaction observed through single arginine mutants by decreasing interprobe distances by 26% (compared to 18% and 23% from each single mutant alone), indicating that the two arginines may act synergistically. The greatest compaction occurred in our mutants that preserved the correct spacing between the arginines with each other (R3(ΔGS10)R11) or the histone fold (R11ΔS15). R3(ΔGS10)R11 decreased the mean interprobe distances by 32% and R11ΔS15 decreased the mean interprobe distances by 40%. Removal of the serine residues, ΔS15 and ΔGS10, had little effect on compaction.

2.3.3.2 Chromatin compaction mediated through H2A arginines is uniform and not strain specific

In addition to the probe set tested in Section 2.3.3.1 (Probe Set A), we analyzed the distribution of interprobe distances from three additional probe sets along the same chromosomal arm for a few of our H2A mutants (Probe Sets B-D, see Figure 2-7A). Our WT strain had a baseline compaction level similar to what has been previously reported[168]. In each of these probe sets, both R11 and R11ΔS15 caused significant chromatin compaction when compared to WT while ΔS15 did not (Figure 2-8A-C). R11 decreased interprobe distance between 17-19% while R11ΔS15 further decreased interprobe distances 19-29%. Plotting the physical distance vs genomic distance, we notice that the compaction mediated by R11 and R11ΔS15 is uniform across the long arm of Chr XVI (Figure 2-8D).

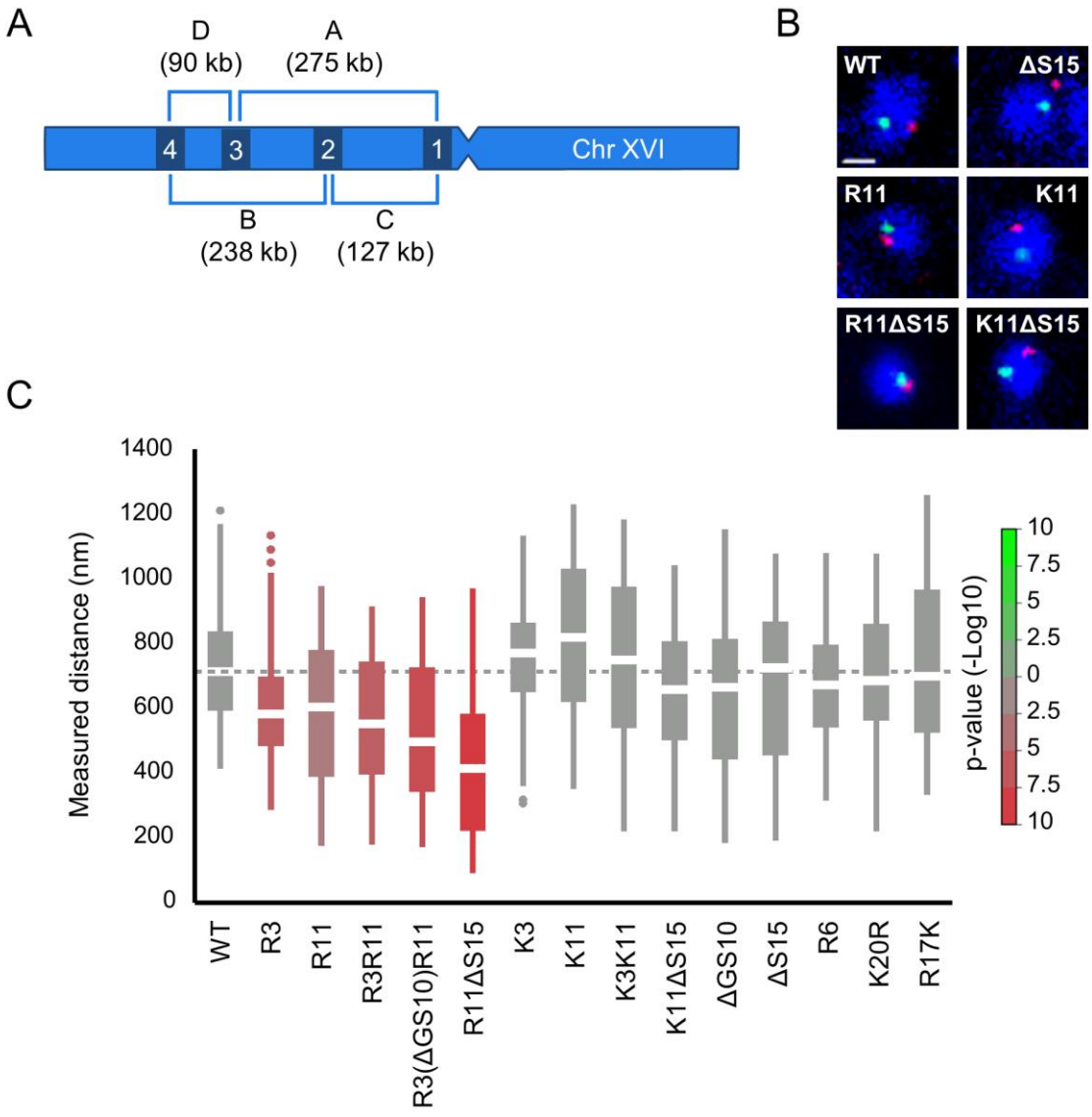

Figure 2-7 - Ectopic expression of H2A NTD arginines causes compaction in yeast. (A) Schematic position of probes on chromosome XVI that were used for FISH. The letters correspond to the probe sets. **(B)** FISH images and **(C)** boxplot of the distributions of interprobe distances for probe set A in the indicated strains. Dashed lines mark the median value for the WT strain. The boxplot whiskers contain 90% of the data. All scale bars are 1 μm. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

In order to determine whether this compaction is specific to the TSY107 background strains or not, we performed FISH using Probe Set A in the context of the FY406 background. FY406 has both of its H2A and H2B genes deleted and instead expresses H2A and H2B on a single plasmid.

Much like in the TSY107 background, both R11 and R11ΔS15 caused a significant decrease in the interprobe distances (16% and 27%, respectively) (Figure 2-9).

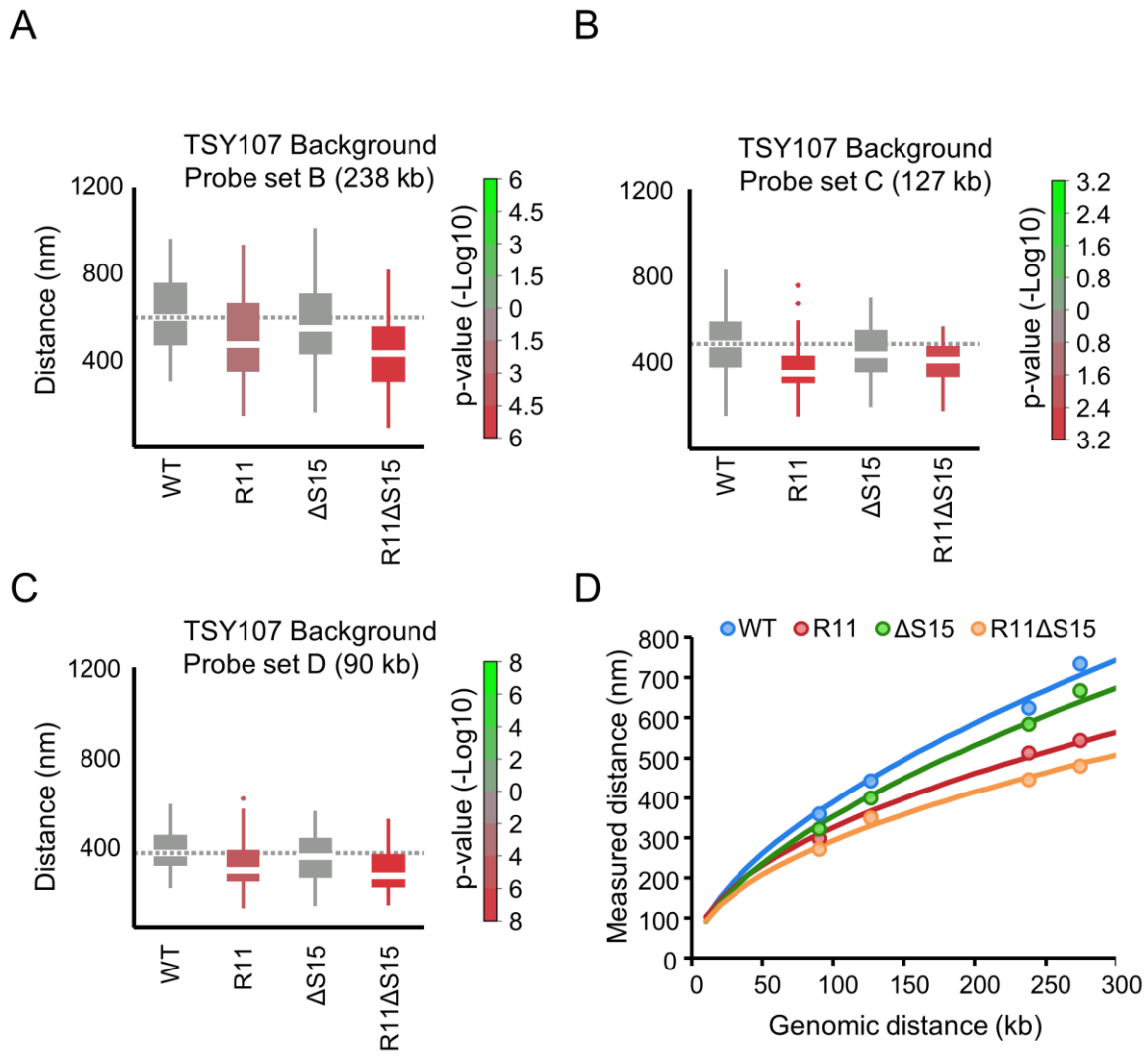

Figure 2-8 – Chromatin compaction mediated through H2A arginines is uniform. (A-C) Boxplots of the distributions of interprobe distances in the indicated probe sets. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively. **(D)** The mean interprobe distances for the indicated yeast strains for probe sets A, B, C, and D are plotted as a function of genomic distance. Solid lines are best fit equations.

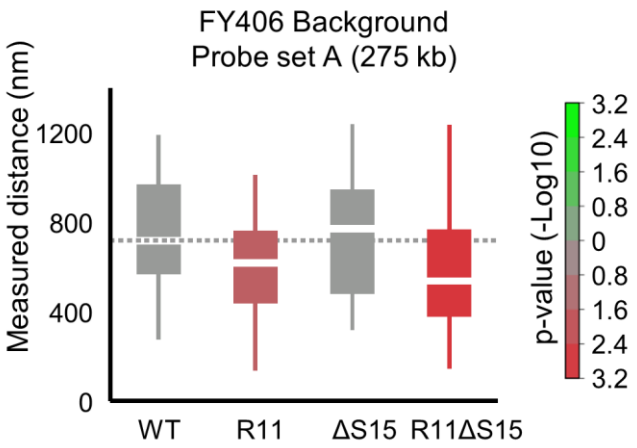

Figure 2-9 – Chromatin compaction mediated through H2A arginines is not strain specific. Boxplot of the distributions of interprobe distances of Probe Set A in H2A mutants based upon FY406. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

2.3.3.3 Chromatin compaction mediated through H2A arginines slightly decreases MNase accessibility but does not alter nucleosomal occupancy

Linker lengths are known to affect the amount of compaction *in vitro* by modulating how the 30-nm fiber forms through entry/exit angles and amount of linker histones that attach to the fiber[124,198]. Indeed, the average linker length is different between yeast and humans (20 bp vs 50 bp)[199]. Therefore, to investigate whether nucleosomal spacing is changed in H2A arginine mutants that exhibit greater compaction, we performed *Micrococcal* nuclease (MNase) digestions of chromatin. Consistent with chromatin that is more compact, we found that accessibility of the enzyme was slightly delayed in the H2A arginine mutants R11 and R11ΔS15, but not ΔS15, resulting in chromatin that was underdigested when compared to WT (Figure 2-10A). However, the spacing between the mono-, di-, and tri-nucleosomes, etc. was similar in all mutants examined, indicating that the genome-wide nucleosomal spacing is unaffected.

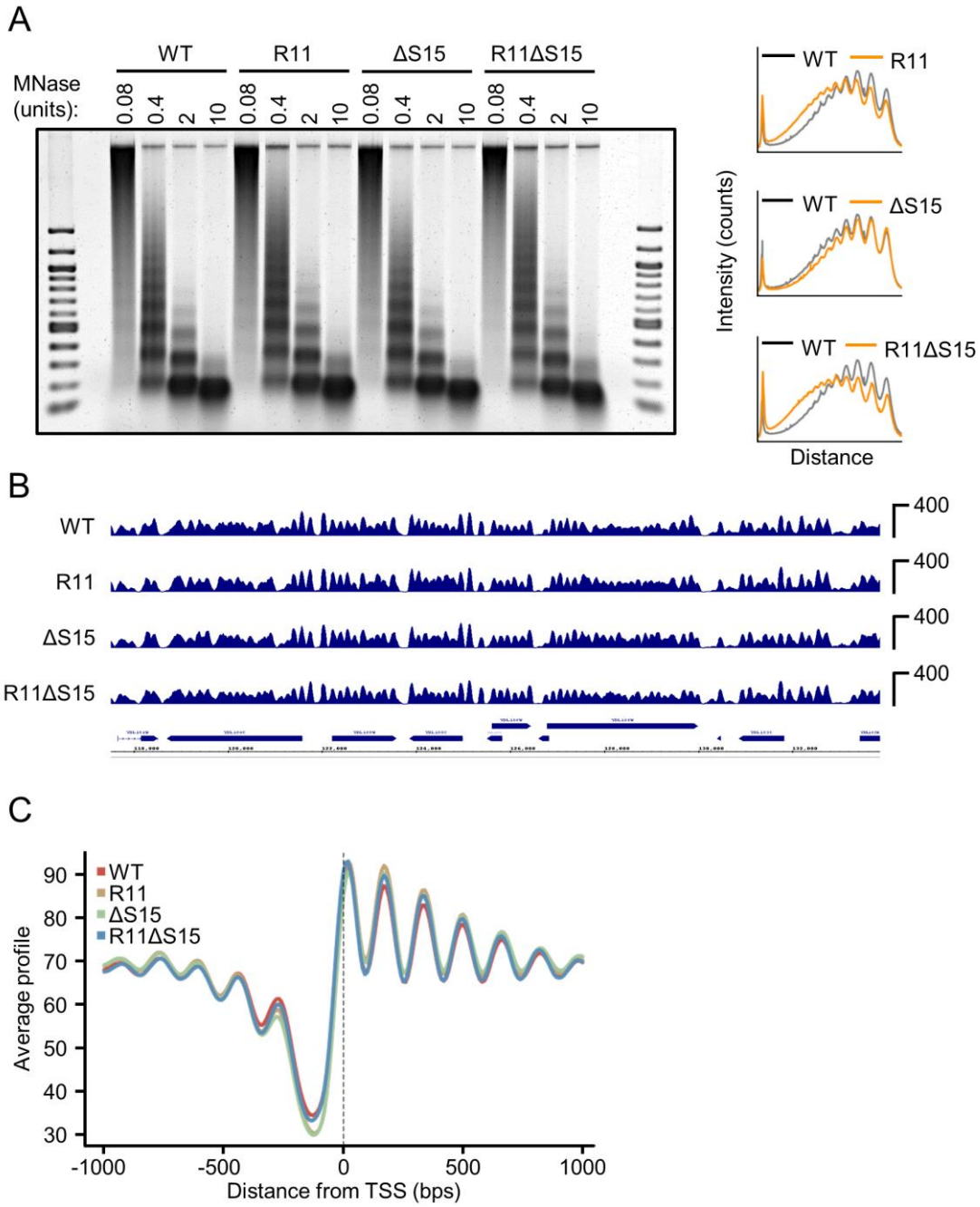

Figure 2-10 – Ectopic expression of H2A NTD arginines causes a slight decrease in MNase accessibility but no difference in nucleosomal occupancy. (A) Agarose gel electrophoresis of MNase-digested chromatin in the indicated strains. The amount of enzyme used is indicated above the gel. The densiometric profiles for 0.4 units of MNase in the given strains are shown to the right. **(B)** Genome browser view of nucleosomal occupancy along a portion of Chr I in the indicated strains from deep sequencing after MNase digestion. **(C)** Average profile of nucleosome occupancy around open reading frames (ORFs) in the indicated strains.

In order to obtain a higher resolution of nucleosomal occupancies throughout the entire genome, we performed MNase digestions followed by deep sequencing (MNase-Seq). Analysis of the results confirmed that there are essentially no differences in the positioning of nucleosomes in our H2A mutants when compared to WT (Figure 2-10B). Meta-analysis of different features of the yeast genome, such as open reading frames (ORFs), tRNAs, autonomously replicating sequences (ARS), long terminal repeats (LTRs), snRNAs, telomeres, centromeres, and Y' elements revealed that there are no differences in nucleosomal occupancies at any of those features between WT and H2A mutants (Figure 2-10C).

The cell cycle is also known to affect the amount of chromatin compaction[197]. However, cell cycle profile analysis revealed little difference between WT and any of our mutant H2A strains. This indicates that difference in the cell cycle do not account for the compaction difference in our H2A NTD arginine mutants.

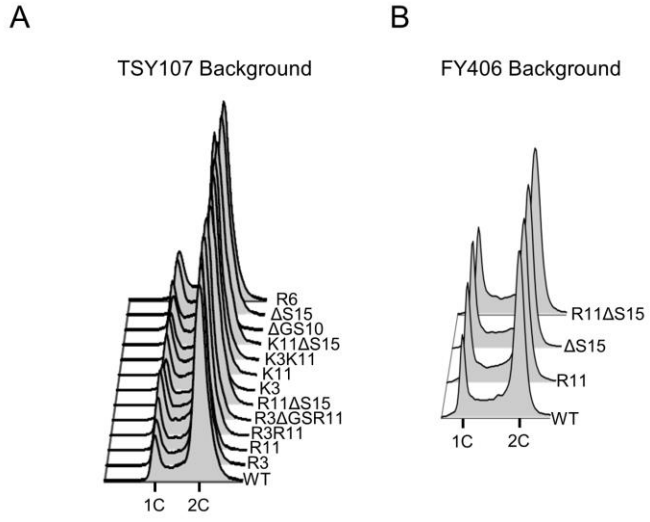

Figure 2-11 – Ectopic expression of H2A NTD arginines does not affect the cell cyle. Cell cycle profiles of the indicated mutant from the TSY107 strain background **(A)** or the FY406 strain background **(B)**.

2.3.3.4 Summary

Together, these data demonstrate that in *S. cerevisiae,* arginines that are added to the H2A NTD cause compaction of chromatin. The positioning of these arginines is important, as strains with R11 correctly positioned 12 residues from the fold (R11ΔS15) cause greater compaction than R11 by itself and strains that have the correct spacing between arginines (R3(ΔGS10)R11) also have greater compaction levels than arginines alone (R3R11). Randomly inserted arginines, such as R6, do not cause compaction, furthering the evidence that the positioning of these arginines matters. The effect is specific to arginines, as lysines in the same positions did not cause compaction. Compaction mediated by H2A arginines is uniform across the chromosome arm and is not strain specific. Finally, it is not caused by differences in nucleosomal occupancy or differences in the cell cycle.

2.3.4 H2A NTD arginines decrease nuclear volume in yeast

2.3.4.1 H2A R11 causes a decrease in nuclear volume

Because chromatin is known to be physically attached to the nuclear envelope[200-202] and may affect its size[175,176], we investigated whether H2A NTD arginines cause changes to nuclear volume. To this end, we tagged a nuclear pore protein, Nup49, in its chromosomal locus with GFP to visualize the nuclear envelope. We then used 3D confocal microscopy to quantify the nuclear volumes in our yeast strains. As compared to WT, strains containing R11 and R3R11 decreased nuclear volume by 21% and 14%, respectively (Figure 2-12). H2A strains with serines removed, ΔS15 and ΔGS10, increased the mean nuclear volume by 14% and 21%, respectively. Combinations of H2A NTD arginines and serine removal, R11ΔS15 and

R3(ΔGS10)R11, subsequently restored nuclear volumes back to WT levels. Strains with H2A NTD lysines, except for K11, increased nuclear volumes, as well as R6 and K20R.

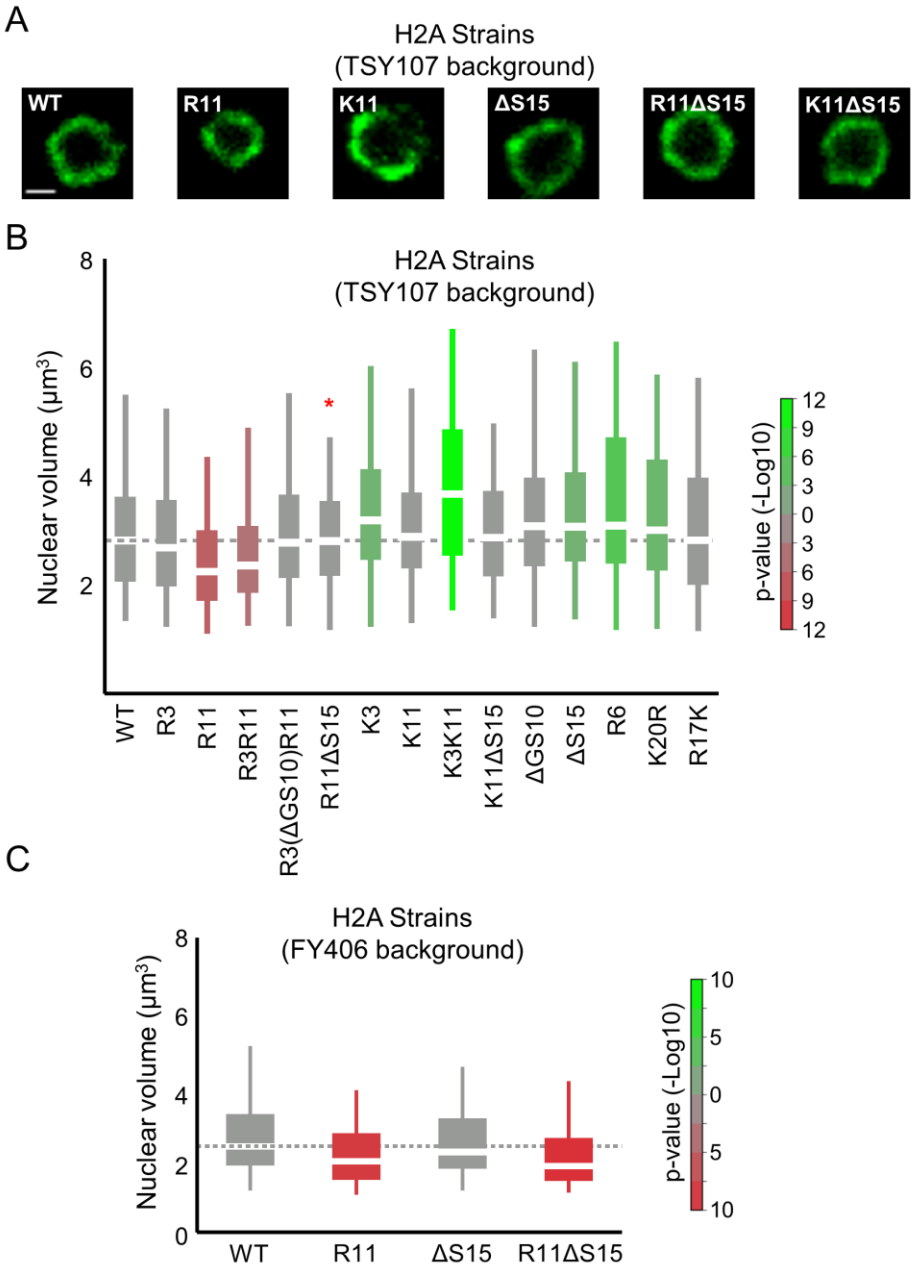

Figure 2-12 – H2A R11 causes a decrease in nuclear volume. (A) Visualization of the nuclear envelope through GFP tagging of Nup49. Boxplot of the distributions of nuclear volumes in the indicated strains in the TSY107 background **(B)** or FY406 background **(C)**. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively. The red star indicates that the mean nuclear volume of R11ΔS15 is significantly smaller than ΔS15 ($p<0.001$).

In all mutants thus far tested, R11 decreases nuclear volume when compared to the appropriate control (Table 2-1). This effect appears to be specific to R11 and unrelated to the levels of chromatin compaction.

Table 2-1 – H2A R11 decreases nuclear volume when compared to appropriate control

Mutant	Control	% Decrease	p-value
R11	WT	-21%	1.4×10^{-5}
R3R11	R3	-11%	1.4×10^{-3}
R11ΔS15	ΔS15	-14%	4.7×10^{-4}
R3(ΔGS10)R11	ΔGS10	-9%	7.7×10^{-2}

2.3.4.2 H2A NTD arginines do not affect cell size

The ratio of the cytoplasmic volume to the nuclear volume (N:C ratio) can be a driving force for the size of the nucleus[175,203]. Therefore, we were interested to know whether H2A NTD arginines were affecting cell size, which then were transmitted as changes in nuclear volume. To measure cellular volume, we stained the cell wall with Concanavalin-A-Rhodamine and visualized the cells under 3D confocal microscopy. Surprisingly, none of the H2A mutants showed any statistical different in mean cellular volume as compared to the WT strain (Figure 2-13).

2.3.4.3 Summary

Together these data show that insertions of arginines into the H2A NTD of yeast cause a decrease in nuclear volume. R11, specifically, has a strong effect on nuclear volume, especially

when compared to the appropriate control. The changes in nuclear volume are not due to changes in the cellular size.

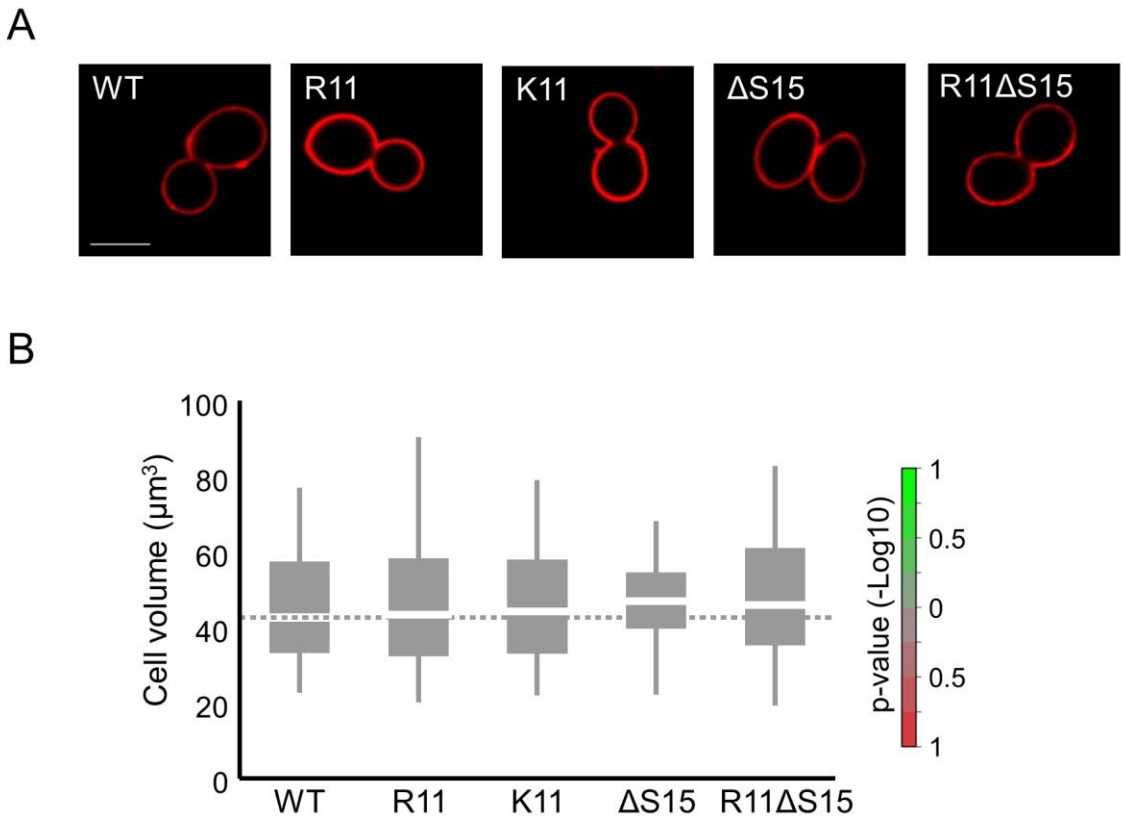

Figure 2-13 – H2A NTD arginines do not change cell size in yeast. (A) Images of cells stained with Concanavalin-A-Rhodamine in the indicated strain. **(B)** Boxplot of the distribution of cellular volumes in the indicated strains. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

2.3.5 Removal of H2A NTD arginines in human cells causes de-compaction of chromatin

2.3.5.1 Removal of R3 or mutation of R11 causes chromatin to de-condense along Chr 1

Because insertion of arginines into the H2A NTD of yeast, an organism with a small genome size and only one arginine natively, caused chromatin compaction, we expected that removal of H2A arginines, specifically R3 and R11, in large genome species, would have the opposite effect and de-compact chromatin. To test this hypothesis, we created an overexpression vector with an

HA tag to either WT H2A or to H2A mutated to remove R3 (ΔR3), replacement of R11 (R11A), or combination of the two (ΔR3R11A). Overexpression was driven by the strong CMV promoter. These plasmids were then transfected into the normal human lung fibroblast cell line, IMR90, or the breast cancer cell line, MDA-MB-453, and FISH was performed using two probes spaced ~0.49 Mbp apart on Chr 1.

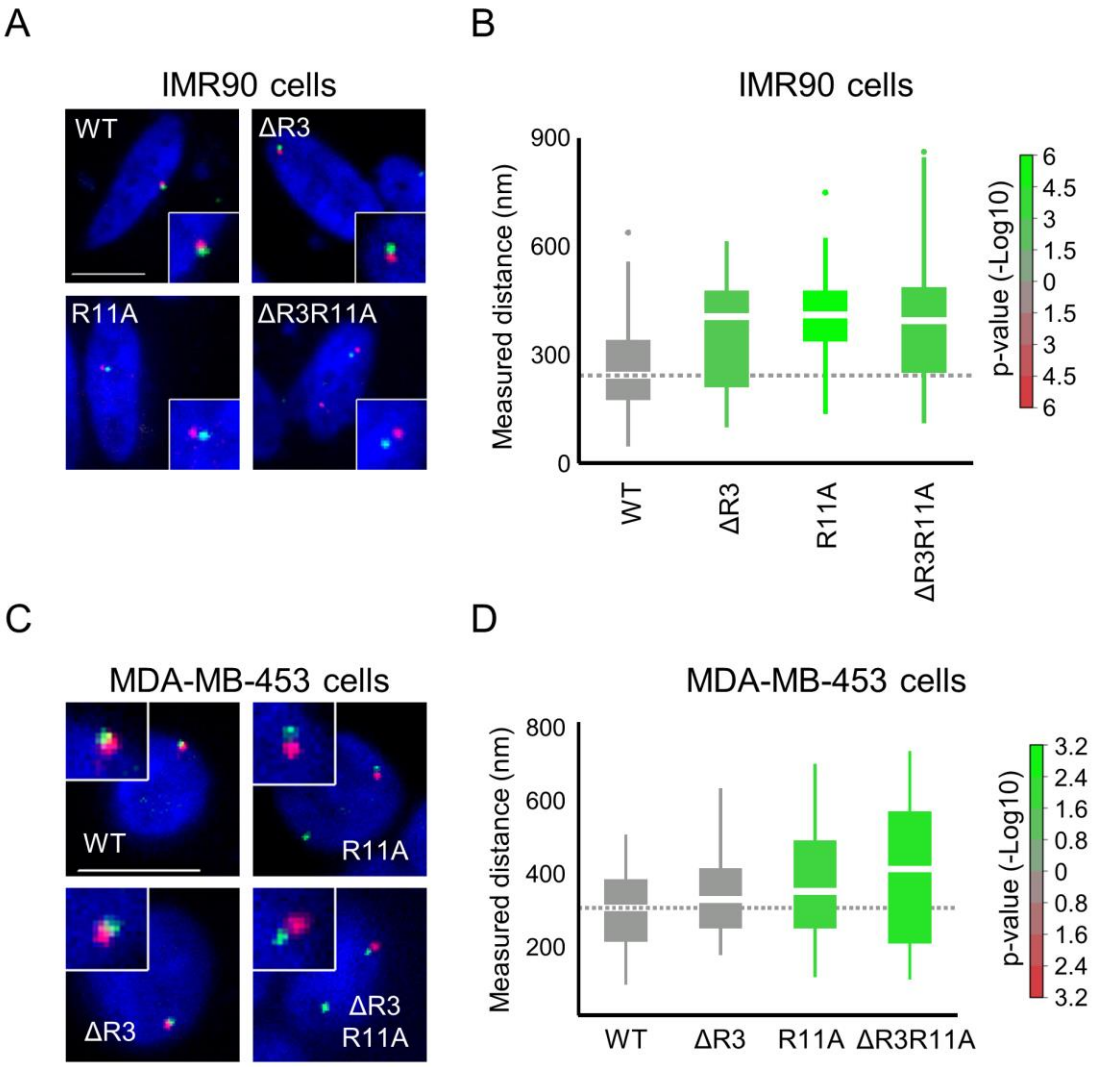

Figure 2-14 – Removal of H2A NTD arginines de-compacts chromatin in human cell lines. Images of FISH probes on Chr 1 in IMR90 cells **(A)** or MDA-MB-453 cells **(C)**. Boxplot of the distribution of interprobe distances in IMR90 cells **(B)** or MDA-MB-453 cells **(D)**. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

As predicted by our hypothesis, IMR90 cells overexpressing any H2A arginine mutant had a distribution of interprobe distances that was larger than WT (Figure 2-14A-B). ΔR3 increased the mean interprobe distance by 23% while R11A and ΔR3R11A increased the mean interprobe distance by 46% and 43%, respectively, as compared to the mean WT distance. In MDA-MB-453 cells, the effects were slightly more modest, as ΔR3, R11A, and ΔR3R11A increased the mean interprobe distance by 15%, 22%, and 31%, respectively (Figure 2-14C-D).

2.3.5.2 Removal of R3 and mutation of R11 causes larger nuclear area in human cells

Because insertion of arginines into the H2A NTD of yeast caused a decrease in nuclear volume, we expect that removal of H2A NTD arginines in a large genome species would increase nuclear volume. To test this hypothesis, we measured the largest nuclear cross-sectional area in cells transfected with WT or mutant H2A plasmids as described above (Section 2.3.5.1). Consistent with our predictions, MDA-MB-453 cells that overexpress R11A, or ΔR3R11A showed a statistically larger cross-sectional nuclear area than WT H2A-overexpressing cells, with increases of 43% and 21%, respectively (Figure 2-15).

2.3.5.3 Summary

Large genome species natively contain arginines in the H2A NTD. Removal of these arginines in the human cell lines IMR90 and MDA-MB-453 causes de-compaction of chromatin as measured by FISH and increases in nuclear area. These two results are in agreement with our results using yeast as a model system (Sections 2.3.3 and 2.3.4), where insertions of H2A NTD arginines caused chromatin compaction and decreases in nuclear volume. This suggests that the effects of H2A NTD arginines are applicable across a wide-range of eukaryotic organisms with varying genome sizes.

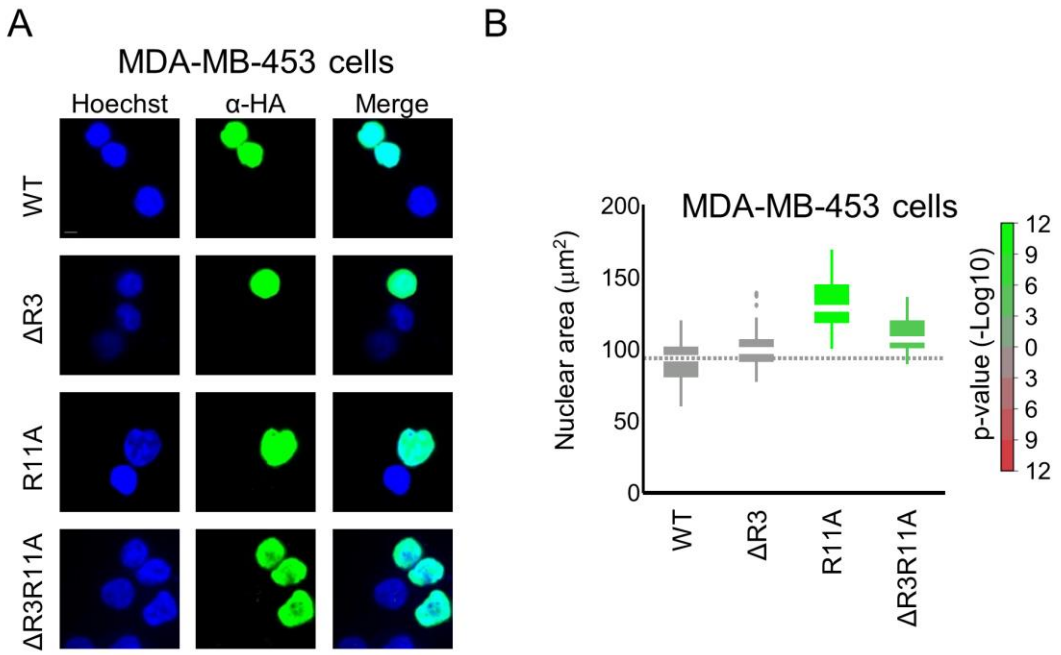

Figure 2-15 – Removal of H2A NTD arginines increases nuclear area in human cells. Images of the nucleus in IMR90 cells **(A)** or MDA-MB-453 cells **(D)**. Boxplot of the distribution of the largest cross-sectional nuclear area in IMR90 cells **(B)** or MDA-MB-453 cells **(E)**. **(D)** Boxplot of the distribution of median fluorescence intensities of each IMR90 cell. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively. A.U., arbitrary units.

2.3.6 H2A R11 directly regulates compaction of nucleosomal arrays *in vitro*

2.3.6.1 Nucleosomal arrays containing ΔR11 are less compact than arrays containing WT H2A

To determine whether the chromatin compaction mediated by H2A NTD arginines, and specifically R11, are through a direct mechanism on the fiber, we used an *in vitro* assembly system to measure compaction. We assembled nucleosomal arrays through step-wise salt dialysis of a DNA template with 12 copies of the 177 bp "601" nucleosome positioning sequence (601-177-12) with histone octamers containing either *X. laevis* WT H2A or ΔR11 H2A. The quality of the assembly was measured by the amount of free DNA following digestion with ScaI (Figure 2-16A) The assembled arrays were incubated in solutions with and without $MgCl_2$ and

then subjected to analytical ultracentrifugation to determine the sedimentation coefficient through measurement of sedimentation velocities with the van Holde-Weischet analysis[204].

Consistent with a more open chromatin structure, ΔR11 arrays sedimented at 31.0 S while WT arrays sedimented at 33.1 S in the absence of $MgCl_2$ (Figure 2-16B). In the presence of 0.8 mM $MgCl_2$, both arrays became more compact with sedimentation values for ΔR11 and WT of 37.4 S and 39.3 S, respectively. These data suggest that arrays containing ΔR11 H2A histones are more open than WT H2A containing histones in the absence and presence of $MgCl_2$. A similar result was produced from an independently assembled chromatin experiment (Figure 2-16C).

2.3.6.2 Summary

These data demonstrate that nucleosomal arrays containing an H2A that removes R11 (ΔR11) are more open than arrays containing WT H2A. This effect is independent of $MgCl_2$ because the difference in sedimentation coefficients is apparent in both the absence and presence of $MgCl_2$. These results suggest that R11 has a direct effect on the folding on a chromatin array because other factors that can affect fiber compaction, such as linker length and histone modifications, were well controlled.

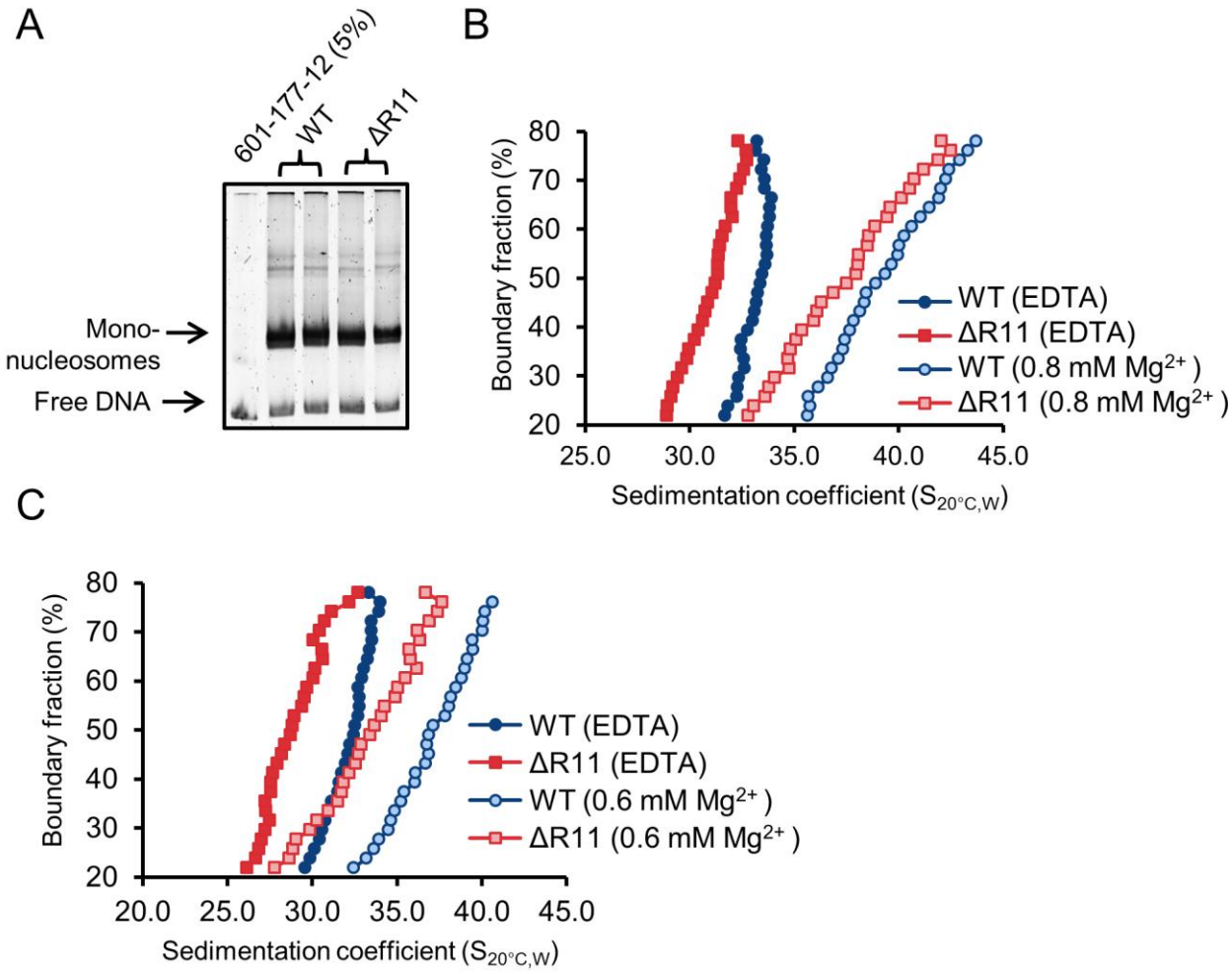

Figure 2-16 – H2A R11 directly modulates chromatin compaction *in vitro*. (A) Polyacrylamide gel electrophoresis (PAGE) of ScaI digested 601-177-12 arrays assembled with either WT H2A or ΔR11 H2A containing octamers. As a control, 5% of the 601-177-12 DNA without octamers was also digested and run. The distribution of sedimentation coefficients determined by von Holde-Weischt analysis plotted against the boundary fraction in the absence or presence of 0.8 mM $MgCl_2$ **(B)** or the absence or presence of 0.6 mM $MgCl_2$ **(C)**. $S_{20°C,W}$ is the sedimentation coefficient corrected to water at 20°C.

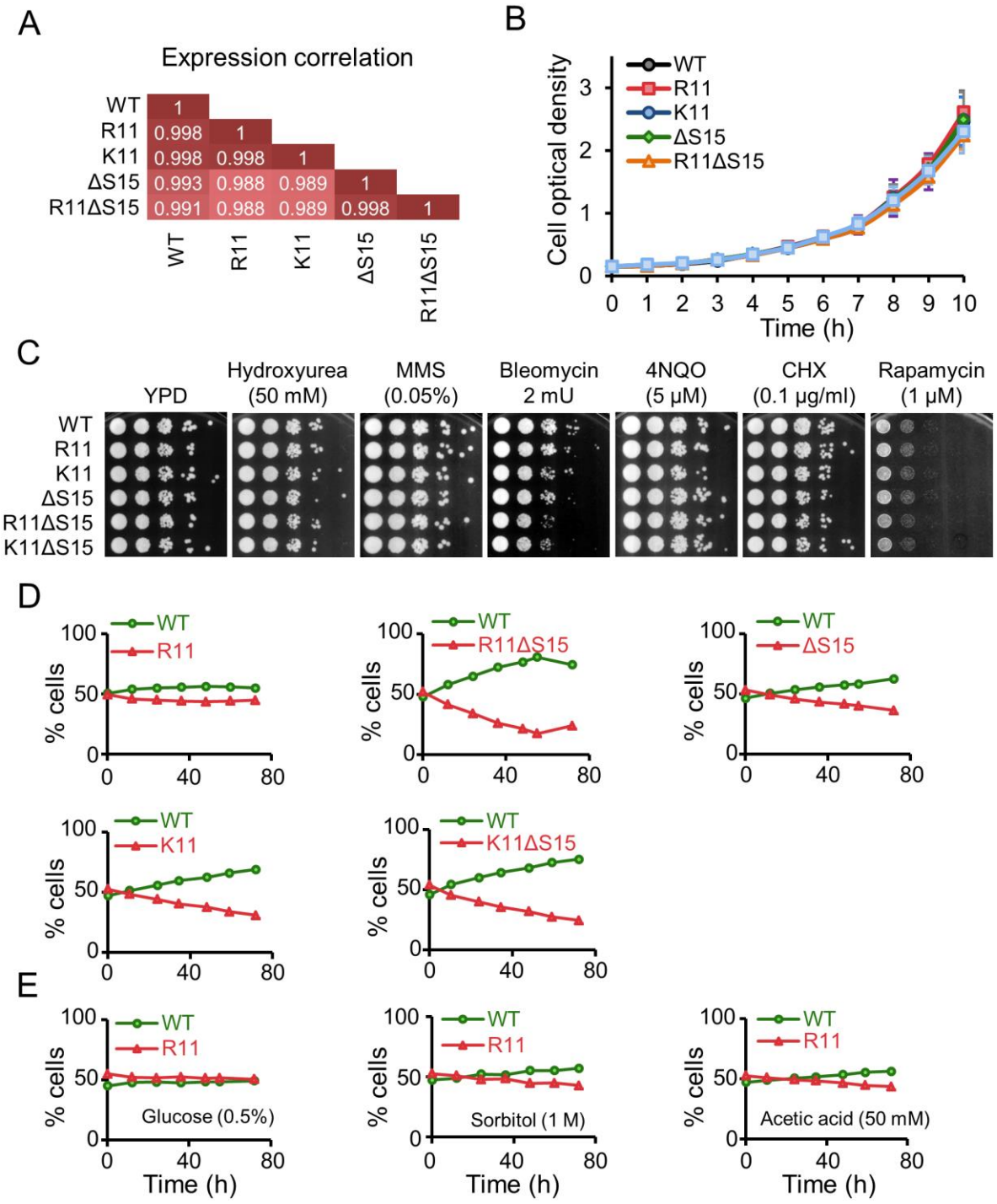

Figure 2-17 – Mutations to the H2A NTD decrease fitness of yeast. (A) Pearson's correlations between global gene expression patterns in the indicated strains when grown in YPD. **(B)** Comparative growth rates of the indicated strains when grown for 10 hrs in YPD. **(C)** Spot tests of the indicated strains grown on YPD or in the presence of the indicated drug. MMS, methyl methanesulfonate. 4NQO, 4-Nitroquinoline 1-oxide. CHX, cycloheximide. The proportion of cells in a co-culture of WT Pgk1p-GFP fusion with the indicated mutant H2A Pgk1p-RFP fusion in YPD **(D)** or the indicated medium **(E)**.

2.3.7 Phenotypic consequences for H2A NTD arginines in yeast

2.3.7.1 Chromatin compaction by H2A NTD arginines does not affect gene expression

Increased chromatin compaction mediated by H2A NTD arginines may interfere with DNA-based processes, such as transcription or replication. To determine if transcription was altered by H2A NTD arginines, we analyzed gene expression patterns in several of our yeast mutants that displayed greater chromatin compaction. Surprisingly we found that expression patterns were remarkably similar in all of our strains with correlations > 0.99 (Figure 2-17A). In the few genes that were more than two fold differentially expressed, no significant gene ontology was enriched. The histone genes were expressed at similar levels in all mutants.

In addition to gene expression, we measured growth rates in YPD and found no significant differences between WT and any mutant H2A strain (Figure 2-17B). There were also no significant sensitivity differences between WT strains and mutant H2A strains on plate supplemented with hydroxyurea, methyl methanesulfonate (MMS), bleomycin, 4-nitroquinoline 1-oxide (4NQO), cycloheximide, and rapamycin, indicating no major defects with DNA replication or repair, protein synthesis, or the TOR signaling pathways (Figure 2-17C).

To measure the fitness of our H2A mutants, we performed competition assays, where WT cells harboring a Pgk1p-GFP fusion were equally mixed and co-cultured with mutant cells harboring a Pgk1p-RFP fusion. The cultures were maintained in exponential growth phase and samples were taken every 12 hrs to be analyzed by flow cytometry for the percentage of cells containing GFP vs RFP. All H2A mutant strains, except R11, were outcompeted by WT cells (Figure 2-17D). Cultures setup with the opposite tags (WT harboring Pgk1p-RFP and mutant strains harboring Pgk1p-GFP) showed similar results (data not shown). This suggests that any mutation to the

H2A NTD, regardless of the effect on compaction or nuclear volume, can decrease the fitness of the yeast cell. Interestingly, R11 grew competitively with WT in all conditions tested, including rich media (YPD), starvation stress (0.5% glucose), pH stress (acetic acid), and osmotic stress (sorbitol) (Figure 2-17D-E).

2.3.7.2 Summary

Chromatin compaction mediated by H2A NTD arginines does not interfere with transcription because all gene expression levels are nearly identical across strains. However, mutations made to the tail of H2A decrease the overall fitness of the strain, although it was not apparent in a short-term growth curve assay. Interestingly, the presence of R11 allows for chromatin compaction without affecting gene expression or fitness levels.

2.3.8 Clinical significance of evolutionarily changed H2A residues

2.3.8.1 A SNP in the H2A histone fold associated with autism does not alter chromatin compaction

The H2A NTD displayed the strongest correlations of protein sequences in relation to increasing genome size, but other residues within the H2A histone fold have similar trends, albeit weaker. In particular, there are two residues in yeast, I44 and S46, which are present in many small genome sized organisms. However, those residues are changed to V44 and A46 in higher eukaryotes, such as humans. Interestingly, in humans, G45, located between these two evolutionarily changed residues, has a SNP associated with autism, which mutates G45 to A45[205].

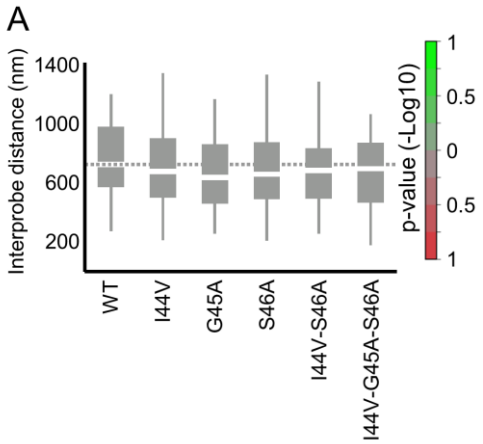

Figure 2-18 – Some H2A mutations in the histone fold do not affect chromatin compaction. Boxplot of the distributions of interprobe distances for probes spaced 275 kbp apart in the indicated strains associated with autism. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

We tested these evolutionarily changed residues, I44V and S46A, as well as the autism SNP, G45A, or combinations thereof, in yeast using the FY406 background strain for differences in chromatin compaction by FISH (Figure 2-18A). No single mutation (I44V, G45A, S46A) or combination mutation (I44V-S46A or I44V-G45A-S46A) had any effect on measured interprobe FISH distances. This indicates that while the SNP is associated with autism, it does not affect chromatin compaction. Additionally, these data reinforce the idea that not every evolutionarily changed residue within histones will affect chromatin compaction as I44V, S46A, and I44V-S46A had no compaction effect. This was previously observed when K20R, K11, and R6 had no effect on interprobe distances (Figure 2-7C).

2.3.8.2 Mutations to R11 seen in cancer may influence chromatin compaction

Alterations to chromatin structure and nuclear volume are often pathologic hallmarks of cancer cells[206] and thus we hypothesized that some cancers may have mutations in the H2A NTD. Using the COSMIC database[207], we found 41 documented missense mutations within the H2A

NTD, of which 29 (71%) affected a residue within an H2A NTD arginine motif (Figure 2-4, Figure 2-19A). Because arginine motifs comprise 55% (12 out of 22 residues) of the H2A NTD, we would expect to observe only 22 mutations within an H2A NTD arginine motif if mutations occurred by random chance ($p<0.05$).

Since R11 was the most frequently mutated residue in the H2A NTD in various cancers (Figure 2-19A) and is the single residue that most affects compaction (Figure 2-7), we were interested to know whether the mutations seen to R11 may affect chromatin structure and nuclear size. We created yeast strains in the FY406 background that mimicked some of the cancer mutations seen at R11 – C11, P11, and H11 – and measured compaction by FISH and nuclear volume (Figure 2-19B-C). Interestingly, H11 was able to compact chromatin to similar levels as R11 but neither C11 nor P11 had interprobe distances different from WT. The effect of H11 may be explained by the fact that histidines can be positively charged (imidazole side chain pKa=6.0) and may mimic the positive charge of R11 in that location. None of the H2A NTD cancer mutations showed an effect on nuclear volume when compared to WT cells (Figure 2-19C).

The results of the H2A cancer mutations need to be considered within the correct context. These mutations are known to occur in various cancers in human cells, which normally contain H2A genes that include R11. Therefore, the correct control strain with which to compare C11, P11, and H11 is the R11 strain. In this case, both C11 and P11 could have the ability to disrupt chromatin compaction mediated by R11 (Figure 2-19B), as they increase the average interprobe distances compared to R11 by 14% and 9%, respectively, with p-values of 0.058 (C11) and 0.042 (P11). Each of the H2A cancer mutations also have the ability to disrupt the nuclear size

regulation of R11 as they increased nuclear volume by 17% (C11), 16% (P11), and 21% (H11), all with $p<0.001$ (Figure 2-19C).

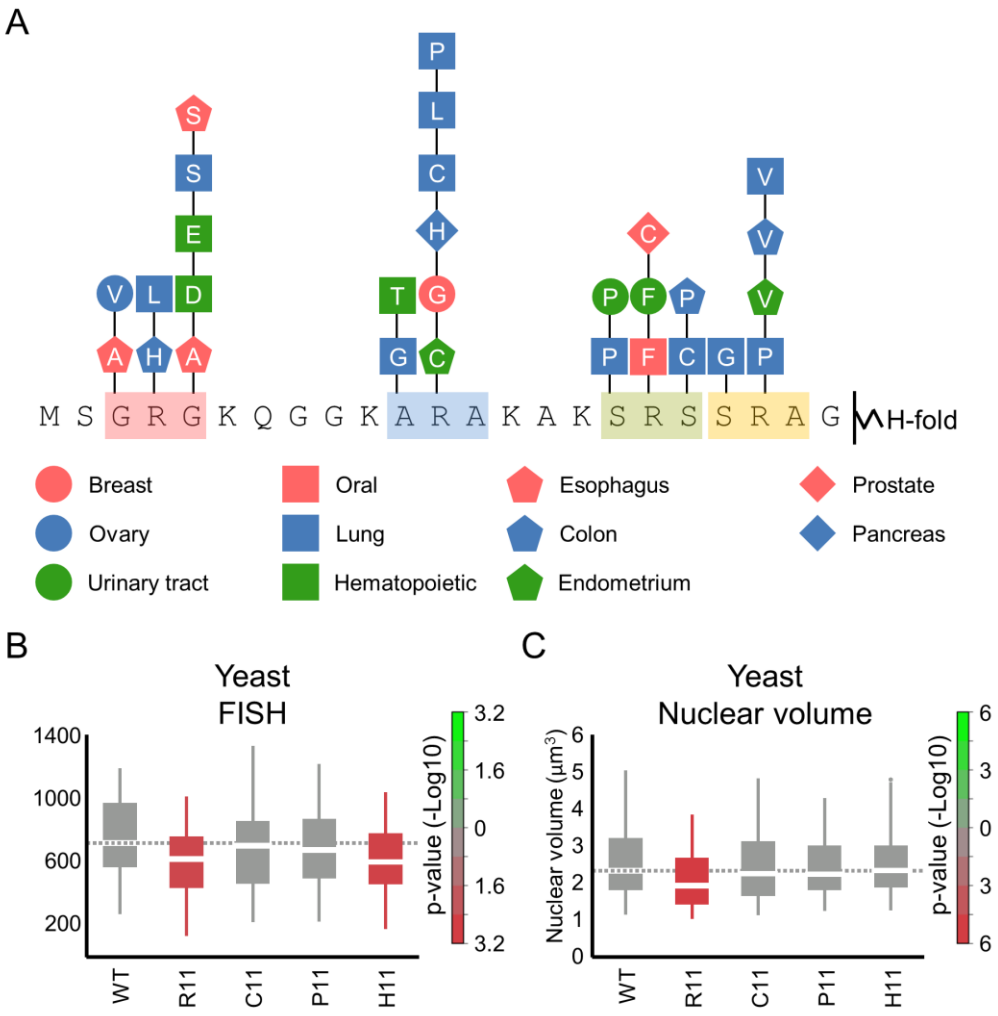

Figure 2-19 – H2A R11 is frequently mutated in cancers and its mutation may affect chromatin compaction. **(A)** Schematic of the H2A NTD showing only the mutations within the arginine motifs found in various cancers as indicated by the colored shapes. The letter within each shape represents the mutated amino acid. Boxplots of the distributions of interprobe distances **(B)** or nuclear volumes **(C)** in the indicated strains. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

2.3.8.3 Summary

Some evolutionary changes within the H2A histone fold, I44 to V44 and S46 to A46, do not affect chromatin compaction as measured by FISH. A SNP associated with autism is located

between those two residues and also does not affect chromatin compaction. Within cancer cells, the R11 motif is frequently mutated. Disruption of R11 with either C11 or P11 may alter chromatin structure, as supported by single copies of H2A disrupting chromatin structure in our overexpression studies.

2.4 Discussion

These results demonstrate that evolutionary adaptations within the histone H2A NTD correlate with increases in genome size and aid in genome compaction. As genome size increases in eukaryotes, the number of arginines in the H2A NTD increases while the numbers of serines and threonines decreases. These changes occur within defined amino acid motifs and in specific positions within the H2A NTD at conserved distances from the globular domain. Insertion of H2A arginines that are found in large organisms into the *S. cerevisiae* H2A NTD, an organism with a small genome, leads to greater compaction of chromatin. Removal of H2A arginines from large genome species has the opposite effect making chromatin less compact. Removal of serines from yeast H2A had marginal effects on chromatin compaction *in vivo*, suggesting that the appearance of arginines in larger genomes may have contributed more to genome compaction than loss of serines or threonines.

The evolutionary adaptation of arginines instead of lysines in the H2A NTD most likely occurred due to the biochemical properties of arginines. Although both arginines and lysines are positively charged at physiological pH, the end of the arginine side chain contains a guandinium group which spreads the positive charge across the three nitrogens through resonance stabilization. On the other hand, lysines contain a side chain that terminates with an amino group and has a much stronger point of positive charge localization. Therefore, as suggested by Rohs *et al.*, it takes less energy to displace water from the arginine side chain in order to make a bond than it does with lysines[208]. Currently this is the most likely explanation for why arginines preferentially mediate protein-DNA and protein-protein interactions and are the only residue within the nucleosome core to make strong contacts with DNA[32,33,208]. This may also explain

why arginines have a much stronger effect on both chromatin compaction and nuclear volume in our yeast system than do lysines.

Examination of the crystal structure revealed that two of the H2A NTD arginines found in higher eukaryotes, R3 and R11, are located near the surface of the nucleosome while the other two H2A NTD arginines, R17 and R20, are more buried. The results from the FISH experiments agree well with this observation. We find that addition of only R3 and R11 to the yeast H2A NTD affects compaction while mutations of K20R and R17K have no effect (Figure 2-7). The strong conservation of the H2A arginine positioning relative to the histone fold region suggests that the physical locations of the H2A arginines within the nucleosome structure are important for their function as well (Figure 2-3). Two of our yeast mutants were designed so that the spacing of the arginines mimics the spacing found in species with large genomes. In higher eukaryotes, R11 is always located 12 residues from the histone fold, but our R11 mutant placed the arginine 13 amino acids from the fold. By removing S15, the R11ΔS15 mutant correctly positions R11 in a location that is 12 amino acids from the histone fold. In addition, there are eight residues between R3 and R11 in higher eukaryotes, but in our R3R11 mutant, there were 10 residues between the arginines. We removed G9 and S10 in the R3(ΔGS10)R11 to correctly space the arginines apart. Both of these mutants showed enhanced compaction relative to their proper controls (R11ΔS15 vs R11 and R3(ΔGS10)R11 vs R3R11) indicating that the proper positioning of the H2A NTD arginines is critical (Figure 2-7).

The amount of compaction in yeast displayed by our H2A NTD arginine mutants through the FISH experiments is very significant, especially when considering that we measure linear compaction. Even a 10% reduction in the linear dimension translates to a 27% reduction in

spherical volume. Over all four probe sets, our R11ΔS15 mutant had an average reduction of interprobe distances of 30% when compared to WT, which leads to a 65% reduction in volume (or volumes that are 35% of the WT levels). Considering that the density of DNA (bp/unit vol) in human cells is 4x the density in yeast cells (Section 1.4.1), R11ΔS15 displays compact levels close to what is seen in higher eukaryotes. It is also significant when considering that yeast mitotic chromosomes only compact to about 50%[197]. This is not to say that H2A mediated compaction is the only way higher eukaryotes will compact their genome, as we know that other mechanisms, such as linker histones and DNA-binding proteins like HP1 and condensin, play more prominent roles chromatin structure within higher eukaryotes[124,127].

Because our *in vitro* nucleosomal array reconstitution experiments were performed with histones that had unmodified tails (except for the desired H2A NTD mutation) using DNA templates with equal linker lengths, we concluded that H2A R11 has a direct effect on chromatin compaction. The internucleosomal contact between R11 and the DNA phosphate backbone in the nucleosomal crystal structure[35] (Figure 2-6) leads us to hypothesize that R11 mediates better stacking of nucleosomes. The *in vitro* experiments of long arrays with short NRLs (167 bps) found that compaction of those arrays depended more on the stacking of nucleosomes rather than linker histones[124]. Because yeast have a very similar NRL (~165 bps), we believe that addition of R11 into the H2A NTD in yeast helps to mediate compaction by tighter nucleosomal stacking. However, we are unable to tell whether this stacking is with local nucleosomes due to increased nucleosomal density or through long range contacts due to increased fiber flexibility.

CHAPTER 3

Co-evolution of H2A and H2B for genome compaction

3.1 Summary

In Chapter 2 we demonstrated that arginines within the H2A NTD have evolved systematically with increasing genome size to aid in chromatin compaction. Within the crystal structure, the H2A NTD is in close physical proximity to the H2B CTD. In this chapter, we outline a few trends within the H2B CTD protein sequence that also evolved with increasing genome size. Individually, these changes can affect chromatin compaction but to a lesser extent than changes within the H2A NTD. However, chromatin compaction is enhanced by simultaneous changes to the H2A NTD and the H2B CTD. We term this region of the nucleosome the H2A-H2B Compaction Domain (ABC Domain) and show that residues within this region have generally shifted towards increasing the positive charge along the surface of the nucleosome.

3.2 Introduction

3.2.1 The *in vivo* role of the H2B CTD

Unlike H2A, H2B does not have a proper C-terminal tail that protrudes from the nucleosome, but it does have a C-terminal α-helix (H2B αC) that is normally included in crystal structures, indicating that it has defined structure[19,35,125]. The most studied residue within the H2B αC is K121 (K123 in yeast). Ubiquitination of K121 is conserved from yeast to humans and is related to transcription through crosstalk with the COMPASS complex that will methylate H3K4 at promoters[209]. It is ubiquitinated by the enzymes Rad6 and Bre1 and de-ubiqutinated by Ubp8 and Upb10[210]. As was mentioned in Section 1.2.2, ubiquitination of H2B K121 has been implicated in chromatin compaction (see section for more details)[68]. Two other residues within the H2B αC have been found to be involved with chromatin structure. Mutation of H2B R102 to alanine enhanced the binding of Sir4 and increased repression of telomeric reporter genes[211]. However, mutation of H2B K111 to alanine had the opposite effect.

Interestingly, in the crystal structure of the tetranucleosome, the H2B αC, and specifically K126, makes an internucleosomal contact with the same residue of a different nucleosome[125]. Because the nucleosomes were crystalized in a zigzag two-start helical conformation, this internuclesomal contact is not between consecutive nucleosomes along the fiber. The *in vivo* implication of this contact is still unknown and is the subject of Chapter 3.

3.3 Results

3.3.1 Evolutionary changes to the H2B CTD

3.3.1.1 The H2A NTD and H2B CTD are physically located near each other in the nucleosome

Crystal structures of the mononucleosome and tetranucleosome show that the H2A NTD protrudes from the top and bottom of the octamer on the opposite side of the DNA entry/exit point[19,35,125,212]. Because H2A NTD arginines can make possible intra- and inter-nucleosomal contacts in this area (Figure 2-5), we investigated other possible residues that are physically located in this vicinity that could have co-evolved with the H2A NTD to aid in compaction. Inspection of the crystal structures revealed that the H2B CTD is situated near the H2A NTD and portions of the two tails parallel one another as they protrude from the octamer (Figure 3-1A).

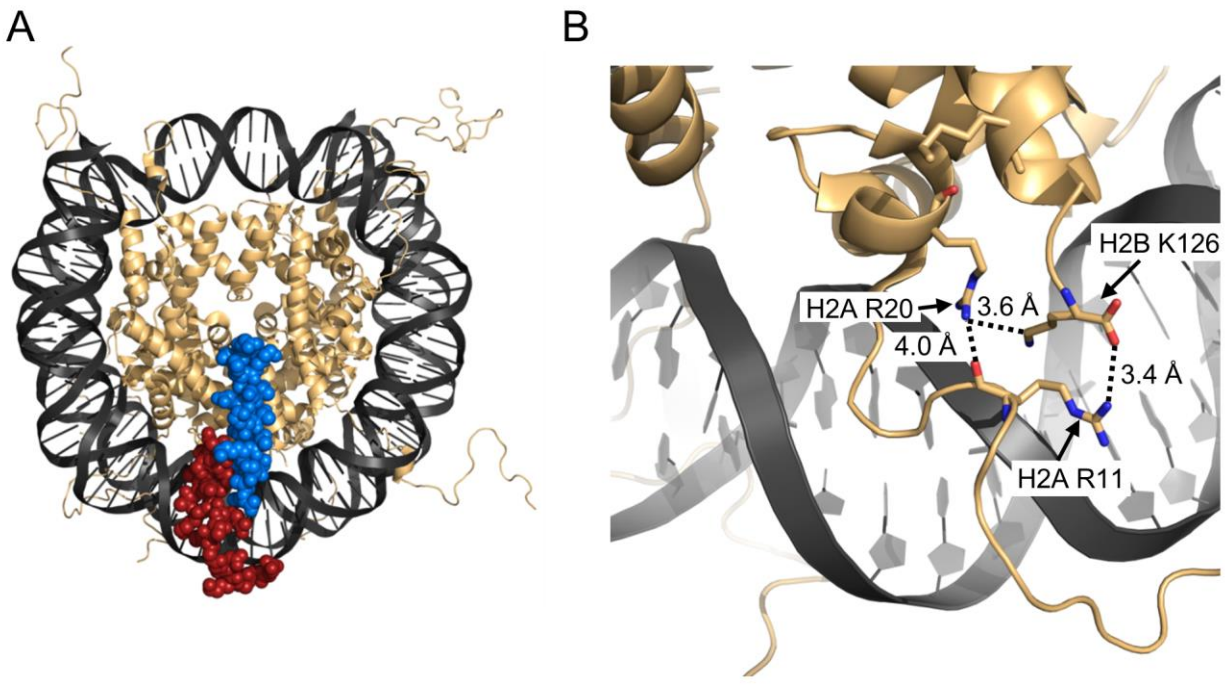

Figure 3-1 – The H2A NTD and H2B CTD occupy adjacent regions in the nucleosome. (A) Crystal structure view from the top of the nucleosome. Red is the H2A NTD. Blue is the H2B CTD. **(B)** Close up view of the region surrounding the C-terminus of H2B. Highlighted residues are colored by atom. Red is oxygen. Blue is nitrogen. Note the orientation of H2B K126 and H2A R11.

Closer examination of the residues within the parallel segments revealed an interesting looping of the H2A NTD to form a pocket of positively charged amino acids in three dimensional space that is not evident from the one dimensional sequence. This region includes two of the evolutionarily changed H2A NTD arginines, R11 and R20, as well as the terminal lysine of the H2B CTD, K126 (Figure 3-1B). The closest atom-to-atom distances between any pair of residues is less than 4 Å, with H2A R20 and H2B K126 located 3.6 Å apart, H2A R11 and H2B K126 located 3.4 Å apart, and H2A R11 and H2A R20 located 4.0 Å apart. In the crystal structure[35], both the positively charged guanidinium group of H2A R20 and the positively charged amino group of H2B K126 point toward the DNA phosphate backbone of the DNA gyre at the -42 position. The guanidinium group of H2A R11 points away from the DNA backbone.

3.3.1.2 H2B CTD lysines and polar residues correlate to genome size

To determine whether amino acids within the H2B CTD have evolved with increasing genome size, we performed residue composition analysis on the 153 species in our dataset that currently had an H2B protein sequence publically available. Analysis of single amino acids revealed that the number of lysines within the H2B CTD increases with increasing genome size (Figure 3-2A). In order for this positive correlation to be become apparent, we had to slightly modify the genome size categories of species in our dataset. We split the "small" genome species (previously species with a genome size of <100 Mbp) into two categories. "Extra-small" now refers to species that have a genome size of <25 Mbp while "small" refers to species with a genome size between 25 Mbp and 100 Mbp. The new genome size categories revealed that the number of lysines increased in the transition from extra-small genome sized species to small genome sized species. Residue counts of other single amino acids do not show strong trends.

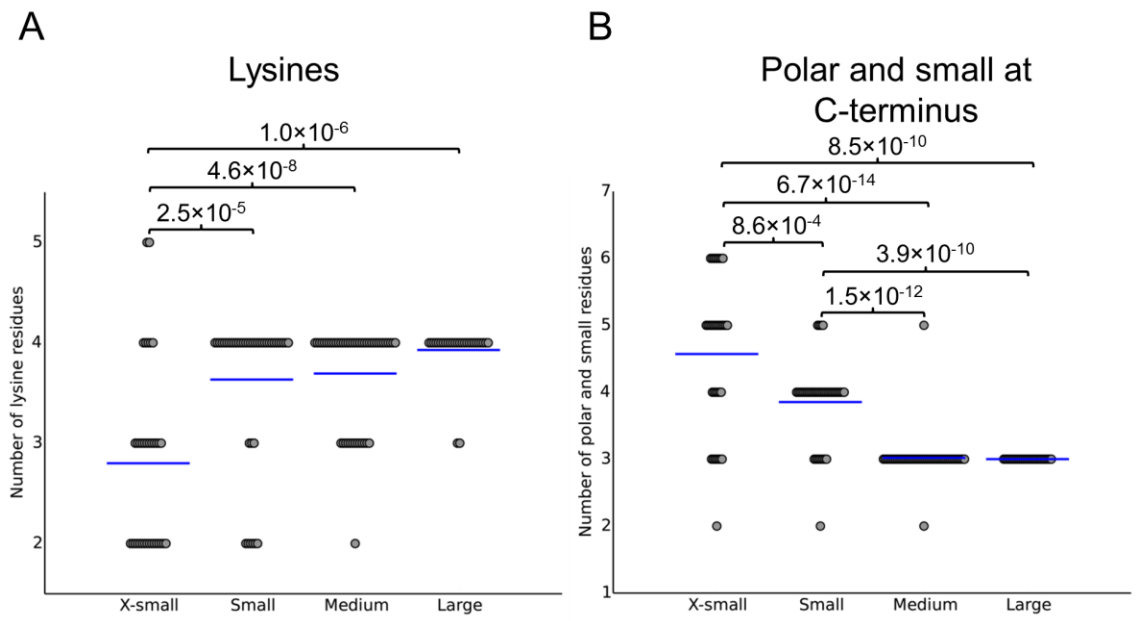

Figure 3-2 – Evolutionary trends in the H2B CTD. Distributions of the number of lysines **(A)** or the number of polar and small residues **(B)** in the H2B CTD. Polar residues include serines (S), threonines (T), asparagines (N), and glutamines (Q). Small residues include alanines (A) and glycines (G). Genome size categories are: X-small (<25 Mbp), Small (25-100 Mbp), Medium (100-1000 Mbp), and Large (>1000 Mbp).

In addition to single amino acid residue composition analysis, we examined the protein sequences of the H2B CTD to determine if there were other trends. Interestingly, we noticed that there was a stretch of polar residues (serines [S], threonines [T], asparagines [N], or glutamines [Q]) and residues with small side chains (alanines [A] or glycines [G]) near the C-terminus that varied with genome size (Figure 3-2B). This is the region of the H2B CTD that parallels the H2A NTD (Figure 3-1A). In this grouped analysis, the number of polar and small amino acids at the C-terminus of H2B decreases with increasing genome size.

The evolutionary trends within the H2B CTD display a striking similarity with the evolutionary trends in the H2A NTD. For instance, both the H2A NTD and H2B CTD show a positive correlation with the number of positively charged residues in regards to genome size. In the H2A NTD, the number of arginines increases with increasing genome size, while in the H2B

CTD, the number of lysines shows an analogous pattern. At the same time, both the H2A NTD and the H2B CTD show a negative correlation between the number of polar residues and genome size. Overall, the trends in the H2A NTD are stronger than the trends in the H2B CTD (compare p-values of 10^{-8} for H2B lysines vs 10^{-22} for H2A arginines).

3.3.1.3 The H2B CTD acquires a lysine at the C-terminus

Because of the similarities between the acquisition of H2A NTD arginines and H2B CTD lysines as genome size increases, we hypothesized that the H2B CTD lysines are also precisely positioned within the tail much like the H2A NTD arginines are positioned. To determine whether this is the case, we first examined representative H2B CTD protein sequences from an organism in each of the four genome size categories (Figure 3-3A). Inspection of these sequences revealed that a glutamine (Q130, yeast numbering) near the C-terminus of H2B in *S. cerevisiae* is mutated to a lysine (K126, human numbering) in *A. capsulatus*, *D. melanogastger*, and *H. sapiens* (Figure 3-3A – orange highlight). Positioned between the Q130/K126 residue and a conserved tyrosine (Y) (See Figure 3-3A – blue highlight), there is a stretch of amino acids consisting of polar and small side chains (Figure 3-3A – green highlight). In *S. cerevisiae* and *A. capsulatus*, this stretch is comprised of four residues and in *D. melanogaster* and *H. sapiens*, this stretch consists of three residues. Much like the conservation of R11 from this histone fold region in the H2A NTD (Figure 2-3), the stretch of four amino acids that changes to three amino acids may be positioning the K126 in a conserved location from the H2B histone fold.

Alignment of the H2B CTD protein sequences in all 153 species demonstrated that the patterns observed previously hold up well across a wide range of eukaryotes (Figure 3-3B). Like yeast, most extra-small genome sized species contain Q130 which is then mutated to a terminal lysine

(K126) in small, medium, and large sized genome species. Most extra-small and small sized genome species contain a stretch of four amino acids near the C-terminus between the conserved tyrosine and terminal lysine which subsequently becomes three amino acids in medium and large sized genome species.

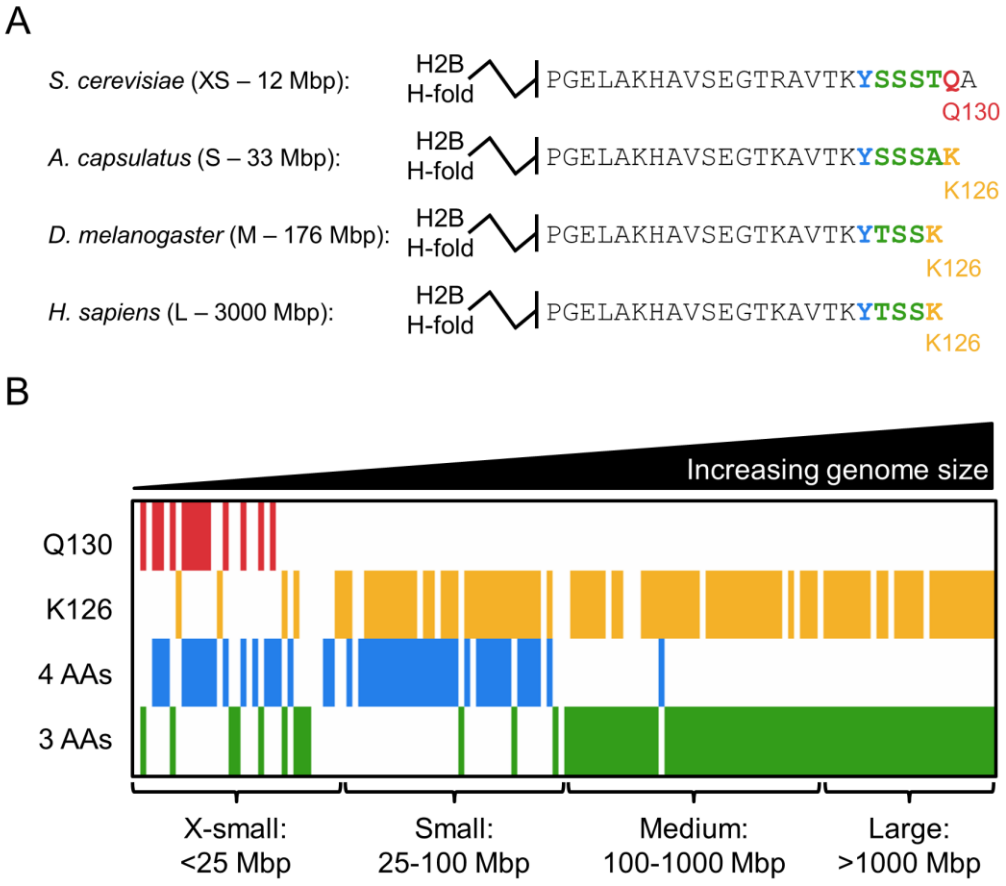

Figure 3-3 – H2B CTD evolutionary changes. (A) H2B CTD protein sequences from the indicated species in each genome size category. XS: Extra-small, S: Small, M: Medium, L: Large. **(B)** Heatmap of the H2B CTD residue composition at the indicated residue (Q130 or K126) or the presence of four amino acids or three amino acids (green highlight of part **(A)**) near the C-terminus. Species are ordered by increasing genome size.

3.3.1.4 Summary

The H2B CTD and the H2A NTD are physically located adjacent to each other in the structure of the nucleosome and parts of each tail parallel one another as they protrude from the core. In

higher eukaryotes, such as *X. laevis*, the H2A NTD loops around to allow R11 and R20 form a pocket of positive charge with the H2B K126. Much like the H2A NTD, the H2B CTD contains a few evolutionarily changed residues the correlate with genome size, albeit in a weaker manner than H2A. A terminal lysine, K126, is acquired in the H2B CTD as genome size gets larger than 25 Mbp. In addition, amino acids with polar or small side chain are lost from the C-terminus of H2B as genome size becomes greater than 100 Mbp.

3.3.2 The H2B CTD affects chromatin compaction but not nuclear volume

3.3.2.1 Evolutionary changes to the H2B CTD can compact chromatin when combined

Because changes to the H2B CTD correlate with increasing genome size, we predicted that they can compact chromatin much like changes to the H2A NTD. To test this prediction, we used a yeast strain (FY406) that had both of its H2A and H2B genes deleted and instead expresses H2A and H2B on a plasmid with their native promoters[213]. We then mutated the H2B gene to mimic some of the evolutionary changes described in Section 3.3.1. To test the effect of the acquisition of a terminal lysine seen in small, medium, and large sized genome sized species, we created a mutant Q130K (see Figure 3-3A). We also removed polar and small residues near the C-terminus of H2B by creating mutants ΔS127 and ΔA131. Finally, we combined several mutations together to make ΔS127Q130K and Q130KΔA131. To determine chromatin compaction, we performed FISH using probes belonging to Probe Set A (see Figure 2-7A) spaced 275 kb apart and measured the interprobe distances as described previously.

When compared to an isogenic WT strain, Q130K decreased median interprobe distances by 12% (Figure 3-4A). Removal of polar or small side chained amino acids had a mixed effect. ΔS127 increased interprobe distances by 3% while ΔA131 decreased distance by 8%. The p-

values for the differences in means between these H2B mutants and WT is p=0.066 for Q130K, p=0.70 for ΔS127, and p=0.30 for ΔA131, indicating that no single H2B CTD mutant by itself has a significant impact on chromatin compaction. However, combination H2B mutants showed a greater effect on compaction than single mutants. ΔS127Q130K and Q130KΔA131 decreased interprobe distances by 25% and 12%, respectively (Figure 3-4A), both with p-values < 0.005.

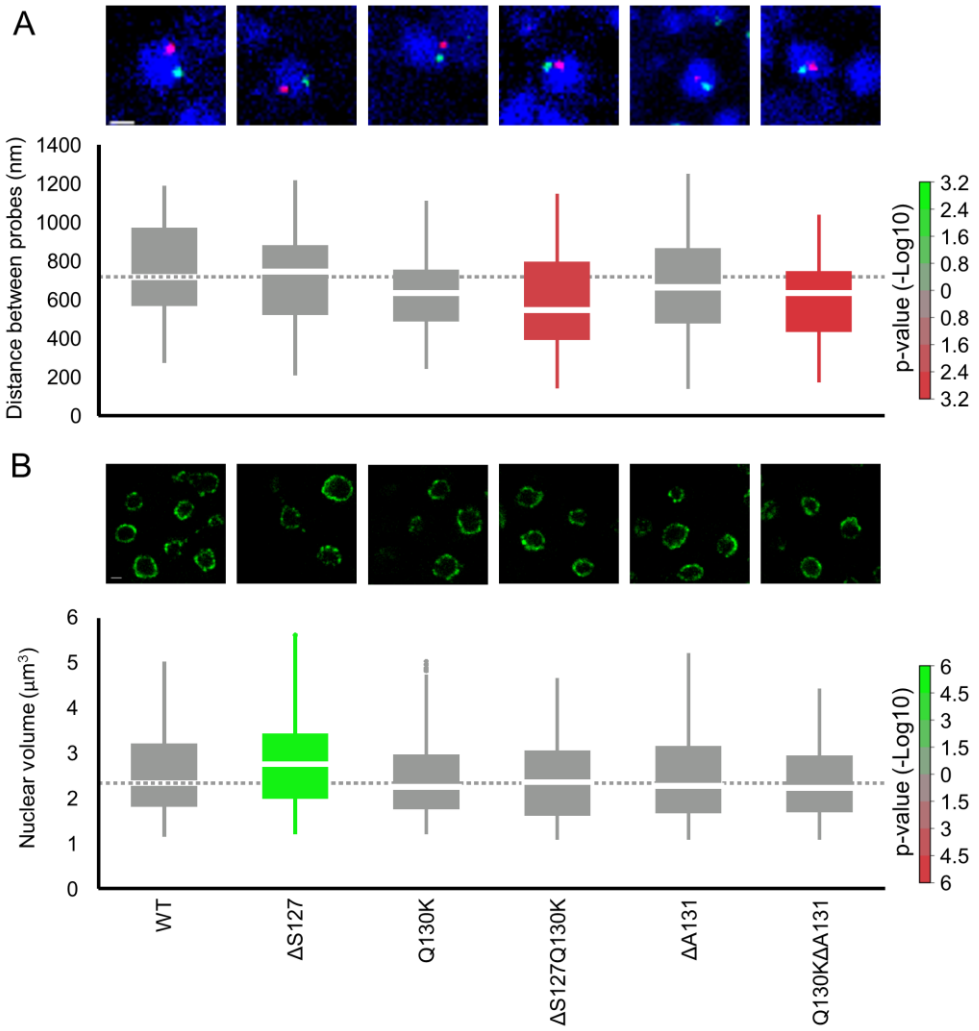

Figure 3-4 – Evolutionary changes to the H2B CTD compacts chromatin. (A) Boxplot of the distributions of interprobe distances measured by FISH. See bottom of figure for strain labels. Representative images of each strain are located above the boxplot. **(B)** Boxplot of the distributions of nuclear volumes for the indicated strains. Representative images are located above the boxplot. Boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

3.3.2.2 Evolutionary changes to the H2B CTD do not affect nuclear volume

To determine if evolutionary changes to the H2B CTD affect nuclear volume, we tagged Nup49 in its genomic locus with GFP in order to visualize the nuclear envelope. Using 3D confocal imaging as described above, we were able to quantify the nuclear volumes of approximately 200 cells in our H2B mutant strains (Figure 3-4B). With the exception of ΔS127, evolutionary changes to the H2B CTD had no effect on nuclear volumes. ΔS127 increased nuclear volumes by 16%.

3.3.2.3 Comparison of chromatin compaction between H2A and H2B single mutants

As mentioned in Section 3.3.1.2, the evolutionary trends of the H2B CTD with genome size were not as significant as the evolutionary trends within the H2A NTD. In order to compare the levels of chromatin compaction mediated by either H2A or H2B, we performed FISH and measured interprobe distances for Probe Set A in the H2A-only mutants R11, ΔS15, R11ΔS15, and K20R in the FY406 strain background. When comparing the percentage difference in means, the percentage difference in medians (Figure 3-5), or p-values, the chromatin compaction mediated by H2A-only mutations is more significant than H2B-only mutations. For instance, when evaluating single mutations, H2A R11 can compact chromatin by 15%, while compaction mediated through H2B Q130K (the largest of any single H2B mutation) is 12%. In addition, with outliers removed, H2B ΔS127Q130K compacted chromatin by 25% while H2A R11ΔS15 compacted chromatin by 29% with a more significant p-value ($\sim 10^{-3}$ vs $\sim 10^{-4}$, respectively).

3.3.2.4 Summary

Single evolutionary trends in the H2B CTD do not have a strong effect on genome compaction. When two H2B CTD mutations are combined (ΔS127Q130K or Q130KΔA131), the compaction

effects are much stronger but still not as strong as H2A NTD mutations. The H2B CTD does not modulate nuclear volume.

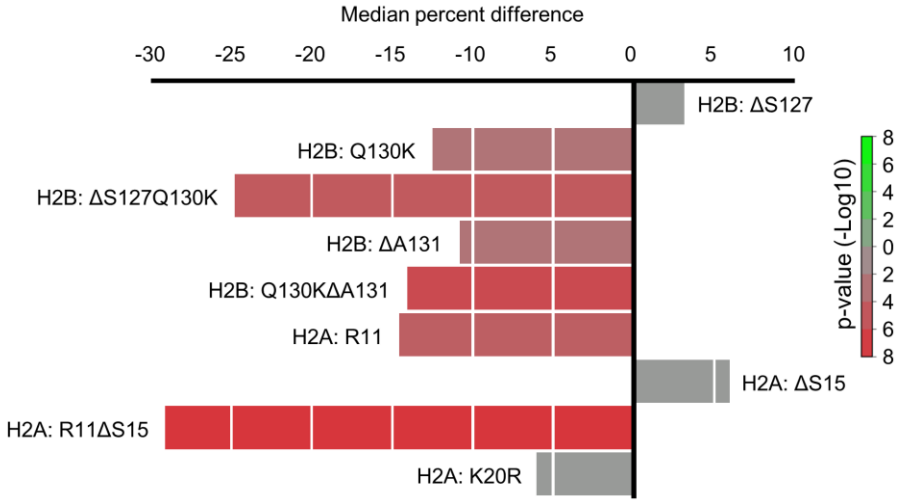

Figure 3-5 – The H2A NTD has a stronger effect on chromatin compaction than the H2B CTD. The median percent difference compared to an isogenic WT of the indicated strain is plotted. The bars are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

3.3.3 The H2A NTD and the H2B CTD co-evolved to aid in genome compaction

3.3.3.1 Combination trends of H2A and H2B correlate strongly with genome size

Both of the protein sequences of the H2A NTD and the H2B CTD correlate to genome size but the genome sizes that their patterns describe are different. For instance, the H2A NTD for both extra-small (<25 Mbp) and small (25-100 Mbp) genome sized organisms contains serines, such as S10 and S15, and one arginine, R17. Medium genome sized organisms (100-1000 Mbp) lose some H2A NTD serines and gain two arginines at H2A positions 3 and 20. Large genome sized organisms gain an additional arginine at R11. On the other hand, in the H2B CTD, extra-small genome sized organisms contain Q130 which is mutated to a lysine (K126) in small genome sized organisms. Both extra-small and small genome sized organisms contain a stretch of four

polar/small residues near the C-terminus of H2B, while medium and large genome sized species contain a stretch of three polar/small residues in the same location.

Table 3-1 – Changes to the H2A NTD and H2B CTD can describe genome sizes in eukaryotes

Histone position	Extra-small	Small	Medium	Large
H2A serines	✓	✓		
H2B K126		✓	✓	✓
H2B polar/small stretch	4 AAs	4 AAs	3 AAs	3 AAs
H2A R3/R20			✓	✓
H2A R11				✓

Therefore, we can use the combination of changes to the H2A NTD and H2B CTD to describe genome sizes of eukaryotes from extra-small to large sized genomes. Table 3-1 illustrates these changes.

3.3.3.2 H2A and H2B act synergistically to increase chromatin compaction

Because evolutionary changes in both the H2A NTD and H2B CTD can independently compact chromatin, we were interested to know whether the two tails can work together to further compact chromatin. If so, what is the mechanism of interaction? Do they act additively or not? To answer these questions, we created yeast strains based on the FY406 background that simultaneously have changes to the H2A NTD and H2B CTD. We combined one H2A NTD mutant selected from R11, ΔS15, R11ΔS15, or K11 with an H2B CTD mutant that was selected from either Q130K or ΔS127Q130K and measured compaction through FISH.

Each combination mutant with concurrent H2A and H2B changes can compact the chromatin to a greater degree compared to either a WT strain or single H2A mutants (Figure 3-6). For

instance, in this strain background, R11 compacts chromatin by 15% compared to WT, but R11-Q130K compacts chromatin by 20% and R11-ΔS127Q130K compacts chromatin by 28%. When comparing to R11, the double histone mutants R11-Q130K and R11-ΔS127Q130K further compact chromatin by 6% and 16%, respectively. The same trend holds for ΔS15, R11ΔS15, and K11 when combined with Q130K or ΔS127Q130K. Interesting, both H2A ΔS15 and H2A K11, by themselves, do not statistically change the compaction levels, but when combined with H2B ΔS127Q130K, both show a statistically significant decrease in interprobe distances. The strongest levels of chromatin compaction were seen in R11ΔS15-ΔS127Q130K mutants, which showed a 33% decrease in interprobe distances.

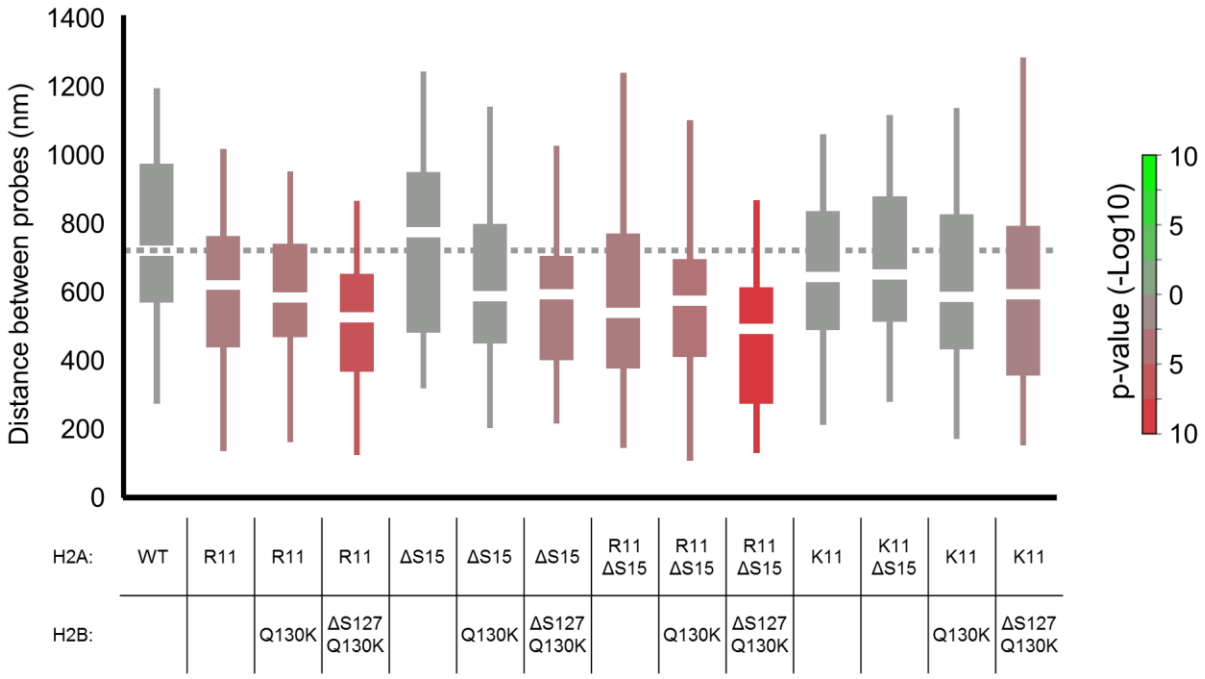

Figure 3-6 – Simultaneous ectopic expression of H2A NTD arginines and H2B CTD lysines compacts chromatin. Boxplot of the distribution of interprobe distances for FISH probes spaced 275 kbp apart. The mutation within H2A and H2B are indicated below the boxes. The boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

Analysis of interprobe distances in FISH experiments using H2A NTD-only mutants, H2B CTD-only mutants, or combination H2A-H2B mutants revealed that compaction mediated by simultaneous H2A-H2B mutations is synergistic but not additive. For example, H2A R11ΔS15 compacts chromatin by 29%, H2B ΔS127Q130K compacts chromatin by 25%, but the R11ΔS15-ΔS127Q130K combination mutant only compacts chromatin by 33% (not the expected 54% if additive). Synergistic but not additive compaction holds for the other H2A and H2B combinations as well, indicating that compaction mediated by H2A and H2B may be through the same or similar mechanisms.

3.3.3.3 Only H2A R11 affects nuclear volume

Unlike their effect on chromatin compaction, most evolutionary changes to the H2B CTD had no effect on nuclear volume (Figure 3-4B). However, H2A and H2B work synergistically to compaction chromatin (Figure 3-6), so it is conceivable that they might do the same for regulation of nuclear volume. To determine if the H2A NTD and H2B CTD cooperate to regulate nuclear size, we performed nuclear volume measurements in the same yeast strains as in the above section (Section 3.3.3.2) that simultaneously express evolutionary changes to H2A and H2B.

Analysis of nuclear volume measurements demonstrated that only strains expressing H2A R11, whether in conjunction with other H2A NTD mutations (such as ΔS15) or H2B CTD mutations, had a decreased nuclear volume (Figure 3-7). Strains that did not express H2A R11, such as ΔS15 or ΔS15-ΔS127Q130K, did not have a mean nuclear volume statistically different from WT. Importantly, the nuclear volume in H2A R11 or H2A R11ΔS15 was not further decreased by the concurrent expression of H2B CTD mutations but rather stayed relatively constant. These

data suggest that H2A R11 regulates nuclear volume independently from other evolutionarily changed residues. It also demonstrates that nuclear volume control can be decoupled from chromatin compaction.

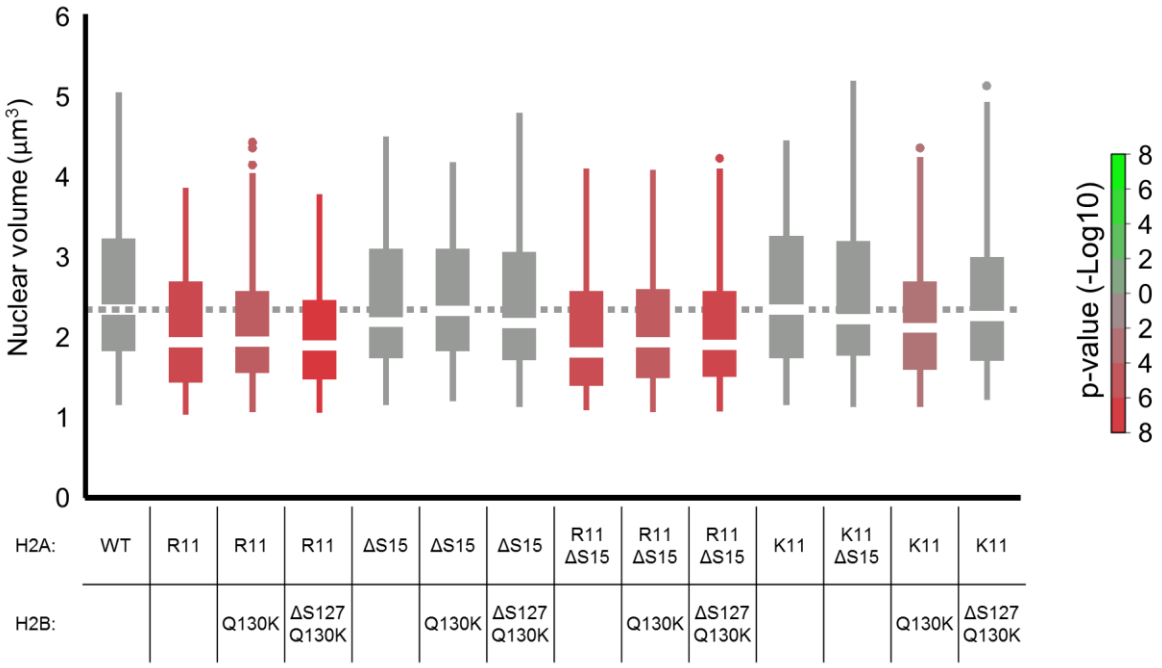

Figure 3-7 – Ectopic expression of H2A R11 decreases nuclear volume in yeast. Boxplot of the distributions of nuclear volume in the indicated strain. The mutation within H2A and H2B are indicated below the boxes. The boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

3.3.3.4 Summary

Analysis of the evolutionary changes to the H2A NTD and the H2B CTD show that they both have co-evolved with increasing genome size. Combinations of simultaneous changes to both regions can represent species of different genome size categories, as well as enhance chromatin compaction. Nuclear volume, on the other hand, seems to be regulated only through the presence of H2A R11.

3.3.4 H2A R20 modulates chromatin compaction through other residues

3.3.4.1 H2A K20R increases chromatin compaction in yeast mutants containing R11

In yeast, H2A K20R had little effect on chromatin compaction by itself (Figure 3-5); however H2A R20 is found in the same physical vicinity as H2A R11 in the crystal structure (Figure 3-1B). Because it has a strong evolutionary trend (Figure 2-2), this led us to predict that H2A R20 may influence other residues surrounding it when put in the correct context. To examine the impact of H2A R20 on chromatin compaction, we designed several yeast mutants in the FY406 background where the K20R mutation was combined with previous H2A arginine and serine mutations and measured chromatin compaction by FISH. As shown in Figure 3-8A, the addition of K20R in the context of R11 or R11ΔS15 (R11K20R and R11ΔS15K20R, respectively) further increased chromatin compaction of those mutants. The median percent decrease of interprobe distances in R11 fell from 15% (R11) to 25% (R11K20R), and in R11ΔS15, it fell from 29% (R11ΔS15) to 36% (R11ΔS15K20R). In the context of ΔS15 alone, the addition of K20R slightly decreased interprobe distances from +6% (ΔS15) to -8% (ΔS15K20R), but neither were statistically significant.

In addition to chromatin compaction, H2A R20 may modulate nuclear volume. We investigated whether the K20R mutation by itself or in combination with other H2A NTD arginine or serine mutations affects nuclear size by measuring nuclear volumes in those same mutants. Although K20R increased chromatin compaction when combined with R11 or R11ΔS15, surprisingly it had no further influence on nuclear volume (Figure 3-8B). As was seen previously (Figure 3-7), only mutations that include H2A R11 decrease nuclear volume in yeast and the decrease is not further impacted by surrounding mutations.

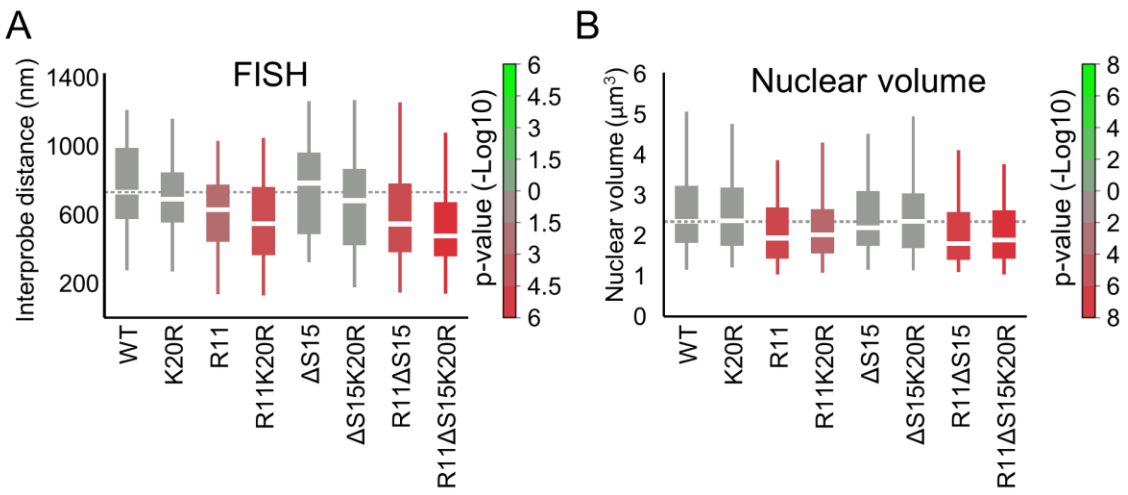

Figure 3-8 – H2A K20R modulates chromatin compaction but not nuclear volume in yeast. (A) Boxplot showing the distribution of interprobe distances measured by FISH of Probe Set A spaced 275 kbp apart in the indicated mutants. **(B)** Boxplot showing the distribution of nuclear volumes in the indicated mutants. The boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

3.3.4.2 Combinations of H2A R11, H2A R20, and H2B K126 cannot further compact chromatin in yeast

Because both H2A R20 (Figure 3-8) and evolutionary changes to the H2B CTD (Figure 3-6) enhanced compaction of H2A R11 mutants, we interested to know whether yeast strains simultaneously containing all three mutations would further compact chromatin. FISH analysis of mutant yeast strains demonstrated that interprobe distances were not further decreased by the simultaneous addition of those three evolutionary changes. Using H2A R11ΔS15 as a basis for comparison, there are no major differences in the level of compaction in R11ΔS15K20R, R11ΔS15-ΔS127Q130K, or R11ΔS15K20R-ΔS127Q130K but all three further compact chromatin relative to R11ΔS15. This same trend occurs in mutants containing only R11 or ΔS15 as well (data not shown).

3.3.4.3 Summary

The physical proximity of H2A R20 with H2A R11 and H2B K126 within the crystal structure suggests that the three residues may act synergistically. H2A R20, in combination with H2A R11, enhances chromatin compaction but has no further effects on nuclear volume. When H2A R11, H2A R20, and H2B K126 are combined, chromatin compaction is not increased more than combinations of H2A R11 with either H2A R20 or H2B K126 alone.

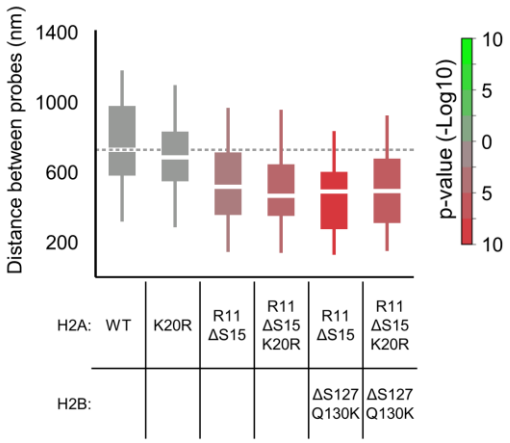

Figure 3-9 – Combinations of H2A R11, H2A R20, and H2B K126 cannot further compact chromatin in yeast. Boxplot showing the distribution of interprobe distances measured by FISH of Probe Set A spaced 275 kbp apart in the indicated mutants. The boxes are colored by significance of their p-value relative to WT according to the scale bar shown to the right. Green and red colors indicate mean values greater or smaller than that of WT, respectively.

3.3.5 The H2A-H2B Compaction Domain (ABC Domain) modulates chromatin compaction across eukaryotes

3.3.5.1 The region physically near the C-terminus of H2B contains the evolutionarily changed residues that most affect chromatin compaction

The results presented here, in Chapter 3 and in conjunction with Chapter 2, have demonstrated that there are specific residues within the H2A NTD and the H2B CTD of eukaryotes that have

co-evolved with increasing genome size. These residues function by increasing chromatin compaction when the genome size is disproportionately larger than nuclear volume, and in the case of R11, help to regulate nuclear size.

Not every evolutionary change to H2A or H2B affects chromatin compaction. In yeast, H2A ΔS15, H2A ΔGS10, H2A K11, and H2B ΔS127 do not statistically alter the interprobe distances measured by FISH (Figure 2-7C, Figure 3-4A). Other mutants that mimic evolutionary changes in the H2A histone fold, I44V, S46A, and I44V-S46A also do not affect compaction (Figure 2-18). Finally there are residues, that by themselves, have no or little impact on compaction (H2A K20R and H2B Q130K) but when put in the correct context (H2A R11K20R or H2B ΔS127Q130K), have greater effects on compaction (Figure 3-4A, Figure 3-8A).

We found that there is a hotspot region of the nucleosome where there are simultaneously many residues that not only change evolutionarily but also affect chromatin compaction. Using the crystal structure where the H2A NTD and H2B CTD were fully crystalized[35], we calculated the nearest atom-to-atom distances for all residues. Table 3-2 summarizes the residues found in this hotspot region. This region, which we call the H2A-H2B Compaction Domain (ABC Domain), contains three out of the four H2A NTD arginines found in higher eukaryotes (R11, R17, and R20), as well as K126 and the stretch of three residues with polar/small side chains. Higher eukaryotes do not have serine 10 that is present in yeast, but in *X. laevis*, whose histones were used in the crystal structure, residue 10 is a threonine (A10 in humans) and T10 is found in the ABC Domain as well. Three residues in this region, H2A R17, H2B Y122 and H2B K121 may have conserved functions. Y122 contains an aromatic ring that interacts very closely with H2A

R17 in possible cation-pi interactions. As mentioned earlier, ubiquitination of K121 de-condenses chromatin and is involved with *trans*-histone crosstalk with H2A R17 and H3 K4[185].

Table 3-2 – Nearest atom-to-atom distances between indicated residue and the H2B C-terminus. Blue highlighted residues shows trends in relation to genome size. In regards to H2B T123, S124, and A125, the actual residue did not change but they are located within the stretch of residues with polar/small side chains that, as a group, correlate to genome size (Figure 3-2 and Figure 3-3).

Histone	Residue	Distance to H2B C-terminus (Å)
H2B	K126	0
H2B	A125 (S125 in humans)	1.37
H2A	R11	3.45
H2A	R20	3.65
H2B	S124	4.20
H2A	A12	5.65
H2B	T123	5.89
H2B	Y122	5.91
H2A	T10 (A10 in humans)	6.78
H2A	K15	6.92
H2B	K121	6.96
H2A	R17	6.96

3.3.5.2 The ABC Domain has increased in positive charge between yeast and Xenopus

We noticed that many of the residues within the ABC Domain contain side chains that are positively charged at physiological pH. To better understand how these changes may have affected the surface of the nucleosome, we took the crystal structure of the *Xenopus laevis* nucleosome where the H2A NTD and H2B CTD were mostly crystalized[35] and used PyMOL[214] to mutate evolutionary changed residues back to the residue found in yeast (see Method section for exact mutagenesis changes). We then used the APBS toolkit[215] to calculate the electrostatic

potential of the H2A NTD and H2B CTD in the original structure and in the mutated structure to resemble the yeast tails (Figure 3-10). When compared to the yeast residues, the residues in the *Xenopus* structure, which represents higher eukaryotes, show an increase of charge density along the surface of the nucleosome. This may be a possible mechanism for better nucleosomal stacking.

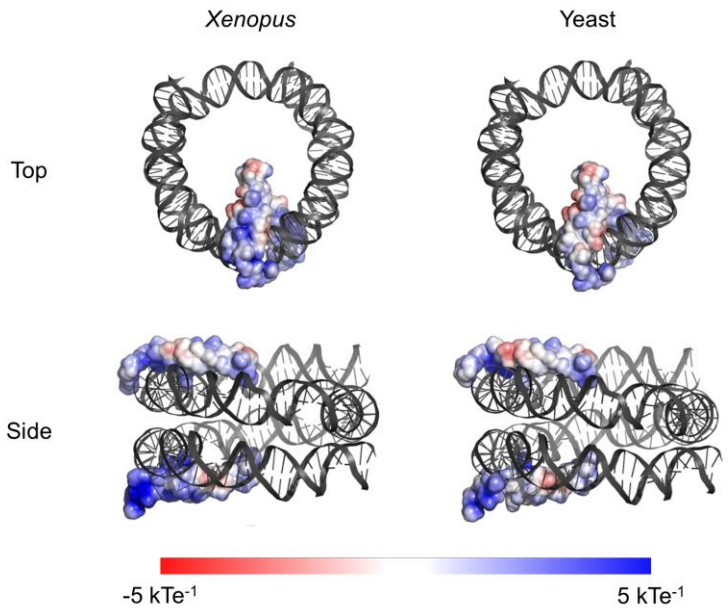

Figure 3-10 – The ABC Domain has increased in positive charge between yeast and *Xenopus*. View of the electrostatic potential of the H2A NTD and H2B CTD with the indicated orientation. The scale bar at the bottom shows the color of the corresponding charge.

3.3.5.3 Summary

There is a hotspot region of the nucleosome that physically contains many residues that change in relation to increasing genome size. These residues seem to be involved with chromatin compaction as outlined here and elsewhere[68,72]. We term this region the H2A-H2B Compaction Domain (ABC Domain). The ABC Domain has generally increased the positive electrostatic potential along the surface of the nucleosome on the opposite side of the DNA entry/exit points, which may lead to better nucleosomal stacking.

3.3.6 The ABC Domain does not alter nucleosomal spacing or gene expression in yeast

3.3.6.1 Nucleosomal spacing is unchanged in H2A and H2B mutants

Although H2A NTD mutations showed no difference in nucleosomal spacing (Figure 2-10B-C), chromatin compaction mediated by simultaneous mutations within the ABC Domain is stronger than H2A NTD arginines alone (Figure 3-6). Therefore, it is conceivable that differences in nucleosomal spacing may explain the increased compaction observed through mutations within the ABC Domain. To determine the occupancy profile of all the nucleosomes within the genome, we performed MNase-Seq in yeast strains with H2A NTD-only mutations, H2B CTD-only mutations, and combination H2A-H2B mutations.

Similar to previous observations, neither mutations within the H2B CTD nor simultaneous H2A-H2B mutations significantly altered the occupancy of nucleosomes throughout the genome when compared to a WT strain (Figure 3-11A). Meta-analysis of several different features within the yeast genome, such as ORFs, ARS, LTRs, tRNAs, snRNAs, telomeres, ncRNAs, Y' elements, retrotransposons, and centromeres also revealed no differences between any of the mutants (Figure 3-11B). Other changes besides occupancy, such as shift or fuzziness, were unchanged in our H2A and H2B mutants as well. These data, along with MNase-Seq analyses in H2A NTD-only strains (Figure 2-10B-C), demonstrate that mutations within the ABC Domain increase chromatin compaction through a mechanism unrelated shifting nucleosome positioning.

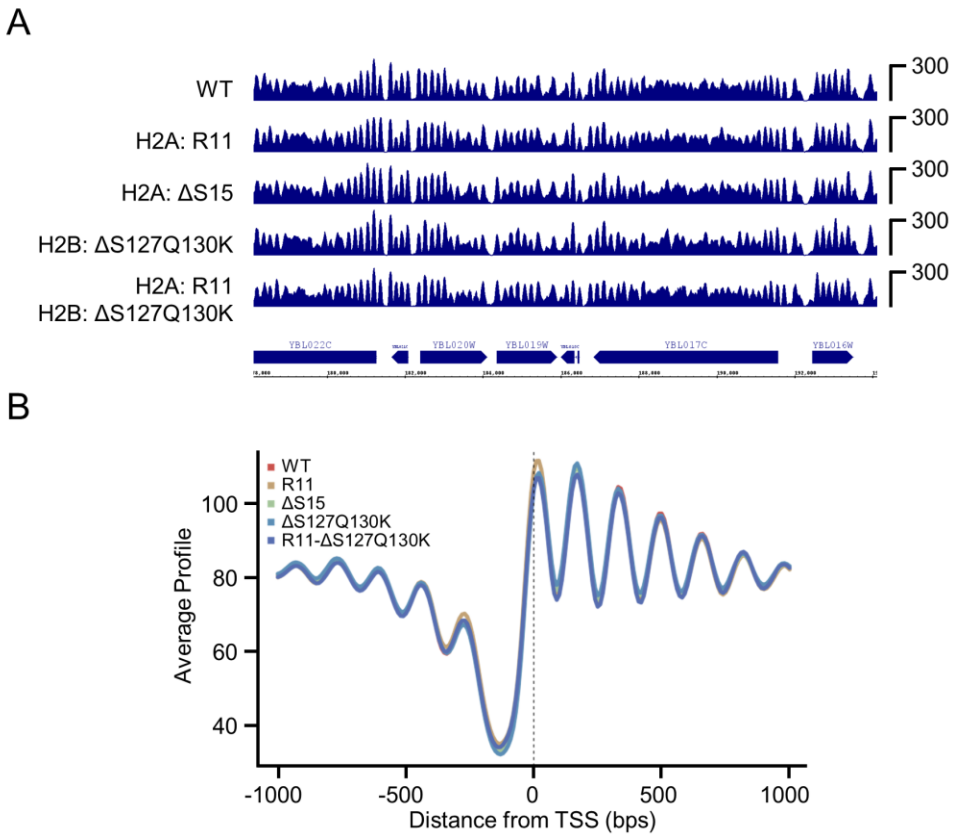

Figure 3-11 – Mutations within the ABC Domain do not affect nucleosomal occupancy. (A) Genome browser view of nucleosomal occupancy along a portion of Chr II in the indicated strains from deep sequencing after MNase digestion. **(B)** Average profile of nucleosome occupancy around open reading frames in the indicated strains.

3.3.6.2 H2B CTD mutations do not affect gene expression

Addition of arginines to the H2A NTD in yeast did not alter the gene expression patterns of exponentially growing cells, although it caused a decrease in fitness (Figure 2-17A,D). Thus it was unlikely that evolutionary changes to the H2B CTD in yeast would shift gene expression patterns. To verify this is the case, we performed mRNA extraction with deep sequencing (mRNA-Seq) in several H2B CTD mutants to quantify the number of transcripts being produced from each gene. Analysis revealed that the gene expression pattern within all the mutants was very similar to the isogenic WT strain with correlations between all mutants >0.995 (Figure 3-12A). The two H2B CTD mutants that displayed to largest difference compared to WT,

ΔS127Q130K and Q130KΔA131, only had a few genes that were more than two fold differentially expressed in either direction with no specific gene ontology. Overall, individual RPKM values were remarkably similar between all strains (Figure 3-12B), indicating that mutations to the H2B CTD do not affect gene expression.

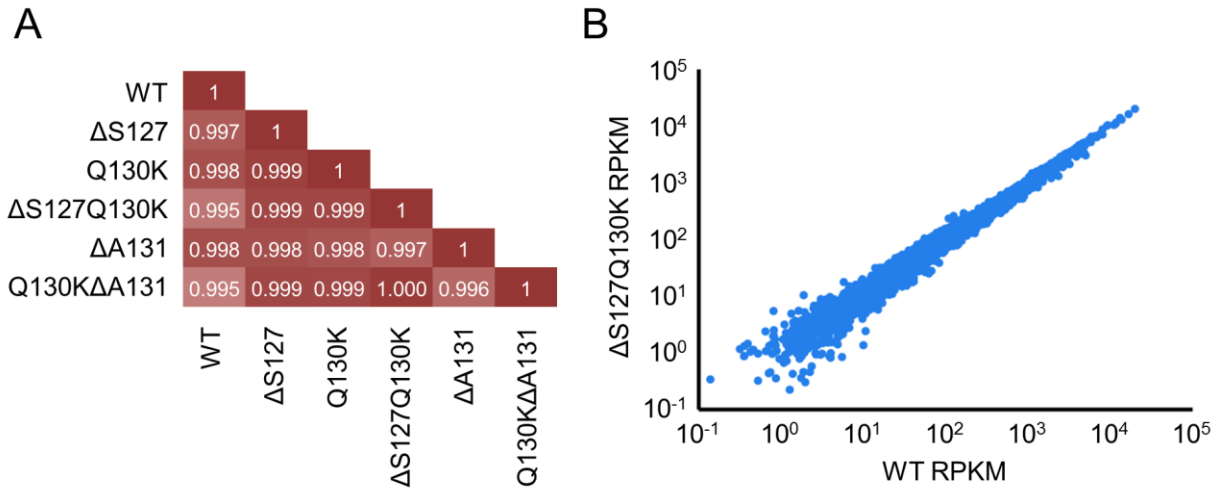

Figure 3-12 – The H2B CTD does not affect gene expression in yeast. (A) Correlations of all RPKM values in the genome between the indicated strains. **(B)** Scatter plot of the RPKM values for individual genes between WT and ΔS127Q130K.

3.3.6.3 Summary

The ABC Domain does not affect nucleosomal occupancy or spacing within yeast strains (see Discussion below on reasons why this may be so). In addition, this domain also does not affect gene expression. This intriguing finding points to a potential mechanism of evolutionary compaction that will not disrupt DNA based processes such as transcription and replication.

3.4 Discussion

In Chapter 3, we have identified a region of the nucleosome, termed the ABC Domain, which has evolved with increasing genome size and acts cooperatively to aid in genome compaction. The ABC Domain is comprised of regions of the H2A NTD and the H2B CTD that partially parallel each other along the nucleosomal surface opposite of the DNA entry/exit. The residues responsible for chromatin compaction in the ABC Domain are located physically close to each other and form a pocket of positive charge. When simultaneously present, these residues increase the positive electrostatic potential of the nucleosome surface and enhance the compaction effects of independent residues alone, indicating that they work collectively.

There are several parallels between the evolutionary changes observed in the H2B CTD and the H2A NTD. For instance, addition of positively charged residues such as arginines and lysines accumulate with increasing genome size while polar residues are lost. However, as was noted in Section 3.3.2.3, residues in the H2A NTD not only have stronger evolutionary trends, but also stronger effects on chromatin compaction. It is not surprising that, in yeast, H2B Q130K had modest effects on compaction by itself. The change from a glutamine to a lysine occurs in the transition from species with extra-small sized genomes to small sized genomes, and thus, if it evolved to aid in genome compaction, there may not have been a need for large compaction effects when genome sizes were still small. Like H2A R11, which had a greater effect on compaction when combined with removal of S15 (R11ΔS15), Q130K also has a greater effect on compaction when combined with removal of S127 (ΔS127Q130K) (see Figure 3-4). Because the positioning of R11 relative to the histone fold was important, ΔS127Q130K positions K130 in yeast to be the same distance from the H2B histone fold as is seen in higher eukaryotes.

The effect on genome compaction mediated by K20R underscores how the ABC Domain acts cooperatively. In both the TSY107 strain background and the FY406 strain background, mutation of K20R, by itself, had very little effects on compaction as measured by FISH. However, when combined with other H2A NTD arginine mutations (R11K20R and R11ΔS15K20R), it enhanced chromatin compaction mediated by those residues.

Our studies on nuclear volume demonstrate that H2A R11 decreases nuclear size both in yeast and in human cells (removal of R11 in human cells results in larger nuclear areas – see Figure 2-15). In yeast, this effect is independent of the level of chromatin compaction as mutants that have greater levels of compaction, such as R11ΔS15, do not show smaller nuclear volumes than R11. Therefore it seems as if nuclear volume control is decoupled from compaction. The ratio of cytoplasmic volume to nuclear volume (N:C ratio) has been found to be relatively constant across eukaryotes[175]. Several factors have been shown to regulate nuclear size in various organisms, such as import of cytoplasmic factors in *Xenopus*[216]. In fission yeast, the amount of DNA did not strongly affect nuclear size, as ploidy up to 16 had little difference in nuclear volume[203]. In contrast, the size of the nucleus in budding yeast expanded throughout the cell cycle, including during S phase when the DNA was replicated[176].

The mechanism governing the increase of cell/nuclear size with increasing genome size, and thus the appropriate nuclear size for a given organism, has been of much debate. Currently there are two theories that argue for an optimal relationship between the amount of DNA and the size of the nucleus such that the DNA molecules themselves have a function beyond coding for proteins. In the nucleoskeletal theory, proposed by Cavalier-Smith[175,217], the DNA plays a structural role in shaping the size of the nucleus. But in this theory, the driving force behind the regulation of

nuclear size is the cell size, whereby larger cells would require a larger nucleus (presumably to make more RNA or to hold more proteins), and the larger nucleus then would require more DNA. The other theory, called the nucleotypic theory, argues that increases to DNA content are propagated through to increases in nuclear and cellular size[218,219]. More DNA requires longer times for replication and may crowd the nucleus to prevent factors, such as cyclins and CDKs, from finding their targets, ultimately slowing down the cell and allowing it to grow more. However, what is common to both of these theories is that the DNA structure and architecture influence nuclear size. Our finding that H2A R11 alone regulates nuclear size would be more compatible with the nucleoskeletal model. For instance, in yeast, chromosomes are known to attach to the nuclear envelope, and if the nucleoskeletal model is correct, would exert forces upon the nucleus to expand its volume[201]. One possibility would be that the presence of H2A R11 weakens interactions with the nuclear envelope, preventing the chromosomes from exerting the usual force, and the nucleus shrinks to its preferred geometry as a result. This hypothesis could explain why the nuclear volume is decreased but equal in all mutants that harbor H2A R11. Obviously this is still an open question and effects of R11 on nuclear envelope attachment would need to be determined experimentally.

We found that the ABC Domain, while affecting genome compaction, does not influence nucleosomal positioning in yeast. This means that compaction mediated by the ABC Domain is not due to changes in linker length. This result is not surprising because the position of nucleosomes in yeast has been shown to be due to several intrinsic factors. Studies out of the Widom group demonstrated the underlying DNA sequence determines much of the occupancy of nucleosomes[220]. In contrast, Struhl and colleagues argue that other processes, such as chromatin remodelers and RNA polymerase, position nucleosomes throughout the genome[221,222]. In their

studies, they created *S. cerevisiae* strains that harbored DNA from other yeast species that intrinsically have different NRLs. They then performed MNase-Seq and found that nucleosomes were positioned on the foreign DNA with NRLs that resembled *S. cerevisiae* and not the yeast strain from which they originated. This indicated that factors within *S. cerevisiae* positioned nucleosomes accordingly. Because we changed neither the underlying DNA sequence nor any chromatin remodelers or transcription factors, the lack of difference in the occupancy of nucleosomes between WT and any ABC Domain mutant is expected.

The most glaring question remaining from these studies is the phenotypic impact of chromatin compaction mediated through the ABC Domain. Because H2A R11 causes both chromatin compaction and decreases in nuclear volume, impacts of compaction may be compensated by the smaller nuclear size. However, some H2B CTD mutants, such as ΔS127Q130K, increase chromatin compaction without affecting nuclear volume and may be good candidates in which to probe phenotypic effects. In addition, any lack of a strong phenotype thus far may be a product of our model system. Yeast are single-celled organisms that are able to adapt to stresses and environmental changes easily. In fact, deletions of the entire NTD of histones have little impact on their ability to grow[181,182]. It is only when two or more NTDs are deleted simultaneously that strong phenotypic effects are observed. Therefore we are developing screening tools to probe for phenotypic effects in the context of increased chromatin compaction as many of our initial studies showed no differences between compacted strains and WT.

CHAPTER 4

Experimental procedures

4.1 Strains and media

The yeast strains used in this study are listed in Table 5-1. Yeast cells were grown in YPD at 30°C unless otherwise noted. C-terminal tagging of yeast proteins was performed as described previously[223]. Mammalian cell lines were maintained at 37°C and 5% CO_2 and cultured with 10% fetal bovine serum (FBS) and DMEM (Gibco).

4.2 Histone sequence database construction and analysis

Sequences were initially extracted from the Entrez database using a keyword search for "histone", and removing non-histone sequences by using keyword searches such as "histone-like", "ubiquitin", and "acetyl", yielding 54,646 results. Blast 2.0[224] was used to align the sequences against the highly conserved histone fold region of the four core histones from *H. sapiens*. Thresholds for true hits were set at >35% identity match and >90% overlap match with the histone fold globular domain region. All duplicate sequences were removed, and further sequence comparisons were made for histone H3 and H2A sequences to filter variants within them. The canonical sequence datasets comprised 672 sequences for histone H3, 357 sequences for histone H4, 518 sequences for histone H2B, and 435 sequences for histone H2A. To further select one canonical sequence for a species among isotypes and variants when annotation was missing, the sequences were compared to the canonical *H. sapiens* and *S. cerevisiae* sequence, and the sequence with highest similarity was selected. Using only completely sequenced species, the final histone sequence dataset included canonical sequences for 160 species from plants, fungi, protozoa, and animals, with genome sizes ranging from 8 to 5600 Mbp.

Sequences for the four core histones were subsequently split into the N-terminal tail, globular domain, and C-terminal tail (in the case of H2A and H2B) sub-sequences. For discovery of

patterns of residue changes according to genome size, each of the sub-sequences was further sub-grouped into small (<100 Mbp), medium (100-1000 Mbp), and large (>1000 Mbp) genome sizes. The frequency of the amino acid residues in each sequence in the sub-groups was determined, and a p-value for the comparison between sub groups was obtained using a Mann-Whitney U Test. Multiple sequence alignment profiles were created using the Muscle sequence comparison tool from Embl-EBI[225,226]. Weblogo3[227,228] was used for motif discovery. Heat maps for residue positions were constructed using Cluster 3.0[229] and Java Treeview[230].

4.3 Yeast H2A and H2B mutagenesis

Site directed mutagenesis was performed using the QuickChange Lightning kit (Agilent) on the pFL142 plasmid (H2A-only) or pJH55 (H2A/H2B). The correct mutation was verified by sequencing.

4.4 Measurement of yeast nuclear volume

Yeast strains were generated that contained a C-terminally tagged Nup49p-GFP fusion. Cells were grown in rich medium to 0.6-0.8x10^7 cells/mL, fixed in growth medium with 4% paraformaldehyde for 15 mins at room temperature, washed twice in PBS, and mounted on a poly-L-lysine-coated slide with mounting medium (Vector Labs). Z-stacks were obtained as described in the microscopy imaging section, and GFP excitation was achieved at 488 nm. Resulting z-stack images were de-convolved using a constrained iterative algorithm from SlideBook 5.5 software and nuclear volumes were measured by masking the inside of each nucleus, which were delineated by the GFP signal. The resulting mask was used to calculate volumes through the SlideBook software. Statistical analysis was performed using the Student's t-test.

4.5 Measurement of yeast cellular volume

Yeast strains were grown in rich medium to 0.6-0.8x10^7 cells/mL, fixed in growth medium with 4% paraformaldehyde for 15 mins at room temperature, washed twice in PBS, and stained with a 1:50 dilution of concanavalin A conjugated with tetramethylrhodamine (Invitrogen) for 15 mins at room temperature. Cells were washed twice in PBS, once in water, and mounted on a poly-L-lysine-coated slide with mounting medium. Z-stacks were obtained as described in the microscopy imaging with mRFP excitation. Cell volume was measured by masking the inside of the RFP signal as described in measurement of yeast nuclear volume.

4.6 FISH probe preparation

For yeast FISH analysis, DNA templates for probes 1, 3, and 4 came from ATCC cosmids 71042, 70912, and 70982 as described elsewhere[197]. DNA templates for Probe 2 were obtained by PCR amplification of a 10kb region starting at coordinate 364647 of chromosome 16 using three primer pairs (Probe2_P1, Probe2_P2, Probe2_P3, Table S16). All DNA templates were digested to smaller fragments using Sau3a (NEB). Fragments were directly labeled using BioPrime labeling kit (Invitrogen) with either ChromaTide Alexa Fluor 488-5-dUTP or ChromaTide Alexa Fluor 568-5-dUTP (Invitrogen).

For human cell FISH analysis, DNA templates for probes came from BACS RP11-252L24 and RP11-195J4 spaced 0.488 Mb apart on chromosome 1. Each BAC was digested into smaller fragments using Sau3a and fragments were directly labeled using BioPrime labeling kit with either ChromaTide Alexa Fluor 488-5-dUTP or ChromaTide Alexa Fluor 568-5-dUTP, as described above.

4.7 Fluorescence *in situ* hybridization analysis in yeast

Yeast strains were grown in rich medium to 0.6-0.8x10^7 cells/mL and fixed in growth medium with 4% paraformaldehyde for 15 mins at room temperature. Cells were then washed twice in growth medium and re-suspended in 2 mL of EDTA-KOH (0.1 M, pH 8.0) and 10 mM DTT and incubated for 10 mins shaking at 30°C. Cells were spun down and re-suspended in 2 mL of YPD + 1.2 M sorbitol with 50 µg/mL of Zymolyase 100-T (Sunrise Science) and 400 U/mL of lyticase (Sigma) and incubated at 30°C for 16 mins with shaking. Spheroblasts were then washed twice in YPD + 1.2 M sorbitol and transferred to a poly-L-lysine-coated slide. After settling for 5 mins, excess liquid was aspirated away and slides were allowed to air dry for 5 additional mins. Slides were washed in methanol for 10 mins and then acetone for 30 secs before air drying. Cells were then dehydrated in a series of cold ethanol washes (70%, 80%, 90%, 100%, 1 min each) and allowed to air dry. Denaturing solution (70% deionized formamide, 2x SSC) was added to the slide and cells were denatured at 75°C for 7-10 mins. Slides were immediately put through another cold ethanol dehydration series and allowed to air dry. Hybridization solution (50% deionized formamide, 2x SSC, 10% dextran sulfate, 100 ng/µL salmon sperm DNA) containing fresh probes was added to the slide and the probes were hybridized for 40-48 hours at 37°C. Slides were then washed in two 5 min washes in 0.05x SSC at 48°C and washed twice in BT Buffer (0.15 M NaHCO3 pH 7.5, 0.1% Tween) for 5 mins at room temperature. Mounting medium containing DAPI was added to the slides and a coverslip was sealed with nail polish.

Inter-probe distances were measured in single projections as described elsewhere[168] by finding the pixel distance between weighted centers of the GFP signal and mRFP signal and converted to nm by the appropriate factor.

4.8 Microscopy imaging

A 3i Marianas SDC confocal microscope equipped with a Zeiss AxioObserver Z1 with a 100×/1.45 NA objective and Yokogawa CSU-22 confocal head was used. Images were captured by a Hamamatsu EMCCD C9100-13 camera controlled by Slidebook 5.0. DAPI, GFP, and mRFP images were acquired by excitation at 360 nm, 488 nm, and 561 nm from a high-speed AOTF laser launch line. A step size of 0.3 μm was used for z-stack acquisition.

4.9 *Micrococcal* nuclease digestion and deep sequencing

Micrococcal nuclease (MNase) digestions were performed on exponentially growing yeast cells as described previously, except that the enzyme was obtained from Sigma-Aldrich (Sigma-Aldrich)[231]. For samples run on an agarose gel, increasing concentrations of MNase were used. For MNase-Seq, a concentration of MNase was used that gave approximately 80% mononucleosomes. The mononucleosomal DNA was gel extracted and libraries were prepared using NuGen Ovation Rapid DR Multiplex kits. Libraries were sequenced using an Illumina HIseq-2000 to obtain 50 bp paired-end reads. The reads were aligned back to the SacCer3 genome using Bowtie2[232] and nucleosome occupancy was calculated using DANPOS[233].

4.10 RNA expression analysis

RNA was extracted from exponentially growing yeast as described previously[234]. PolyA-RNA was prepared, labeled and hybridized to Affymetrix Gene ChIP Yeast Genome 2.0 array by the UCLA clinical microarray core facility and data normalized according to manufacturer's indications. The data are accessible at Gene Expression Omnibus with accession number GSE50440. For RNA-Seq, libraries were prepared and analyzed as previously described[235].

4.11 DNA template and histone preparation for *in vitro* studies

A plasmid containing 12 tandem 177 bp repeats of the high affinity 601 sequence was obtained from Craig L. Peterson's laboratory[57]. DNA arrays were prepared as described previously[236]. After excision with EcoRV, the arrays were gel purified. QuikChange Lightning Site-Directed Mutagenesis (Agilent) was used to create H2A ΔR11 using primers as listed in Table S16. Recombinant *X. laevis* histones were expressed in bacteria and purified. Equimolar amounts of all histones were co-folded to form octamers. Intact octamers were purified from aggregates and free H2A-H2B dimers using Pharmacia Superdex 200 gel filtration column.

4.12 Nucleosome assembly

Recombinant histone octamers and the 601-177-12 DNA template[237] were combined in stoichiometric amounts where 1.0 equivalent of histone octamers and 1.0 equivalents of DNA template were mixed in 2.0 M NaCl. Nucleosome arrays were assembled by step-wise salt dialysis in decreasing NaCl concentration: 1.6 M, 1.2 M, 1.0 M, 0.6 M, 0.4 M, 0.1 M and 0.025 M (in 10 mM Tris pH 8.0, 0.25 mM EDTA), followed by exchanges with 2.5 mM NaCl and 10 mM Tris pH 8.0 without EDTA. Each dialysis step was performed at 4°C for 4 hr to overnight. Partially assembled chromatin was eliminated by precipitation in 4.0 mM $MgCl_2$[56]. The extent of array saturation was assessed by ScaI digestion (200 ng total DNA/chromatin, 3 units ScaI, 50 mM NaCl, 50 mM Tris pH 7.4, 0.5 mM $MgCl_2$), performed for 16 hrs at room temperature followed by 1hr at 37°C, and subsequent analysis using a 5% native polyacrylamide gel.

4.13 Analytical ultracentrifugation

Nucleosome arrays were allowed to equilibrate at room temperature in buffer (2.5 mM NaCl, 10 mM Tris-HCl pH 8.0) containing either 0.1 mM EDTA or 0.6 and 0.8 mM $MgCl_2$. Samples

were centrifuged at 20,000 RPM on a Beckman Optima XL-I analytical ultracentrifuge using an An60 Ti rotor after a 1 hr equilibration at 20°C under vacuum. Time-dependent sedimentation was monitored at 260 nm. Boundaries were analyzed by the method of van Holde and Weichet[204,238].

4.14 Combined immunofluorescence and fluorescence *in-situ* hybridization in human cells

WT or mutant H2A histones were cloned by PCR into prokaryotic expression vector pCMV-HA (Clontech) between EcoRI and NotI sites. Human cells (HEK293, IMR90 and MDA-MB-453) were grown on glass coverslips in 24-well plates in DMEM containing 10% fetal bovine serum and transfected with the indicated HA-H2A expression plasmids using BioT transfection reagent (Bioland Scientific). Cells were grown for 48 hrs post-transfection. Transfected cells were fixed with ice-cold methanol for 15 min at -20°C followed by washing with PBS-T. Cells were then blocked in 5% BSA and incubated with anti-HA antibody (1:250 dilution, Abcam, ab9110). Cells were washed and incubated with secondary antibody (1:500 Alexa Fluor 488 goat anti-rabbit, Invitrogen, A11008 or 1:250 Alexa Fluor 647 goat anti-rabbit, Invitrogen, A21245).

For immunofluorescence, cells were washed and then incubated with Hoechst stain (0.001 mg/mL in PBS). After final washes, cover slips were mounted and imaged. Fluorescence was visualized as above except with the use of 63X magnification. Hoechst nuclear stain, in HA-H2A-expressing cells, was used to measure lengths of the long and short nuclear axes. Estimated nuclear cross-sectional area was calculated using the following formula: Area = $(D_1/2)*(D_2/2)*\pi$, where D_1 and D_2 are long and short axis lengths, respectively.

For FISH, cells were washed, following secondary antibody incubation, in CSK buffer (100 mM NaCl, 300 mM sucrose, 3mM $MgCl_2$, 10 mM PIPES pH 6.8) and permeabilized in CSKT buffer

(CSK+0.5% Triton X-100) before being fixed for 10 mins in 4% paraformaldehyde in PBS at room temperature. Cells were immediately put through a cold ethanol dehydration series (5 mins each at 85%, 95%, and 100%) and allowed to air dry. Cells were rehydrated in 2x SSC for 5 mins and then RNase-treated for 30 mins at 37°C in a humid chamber. Cells were washed with 2x SSC, and denatured at 80°C for 15-20 mins with 70% deionized formamide and 2x SSC. They were immediately cooled with cold 2x SSC, and put through another cold ethanol dehydration series. Probes were added to cells and allowed to hybridize for 48 hrs. After hybridization, cells were washed with 50% formamide in 2x SSC, 2x SSC, and 1x SSC containing DAPI. Slides were mounted and imaged as described above.

4.15 Competition assays

Two sets of yeast strains were generated in which Pgk1p was C-terminally fused with either GFP or RFP. GFP-labeled WT H2A strains were co-cultured with RFP-labeled mutant H2A strains at a 1:1 ratio and at an optical density of ~0.4. Corresponding co-cultures with switched fluorescent labels were also made. Cultures were incubated at 30°C for 72 hours and were diluted every 6-12 hrs to maintain cells in exponential growth phase. Samples were collected every 12 hrs for analysis by flow cytometry. Collected cells were fixed in 70% ethanol, washed and re-suspended in 50 mM sodium citrate, pH 7.0, and mildly sonicated to disrupt aggregates. GFP- and RFP- labeled cells were counted using a Becton Dickinson FACScan cytometer, and the proportion of each in the population was calculated.

4.16 Cell cycle analysis

Cell cycle analysis of exponentially growing cells was performed essentially as described previously[239], except that cells were stained with 1 μM SYTOX Green (Molecular Probes).

4.17 Spot tests

Approximately $1.0x10^7$ exponentially growing yeast cells were collected and re-suspended in 100 µl of H_2O and 10-fold serially diluted. Subsequently, 5 µl were spotted on agar plates containing media and drugs as indicated in the figures and incubated at 30^oC for 2-6 days.

4.18 Nucleosome surface electrostatic calculations

PyMOL[214] was used to make changes to the 1KX5 structure to mimic the residues found in small genome organisms[35]. The follow mutations occurred in H2A: R20K, R11A, R3A, Q6A, K15Q, A14S, K13A, and A12K. H2B K126Q was mutated. H2A residues 1-23 and H2B residues 100-126 were extracted into PDB files and a PQR file was generated[215]. The resulting structure was then analyzed by the APBS toolkit for electrostatic potentials and projected onto the structure. 1KX5 DNA was overlaid.

CHAPTER 5

Tables

Table 5-1 – Yeast strains used in this study

Name	Mutant name	Description	Reference
FLY142		*Matα, hta1-1, hta2-1, ura3-52, his3*, pFL142-*HIS3*	Lefant et al., 1996
TSY107	Parental	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pJC102-*URA3*	Schuster et al., 1986
FY406	Parental	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pSAB6*	Hirschhorn et al., 1995
AOY001	WT	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pFL142-*HIS3*	this study
AOY002	R3	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pR3-*HIS3*	this study
AOY004	R11	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pR11-*HIS3*	this study
AOY005	K3	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pK3-*HIS3*	this study
AOY009	K20R	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pK20R-*HIS3*	this study
AOY011	R17K	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pR17K-*HIS3*	this study
AOY013	K11	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pK11-*HIS3*	this study
AOY014	K3K11	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pK3K11-*HIS3*	this study
AOY015	R3R11	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pR3R11-*HIS3*	this study
AOY020	R6	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pR6-*HIS3*	this study
AOY022	ΔGS10	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pΔGS10-*HIS3*	this study
AOY023	R3(ΔGS10) R11	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pR3Δ(GS)R11-*HIS3*	this study
AOY024	ΔS15	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pΔS15-*HIS3*	this study
AOY025	R11ΔS15	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pR11ΔS15-*HIS3*	this study
AOY029	K11ΔS15	*Mat*a, *hta1-1, hta2-1, ura3-52, his3*, pK11ΔS15-*HIS3*	this study
BBY011	WT	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pJH55*	this study
BBY013	ΔS15	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔS15*	this study
BBY022	R11	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11*	this study
BBY023	R11ΔS15	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11ΔS15*	this study
BBY001	ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔS127Q130K*	this study
BBY003	Q130KΔA131	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pQ130KΔA131*	this study
BBY005	ΔS127	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔS127*	this study
BBY007	Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pQ130K*	this study
BBY009	ΔA131	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔA131*	this study
BBY024	R11-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11Q130K*	this study
BBY025	ΔS15-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔS15Q130K*	this study
BBY026	R11ΔS15-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11ΔS15Q130K*	this study
BBY030	K11	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pK11*	this study
BBY031	K11-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pK11Q130K*	this study
BBY033	R11-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11*ΔS15Q130K	this study
BBY034	ΔS15-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔS15ΔS127Q130K*	this study
BBY035	R11ΔS15-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11ΔS15ΔS127Q130K*	this study
BBY036	K11-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pK11ΔS127Q130K*	this study

BBY040	I44V	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pI44V*	this study
BBY041	G45A	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pG45A*	this study
BBY042	S46A	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pS46A*	this study
BBY043	K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pK20R*	this study
BBY044	R11K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11K20R*	this study
BBY045	ΔS15K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔS15K20R*	this study
BBY046	R11ΔS15 K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11ΔS15K20R*	this study
BBY048	R11K20R-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11K20RΔS127Q130K*	this study
BBY049	ΔS15K20R-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pΔS15K20RΔS127Q130K*	this study
BBY050	R11ΔS15 K20R-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pR11ΔS15K20RΔS127Q130K*	this study
BBY054	K11ΔS15	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pK11ΔS15*	this study
BBY055	I44V-S46A	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pI44VS46A*	this study
BBY056	I44V-G45A-S46A	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, pI44VG45AS46A*	this study
BMY003	Parental	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pJC102-*URA3*	this study
BMY004	WT	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pFL142-*HIS3*	this study
BMY005	R3	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pR3-*HIS3*	this study
BMY007	R11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pR11-*HIS3*	this study
BMY008	K3	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pK3-*HIS3*	this study
BMY012	K20R	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pK20R-*HIS3*	this study
BMY014	R17K	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pR17K-*HIS3*	this study
BMY016	K11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pK11-*HIS3*	this study
BMY017	K3K11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pK3K11-*HIS3*	this study
BMY018	R3R11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pR3R11-*HIS3*	this study
BMY038	R6	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pR6-*HIS3*	this study
BMY039	ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pΔS15-*HIS3*	this study
BMY040	ΔGS10	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pΔGS10-*HIS3*	this study
BMY041	R3(ΔGS10) R11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pR3ΔGS10R11-*HIS3*	this study
BMY043	R11ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pR11ΔS15-*HIS3*	this study
BMY045	K11ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Nup49-GFP(kanMX6)*, pK11ΔS15-*HIS3*	this study
BMY501	WT	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-GFP(kanMX6)*, pFL142-*HIS3*	this study
BMY502	R11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-GFP(kanMX6)*, pR11-*HIS3*	this study
BMY503	ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-GFP(kanMX6)*, pΔS15-*HIS3*	this study
BMY504	R11ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-GFP(kanMX6)*, pR11ΔS15-*HIS3*	this study
BMY505	K11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-GFP(kanMX6)*, pK11-*HIS3*	this study
BMY509	K11ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-GFP(kanMX6)*, pK11ΔS15-*HIS3*	this study
BMY511	WT	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-RFP(kanMX6)*, pFL142-*HIS3*	this study
BMY512	R11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-RFP(kanMX6)*, pR11-*HIS3*	this study
BMY513	ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-RFP(kanMX6)*, pΔS15-*HIS3*	this study
BMY514	R11ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-RFP(kanMX6)*, pR11ΔS15-*HIS3*	this study

BMY515	K11	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-RFP(kanMX6)*, pK11-*HIS3*	this study
BMY519	K11ΔS15	*Mat*a, *hta1-1*, *hta2-1*, *ura3-52*, *his3*, *Pgk1-RFP(kanMX6)*, pK11ΔS15-*HIS3*	this study
BMY1000	Parental	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pSAB6*	
BMY1001	WT	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pJH55*	this study
BMY1021	ΔS15	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pΔS15*	this study
BMY1020	R11	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11*	this study
BMY1022	R11ΔS15	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11ΔS15*	this study
BMY1004	ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pΔS127Q130K*	this study
BMY1006	Q130KΔA131	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pQ130KΔA131*	this study
BMY1002	ΔS127	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pΔS127*	this study
BMY1003	Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pQ130K*	this study
BMY1005	ΔA131	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pΔA131*	this study
BMY1030	R11-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11Q130K*	this study
BMY1031	ΔS15-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pΔS15Q130K*	this study
BMY1032	R11ΔS15-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11ΔS15Q130K*	this study
BMY1023	K11	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pK11*	this study
BMY1027	K11-Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pK11Q130K*	this study
BMY1033	R11-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11*ΔS15Q130K	this study
BMY1034	ΔS15-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pΔS15ΔS127Q130K*	this study
BMY1035	R11ΔS15-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11ΔS15ΔS127Q130K*	this study
BMY1036	K11-ΔS127Q130K	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pK11ΔS127Q130K*	this study
BMY1037	K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pK20R*	this study
BMY1038	R11K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11K20R*	this study
BMY1039	ΔS15K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pΔS15K20R*	this study
BMY1040	R11ΔS15 K20R	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pR11ΔS15K20R*	this study
BMY1045	K11ΔS15	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pK11ΔS15*	this study
BMY1046	C11	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pC11*	this study
BMY1047	P11	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pP11*	this study
BMY1048	H11	*Mat*a, *(hta1-htb1)Δ::LEU2, (hta2-htb2)Δ::TRP1, Nup49-GFP(kanMX6), pH11*	this study

Table 5-2 – Summary of FISH data in the TSY107 background

Yeast FISH - TSY107 Background – Probe Set A							
Strain	nm				% change	p-value	No. cells
	Minimum	Maximum	Mean	Median			
WT	364	1487	757	714	0	1.0E+00	63
R3	178	1377	619	584	-18	9.5E-04	90
R11	22	1101	584	605	-23	8.6E-04	49
R3R11	79	1102	557	553	-26	8.2E-06	72
R3(ΔGS10)R11	107	1410	516	497	-32	2.1E-06	51
R11ΔS15	67	1378	455	415	-40	3.9E-08	57
K3	272	1378	749	772	-1	8.6E-01	64
K11	236	1421	808	820	7	3.1E-01	56
K3K11	164	1430	747	751	-1	8.3E-01	71
K11ΔS15	124	1355	665	659	-12	9.2E-02	42
ΔGS10	66	1084	649	666	-14	3.2E-02	60
ΔS15	30	1236	668	726	-12	9.4E-02	49
R6	106	1410	674	672	-11	5.6E-02	73
K20R	205	1539	708	686	-6	3.0E-01	64
R17K	230	1598	743	699	-2	7.9E-01	54
Yeast FISH - TSY107 Background – Probe Set B							
WT	259	1372	626	599	0	1.0E+00	59
R11	64	1231	514	476	-18	1.4E-02	60
ΔS15	88	1452	586	549	-6	2.8E-01	53
R11ΔS15	66	934	446	433	-29	2.3E-05	51
Yeast FISH - TSY107 Background – Probe Set C							
WT	91	1192	493	485	0	1.0E+00	75
R11	74	1481	399	349	-19	1.4E-03	50
ΔS15	153	896	456	437	-8	1.9E-01	55
R11ΔS15	89	925	398	410	-19	5.4E-03	48
Yeast FISH - TSY107 Background – Probe Set D							
WT	65	1029	361	342	0	1.0E+00	85
R11	41	802	299	261	-17	4.8E-05	70
ΔS15	33	648	323	324	-11	8.7E-02	51
R11ΔS15	16	748	273	237	-24	1.6E-07	62

Table 5-3 – Summary of FISH data for Probe Set A in yeast strains with the FY406 background

Yeast FISH - FY406 Background - Probe Set A								
Strain	nm				% change			
	Min	Max	Mean	Median	Mean	Median	p-value	No. cells
WT	172	1274	721	718	0	0	1.0E+00	79
ΔS127	67	1317	709	746	-3	3	7.0E-01	56
Q130K	132	1256	641	634	-12	-12	6.6E-02	62
ΔS127Q130K	70	1401	601	545	-17	-25	1.7E-02	54
ΔA131	91	1557	678	663	-7	-8	3.0E-01	72
Q130KΔA131	109	1764	619	635	-15	-12	1.3E-02	93
R11	56	1158	609	617	-16	-15	7.5E-03	86
ΔS15	165	1523	742	770	2	7	7.8E-01	62
R11ΔS15	23	1450	581	536	-20	-26	5.1E-03	61
K11	93	1401	642	641	-12	-11	5.7E-02	83
K11ΔS15	128	1279	676	647	-7	-10	2.3E-01	84
R11-Q130K	67	1543	594	581	-18	-20	2.4E-03	82
ΔS15Q130K	118	1486	637	585	-13	-19	5.0E-02	83
R11-ΔS15Q130K	60	1346	566	572	-22	-21	8.7E-04	66
K11-ΔS127Q130K	39	1574	606	591	-17	-18	1.9E-02	73
R11-ΔS127Q130K	58	1014	499	524	-31	-28	6.2E-07	64
ΔS15-ΔS127Q130K	112	1802	598	591	-18	-18	4.9E-03	73
R11ΔS15-ΔS127Q130K	50	1070	478	491	-34	-32	8.2E-09	86
K20R	86	1850	700	680	-4	-6	5.2E-01	91
R11K20R	76	1045	553	537	-24	-26	1.2E-04	75
ΔS15K20R	110	1891	685	670	-6	-7	3.9E-01	76
R11ΔS15K20R	70	1316	542	467	-26	-35	1.0E-04	71
R11K20R-ΔS127Q130K	83	1325	566	570	-22	-21	1.2E-04	91
ΔS15K20R-dS127Q130K	80	1383	633	643	-13	-11	5.0E-02	71
R11ΔS15K20R-ΔS127Q130K	68	1378	541	517	-26	-28	2.7E-05	95
I44VS46A	103	1436	692	680	-5	-6	4.5E-01	69
I44VG45AS46A	143	1297	664	690	-9	-5	1.8E-01	68
G45A	39	1399	662	630	-9	-13	1.6E-01	72
I44V	132	1958	723	671	-1	-7	9.1E-01	89
S46A	56	1464	692	653	-5	-10	4.6E-01	83
C11	107	1915	705	703	-3	-3	6.6E-01	77
H11	126	1298	613	600	-16	-17	5.4E-03	96
P11	89	1579	698	676	-4	-7	5.1E-01	101

Table 5-4 – Summary of nuclear volumes in TSY107 background

Yeast nuclear volume (TSY107 Background)							
Nuclear Vol	μm^3				% change	p-value	No. cells
	Minimum	Maximum	Mean	Median			
WT	0.6	6.9	2.9	2.8	0	1.0E+00	188
R3	0.6	6.5	2.8	2.6	-5	4.1E-01	177
R11	0.7	7.0	2.3	2.1	-20	5.9E-05	201
R3R11	0.4	6.9	2.5	2.3	-16	3.0E-03	180
R3(ΔGS10)R11	0.7	8.1	3.1	3.0	6	3.7E-01	196
R11ΔS15	0.8	6.2	2.8	2.7	0	4.2E-01	150
K3	0.5	9.1	3.3	3.1	13	9.4E-03	181
K11	0.7	10.8	3.2	2.9	3	2.6E-01	191
K3K11	1.0	9.3	3.8	3.6	31	5.4E-08	172
K11ΔS15	0.8	8.9	3.0	2.9	2	6.6E-01	186
ΔGS10	0.8	12.6	3.5	3.1	10	3.0E-02	198
ΔS15	0.8	8.8	3.3	3.0	9	9.7E-03	202
R6	0.8	10.3	3.4	3.1	10	1.0E-03	201
K20R	0.7	9.3	3.3	3.0	7	1.5E-02	199
R17K	0.7	8.3	3.0	2.5	0	2.7E-01	199
Yeast cellular volume (TSY107 Background)							
WT	11.4	92.8	46.1	42.6	0	1.0E+00	178
R11	12.1	155.9	48.2	43.4	4	2.9E-01	199
K11	10.7	114.1	44.2	44.2	-4	5.0E-01	209
ΔS15	13.7	123.5	47.0	47.0	2	5.0E-01	83
R11ΔS15	13.0	99.5	46.0	46.0	0	4.3E-01	104

Table 5-5 – Summary of nuclear volumes in FY406 background

Yeast nuclear volume (FY406 background)								
Strain	μm^3				% change			
	Min	Max	Mean	Median	Mean	Median	p-value	No. cells
WT	1.01	5.09	2.48	2.26	0	0	1.0E+00	268
ΔS127	1.02	8.40	2.93	2.76	10	18	3.5E-02	193
Q130K	1.01	8.57	2.63	2.25	-1	-3	7.7E-01	200
ΔS127Q130K	1.01	7.05	2.50	2.36	-6	1	1.6E-01	202
ΔA131	1.00	7.66	2.58	2.27	-3	-3	5.0E-01	196
Q130KΔA131	1.04	6.02	2.45	2.23	-8	-4	5.8E-02	196
R11	1.01	7.39	2.15	1.93	-19	-17	4.7E-06	195
ΔS15	1.02	8.35	2.56	2.18	-4	-7	3.8E-01	194
R11ΔS15	1.03	6.79	2.15	1.80	-19	-23	8.6E-06	191
K11	1.00	6.04	2.58	2.34	-3	0	5.1E-01	195
K11ΔS15	1.00	12.21	2.66	2.22	0	-5	9.9E-01	196
R11-Q130K	1.00	6.73	2.23	1.94	-16	-17	1.1E-04	212
ΔS15-Q130K	1.06	7.90	2.52	2.32	-5	-1	2.2E-01	200
R11-ΔS15Q130K	1.01	6.19	2.22	1.93	-17	-17	8.2E-05	196
K11-Q130K	1.02	5.86	2.33	2.11	-12	-10	3.4E-03	195
R11-ΔS127Q130K	1.02	5.12	2.10	1.89	-21	-19	2.8E-07	193
ΔS15-ΔS127Q130K	1.02	12.48	2.51	2.17	-6	-7	1.9E-01	210
R11ΔS15-ΔS127Q130K	1.03	4.97	2.16	1.90	-19	-19	2.9E-06	204
K11ΔS127Q130K	1.05	9.35	2.57	2.26	-4	-3	4.3E-01	184
K20R	1.04	7.48	2.62	2.35	-1	1	7.5E-01	192
R11K20R	1.01	8.03	2.28	2.01	-14	-14	1.2E-03	183
ΔS15K20R	1.07	7.91	2.60	2.34	-2	0	5.9E-01	189
R11ΔS15K20R	1.02	6.00	2.13	1.88	-20	-19	2.0E-06	191
R11K20R-ΔS127Q130K	1.01	6.51	2.25	2.03	-16	-13	4.0E-04	182
ΔS15K20R-ΔS127Q130K	1.01	5.95	2.49	2.24	-7	-4	1.3E-01	186
R11ΔS15K20R-ΔS127Q130K	1.00	6.43	2.20	1.81	-17	-22	5.1E-05	194
C11	1.10	7.55	2.52	2.25	-5	-3	2.4E-01	184
P11	1.04	6.22	2.49	2.24	-7	-4	1.1E-01	204
H11	1.05	6.30	2.57	2.34	-3	0	4.5E-01	181

Table 5-6 – Summary of FISH in human cells

Human FISH – IMR90 cells							
IMR90	nm				% change	p-value	No. cells
	Minimum	Maximum	Mean	Median			
WT	35	1079	294	243	0	1.0E+00	47
ΔR3	57	702	363	406	23	6.4E-02	53
R11A	28	1087	430	411	46	3.1E-03	34
ΔR3R11A	70	1161	420	394	43	7.5E-03	54
Human FISH – MDA-MB-453 cells							
WT	33	745	296	293	0	1.0E+00	49
ΔR3	42	740	341	315	15	1.9E-01	45
R11A	97	793	362	338	22	1.7E-02	60
ΔR3R11A	46	764	388	399	31	5.9E-03	35

Human nuclear area - IMR90 cells						
HEK293	μm^2			% change	p-value	No. cells
	Minimum	Maximum	Mean			
WT	58	158	96	0	1.0E+00	52
ΔR3	65	161	106	10	7.0E-03	60
R11A	68	231	116	21	4.4E-05	65
ΔR3R11A	76	251	121	26	4.6E-07	52
Human nuclear area - MDA-MB-453 cells						
WT	57	137	92	0	1.0E+00	36
ΔR3	67	165	103	12	3.0E-02	34
R11A	92	172	132	43	4.8E-11	33
ΔR3R11A	78	154	111	21	3.1E-05	33

REFERENCES

1 Fishman, D. & Patterson, G. Light scattering studies of supercoiled and nicked DNA. *Biopolymers* **38**, 535-552 (1996).

2 Hancock, R. A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus. *Journal of structural biology* **146**, 281-290 (2004).

3 Thaw, P. *et al.* Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic acids research* **34**, 1439-1449 (2006).

4 Bianchi, M. E. & Agresti, A. HMG proteins: dynamic players in gene regulation and differentiation. *Current opinion in genetics & development* **15**, 496-506 (2005).

5 Finkel, S. & Johnson, R. The Fis protein: it's not just for DNA inversion anymore. *Molecular microbiology* **6**, 3257-3265 (1992).

6 Schneider, R. *et al.* An architectural role of the Escherichia coli chromatin protein FIS in organising DNA. *Nucleic acids research* **29**, 5107-5114 (2001).

7 de los Rios, S. & Perona, J. J. Structure of the< i> Escherichia coli</i> Leucine-responsive Regulatory Protein Lrp Reveals a Novel Octameric Assembly. *Journal of molecular biology* **366**, 1589-1602 (2007).

8 Grove, A. Functional evolution of bacterial histone-like HU proteins. *Current issues in molecular biology* **13**, 1 (2011).

9 Zhang, J., Zeuner, Y., Kleefeld, A., Unden, G. & Janshoff, A. Multiple Site-Specific Binding of Fis Protein to Escherichia coli nuoA-N Promoter DNA and its Impact on DNA Topology Visualised by Means of Scanning Force Microscopy. *Chembiochem* **5**, 1286-1289 (2004).

10 Dame, R. T. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular microbiology* **56**, 858-870 (2005).

11 Howard, M., Sandman, K., Reeve, J. & Griffith, J. HMf, a histone-related protein from the hyperthermophilic archaeon Methanothermus fervidus, binds preferentially to DNA containing phased tracts of adenines. *Journal of bacteriology* **174**, 7864-7867 (1992).

12 Starich, M. R., Sandman, K., Reeve, J. N. & Summers, M. F. NMR Structure of HMfB from the Hyperthermophile,< i> Methanothermus fervidus</i>, Confirms that this Archaeal Protein is a Histone. *Journal of molecular biology* **255**, 187-203 (1996).

13 Sandman, K., Krzycki, J., Dobrinski, B., Lurz, R. & Reeve, J. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most

closely related to histones. *Proceedings of the National Academy of Sciences* **87**, 5788-5791 (1990).

14 Reeve, J. *et al.* Archaeal histones: structures, stability and DNA binding. *Biochemical Society Transactions* **32**, 227-230 (2004).

15 Musgrave, D. R., Sandman, K. M. & Reeve, J. N. DNA binding by the archaeal histone HMf results in positive supercoiling. *Proceedings of the National Academy of Sciences* **88**, 10397-10401 (1991).

16 Ammar, R. *et al.* Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* **1** (2012).

17 Sandman, K. & Reeve, J. N. Archaeal histones and the origin of the histone fold. *Current opinion in microbiology* **9**, 520-525 (2006).

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

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

20 Wong, J. T., New, D., Wong, J. & Hung, V. Histone-like proteins of the dinoflagellate Crypthecodinium cohnii have homologies to bacterial DNA-binding proteins. *Eukaryotic cell* **2**, 646-650 (2003).

21 Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-352 (1997).

22 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell research* **21**, 381-395 (2011).

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

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

25 Braunstein, M., Sobel, R. E., Allis, C. D., Turner, B. M. & Broach, J. R. Efficient transcriptional silencing in Saccharomyces cerevisiae requires a heterochromatin histone acetylation pattern. *Molecular and Cellular Biology* **16**, 4349-4356 (1996).

26 Oki, M., Aihara, H. & Ito, T. in *Chromatin and Disease* 323-340 (Springer, 2007).

27 Kim, J. *et al.* Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO reports* **7**, 397-403 (2006).

28 Mujtaba, S., Zeng, L. & Zhou, M. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **26**, 5521-5527 (2007).

29 Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B. J. & Grunstein, M. Histone acetylation regulates the time of replication origin firing. *Molecular cell* **10**, 1223-1233 (2002).

30 Fischle, W. *et al.* Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116-1122 (2005).

31 Durrin, L. K., Mann, R. K., Kayne, P. S. & Grunstein, M. Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. *Cell* **65**, 1023-1031 (1991).

32 Luger, K. & Richmond, T. J. DNA binding within the nucleosome core. *Current opinion in structural biology* **8**, 33-40 (1998).

33 West, S. M., Rohs, R., Mann, R. S. & Honig, B. Electrostatic interactions between arginines and the minor groove in the nucleosome. *Journal of Biomolecular Structure and Dynamics* **27**, 861-866 (2010).

34 Tan, S. & Davey, C. A. Nucleosome structural studies. *Current opinion in structural biology* **21**, 128-136 (2011).

35 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 1.9 Å resolution. *Journal of molecular biology* **319**, 1097-1113 (2002).

36 Olins, A. L. & Olins, D. E. Spheroid chromatin units (ν bodies). *Science* **183**, 330-332 (1974).

37 Woodcock, C., Safer, J. & Stanchfield, J. Structural repeating units in chromatin: I. Evidence for their general occurrence. *Experimental cell research* **97**, 101-110 (1976).

38 Woodcock, C. in *Journal of Cell Biology*. A368-A368 (ROCKEFELLER UNIV PRESS 1114 FIRST AVE, 4TH FL, NEW YORK, NY 10021).

39 Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nature structural & molecular biology* **12**, 46-53 (2005).

40 Mozziconacci, J. & Victor, J.-M. Nucleosome gaping supports a functional structure for the 30nm chromatin fiber. *Journal of structural biology* **143**, 72-76 (2003).

41 Mozziconacci, J., Lavelle, C., Barbi, M., Lesne, A. & Victor, J.-M. A physical model for the condensation and decondensation of eukaryotic chromosomes. *FEBS letters* **580**, 368-372 (2006).

42 Lesne, A. & Victor, J.-M. Chromatin fiber functional organization: some plausible models. *The European Physical Journal E* **19**, 279-290 (2006).

43 Kuryan, B. G. *et al.* Histone density is maintained during transcription mediated by the chromatin remodeler RSC and histone chaperone NAP1 in vitro. *Proceedings of the National Academy of Sciences* **109**, 1931-1936 (2012).

44 Kireeva, M. L. *et al.* Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Molecular cell* **9**, 541-552 (2002).

45 Alilat, M., Sivolob, A., Révet, B. & Prunell, A. Nucleosome dynamics IV. protein and DNA contributions in the chiral transition of the tetrasome, the histone (H3-H4)< sub> 2</sub> tetramer-DNA particle. *Journal of molecular biology* **291**, 815-841 (1999).

46 Bancaud, A. *et al.* Nucleosome chiral transition under positive torsional stress in single chromatin fibers. *Molecular cell* **27**, 135-147 (2007).

47 Szent-Györgyi, C., Finkelstein, D. B. & Garrard, W. T. Sharp boundaries demarcate the chromatin structure of a yeast heat-shock gene. *Journal of molecular biology* **193**, 71-80 (1987).

48 Hansen, J. C., Ausio, J., Stanik, V. H. & Van Holde, K. Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. *Biochemistry* **28**, 9129-9136 (1989).

49 Hansen, J. C. Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annual review of biophysics and biomolecular structure* **31**, 361-392 (2002).

50 Schwarz, P. M., Felthauser, A., Fletcher, T. M. & Hansen, J. C. Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry* **35**, 4009-4015 (1996).

51 Fletcher, T. M. & Hansen, J. C. Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms. *Journal of Biological Chemistry* **270**, 25359-25362 (1995).

52 Garcia-Ramirez, M., Dong, F. & Ausio, J. Role of the histone" tails" in the folding of oligonucleosomes depleted of histone H1. *Journal of Biological Chemistry* **267**, 19587-19595 (1992).

53 Liu, Y. *et al.* Influence of histone tails and H4 tail acetylations on nucleosome–nucleosome interactions. *Journal of molecular biology* **414**, 749-764 (2011).

54 Tse, C. & Hansen, J. C. Hybrid trypsinized nucleosomal arrays: identification of multiple functional roles of the H2A/H2B and H3/H4 N-termini in chromatin fiber compaction. *Biochemistry* **36**, 11381-11388 (1997).

55 Kalashnikova, A. A., Porter-Goff, M. E., Muthurajan, U. M., Luger, K. & Hansen, J. C. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *Journal of The Royal Society Interface* **10**, 20121022 (2013).

56 Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *Journal of molecular biology* **327**, 85-96 (2003).

57 Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847 (2006).

58 Sinha, D. & Shogren-Knaak, M. A. Role of direct interactions between the histone H4 Tail and the H2A core in long range nucleosome contacts. *Journal of Biological Chemistry* **285**, 16572-16581 (2010).

59 Robinson, P. J. *et al.* 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *Journal of molecular biology* **381**, 816-825 (2008).

60 Allahverdi, A. *et al.* The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic acids research* **39**, 1680-1691 (2011).

61 Suka, N., Suka, Y., Carmen, A. A., Wu, J. & Grunstein, M. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Molecular cell* **8**, 473-479 (2001).

62 Braunstein, M., Rose, A., Holmes, S., Allis, C. & Broach, J. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes & Development* **7**, 592-604 (1993).

63 Kurdistani, S. K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nature Reviews Molecular Cell Biology* **4**, 276-284 (2003).

64 Millar, C. B. & Grunstein, M. Genome-wide patterns of histone modifications in yeast. *Nature Reviews Molecular Cell Biology* **7**, 657-666 (2006).

65 Lu, X. *et al.* The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nature structural & molecular biology* **15**, 1122-1124 (2008).

66 Robzyk, K., Recht, J. & Osley, M. A. Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**, 501-504 (2000).

67 West, M. H. & Bonner, W. M. Histone 2B can be modified by the attachment of ubiquitin. *Nucleic acids research* **8**, 4671-4680 (1980).

68 Fierz, B. *et al.* Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nature chemical biology* **7**, 113-119 (2011).

69 Xiao, T. *et al.* Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Molecular and cellular biology* **25**, 637-651 (2005).

70 Minsky, N. *et al.* Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nature cell biology* **10**, 483-488 (2008).

71 Kim, J. *et al.* RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**, 459-471 (2009).

72 Wang, C.-Y. *et al.* The C-terminus of histone H2B is involved in chromatin compaction specifically at telomeres, independently of its monoubiquitylation at lysine 123. *PloS one* **6**, e22209 (2011).

73 Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J. & Maltais, L. J. The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-498 (2002).

74 Hayashi, H., Nomoto, M. & Iwai, K. Tetrahymena histone H4. Complete amino acid sequences of two variants. *Journal of biochemistry* **96**, 1449 (1984).

75 Siegel, T. N. *et al.* Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes & development* **23**, 1063-1076 (2009).

76 Moosmann, A. *et al.* Histone variant innovation in a rapidly evolving chordate lineage. *BMC evolutionary biology* **11**, 208 (2011).

77 Gatewood, J., Cook, G., Balhorn, R., Schmid, C. & Bradbury, E. Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. *Journal of Biological Chemistry* **265**, 20662-20666 (1990).

78 Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478 (2009).

79 Collins, K. A., Castillo, A. R., Tatsutani, S. Y. & Biggins, S. De novo kinetochore assembly requires the centromeric histone H3 variant. *Molecular biology of the cell* **16**, 5649-5660 (2005).

80 Santaguida, S. & Musacchio, A. The life and miracles of kinetochores. *The EMBO journal* **28**, 2511-2531 (2009).

81 Furuyama, T. & Henikoff, S. Centromeric nucleosomes induce positive DNA supercoils. *Cell* **138**, 104-113 (2009).

82 Dalal, Y., Furuyama, T., Vermaak, D. & Henikoff, S. Structure, dynamics, and evolution of centromeric nucleosomes. *Proceedings of the National Academy of Sciences* **104**, 15974-15981 (2007).

83 Dalal, Y., Wang, H., Lindsay, S. & Henikoff, S. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS biology* **5**, e218 (2007).

84 Tachiwana, H. *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232-235 (2011).

85 Chow, C. M. *et al.* Variant histone H3. 3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO reports* **6**, 354-360 (2005).

86 Hake, S. B. & Allis, C. D. Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proceedings of the National Academy of Sciences* **103**, 6428-6435 (2006).

87 Bönisch, C. & Hake, S. B. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic acids research* **40**, 10719-10741 (2012).

88 Lindseyt, G. G., Orgeig, S., Thompson, P., Davies, N. & Maeder, D. L. Extended C-terminal tail of wheat histone H2A interacts with DNA of the "linker" region. *Journal of molecular biology* **218**, 805-813 (1991).

89 West, M. H. & Bonner, W. M. Histone 2A, a heteromorphous family of eight protein species. *Biochemistry* **19**, 3238-3245 (1980).

90 Thatcher, T. H. & Gorovsky, M. A. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic acids research* **22**, 174-179 (1994).

91 Faast, R. *et al.* Histone variant H2A. Z is required for early mammalian development. *Current Biology* **11**, 1183-1187 (2001).

92 van Daal, A. & Elgin, S. A histone variant, H2AvD, is essential in Drosophila melanogaster. *Molecular biology of the cell* **3**, 593-602 (1992).

93 Jackson, J. & Gorovsky, M. Histone H2A. Z has a conserved function that is distinct from that of the major H2A sequence variants. *Nucleic acids research* **28**, 3811-3816 (2000).

94 Carr, A. *et al.* Analysis of a histone H2A variant from fission yeast: evidence for a role in chromosome stability. *Molecular and General Genetics MGG* **245**, 628-635 (1994).

95 Issel-Tarver, L. *et al.* Saccharomyces Genome Database. *Methods in enzymology* **350**, 329 (2002).

96 Cherry, J. M. *et al.* SGD: Saccharomyces genome database. *Nucleic acids research* **26**, 73-79 (1998).

97 Hardy, S. *et al.* The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A. Z deposition. *PLoS genetics* **5**, e1000687 (2009).

98 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).

99 Zlatanova, J. & Thakar, A. H2A. Z: view from the top. *Structure* **16**, 166-179 (2008).

100 Santisteban, M. S., Kalashnikova, T. & Smith, M. M. Histone H2A. Z regulates transcription and is partially redundant with nucleosome remodeling complexes. *Cell* **103**, 411-422 (2000).

101 Meneghini, M. D., Wu, M. & Madhani, H. D. Conserved histone variant H2A. Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**, 725-736 (2003).

102 Zhang, H. *et al.* The Yaf9 component of the SWR1 and NuA4 complexes is required for proper gene expression, histone H4 acetylation, and Htz1 replacement near telomeres. *Molecular and cellular biology* **24**, 9424-9436 (2004).

103 Fan, J. Y., Rangasamy, D., Luger, K. & Tremethick, D. J. H2A. Z alters the nucleosome surface to promote HP1α-mediated chromatin fiber folding. *Molecular cell* **16**, 655-661 (2004).

104 Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C. & Tremethick, D. J. The essential histone variant H2A. Z regulates the equilibrium between different chromatin conformational states. *Nature Structural & Molecular Biology* **9**, 172-176 (2002).

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

106 Pinto, D. M. S. & Flaus, A. in *Genome Stability and Human Diseases* 55-78 (Springer, 2010).

107 Shroff, R. *et al.* Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. *Current biology* **14**, 1703-1711 (2004).

108 Chadwick, B. P. & Willard, H. F. A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *The Journal of cell biology* **152**, 375-384 (2001).

109 Soboleva, T. A. *et al.* A unique H2A histone variant occupies the transcriptional start site of active genes. *Nature structural & molecular biology* **19**, 25-30 (2012).

110 Bao, Y. *et al.* Nucleosomes containing the histone variant H2A. Bbd organize only 118 base pairs of DNA. *The EMBO journal* **23**, 3314-3324 (2004).

111 Doyen, C. M. *et al.* Dissection of the unusual structural and functional properties of the variant H2A. Bbd nucleosome. *The EMBO journal* **25**, 4234-4244 (2006).

112 Zhou, J., Fan, J. Y., Rangasamy, D. & Tremethick, D. J. The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nature structural & molecular biology* **14**, 1070-1076 (2007).

113 Consortium, U. The universal protein resource (UniProt). *Nucleic acids research* **36**, D190-D195 (2008).

114 Costanzi, C. & Pehrson, J. R. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393**, 599-601 (1998).

115 Buschbeck, M. *et al.* The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nature structural & molecular biology* **16**, 1074-1079 (2009).

116 Timinszky, G. *et al.* A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nature structural & molecular biology* **16**, 923-929 (2009).

117 Muthurajan, U. M., McBryant, S. J., Lu, X., Hansen, J. C. & Luger, K. The linker region of macroH2A promotes self-association of nucleosomal arrays. *Journal of Biological Chemistry* **286**, 23852-23864 (2011).

118 Lantermann, A. B. *et al.* Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nature structural & molecular biology* **17**, 251-257 (2010).

119 Ronald Morris, N. Nucleosome structure in Aspergillus nidulans. *Cell* **8**, 357-363 (1976).

120 Athey, B. D., Smith, M. F., Rankert, D. A., Williams, S. P. & Langmore, J. P. The diameters of frozen-hydrated chromatin fibers increase with DNA linker length: evidence in support of variable diameter models for chromatin. *The Journal of cell biology* **111**, 795-806 (1990).

121 PEARSON, E. C., BATES, D. L., PROSPERO, T. D. & THOMAS, J. O. Neuronal nuclei and glial nuclei from mammalian cerebral cortex. *European Journal of Biochemistry* **144**, 353-360 (1984).

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

123 Robinson, P. J., Fairall, L., Huynh, V. A. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences* **103**, 6506-6511 (2006).

124 Routh, A., Sandin, S. & Rhodes, D. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proceedings of the National Academy of Sciences* **105**, 8872-8877 (2008).

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

126 Grigoryev, S. A. Nucleosome spacing and chromatin higher-order folding. *Nucleus* **3**, 493-499 (2012).

127 Woodcock, C. L., Skoultchi, A. I. & Fan, Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Research* **14**, 17-25 (2006).

128 BATES, D. L., BUTLER, P. J. G., PEARSON, E. C. & THOMAS, J. O. Stability of the Higher-Order Structure of Chicken-Erythrocyte Chromatin in Solution. *European Journal of Biochemistry* **119**, 469-476 (1981).

129 KASINSKY, H. E., LEWIS, J. D., DACKS, J. B. & Ausio, J. Origin of H1 linker histones. *The FASEB Journal* **15**, 34-42 (2001).

130 Roque, A., Iloro, I., Ponte, I., Arrondo, J. L. R. & Suau, P. DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. *Journal of Biological Chemistry* **280**, 32141-32147 (2005).

131 Zhou, Y.-B., Gerchman, S. E., Ramakrishnan, V., Travers, A. & Muyldermans, S. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* **395**, 402-405 (1998).

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

133 Shen, X. & Gorovsky, M. A. Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell* **86**, 475-483 (1996).

134 Hellauer, K., Sirard, E. & Turcotte, B. Decreased expression of specific genes in yeast cells lacking histone H1. *Journal of Biological Chemistry* **276**, 13587-13592 (2001).

135 Dasso, M., Dimitrov, S. & Wolffe, A. P. Nuclear assembly is independent of linker histones. *Proceedings of the National Academy of Sciences* **91**, 12477-12481 (1994).

136 Carruthers, L. M., Bednar, J., Woodcock, C. L. & Hansen, J. C. Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding. *Biochemistry* **37**, 14776-14787 (1998).

137 Kruithof, M. *et al.* Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature structural & molecular biology* **16**, 534-540 (2009).

138 Grigoryev, S. A., Arya, G., Correll, S., Woodcock, C. L. & Schlick, T. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proceedings of the National Academy of Sciences* **106**, 13317-13322 (2009).

139 Barbera, A. J. *et al.* The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311**, 856-861 (2006).

140 Roussel, L., Erard, M., Cayrol, C. & Girard, J. P. Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A–H2B acidic pocket. *EMBO reports* **9**, 1006-1012 (2008).

141 Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562-566 (2010).

142 Armache, K.-J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334**, 977-982 (2011).

143 Kato, H. *et al.* Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proceedings of the National Academy of Sciences* **108**, 12283-12288 (2011).

144 Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120-124 (2001).

145 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120 (2001).

146 Nakayama, J.-i., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-113 (2001).

147 Jacobs, S. A. *et al.* Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *The EMBO journal* **20**, 5232-5241 (2001).

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

149 Sadaie, M. *et al.* Balance between distinct HP1 family proteins controls heterochromatin assembly in fission yeast. *Molecular and cellular biology* **28**, 6973-6988 (2008).

150 Azzaz, A. M. *et al.* HP1Hsα Promotes Nucleosome Associations that Drive Chromatin Condensation. *Journal of Biological Chemistry*, jbc. M113. 512137 (2014).

151 Hahn, M. *et al.* Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. *Genes & development* **27**, 859-872 (2013).

152 Jürgens, G. A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature* **316**, 153-155 (1985).

153 Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature reviews Molecular cell biology* **10**, 697-708 (2009).

154 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).

155 Anderson, D. E., Losada, A., Erickson, H. P. & Hirano, T. Condensin and cohesin display different arm conformations with characteristic hinge angles. *The Journal of cell biology* **156**, 419-424 (2002).

156 Tada, K., Susumu, H., Sakuno, T. & Watanabe, Y. Condensin association with histone H2A shapes mitotic chromosomes. *Nature* **474**, 477-483 (2011).

157 Yan, J. *et al.* Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* **154**, 801-813 (2013).

158 Van Hoide, K., Sahasrabuddhe, C. G. & Shaw, B. R. A model for particulate structure in chromatin. *Nucleic Acids Research* **1**, 1579-1586 (1974).

159 Finch, J. & Klug, A. Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences* **73**, 1897-1901 (1976).

160 Worcel, A., Strogatz, S. & Riley, D. Structure of chromatin and the linking number of DNA. *Proceedings of the National Academy of Sciences* **78**, 1461-1465 (1981).

161 Woodcock, C., Frado, L.-L. & Rattner, J. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *The Journal of cell biology* **99**, 42-52 (1984).

162 Dorigo, B. *et al.* Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **306**, 1571-1573 (2004).

163 Depken, M. & Schiessel, H. Nucleosome shape dictates chromatin fiber structure. *Biophysical journal* **96**, 777-784 (2009).

164 Emanuel, M., Radja, N. H., Henriksson, A. & Schiessel, H. The physics behind the larger scale organization of DNA in eukaryotes. *Physical biology* **6**, 025008 (2009).

165 Horowitz, R., Agard, D., Sedat, J. & Woodcock, C. The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *The Journal of cell biology* **125**, 1-10 (1994).

166 Eltsov, M., MacLellan, K. M., Maeshima, K., Frangakis, A. S. & Dubochet, J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proceedings of the National Academy of Sciences* **105**, 19732-19737 (2008).

167 Hiratani, I. *et al.* Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome research* **20**, 155-169 (2010).

168 Bystricky, K., Heun, P., Gehlen, L., Langowski, J. & Gasser, S. M. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16495-16500 (2004).

169 Dekker, J. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *Journal of biological chemistry* **283**, 34532-34540 (2008).

170 Tremethick, D. J. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128**, 651-654 (2007).

171 Maeshima, K., Hihara, S. & Eltsov, M. Chromatin structure: does the 30-nm fibre exist< i> in vivo</i>? *Current opinion in cell biology* **22**, 291-297 (2010).

172 Fussner, E., Ching, R. W. & Bazett-Jones, D. P. Living without 30nm chromatin fibers. *Trends in biochemical sciences* **36**, 1-6 (2011).

173 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science* **326**, 289-293 (2009).

174 Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research* **19**, 37-51 (2011).

175 Cavalier-Smith, T. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Annals of Botany* **95**, 147-175 (2005).

176 Jorgensen, P. *et al.* The size of the nucleus increases as yeast cells grow. *Molecular biology of the cell* **18**, 3523-3532 (2007).

177 Maul, G. G. & Deaven, L. Quantitative determination of nuclear pore complexes in cycling cells with differing DNA content. *The Journal of cell biology* **73**, 748-760 (1977).

178 Olmo, E. Nucleotype and cell size in vertebrates: a review. *Basic and applied histochemistry* **27**, 227-256 (1982).

179 Gregory, T. R.

180 Vinogradov, A. E. Genome size and chromatin condensation in vertebrates. *Chromosoma* **113**, 362-369 (2005).

181 Schuster, T., Han, M. & Grunstein, M. Yeast histone H2A and H2B amino termini have interchangeable functions. *Cell* **45**, 445-451 (1986).

182 Kim, J.-A., Hsu, J.-Y., Smith, M. M. & Allis, C. D. Mutagenesis of pairwise combinations of histone amino-terminal tails reveals functional redundancy in budding yeast. *Proceedings of the National Academy of Sciences* **109**, 5779-5784 (2012).

183 Recht, J., Dunn, B., Raff, A. & Osley, M. A. Functional analysis of histones H2A and H2B in transcriptional repression in Saccharomyces cerevisiae. *Molecular and cellular biology* **16**, 2545-2553 (1996).

184 Parra, M. A. & Wyrick, J. J. Regulation of gene transcription by the histone H2A N-terminal domain. *Molecular and cellular biology* **27**, 7641-7648 (2007).

185 Zheng, S., Wyrick, J. J. & Reese, J. C. Novel trans-tail regulation of H2B ubiquitylation and H3K4 methylation by the N terminus of histone H2A. *Molecular and cellular biology* **30**, 3635-3645 (2010).

186 Vogelauer, M., Wu, J., Suka, N. & Grunstein, M. Global histone acetylation and deacetylation in yeast. *Nature* **408**, 495-498 (2000).

187 Jiang, L. *et al.* Global Assessment of Combinatorial Post-translational Modification of Core Histones in Yeast Using Contemporary Mass Spectrometry LYS4 TRIMETHYLATION CORRELATES WITH DEGREE OF ACETYLATION ON THE SAME H3 TAIL. *Journal of biological chemistry* **282**, 27923-27934 (2007).

188 Di Lorenzo, A. & Bedford, M. T. Histone arginine methylation. *FEBS letters* **585**, 2024-2031 (2011).

189 Waldmann, T. *et al. Methylation of H2AR29 is a novel repressive PRMT6 target*. (Bibliothek der Universität Konstanz, 2011).

190 Zheng, C. & Hayes, J. J. Intra-and inter-nucleosomal protein-DNA interactions of the core histone tail domains in a model system. *Journal of Biological Chemistry* **278**, 24217-24224 (2003).

191 Gordon, F., Luger, K. & Hansen, J. C. The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. *Journal of Biological Chemistry* **280**, 33701-33706 (2005).

192 Luger, K. & Richmond, T. J. The histone tails of the nucleosome. *Current opinion in genetics & development* **8**, 140-146 (1998).

193 Papamokos, G. V. *et al.* Structural Role of RKS Motifs in Chromatin Interactions: A Molecular Dynamics Study of HP1 Bound to a Variably Modified Histone Tail. *Biophysical journal* **102**, 1926-1933 (2012).

194 Kouzarides, T. Histone methylation in transcriptional control. *Current opinion in genetics & development* **12**, 198-209 (2002).

195 Lee, J.-H. *et al.* PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. *Journal of Biological Chemistry* **280**, 3656-3664 (2005).

196 Lenfant, F., Mann, R., Thomsen, B., Ling, X. & Grunstein, M. All four core histone N-termini contain sequences required for the repression of basal transcription in yeast. *The EMBO journal* **15**, 3974 (1996).

197 Guacci, V., Koshland, D. & Strunnikov, A. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell* **91**, 47 (1997).

198 Szerlong, H. J. & Hansen, J. C. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure This paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal's usual peer review process. *Biochemistry and Cell Biology* **89**, 24-34 (2010).

199 Grigoryev, S. Nucleosome spacing and chromatin higher-order folding. *Nucleus* **3**, 9-15 (2012).

200 Furuyama, S. & Biggins, S. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proceedings of the National Academy of Sciences* **104**, 14706-14711 (2007).

201 Zimmer, C. & Fabre, E. Principles of chromosomal organization: lessons from yeast. *The Journal of cell biology* **192**, 723-733 (2011).

202 Therizols, P., Duong, T., Dujon, B., Zimmer, C. & Fabre, E. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proceedings of the National Academy of Sciences* **107**, 2025-2030 (2010).

203 Neumann, F. R. & Nurse, P. Nuclear size control in fission yeast. *The Journal of cell biology* **179**, 593-600 (2007).

204 Weischet, W., Tatchell, K., Van Holde, K. & Klump, H. Thermal denaturation of nucleosomal core particles. *Nucleic acids research* **5**, 139-160 (1978).

205 Stenson, P. D. *et al.* The human gene mutation database: 2008 update. *Genome Med* **1**, 13 (2009).

206 Edens, L. J., White, K. H., Jevtic, P., Li, X. & Levy, D. L. Nuclear size regulation: from single cells to development and disease. *Trends in cell biology* (2012).

207 Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic acids research* **39**, D945-D950 (2011).

208 Rohs, R. *et al.* The role of DNA shape in protein–DNA recognition. *Nature* **461**, 1248-1253 (2009).

209 Dover, J. *et al.* Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *Journal of Biological Chemistry* **277**, 28368-28371 (2002).

210 Gardner, R. G., Nelson, Z. W. & Gottschling, D. E. Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. *Molecular and cellular biology* **25**, 6123-6139 (2005).

211 Kyriss, M. N., Jin, Y., Gallegos, I. J., Sanford, J. A. & Wyrick, J. J. Novel functional residues in the core domain of histone H2B regulate yeast gene expression and silencing and affect the response to DNA damage. *Molecular and cellular biology* **30**, 3503-3518 (2010).

212 White, C. L., Suto, R. K. & Luger, K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *The EMBO journal* **20**, 5207-5218 (2001).

213 Hirschhorn, J. N., Bortvin, A. L., Ricupero-Hovasse, S. L. & Winston, F. A new class of histone H2A mutations in Saccharomyces cerevisiae causes specific transcriptional defects in vivo. *Molecular and cellular biology* **15**, 1999-2009 (1995).

214 DeLano, W. L. The PyMOL molecular graphics system. (2002).

215 Unni, S. *et al.* Web servers and services for electrostatics calculations with APBS and PDB2PQR. *Journal of computational chemistry* **32**, 1488-1491 (2011).

216 Levy, D. L. & Heald, R. Nuclear Size Is Regulated by Importin α and Ntf2 in Xenopus. *Cell* **143**, 288-298 (2010).

217 Cavalier-Smith, T. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *Journal of Cell Science* **34**, 247-278 (1978).

218 Martin, P. Variation in the amounts of nucleic acids in the cells of different species of higher plants. *Experimental cell research* **44**, 84-94 (1966).

219 Bennett, M. Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **181**, 109-135 (1972).

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

221 Zhang, Y. *et al.* Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nature structural & molecular biology* **16**, 847-852 (2009).

222 Hughes, A. L., Jin, Y., Rando, O. J. & Struhl, K. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Molecular cell* **48**, 5-15 (2012).

223 Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953-961 (1998).

224 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10**, 421 (2009).

225 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797 (2004).

226 Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* **5**, 113 (2004).

227 Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome research* **14**, 1188-1190 (2004).

228 Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic acids research* **18**, 6097-6100 (1990).

229 de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-1454 (2004).

230 Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248 (2004).

231 Rando, O. J. Genome-wide mapping of nucleosomes in yeast. *Methods in enzymology* **470**, 105-118 (2010).

232 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).

233 Chen, K. *et al.* DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome research* **23**, 341-351 (2013).

234 Schmitt, M. E., Brown, T. A. & Trumpower, B. L. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic acids research* **18**, 3091 (1990).

235 Ferrari, R. *et al.* Reorganization of the host epigenome by a viral oncogene. *Genome research* **22**, 1212-1221 (2012).

236 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods in enzymology* **304**, 3-19 (1999).

237 Lowary, P. & 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 (1998).

238 Hansen, J. C. & Turgeon, C. L. in *Chromatin Protocols* 127-141 (Springer, 1999).

239 Zou, L., Mitchell, J. & Stillman, B. CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. *Molecular and cellular biology* **17**, 553-563 (1997).