

Figure S1. Similarity between *in vivo* and *in vitro* assembled nucleosomes around promoters and terminators (defined as 1 kb regions centered by TSSs or TTSs).

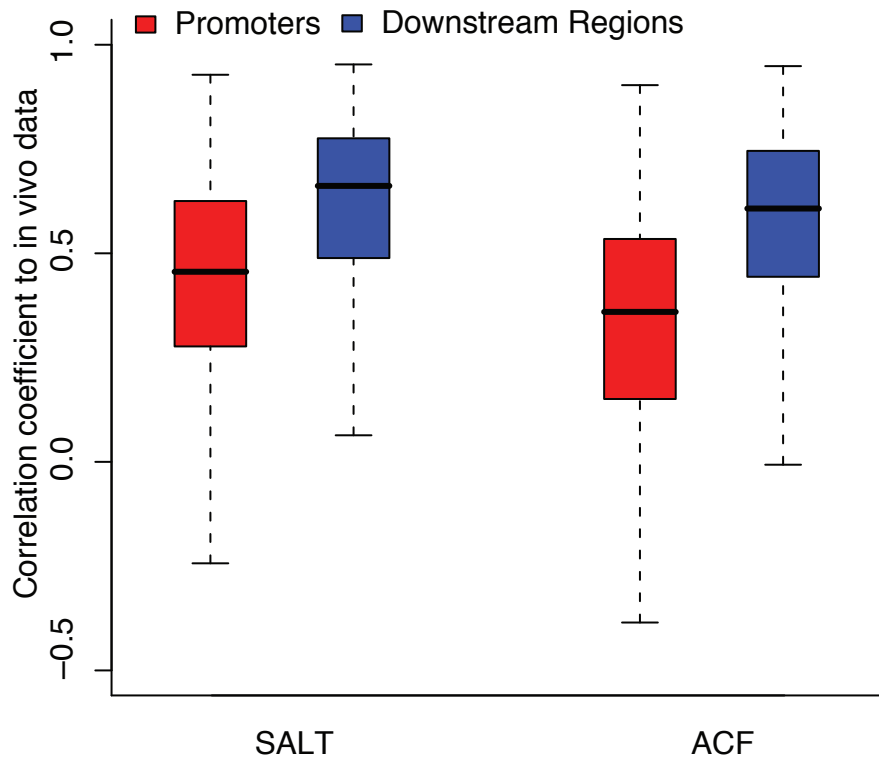


Figure S2. Bar plot shows the comparison of maximum positioning degree at +1 through +10 nucleosome locations between salt dialysis assembled chromatin and the *in vivo* nucleosome data (YPD). X-axis shows the approximate locations or +1 through +10 nucleosomes derived from average nucleosome density profile.

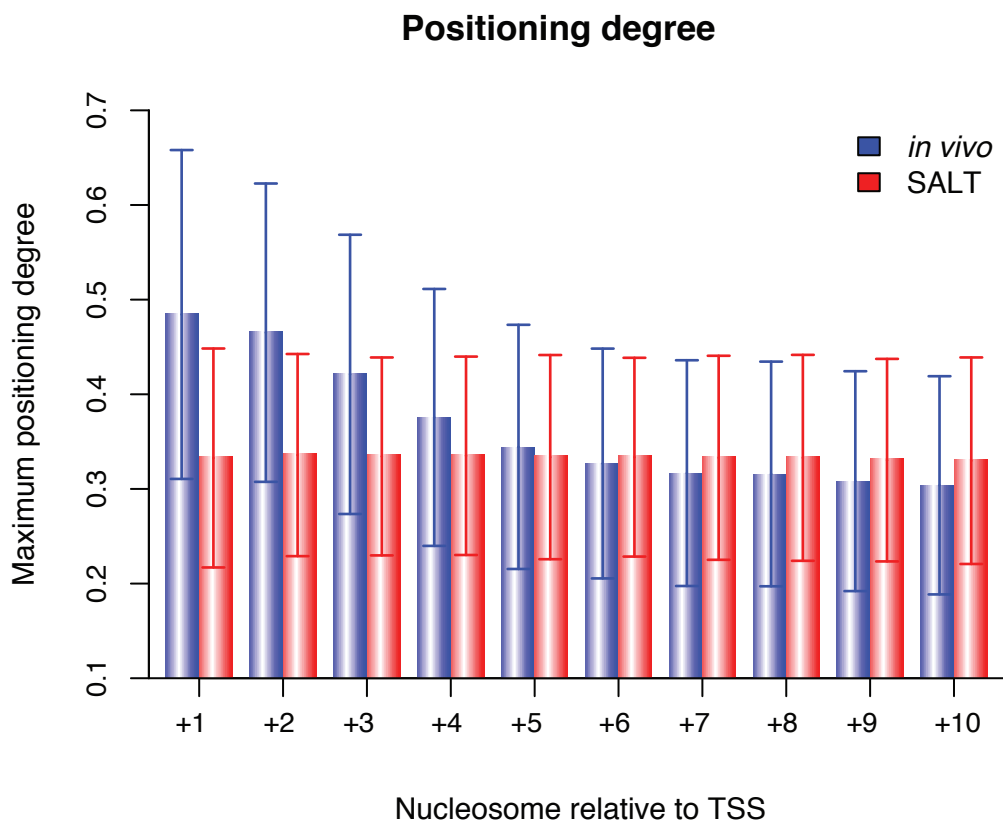


Figure S3. Comparison of statistical positioning between *in vivo* nucleosomes (YP ethanol medium) and other data sets (eliminating random influence). For each sample, we generated control data with the same number of randomly tags in gene bodies, and similar to the approaches of generating Fig. 4b-d, distributions of the tag numbers of control samples were also calculated. Plots show the differences between Fig. 4b-d and the distribution of corresponding control samples.

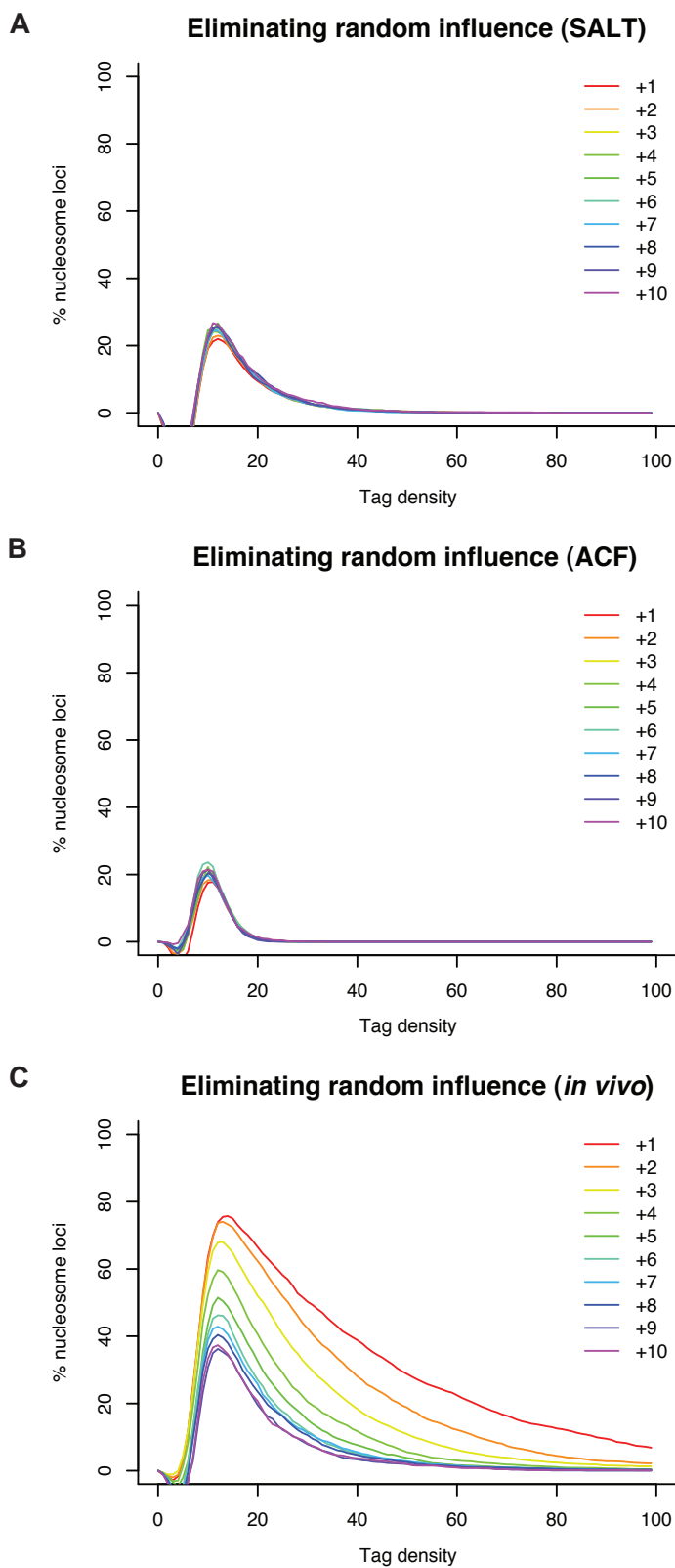


Figure S4. *In vivo* nucleosome density profiles around the binding sites of transcription factors (a) Rap1, (b) Mbp1, (c) Cbf1, (d) Swi6, (e) Msn4 and (f) Pho2.

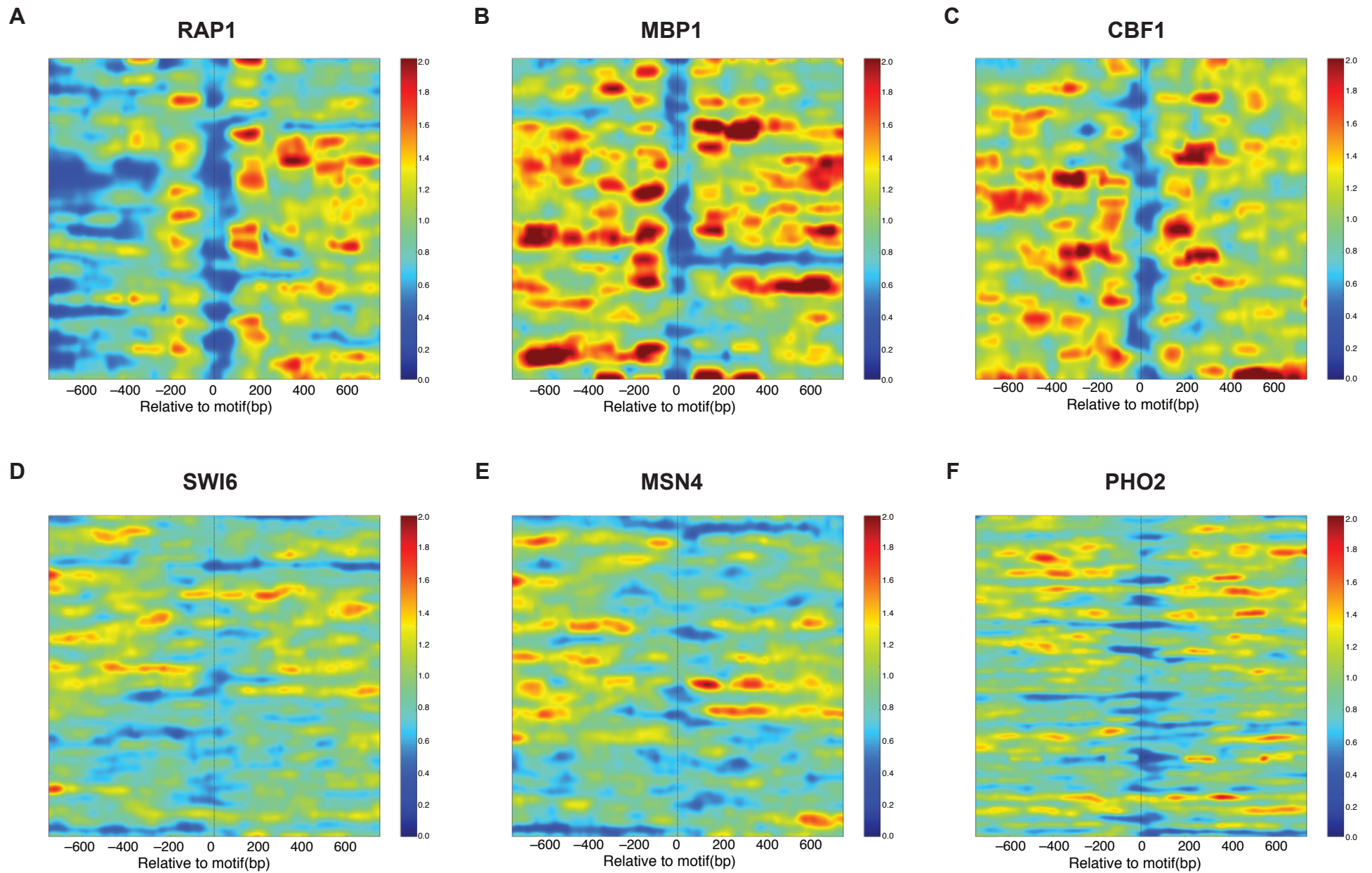


Figure S5. Analysis of nucleosomes generated by salt dialysis by Kaplan *et al.* *in vitro* Nucleosome density profile around TTSs of genes with isolated 5' ends for (a) *in vitro* data (Kaplan *et al.*), and (b) the *in vivo* nucleosome data (YPD). (c) Start-to-start distances of tags in the same strand. We randomly sampled 3.27 M for *in vitro* data (Kaplan *et al.*), and we aligned “1-pile” nucleosome 5' end set to plot the distribution.

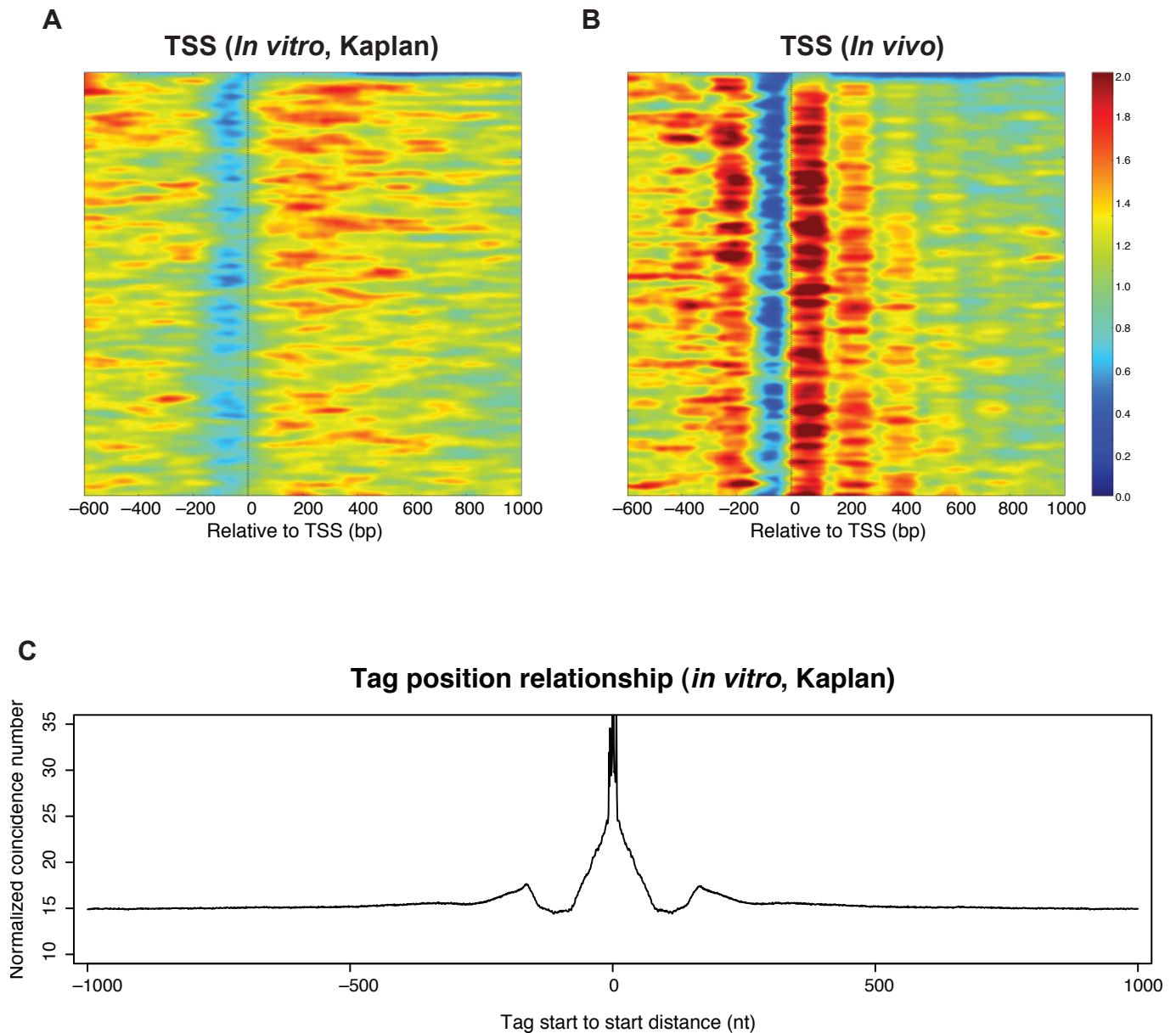
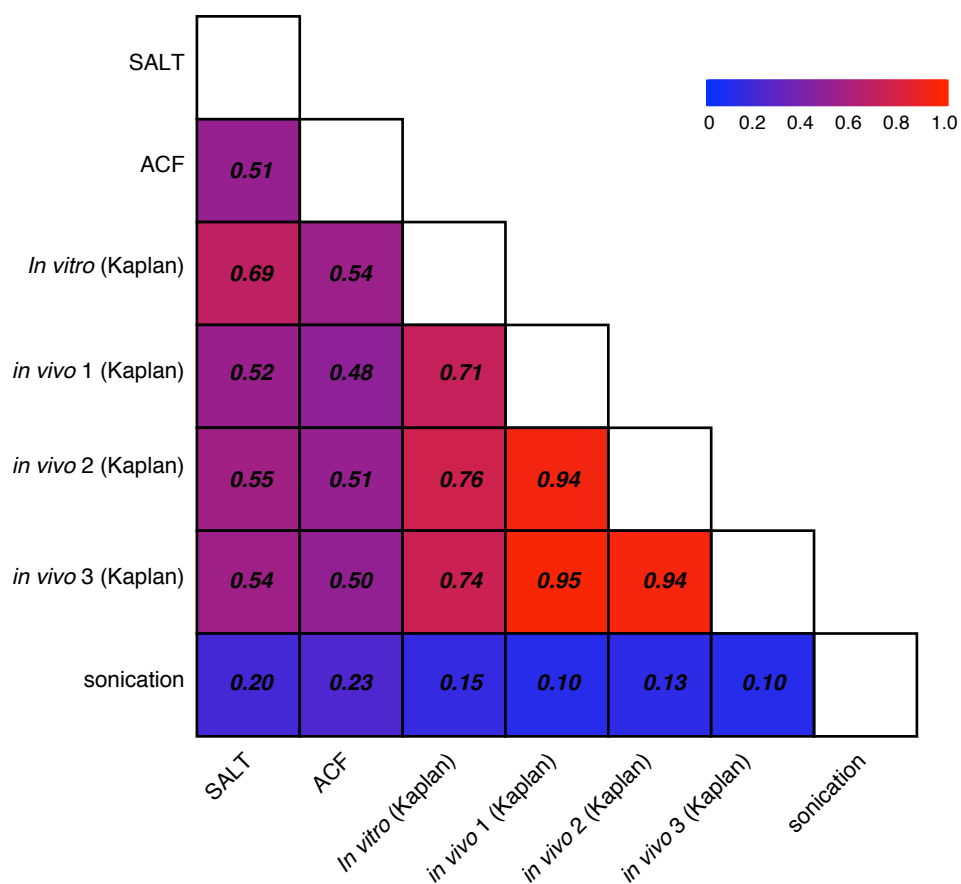


Figure S6. Overall similarity among datasets. For each dataset, each tag was extended to 146 bp and then piled-up. Similarity of nucleosome density profiles in mappable genomic regions is measured by Spearman's rank correlation coefficient.



Supplementary Methods

In vitro nucleosome assembly and mapping

Yeast DNA was prepared from exponentially growing cells that were treated with Zymolyase. The spheroplasts were in Qiagen buffer G2 with 200 μ g/ml RNase A were incubated with Proteinase K (8mg, Qiagen) for 30 minutes at 50°C. After removal of cellular debris by centrifugation, DNA was precipitated with ethanol and isolated by spooling, and was further purified using a Qiagen Genomic-tip 500/G according to the manufacturer's instructions. Purified genomic DNA preparations from *Saccharomyces cerevisiae* as well as from *Escherichia coli* were sonicated separately to yield fragments that ranged in length from 5 to 10 kb. The resulting yeast and *E. coli* DNA samples were combined in a 3:1 mass ratio, and assembled into chromatin by using a purified system containing recombinant *Drosophila* NAP-1 and ACF as well as purified native histones from *Drosophila* embryos as described previously¹. Three independent chromatin assembly reactions were performed on the same DNA mixture. Chromatin was extensively digested with micrococcal nuclease to yield core particles. The ~147 bp DNA fragments derived from the core particles were purified by agarose gel electrophoresis. Mononucleosomal DNA was treated with calf intestinal alkaline phosphatase (New England Biolabs) to remove 3' phosphate groups left after the micrococcal nuclease treatment.

Libraries were prepared according to Illumina's instructions accompanying the DNA Sample Kit (Part# 0801-0303). Briefly, DNA was end-repaired using a combination of T4 DNA polymerase, *E. coli* DNA Pol I large fragment (Klenow polymerase) and T4 polynucleotide kinase. The blunt, phosphorylated ends were treated with Klenow fragment (3' to 5' exo minus) and dATP to yield a protruding 3' 'A' base for ligation of Illumina's adapters which have a single 'T' base overhang at the 3' end. After adapter ligation DNA was PCR amplified with Illumina primers for 16 cycles and library fragments of ~250 bp (insert

plus adaptor and PCR primer sequences) were band isolated from an agarose gel. The purified DNA was captured on an Illumina flow cell for cluster generation. Libraries were sequenced on the Genome Analyzer following the manufacturer's protocols. Sequence tags were obtained and aligned to the *Saccharomyces cerevisiae* from SGD² (Apr 2008 build) and *Escherichia coli* K12 MG1655 (U00096) genomes using the Illumina/Solexa Analysis Pipeline. The analysis allowed 2 mismatches per mapped read and only uniquely aligned reads were retained.

Other data sets

Yeast gene annotations were derived from expression tiling array results³, which were also used in a recent study⁴. *In vivo* yeast gene expression level was measured by mRNA abundance by combining multiple microarray datasets⁵. Transcription factor binding sites in yeast were obtained from MacIsaac *et al.*⁶, and binding p-value cutoff 0.005 and moderate conservation cutoff were used to define binding sites. Three yeast *in vivo* nucleosome maps were taken from a recent study⁷, with 24.1 M (YPD), 15.3 M (YPEtOH) and 12.5 M (YPGal) uniquely mapped sequencing tags separately. The same study also has an independent *in vitro* nucleosome map⁷, which was discussed in this study. Genome coordinates of all yeast data sets used in this study were converted to Apr 2008 build of SGD².

For each 28-mer in the yeast genome, we examined whether it can be uniquely mapped to the mixture of yeast and *E. coli* genomes. Mappable ratio was defined for each genomic region (with length larger than 28 bp) as: # (uniquely mapped 28-mers) / # (all 28-mers in the region). In yeast genome, 93.5% of 100 bp windows have mappable ratio ≥ 0.8 , and only these mappable regions were used in this study to remove the influence of sequencing tag mapping.

Nucleosome density profile

In this study, each sequencing tag represents one end of a mono-nucleosomal DNA fragment. Therefore, we extended each tag to 146 bp in the 3' direction to represent a whole mono-nucleosome, and then we piled up all extended sequencing tags in a sample to obtain the nucleosome density profile along the genome. For each sample, we generated heat maps of nucleosome density profiles around certain crucial features (including transcription start sites, TSSs; transcription termination sites, TTSs; transcription factor binding sites, TFBSs) with following approaches. First, in order to derive reliable nucleosome density profiles, only a subset of the TSS or TTS entries were selected as follows. For heat map around TSSs, we picked up 1,752 TSSs of genes with isolated promoters (, whose [-1kb, 0] to TSSs have no overlapping with other genes); while for heat map around TTSs, we chose 1,548 TTSs of genes with isolated terminator regions (, whose [0, +1kb] to TTSs have no overlapping with other genes). Secondly, we aligned nucleosome density profiles by given features, with upstream regions in the left side and downstream in the right; in order to do comparison between samples with different sequencing depth, we normalized the density profiles to enrichment ratios. Finally, we ranked the entries either by gene expression levels (TSSs and TTSs) or by genomic loci (TFBSs), and then we performed Gaussian kernel smoothing in heat maps.

Sequence dinucleotide periodicity

Similar to a recent study in *C. elegans*⁸, for each sample, we defined “1-pile” nucleosome 5' end set as the genomic loci (strand distinguishable) where one or more sequencing tags' 5' ends are, and accordingly, “5-pile” set as genomic loci with five or more tags' 5' ends. For *in vitro* nucleosome samples, we aligned “1-pile” nucleosome 5' end set to generate AA/TT/AT dinucleotide fraction pattern; while for *in vivo* nucleosome sample (YPD), “5-pile” set were aligned to generate dinucleotide fraction pattern. Power spectrum analysis is a popular technique to detect the power in each frequency components of a numerical signal. Discrete Fourier Transform (DFT) was performed on [+11, +160] interval

of AA/TT/AT fraction pattern to measure the power of 10 bp periodicity for each sample, and the fraction pattern was first subtracted its mean to remove the DC component.

We applied a similar approach used in a recent study⁹ to perform power spectrum analysis of AA/TT/AT dinucleotide pattern for yeast and *E. coli* genomic DNA. We first converted yeast or *E. coli* genome DNA sequence to a binary sequence, according to whether AA/TT/AT present at each dinucleotide position. Then we split the binary sequence into 1,024 bp fragments, and we applied DFT on each fragment. The averaged result of all fragments was used to show the power spectrum of AA/TT/AT dinucleotide pattern in genomic DNA.

Tag position relationship

To compare the position relationship of sequencing tags between different samples, we randomly sampled the same number of tags of *in vitro* salt dialysis nucleosome data (3.27 M) for both *in vivo* (YPD) nucleosome data and *in vitro* ACF-assembled nucleosome data. Similar to a recent study in *C. elegans*⁸, for each sample, we aligned “1-pile” nucleosome 5’ end set to plot the distribution of the start-to-start distances of tags in the same strand.

Nucleosome positioning degree

In principle, a nucleosome will exclusively occupy a DNA fragment with length ~160 bp (nucleosome DNA + linker), and in a cell population, the distribution of nucleosome centers in a genomic locus reflects the local nucleosome positioning degree, while more disperse distribution corresponds to less positioning degree. In this study, we proposed a straightforward definition of nucleosome positioning degree as the local nucleosome center proportion. Two steps are performed to infer the positioning degree in each genomic locus. 1) The first nucleotide in the 5’ end of each sequencing tag is shifted 73 bp towards its downstream, and then it presents the nucleosome center. 2) For each genomic location, we count the number of nucleosome centers within a 20 bp window centered by the location (to

allow for imprecision of MNase cleavage), denoted as $N_{20\text{ bp}}$, and we also count the number of nucleosome centers in the 160 bp window centered by the location (160 bp is used as the length of exclusive DNA fragment occupied by mono-nucleosome), denoted as $N_{160\text{ bp}}$. Nucleosome positioning degree in each genomic location is then calculated as $N_{20\text{ bp}} / \max(N_{160\text{ bp}}, N_{\text{average}})$, where N_{average} is used to eliminate the biases raised from small samples, calculated as $160 \times (\text{total tag number}) / (\text{mappable genome size})$. The values of positioning degree range from 0 to 1.

The maximum positioning degree value within a 160 bp window is the positioning degree of most “positioned” nucleosomes contained in this region, and thus the percentage of genomic regions with maximum positioning degree larger than certain threshold (which means the percentage of genomic regions containing “positioned” nucleosomes above certain threshold) is a direct measurement to reflect the nucleosome positioning status globally. In this study, in order to compare the nucleosome positioning status between samples, we randomly sampling the same number of tags of *in vitro* salt dialysis nucleosome data (3.27 M) for both *in vivo* (YPD) nucleosome data and *in vitro* ACF-assembled nucleosome data, and we also generated control data with same number of random tags in mappable genomic regions.

Nucleosome pattern in genes

To perform the analysis of nucleosome pattern in genes, we first determined the approximate locations of +1 through +10 nucleosomes relative to TSS from *in vivo* nucleosome data. In details, we picked up 3,774 non-overlapping genes (without overlapping to other genes), and we aligned the nucleosome density profiles by TSSs. The average nucleosome density profile along gene body was then generated, and the approximate locations of +1 through +10 nucleosomes were obtained from the average profile, for example, the *in vivo* +1 nucleosomes are roughly centered at 54 bp downstream TSS.

In addition to the approximate locations from average nucleosome density profile, we also detected the precise locations of *in vivo* +1 through +10 nucleosomes for each non-overlapping gene as follows. We re-implemented and slightly modified the algorithm of detecting “positioned” nucleosomes with stability score in a recent study¹⁰. In that study, tags mapped to plus and minus strands were processed separately, while in our approach, we shifted the first 5’ end nucleotide of each tag 73 bp towards its downstream, and combined both plus and minus strand tags together. We applied the algorithm to *in vivo* nucleosome data (YPEtOH), and we identified ~70,000 “positioned” nucleosomes with stability score larger than 0.2. Then for *in vivo* data (YPEtOH), the precise location of certain nucleosome (e.g. +1 nucleosome) for a given gene is defined as the location of “positioned” nucleosome with the largest stability score (0.2 or larger) within the 100 bp window centered at the approximate nucleosome location. In this study, some genes may not have all +1 through +10 nucleosomes, due to the limited gene length, or the lacking of “positioned” nucleosomes around some loci. Same approaches are applied to detect the gene-based +1 nucleosome center locations for both *in vitro* and *in vivo* nucleosome data.

Supplementary References

1. Fyodorov, D.V. & Kadonaga, J.T. Chromatin assembly in vitro with purified recombinant ACF and NAP-1. *Methods Enzymol* **371**, 499-515 (2003).
2. Cherry, J.M. et al. SGD: Saccharomyces Genome Database. *Nucleic Acids Res* **26**, 73-9 (1998).
3. David, L. et al. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A* **103**, 5320-5 (2006).
4. Mavrich, T.N. et al. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Research* (2008).
5. Beyer, A., Hollunder, J., Nasheuer, H.P. & Wilhelm, T. Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol Cell Proteomics* **3**, 1083-92 (2004).
6. MacIsaac, K.D. et al. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**, 113 (2006).
7. Kaplan, N. et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* (2008).
8. Valouev, A. et al. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Research* (2008).
9. Chen, K. et al. A novel DNA sequence periodicity decodes nucleosome positioning. *Nucleic Acids Res* **36**, 6228-36 (2008).
10. Shivaswamy, S. et al. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol* **6**, e65 (2008).