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: Siavash Kurdistani
Tel: 310.794.5194
skurdistani@mednet.ucla.edu

Competing interests

The authors declare that no competing interests exist.

Abstract

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 is unknown whether histones themselves have also evolved to regulate chromatin compaction. Analysis of histone sequences from 160 eukaryotes revealed that the H2A N-terminus 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.

Introduction

Genome size, defined as the haploid DNA content of a cell, has increased as eukaryotes evolved from single-cell species to more complex, multicellular organisms. Within the same evolutionary timeframe, nuclear volume has also increased but at a slower rate than genome size expansion (Maul and Deaven, 1977; Olmo, 1982). While the ratio of nuclear to cell size has remained essentially constant in eukaryotes (Cavalier-Smith, 2005), the disproportional increase in genome size relative to the nuclear volume has required organisms with larger genomes to compact their chromatin to greater extents than organisms with small sized genomes. Indeed there is a positive correlation between genome size and native chromatin compaction as measured by dye incorporation into chromatin (Vinogradov, 2005). In most eukaryotes, the genome is organized into chromatin by the repeating nucleosomal structure (Luger et al., 1997). The nucleosomes stack and fold into higher order structures, serving to systematically compact the genome (Duan et al., 2010; Lieberman-Aiden et al., 2009) and to regulate molecular processes that are based on DNA (Celeste et al., 2002; Fischle et al., 2005; Fussner et al., 2011; Kouzarides, 2007; Vogelauer et al., 2002).

The surface of the histone octamer has fourteen DNA interaction sites. Each interaction is mediated by an arginine residue that intercalates into the minor groove of the DNA to stabilize the nucleosomal structure (Luger et al., 1997; West et al., 2010). Arginine is the most commonly used amino acid for interaction with DNA due to its positive charge and the lower energetic cost compared to lysine for displacing water when intercalating into the minor groove (Rohs et al., 2009). Nucleosomes mediate chromatin compaction through a variety of mechanisms. For instance, nucleosomes form higher order structures through inter-nucleosomal contacts between the histone H4 N-terminal domain (NTD) and the acidic patch of H2A and between two H2B C-terminal domains (CTD) (Dorigo et al., 2003; Dorigo et al., 2004; Gordon et al., 2005; Luger et al., 1997; Schalch et al., 2005). Histone variants, such as H2A.Z or H2A.Bbd, as well as post-translational modifications of histones, such as H4K16ac, can further regulate the degree of compaction (Bao et al., 2004; Chandrasekharan et al., 2009; Fierz et al., 2011; Kim et al., 2009; Shogren-Knaak et al., 2006; Suto et al., 2000; Zhou et al., 2007). Polycomb complexes compact large domains of chromatin (Eskeland et al., 2010) and are important for proper development.

Histones of the H1 family promote additional compaction by binding between nucleosomes to linker DNA near the DNA entry/exit site on nucleosomes and stabilize the intrinsic ability of nucleosomal arrays to fold *in vitro* (Carruthers et al., 1998; Robinson et al., 2008; Szerlong and Hansen, 2010). Linker histones may affect chromatin compaction globally (Fan et al., 2005), at specific stages of the cell cycle such as mitosis (Maresca et al., 2005) or at specific regions of the genome (Li et al., 2012). In contrast to canonical histones, the sequence of linker histones are much less conserved (Caterino and Hayes, 2010) and ectopic expression of human linker histones in the budding yeast even at low levels is lethal for the cell (Miloshev et al., 1994). Finally, structural proteins such as condensin also contribute to chromatin condensation (Tada et al., 2011). Many of these modulatory mechanisms are dynamic in nature (Luger et al., 2012) and may help explain why multicellular organisms can compact chromatin to different degrees in different cell types. However, despite the existence of these mechanisms for genome compaction in higher eukaryotes, it has not been known whether the canonical histones themselves have evolved sequence features that also contribute to the generally increased chromatin compaction observed in organisms with larger genomes.

Here we provide evidence from analysis of 160 fully-sequenced eukaryotic genomes that arginine (R) residues at specific positions in the N-terminal tail of histone H2A—which protrudes from the nucleosome on the opposite side of DNA entry/exit site—have co-evolved with increasing genome size with a concomitant decrease in serine/threonines. Although increases in genome size are associated with phylogenetic evolution from protozoa to fungi to more complex plants and animals, we present genetic and molecular evidence from the budding yeast and human cells as well as *in vitro* biochemical data to demonstrate that the evolutionary changes in H2A directly regulate chromatin compaction *in vivo* and *in vitro* with consequences for the nuclear volume. The evolutionary changes in H2A regulate chromatin compaction in yeast and human cells, revealing a surprising flexibility in the dynamics of the chromatin fiber that has been conserved across distant eukaryotes. This previously unrecognized structural feature of the nucleosome has evolved to enable greater chromatin compaction when genome size is disproportionately larger than the nuclear volume. Our findings also suggest that the reported mutations in histone H2A NTD may contribute to the altered chromatin compaction that is commonly observed in cancer cells (Zink et al., 2004).

Results

H2A acquires specifically positioned arginines as genome size increases

To determine whether specific residues in the four core histones have co-evolved with increasing genome size, we performed residue composition analysis of canonical histone protein sequences from 160 fully sequenced eukaryotes with genome sizes ranging from 8-5600 Mbp encompassing protozoa, fungi, plants and animals. The canonical histone proteins for each organism were defined based on at least 90% overlap and 35% identity with the histone fold domain of the corresponding human sequence (see Methods). Each organism was categorized as having a small (<100 Mbp), medium (100-1000 Mbp), or large (>1000 Mbp) genome (Fig 1 – figure supplement 1A). Of the canonical histones, the H2A NTD showed the most statistically significant variability in amino acid residues, where the number of arginines increased with increasing genome size (Fig 1A), while the number of serines (S) and threonines (T) decreased (Fig 1B). Other amino acid residues in the H2A NTD, including lysines (K), did not correlate with genome size (Fig 1 – figure supplement 1B).

The acquisition of arginines and loss of serines/threonines in the H2A NTD with increasing genome size occur at specific positions and in sequential order. For instance, the human H2A NTD contains arginine residues at positions 3 and 11 that are absent in yeast and at position 20 which is correspondingly a lysine in yeast (Fig 1C). In contrast, the human sequence lacks S10 and S15 that are observed in the yeast H2A NTD (Fig 1C). Alignment of all H2A NTD sequences also revealed similar trends across all eukaryotes studied here. The heat map in figure 1D shows the occurrence of arginines and serines/threonines in the H2A NTD as a function of genome size (see Fig 1 - source data 1 for raw data and Fig 1 – figure supplement 1C for statistical analysis). At position 3, an arginine (R3) is predominantly present in medium and large species but lacking in small species. At position 11, a lysine (K11) is observed in species with medium genomes which evolves to an arginine (R11) mainly in organisms with large genomes. R17 is present in most organisms examined, suggesting a very conserved function for this residue (Zheng et al., 2010). At position 20, small genomes contain predominantly a lysine

residue, which converts to arginine in medium and large genomes. In contrast, serines/threonines at positions 10 and 15 are found primarily in organisms with small genomes and much less so in organisms with medium and large genomes (Fig 1D). Additionally, each of the four H2A NTD arginines is surrounded by a conserved motif (Fig 1E). The residues surrounding R3 and R17 are mainly glycine and serine respectively. At position 11, the motif varies based on genome size. Species with medium sized genomes contain VKG and those with large genomes contain ARA. The same is true of position 20, where AKA is present in organisms with small genomes and (S/T)RA in larger genome species (Fig 1E).

Interestingly, except for R3, the positions of all the other evolutionarily varying residues in the H2A NTD are strongly conserved relative to the histone fold domain and not the N-terminus (Fig 1F). When counting conventionally from the N-terminus, amino acids R11, R17 and R20—which are numbered based on the human sequence—were not observed consistently at the same positions in other organisms. However these residues are respectively 12, 6 and 3 amino acids away from the histone fold in most species (note the vertical axes in Fig 1F). S10 and S15—which are numbered based on the yeast sequence—also show more uniform positioning when counted from the histone fold. Altogether, as genome size increases, arginines appear in conserved positions within the H2A NTD relative to the histone fold and serines and threonines are lost.

Arginines 3 and 11 in the H2A NTD increase chromatin compaction

To determine whether arginines and serines/threonines of the H2A NTD affect chromatin compaction *in vivo,* we took advantage of a strain of *S. cerevisiae*, that has both chromosomal copies of H2A deleted and carries a single copy of H2A on a plasmid (TSY107) to construct mutant strains containing single or multiple insertions of arginines into their conserved motifs, deletions of serines, or combinations thereof (see Table 1 for specific amino acid changes and Supplementary file 1A for a description of the mutant strains). Two mutants, R3(ΔGS10)R11 and R11ΔS15, were also designed such that the spacing between R3 and R11 or R11 and the histone fold, respectively, is the same as in the H2A NTD of organisms with large genomes (Fig

178 1C, 1F). As a control for positive charge, mutant strains with lysines inserted in the same
179 positions as arginines were also generated.
180
181 To test the effects of H2A NTD changes on chromatin compaction, the physical distance
182 between two probes on chromosome XVI spaced 275 kb apart was assessed in each of the H2A
183 mutants using fluorescent *in situ* hybridization (FISH) (Fig 2A) (Bystricky et al., 2004; Guacci et
184 al., 1997). The probes were differentially labeled and visualized by confocal microscopy. The
185 distance between the probes was measured in a single plane in which both probes were present
186 within each nucleus (Bystricky et al., 2004). When compared to the isogenic wild type (WT),
187 addition of a single arginine at position 3 (R3) or 11 (R11) to the H2A NTD was sufficient to
188 significantly decrease the average interprobe distance by 18% and 15%, respectively (Fig 2B-C,
189 Table 1). The average interprobe distance was further decreased by 22% when both arginines
190 were present (R3R11) and even more so (30%) in R3(ΔGS10)R11. Deleting G9S10 (ΔGS10)
191 alone caused slightly increased compaction with low statistical significance (Table 1). The
192 largest decrease in interprobe distance (41%) was observed in the R11ΔS15 mutant, which
193 places R11 twelve amino acids from the histone fold, the same position as in organisms with
194 large genomes. Removal of S15 (ΔS15) alone had no effect. The effect of arginines was not
195 simply due to increasing the positive charge of the H2A NTD, as insertions of lysines at
196 positions 3 and 11 did not significantly affect the interprobe distances (Fig 2C and
197 Supplementary file 2). Although lysines are found at these positions in certain species (Fig 1D),
198 the lack of potential compaction by lysines may be due to the absence of other evolutionary
199 changes in yeast histones (see Fig 1E). Additionally, R17K or K20R mutations did not affect
200 compaction, nor did a randomly inserted arginine at position 6 (R6) (Fig 2C and Supplementary
201 file 2), suggesting that not every arginine in the H2A NTD contributes to chromatin compaction.
202
203 We further confirmed the effects of R11 on chromatin compaction using three additional
204 probe sets (Fig 2A). The level of compaction seen in our WT strain is similar to what has been
205 previously reported in yeast using a different strain background (Bystricky et al., 2004). The
206 interprobe distances for all probe sets were significantly decreased in R11 compared to WT and
207 even more so in R11ΔS15 but not ΔS15 alone (Fig 2 – figure supplement 1 A-C and
208 Supplementary file 2). Plotting the physical vs. genomic distances for all probe sets revealed

uniform compaction across large genomic distances (Fig 2D). The effect of R11 on chromatin compaction was not strain-specific as H2A R11 and R11ΔS15, but not ΔS15, caused chromatin compaction in a different strain background (Fig 2E; Supplementary file 2). We therefore conclude that H2A arginines at positions 3 and 11, especially when R11 is placed at the evolutionarily-conserved position relative to the histone fold, increase the degree of chromatin compaction.

Chromatin is differentially compacted at different cell cycle stages (Guacci et al., 1997). Cell cycle profile analysis showed little difference between the strain harboring WT H2A and any of the mutant strains (Fig 2 – figure supplement 1D-E), indicating that the observed differences in chromatin compaction are not due to altered cell cycle profiles. Chromatin compaction may also be influenced by nucleosomal spacing; indeed the linker DNA length is larger in human cells than in yeast (Grigoryev, 2012). We find that there are essentially no differences in nucleosomal density in H2A arginine mutants using Micrococcal nuclease (MNase) digestion (Fig 2F, Figure 2 – figure supplement 1F), indicating that the average nucleosomal spacing is not affected by these mutations. But the more compact mutants displayed decreased accessibility to MNase as indicated by the delayed appearance of nucleosomal digestion pattern (Fig 2F, Figure 2 – figure supplement 1F).

H2A arginines and serines affect nuclear volume in yeast

Since chromatin structure may influence volume of the nucleus (Cavalier-Smith, 2005), we asked whether nuclear volume was affected by H2A-mediated chromatin compaction. We tagged a nuclear pore protein, Nup49, in its chromosomal locus with GFP to visualize the nuclear membrane and used confocal microscopy to capture three-dimensional images of the nucleus to quantify volumes of ≥150 cells per H2A mutant (Fig 3A-B, Table 1, Supplementary file 3, see Methods for volume calculations). As compared to WT cells, H2A mutants containing R11 or R3R11, both of which contain more compact chromatin, displayed significantly decreased nuclear volumes. The average nuclear volume in the R3 mutant was also less than WT but did not reach statistical significance. Interestingly, H2A mutants from which serines 10 and 15 were removed displayed larger nuclear volumes. Simultaneous insertions of arginines

into these strains (R3(ΔGS10)R11 and R11ΔS15) decreased their nuclear volume (R11ΔS15 $p<0.001$ compared to ΔS15), restoring them to levels similar to WT. The control strains with either lysines or R6 had nuclear volumes similar to or larger than WT. Neither arginines nor serines had any effect on total cell size as measured by concanavalin A staining (Fig 3 – figure supplement 1A-B, Supplementary file 3). In the FY406 strain background, ΔS15 did not cause an increase in nuclear volume; and thus both R11 and R11ΔS15 strains exhibited smaller nuclear volumes than isogenic WT (Fig 3C, Supplementary file 3). These data suggest that modulation of chromatin compaction through the H2A NTD, especially in presence of R11, affects the nuclear volume but this effect may be indirect (see human data below).

Loss of H2A arginines causes de-compaction of chromatin in human cells

Since the H2A NTD in large genomes contains both R3 and R11, we expected that their removal would cause de-compaction of chromatin. To test this prediction, we ectopically expressed WT or mutant H2A in several human cell lines and measured the distances between probes 0.49 Mbp apart on chromosome 1 by FISH, as well as the largest nuclear cross-sectional areas (see Methods). The H2A gene was HA-tagged and mutated to remove R3 (ΔR3), to replace R11 with alanine (R11A) or lysine (R11K), or to combine two mutations (ΔR3R11A). The H2A constructs were overexpressed using the strong CMV promoter in the normal human IMR90 fibroblasts, the breast cancer cell line MDA-MB-453, or the HEK293 cells. Cells overexpressing ΔR3, R11A, or ΔR3R11A H2A mutants had increased interprobe distances, indicating de-compaction of chromatin. Expression of H2A R11K had modest effects on chromatin de-compaction with marginal statistical significance (Fig 4A-B, Fig 4 – figure supplement 1A-B, Supplementary file 4). Cells expressing any of the H2A mutants displayed larger nuclear areas, suggesting that nuclear size is increased (Fig 4C-D, Fig 4 – figure supplement 1C-F, Supplementary file 5). Equal degrees of overexpression were confirmed by immunofluorescence analysis with an anti-HA antibody and detection of HA-H2A by western blotting (Fig 4D, Fig 4 – figure supplement 1G). Ectopic expression of a C-terminally FLAG-tagged H2A mutant missing residues 1-12 (Δ1-12) also caused significant de-compaction of chromatin and increased nuclear area despite being expressed at a lower level than WT (Fig 4E-

F). These data demonstrate that consistent with our predictions, the H2A NTD, especially arginines 3 and 11, function to compact chromatin in human cells.

H2A R11 regulates compaction of nucleosomal arrays *in vitro*

Because R11 compacts chromatin *in vivo*, we investigated whether this effect is directly on the chromatin fiber. We used step-wise salt dialysis to assemble nucleosomal arrays with a DNA template containing twelve copies of the 177 bp "601" nucleosome positioning sequence (601-177-12) and recombinant *X. laevis* histone octamers that contain either WT H2A or one with R11 deleted (ΔR11). We assembled nucleosomal arrays at different octamer-to-template ratios (0.9, 1, and 1.1 octamer to 1 template) and monitored the quality of the arrays by $MgCl_2$ precipitation and restriction digest analysis using ScaI. We found that a 1:1 octamer-to-template ratio gave the best results as the ScaI digest demonstrated well assembled arrays compared to the 5% free DNA loaded as a comparison (Fig 5A). We used analytical ultracentrifugation to determine the sedimentation velocity combined with van Holde–Weischet analysis (Weischet et al., 1978) to ascertain the distribution of sedimentation coefficients (S) for each nucleosomal array in the absence or presence of 0.8 mM $MgCl_2$, a concentration of the divalent cation that promotes intra-molecular folding of nucleosomal arrays (Schwarz and Hansen, 1994). In the absence of Mg^{2+}, arrays containing WT H2A sedimented with a coefficient of 33.1, which is a value that has been previously shown for similar arrays (Dorigo et al., 2003; Shogren-Knaak et al., 2006; Zhou et al., 2007). In contrast, arrays missing R11 adopted a more extended conformation with a smaller sedimentation coefficient of 31.0 (Fig 5B). Addition of Mg^{2+} increased compaction of both arrays and shifted the sedimentation coefficients to 39.3 and 37.4 for WT and ΔR11 H2A, respectively (Fig 5B). A second independent chromatin assembly and ultracentrifuge analysis confirmed these results (Figure 5 – figure supplement 1). Thus, in the absence of R11 in the H2A NTD, nucleosomal arrays adopt a less compact conformation even in the presence of divalent cations, showing that R11 directly increases chromatin compaction.

Compaction of chromatin by H2A NTD arginines does not alter global gene expression in yeast

To determine whether chromatin compaction through H2A arginines interferes with transcription regulation, we examined gene expression patterns in the H2A yeast mutants. Remarkably, there was a high level of correlation (≥0.99) between all strains examined (Fig 6A), and no specific gene ontology was found among the genes that were differentially expressed by two-fold or more. The expression levels of the histone genes were similar, indicating that altered levels of histone genes expression do not account for the changes in chromatin compaction. These data indicate that compaction of chromatin by H2A does not significantly alter global gene expression in exponentially growing cells.

All strains also showed similar growth rates in rich media (Fig 6B) and no significant differences in sensitivity to 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 (Fig 6C). But in competition growth assays in which equal amounts of WT and H2A mutant cells harboring the PGK1 gene fused to either GFP or RFP were co-cultured, the H2A mutants regardless of any effect on chromatin compaction, were outcompeted (Fig 6D). This suggests that changes in the H2A NTD sequence can affect the overall fitness of the cell.

H2A NTD arginines and their surrounding residues are mutated in cancer and affect chromatin compaction

Deregulated chromatin compaction is often a pathological hallmark of cancer cells (Edens et al., 2012), although the underlying mechanisms are not well-understood. A survey of the COSMIC database (Forbes et al., 2011), as of the time of writing, revealed 41 documented missense mutations within the H2A NTD with 29 (71%) affecting a residue within one of the four arginine motifs (Fig 7A). R11, which had the strongest effect of any single arginine residue on chromatin compaction, is the most commonly mutated residue in the H2A NTD. We tested the effects of three of these mutations, R11C, H and P, and found that ectopic expression of each in normal human fibroblasts decreases chromatin compaction significantly with R11P having the strongest effect (Fig 7B-C). These cancer mutations have little effect on increasing nuclear area, however (Fig 7D-E), in contrast to R11A (Fig 4D). It is unclear to what extent the H2A mutants

have to be expressed in cancer cells relative to the seventeen canonical H2A genes in the human genome to affect chromatin compaction. But our data suggest that over-expression of an H2A mutant has the potential to disrupt chromatin compaction in cancer.

Discussion

In this study we describe evolutionary adaptations of the histone H2A whereby single arginines in the NTD function to dramatically affect the degree of genome compaction. This mechanism is distinct from several other known chromatin compaction mechanisms in higher eukaryotes (Bednar et al., 1998; Dorigo et al., 2003; Fierz et al., 2011; Shogren-Knaak et al., 2006; Zhou et al., 2007) in that it involves the histone proteins themselves. The H2A-mediated chromatin compaction thus provides a novel but potentially complementary mechanism for genome compaction. Organisms with small genomes but also very small cell size may face similar physical constraints as those with larger genomes, and may therefore use arginine-containing H2A as a means for chromatin compaction. For instance, *Ostreococcus tauri* which possesses an R3-containing H2A, is a free-living unicellular algae that has a very small genome of 12.6 Mbp but a cell diameter of 0.8 μm (Palenik et al., 2007). *S. cerevisiae*, which does not contain an H2A with R3, has a similarly sized genome but has a cell diameter that is approximately five times larger. Furthermore, certain organisms such as *Oikopleura dioica*, which has one of the smallest genomes in animals, have distinctive life cycles and possess H2A genes with and without arginines, which may enable them to dynamically regulate genome compaction at different stages of their life cycles (Moosmann et al., 2011) (for species with H2A isoforms see attached spreadsheet). So, the H2A arginines may have evolved in circumstances when the genome size became disproportionately large compared to nuclear volume. Interestingly, the toad, *Bufo gargarizans,* which has a genome size that is twice as large as the human genome, possesses an H2A gene with not only R3 and R11 but also glutamine 6 replaced with an arginine, suggesting that additional arginines in the H2A tail may enable further compaction in organisms with even larger genomes.

To better understand the three dimensional positions of the H2A NTD arginines, we examined a crystal structure of the mono-nucleosome in which R3, R11, R17, and R20 were all

363 simultaneously crystalized, and visualized interactions between nucleosomes within the crystal
364 lattice (Davey et al., 2002). Interestingly, while R17 and R20 are more buried within the
365 octamer, R3 and R11 are situated close to the DNA backbone. R3 is at 2.87 Å from the DNA
366 and could potentially bind the DNA gyre as DNA wraps around the histone octamer. R11 forms
367 close contacts with the DNA phosphate backbone of self and neighboring nucleosomes (4.09,
368 2.90 Å, respectively). Although these interactions may have helped form the crystal lattice, they
369 also suggest a possible mechanism for tighter nucleosomal stacking *in vivo* through shielding of
370 the DNA negative charge (Fig 1 – figure supplement 2A-B) (Davey et al., 2002). Thus, the
371 evolutionary appearance of arginines in the H2A NTD sequence at positions 3 and 11
372 corresponds to strategic positioning of R3 and R11 within the nucleosome structure that may
373 enable interactions with the DNA, leading to more compact chromatin.
374
375 Our *in vivo* data in yeast cells demonstrate that interprobe distances shorten in the
376 presence of the H2A arginines R3 and R11. While the mechanism of this shortening is still
377 unknown, the two most likely explanations are due to linear chromosomal compaction or
378 increased chromatin looping (Bohn and Heermann, 2010). However, our data is more consistent
379 with increased linear compaction due to several reasons. First, our analysis of multiple probes
380 along Chr XVI in yeast demonstrates a uniform compaction between all probe pairs examined.
381 Second, our *in vivo* data within human cells lines shows de-compaction of chromatin in the
382 absence of R11. If chromatin looping was the mechanism, loops would have to be disassembled
383 in human nuclei independent of factors such as CTCF and condensin. Third, our *in vitro* data
384 show that R11 alone affects chromatin compaction even in absence of divalent cations. Because
385 the *in vitro* experiments were performed with unmodified histones in arrays with equal linker
386 lengths, this strongly points to a direct effect of H2A NTD arginines on chromatin compaction
387 that occurs in short arrays. Both R3 and R11 are at contact distances from the DNA, and R11
388 may also contact the DNA backbone of the neighboring nucleosome (Fig 1 – figure supplement
389 2). These intra- and inter-nucleosomal interactions with the arginines and the DNA may serve to
390 neutralize the negative charge of the DNA backbone, leading to enhanced stacking of
391 nucleosomes and hence increased compaction. Consistent with this model, the other two
392 arginine residues in the H2A NTD, R17 and R20, which are more buried from the surface, do not
393 affect compaction by themselves. However, their functions may be to modulate the effects of the

surrounding residues. Although all canonical H2A genes contain R3 and R11 in humans, the cell may still be able to dynamically regulate chromatin compaction by these arginines. For instance, arginines may be subject to posttranslational modifications, such a methylation which makes the arginine residue bulkier, or citrullination which removes the positive charge (Di Lorenzo and Bedford, 2011; Hagiwara et al., 2005; Waldmann et al., 2011; Wang et al., 2001). Interestingly, the H2A NTD is situated in close proximity to the H2B CTD which when ubiquitylated, disrupts chromatin compaction *in vitro* (Fierz et al., 2011), lending support to the ability of this region of the nucleosome to modulate chromatin compaction.

The inability of lysines, especially at position 11, to increase chromatin compaction suggests exquisite structural constraints for H2A-mediated chromatin compaction. Although lysines and arginines both are positively charged, the positive charge of arginine is due to the presence of a guanidinium group which is structurally different from the positive charge of an amino group of a lysine residue. In this regard, it is interesting to note that only arginines contact DNA as it wraps around the nucleosome core (Luger et al., 1997); and arginines preferentially bind the minor groove of DNA compared to lysines (Rohs et al., 2009). The context in which lysines appear in evolution may be important as well. We did not observe a lysine at position 3 in our list of organisms, and K11 was present in organisms with medium sized genomes and surrounded mainly by the motif VKG (Fig 1D-E). When tested in our *S. cerevisiae* strains, K11 was in the context of AKA. So, it is conceivable that additional amino acid changes would be required for lysines in the H2A NTD to increase genome compaction.

The nucleoskeletal theory proposes that chromatin structure influences the shape of the nucleus and thus is a major determinant of nuclear volume (Cavalier-Smith, 2005), although the amount of DNA *per se* does not affect nuclear volume (Neumann and Nurse, 2007). Non-chromatin components such as nuclear import factors from the cytoplasm may also modulate nuclear size (Levy and Heald, 2010). Our data suggest that in particular cases, the effects of H2A NTD mutations on chromatin compaction are linked to nuclear volume, although not in a straightforward relationship. While arginines at positions 3 and 11 increase chromatin compaction and reduce nuclear volume, lysines at the same positions have no effect on chromatin compaction yet increase nuclear volume. Removal of serines at positions 10 or 15 has

little effect on compaction but also increase nuclear volume. In human cells, expression of all R11 mutants (R11A, C, H, and P) decreased compaction but only R11A also affected nuclear area. Although we do not observe a clear-cut relationship between nuclear volume and chromatin compaction, our data identify a region of the nucleosome that is directly or indirectly linked to nuclear volume control mechanisms.

Since alterations in chromatin structure often cause changes in transcription (Parra and Wyrick, 2007), we were surprised that mutant H2A-containing yeast had very similar gene expression profiles as WT cells, grew at similar rates, did not have altered cell cycle profiles, and were not sensitive to DNA damaging drugs or environmental challenges. These data raise the possibility that H2A-mediated compaction of chromatin may have evolved as a mechanism to enable regulation of chromatin compaction without having to make compensatory changes to all other processes that are also based on DNA such as transcription. Nevertheless, it remains to be determined what molecular or cellular processes govern the optimal level of chromatin compaction and nuclear volume for an organism. The stable alterations of chromatin compaction in eukaryotic model organisms through genetic manipulation of H2A should facilitate further experiments to uncover these processes.

Materials and Methods

Strains and media

The yeast strains used in this study are listed in Supplementary file 1A. Yeast cells were grown in YPD at 30°C unless otherwise noted. C-terminal tagging of yeast proteins was performed as described previously(Longtine et al., 1998). Mammalian cell lines were maintained at 37°C and 5% CO_2 and cultured with 10% fetal bovine serum (FBS) and DMEM (Gibco).

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 (Camacho et al., 2009) 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 (Edgar, 2004a, b). Weblogo3 (Crooks et al., 2004; Schneider and Stephens, 1990) was used for motif discovery. Heat maps for residue positions were constructed using Cluster 3.0 (de Hoon et al., 2004) and Java Treeview(Saldanha, 2004).

Yeast H2A mutagenesis

Site directed mutagenesis was performed using the QuickChange Lightning kit (Agilent) on the pFL142 plasmid. Supplementary file 1B contains all the plasmids that were used and constructed in this study. The sequences of primers are listed in Supplementary file 1C. The correct mutation was verified by sequencing.

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 min 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 5.0 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.

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 min at room temperature, washed twice in PBS, and stained with a 1:50 dilution of concanavalin A conjugated with tetramethylrhodamine (Invitrogen) for 15 min 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.

FISH probes

For yeast FISH analysis, DNA templates for probes 1, 3, and 4 came from ATCC cosmids 71042, 70912, and 70982 as described elsewhere(Guacci et al., 1994). 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, Supplementary file 1C). 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.

Fluorescent *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 min 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 min 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 min 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 min, excess liquid was aspirated away and slides were allowed to air dry for 5 additional min. Slides were washed in methanol for 10 min and then acetone for 30 sec 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 min. 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 min 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 (Bystricky et al., 2004) by finding the pixel distance between weighted centers of the green signal and red signal and converted to nm by the appropriate factor.

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 5.5. DAPI, GFP, mRFP, and Far-red images were acquired by excitation at 360 nm, 488 nm, 561 nm, and 640 nm from a high-speed AOTF laser launch line. A step size of 0.3 (yeast) or 0.5 (human) μm was used for z-stack acquisition.

Micrococcal nuclease digestion

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)(Rando, 2010).

RNA expression analysis

RNA was extracted from exponentially growing yeast as described previously(Schmitt et al., 1990). 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.

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 (Shogren-Knaak et al., 2006). DNA arrays were prepared as described previously (Luger et al., 1999). 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 Supplementary file 1C. Recombinant *X. laevis* histones were expressed in bacteria and purified as described previously (Luger et al., 1999). 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.

Nucleosome array assembly

Recombinant histone octamers and the 601-177-12 DNA template (Lowary and Widom, 1998) 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 $MgCl_2$(Dorigo et al., 2003). 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 1 hr at 37°C, and subsequent analysis using a 5% native polyacrylamide gel(Luger et al., 1999).

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 (Hansen and Turgeon, 1999; Weischet et al., 1978).

Combined immunofluorescence and fluorescent *in-situ* hybridization in human cells

N-terminally HA-tagged WT H2A of *X. laevis* was cloned by PCR into mammalian expression vector pCMV-HA (Clontech) between EcoRI and NotI sites. C-terminally Myc-FLAG-tagged human H2A, in a mammalian expression vector, was obtained from OriGene (RC200688). Site directed mutagenesis was performed using the QuickChange Lightning kit (Agilent) on these expression plasmids. 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 H2A expression plasmids using BioT transfection reagent (Bioland Scientific) or Lipofectamine LTX with Plus reagent (Life Technologies). Cells were grown for 48 hrs post-transfection. For immunofluorescence only, transfected cells were fixed with ice-cold methanol for 15 min at -20°C followed by washing with PBS-T. For combined immunofluorescence and FISH, transfected cells were fixed with 4% paraformaldehyde in PBS for 10 min at room temperature followed by washing with PBS. Cells were then permeabilized in 0.5% Triton X-100 in PBS for 10 min at room temperature followed by washing with PBS. Cells were blocked in 5% BSA and incubated with anti-HA antibody (1:250 dilution, Abcam, ab9110) or anti-FLAG antibody (1:1000 dilution, Sigma F1804). Cells were washed and incubated with secondary antibody (1:500 Alexa Fluor 488 goat anti-rabbit, Invitrogen, A11008, 1:250 Alexa Fluor 647 goat anti-rabbit, Invitrogen, A21245, 1:500 Alexa Fluor 488 goat anti-mouse, Invitrogen, A11001, or 1:100 Alexa Fluor 647 goat anti-mouse, Invitrogen, A21235). 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.

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 min in 4% paraformaldehyde in PBS at room temperature. Cells were immediately put through a cold ethanol dehydration series (5 min each at 85%, 95%, and 100%) and allowed to air dry. Cells were rehydrated in 2x

SSC for 5 min and then RNase-treated for 30 min at 37°C in a humid chamber. Cells were washed with 2x SSC, and denatured at 80°C for 15-20 min 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, imaged, and analyzed as described above. Nuclear staining, in H2A-expressing cells, was used to measure lengths of the long and short orthogonal 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.

Competition assays

Two sets of yeast strains were generated in which Pgk1p was C-terminally fused with either GFP or RFP (Supplementary file 1A). 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.

Cell cycle analysis

Cell cycle analysis of exponentially growing cells was performed essentially as described previously (Zou et al., 1997), except that cells were stained with 1 μM SYTOX Green (Molecular Probes).

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.

Acknowledgements

We acknowledge the support of Martin Phillips and the UCLA-Department of Energy Biochemistry Instrumentation Facility for analytical ultracentrifuge experiments, the Jonsson Comprehensive Cancer Center flow cytometry core facility (supported by grants P30 CA016042 and 5P30 AI028697) and the UCLA Broad Stem Cell Center Sequencing Core. We thank Michael Grunstein and Fred Winston for providing yeast strains TSY107 and FY406, respectively. BM was supported partially by a Howard Hughes Medical Institute Medical Student Fellowship and a Philip Whitcome Pre-doctoral Training Program grant. TS was supported partially by a UCLA Cancer Cell Biology Postdoctoral Training grant. This work was funded by an NIH Director's Innovator Award to SKK.

673 **References**

674

675 Bao, Y., Konesky, K., Park, Y.-J., Rosu, S., Dyer, P.N., Rangasamy, D., Tremethick, D.J.,
676 Laybourn, P.J., and Luger, K. (2004). Nucleosomes containing the histone variant H2A. Bbd
677 organize only 118 base pairs of DNA. The EMBO journal *23*, 3314-3324.

678 Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J., and
679 Woodcock, C.L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural
680 motif that directs the higher-order folding and compaction of chromatin. Proceedings of the
681 National Academy of Sciences *95*, 14173-14178.

682 Bohn, M., and Heermann, D.W. (2010). Diffusion-driven looping provides a consistent
683 framework for chromatin organization. PLoS One *5*, e12218.

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

688 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.
689 (2009). BLAST+: architecture and applications. BMC bioinformatics *10*, 421.

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

693 Caterino, T.L., and Hayes, J.J. (2010). Structure of the H1 C-terminal domain and function in
694 chromatin condensation This paper is one of a selection of papers published in a Special Issue
695 entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has
696 undergone the Journal's usual peer review process. Biochemistry and Cell Biology *89*, 35-44.

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

699 Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova,
700 O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., and Difilippantonio, M.J. (2002). Genomic
701 instability in mice lacking histone H2AX. Science Signaling *296*, 922.

702 Chandrasekharan, M.B., Huang, F., and Sun, Z.-W. (2009). Ubiquitination of histone H2B
703 regulates chromatin dynamics by enhancing nucleosome stability. Proceedings of the National
704 Academy of Sciences *106*, 16686-16691.

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

707 Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent
708 mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. Journal
709 of molecular biology *319*, 1097-1113.

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

712 Di Lorenzo, A., and Bedford, M.T. (2011). Histone arginine methylation. FEBS letters *585*,
713 2024-2031.

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

716 Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004).
717 Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-
718 1573.

719 Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S.,
720 Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature
721 *465*, 363-367.

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

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

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

728 Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y.,
729 Skoultchi, A.I., and Wutz, A. (2010). Ring1B compacts chromatin structure and represses gene
730 expression independent of histone ubiquitination. Molecular cell *38*, 452-464.

731 Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A.,
732 Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global
733 chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199-1212.

734 Fierz, B., Chatterjee, C., McGinty, R.K., Bar-Dagan, M., Raleigh, D.P., and Muir, T.W. (2011).
735 Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. Nature
736 chemical biology *7*, 113-119.

737 Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J.,
738 Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1–chromatin binding by
739 histone H3 methylation and phosphorylation. Nature *438*, 1116-1122.

740 Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R.,
741 Leung, K., and Menzies, A. (2011). COSMIC: mining complete cancer genomes in the
742 Catalogue of Somatic Mutations in Cancer. Nucleic acids research *39*, D945-D950.

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

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

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

749 Guacci, V., Hogan, E., and Koshland, D. (1994). Chromosome condensation and sister
750 chromatid pairing in budding yeast. The Journal of cell biology *125*, 517-530.

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

Hagiwara, T., Hidaka, Y., and Yamada, M. (2005). Deimination of histone H2A and H4 at arginine 3 in HL-60 granulocytes. Biochemistry *44*, 5827-5834.

Hansen, J.C., and Turgeon, C.L. (1999). Analytical ultracentrifugation of chromatin. In Chromatin Protocols (Springer), pp. 127-141.

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

Kim, J., Guermah, M., McGinty, R.K., Lee, J.-S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W., and Roeder, R.G. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell *137*, 459-471.

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

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

Li, J.-Y., Patterson, M., Mikkola, H.K., Lowry, W.E., and Kurdistani, S.K. (2012). Dynamic distribution of linker histone h1. 5 in cellular differentiation. PLoS genetics *8*, e1002879.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., and Dorschner, M.O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. science *326*, 289-293.

Longtine, M.S., McKenzie III, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Lowary, P., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. Journal of molecular biology *276*, 19-42.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nature Reviews Molecular Cell Biology *13*, 436-447.

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

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

Maresca, T.J., Yan, J., Skoko, D., Heald, R., and Marco, J. (2005). Histone H1 is essential for mitotic chromosome architecture and segregation in Xenopus laevis egg extracts. Journal of Cell Biology *169*, 859-869.

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

790 Miloshev, G., Venkov, P., van Holde, K., and Zlatanova, J. (1994). Low levels of exogenous
791 histone H1 in yeast cause cell death. Proceedings of the National Academy of Sciences *91*,
792 11567-11570.

793 Moosmann, A., Campsteijn, C., Jansen, P.W., Nasrallah, C., Raasholm, M., Stunnenberg, H.G.,
794 and Thompson, E.M. (2011). Histone variant innovation in a rapidly evolving chordate lineage.
795 BMC evolutionary biology *11*, 208.

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

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

800 Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C.,
801 Jorgensen, R., Derelle, E., and Rombauts, S. (2007). The tiny eukaryote Ostreococcus provides
802 genomic insights into the paradox of plankton speciation. Proceedings of the National Academy
803 of Sciences *104*, 7705-7710.

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

806 Rando, O.J. (2010). Genome-wide mapping of nucleosomes in yeast. Methods in enzymology
807 *470*, 105-118.

808 Robinson, P.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D.
809 (2008). 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker
810 histone eviction. Journal of molecular biology *381*, 816-825.

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

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

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

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

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

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

823 Schwarz, P.M., and Hansen, J.C. (1994). Formation and stability of higher order chromatin
824 structures. Contributions of the histone octamer. Journal of Biological Chemistry *269*, 16284-
825 16289.

826 Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006).
827 Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*,
828 844-847.

Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., and Hughes, T.R. (2006). Mapping pathways and phenotypes by systematic gene overexpression. Molecular cell *21*, 319-330.

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

Szerlong, H.J., and Hansen, J.C. (2010). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. Biochemistry and Cell Biology *89*, 24-34.

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

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

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

Waldmann, T., Izzo, A., Kamieniarz, K., Richter, F., Vogler, C., Sarg, B., Lindner, H., Young, N.L., Mittler, G.M.J., and Garcia, B.A. (2011). Methylation of H2AR29 is a novel repressive PRMT6 target (Bibliothek der Universität Konstanz).

Wang, H., Huang, Z.-Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B.D., Briggs, S.D., Allis, C.D., Wong, J., and Tempst, P. (2001). Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. Science *293*, 853-857.

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

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

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

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

Zink, D., Fischer, A.H., and Nickerson, J.A. (2004). Nuclear structure in cancer cells. Nature reviews cancer *4*, 677-687.

Zou, L., Mitchell, J., and Stillman, B. (1997). 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.

Figure 1: Histone H2A N-terminal sequence has co-evolved with genome size. Violin plots of the number of (**A**) arginines or (**B**) serines/threonines in the H2A NTD for species with small, medium, and large genomes. Plot widths correspond to species frequency within each group. (**C**) H2A NTD sequences for *S. cerevisiae* and *H. sapiens*. (**D**) Heat map of H2A NTD residue composition at the indicated positions ordered by genome size. Example species are shown with kingdom and genome size information. (**E**) Protein sequence motifs surrounding the four H2A NTD arginine residues. (**F**) Positioning of evolutionarily variable residues relative to the H2A N-terminus (left) or histone fold (right). See also Figure 1 – figure supplements 1 and 2.

Figure 2: 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. (**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. (**E**) 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 if the mean of the indicated strain is significantly different from WT. Red stars denote level of significance: * $p<0.01$; ** $p<0.001$; *** $p<0.0001$ (For exact values, see Supplementary file 2). (**F**) Agarose gel electrophoresis of MNase-digested chromatin in the indicated strains including the densitometric profiles comparing the WT to each of the mutant H2A strains for a given amount of enzyme. See also Figure 2 – figure supplement 1.

Figure 3: Ectopic expression of H2A NTD arginines decreases nuclear volume in yeast. (**A**) Images of the nuclear envelope, as visualized by Nup49p-GFP, and boxplot of the distributions of nuclear volumes in the indicated strains in the TSY107 background (**B**) or the FY406 background (**C**). 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 if the mean of the indicated strain is significantly different from WT. Red stars denote level of significance: * $p<0.01$; ** $p<0.001$; *** $p<0.0001$ (Supplementary file 3). Red dagger (†) indicates that mean nuclear volume of R11ΔS15 is significantly smaller than its isogenic WT strain (ΔS15; $p<0.001$). See also Figure 3 – figure supplement 1.

Figure 4: Loss of H2A NTD arginines decreases chromatin compaction in human cells. (**A**) FISH images of probes on chromosome 1 in normal primary IMR90 fibroblasts with HA-tagged WT or mutant H2A overexpressed as indicated. (**B**) Boxplot of the distributions of inter-probe distances. Note that R11K was only marginally significant at p=0.023. (**C**) Immunofluorescence images of IMR90 cells overexpressing HA-tagged WT or mutant H2A as indicated. (**D**) Top: Boxplot of the distributions of largest nuclear cross-sectional areas in the indicated H2A overexpressing cells. Bottom: Boxplot of the distributions of α-HA fluorescence intensities. (**E**) Left: FISH images, as in (A), of IMR90 cells expressing a C-terminal FLAG-tagged WT or tailless (Δ1-12) H2A. Right: Boxplot of the distributions of inter-probe distances. (**F**) Top: Immunofluorescence images of IMR90 cells overexpressing FLAG-tagged WT or tailless H2A. Bottom: Boxplot of nuclear areas and fluorescence intensities, as indicated. Dashed lines mark the median value for the WT strain. All scale bars are 10 μm. Boxes are colored if the mean of the indicated strain is significantly different from WT. Red stars denote level of significance: * $p<0.01$; ** $p<0.001$; *** $p<0.0001$ (Supplementary files 4 and 5). See also Figure 4 – figure supplement 1.

Figure 5: H2A NTD R11 directly modulates the compaction of chromatin fibers *in vitro*. (**A**) Polyacrylamide gel electrophoresis (PAGE) of ScaI-digested 601-177-12 DNA template assembled with octamers containing recombinant WT or ΔR11 H2A. As a control, 5% of the 601-177-12 DNA without octamers was also digested. (**B**) The distribution of sedimentation coefficients determined by von-Holde Weischet analysis plotted against the percent boundary fraction in the absence or presence of 0.8 mM $MgCl_2$ as indicated. $S_{20°C,W}$ is the sedimentation coefficient corrected to water at 20°C. See also Figure 5 – figure supplement 1.

Figure 6: Mutations to H2A NTD decrease the fitness of yeast. (**A**) Pearson correlations between the global gene expressions of the indicated strains grown in YPD. Correlations are calculated from an average of at least two experiments. (**B**) Growth curves of the indicated H2A yeast strains over 10 hrs in YPD. (**C**) Spot tests with ten-fold serial dilutions for the indicated strains in the presence of different drugs. (**D**) The proportion of yeast cells in a co-culture of WT and the indicated mutant H2A carrying Pgk1 gene fusion to GFP (green) or RFP (red) as indicated by color.

Figure 7: Mutations of H2A NTD found in cancers decreases chromatin compaction in human cells. (**A**) Schematic of the H2A NTD showing only the mutations within the arginine motifs found in various cancers as indicated by the colored shapes (Forbes et al., 2011). The letter within each shape represents the mutated amino acid. (**B**) FISH images of probes on chromosome 1 in normal primary IMR90 fibroblasts with HA-tagged WT or mutant H2A overexpressed as indicated. (**C**) Boxplot of the distributions of inter-probe distances. (**D**) Immunofluorescence images of IMR90 cells overexpressing HA-tagged WT or mutant H2A as indicated. Anti-HA primary and Alexa Fluor 647-conjugated secondary antibodies were used to determine expression in FISH images and for measurement of nuclear areas. (**E**) Top: Boxplot of the distributions of largest nuclear cross-sectional areas in the indicated H2A overexpressing cells. (**F**) Boxplot of the distributions of α-HA fluorescence intensities. Dashed lines mark the median value for the WT strain. All scale bars are 10 μm. Boxes are colored if the mean of the indicated strain is significantly different from WT. Red stars denote level of significance: * $p<0.01$; *** $p<0.0001$ (Supplementary files 4 and 5).

Table 1: List of H2A mutations, sequence changes and their effects on chromatin compaction and nuclear volume.

Yeast					
		FISH		Nuclear volume	
H2A Mutant	H2A NTD Protein sequence	% change	p-val	% change	p-val
WT	SG-GKG-GKAGSA-AKASQSRSAKAG	-	1.0E+00	-	1.0E+00
R3	SG**R**GKG-GKAGSA-AKASQSRSAKAG	**-18**	9.5E-04	-5	4.1E-01
R11	SG-GKG-GKAGSA**R**AKASQSRSAKAG	**-15**	8.6E-04	**-20**	5.9E-05
R3R11	SG**R**GKG-GKAGSA**R**AKASQSRSAKAG	**-22**	8.2E-06	**-16**	3.0E-03
R3(ΔGS10)R11	SG**R**GKG-GKA··A**R**AKASQSRSAKAG	**-30**	2.1E-06	+6	3.7E-01
R11ΔS15	SG-GKG-GKAGSA**R**AKA·QSRSAKAG	**-41**	3.9E-08	**-9**†	4.7E-04
K3	SG**K**GKG-GKAGSA-AKASQSRSAKAG	+9	8.6E-01	**+13**	9.4E-03
K11	SG-GKG-GKAGSA**K**AKASQSRSAKAG	+16	3.1E-01	+3	2.6E-01
K3K11	SG**K**GKG-GKAGSA**K**AKASQSRSAKAG	+6	8.3E-01	**+31**	5.4E-08
K11ΔS15	SG-GKG-GKAGSA**K**AKA·QSRSAKAG	-7	9.2E-02	+2	6.6E-01
ΔGS10	SG-GKG-GKA··A-AKASQSRSAKAG	-6	3.2E-02	+10	3.0E-02
ΔS15	SG-GKG-GKAGSA-AKA·QSRSAKAG	+3	9.4E-02	**+9**	9.7E-03
R6	SG-GKG**R**GKAGSA-AKASQSRSAKAG	-5	5.6E-02	**+10**	1.0E-03
K20R	SG-GKG-GKAGSA-AKASQSRSA**R**AG	-3	3.0E-01	+7	1.5E-02
R17K	SG-GKG-GKAGSA-AKASQS**K**SAKAG	-1	7.9E-01	0	2.7E-01
Human – HA Tag					
WT	SGRGKQGGKTRAKAKSRSSRAG	-	1.0E+00	-	1.0E+00
ΔR3	SG·GKQGGKTRAKAKSRSSRAG	**+39**	8.3E-03	**+42**	1.3E-08
R11K	SGRGKQGGKT**K**AKAKSRSSRAG	**+20**	2.3E-02	**+14**	1.2E-03
R11A	SGRGKQGGKT**A**AKAKSRSSRAG	**+43**	1.0E-05	**+21**	5.7E-07
ΔR3R11A	SG·GKQGGKT**A**AKAKSRSSRAG	**+35**	3.5E-03	**+18**	5.9E-04
Human – FLAG Tag					
WT	SGRGKQGGKARAKAKSRSSRAG	-	1.0E+00	-	1.0E+00
Δ1-12	·····KAKSRSSRAG	**+47**	4.9E-03	**+18**	2.8E-04

- The – marks indicate spacing for sequence alignment purposes. The inserted residues are bold typed and underlined. Deletions are indicated by ·.
- Percent (%) change refers to the difference in median values relative to WT unless otherwise indicated; the statistically significant differences are bold typed. P-values were calculated using the t-test (yeast) and Mann-Whitney U test (human).
- †percent change was calculated relative to isogenic WT control (ΔS15).

Figure 1 – figure supplement 1: Phylogenetic distribution of species analyzed in this paper. (**A**) The bar graph indicates the number and proportion of organisms in our dataset that belong to the indicated phylogenetic kingdoms for each genome size category. (**B**) Violin plots of the number of lysines in the H2A NTD grouped by genome size as in figure 1. (**C**) Boxplot of genome sizes of the organisms which have an H2A without (left) or with (right) the indicated residue. P-values (Mann-Whitney U test) of the difference in means between the absence or presence of the indicated residue are indicated.

Figure 1 – figure supplement 2: H2A arginines 3 and 11 are situated adjacent to DNA within the nucleosome. Structure of a di-nucleosome obtained from the crystal lattice of a mono-nucleosome structure is shown from (**A**) the side or (**B**) close up highlighting potential intra- and inter-nucleosomal 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 (Davey et al., 2002).

Figure 2 – figure supplement 1: Ectopic expression of H2A NTD arginines causes compaction in yeast. (**A-C**). Boxplot of the distributions of interprobe distances in the indicated H2A mutant strains for probe sets B, C, and D as shown in figure 3A. Boxes are colored if the mean of the indicated strain is significantly different from WT. Red stars denote level of significance: * $p<0.01$; ** $p<0.001$; *** $p<0.0001$ (Supplementary file 2). Cell cycle analysis of yeast strains in the TSY107 (**D**) or FY406 (**E**) background. 1C and 2C refer to G1 and G2 DNA content, respectively. Note that the WT strains (TSY107, FY406) carry one copy of the H2A gene on a plasmid with the two chromosomal copies deleted (Hirschhorn et al., 1995; Schuster et al., 1986). Dosage alterations of the H2A protein cause G2/M arrest (Sopko et al., 2006) which is evident in our WT strains. Nonetheless, all the mutants display highly similar cell cycle profiles. (**F**) Agarose gel electrophoresis of MNase-digested chromatin in the indicated strains. The amount of enzyme used to digest chromatin is indicated. Also shown are the densitometric profiles of the agarose gel that compares WT to each of the indicated mutant H2A strains for a given amount of enzyme.

Figure 3 – figure supplement 1: H2A arginines do not affect cell size. (**A**) Images of the cell wall, as visualized by concanavalin A staining, and (**B**) boxplot of the distributions of cellular volumes in the indicated yeast strains. The scale bar is 5 μm. Boxes are colored if the mean of the indicated strain is significantly different from WT. Red stars denote level of significance: * $p<0.01$; ** $p<0.001$; *** $p<0.0001$.

Figure 4 – figure supplement 1: Loss of H2A NTD arginines decreases chromatin compaction in human cells. (**A**) FISH images of probes on chromosome 1 in MDA-MB-453 cells with either WT or mutant HA-tagged H2A overexpressed. (**B**) Boxplot of the distributions of interprobe distances. Immunofluorescence images of (**C**) MDA-MB-453 or (**E**) HEK293 cells overexpressing WT or mutant HA-tagged H2A. Boxplot of the distributions of largest nuclear cross-sectional areas in (**D**) MDA-MB-453 or (**F**) HEK293 for the indicated H2A over-expressing cells. (**G**) Western blot of lysates from HEK293 cells overexpressing the indicated WT or mutant HA-tagged H2A. All scale bars are 10 μm. Dashed lines mark the median value for the WT strain. Boxes are colored if the mean of the indicated strain is significantly different from WT. Red stars denote level of significance: * $p<0.01$; *** $p<0.0001$.

Figure 5 – figure supplement 1: H2A NTD R11 directly modulates the compaction of chromatin fibers *in vitro*. The distribution of sedimentation coefficients determined by von-Holde Weischet analysis plotted against the percent boundary fraction in absence or presence of 0.6 mM $MgCl_2$ as indicated. $S_{20°C,W}$ is the sedimentation coefficient corrected to water at 20°C.

Supplementary Files

Supplementary File 1 - Tables of yeast strains, plasmids, and primers

Supplementary File 2 - Table of yeast FISH results

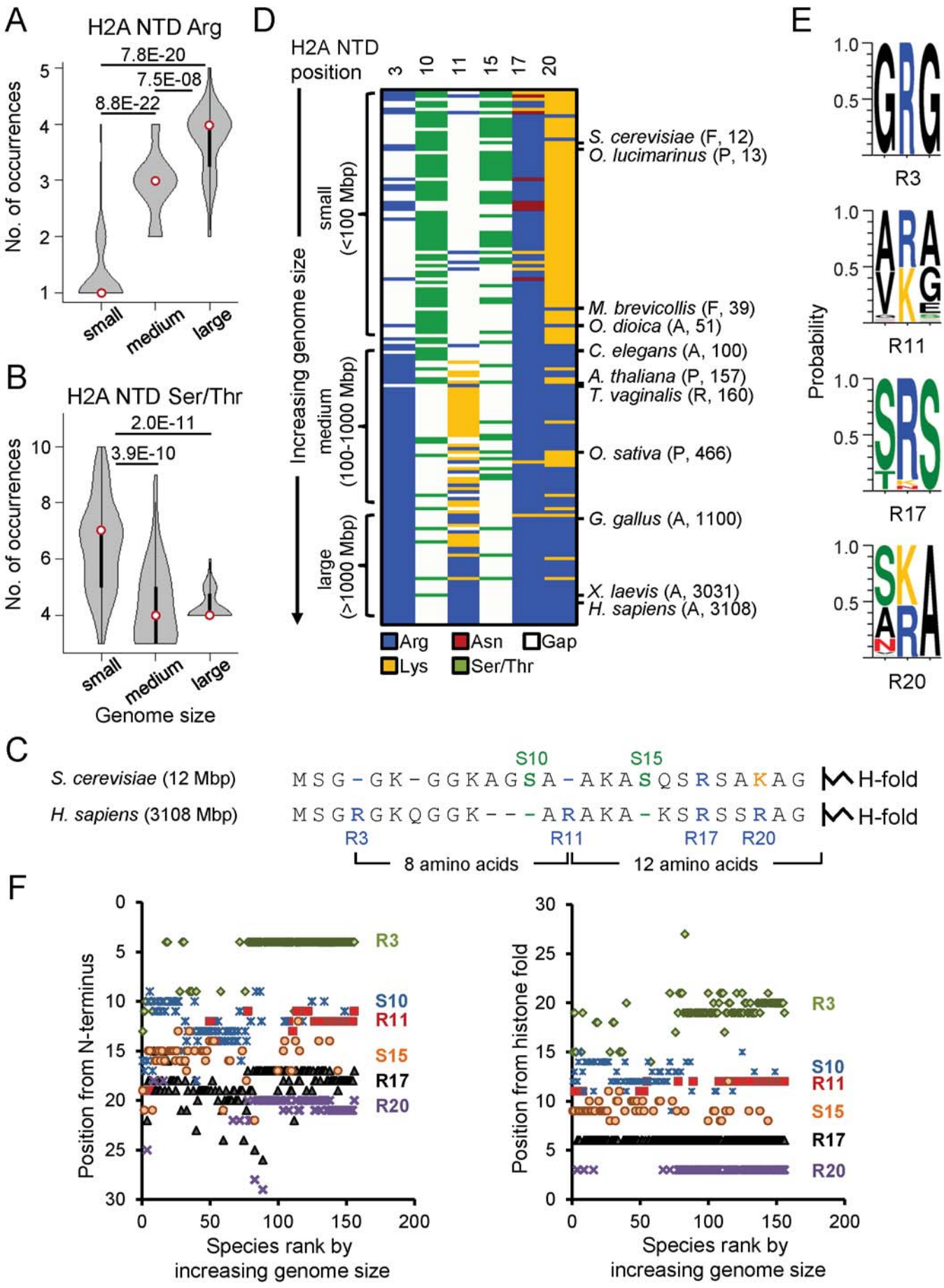
Supplementary File 3 - Table of yeast nuclear and cellular volumes

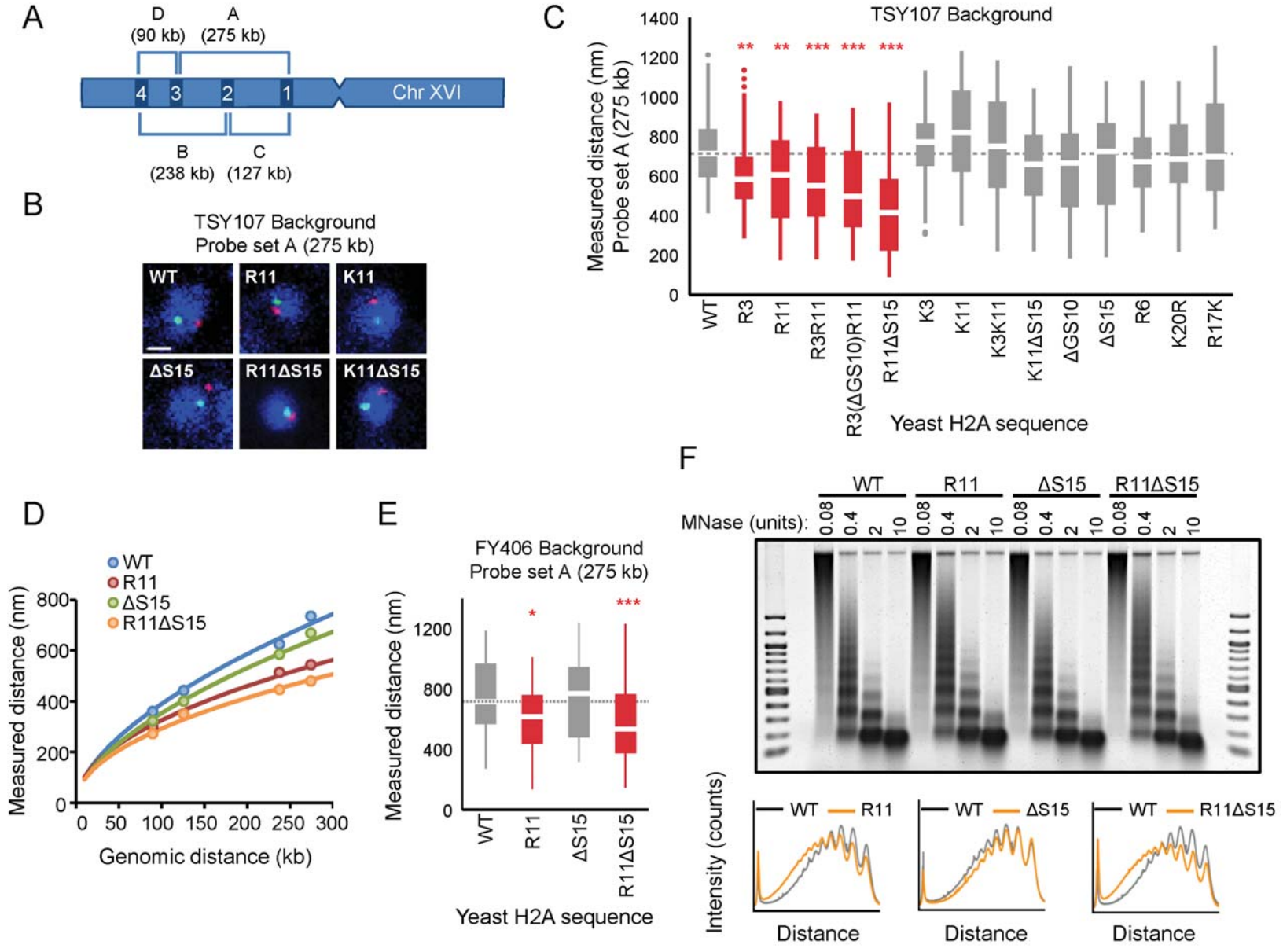
Supplementary File 4 - Table of human FISH results

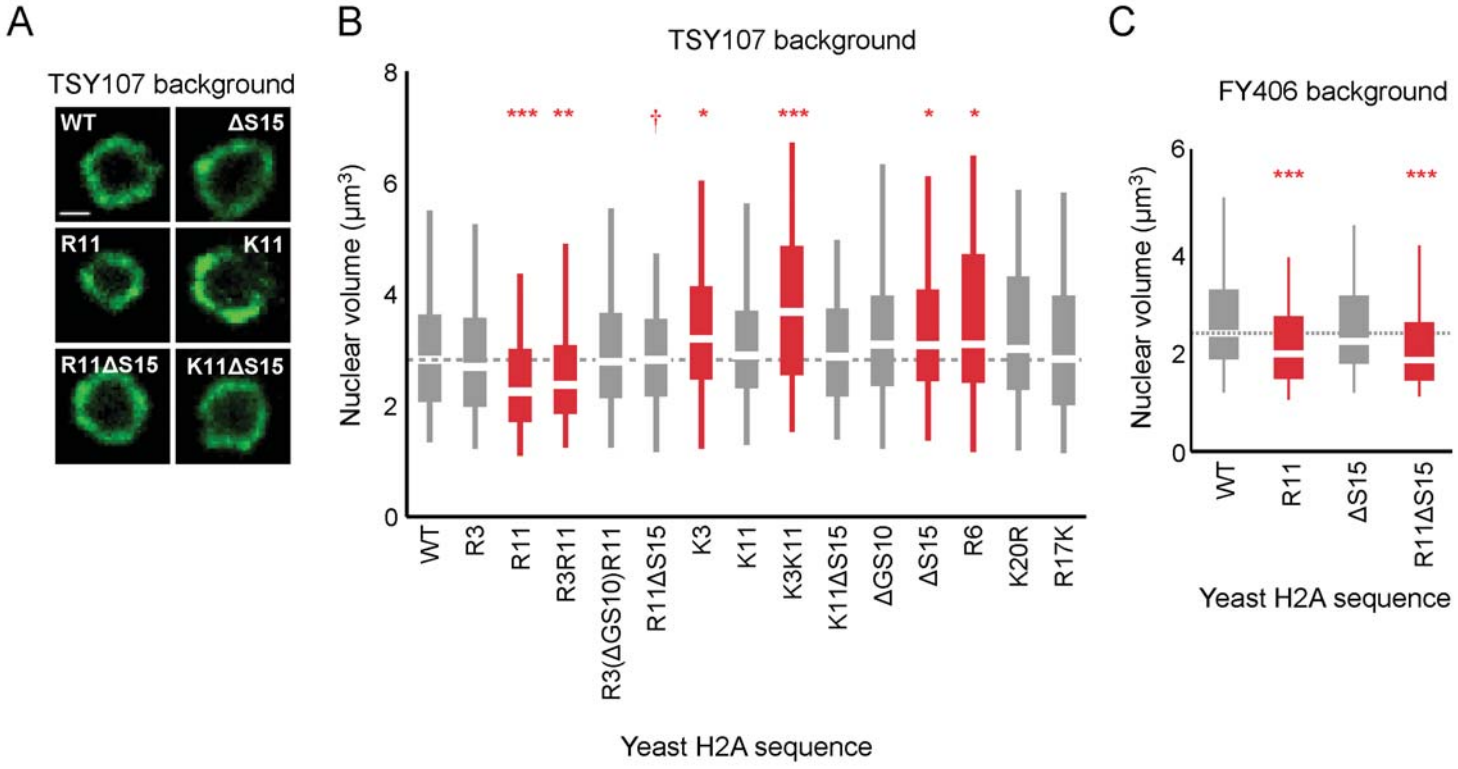
Supplementary File 5 - Table of human nuclear area results

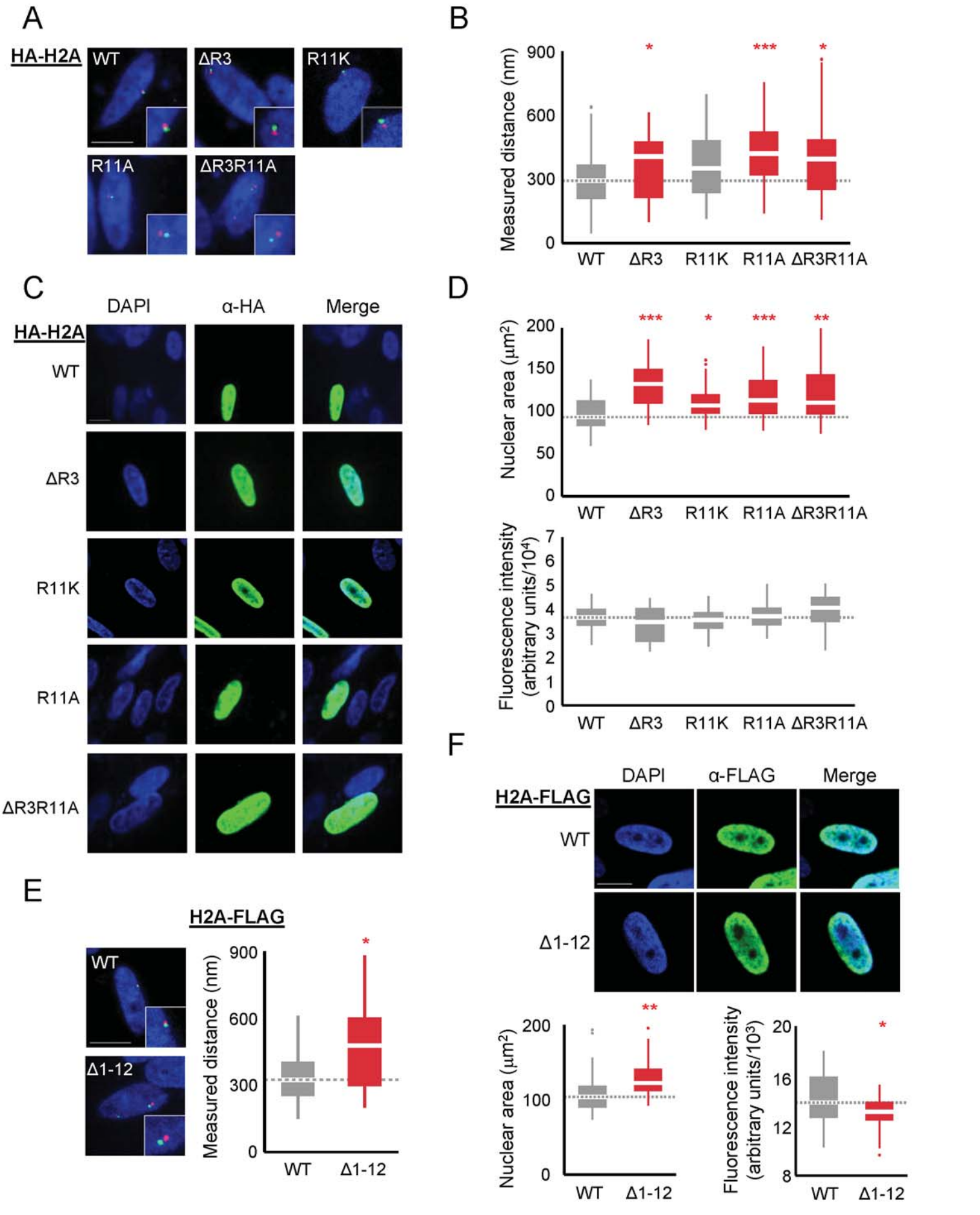
Source Data Files

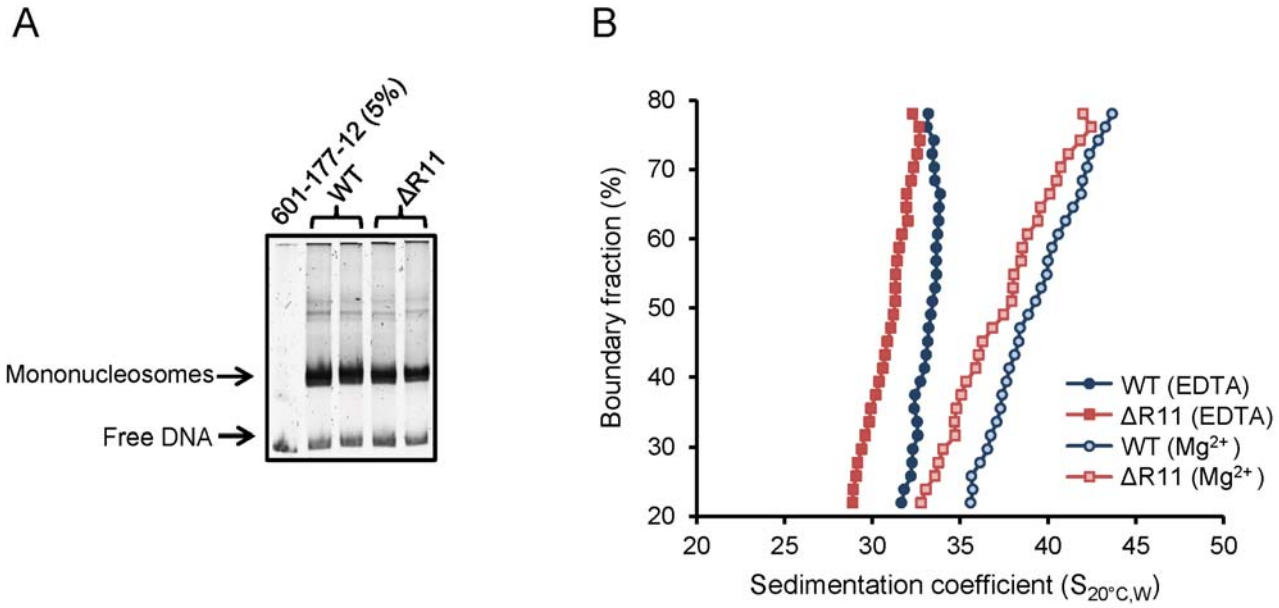
Figure 1 - source data file 1 – H2A multiple sequence alignments, heat map data, and canonical H2A isoforms.

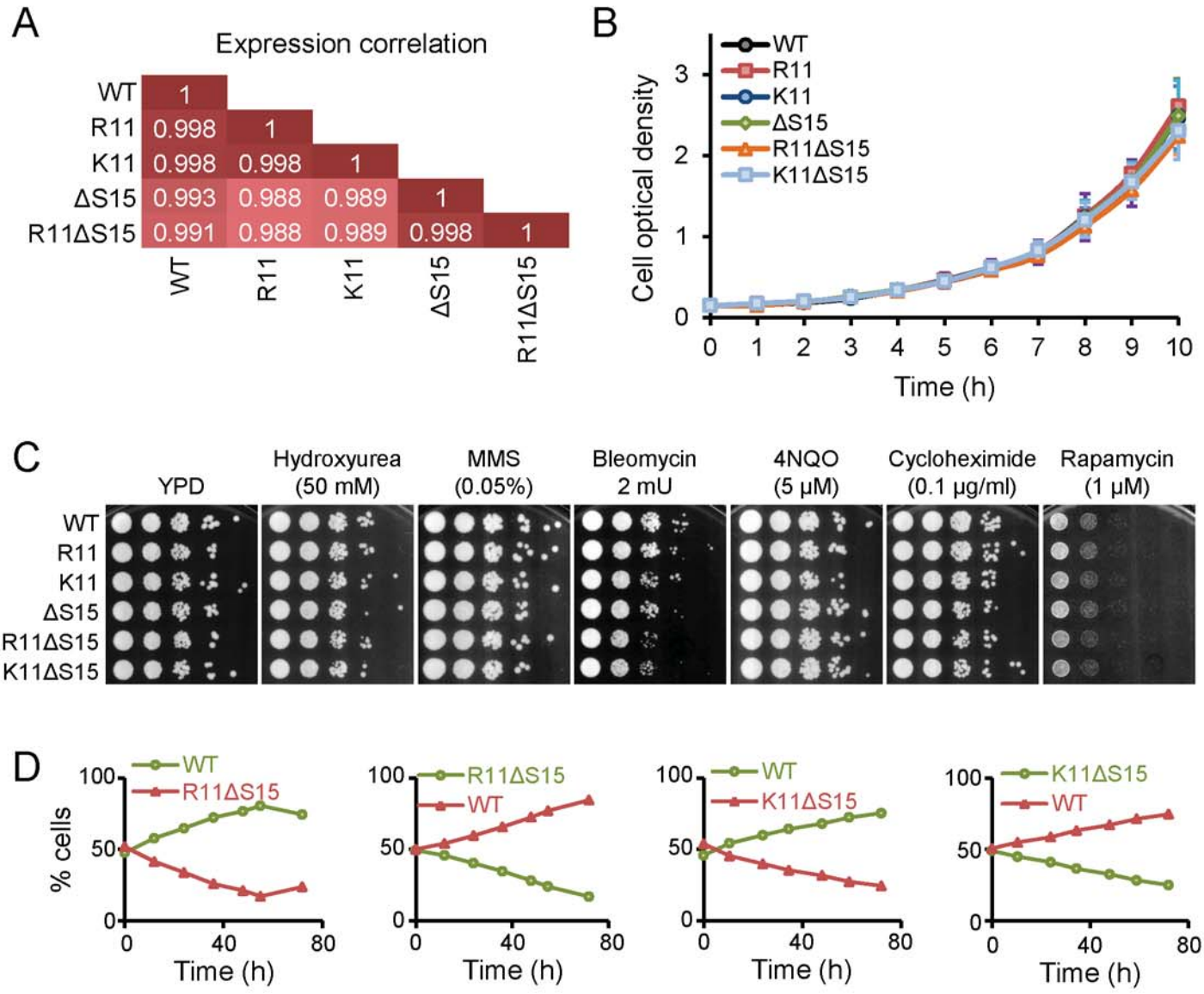

A
H2A NTD Arg
7.8E-20
7.5E-08
8.8E-22
No. of occurrences
small
medium
large
B
H2A NTD Ser/Thr
2.0E-11
3.9E-10
No. of occurrences
small
medium
large
Genome size
D
H2A NTD position
3
10
11
15
17
20
Increasing genome size
small (<100 Mbp)
medium (100-1000 Mbp)
large (>1000 Mbp)
S. cerevisiae (F, 12)
O. lucimarinus (P, 13)
M. brevicollis (F, 39)
O. dioica (A, 51)
C. elegans (A, 100)
A. thaliana (P, 157)
T. vaginalis (R, 160)
O. sativa (P, 466)
G. gallus (A, 1100)
X. laevis (A, 3031)
H. sapiens (A, 3108)
Arg
Asn
Gap
Lys
Ser/Thr
E
Probability
R3
R11
R17
R20
C
S10
S15
S. cerevisiae (12 Mbp)
M S G – G K – G G K A G S A – A K A S Q S R S A K A G
H-fold
H. sapiens (3108 Mbp)
M S G R G K Q G G K – – – A R A K A – K S R S S R A G
H-fold
R3
R11
R17
R20
8 amino acids
12 amino acids
F
Position from N-terminus
R3
S10
R11
S15
R17
R20
Species rank by increasing genome size
Position from histone fold
R3
S10
R11
S15
R17
R20
Species rank by increasing genome size

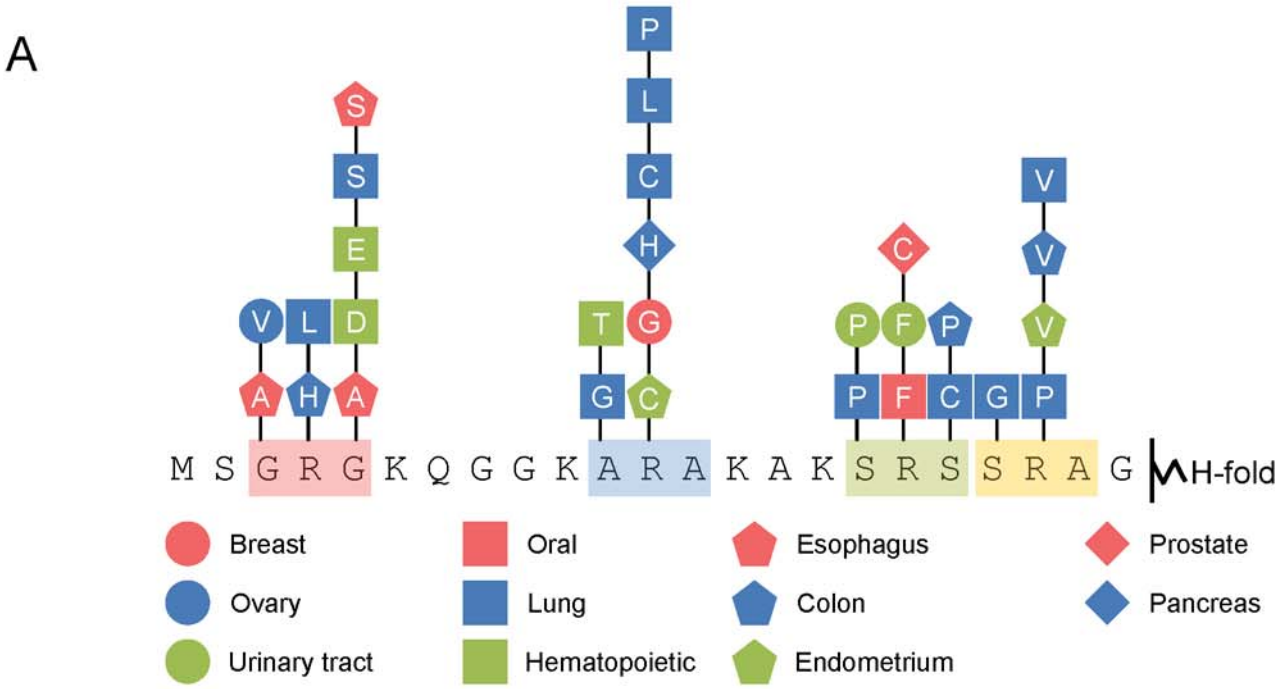

A
D
(90 kb)
A
(275 kb)
4
3
2
1
Chr XVI
B
(238 kb)
C
(127 kb)
B
TSY107 Background
Probe set A (275 kb)
WT
R11
K11
ΔS15
R11ΔS15
K11ΔS15
C
TSY107 Background
Measured distance (nm)
Probe set A (275 kb)
1400
1200
1000
800
600
400
200
0
**
**
***
***
***
WT
R3
R11
R3R11
R3(ΔGS10)R11
R11ΔS15
K3
K11
K3K11
K11ΔS15
ΔGS10
ΔS15
R6
K20R
R17K
Yeast H2A sequence
D
WT
R11
ΔS15
R11ΔS15
Measured distance (nm)
800
600
400
200
0
0
50
100
150
200
250
300
Genomic distance (kb)
E
FY406 Background
Probe set A (275 kb)
Measured distance (nm)
1200
800
400
0
*
***
WT
R11
ΔS15
R11ΔS15
Yeast H2A sequence
F
WT
R11
ΔS15
R11ΔS15
MNase (units):
0.08
0.4
2
10
Intensity (counts)
WT
R11
WT
ΔS15
WT
R11ΔS15
Distance
Distance
Distance

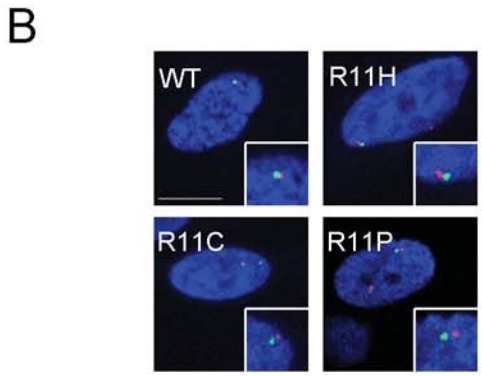

A
TSY107 background
WT
ΔS15
R11
K11
R11ΔS15
K11ΔS15
B
TSY107 background
Nuclear volume (μm³)
0
2
4
6
8
*** ** † * *** * *
WT
R3
R11
R3R11
R3(ΔGS10)R11
R11ΔS15
K3
K11
K3K11
K11ΔS15
ΔGS10
ΔS15
R6
K20R
R17K
Yeast H2A sequence
C
FY406 background
Nuclear volume (μm³)
0
2
4
6
*** ***
WT
R11
ΔS15
R11ΔS15
Yeast H2A sequence

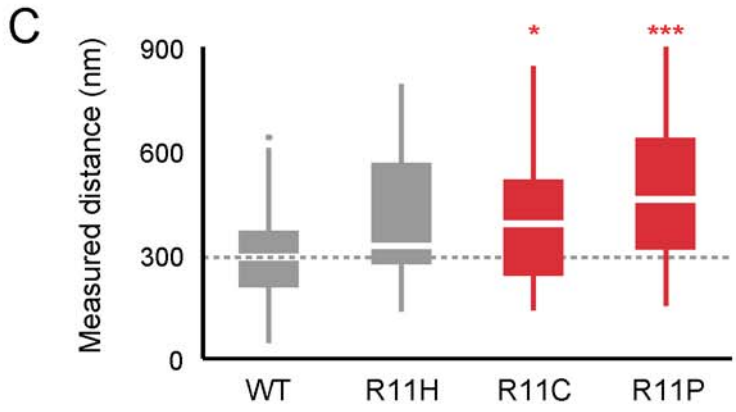

A
HA-H2A
WT
ΔR3
R11K
R11A
ΔR3R11A
B
Measured distance (nm)
900
600
300
0
*
***
*
WT
ΔR3
R11K
R11A
ΔR3R11A
C
DAPI
α-HA
Merge
HA-H2A
WT
ΔR3
R11K
R11A
ΔR3R11A
D
Nuclear area (μm²)
200
150
100
50
0
***
*
***
**
WT
ΔR3
R11K
R11A
ΔR3R11A
Fluorescence intensity (arbitrary units/10⁴)
7
6
5
4
3
2
1
0
WT
ΔR3
R11K
R11A
ΔR3R11A
E
H2A-FLAG
WT
Δ1-12
Measured distance (nm)
900
600
300
0
*
WT
Δ1-12
F
DAPI
α-FLAG
Merge
H2A-FLAG
WT
Δ1-12
Nuclear area (μm²)
200
100
0
**
WT
Δ1-12
Fluorescence intensity (arbitrary units/10³)
20
16
12
8
*
WT
Δ1-12

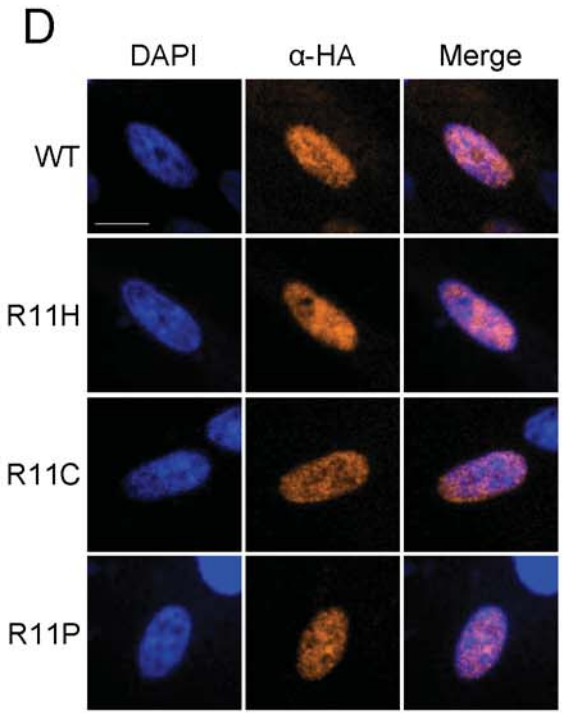

A
601-177-12 (5%)
WT
ΔR11
Mononucleosomes
Free DNA
B
Boundary fraction (%)
80
70
60
50
40
30
20
20
25
30
35
40
45
50
Sedimentation coefficient ($S_{20°C,W}$)
WT (EDTA)
ΔR11 (EDTA)
WT (Mg^{2+})
ΔR11 (Mg^{2+})

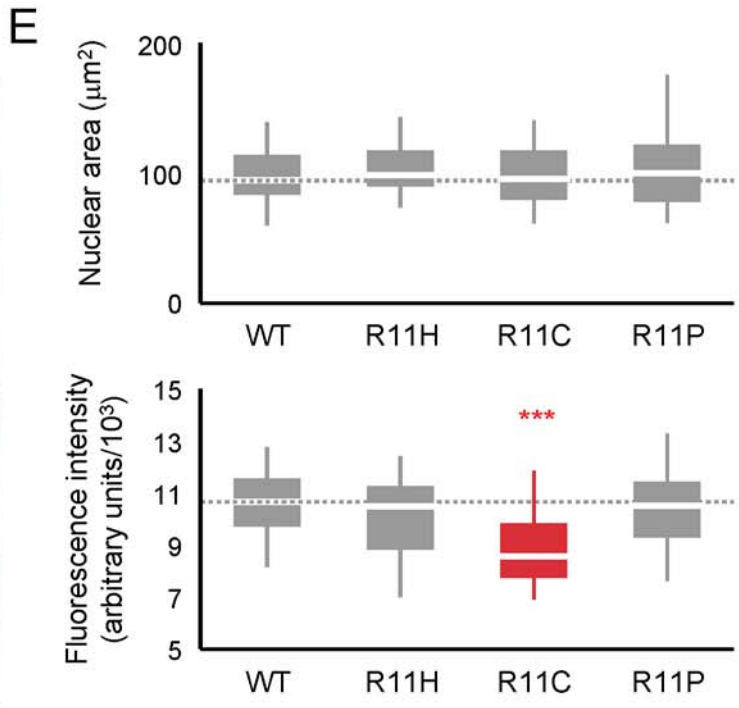

A
Expression correlation
WT 1
R11 0.998 1
K11 0.998 0.998 1
ΔS15 0.993 0.988 0.989 1
R11ΔS15 0.991 0.988 0.989 0.998 1
WT
R11
K11
ΔS15
R11ΔS15
B
WT
R11
K11
ΔS15
R11ΔS15
K11ΔS15
Cell optical density
Time (h)
C
YPD
Hydroxyurea (50 mM)
MMS (0.05%)
Bleomycin 2 mU
4NQO (5 μM)
Cycloheximide (0.1 μg/ml)
Rapamycin (1 μM)
WT
R11
K11
ΔS15
R11ΔS15
K11ΔS15
D
% cells
WT
R11ΔS15
R11ΔS15
WT
WT
K11ΔS15
K11ΔS15
WT
Time (h)


A
M S G R G K Q G G K A R A K A K S R S S R A G
H-fold
Breast
Ovary
Urinary tract
Oral
Lung
Hematopoietic
Esophagus
Colon
Endometrium
Prostate
Pancreas


B
WT
R11H
R11C
R11P


C
Measured distance (nm)
900
600
300
0
*
***
WT
R11H
R11C
R11P


D
DAPI
α-HA
Merge
WT
R11H
R11C
R11P


E
Nuclear area (μm²)
200
100
0
WT
R11H
R11C
R11P
Fluorescence intensity (arbitrary units/10³)
15
13
11
9
7
5
***
WT
R11H
R11C
R11P