7176–7189 Nucleic Acids Research, 2012, Vol. 40, No. 15 Published online 9 May 2012
doi:10.1093/nar/gks351

Chromatin architectures at fission yeast transcriptional promoters and replication origins

Robert M. Givens[1], William K. M. Lai[2], Jason M. Rizzo[2], Jonathan E. Bard[2], Piotr A. Mieczkowski[3], Janet Leatherwood[4], Joel A. Huberman[1,*] and Michael J. Buck[2,*]

[1]Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY 14263, [2]Department of Biochemistry and Center of Excellence in Bioinformatics and Life Sciences, SUNY Buffalo, 701 Ellicott St., Buffalo, NY 14203, [3]Department of Genetics, School of Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599 and [4]Department of Molecular Genetics and Microbiology, SUNY Stony Brook, Stony Brook, NY 11794, USA

Received February 25, 2012; Revised March 29, 2012; Accepted April 9, 2012

ABSTRACT

We have used micrococcal nuclease (MNase) digestion followed by deep sequencing in order to obtain a higher resolution map than previously available of nucleosome positions in the fission yeast, *Schizosaccharomyces pombe*. Our data confirm an unusually short average nucleosome repeat length, ~152 bp, in fission yeast and that transcriptional start sites (TSSs) are associated with nucleosome-depleted regions (NDRs), ordered nucleosome arrays downstream and less regularly spaced upstream nucleosomes. In addition, we found enrichments for associated function in four of eight groups of genes clustered according to chromatin configurations near TSSs. At replication origins, our data revealed asymmetric localization of pre-replication complex (pre-RC) proteins within large NDRs—a feature that is conserved in fission and budding yeast and is therefore likely to be conserved in other eukaryotic organisms.

INTRODUCTION

In eukaryotic organisms, DNA is complexed with histones and other specialized proteins to form chromatin fibers, which in turn are folded into chromosomes. Assembly into chromatin serves to linearly compact the DNA, allowing it to fit within the nucleus, protects it from damage and regulates access to genetic information. The basic structural unit of eukaryotic chromatin is the nucleosome, which consists of ~147 bp of DNA wrapped around a histone octamer, two each of the four core histones—H2A, H2B, H3 and H4 (1–3).

We were interested in relationships between chromatin structure and control of both gene expression and replication initiation in the fission yeast, *Schizosaccharomyces pombe*, which is an important eukaryotic model organism. We wanted to know whether similar chromatin architectures at transcriptional start sites (TSSs) are used for genes encoding proteins involved in similar functions. We also wanted to learn whether DNA replication origins have distinctive chromatin architecture.

To obtain information about chromatin structure, we employed cryo-grinding to liberate well-preserved nuclei from inside cell walls. Then we used micrococcal nuclease (MNase) to cut chromatin between nucleosomes, and we precisely identified the 5′-ends of the resulting mono-nucleosome-sized DNA fragments by next-generation sequencing (4).

Our new higher resolution data support and extend some aspects of the findings of a microarray-based analysis recently reported by Lantermann *et al.* (5,6), including the unusually short inter-nucleosome repeat distance [originally described by Godde and Widom (7)] and ordered arrays of nucleosomes over transcribed DNA with less regular spacing upstream. In contrast, our computer-aligned, deep-sequencing results reveal a more defined and elaborate composite profile for chromatin near fission yeast DNA replication origins. Aligned chromatin profiles for TSSs and replication origins both exhibit potentially fundamental similarities to, and intriguing differences from, their counterparts in the budding yeast, *Saccharomyces cerevisiae*.

Among our unique findings is that clustering analysis of our high-resolution chromatin profiles at TSSs resolves sets of genes with distinctly associated biological functions. We have also discovered that, at fission yeast replication origins as in budding yeast (8,9), the proteins of the pre-replication complex (pre-RC), which are needed to

*To whom correspondence should be addressed. Email: joel.huberman@post.harvard.edu
Correspondence may also be addressed to Michael J. Buck. Tel: +716 881 7569; Email: mjbuck@buffalo.edu

initiate DNA replication, are usually bound to DNA at one side of a large nucleosome-depleted region (NDR).

MATERIALS AND METHODS

Cell growth, nuclear isolation by cryogrinding and preparation of MNase-cut DNA fragments

For a detailed description of these methods, see Givens *et al*. (4). Here we provide a summary. We used *S. pombe* strain D18 (10). The cells were grown in minimal medium [EMM (11)] at 25°C. In some cases, cells were fixed with 1.5% formaldehyde for 15 min prior to harvest. Cells were flash-frozen, then ground in liquid nitrogen.

When ready, each gram of ground material was slowly mixed at 0°C into 4 ml of *In-Nucleo* Chromatin Analysis (INCA) buffer (1.2 M Sorbitol, 100 mM NaCl, 50 mM HEPES-pH 7.4, 5 mM $CaCl_2$, 2 mM $MgCl_2$, with 1 mM 2-Mercapto-EtOH and 0.5 mM spermidine added immediately prior to use, all at 0°C). Intact nuclei and nucleus-bearing cell fragments were then separated from intact cells, aggregated cell debris, small cell debris and soluble material by differential centrifugation. The final washed, crude nuclear pellets at 0° were resuspended in 1 ml of INCA buffer per 5×10^9 initial cell equivalents at 25°C.

Then an equal volume of INCA buffer at 25°C containing 300 U/ml MNase was added, incubation was continued for 12 min, digestion was terminated and DNA was purified. Mono-nucleosome-sized DNA was recovered by preparative gel electrophoresis. DNA was dissolved in TE, pH 8.0, 40 µl per original 1×10^{10}–4×10^{10} cell equivalents.

Sequencing and sequence processing

Overall, we prepared four data sets from both log phase and stationary phase growth: log-fixed narrow (LFN), log-fixed broad (LFB), log-unfixed narrow (LUN) and stationary-unfixed narrow (SUN). Nucleosome DNA was sequenced by an Illumina Genome Analyzer IIx as described previously (12–14). Sequencing reads were aligned to the August, 2008, build of the *S. pombe* genome (http://www.sanger.ac.uk/Projects/S_pombe/) using the BowTie algorithm (15) allowing only unique matches with up to two mismatches.

We processed the data by three methods:

(1) MNase protection was calculated for the data set of aligned sequence reads by extrapolating (extending) each read to a final length of 120 nt, then adding up the number of extended reads crossing each position in the genome and dividing the number of reads at each position by the genome-wide average read count per base pair. This ratio was converted into continuous space (log_2 ratio). Each data set was then standardized at single-base-pair resolution to have a mean occupancy of 0 and a standard deviation of 1 (Z-score statistic).
(2) Individual nucleosome calls were made using the template-filtering algorithm (14) with default parameters, the seven standard defined templates and a minimum and maximum allowable nucleosome width of 80 bp and 200 bp, respectively.
(3) Conditional positioning probability was calculated as described by (16), using the program ArchTEx (17). The tag counts were first smoothed using a Gaussian kernel and absolute position was determined at 1-bp resolution. Conditional positioning was then determined by using a 140-bp sliding window across the genome, with each conditional score determined as the smoothed read count at the center of the window divided by the sum of the smoothed-read-count values within the window at 1-bp resolution (16).

Mappable regions of the genome were determined by generating a data set containing all possible 36-bp sequences (12.5 million 36-mer sequences) and aligning them back to the *S. pombe* reference genome with BowTie using the same parameters as our experimental data sets. The resulting alignment files distinguish those start coordinates that can be aligned uniquely in the *S. pombe* genome (mappable, indicated by '1') from those that cannot (unmappable, indicated by '0'). The files (called 'Bowtie Alignment Files') are available for each of our experiments at http://www.acsu.buffalo.edu/~mjbuck/Fission_Yeast_chromatin.html.

Cross-correlation coefficients for all of the sequencing data sets were calculated by determining the Pearson correlation between the forward and reverse sequence read counts at 1-bp resolution using the program ArchTEx (17). The reverse reads were shifted 1 bp upstream, and the cross correlation was calculated between the forward and reverse reads for all distances up to 1 kb between forward and reverse reads.

Clustering

K-medioid clusterings were carried out using Cluster (18,19) using $k = 8$ (Figure 3) or $k = 10$ (Supplementary Figure S9), with Euclidean distance as the similarity metric. For Figure 3, the analysis was performed using windows of −500 bp to +500 bp around TSSs. For Supplementary Figure S9, in addition to full weight being given to a window of −500 bp to +500 bp around each origin center, the next 350 nucleotides on both sides were assigned weights of 0.5, and the outermost 150 nucleotides on both sides were assigned weights of 0. Raw data from Cluster were imported into Java TreeView (20) in order to generate the heat maps and scales.

ArchAlign DNA replication origin alignment

Alignment of MNase protection of origin centers was performed using the chromatin architecture alignment algorithm, ArchAlign (21). Alignment was performed on log2-ratio, 120-bp-extrapolated data from the LFN sample on 2001-bp windows at 10-bp resolution with a shifting window size of 1500 bp and reversals enabled.

Data availability

All of our raw data are available in the NCBI GEO database, accession number GSE28071. Our processed

data are also available at http://www.acsu.buffalo.edu/~mjbuck/Fission_Yeast_chromatin.html. In addition to raw data, processed data, including relative MNase-protection profiles and template-filtered nucleosome positions for each of the three chromosomes are provided. This website also offers scalable graphs for each chromosome showing MNase-protection profiles, histone H3 amounts (22,23), pre-RC protein amounts (24) and probability of *in vitro* nucleosome formation (12) at each nucleotide position in the genome. In addition, all template-filtered nucleosome positions have been integrated directly into the Pombe community database Pombase (http://www.pombase.org/).

RESULTS

Experimental strategy

We used MNase digestion with next-generation sequencing (MNase-Seq) to map chromatin structure in *S. pombe* during both log and stationary phase culture. To accomplish this, we developed a cryofixation method to protect chromatin integrity during isolation in the absence of treatment with a covalent crosslinking agent, such as formaldehyde (4). This allowed us to compare formaldehyde-fixed and unfixed samples. Briefly, we flash-froze cells in liquid nitrogen (−196°C), then broke the cell walls at the same low temperature in a precision-adjustable motorized mortar-and-pestle grinding device. The nuclei were then treated with MNase in a specially formulated buffer which approximates the physiological and hydration state of the nucleus. The resulting nucleosomal ladder was then separated on an agarose gel, and the mono-nucleosome-size band was excised and sequenced using an Illumina Genome Analyzer II. A narrow band (150–220 bp), or a broad band (100–300 bp), was excised. Four samples were sequenced (Supplementary Figure S1): LFN, LFB, LUN and SUN. Control incubations without MNase demonstrated the absence both of endogenous nuclease activity and of significant mechanical shearing by the cryogrinding procedure (Supplementary Figure S1B). We suspect that the smears between the bands of the nucleosome ladders are due to irregular spacings between nucleosomes in some portions of the genome, combined with the fact that we intentionally did not digest to completion.

Nucleosome *positions* derived from all of these samples were very similar (Supplementary Figure S2A). Nevertheless, both biological variations (such as cell growth rate) and technical variations (such as extent of MNase digestion (25) or usage of formaldehyde fixation (4)) may have contributed to the observed deviations in the *extent of MNase protection* at each position (Supplementary Figure S2B). For clarity we have limited our discussion to our LFN sample and will highlight only chromatin architectures seen in all four of our experiments.

Since the optimal method for interpreting MNase-Seq experiments has yet to be clearly demonstrated, we used four methods to analyze our raw data (See Supplementary Figure S3 for overview). First, we examined the ends of the sequence reads on individual strands (Figure 1A). This representation of the data is completely unprocessed and is an exact representation of the raw aligned sequencing output. Second, based on the fission yeast genome sequence, we used a computer to extrapolate each 36-nt tag to a total length of 120 nt. Although nucleosomes are 147 nt, we found that extrapolation to 120 nt provided better demarcation of nucleosome borders with no effect on occupancy comparisons. Then, we counted the number of times each unique position in the genome was included within a 120-nt extrapolated sequence, and we determined the $\log_2$ ratio of this number to the genome average, to generate a plot of 'MNase protection' across the genome (Figure 1B, orange line). We prefer to use the term 'MNase protection' to describe these data rather than 'nucleosome occupancy' because MNase protection may alternatively be conferred by non-nucleosomal protein complexes or by higher order structure in the largely intact nuclei. Our third approach was 'template filtering', as recently described by Weiner *et al.* (14). This method employs a set of seven templates matching frequently found distributions of sequence tags at MNase-generated nucleosome ends to extract information about positions, sizes and occupancies directly from raw data. Lastly, the raw data were also converted into conditional nucleosome positioning probabilities (16). These values represent the relative probability of a nucleosome at a given position compared to other positions within a 140-bp window centered on that position. This measurement is significantly different from our MNase protection results, because it standardizes all locations by the measured MNase protection (nucleosome occupancy) for that window. Therefore, a region with low nucleosome occupancy but only one possible nucleosomal configuration would have a high conditional nucleosome positioning value. Conditional nucleosome positioning calculations are also more robust than nucleosome occupancy with regard to technical differences among MNase-seq experiments, and they allow more accurate comparisons between datasets derived from different laboratories.

Nucleosome positions in *S. pombe* are well conserved between fixed and unfixed cells in log and stationary growth conditions

As indicated above, the positions of nucleosomes in our different preparations of *S. pombe* chromatin were similar, suggesting extensive conservation of chromatin structure between log and stationary growth states, regardless of formaldehyde fixation and regardless of whether the gel explant defining 'mononucleosome' size was narrow or broad. For this reason, we felt that a database of nucleosome positions in all of our experimental samples might facilitate future chromatin studies in *S. pombe*. We used template filtering (see 'Materials and Methods' section) to develop such a database. We also developed a reference dataset composed of all the nucleosome positions that were well conserved within our four experimental preparations (examples in Supplementary Figure S4). Reference conserved nucleosomes were identified as nucleosome positions appearing in at least three of our four data sets and occurring in both log and stationary growth states.

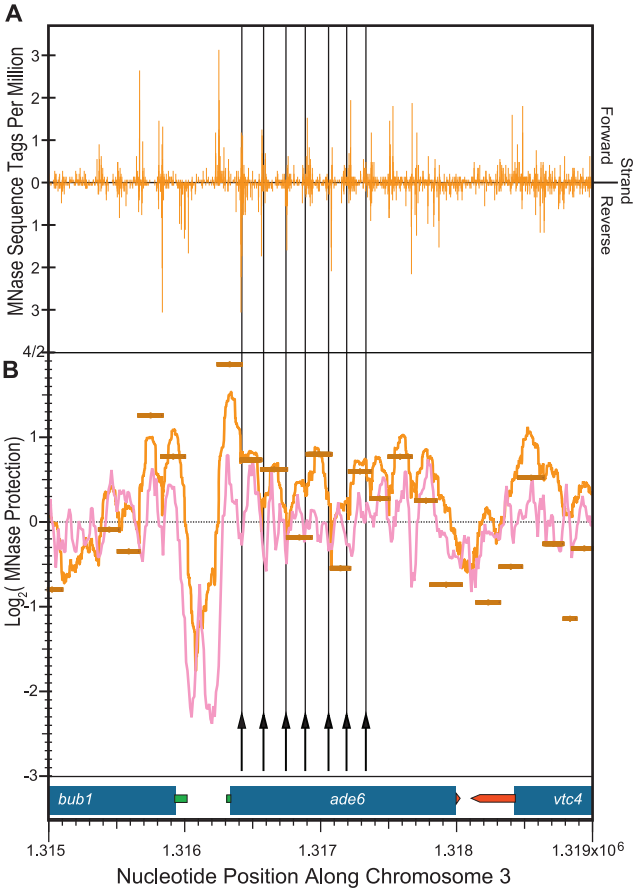

Figure 1. Chromatin profiles of the *S. pombe ade6* gene region. The horizontal axis and transcription diagram at the bottom of this figure are common to (**A** and **B**). Thick blue bars represent the CoDing Sequences (CDS) of the indicated genes. The narrow green bars represent the 5′-UTRs of these genes, while the narrow red pointed bars represent their 3′-UTRs. (A) Histogram (vertical orange bars) displaying the frequencies (tags at the indicated position per million tags genome-wide) of 36-nt sequence tags in mononucleosomal DNA (LFN sample), which map uniquely to the forward (upward scale) and reverse (downward scale) strands of a 4-kbp region of *S. pombe* chromosome 3 centered on the *ade6* gene. (B) Profile of MNase protection (orange line, left-hand vertical axis) defined as the $\log_2$ ratio of the number of 3′ extended sequence tags at each nucleotide position to the genome average (horizontal black, dotted line). The pink line shows MNase protection based on the microarray signals published by Lantermann *et al.* (6). The vertical black arrows and lines mark the positions of MNase cuts mapped within the *ade6* gene by Lantermann *et al.* (5) using the traditional indirect-end-labeling method. The horizontal orange boxes represent nucleosome positions, sizes, and occupancy, calculated directly from the LFN sequence-tag data shown in (A) by the 'Template Filtering' algorithm of Weiner *et al.* (14). The center of each called nucleosome position is indicated by the vertical tick mark superimposed on the corresponding box.

Both strict (dyads within ±10 bp of each other) and lenient (dyads within ±10–30 bp) cutoffs were applied. An average of 72 146 nucleosomes were called for the *S. pombe* genome across all four data sets. Approximately 62% (45 089) met at least the lenient criteria for conservation. This reference nucleosome data set and the four experimental nucleosome data sets used to compile it are all available for general use on Pombase (http://www.pombase.org/).

Nucleosome locations are accurately identified by MNase-seq at the *ade6* locus

To confirm that our approach was able to accurately map nucleosomes, we examined the previously characterized *ade6* gene region on chromosome 3. The locations of nucleosomes in this region have been determined by gel electrophoretic analysis of MNase-generated fragment sizes (indirect end labeling (5,26–28); and, more recently, by micro-array analysis (6,29).

Figure 1 demonstrates that our raw and processed data for this locus (LFN sample) are in excellent agreement with independent results from earlier studies (5,6) [the recent MNase-microarray (MNase-chip) results from De Castro *et al.* (30) are similar to those of Lantermann *et al.* (6) and are not shown here]. The coding region of *ade6* is marked by distinct spikes in tag frequency on the forward and reverse strands at regular intervals of ~130–170 bp, consistent with the presence of nine positioned nucleosomes. The vertical black lines that extend up into Figure 1A from the arrows in Figure 1B show the borders of MNase fragments previously determined by gel electrophoresis and indirect end labeling (5). Note the excellent correlation between the fragment borders mapped by indirect end labeling and those identified by deep sequencing. The orange line in Figure 1B shows the 'MNase protection' profile for our LFN sample, and the dark orange boxes show the nucleosome positions/occupancy calculated by the template filtering algorithm. Note that the MNase protection profile and template-filtered nucleosome positions are independently calculated directly from the raw data without any computational or plot smoothing.

The pink line shows the results for this region of an earlier genome-wide study (6), which used hybridization to microarrays rather than sequencing (MNase-chip). The fact that our results for the *ade6* region, whether expressed as raw sequence tag counts (Figure 1A) or as MNase protection or called nucleosomes (Figure 1B), are in excellent agreement with both the earlier indirect end-labeling studies [black arrows in Figure 1; (5)] and with the more recent genome-wide microarray studies [pink line, Figure 1B; (6)] suggests that our results are also likely to provide an accurate picture of nucleosome positions elsewhere in the genome. However, we note that in some other regions of the genome, particularly at replication origins, there is significant disagreement regarding the extent of nucleosome depletion in NDRs between our results and the microarray findings of both Lantermann *et al.* (6)and De Castro *et al.* (30). These exceptions will be discussed below.

Similarities and differences between nucleosomal arrays in fission and budding yeasts

Previous investigators have shown that both budding and fission yeasts differ significantly from animals and plants, because, within arrays of regularly repeating nucleosomes, they have a relatively short inter-nucleosomal repeat distance and lack classic linker histones (7,31,32). Despite their short repeat distances, yeast polynucleosomes can still condense *in vitro* into 30-nm chromatin filaments typical of

other eukaryotes (3,31). We were therefore interested in directly comparing the characteristics of nucleosomal arrays between fission and budding yeasts.

We began our analysis by comparing nucleosome distributions at TSSs and TTSs between one of our data sets (LFN) and a similarly prepared dataset from budding yeast [the BY2 experiment of Weiner *et al.* (14), in which mild MNase digestion conditions, similar to ours, were employed]. Figure 2A depicts MNase-protection data (orange line, fission yeast; purple line, budding yeast) averaged over a 2001-bp window centered on the TSS for 3775 fission yeast genes [TSS coordinates from (6)] and for 4800 budding yeast genes [coordinates from (33)]. Figure 2B shows the same regions when the data are converted into conditional probability of nucleosome dyad positioning (16). In both yeasts, there is a strong MNase-protection peak, indicating a positioned '+1' nucleosome, ~50 bp downstream of the TSS, as previously documented in budding yeast (34) and more recently in fission yeast (6). About 50–100 bp upstream of the TSS, there is a trough in the protection signal, frequently called a 'Nucleosome-Depleted Region' [NDR; (35)]. The promoter-associated NDR is itself immediately preceded by a protection peak consistent with a positioned '−1' nucleosome. For the MNase-protection profiles (Figure 2A), budding yeast exhibits a series of positioned

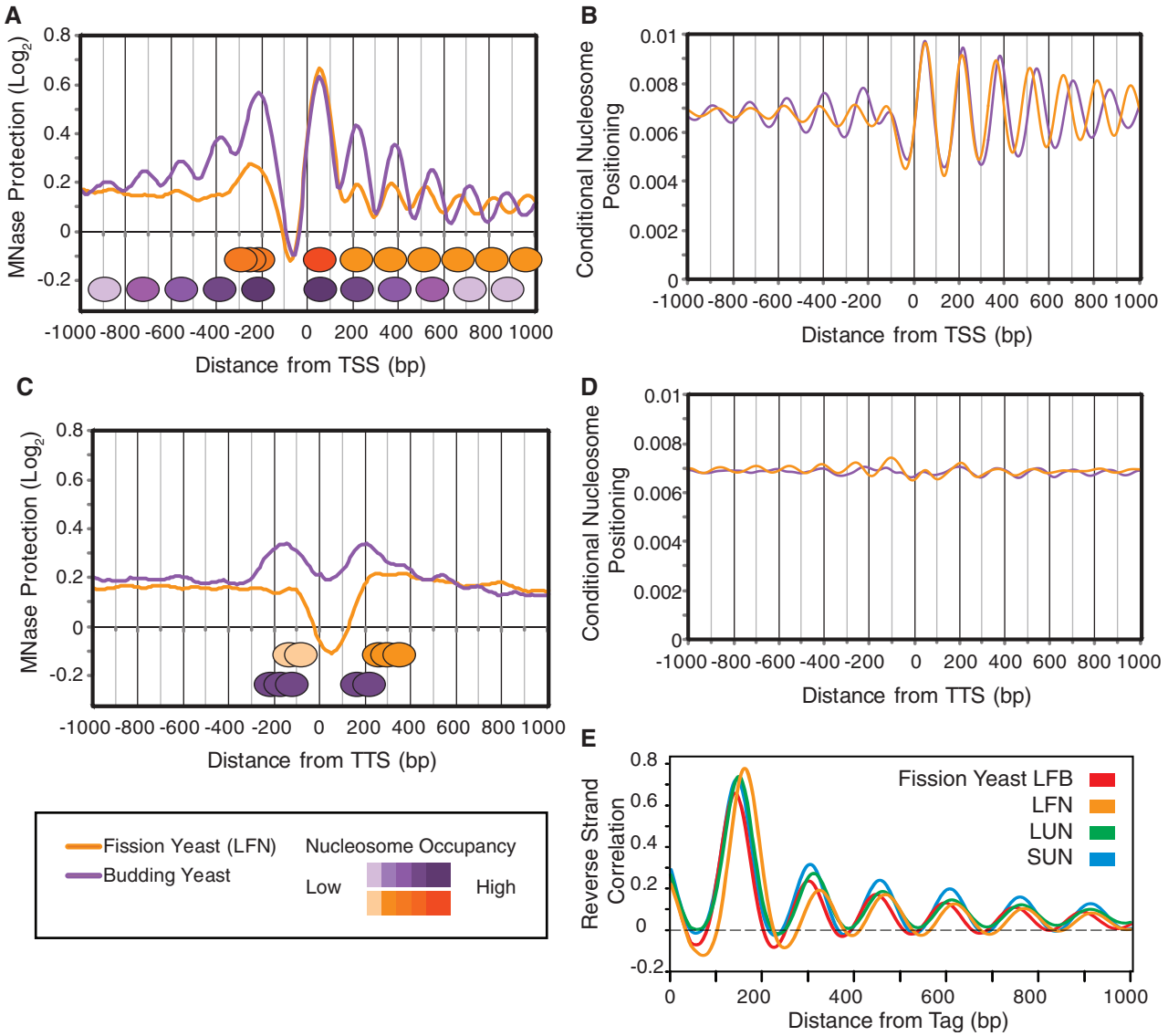

Figure 2. Comparison of chromatin at TSSs and TTSs, and nucleosome repeat lengths for fission yeast. (**A**) Profiles of MNase protection over 2001-bp windows centered on TSSs. Orange line: *S. pombe* LFN sample. Purple line: *S. cerevisiae*. The locations and occupancies of the average nucleosomes flanking TSSs are shown below the curve. (**B**) Conditional probabilities of nucleosomes flanking TSSs. Conditional positioning is a measure of the probability of a nucleosome at a particular position compared to the sum of all nucleosomal configurations within 140 bp of that position. (**C**) Profiles of MNase protection over 2001-bp windows centered on TTSs with the locations and occupancies of the average nucleosomes. (**D**) Conditional probabilities of nucleosomes flanking TTSs. (**E**) Cross-correlations between forward- and reverse-strand MNase fragment sequence tags for all four of our samples. Genome-wide Pearson correlation coefficients were determined between the number of forward sequence tags at each mappable position in the genome and the number of reverse sequence tags at each position downstream for 1000 bp.

peaks extending in both directions, with declining magnitudes, away from both the −1 and +1 nucleosomes. In fission yeast such a series is found only over the gene-coding region on the right. However, the conditional probability profiles (Figure 2B) display a regular array of nucleosomes both upstream and downstream of the TSS for both yeasts. It is important to note that the magnitude of the nucleosome signals is decreased upstream of the TSS for both yeasts, thus demonstrating that nucleosome positioning is not as strongly defined upstream as downstream of the TSS. In fission yeast the upstream positioning is further reduced compared to budding yeast.

Similar to TSSs, transcriptional termination sites (TTSs) in budding yeast have previously been shown to have an NDR (13). For fission yeast there is an evident NDR, but weak evidence for positioned nucleosomes either upstream or downstream (Figure 2C). In addition, the NDR for fission yeast is significantly deeper compared to budding yeast at this level of MNase digestion (Figure 2C). Conditional positioning results confirm that nucleosomes are only weakly positioned at TTSs in both yeasts (Figure 2D).

It is evident from both the MNase protection (Figure 2A) and the conditional probability data (Figure 2B) that the spacing between nucleosomes downstream of the TSS is shorter in fission yeast than in budding yeast. In order to evaluate the generality of this short repeat length over the entire fission yeast genome, we performed a genome-wide cross-correlation analysis of the distances between forward and reverse strand tags for all of our samples (Figure 2E). The center of the first peak, at 160–170 bp, is a measure of the most frequent length of DNA recovered from the sampled mono-nucleosome gel band. The subsequent peaks show the most frequent distances to the distal borders of subsequent nucleosomes. The distances between these subsequent peaks provide a measure of the spacing between nucleosome centers. For all of our experiments averaged together, the inter-nucleosome repeat distance (distance between nucleosome centers) is 152 ± 2 bp (95% confidence interval). For budding yeast, we estimated a 164-bp inter-nucleosome repeat distance by cross-correlation, demonstrating that the inter-nucleosome repeat distance in fission yeast is significantly shorter then budding yeast.

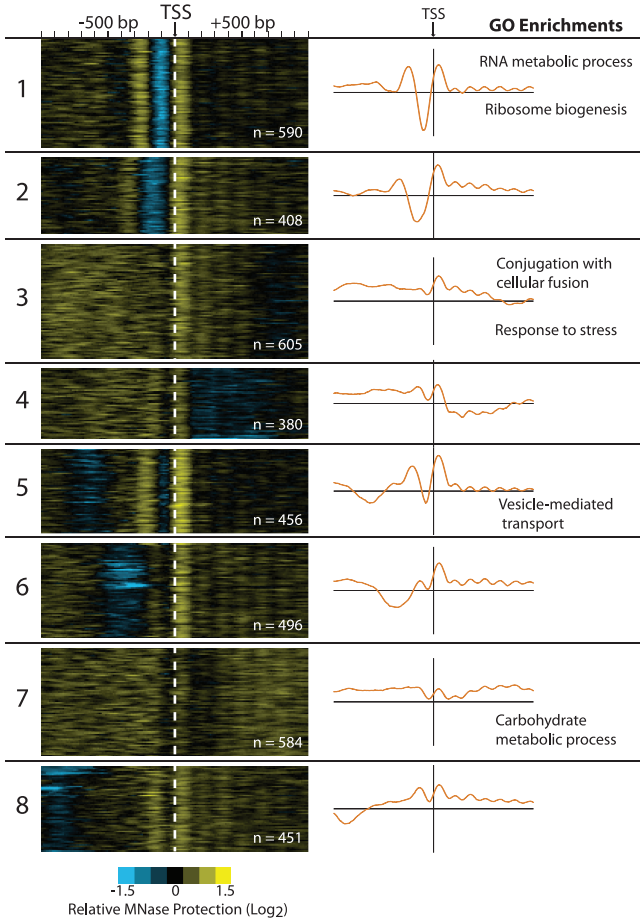

Figure 3. MNase-protection profiles at fission yeast TSSs distinguish sets of functionally associated genes. The eight data panels represent groupings defined by k-medioids clustering of TSSs. The analysis was performed using windows of −500 bp to +500 bp around TSSs. The number of genes (*n*) comprising each group is shown. The colors (blue to black to yellow) denote relative MNase protection according to the scale at the bottom. Distance from the TSS (vertical dashed white line) is shown in base pairs on the scale at the top of the column. Cluster-specific averaged TSS-centered profiles over 2001-bp windows are shown for each labeled group. Significant ($P < 0.05$; multiple testing corrected) enrichments of Gene Ontology (GO) slim terms for biological processes are shown.

Sets of functionally associated fission yeast genes can be distinguished based on chromatin profiles near TSSs

To further explore the relationship between chromatin structure and gene function we used unsupervised k-medioid clustering analysis on the LFN chromatin data for regions surrounding TSSs (Figure 3). Of the eight distinct cluster groups identified, four were significantly ($P < 0.05$) enriched for at least one GO slim term process [a term used by the Gene Ontology consortium (http://www.geneontology.org/) to describe a biological process] according to a multiple-testing-corrected hypergeometric distribution test. Incorporation of our additional data sets (LFB, LUN and SUN) failed to improve identification of GO terms. Furthermore, clustering with other values of k ($3 < k < 10$) had no qualitative effect on the observed GO slim term enrichments.

In all clusters there is a peak in MNase protection at the +1 nucleosome and an ordered array of downstream nucleosomes, as observed in the cumulative average plot (Figure 2A). Upstream of the TSS, each group separates into unique architectures likely representing different chromatin regulatory processes. Cluster 1 is enriched for ribosome biogenesis ($P = 9 \times 10^{-6}$) and RNA metabolic process ($P = 1 \times 10^{-10}$). This cluster contains the most consistently positioned −1 nucleosome and the deepest NDR. Cluster 4 is enriched for conjugation with cellular fusion ($P = 9 \times 10^{-5}$) and response to stress ($P = 8 \times 10^{-5}$). Cluster 5 is enriched for vesicle-mediated transport ($P = 5 \times 10^{-6}$), and cluster 7 is enriched for carbohydrate metabolic process ($P = 8 \times 10^{-5}$).

A deep, wide NDR at fission yeast replication origins

The organization of nucleosomes at DNA replication origins has recently been elucidated in the budding yeast, *Saccharomyces cerevisiae*, by Eaton *et al.* (9) and Berbenetz *et al.* (8). These investigations were facilitated by the fact that every budding yeast replication origin contains an essential sequence motif, the T-rich 11-bp ARS Consensus Sequence [ACS; reviewed in (36)], which binds to the heterohexameric Origin Recognition Complex (ORC). Bound ORC in turn helps to recruit the heterohexameric MiniChromosome Maintenance (MCM) DNA helicase and other proteins to form the Pre-RC (37), thus defining a potential replication origin.

When these investigators (8,9) aligned and oriented >200 budding yeast origins by their ACS motifs, they found that the averaged MNase protection profile consisted of an NDR of about 130 bp flanked on both sides by well-positioned nucleosomes each bordering a regular series of peaks tapering away from the origin. The ACS is located eccentrically along one slope of the NDR trough, with its T-rich strand pointing toward the center.

Unfortunately, *S. cerevisiae* and its close relatives among the budding yeasts are the *only* eukaryotic organisms known to have unique essential origin sequence motifs (such as the ACS), by which their replication origins can be aligned. Fission yeast ORC lacks sequence specificity but instead binds selectively to AT-rich DNA by means of a multiple-AT-hook domain in its Orc4 subunit (38–41). Consequently fission yeast replication origins generally correlate with extended zones of exceedingly AT-rich DNA referred to as 'AT-islands' (42) and frequently contain multiple ORC-binding sites (38,40,41,43,44).

Previous studies have mapped fission yeast origins on the basis of AT content ['AT islands' (42)], on the basis of extent of DNA synthesis in early S phase measured by hybridization to probes in a microarray (24,45–47), and on the basis of ChIP-chip measurements of abundance of pre-RC proteins (24). Of these methods, only measurements of AT content and of pre-RC protein abundance have sufficient resolution to localize origins within a few hundred bp of their true positions. In order to maximize our chance of success, we decided to combine these two methods. We identified a set of 217 fission yeast origins (Supplementary Table S1) that were sufficiently unique in nucleotide sequence that they were mappable by our MNase-sequencing method. In addition, each of them had previously been identified both as an AT island and as a site of pre-RC binding, and in each case the AT-island position (42,47) was within 500 bp of the corresponding pre-RC position (24). Of these origins, 157 were found to be efficient and early-firing by Hayashi *et al.* (24), while the other 60 were found to be inefficient and/or late-firing.

In an initial attempt to see whether there are any conserved chromatin features at replication origins, we aligned these 217 origins by their AT-island positions and averaged their MNase protection (LFN sample) at each point in a 2001-bp window centered on their AT islands. The resulting profile (orange line in Figure 4A) revealed a wide, deep NDR, but no other significant features.

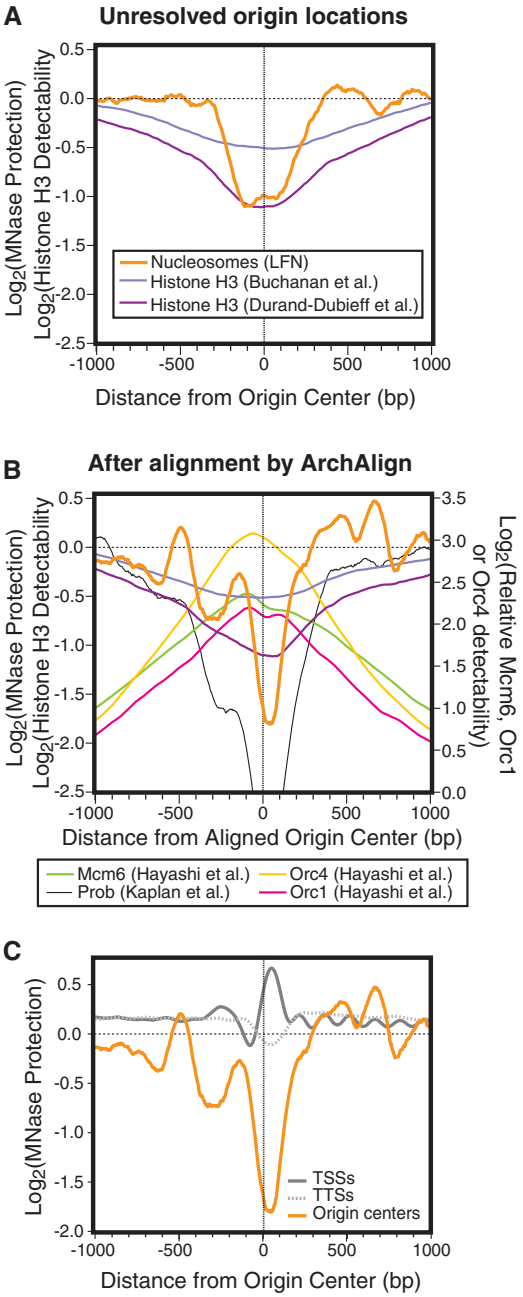

Figure 4. Average distribution profiles of MNase protection, histone H3, and pre-RC proteins around 217 replication origins. (**A**) For this graph, 'origin centers' are AT-island positions [Supplementary Table S1; (42,47)]. The average protection profile for replication origins aligned by their AT-island positions, but not oriented, is plotted for the LFN sample (orange line). Average $\log_2$ ChIP-chip signals relative to genome average for histone H3 are also plotted. The histone H3 data are from Buchanan *et al.* [(22); blue-gray line] and Durand-Dubief *et al.* [(23); purple line]. (**B**) Orange line: the average LFN protection profile around origin centers after alignment by ArchAlign. Blue-gray and purple lines: histone H3 data as in (A), but centered on ArchAlign-defined origin centers. Magenta, yellow and green lines: average log2 ChIP-chip signals relative to genome average for the pre-RC proteins, ORC1, ORC4 and MCM6, respectively. Black line: predicted nucleosome occupancy, based on algorithm of Kaplan *et al.* (12), plotted on left-hand vertical axis. (**C**) Comparison of NDRs at origins with those of TSSs and TTSs. Orange line: average $\log_2$ MNase protection data for the LFN sample around replication origins, as in (B). Solid gray line: profile for the LFN data around TSSs. Dashed gray line: profile for the LFN data around TTSs.

To test whether alignment by MNase protection rather than by AT-island position might reveal additional features, we used the chromatin architecture alignment method, ArchAlign (21). ArchAlign attempts to align a set of functionally related chromatin regions by sliding them sideways and, if necessary, reorienting them, in order to maximize the similarity for the chromatin data set across all regions. ArchAlign has been extensively validated and shown to accurately align *S. cerevisiae* replication origins after random shuffling (21). When applied to this population of 217 well-localized fission yeast origins, ArchAlign generated interesting results, which are shown as the orange lines in Figure 4B and C. We wish to emphasize that 'origin centers', which are aligned at the centers of the horizontal axes in Figure 4B and C, probably have no intrinsic biological function; they are simply the centers of the regions which the ArchAlign algorithm selected in order to maximize the correlation between MNase-protection profiles at origins.

Comparing the orange LFN MNase protection profile in Figure 4B and C with that in Figure 4A shows that alignment and orientation by chromatin features led to a narrower and deeper NDR. It also led to peaks and valleys of lesser magnitude in the regions flanking the NDR. These features are also evident in the aligned MNase-protection profiles from our other samples (red, green and blue lines in Supplementary Figure S6).

Curiously, the major characteristic of the LFN MNase-protection average origin profile—the deep NDR—is not evident in the microarray-based MNase-chip data from Lantermann *et al.* (6) aligned and plotted on the same axes (Supplementary Figures S6 and S7, pink line). While the flanking peaks are present, albeit at reduced signal level, the NDR itself is far less distinct than that observed for our fission yeast samples and those seen in published budding yeast origin chromatin profiles (8,9). Recently, De Castro *et al.* (30) published similar MNase-chip data, on the basis of which they concluded that fission yeast, unlike budding yeast, does not have NDRs at its replication origins. An example of an origin (the one at AT1033) where we detect a pronounced NDR but Lantermann *et al.* (6) and De Castro *et al.* (30) do not is shown in Supplementary Figure S10A.

We do not know why our fission yeast origin chromatin profiles, obtained by MNase-seq, reveal pronounced NDRs, while those of both Lantermann *et al.* (6) and De Castro *et al.* (30), obtained by MNase-chip, do not. However, the following arguments suggest that NDRs are true features of fission yeast replication origins:

(1) To independently test for the presence of a NDR at replication origins we examined histone occupancy measurements from two recent ChIP-chip data sets (22,23). The ChIP-chip method relies on immunodetection of proteins rather than nuclease protection. The average profiles for histone H3 distribution at replication origins (Figure 4A and B, blue–gray and purple lines) reveal that histones are relatively depleted within 1 kb on both sides of AT islands or origin centers (all values are less than the genome average, 0 on the left axis), and the lowest histone abundance is near the center of the NDR. Depending on the data set, the average histone depletion at the NDR centers ranges from ~1.4- to 2-fold, relative to the genome average.
(2) Additional Figure 1, which is available as a multi-page PDF file from http://www.acsu.buffalo.edu/~mjbuck/Fission_Yeast_chromatin.html, shows—for each of the 217 origins—the data on which the average profiles in Figure 4 and Supplementary Figure S6 were based. Examination of the histone H3 levels in Additional Figure 1 reveals that there is significant histone H3 depletion at every one of the 217 origins; there is not a single exception.
(3) Figure 4C compares the LFN MNase protection profile generated around replication origins (orange line) with the protection profiles, from the same sample and displayed at the same scale, around TSSs (solid gray line) and TTSs (dashed gray line). The fact that the origin NDR is much deeper than the TSS or TTS NDR means that the NDR at origins cannot be explained by possible overlap between origins and TSSs or TTSs.
(4) The fact that the LFN NDR (Figure 4B and 4C, orange line from approximately −100 to +180 bp relative to origin center) coincides with a strong minimum in the calculated probability of nucleosome formation [Figure 4B, black line; (12)] is consistent with the NDR having low nucleosome occupancy. In other words, the nucleotide sequence at our origin NDRs is especially unfavorable for nucleosome formation, so it is not surprising that we detect an NDR there.
(5) We further confirmed the presence of NDRs at the majority of origins by determining how many origins contain an 80-bp or larger gap between neighboring nucleosomes (based on template-filtered nucleosome boundaries). Of the 217 origins, 181 (83%) contained a 80-bp or larger gap while only 49% of random 1-kb windows contained such gaps. In addition, the nucleosome density at origins (4.6 nucleosomes per kb) was significantly ($P < 0.05$) lower than the density at random genomic regions (5.6 nucleosomes per kb).

Asymmetric distribution of pre-RC proteins in fission yeast replication origin NDRs

A striking feature of our alignment (Figure 4B) is that the peaks and centers of mass of the average MCM6, ORC1 and ORC4 binding distributions (light-green, magenta and yellow lines, respectively) are all located to the left of origin center (Supplementary Table S2), while the center of the NDR is 44 bp to the right of origin center. Thus, the pre-RC proteins bind to origins, on average, at distances of ~60–150 bp away from the centers of their NDRs.

To test whether the asymmetric distribution of pre-RC proteins within the origin NDR (Figure 4B) might be an artifact of using the ArchAlign algorithm, we aligned and oriented the same 217 replication origins by an entirely independent procedure, based primarily on visual estimation of the position of minimum predicted probability of nucleosome formation within the origin [(12); see Supplementary Materials]. The results, shown in

Supplementary Figure S8, are strikingly similar to those in Figure 4B (except for reduced definition of the peaks in MNase protection flanking the NDR). Thus, the asymmetric location of pre-RC binding within the origin-associated NDR is a robust observation independent of the method used to align the origins.

These observations are reminiscent of the finding by Eaton *et al.* (9) and Berbenetz *et al.* (8) that the ORC-binding site in budding yeast replication origins, the ACS, is positioned asymmetrically within the origin-associated NDR. Consistent with the fact that fission yeast replication origins are larger than those of budding yeast (48), the distances between the NDR center and pre-RC binding sites are larger in fission yeast (~60–160 bp; Supplementary Table S2) than is the distance between the ACS and NDR center in budding yeast [~30 bp; (8)].

Chromatin architecture at replication origins is heterogeneous

To further investigate chromatin architecture at fission yeast replication origins we examined the MNase protection profiles of three origins that had been well characterized in previous genetic studies. The lowest panels (v) in Figure 5 show cartoons of the three fission yeast DNA replication origins, (A) *ars1*, (B) *ars2004* and (C) *ars3002*, and their flanking genes in windows of 3 kbp surrounding each origin. The pink boxes in panels (v) show the regions found by genetic analysis to be important for origin activity within plasmids containing these DNA stretches (48–50). The smaller, elevated pink boxes show the sub-regions most important for plasmid origin activity. The underlying light yellow boxes show positions where fission yeast ORC binds to DNA *in vitro* (40,41,43). There are two or more such binding sites within each origin. Note that at least one ORC-binding site per origin corresponds to a region especially important for origin activity. Thin black vertical guide lines extend upward from the transcription and origin boundaries in the bottom panels (v) to the upper panels, which display the distributions of MNase protection, histones and pre-RC proteins.

As in Figure 1A, the topmost panels (i) in Figure 5 show the distributions of MNase-generated sequence tags in the top and bottom strands of our LFN sample. Panels (i) demonstrate that higher tag frequencies often occur at regular intervals in both gene and origin regions, suggesting that nucleosomes (and/or possibly other MNase-resistant structures) can be relatively well positioned through origin regions in the LFN sample.

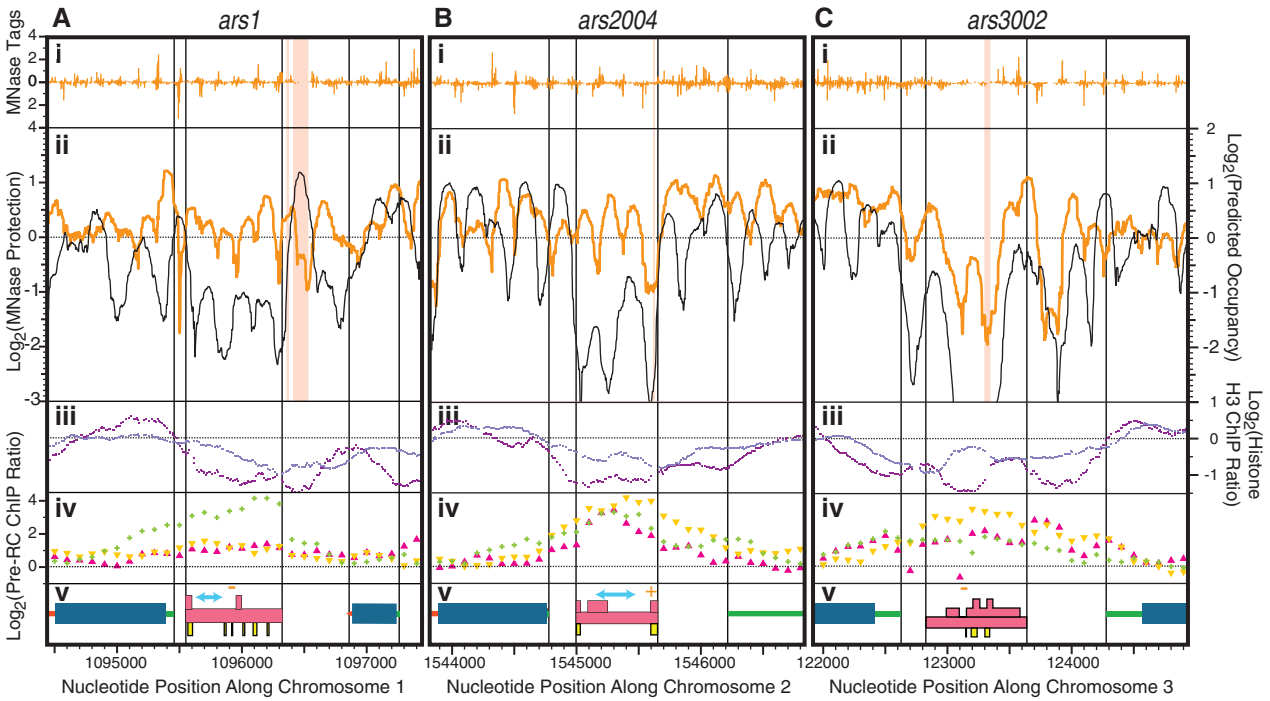

Figure 5. Chromatin structures of three well-studied replication origins. Each vertical panel shows 3 kb surrounding (**A**) *ars1,* (**B**) *ars2004,* (**C**) *ars3002*. Panels (i): (orange lines) number of MNase sequence tags per million in the LFN sample. Tags mapping to the top (bottom) strand are shown above (below) the 0 line. Panels (ii): MNase protection relative to genome average, LFN sample (orange line). Pale pink-orange vertical bars: regions of reduced mappability due to sequence repetitions. Black line: predicted nucleosome occupancy, based on algorithm of Kaplan *et al.* (12), right-hand scale. Panels (iii): ChIP-chip signal ratios relative to genome average for histone H3. Data from Buchanan *et al.* [(22); blue-gray line] and Durand-Dubief *et al.* [(23); purple line]. Panels (iv): ChIP-chip signal ratios for ORC1 (magenta triangles), ORC4 (inverted yellow triangles) and MCM6 (yellow–green crosses). Data from Hayashi *et al.* (24). Panels (v): coding regions (thick blue bars), 5′-UTRs (thin green bars) and 3′-UTRs (thin red bars) of the genes flanking the depicted origins. Pink boxes: regions defined genetically as important for ARS activity; upper boxes show small especially important regions. Lower, yellow boxes: *in vitro* ORC binding sites. Small orange minus and plus signs: positions of origin centers and orientations according to ArchAlign (21). Double-headed light cyan arrows in (**A**v) and (**B**v): positions from which replication forks diverge (51,52). Thin black vertical guide lines: boundaries of TSSs, TTSs and genetically defined origins. See the main text for details.

The MNase protection profiles generated by our LFN sample (orange lines in panels ii) suggest average levels of MNase protection across most of *ars1* (Figure 5A) and *ars2002* (Figure 5B) but reduced MNase protection across most of *ars3002* (Figure 5C). In other words, although most of *ars3002* appears to be a NDR, and there is a small NDR at the right end of *ars2004*, no NDR is evident in the MNase protection profile for *ars1*. The fact that *ars1* does not display an NDR in its MNase protection profile is somewhat surprising, because ChIP-chip measurements of histone H3 (panels iii) indicate that all of *ars1* has less H3 than genome average. The other two origins, *ars2004* and *ars3002*, are also deficient in histone H3. As indicated above, Additional Figure 1 (available at http://www.acsu.buffalo.edu/~mjbuck/Fission_Yeast_chromatin.html) shows that every one of the 217 origins we studied is deficient in histone H3.

Functional replication origins must recruit the pre-RC protein complexes, ORC and MCM. In 2007, Hayashi *et al.* (24) reported ChIP-chip measurements of the relative abundances of the Orc1, Orc4 and Mcm6 proteins throughout the fission yeast genome. Their results for the abundances of these three proteins in *ars1*, *ars2004*, and *ars3002* are shown in panels (iv) of Figure 5. All three of these pre-RC proteins were detected at levels above the genome average (indicated by the horizontal dotted line at 0, left-side axis) at all three origins (24). Additional Figure 1 shows that the same is true for each of our other 217 studied origins.

Aside from paucity of histone H3 and elevated levels of pre-RC proteins, these three origins do not appear to have a shared chromatin architecture. To further examine the extent of heterogeneity at fission yeast origins, we performed k-medioids clustering to divide the 217 origins, after alignment by ArchAlign, into groups based on their MNase-protection profiles. The heat map in Supplementary Figure S9 shows clustering into 10 groups and illustrates the heterogeneity of chromatin architecture at these origins. This heterogeneity is further emphasized by comparing the MNase-resistance, histone H3 and pre-RC protein profiles surrounding all 217 origins (Additional Figure 1; http://www.acsu.buffalo.edu/~mjbuck/Fission_Yeast_chromatin.html). Supplementary Table S3 lists all 217 origins, within their cluster groups, in their heat-map order (top to bottom). Although nearly all origins have a strong NDR in their MNase protection profiles, close to their origin centers, a small subset (including *ars1*) do not.

Although regularly spaced nucleosomes flanking the NDR, which are a prominent feature of budding yeast replication origins (8,9), are not evident in Supplementary Figure S9, nucleosomes do appear to be located preferentially in the vicinities of the dark orange vertical lines. It is evident from Supplementary Figure S9 that the peaks in the orange MNase-protection profile in Figure 4B and C are due to preferential nucleosome positioning in some, but not all, cluster groups.

We found no significant enrichment for early- or late-firing origins in any of the cluster groups, irrespective of the number of clusters (5–10). These findings suggest that origin activity (early/late) is probably not determined solely by nucleosome position and occupancy in fission yeast. Further analysis with additional chromatin data sets (for example, histone modifications) will be needed to determine the extent to which origin activity is dependent on chromatin architecture.

DISCUSSION

Unusually short inter-nucleosome repeat length

We found the average *S. pombe* inter-nucleosome repeat length (Figure 2) to be 152 ± 2 bp (95% confidence interval), which is within experimental error of the 154 bp measured by Lantermann *et al.* (6) using microarray analysis of nucleosome-size DNA. It is also close to the interval reported by Godde and Widom [156 ± 2 bp; (7)] based on measurement of the distances between peaks in gel-electrophoretic nucleosome ladders.

All of these measurements are somewhat shorter than the inter-nucleosome distance for budding yeast [164–167 bp; our group and (2,6)] and much shorter than the typical inter-nucleosome repeat length for vertebrates [180–210 bp; (2)], although rabbit cerebral cortical neurons and ox glial neurons have a yeast-like inter-nucleosome repeat [~162 bp; (2)]. So far as we are aware, *S. pombe*, along with another fungus [*Aspergillus nidulans*; (2)] has the shortest inter-nucleosome repeat length of any eukaryotic organism.

The short inter-nucleosome repeat length in fission and budding yeasts is probably related to their lack of bulk 'linker' histones similar to mammalian H1 or avian H5. The short repeat length likely contributes to the ability of yeast poly-nucleosomes to form "higher order structures" (such as 30-nm fibers) (31,53–55) in the absence of linker histones in vitro, while vertebrate poly-nucleosomes can only do so in their presence. The distinctively short segments of linker DNA between yeast nucleosomes would present only a minimal electrorepulsive or steric barrier to inter-nucleosomal packing interactions (31). In contrast, the greater length of DNA separating vertebrate nucleosomes may necessitate linker-histone binding to bring core particles close enough together to permit inter-nucleosome packing interactions.

Weakly arrayed nucleosomes upstream and strongly arrayed nucleosomes downstream of NDRs at TSSs

Figure 2 shows that nucleosomes downstream of TSSs are strongly positioned in an organized array, while upstream nucleosomes are weakly positioned. The upstream positioning is so weak that it was evident only when we analyzed our data by examining conditional positioning probability. Although our conditional positioning analysis identified a slight positioning signal upstream of TSSs for all of our samples, in every case the upstream positioning signals pale in comparison to the robust positioning signals downstream of TSSs (Supplementary Figure S2A). Our results are in contrast to those of Lantermann *et al.* (6), who used DNA-microarray hybridization and did not observe any nucleosome positioning at all upstream of TSSs. It is likely that our use of

next-generation sequencing and samples from highly intact, cryo-fixed nuclei allowed us to detect the weak upstream positioning signals.

An NDR at TTSs

We also found a lack of regular nucleosomal arrays in the vicinity of TTSs (Figure 2C and D), either upstream (over genes) or downstream (intergenic). However, NDRs were evident slightly downstream of TTSs. The 'valleys' at these NDRs were seen in all of our data sets (Supplementary Figure S5) and were as deep as or deeper than the NDRs at the TSSs in the same sample (Supplementary Figure S2B). Interestingly, Lantermann *et al.* (6), who used a more extensive MNase digestion, did not detect significant NDR valleys at TTSs (their Supplementary Figure S5b). We are confident in our results, which are consistent with budding yeast, but the reasons for this difference between our results and those of Lantermann *et al.* (6) deserve further investigation in the future.

Ribosomal biogenesis genes have similar chromatin architectures in budding and fission yeast

The most striking chromatin profiles are those around the TSSs of cluster #1 (Figure 3). These display especially deep and wide NDRs and obvious −1 and +1 nucleosome signals. They are strongly ($P = 9 \times 10^{-6}$) associated with the GO slim term, 'ribosome biogenesis'. Similar profiles are also evident at a cluster of budding yeast genes that is also associated ($P = 6 \times 10^{-14}$) with the GO slim term, 'ribosome biogenesis' (35). The fact that the chromatin profiles around genes involved in making ribosomes are so well preserved in distantly related fission and budding yeasts suggests that similar profiles may also be present around ribosome biogenesis genes in other eukaryotic organisms. The reasons for these similarities are unclear but present an interesting example of possible evolutionary conservation of chromatin architecture.

NDRs at replication origins are more easily detected by histone immunoprecipitation than by MNase sensitivity

Our measurements of MNase protection in the vicinity of DNA replication origins revealed considerable heterogeneity. Although most replication origins were associated with a NDR, which in some cases was unusually large, some origins were not so associated (Supplementary Figure S9). These included *ars1* (Figure 5A). The average NDR for all 217 origins proved to be significantly wider and deeper than the average NDRs for TSSs and TTSs (Figure 4C).

Interestingly, results of the two previously published genome-wide studies of MNase protection in fission yeast (both of which employed MNase-chip rather than MNase-seq) suggest that replication origins are only rarely associated with NDRs (6,30). In most regions of the genome, our MNase-protection profiles (obtained by MNase-seq) are similar to the MNase-protection results obtained by MNase-chip (example in Figure 1). It is primarily at replication origins that the two types of study are in disagreement (Supplementary Figure S7). The disagreement is only partial, because all three studies concur that some origins (such as *ars1*) are not associated with NDRs. Furthermore, we have done a genome-wide comparison of the MNase-chip data from Lantermann *et al.* (6) with our MNase-seq results, and we have found that at a few origins the NDR detected by Lantermann *et al.* is at least as wide and deep as the NDR that we detected. The origin at AT1015 is an example (Supplementary Figure S10B). Although we have not done an extensive comparison of our data with the results from De Castro *et al.* (30), we note that the MNase protection results from De Castro *et al.* are very similar to those from Lantermann *et al.*, so we predict that there will prove to be similar examples of concurrence between our data and the results of De Castro *et al.* at a few replication origins. In summary, all three data sets detect NDRs at a subset of replication origins. That subset is much larger (most origins) in our case than in the other two cases (only a few origins), but the difference is one of degree.

In contrast to the variable proportion of NDR-associated replication origins when the presence or absence of nucleosomes is assayed by MNase protection, two studies employing an alternative method for detecting nucleosome abundance—immunoprecipitation of chromatin fragments containing histone H3 followed by microarray analysis of the DNA from those H3-containing fragments (ChIP-chip)—suggest that all of the 217 origins that we studied coincide with apparently histone-depleted regions (22,23). Results from these two studies are plotted in panels (iii) of Figure 5. Note in Figure 5A that *ars1* appears to be histone-deficient throughout its length, even though an NDR is not evident in its MNase protection profile. The histone H3 results, along with our MNase-seq results and the MNase-chip results of Lantermann *et al.*, are plotted for each of the 217 origins in Additional Figure 1 (a multi-page PDF file available at http://www.acsu.buffalo.edu/~mjbuck/Fission_Yeast_chromatin.html). Every one of the 217 origins is associated with a histone-H3-deficient region, even if it is not associated with an NDR.

Thus there is significant disagreement between methods based on MNase protection and methods based on histone immunoprecipitation with regard to histone levels at replication origins. It is not yet entirely clear which set of results is closer to the truth, because neither type of method is free from uncertainties. For example, non-histone-containing protein complexes (such as the pre-RC proteins) could confer MNase protection within a region lacking nucleosomes. As a counter example, proteins binding to the surfaces of nucleosomes could shield histone H3 within nucleosomes from detection by antibodies.

We acknowledge these uncertainties, but we think that the evidence provided by immunoprecipitation—suggesting that replication origins in fission yeast are consistently deficient in nucleosomes—is more likely to be correct, for the following reason. Several years ago, Kaplan *et al.* (12) measured the probability of in vitro nucleosome formation (in the absence of transcription factors, using only purified calf thymus histones and naked budding yeast

DNA) throughout the budding yeast genome. Then they developed a computer algorithm to predict such probability in any other genome based on its DNA sequence. We have applied that algorithm to the fission yeast genome, and we have found that every one of our 217 studied origins is unlikely, in the absence of other factors, to be a good substrate for nucleosome formation (thin black line in Figures 4B and 5, Supplementary Figures S6–S8 and Additional Figure 1). In our view, the consistency between the low predicted probability of nucleosome formation at every origin and the immunoprecipitation measurement indicating low histone density at every origin outweighs the variable, inconsistent extent of MNase protection at origins, measured either by MNase-seq or MNase-chip.

The disparities among studies cited in the preceding paragraphs raise two very interesting questions. First, what are the features of fission yeast replication origins that lead to disagreements between ChIP-chip measurements of histone abundance on the one hand and MNase-protection measurements of nucleosome abundance on the other? Second, what are the features of fission yeast replication origins that lead to conflict between MNase-protection analyses by MNase-seq and by MNase-chip? The answers to these questions, when obtained, will undoubtedly reveal interesting properties of fission yeast replication origins and possibly of all eukaryotic replication origins.

Correlation between NDR and pre-RC positions at fission yeast DNA replication origins

We used the recently developed computer algorithm, ArchAlign (21) to generate a high-quality alignment of MNase protection profiles at fission yeast replication origins (Figure 4B and 4C). The results from this method were reinforced by similar results obtained by alignment to preRC-proximal AT-islands (Figure 4A) and by an independent alignment by eye (Supplementary Figure S8). All three alignments reveal that the average fission yeast replication origin contains a substantial NDR (when measured by MNase-seq) that is much wider and deeper than the NDRs associated (in our experiments) with TSSs and TTSs (Figure 4C). The incorporation of published ChIP-chip data into both our ArchAlign and visual origin centering analyses permitted us to further conclude that proteins of the pre-RC (ORC and MCM; (24)) are maximally bound at one edge of the NDR, 60-160 bp away from its center (Figure 4B, Supplementary Figure S8, and Supplementary Table S2).

The above conclusions apply to fission yeast replication origins *on average*. Figure 5 and Additional Figure 1, Supplementary Figure S9 show that in fact there is a great deal of heterogeneity in the MNase protection profiles of individual fission yeast replication origins. That alignment of these heterogeneous origins permitted the conclusions listed in the preceding paragraph suggests that chromatin-structure-based alignment of multiple origins in other eukaryotic organisms (such as vertebrates, which, like fission yeast, lack conserved origin sequence motifs) may also reveal interesting origin features.

Comparison with budding yeast replication origins

Detailed comparison of our measurements of chromatin architecture near fission yeast replication origins (Figure 4 and Supplementary Figure S9) with corresponding measurements by Eaton *et al*. (9) and Berbenetz *et al*. (8) for budding yeast origins reveal two features that are similar and therefore may be evolutionarily conserved and several features that differ between these two very distantly related yeasts. The similar features include (i) a wide, deep NDR at most origins and (ii) the fact that pre-RC proteins bind at one side of this NDR. Differences include (i) arrays of regularly spaced nucleosomes on both sides of the budding yeast origin NDR but irregularly spaced nucleosomes flanking the fission yeast origin NDR, (ii) greater width of the fission yeast NDR and (iii) the presence of a second NDR about two nucleosomes' width beyond the primary NDR at a majority of fission yeast origins (Supplementary Figure S9), while only a minority of budding yeast origins have a second NDR (8).

The fact that fission and budding yeast replication origins are both characterized by the binding of pre-RC proteins to one side within an NDR and that they diverged from each other near the time when fungi diverged from animals (56) suggests the possibility that animal replication origins may also share this potentially functionally relevant configuration. In this context, it is interesting that recent studies (57,58) revealed that ORC binds to regions of low predicted and actual nucleosome occupancy in both fruit-fly and Chinese-hamster cells. Although these studies lacked sufficient resolution to permit determination of the positions of ORC in the vicinity of NDRs, the accelerating advance of technology promises adequate resolution and scope will be achieved in animal studies within the near future. Even now it is possible to infer that, in all these cases (budding yeast, fission yeast and animals), regions of low nucleosome occupancy are more likely to serve as replication origins, probably because pre-RC proteins can better compete with nucleosomes for binding to DNA within such regions (8,9,57,58).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–10, Supplementary Methods and Supplementary Reference [59].

ACKNOWLEDGEMENTS

R.M.G. is grateful for the support of Lt. Col. and Mrs C. M. Givens, R. Tafari, and the NY State Department of Labor during data analysis and article preparation.

FUNDING

National Institutes of Health [GM70566 to J.A.H.]; the National Science Foundation [IIS1016929 to M.J.B.]; a PhRMA predoctoral fellowship in Informatics (to J.M.R.). Funding for open access charge: Department of Biochemistry, SUNY at Buffalo.

Conflict of interest statement. None declared.

REFERENCES

1. Campos,E.I. and Reinberg,D. (2009) Histones: annotating chromatin. *Annu. Rev. Genet.*, **43**, 559–599.
2. van Holde,K.E. (1989) *Chromatin*. Springer, New York.
3. Woodcock,C.L. and Ghosh,R.P. (2010) Chromatin higher-order structure and dynamics. *Cold Spring Harb. Perspect. Biol.*, **2**, a000596.
4. Givens,R.M., Mesner,L.D., Hamlin,J.L., Buck,M.J. and Huberman,J.A. (2011) Integrity of chromatin and replicating DNA in nuclei released from fission yeast by semi-automated grinding in liquid nitrogen. *BMC Res. Notes*, **4**, 499.
5. Lantermann,A., Strålfors,A., Fagerström-Billai,F., Korber,P. and Ekwall,K. (2009) Genome-wide mapping of nucleosome positions in *Schizosaccharomyces pombe*. *Methods*, **48**, 218–225.
6. Lantermann,A.B., Straub,T., Strålfors,A., Yuan,G.-C., Ekwall,K. and Korber,P. (2010) *Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccharomyces cerevisiae*. *Nat. Struct. Mol. Biol.*, **17**, 251–257.
7. Godde,J.S. and Widom,J. (1992) Chromatin structure of *Schizosaccharomyces pombe*: A nucleosome repeat length that is shorter than the chromatosomal DNA length. *J. Mol. Biol.*, **226**, 1009–1025.
8. Berbenetz,N.M., Nislow,C. and Brown,G.W. (2010) Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet.*, **6**, e1001092.
9. Eaton,M.L., Galani,K., Kang,S., Bell,S.P. and MacAlpine,D.M. (2010) Conserved nucleosome positioning defines replication origins. *Genes Dev.*, **24**, 748–753.
10. Grimm,C., Kohli,J., Murray,J. and Maundrell,K. (1988) Genetic engineering of *Schizosaccharomyces pombe*: a system for gene disruption and replacement using the *ura4* gene as a selectable marker. *Mol. Gen. Genet.*, **215**, 81–86.
11. Moreno,S., Klar,A., Nurse,P. and Christine Guthrie and Gerald,R.F. (1991) *Methods in Enzymology*, Vol. 194. Academic Press, pp. 795–823.
12. Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Tillo,D., Field,Y., LeProust,E.M., Hughes,T.R., Lieb,J.D., Widom,J. *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**, 362–366.
13. Shivaswamy,S., Bhinge,A., Zhao,Y., Jones,S., Hirst,M. and Iyer,V.R. (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.*, **6**, e65.
14. Weiner,A., Hughes,A., Yassour,M., Rando,O.J. and Friedman,N. (2010) High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res.*, **20**, 90–100.
15. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
16. Kaplan,N., Hughes,T.R., Lieb,J.D. and Segal,E. (2010) Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. *Genome Biol.*, **11**, 140.
17. Lai,W.K., Bard,J.E. and Buck,M.J. (2012) ArchTEx: accurate extraction and visualization of Next-Generation Sequence data. *Bioinformatics*, **28**, 1021–1023.
18. de Hoon,M.J., Imoto,S., Nolan,J. and Miyano,S. (2004) Open source clustering software. *Bioinformatics*, **20**, 1453–1454.
19. Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad Sci. USA*, **95**, 14863–14868.
20. Saldanha,A.J. (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.
21. Lai,W. and Buck,M.J. (2010) ArchAlign: Coordinate-free chromatin alignment reveals novel architectures. *Genome Biol.*, **11**, R126.
22. Buchanan,L., Durand-Dubief,M., Roguev,A., Sakalar,C., Wilhelm,B., Strålfors,A., Shevchenko,A., Aasland,R., Shevchenko,A., Ekwall,K. *et al.* (2009) The *Schizosaccharomyces pombe* JmjC-protein, Msc1, prevents H2A.Z localization in centromeric and subtelomeric chromatin domains. *PLoS Genet.*, **5**, e1000726.
23. Durand-Dubief,M., Persson,J., Norman,U., Hartsuiker,E. and Ekwall,K. (2010) Topoisomerase I regulates open chromatin and controls gene expression *in vivo*. *EMBO J.*, **29**, 2126–2134.
24. Hayashi,M., Katou,Y., Itoh,T., Tazumi,A., Yamada,Y., Takahashi,T., Nakagawa,T., Shirahige,K. and Masukata,H. (2007) Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *EMBO J.*, **26**, 1327–1339.
25. Rizzo,J.M., Mieczkowski,P.A. and Buck,M.J. (2011) Tup1 stabilizes promoter nucleosome positioning and occupancy at transcriptionally plastic genes. *Nucleic Acids Res.*, **39**, 8803–8819.
26. Bernardi,F., Koller,T. and Thoma,F. (1991) The *ade6* gene of the fission yeast *Schizosaccharomyces pombe* has the same chromatin structure in the chromosome and in plasmids. *Yeast*, **7**, 547–558.
27. Bernardi,F., Zatchej,M. and Thoma,F. (1992) Species specific protein-DNA interactions may determine the chromatin units of genes in *S. cerevisiae and in S. pombe*. *EMBO J.*, **11**, 1177–1185.
28. Mizuno,K., Emura,Y., Baur,M., Kohli,J., Ohta,K. and Shibata,T. (1997) The meiotic recombination hot spot created by the single-base substitution ade6-M26 results in remodeling of chromatin structure in fission yeast. *Genes Dev.*, **11**, 876–886.
29. Song,J.S., Liu,X., Liu,X.S. and He,X. (2008) A high-resolution map of nucleosome positioning on a fission yeast centromere. *Genome Res.*, **18**, 1064–1072.
30. de Castro,E., Soriano,I., Marin,L., Serrano,R., Quintales,L. and Antequera,F. (2012) Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. *EMBO J.*, **31**, 124–137.
31. Kobori,T., Iwamoto,S., Takeyasu,K. and Ohtani,T. (2007) Chromatin dynamics of unfolding and refolding controlled by the nucleosome repeat length and the linker and core histones. *Biopolymers*, **85**, 295–307.
32. Lohr,D., Corden,J., Tatchell,K., Kovacic,R.T. and Van Holde,K.E. (1977) Comparative subunit structure of HeLa, yeast, and chicken erythrocyte chromatin. *Proc. Natl Acad. Sci. USA*, **74**, 79–83.
33. David,L., Huber,W., Granovskaia,M., Toedling,J., Palm,C.J., Bofkin,L., Jones,T., Davis,R.W. and Steinmetz,L.M. (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA*, **103**, 5320–5325.
34. Jiang,C. and Pugh,B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, **10**, 161–172.
35. Lee,W., Tillo,D., Bray,N., Morse,R.H., Davis,R.W., Hughes,T.R. and Nislow,C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.*, **39**, 1235–1244.
36. Chang,F., Theis,J.F., Miller,J., Nieduszynski,C.A., Newlon,C.S. and Weinreich,M. (2008) Analysis of chromosome III replicators reveals an unusual structure for the ARS318 silencer origin and a conserved WTW sequence within the origin recognition complex binding site. *Mol. Cell. Biol.*, **28**, 5071–5081.
37. Masai,H., Matsumoto,S., You,Z., Yoshizawa-Sugata,N. and Oda,M. (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu. Rev. Biochem.*, **79**, 89–130.
38. Chuang,R.Y., Chretien,L., Dai,J. and Kelly,T.J. (2002) Purification and characterization of the *Schizosaccharomyces pombe* origin recognition complex: interaction with origin DNA and Cdc18 protein. *J. Biol. Chem.*, **277**, 16920–16927.
39. Chuang,R.-Y. and Kelly,T.J. (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc. Natl Acad. Sci. USA*, **96**, 2656–2661.
40. Kong,D. and DePamphilis,M.L. (2001) Site-specific DNA binding of the *Schizosaccharomyces pombe* origin recognition complex is determined by the Orc4 subunit. *Mol. Cell. Biol.*, **21**, 8095–8103.
41. Lee,J.-K., Moon,K.-Y., Jiang,Y. and Hurwitz,J. (2001) The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc. Natl Acad. Sci. USA*, **98**, 13589–13594.

42. Segurado,M., Luis,A.d. and Antequera,F. (2003) Genome-wide distribution of DNA replication origins at A + T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep.*, **4**, 1048–1053.
43. Takahashi,T. and Masukata,H. (2001) Interaction of fission yeast ORC with essential adenine/thymine stretches in replication origins. *Genes Cells*, **6**, 837–849.
44. Takahashi,T., Ohara,E., Nishitani,H. and Masukata,H. (2003) Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast. *EMBO J.*, **22**, 964–974.
45. Feng,W., Collingwood,D., Boeck,M.E., Fox,L.A., Alvino,G.M., Fangman,W.L., Raghuraman,M.K. and Brewer,B.J. (2006) Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat. Cell Biol.*, **8**, 148–155.
46. Heichinger,C., Penkett,C.J., Bähler,J. and Nurse,P. (2006) Genome-wide characterization of fission yeast DNA replication origins. *EMBO J.*, **25**, 5171–5179.
47. Mickle,K.L., Ramanathan,S., Rosebrock,A., Oliva,A., Chaudari,A., Yompakdee,C., Scott,D., Leatherwood,J. and Huberman,J.A. (2007) Checkpoint-independence of most DNA replication origins in fission yeast. *BMC Mol. Biol.*, **8**, 112.
48. Dubey,D.D., Kim,S.-M., Todorov,I.T. and Huberman,J.A. (1996) Large, complex modular structure of a fission yeast DNA replication origin. *Curr. Biol.*, **6**, 467–473.
49. Clyne,R.K. and Kelly,T.J. (1995) Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *EMBO J.*, **14**, 6348–6357.
50. Okuno,Y., Satoh,H., Sekiguchi,M. and Masukata,H. (1999) Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Mol. Cell Biol.*, **19**, 6699–6709.
51. Gomez,M. and Antequera,F. (1999) Organization of DNA replication origins in the fission yeast genome. *EMBO J.*, **18**, 5683–5690.
52. Okuno,Y., Okazaki,T. and Masukata,H. (1997) Identification of a predominant replication origin in fission yeast. *Nucleic Acids Res.*, **25**, 530–536.
53. Bystricky,B., Heun,P., Gehlen,L., Langowski,J. and Gasser,S.M. (2004) Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl Acad. Sci. USA*, **101**, 16495–16500.
54. Erard,M. and Barker,D.G. (1985) Electron microscopic studies of condensed mitotic chromosomes in the fission yeast, *Schizosaccharomyces pombe*. *Biol. Cell*, **55**, 27–34.
55. Lowary,P.T. and Widom,J. (1989) Higher-order structure of *Saccharomyces cerevisiae* chromatin. *Proc. Natl Acad. Sci. USA*, **86**, 8266–8270.
56. Sipiczki,M. (2004) In: Egel,R. (ed.), *The Molecular Biology of Schizosaccharomyces pombe: Genetics, Genomics and Beyond*. Springer, Berlin, pp. 431–443.
57. Lubelsky,Y., Sasaki,T., Kuipers,M.A., Lucas,I., Le Beau,M.M., Carignon,S., Debatisse,M., Prinz,J.A., Dennis,J.H. and Gilbert,D.M. (2011) Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone. *Nucleic Acids Res.*, **39**, 3141–3155.
58. MacAlpine,H.K., Gordân,R., Powell,S.K., Hartemink,A.J. and MacAlpine,D.M. (2010) *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.*, **20**, 201–211.
59. Field,Y., Kaplan,N., Fondufe-Mittendorf,Y., Moore,I.K., Sharon,E., Lubling,Y., Widom,J. and Segal,E. (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.*, **4**, e1000216.