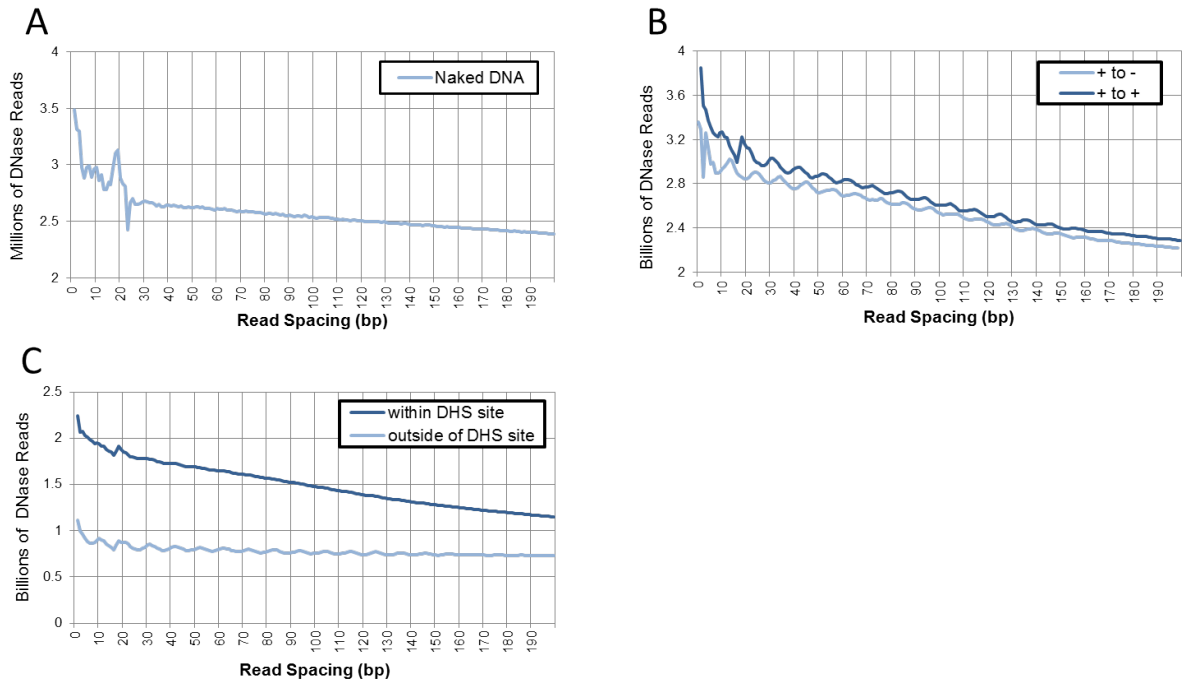
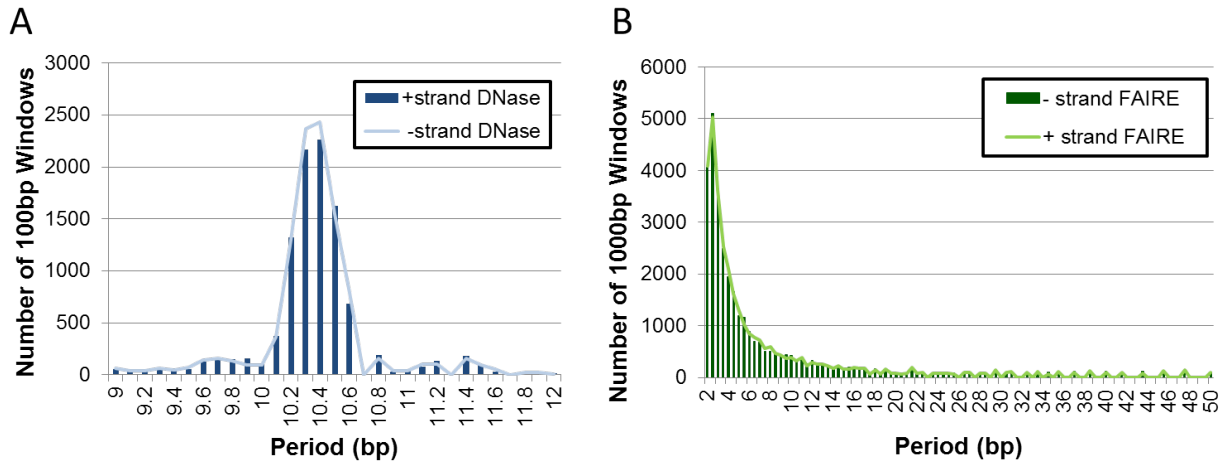
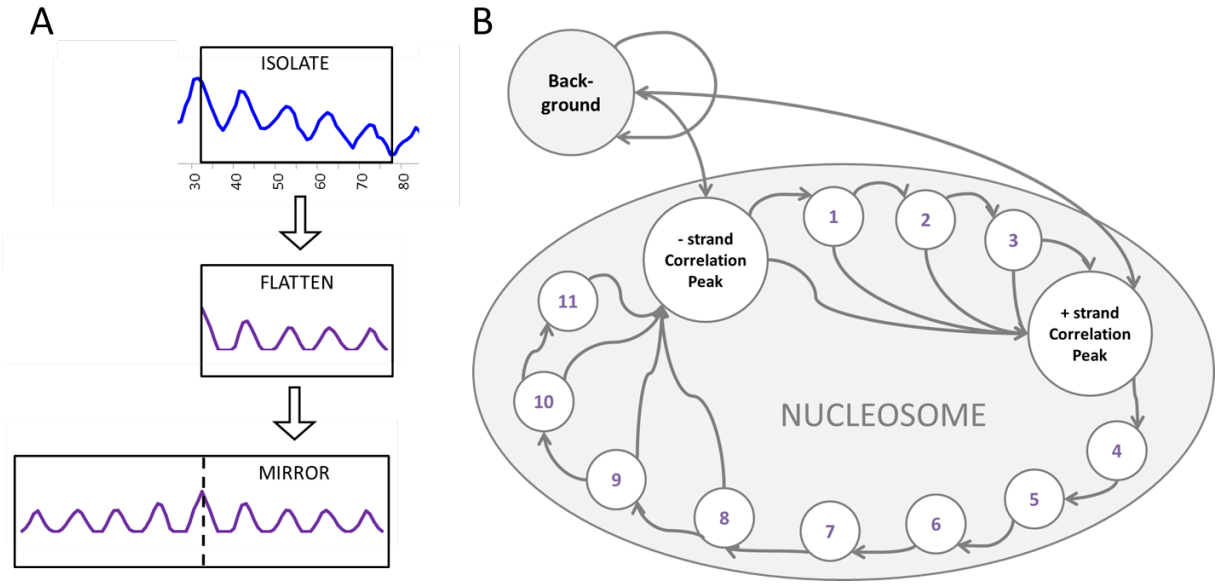
SUPPLEMENTAL FIGURES AND TABLES



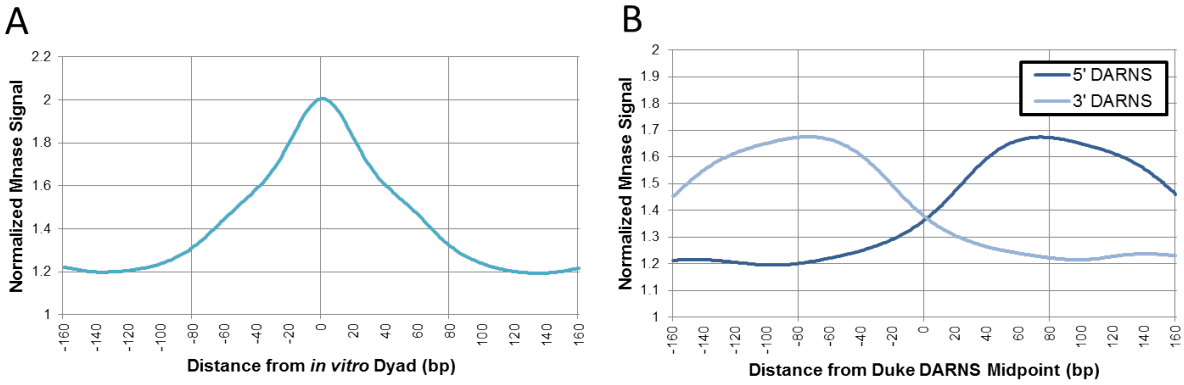
Supplemental Figure S1. Additional oscillation plots from combined DNase-seq data. (A) Oscillation plot showing that the ~10bp period is not evident in DNase-seq data from naked DNA (B) Oscillation plot showing ~10bp period in positive strand DNase read spacing and ~10bp period offset by ~3bp when positive and negative strands reads are compared (C) Oscillation plot showing 10bp period in DNase reads outside of DHS sites but no oscillation in DNase reads overlapping DHS sites identified from each cell-line



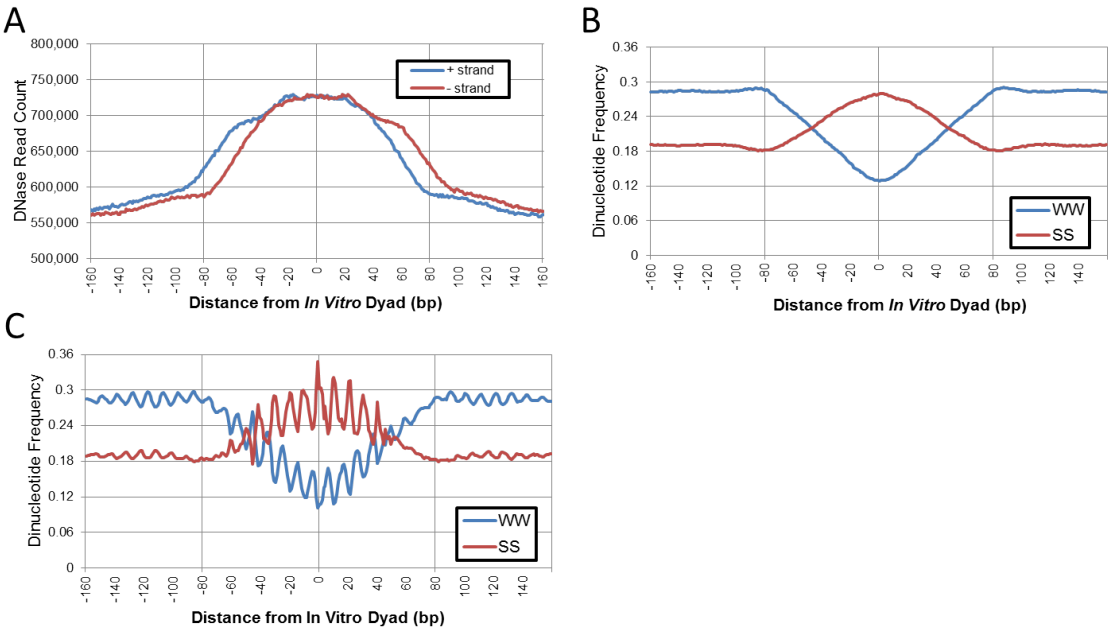
Supplemental Figure S2. Additional Figures for Fourier Analysis. (A) Zoom-in of Figure 2B showing Fourier Analysis on DNase data and indicating average period length between 10.3 and 10.4bp (B) Fourier analysis on chromosome 1 of FAIRE data does not reveal a meaningful period.



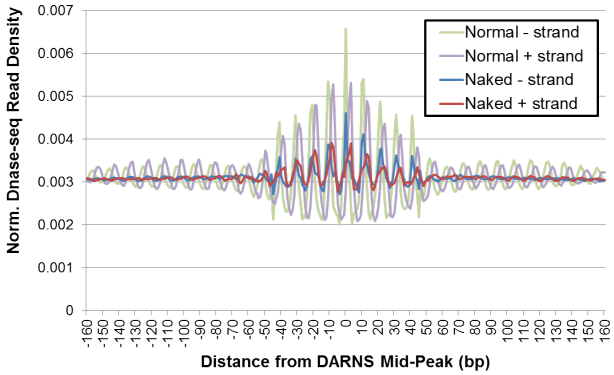
Supplemental Figure S3. Designing the HMM to identify DARNS. (A) Expected Pattern of DNase Digestion around the Nucleosome (91 bases in length) derived from DNase oscillation plot (Fig. 1B) (B) Diagram of states in DARNS HMM. Numbered states represent where non-peak emissions (0) are expected.



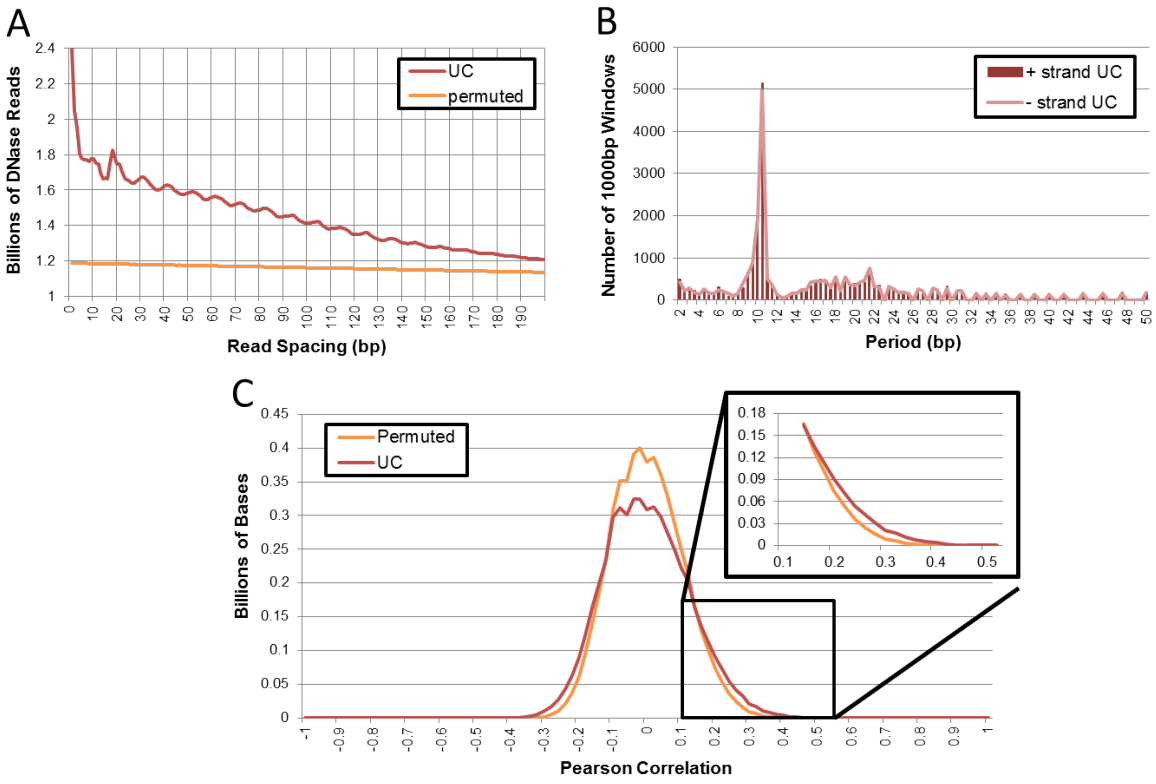
Supplemental Figure S4. MNase signal density around *in vitro* dyads and DARNS. (A) Distribution of Lymphoblast MNase-seq signal around *in vitro* dyads (Valouev et al., 2011) shows corresponding levels of enrichment (B) Distribution of Lymphoblast MNase-seq signal around DARNS designated as the 5' end or 3' end of the nucleosome relative to in vitro dyad. Note that each subset shows MNase-seq signal enrichment on the corresponding side of center



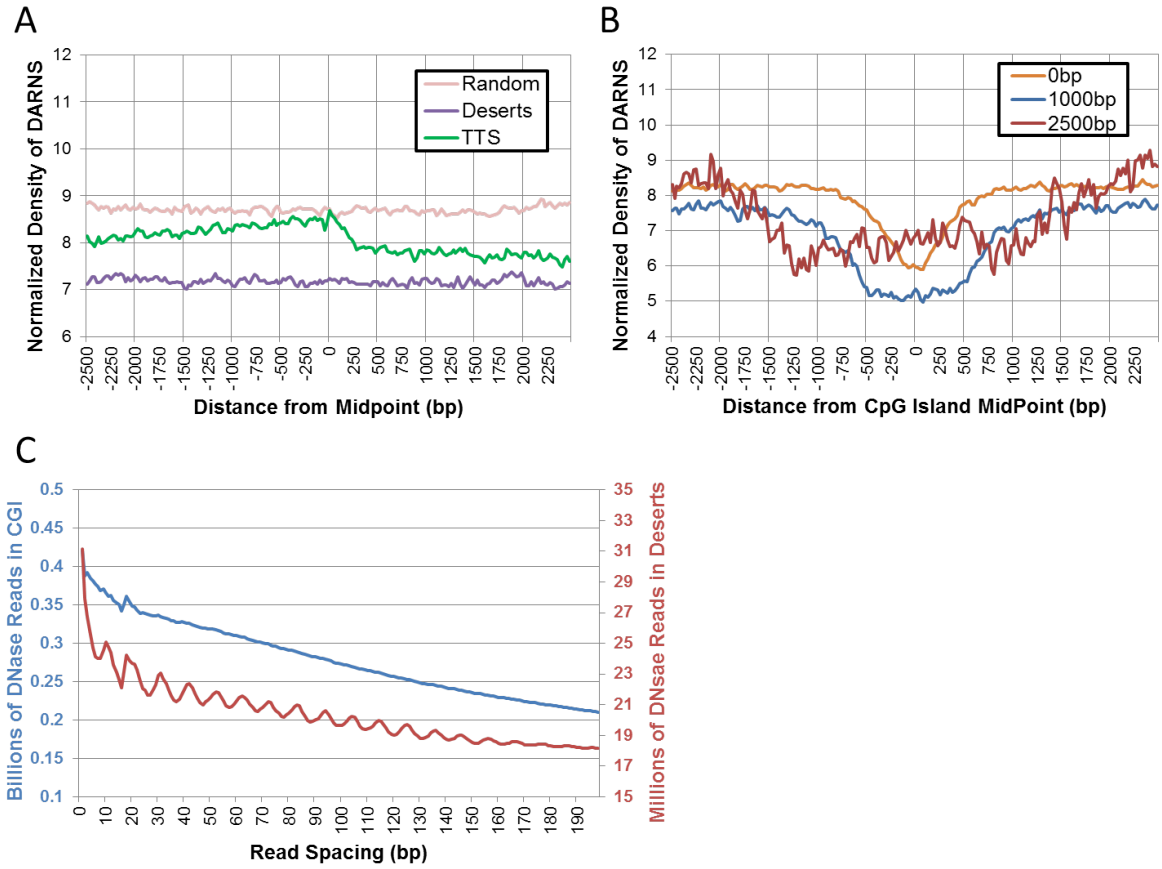
Supplemental Figure S5 - **Properties of *in vitro* dyads.** (A) Distribution of DNase Read Count around in vitro dyads shows crest at center. (B) Distribution of dinucleotide frequencies (W=A/T; S=C/G) (C) same as B) but aligned by nearest negative strand correlation peak



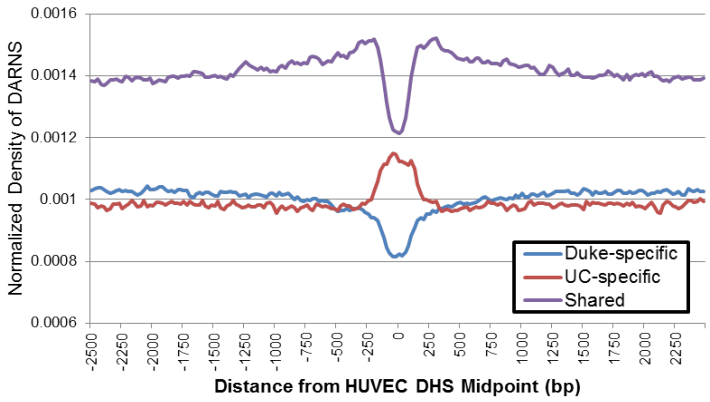
Supplemental Figure S6 – **Normalized DNase-seq reads from naked DNA around DARNS.** Some bias is expected in DNase-seq reads from naked DNA because of the preference of DNaseI to digest at C/G dinucleotides. Since these dinucleotides are periodic in nucleosomes – particularly those covered by DARNS (Figure 5C,D) – we expect this bias to lead to a periodic pattern of digestion. However, the intensity is reduced in DNaseI digested naked genomic DNA derived from K562 when compared with DNaseI digested nuclear K562 DNA using the standard DNase-seq protocol. Read density is normalized by the total number of reads in the window.



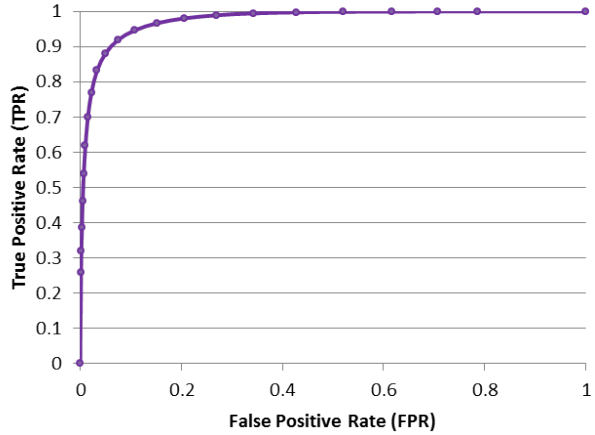
Supplemental Figure S7. University of Chicago (UC) DNase-seq data display similar features as Duke DNase-seq data. (A) Oscillation Plot for UC data shows 10bp period but not in permuted. (B) Fourier Analysis on chromosome 1 of UC reveals a dominant ~10.5bp period with a harmonic shadow at ~21bp. (C) Distribution of correlations in UC has more high scores than permuted data.



Supplemental Figure S8. DARNS around Additional Genomic Features. (A) Normalized Density of DARNS around transcription termination sites (TTS), gene deserts, and random intergenic sites (B) Normalized Density of DARNS around all CpG islands (0), and those that are greater than 1000bp or 2500bp in size. (C) Oscillation plot of DNase-seq reads in gene deserts (Deserts=red) exhibit the ~10bp period associated with nucleosomes but not in CpG islands (CGI=blue).



Supplemental Figure S9. DARNS surrounding HUVEC DHS sites. Normalized distribution of Duke-specific, UC-specific, and shared DARNS around DHS sites identified from a single cell-line (HUVEC). Note the depletion of Duke-specific and shared DARNS but enrichment in UC-specific DARNS.



Supplemental Figure S10. Accuracy of Regression Classifier
The ROC curve for the multivariate linear regression classifier, AUC = 0.975.

Use	Definition	Legend	Text
Figure 6ABCD	Alignable Intergenic Sites	Random	Random Intergenic Sites
Figure 6CD	Intersection 49	Ubiquitous DHS	DHS in every cell type
Figure 6EFG	Union 49 – Union 42	LCL-specific	DHS in LCLs but not other cell types

Figure 1C, 6EFG	Union 49 – Union LCL	Non-LCL	DHS in non-LCL
Figure 1C	Intersection LCL	Ubiquitous LCL	DHS in every LCL

Supplemental Table S1. Definitions of cell type DHS used in Figures 1 & 6.

SUPPLEMENTAL MATERIAL

Regression Classifier. For each of the following 16 features we calculated the Kolmogorov-Smirnov p-values indicating the chance that the values for the "true" DARNS (from DNase-seq data) came from the same distribution as the "random" DARNS (from permuted data): length (bp), number of cycles through HMM "nucleosome" states (0.5 for each peak state passed), sum of reads within the DARNS, read density, value of maximum positive strand correlation peak, value of maximum negative strand correlation peak, mean of positive strand correlation peak values, mean of negative strand correlation peak values, sum of positive strand correlation peak values, sum of negative strand correlation peak values, spacing to nearest upstream DARNS (bp), spacing to nearest downstream DARNS (bp), the HMM posterior probability of the path through the background state, and the HMM posterior probability of the path through the nucleosome set of states. These features were input for a multivariate linear regression analysis used as a classifier to distinguish between true (1) and random DARNS (0). We divided DARNS into 10 equal sized sets and trained the classifier on one tenth of the data, then predicted the class for the remainder. We repeated this 10 times and averaged the nine test scores to determine a confidence value for each DARNS. Then, we plotted the ROC curve (True Positive Rate vs. False Positive Rate) for the performance of the classifier and calculated the Area under the Curve (AUC) (Supplemental Fig. S10). Using the results of the classifier, we estimated the false discovery rate (FDR) from the percentage of DARNS that scored below 0.5.