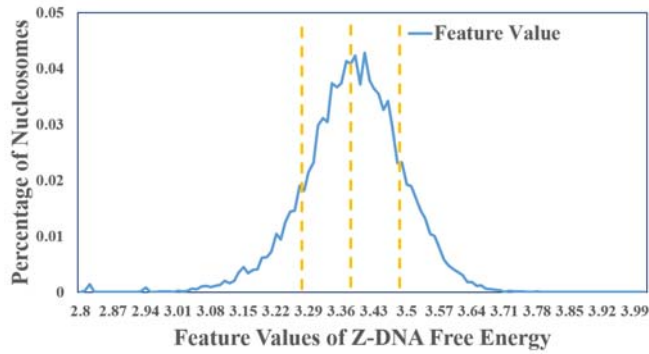# Supplementary Figure 1
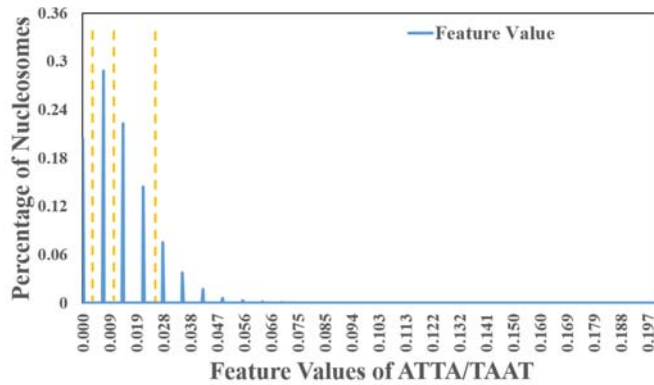
A)

B)

Feature values are discretized based on their empirical distribution. (A) Empirical distributions of values for feature "Z-DNA free energy" can be fit with Gaussian model, the dotted yellow lines show the three cutoffs using $\mu - \sigma$, μ and $\mu + \sigma$. (B) The empirical distribution of feature values for "ATTA/TAAT" is not likely to be modelled with Gaussian distribution, thus (mean - min) / 2, mean, (max - mean) / 2 (dotted yellow lines) are used as the cutoffs to discretize them into 4 levels.
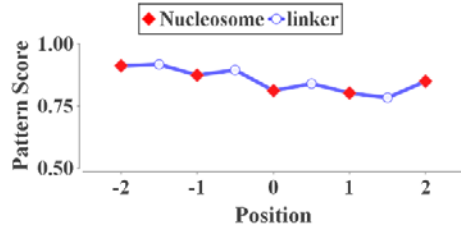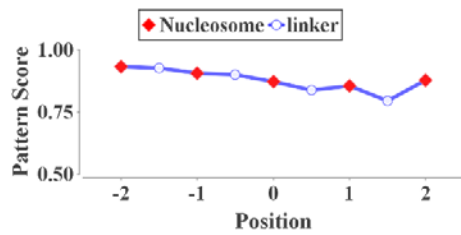
Supplementary Figure 2

A)

B)
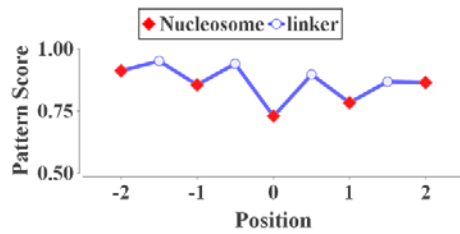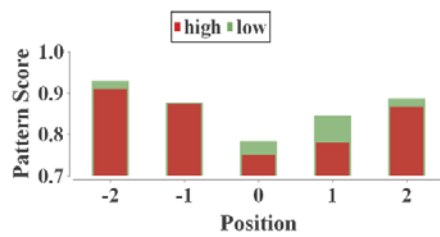
Pattern distribution in activated human T cells data is location and transcription-dependent. (A) Average pattern scores for nucleosomes and linkers around the TSS. (B) Average pattern scores for nucleosomes in 1000 most highly expressed genes and 1000 lowest expressed genes.
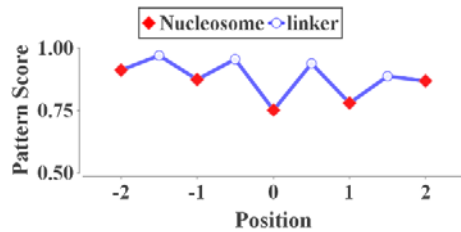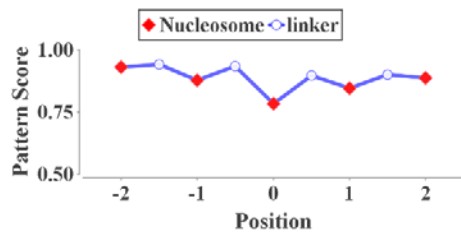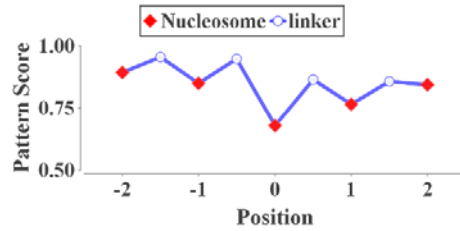
Supplementary Figure 3

A)

B)

C)

D)

Pattern distribution in human T cells data is transcription-dependent. (A) Average pattern score for nucleosomes and linkers in 1000 most highly expressed genes in human resting status. (B) Average pattern score for nucleosomes and linkers in 1000 lowest expressed genes in human resting status. (C) Average pattern score for nucleosomes and linkers in 1000 most highly expressed genes in human activated status. (D) Average pattern score for nucleosomes and linkers in 1000 lowest expressed genes in human activated status.

Supplementary Figure 4

A)

B)
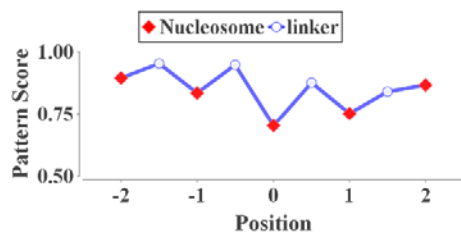
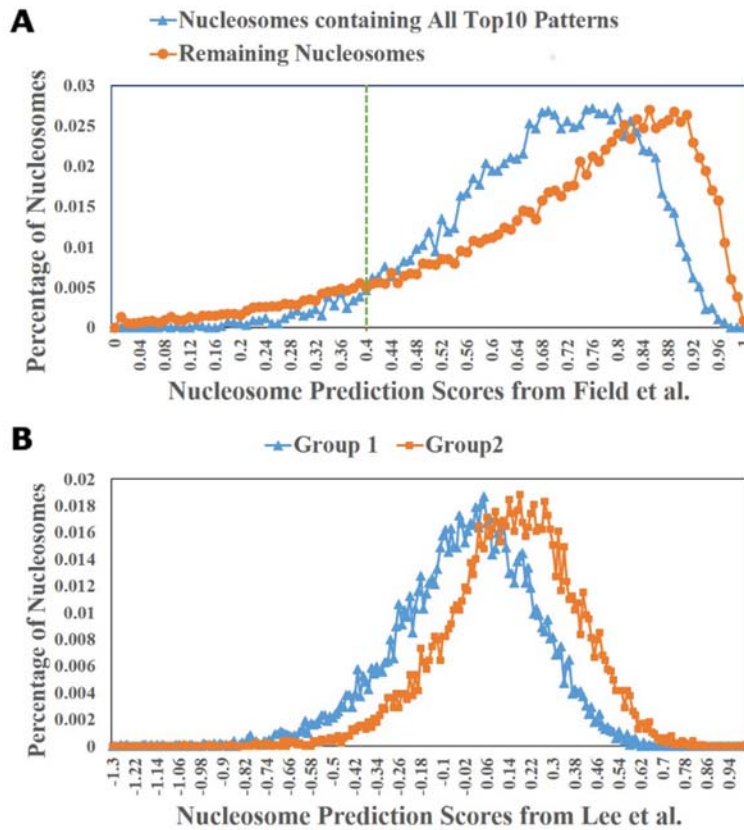Pattern distribution of perturbed genes is location and transcription-dependent in activated human T cells data. (A) Average pattern score for nucleosomes and linkers in repressed genes in activated status. (B) Average pattern score for nucleosomes and linkers in induced genes in activated status.

Supplementary Figure 5

Comparisons of the Field's and Lee's score distributions between the yeast nucleosomes containing the top feature patterns and the remaining yeast nucleosomes (A) Comparison based on Field's score distribution. Nucleosomes containing all top 10 features contain less low scores predicted by Field's model than the other nucleosomes. ~12.7% of the 6438 nucleosomes with low Field scores (<0.5, dotted yellow line) will not be predicted well using periodical dinucleotide feature, while can still be predicted considering the feature patterns enriched. (B) Comparison based on the Lee's score distribution. Nucleosomes containing pattern "Z-DNA free energy" level 2, "slide" level 1, "CCGCC/GGCGG" level 0 (Group 2) have averagely higher regression scores than nucleosomes containing pattern "Z-DNA free energy " level 1, "slide" level 2, "CCGCC/GGCGG" level 0 (Group 1).