Mapping nucleosome positions using DNase-seq:

Supplemental material

Jianling Zhong[1,2], Kaixuan Luo[1,2], Peter S. Winter[3,4], Gregory E. Crawford[1,3,5,6], Edwin S. Iversen[1,7], and Alexander J. Hartemink [*1,2,6,7]

[1] *Program in Computational Biology and Bioinformatics*

[2] *Department of Computer Science*

[3] *Program in Genetics and Genomics*

[4] *Department of Pharmacology and Cancer Biology*

[5] *Department of Pediatrics, Division of Medical Genetics*

[6] *Center for Genomic and Computational Biology*

[7] *Department of Statistical Science*

Duke University, Durham, North Carolina 27708, USA

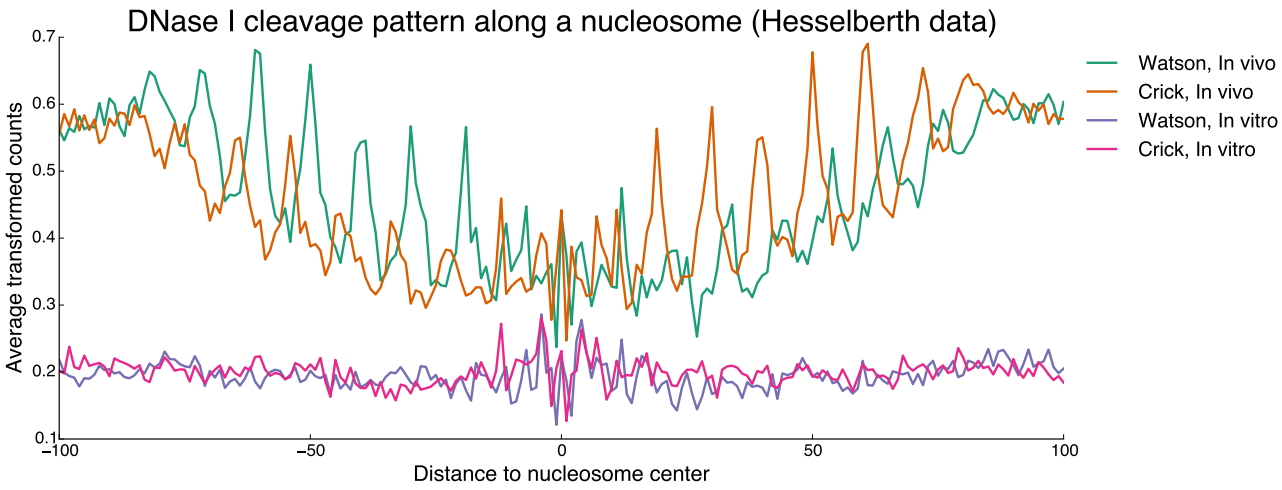
January 13, 2016

[*]Corresponding author. E-mail: amink@cs.duke.edu; office: (919) 660-6514; fax (919) 660-6519

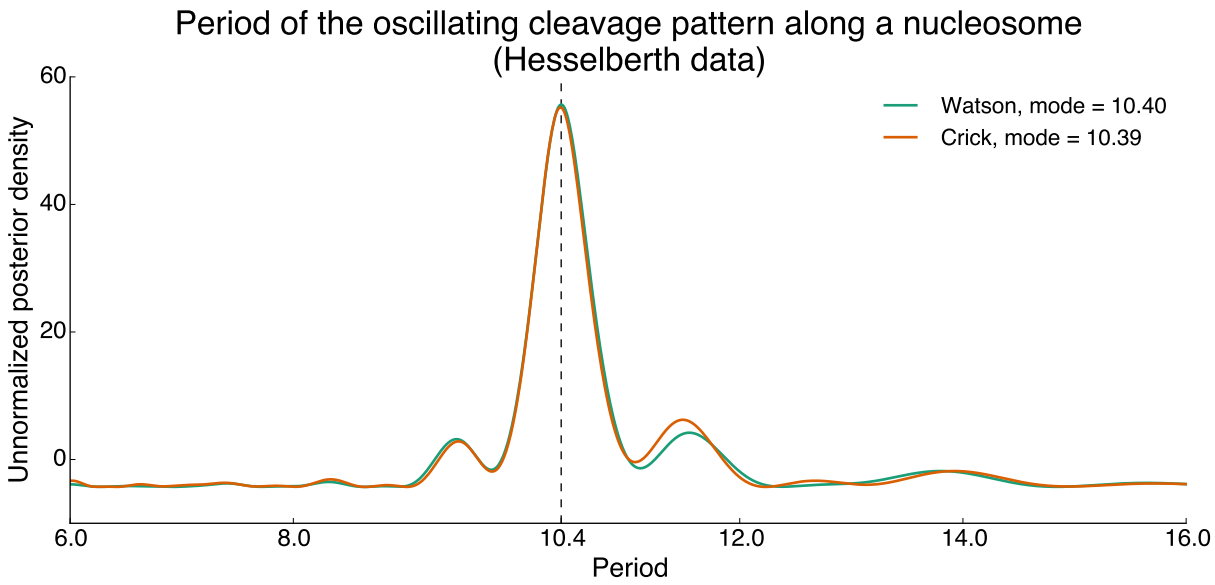
1 DNase-seq experimental protocol

Intact nuclei were prepared as before (Henikoff et al., 2011) with some alterations. Unfixed yeast cells (asynchronous W303 cells grown in YPD to OD_{600} 0.7) were centrifuged at 2000 rpm for 5 min, washed with sterile water, and resuspended in 20 mL buffer Z (0.56 M sorbitol, 7.4 pH 50 mM Tris, autoclaved). After resuspension, 14 μL β-mercaptoethanol and 500 μL of 10 mg/mL zymolyase dissolved in buffer Z were added. Samples were incubated on the benchtop for 30 min and inverted every few minutes. Cells were then centrifuged at 1500 rpm for 6 min at 4 °C and resuspended in 2.5 mL modified NP buffer (1 M sorbitol, 50 mM NaCl, 7.4 pH 10 mM Tris, 5 mM $MgCl_2$) supplemented with 0.5 mM spermidine, 0.007% β-mercaptoethanol, and 0.075% NP-40. Next, dilutions of DNase I were prepared on ice to determine the best digestion conditions. In 1.5 mL tubes, 400 μL of cell mixture from above was added to 12 μL of the following DNase I solutions: 0.03, 0.1, 0.3, and 1 U/μL. Samples were inverted once and incubated at 37 °C for 16 min. After incubation, 100 μL of stop buffer (5% SDS, 50 mM EDTA) was added to terminate the reaction. Proteinase K (0.2 mg/mL) was added to each tube, and the tubes were then inverted to mix and placed at 65 °C overnight. The next day, DNA was recovered by phenol extraction and isopropanol precipitation and run on a 0.8% agarose gel for 3 hours at 85 V to check sample digestion. The samples that had the optimal amount of DNase I digestion were then prepared for sequencing following the Crawford DNase-seq protocol (Song and Crawford, 2010).

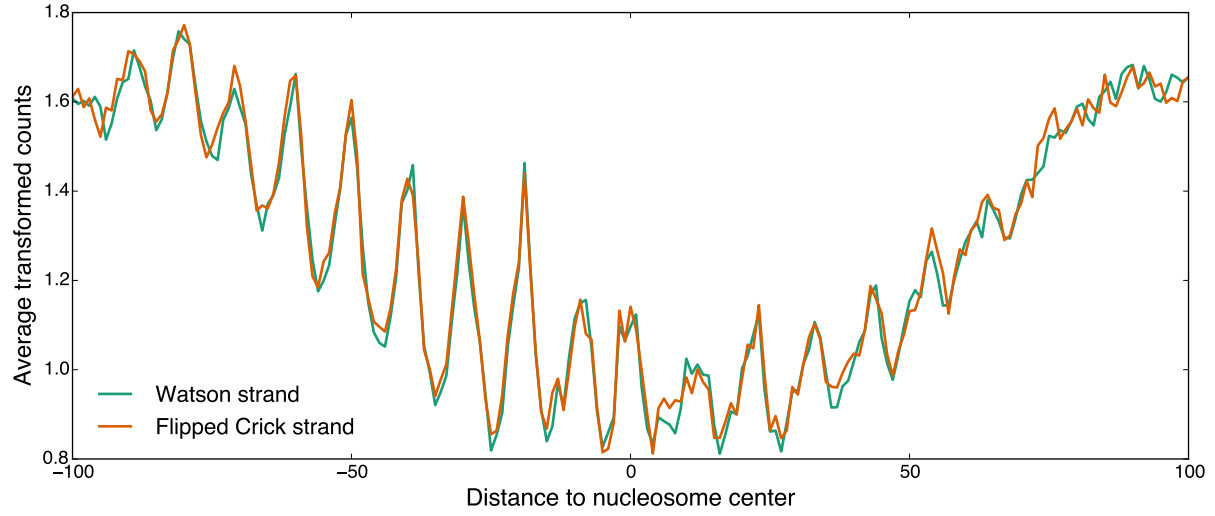
2 Supplemental figures



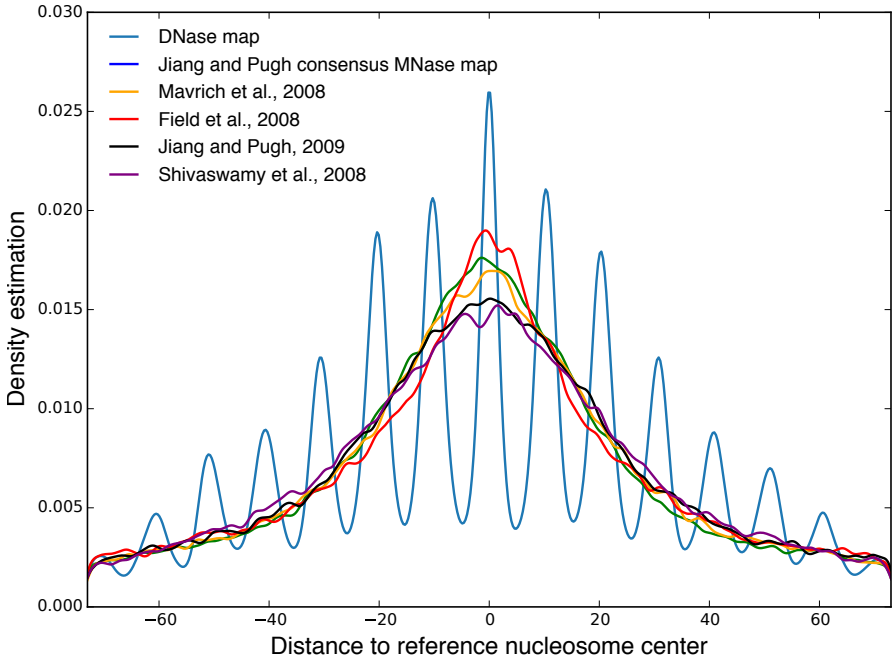
Supplemental Figure 1: Strand-specific cleavage profile of DNase I along the nucleosome, computed by averaging DNase-seq counts in the Hesselberth data (transformed by the inverse hyperbolic sine function) within the 2,000 most strongly positioned nucleosome sites in the yeast genome. The same average is calculated on both *in vivo* (green and orange) and *in vitro* (purple and red) data. The distinct DNase I cleavage profile observed in the *in vivo* data is completely absent in the *in vitro* data. We observe that the DNase I digestion pattern near the dyad is slightly noisier in the Hesselberth data (both *in vivo* and *in vitro*) than in our data, perhaps due to slight differences in the protocols. This figure is similar to Figure 1A, but is computed using Hesselberth data.



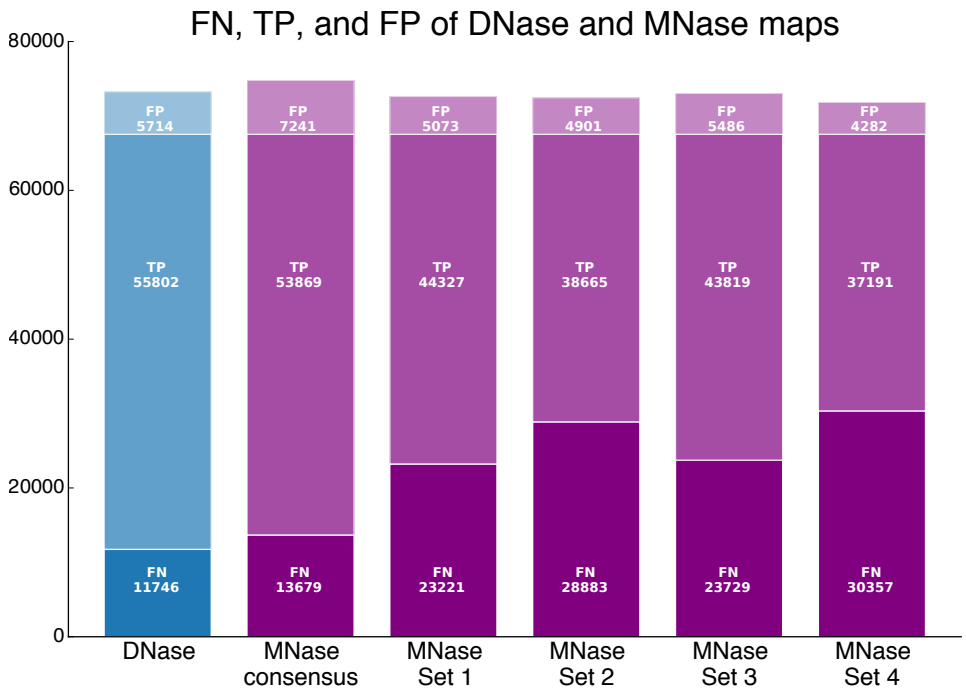
Supplemental Figure 2: Posterior density of the period of oscillation in the Hesselberth data, as determined by Bayesian harmonic regression. The most probable period *a posteriori* for each of the two different strands is around 10.4 bp. This figure is similar to Figure 1C, but is computed using Hesselberth data.



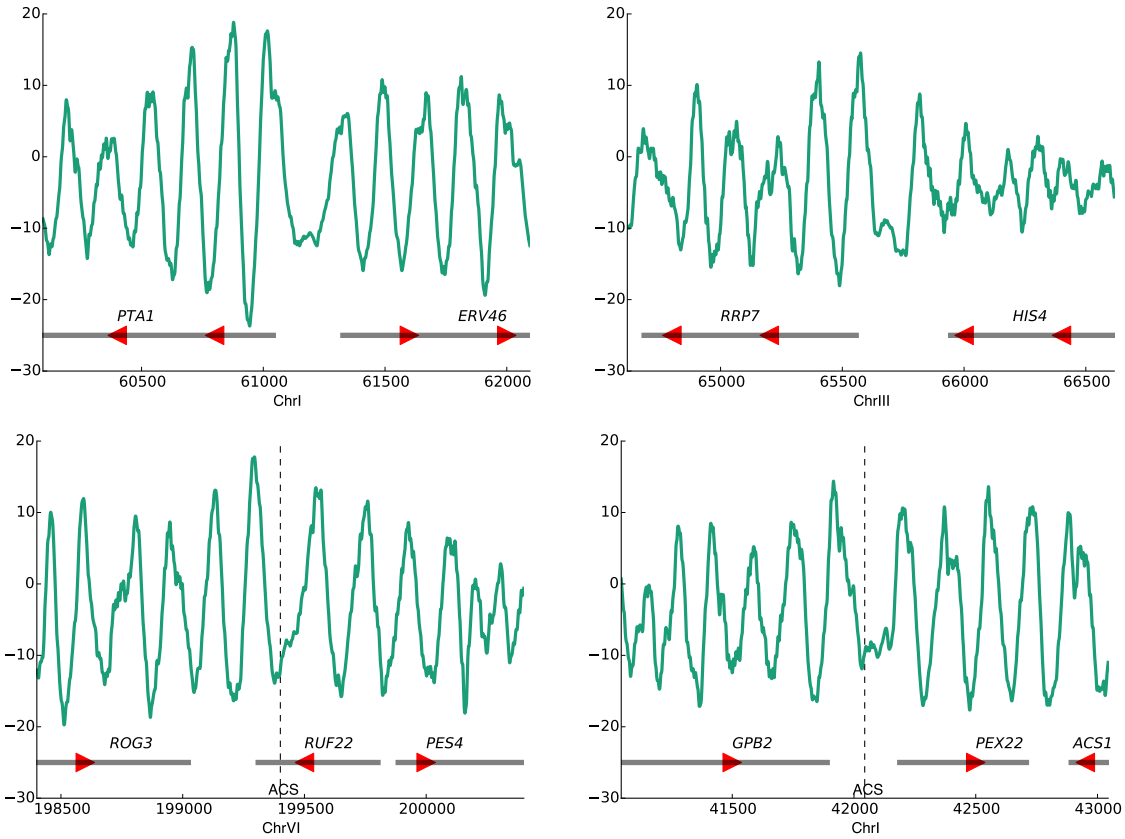
Supplemental Figure 3: DNase I cleavage profiles in our data along the nucleosome for the Watson and Crick strands are almost exactly mirror symmetric with each other. This figure is similar to Figure 1A, but with the Crick strand profile flipped around the line $x = 0$.



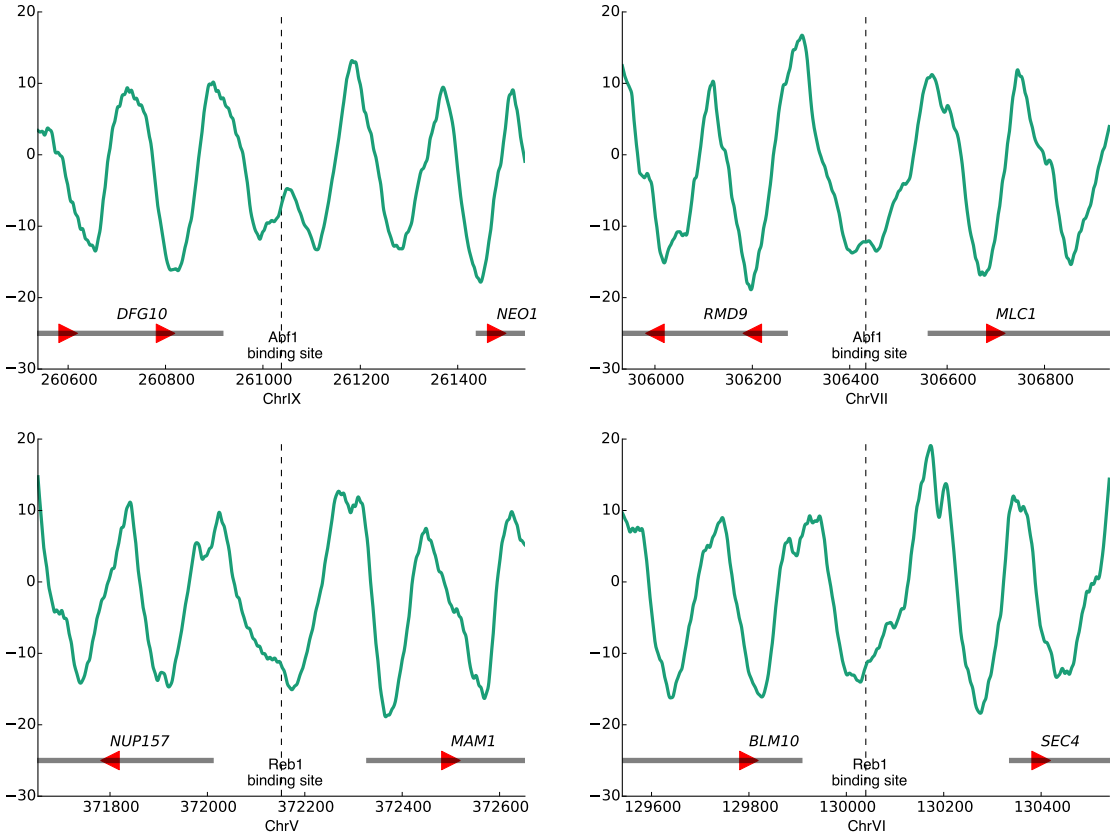
Supplemental Figure 4: This figure reproduces the results shown in Figure 2A, but adds the performance of the individual datasets that form the basis of the MNase consensus map. Using the nucleosome map from Brogaard et al. (2012) as a reference, we calculate the distribution of center-to-center distances between our DNase-seq–based nucleosome map and the Brogaard reference map (blue), and between the consensus MNase-seq–based nucleosome map and the Brogaard reference map (green), as well as between each of MNase-seq–based nucleosome maps and the Brogaard reference map (orange, red, black, and purple).



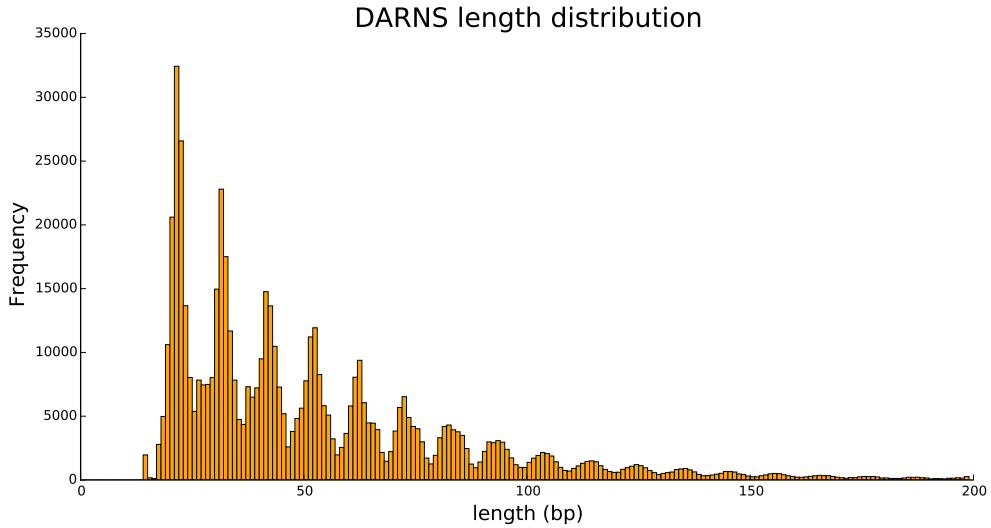
Supplemental Figure 5: This figure reproduces the results shown in Figure 2B, but adds the performance of the individual datasets that form the basis of the MNase consensus map. Using the nucleosome map from Brogaard et al. (2012) as a reference, we calculate the number of true positives (TP), false negatives (FN), and false positives (FP) of our DNase-seq–based map, the consensus MNase-seq–based map, and each of the four individual MNase-seq–based maps. The consensus MNase-seq–based map is compiled by Jiang and Pugh (2009) based on the four individual MNase-seq–based maps. Set 1 is from Mavrich et al. (2008). Set 2 is from Field et al. (2008). Set 3 is from Jiang and Pugh (2009). Set 4 is from Shivaswamy et al. (2008).



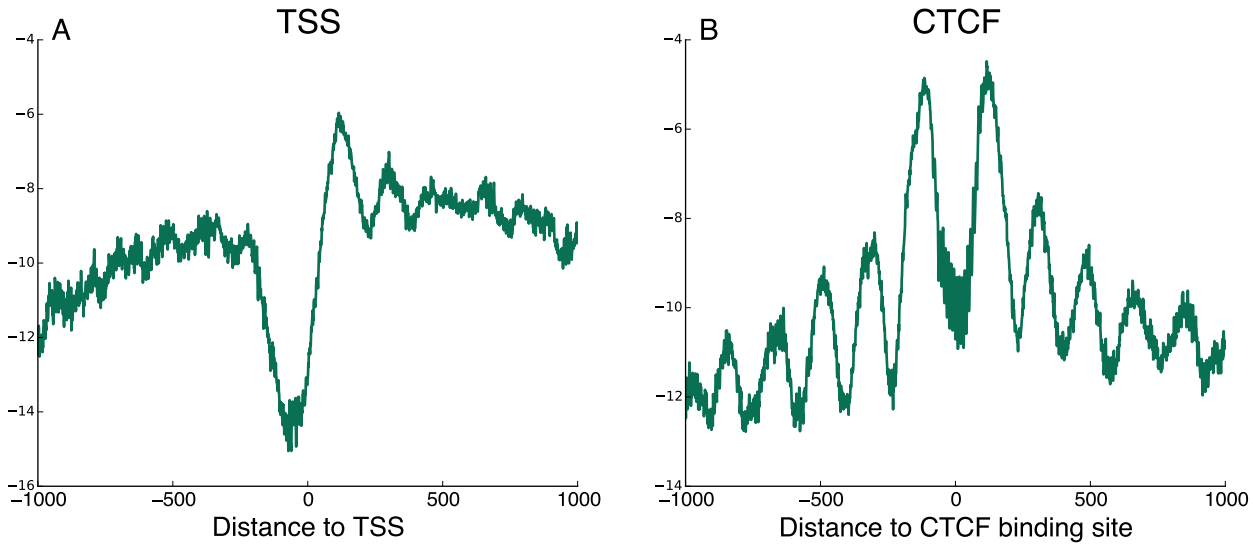
Supplemental Figure 6: Additional example regions showing nucleosome scores around TSSs (top panels) and ACSs (bottom panels). The composite score pattern shown in Figure 3 is also observed at individual genomic loci.



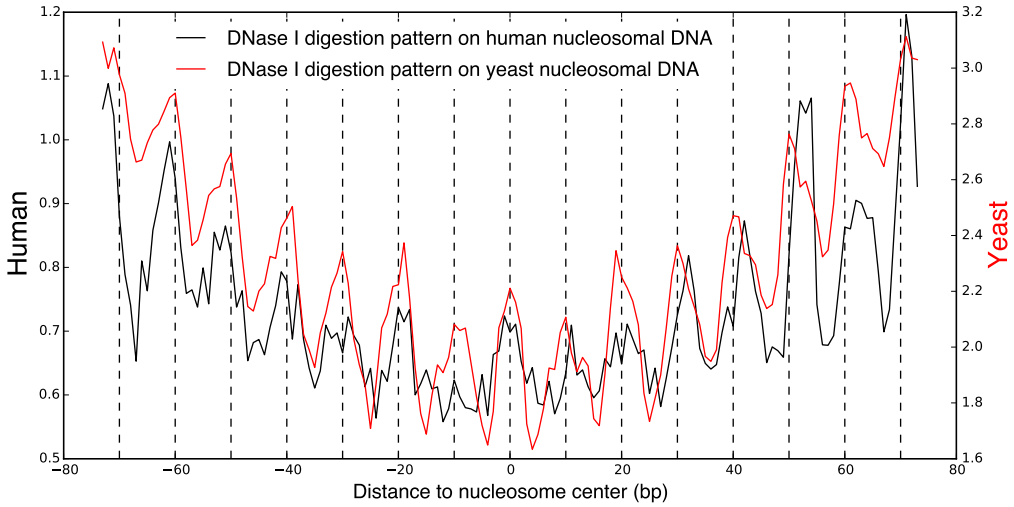
Supplemental Figure 7: Additional example regions showing nucleosome scores around Abf1 (top panels) and Reb1 (bottom panels) binding sites. The composite score pattern shown in Figure 3 is also observed at individual genomic loci.



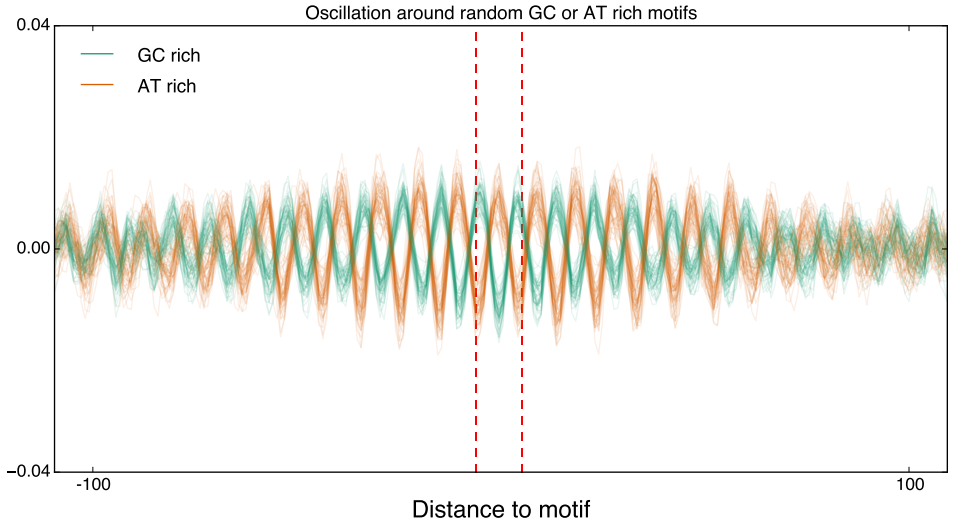
Supplemental Figure 8: The distribution of DARNS lengths, taken from Winter et al. (2013). Note that the vast majority of DARNS are shorter than 100 bp, and the mode is only around 20 bp.



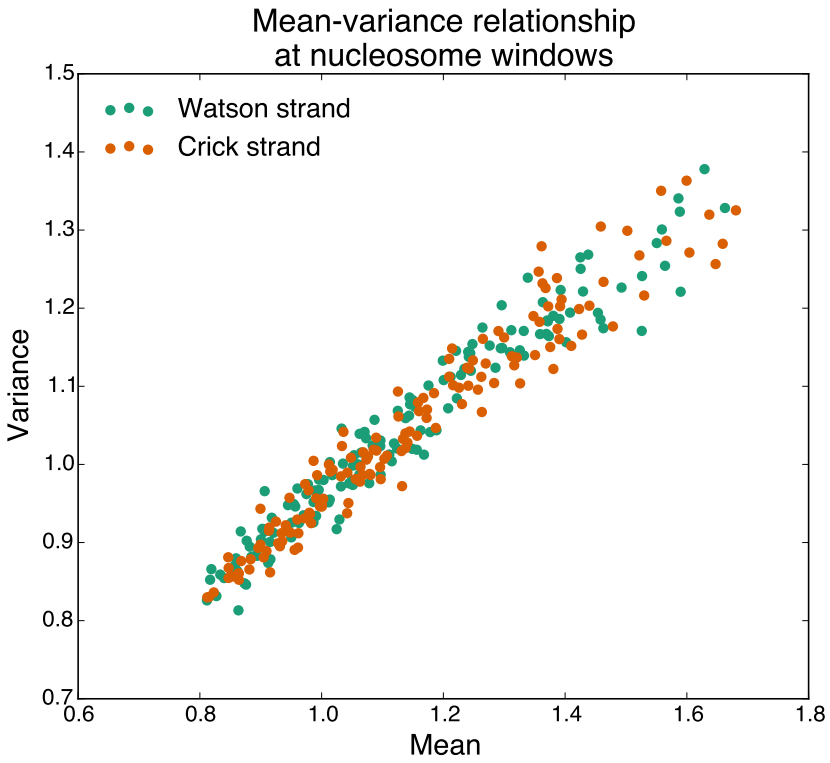
Supplemental Figure 9: Median nucleosome scores around human TSSs (A) and CTCF binding sites (B).



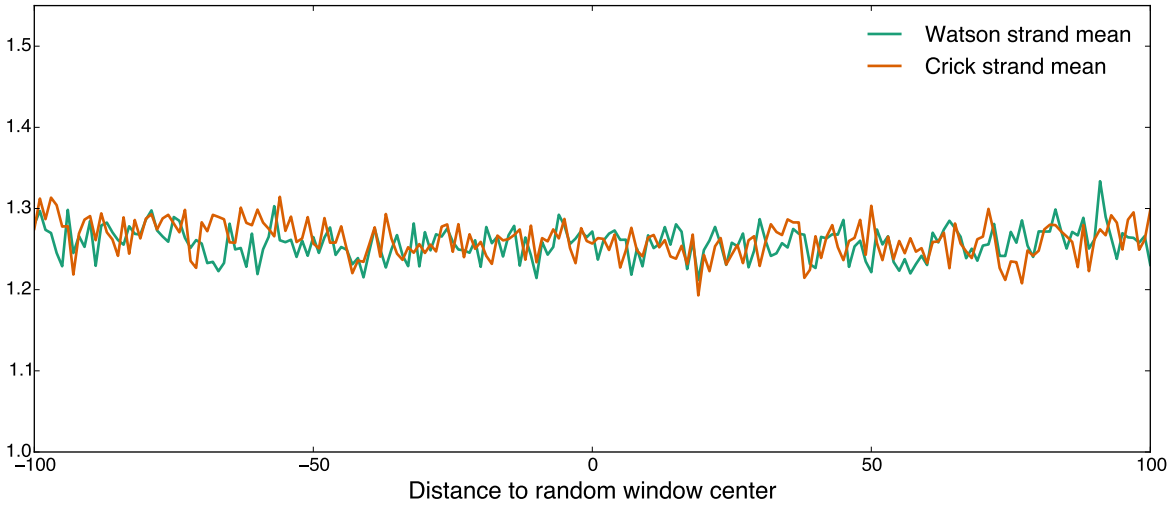
Supplemental Figure 10: DNase I digestion profiles on human nucleosomal DNA (black) and yeast nucleosomal DNA (red). The digestion profiles shown in the figure are an average of the forward and reverse oscillation profiles.



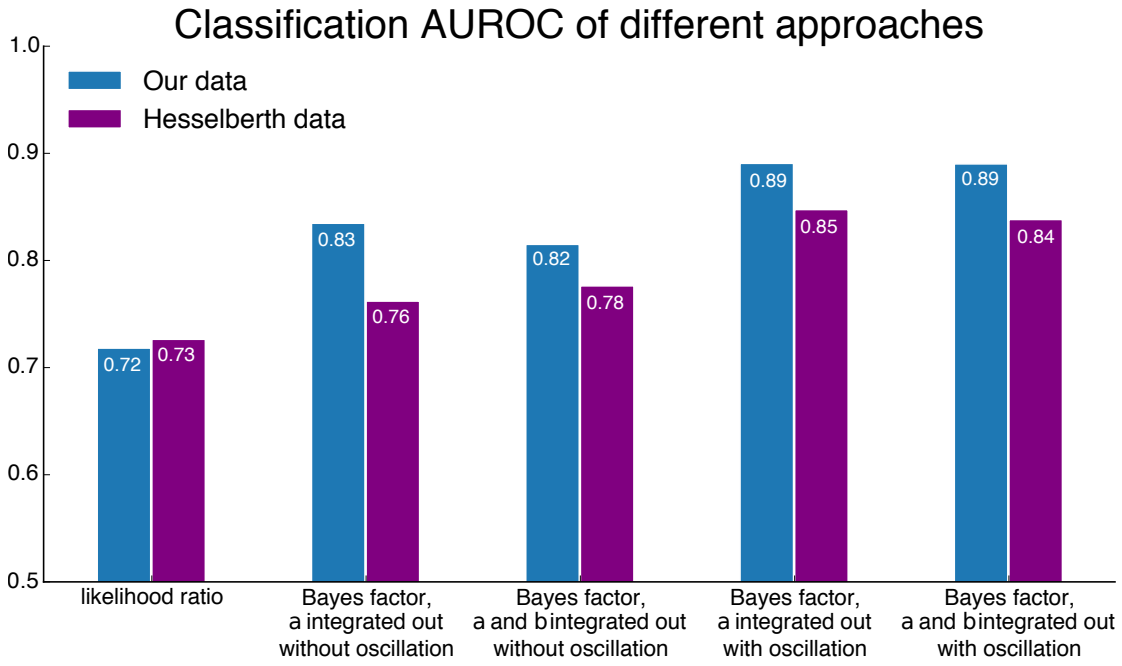
Supplemental Figure 11: We randomly generated five GC (or AT) rich motifs, each with a length of 13 bp (the length of the Abf1 motif). We pooled together all motif matches for these five motifs, filtering out overlapping matches, and then randomly selected 2,600 motif matches from this pool (2,600 being the number of motif matches for Abf1). We calculated and plotted the average oscillation around these motif matches as one curve above. We then repeated this entire process 50 times, resulting in 50 curves for the GC (green) and 50 for the AT (orange) rich motifs. The anti-correlated curves are consistent with nucleosome sequence preferences for GC and AT nucleotides, which are known to oscillate out of phase with one another.



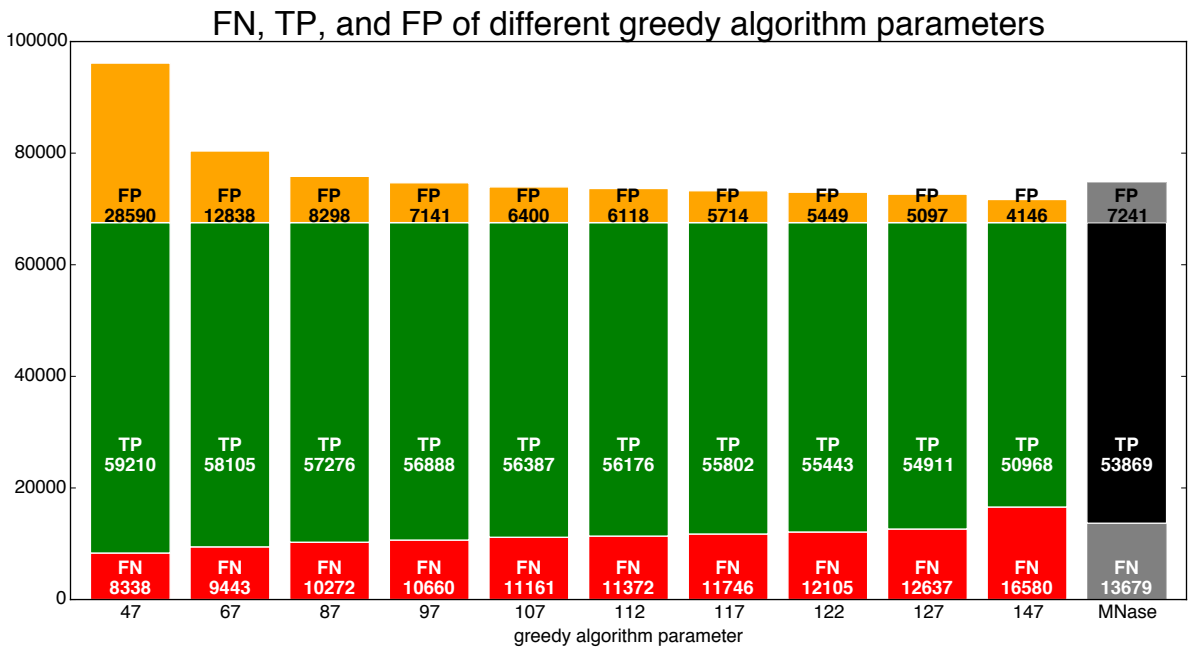
Supplemental Figure 12: Mean and variance of DNase-seq counts at each position along the nucleosome are proportional to each other after inverse hyperbolic sine transformation. Each point corresponds to the mean and variance of one of the 147 positions in the nucleosome window along one of the two strands.



Supplemental Figure 13: DNase I cleavage profiles (average transformed DNase-seq counts) at 2,000 random genomic windows are essentially flat, modulo random sampling noise.



Supplemental Figure 14: Performance (as measured by AUROC) of different classification approaches on both our data and Hesselberth data. Classifying nucleosomal vs. non-nucleosomal locations on the basis of a simple likelihood ratio has the worst performance. Bayes factor approaches that integrate out both the level parameter a and the curvature parameter b perform similarly to approaches that only integrate out a. Approaches that include the oscillation pattern have better performance than approaches without the oscillation pattern. Based on this analysis, we ultimately employed a Bayes factor approach with only a integrated out and with the oscillation pattern (the fourth set of bars).



Supplemental Figure 15: To explore how sensitive our greedy algorithm might be to various settings of the overlap parameter, we tested parameter values different from 117 bp. The same false negative (FN), true positive (TP), and false positive (FP) values as in Figure 2B were calculated and are shown here. The results were largely insensitive for a range of reasonable parameter values between 97 bp and 127 bp. For reference, performance of the consensus MNase-seq–based nucleosome map is reproduced in the last column.

References

Brogaard, K., Xi, L., Wang, J.-P., and Widom, J., 2012. A map of nucleosome positions in yeast at base-pair resolution. *Nature*, **486**(7404):496–501.

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**(11):e1000216.

Henikoff, J., Belsky, J., Krassovsky, K., MacAlpine, D., and Henikoff, S., 2011. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci U S A*, **108**(45):18318–18323.

Jiang, C. and Pugh, B. F., 2009. A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. *Genome Biol*, **10**(10):R109.

Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., Schuster, S. C., Albert, I., and Pugh, B. F., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res*, **18**(7):1073–1083.

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**(3):e65.

Song, L. and Crawford, G. E., 2010. DNase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protocols*, **2010**(2):pdb.prot5384.

Winter, D. R., Song, L., Mukherjee, S., Furey, T. S., and Crawford, G. E., 2013. DNase-seq predicts regions of rotational nucleosome stability across diverse human cell types. *Genome Res*, **23**(7):1118–1129.