LETTERS

Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome

Istvan Albert[1], Travis N. Mavrich[1,2], Lynn P. Tomsho[1], Ji Qi[1], Sara J. Zanton[1,2], Stephan C. Schuster[1] & B. Franklin Pugh[1,2]

The nucleosome is the fundamental building block of eukaryotic chromosomes. Access to genetic information encoded in chromosomes is dependent on the position of nucleosomes along the DNA. Alternative locations just a few nucleotides apart can have profound effects on gene expression[1]. Yet the nucleosomal context in which chromosomal and gene regulatory elements reside remains ill-defined on a genomic scale. Here we sequence the DNA of 322,000 individual *Saccharomyces cerevisiae* nucleosomes, containing the histone variant H2A.Z, to provide a comprehensive map of H2A.Z nucleosomes in functionally important regions. With a median 4-base-pair resolution, we identify new and established signatures of nucleosome positioning. A single predominant rotational setting and multiple translational settings are evident. Chromosomal elements, ranging from telomeres to centromeres and transcriptional units, are found to possess characteristic nucleosomal architecture that may be important for their function. Promoter regulatory elements, including transcription factor binding sites and transcriptional start sites, show topological relationships with nucleosomes, such that transcription factor binding sites tend to be rotationally exposed on the nucleosome surface near its border. Transcriptional start sites tended to reside about one helical turn inside the nucleosome border. These findings reveal an intimate relationship between chromatin architecture and the underlying DNA sequence it regulates.

Chromatin is composed of repeating units of nucleosomes in which ~147 base pairs (bp) of DNA is wrapped ~1.7 times around the exterior of a histone protein complex[2]. A nucleosome has two fundamental relationships with its DNA[3]. A translational setting defines a nucleosomal midpoint relative to a given DNA locus. A rotational setting defines the orientation of DNA helix on the histone surface. Thus, DNA regulatory elements may reside in linker regions between nucleosomes or along the nucleosome surface, where they may face inward (potentially inaccessible) or outward (potentially accessible). Recent discoveries of nucleosome positioning sequences throughout the *S. cerevisiae* (yeast) genome suggest that nucleosome locations are partly defined by the underlying DNA sequence[4,5]. Indeed, a tendency of AA/TT dinucleotides to recur in 10-bp intervals and in counterphase with GC dinucleotides generates a curved DNA structure that favours nucleosome formation[3]. Genome-wide maps of nucleosome locations have been generated[6,7], but not at a resolution that would define translational and rotational settings. To acquire a better understanding of how genes are regulated by nucleosome positioning, we isolated and sequenced H2A.Z-containing nucleosomes from *S. cerevisiae*. Such nucleosomes are enriched at promoter regions[8–11], and thus maximum coverage of relevant regions can be achieved with fewer sequencing runs. With this high resolution map we sought to address the following questions: (1) what are the DNA signatures of nucleosome positioning *in vivo*? (2) How many translational and rotational settings do nucleosomes occupy? (3) Do chromosomal elements possess specific chromatin architecture? (4) What is the topological relationship between the location of promoter elements and the rotational and translational setting of nucleosomes?

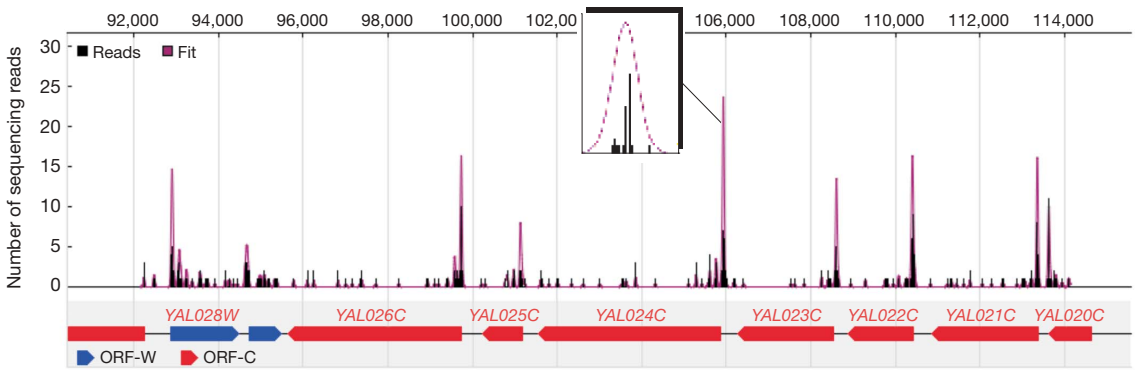

Figure 1 | Distribution of H2A.Z nucleosomal DNA at an arbitrary region of the yeast genome. Any region of the genome can be viewed in this way at http://nucleosomes.sysbio.bx.psu.edu. An enlarged view of a peak is shown in the inset, where each vertical bar corresponds to the number of sequencing reads located at individual chromosomal coordinates. The locations of ORFs are shown below the peaks. Additional browser shots are shown in Supplementary Fig. 1.

[1]Center for Comparative Genomics and Bioinformatics, [2]Center for Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

Nucleosome positions were trapped in yeast by formaldehyde crosslinking. H2A.Z nucleosome core particles were immunoprecipitated from micrococcal nuclease (MNase)-digested chromatin and gel purified (Supplementary Figs 1 and 2); core particles (322,000) were individually sequenced in parallel by pyrosequencing. Most sequencing reads had highly clustered locations throughout the genome and were strongly biased towards the 5′ end of genes (Fig. 1), as expected. In generating the map, we made adjustments for sequence bias inherent in MNase cutting (Supplementary Fig. 3). The distribution of sequencing reads was then smoothed at two levels: fine and coarse (see Fig. 2b), corresponding to individual translational settings and the average setting, respectively (provided in Supplementary Table 1). Independent determinations made on the Watson and Crick strands (left and right borders) concurred, producing a median error of 4 bp (Supplementary Fig. 4a). Comparisons with other experimentally and computationally derived maps[4,5,9,11] are presented in Supplementary Fig. 4b–f.

On average, 25 sequencing reads per nucleosome were used to assign the genomic location of ~10,000 H2A.Z nucleosomes. Another 30,000 nucleosomes, containing low levels of H2A.Z, were assigned with an average of 3–4 reads. Even nucleosomes identified by a single read had comparatively low error (Supplementary Fig. 4d), reflecting widespread incorporation of low H2A.Z levels. This was further examined at the highly repetitive ribosomal DNA locus, which should amplify this noise by a factor of 150–200, corresponding to the number of rDNA repeats. Indeed, reads at the rDNA locus clustered into nucleosome-sized arrays (Supplementary Fig. 5). When normalized to the number of repeats present, the read count was similar to the low level observed throughout the genome. This low genome-wide level of H2A.Z might represent biological noise (that is, real but not necessarily meaningful).

Nucleosome positioning sequences have been largely defined by dinucleotide patterns present in a relatively small number of sequenced nucleosomes. Positioning sequences have not been defined by nucleosomes trapped at their *in vivo* location, which could differ from thermodynamically favoured positions. We examined >8,000 well-defined nucleosome positions, and found an AA/TT and GC dinucleotide pattern remarkably similar to the thermodynamically favoured pattern (Fig. 2a and Supplementary Fig. 6)[3,4]. Strikingly, a relatively strong enrichment of GC dinucleotides and a deficiency of AT dinucleotides was detected 3–4 bp from the nucleosome border; this is where the highly regulated histone H3 tails emerge from the DNA gyres[12], which might indicate a histone–DNA regulatory interaction. The deficiency of AT was not evident 1 bp in from the border, suggesting that the intrinsic MNase specificity for AT was not artificially depleting the region of AT.

Coarse-grain and fine-grain smoothing of the read distributions allowed several nucleosomal parameters to be assessed (Fig. 2b). First we investigated the basis of multiple fine-grain peaks within each coarse-grain peak. Multiple peaks might reflect nucleosome delocalization over a rather narrow range. Indeed, the fine-grain positions

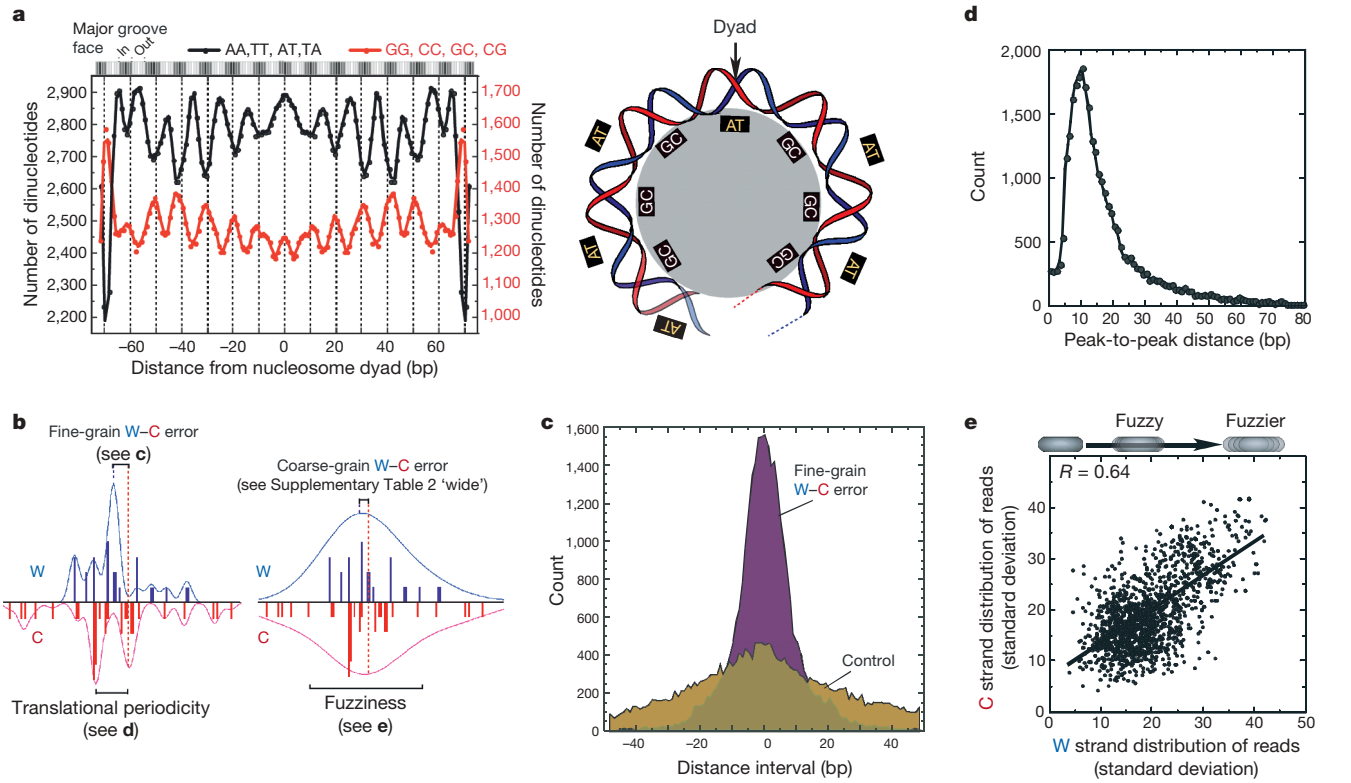

Figure 2 | Rotational and translational settings of H2A.Z nucleosomes. **a**, Smoothed frequency distribution of the combined number of AA, TT, AT, TA or GG, CC, GC, CG dinucleotides at each base pair along the 147 bp of nucleosomal DNA. The schematic shows the rotational setting of the two sets of dinucleotides. **b**, Bar graph depicting the number of reads mapped to each coordinate on the Watson (W, in blue) and Crick (C, in red) strands for an arbitrary nucleosome. Bars are separated in 1-bp increments. Fine-grain (left panel) and coarse-grain smoothing (right panel) are shown for the same data. **c**, Each coarse-grain peak was centred midway between its matched W–C pairs. Closest W–C fine-grain peaks were paired, and their base pair separation calculated. Shown in purple is a smoothed frequency distribution (bin size, 1 bp; 'count', number of fine-grain nucleosome positions) of their distances (error). As a control, the entire distribution of fine-grain C peaks within a single coarse-grain peak was rotated 180° around the centre of its coarse-grain peak, then paired off with nearest fine-grain W peaks (yellow distribution). The relatively short distance between paired W–C peaks compared with the control indicates that MNase heterogeneity, which would be non-identical on the W and C strands, cannot account for the distribution of fine-grain peaks. **d**, Peak-to-peak distances for peaks located on the same strand were measured, using the highest six peaks within each coarse-grain peak. The number of measurements at the indicated distance intervals is plotted as a smoothed frequency distribution. **e**, Scatter plot of the standard deviation of read coordinates located under a single coarse-grain peak, comparing Watson and Crick matched pairs. A total read count was required to exceed 35, to be included.

determined independently on the Watson and Crick strands were highly co-incident compared to a control set (Fig. 2c), suggesting that these peaks largely represent alternative overlapping translational settings.

The 10-bp periodicity of dinucleotides inherent in nucleosome positioning sequences theoretically restricts DNA to a single rotational setting on the histone surface. In the absence of mitigating factors, this limits nucleosomes to quantum translational settings of 10 bp. To test this prediction, we measured peak-to-peak distances between adjacent translational settings, and found the most frequent positioning interval to be 10 bp (Fig. 2d). Thus, on a genomic scale nucleosomes tend towards a primary rotational setting, resulting in discrete translational settings having 10-bp intervals.

The standard deviation of read locations within each coarse-grain nucleosome provides another metric of delocalization or 'fuzziness'. Figure 2d shows that fuzzier nucleosomes measured on the Watson strand were verifiable when examined on the Crick strand. Approximately 10% of all major H2A.Z nucleosomes were 'very fuzzy', tending towards 3–4 predominant translational settings. Most others had 1–3 predominant settings.

We next explored the biological significance of multiple translational settings by examining the properties of genes having very fuzzy H2A.Z nucleosomes (Supplementary Tables 2 and 3). Indeed, these genes tended to be TATA-containing and positively regulated by chromatin remodelling factors including SAGA, SWI/SNF and SWR-C, the latter being H2A.Z-specific. They tended to be negatively regulated by histone modifying factors such as Hda1 deacetylase and Set1 methylase, and by the SSN6–TUP1 chromatin repressor complex. Thus, we find a strong link between chromatin regulation and multiple translational settings, suggesting that nucleosome repositioning by chromatin regulators may have important global roles in regulating gene expression, particularly at TATA-containing promoters.

We also examined 'very wide' nucleosomes, which have coarse-grain Watson positions that are farther away from their matching Crick positions than expected by chance (typically >20 bp). Interestingly, genes having very wide nucleosomes also tended to be controlled by chromatin regulators, with the SSN6–TUP1 complex being particularly involved (Supplementary Table 2). SSN6–TUP1 binds to nucleosomes and forms a repressive chromatin domain[13]. Conceivably, its presence might provide additional MNase protection to nucleosomes, creating an appearance of unusually wide nucleosomes.

The geographical landscape of H2A.Z nucleosomes in the vicinity of chromosomal elements is presented in Fig. 3. With the exception of TATA-containing promoters, a characteristic H2A.Z chromosomal architecture was found to encompass distinct classes of

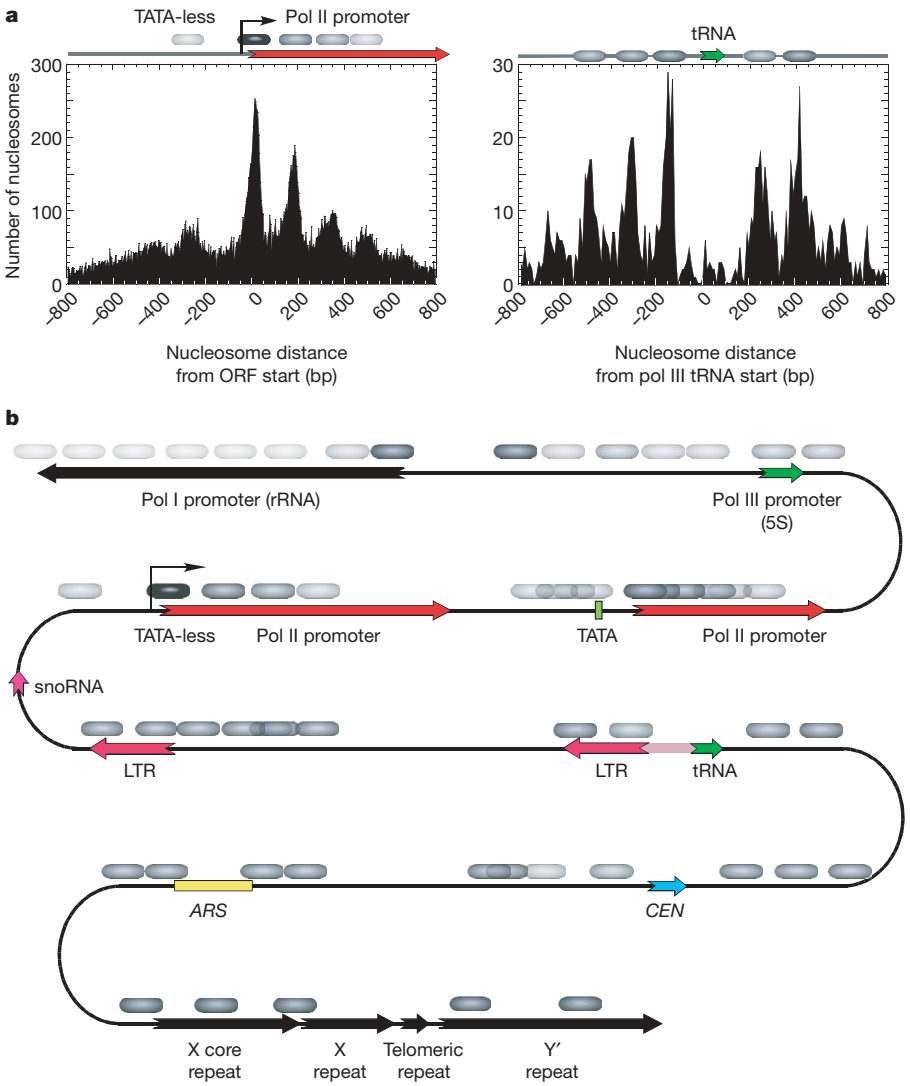

Figure 3 | Distribution of H2A.Z nucleosomes in and around chromosomal elements. **a**, Number of nucleosomes located at varying distances from either an ORF start site (left panel) or a tRNA start site (right panel) were binned in 10-bp intervals, then plotted as a smoothed frequency distribution. Plots for additional chromosomal elements are shown in Supplementary Figs 5 and 8a–m. **b**, Generalized schematic of nucleosome positions in and around chromosomal elements. The darkness of the ovals represents the relative level of H2A.Z.

chromosomal elements. Telomeric repeats and centromeric regions contained H2A.Z nucleosomes in fixed locations and spaced in ~200 bp intervals rather than the canonical 165 bp. Spacing on parts of the Y′ telomeric elements is consistent with an alternating array of standard and H2A.Z-containing nucleosomes. The specialized centromeric nucleosome was devoid of H2A.Z, presumably containing H2A instead. Thus, both H2A and H2A.Z may have geographically distinct roles in centromere function[14,15]. Origins of replication (ARSs) lacked H2A.Z, which is consistent with studies on individual ARSs[16,17]. However, H2A.Z nucleosomes were positioned on flanking genomic DNA, suggesting that the ARS, or it cognate factors, establishes an H2A.Z-containing chromatin architecture surrounding the ARS, perhaps functionally analogous to the chromatin architecture found at promoter regions.

Ribosomal RNA is transcribed by RNA polymerase (pol) I from a single type of promoter. Two H2A.Z nucleosomes were present at the pol I promoter (Fig. 3b), at +13 and −415, establishing a 268-bp nucleosome-free region akin to that at pol II promoters[6]. An H2A.Z nucleosome was also positioned over the start of the 18S rRNA. Because 18S rRNA is generated by RNA cleavage rather than transcription initiation, the significance of this H2A.Z positioning is unclear. H2A.Z nucleosomes may have an important role at pol I promoters, possibly through their established function in counteracting silent information regulator (SIR)-mediated silencing.

The arrangement of nucleosomes at pol II promoters is now established[4–6,9] as H2A.Z nucleosomes flanking a nucleosome-free promoter region[8–11]. Computational modelling predicts that TATA-containing and TATA-less promoters have distinct nucleosomal architecture[5]. Indeed, TATA-less promoters, which constitute the vast majority of the genome[18], showed the canonical architecture including 165-bp inter-nucleosomal spacing (Fig. 3). In contrast, TATA-containing promoters lacked a canonical architecture (Supplementary Fig. 8f). This scattered distribution was not due to a complete delocalization of individual nucleosomes because nucleosomes were positioned at individual promoters, albeit fuzzy. Rather, each TATA-containing promoter has a tailored chromatin architecture, which may explain their greater dependency on chromatin remodelling[5,18].

Highly transcribed genes tend to be depleted of nucleosomes[19–21]. We examined whether this might be due to loss of nucleosomes at specific locations in the promoter region, or whether there was a general depletion. Three groups of genes were examined, including the highly transcribed ribosomal protein genes, the less-transcribed ribosome biogenesis genes, and all highly transcribed genes (Supplementary Fig. 8h–j)[22]. In all cases, the translational positioning of nucleosomes was unperturbed over the open reading frame (ORF) start region as was the case with most genes and individual genes[23]. Instead, H2A.Z levels were reduced, reflecting nucleosome depletion. Transcription of the ORF start region might remove the intervening nucleosome. But, once pol II has passed, the nucleosome returns to its original unperturbed location, perhaps guided by strong positioning sequences.

Pol III-transcribed transfer RNA genes tended to reside in the midst of ~250 bp H2A.Z-free regions, which were flanked by several uniformly positioned nucleosomes with a low H2A.Z content (Fig. 3). Typically two of these flanking nucleosomes housed a long terminal repeat (LTR) from a Ty retrotransposon (Supplementary Fig. 8m), which is often physically linked to tRNA genes[24]. Thus, H2A.Z might have undiscovered roles in transposition.

We mapped the location of promoter elements relative to the nearest nucleosome, the orientation of which is defined by the nearest ORF. As shown in Fig. 4a, TATA boxes were scattered across the downstream border of the first upstream nucleosome, which is consistent with the heterogeneous chromatin architecture of TATA-containing promoters[5] and their dependence on chromatin modifying and remodelling factors[18]. In contrast, transcriptional start sites were concentrated ~13 bp inside the upstream border of the +1 nucleosome, which tended to be less fuzzy than other nucleosomes (Fig. 4b). This location places the histone H3 amino-terminal tail just upstream of the transcriptional start site, where it could regulate transcription initiation (see Fig. 4e). Indeed, the highly regulated H3 K36 site resides just outside of the DNA gyre, placing it near the start site.

Evolutionarily conserved binding sites for over 100 gene-specific transcription factors have been identified in promoter regions throughout the yeast genome[25,26]. Only a small fraction of these sites have been reported to bind transcription factors[27,28], which brings into question the physiological significance of the vast majority of unbound sites. Irrelevant sites should be randomly distributed without regard to chromatin structure. However, we find that conserved but unbound transcription factor binding sites tend to have a topological relationship with their nucleosome neighbour (Fig. 4c). These sites tended to reside near nucleosome borders, and were concentrated in 10-bp periodicities that aligned with the rotationally exposed DNA major groove on the histone surface. This pattern suggests that many transcription factors recognize their cognate sites on the nucleosome surface. Their placement near nucleosome borders positions them to control the translational setting of nucleosomes. Consistent with this finding, sites that were bound by

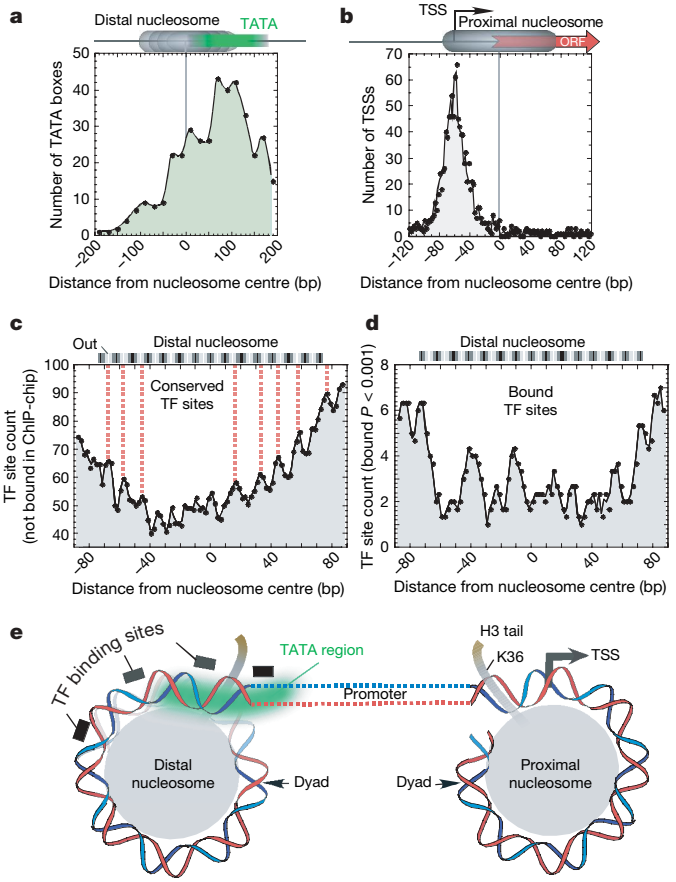

Figure 4 | Distribution of transcriptional regulatory elements along the nucleosome surface. a–d, The distance from each element to its nearest nucleosome centre was measured, binned, and plotted as a smooth distribution. Locations of TATA boxes, transcriptional start sites (TSSs) and transcription factor (TF) binding sites are published[18,29,30]. The topological relationships of individual unbound TF sites with nucleosomes are presented in Supplementary Fig. 9. ChIP-chip, chromatin immunoprecipitation performed on a genome-wide scale. The data sources in panels **c** and **d** were files 'nobind_c3.gff' (stringent conservation and no binding *P*-value cutoff) and 'p001_c3.gff' (stringent conservation and binding $P < 0.001$), respectively, downloaded from http://fraenkel.mit.edu/improved_map/latest_maps.html. **e**, Representation of the rotational setting of TF binding sites and TSSs on the nucleosome surface.

sequence-specific regulators[27–29] tended to reside just outside of the nucleosome border (Fig. 4d). Thus, a potential consequence of transcription factors binding to nucleosomal DNA is the translational displacement of the nucleosome, which might be important for gene activation (Fig. 4e). The H2A.Z nucleosome map presented here reveals a tight regulatory relationship between promoter elements and the topology of nucleosome borders.

METHODS

S. cerevisiae (BY4741 *HTZ1–TAP*) was grown to mid-log phase in rich media, then subjected to formaldehyde crosslinking[6]. *HTZ1–TAP* encodes carboxy-terminally TAP-tagged H2A.Z. Collected cells were disrupted, and non-nucleosomal DNA eliminated with MNase. H2A.Z–TAP nucleosomes were immunopurified with immunoglobulin G-sepharose and eluted with the Tobacco Etch Virus (TEV) protease. The resulting 125–170-bp DNA was gel-purified and sequenced with the Genome Sequencer 20. Sequencing reads were mapped to the reference *S. cerevisiae* genome at http://www.yeastgenome.org. Additional details can be found in Supplementary Information.

Received 20 October 2006; accepted 26 January 2007.

1. Martinez-Campa, C. *et al.* Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1. *Mol. Cell* **15,** 69–81 (2004).
2. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423,** 145–150 (2003).
3. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191,** 659–675 (1986).
4. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442,** 772–778 (2006).
5. Ioshikhes, I. P., Albert, I., Zanton, S. J. & Pugh, B. F. Nucleosome positions predicted through comparative genomics. *Nature Genet.* **38,** 1210–1215 (2006).
6. Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309,** 626–630 (2005).
7. Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122,** 517–527 (2005).
8. Li, B. *et al.* Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl Acad. Sci. USA* **102,** 18385–18390 (2005).
9. Raisner, R. M. *et al.* Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. *Cell* **123,** 233–248 (2005).
10. Zhang, H., Roberts, D. N. & Cairns, B. R. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123,** 219–231 (2005).
11. Guillemette, B. *et al.* Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol.* **3,** e384 (2005).
12. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–260 (1997).
13. Malave, T. M. & Dent, S. Y. Transcriptional repression by Tup1–Ssn6. *Biochem. Cell Biol.* **84,** 437–443 (2006).
14. Pinto, I. & Winston, F. Histone H2A is required for normal centromere function in *Saccharomyces cerevisiae*. *EMBO J.* **19,** 1598–1612 (2000).
15. Krogan, N. J. *et al.* Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4. *Proc. Natl Acad. Sci. USA* **101,** 13513–13518 (2004).
16. Lipford, J. R. & Bell, S. P. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol. Cell* **7,** 21–30 (2001).
17. Nieduszynski, C. A., Knox, Y. & Donaldson, A. D. Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev.* **20,** 1874–1879 (2006).
18. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116,** 699–709 (2004).
19. Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O. & Schreiber, S. L. Global nucleosome occupancy in yeast. *Genome Biol.* **5,** R62 (2004).
20. Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Removal of promoter nucleosomes by disassembly rather than sliding *in vivo*. *Mol. Cell* **14,** 667–673 (2004).
21. Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genet.* **36,** 900–905 (2004).
22. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95,** 717–728 (1998).
23. Gavin, I. M. & Simpson, R. T. Interplay of yeast global transcriptional regulators Ssn6p–Tup1p and Swi–Snf and their effect on chromatin structure. *EMBO J.* **16,** 6263–6271 (1997).
24. Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A. & Voytas, D. F. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* **8,** 464–478 (1998).
25. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254 (2003).
26. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301,** 71–76 (2003).
27. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431,** 99–104 (2004).
28. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298,** 799–804 (2002).
29. MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7,** 113 (2006).
30. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA* **103,** 5320–5325 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Albert for computational support, and S. Tan, J. Reese, D. Gilmour, and Y. Wang for many helpful discussions. This work was supported by a grant from NIH.

Author Contributions I.A. developed computational approaches to derive nucleosome maps from the read locations and developed the associated browser; T.N.M. prepared and purified the nucleosomes; L.P.T. constructed libraries and sequenced nucleosomal DNA; J.Q. mapped sequencing reads to the yeast genome; S.J.Z. produced Supplementary Tables 2 and 3; S.C.S. managed and directed the DNA sequencing phase, and analysed the sequencing data; and B.F.P. directed the project, analysed the processed data, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.F.P. (bfp2@psu.edu).