Supplementary mateirals

Section 1: agreement with promoter nucleosomes determined by Ozsolak et al.

A recent work by Ozsolak *et al*[1] has also profiled DNA protection patterns established by MNase digestion, and identified putative fixed nucleosome positions across promoter regions in a set of seven human cell lines. This set did not include K562 or HeLa cells. Depending on the cell line, 38 to 58 nucleosome positions predicted by Ozsolak *et al.* fall within the HOX clusters examined in this work. Our predictions of nucleosome positions in K562 cells most closely match MP cell line measured by Ozsolak *et al.*, agreeing on 32 of 52 positions. Similarly, nucleosome calls in HeLa cells match 21 of 38 nucleosomes predicted for the MALME cell line. This agreement is consistent given 70% concordance estimated by Oszolak *et al*.

Section 2: agreement with high-throughput sequencing data by Schones et al.

Mononucleosome fragments from human CD4+ T-cells were recently analyzed using high-throughput sequencing by Schones *et al* [2]. We have examined the average patterns of sequence tag occurrence around fixed nucleosome positions identified in our work (Figure 9), and find that positions of the central and adjacent nucleosomes are associated with increased tag density. Examination of different HOX regions reveals that prominent peaks of sequenced tag density in most cases coincide with the predicted fixed nucleosome positions (Figure 10). However the majority of fixed nucleosome positions identified by our analysis are found in regions where the number of sequence tags is not sufficient to distinguish nucleosome positions (Figure 11).

Section 3: Localized differences between cell lines

The differences between predicted fixed nucleosome positions incorporate uncertainty of the HMM algorithm in distinguishing nucleosome peaks. To focus on the regions of significant difference between the two cell lines we calculated a statistical difference score between chromatin profiles which takes into account variability within and between replicate measurements.

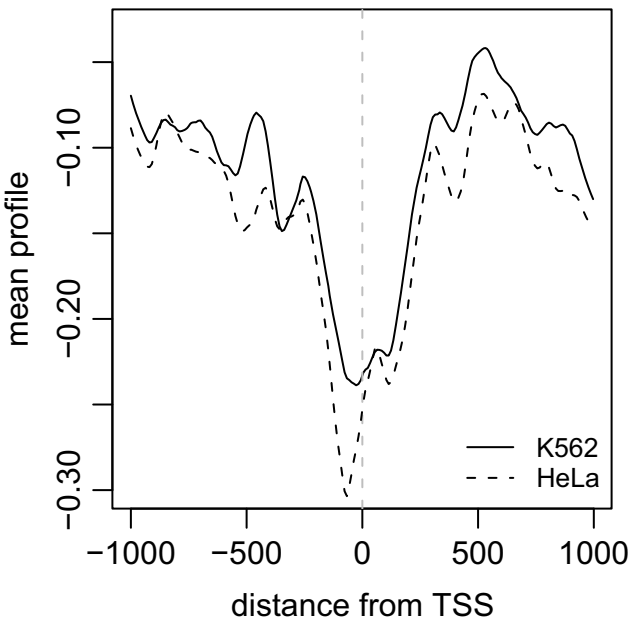
The difference score was calculated using 20-probe (100bp) running window average of the small window P values (SWP): two-sided t-test P values calculated for each position based on the comparison of all replicate data from the two cell lines for the three-probe group around a given position. The regions of significant difference were determined as clusters of probes with the difference scores in the 90% quantile and a minimum size of 60bp.

As expected, the bodies of genes with different expression states on average exhibit significantly more changes in the chromatin structure than genes whose transcriptional states remain similar (P value 1.1e-3). A large fraction (12%) of significant differences regions falls into MHS. In 24% of these cases the change is observed even though MHS is present in both cell lines. The concordance of the predicted fixed nucleosome positions between the two cell lines (greater than 70%) indicates that nucleosome positioning is relatively stable given the expected error rate of nucleosome predictions (20%). Even among the set of significant difference regions 33% show the same fixed nucleosome positioning pattern. The majority of the remaining cases involve individual nucleosome differences, and at a few regions a shift in the periodic nucleosome pattern can be observed between the two cell lines (see Figure 8).

[1] ***F. Ozsolak, J. S. Song, X. S. Liu et al., Nat Biotechnol 25 (2), 244 (2007).***

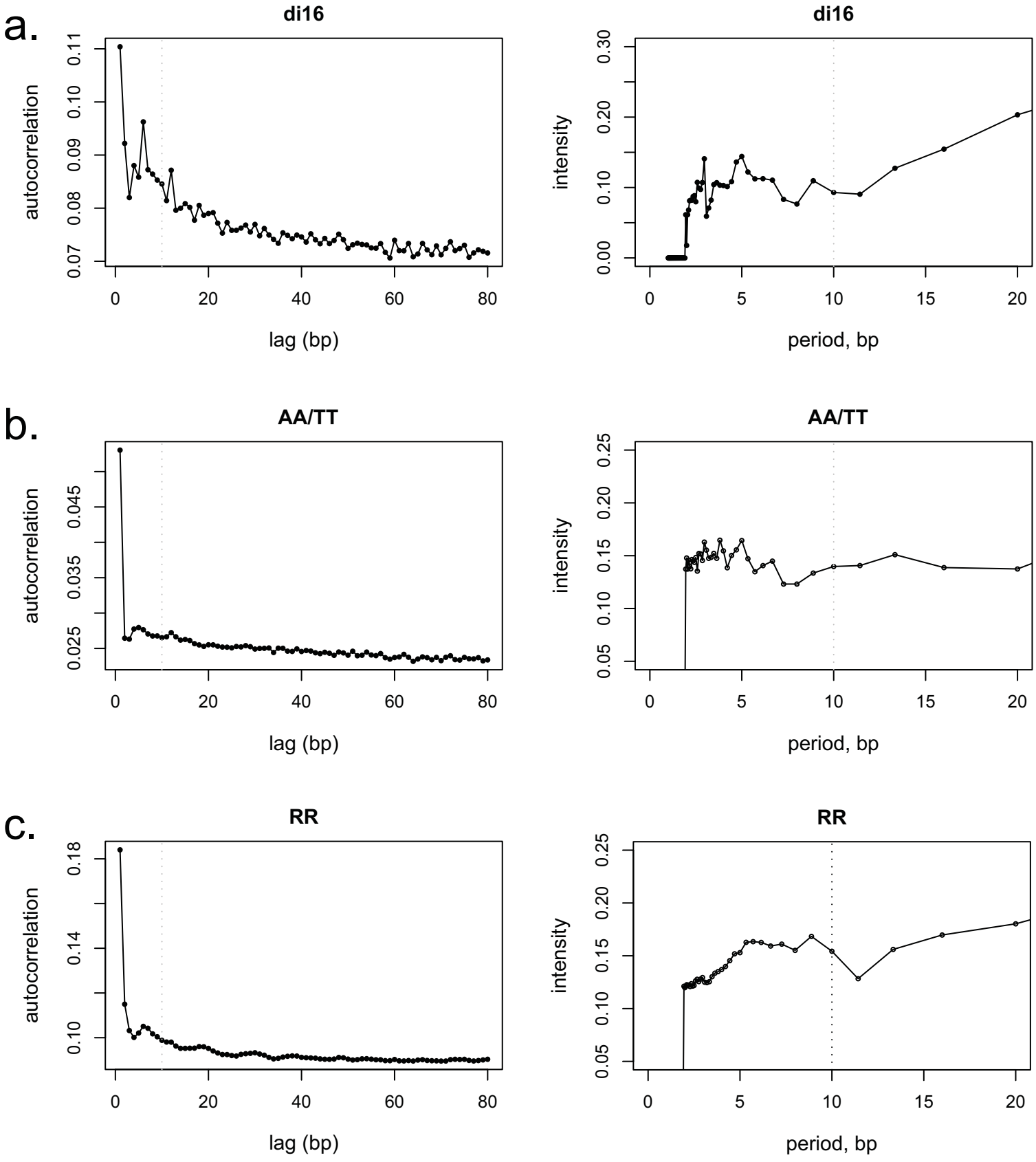
[2] ***D. E. Schones, K. Cui, S. Cuddapah et al., Cell 132 (5), 887 (2008).***

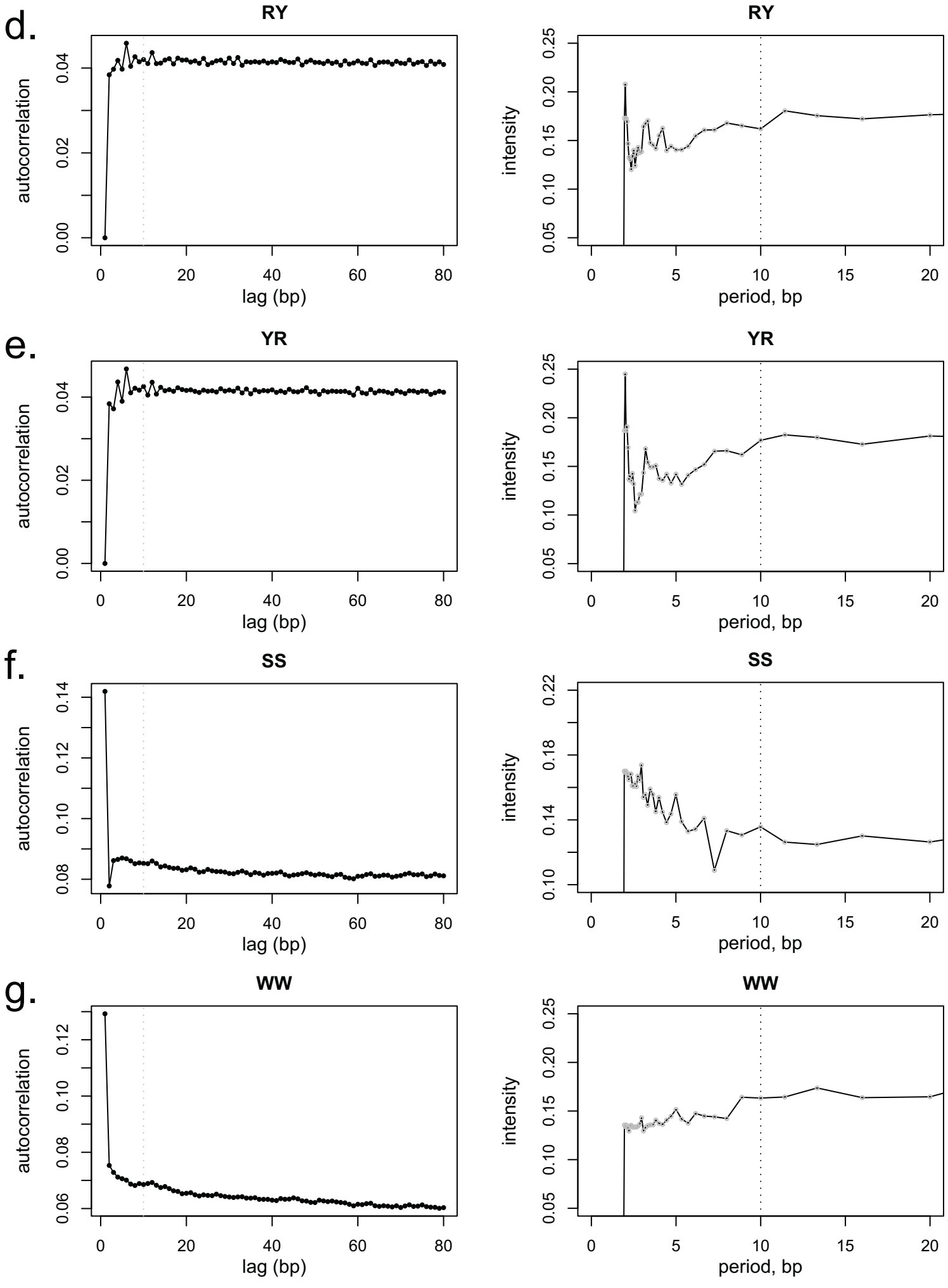
Figure 1

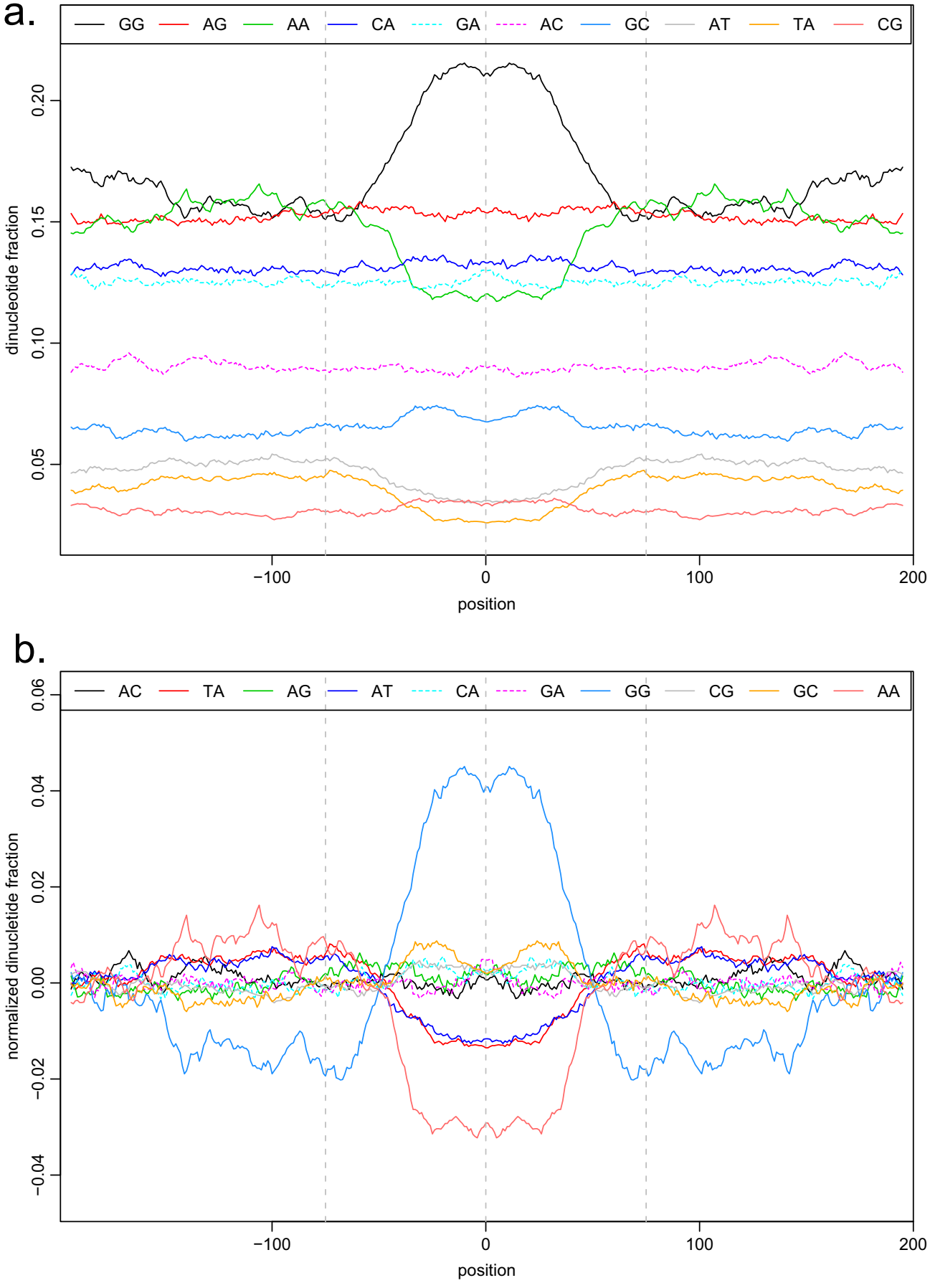


Micrococcal nuclease hypersensitive sites (MHS) are observed near transcription start sites. Average preextraced chromatin profiles around transcription start sites (TSS) are shown for HeLa and K562 cells. The 0 position corresponds to the TSS, with all genes oriented to transcribe from left to right.

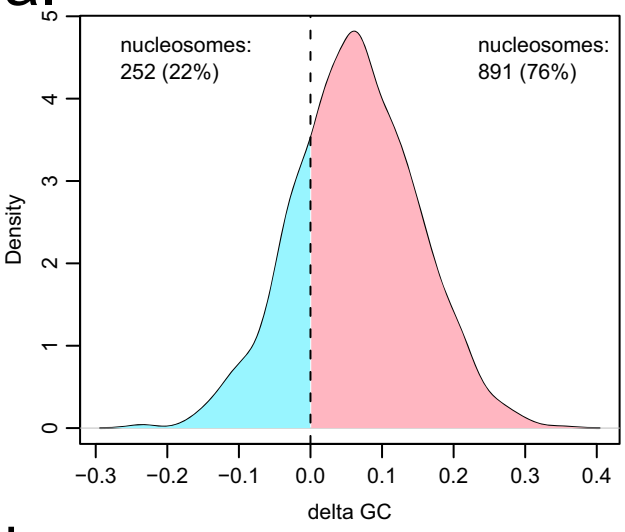
Dinucleotide periodicity near the predicted fixed nucleosome positions. Autocorrelation function (left) and corresponding frequency power spectrum (right) is shown for the following dinucleotides: **a.** combined autocorrelation for all 16 possible dinucleotides. **b-g**. Autocorrelations for individual dinucleotides. R denotes purines, Y - pyrimidines, W - weak (A,T), S - strong (C,G). Dotted vertical lines indicate position of expected 10bp periodicity.



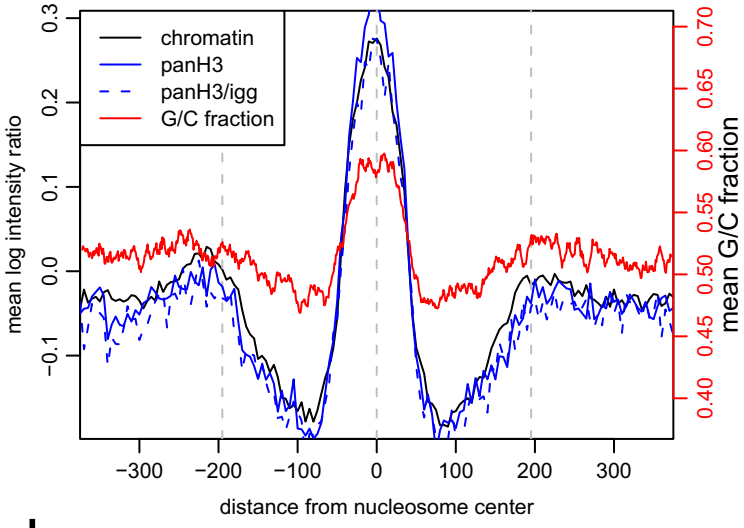

d.
RY
autocorrelation
0.00
0.02
0.04
0
20
40
60
80
lag (bp)
RY
intensity
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
period, bp
e.
YR
autocorrelation
0.00
0.02
0.04
0
20
40
60
80
lag (bp)
YR
intensity
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
period, bp
f.
SS
autocorrelation
0.08
0.10
0.12
0.14
0
20
40
60
80
lag (bp)
SS
intensity
0.10
0.14
0.18
0.22
0
5
10
15
20
period, bp
g.
WW
autocorrelation
0.06
0.08
0.10
0.12
0
20
40
60
80
lag (bp)
WW
intensity
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
period, bp

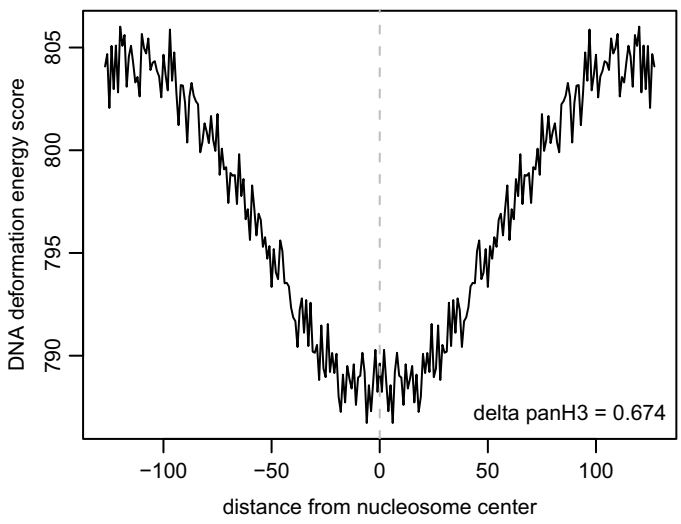


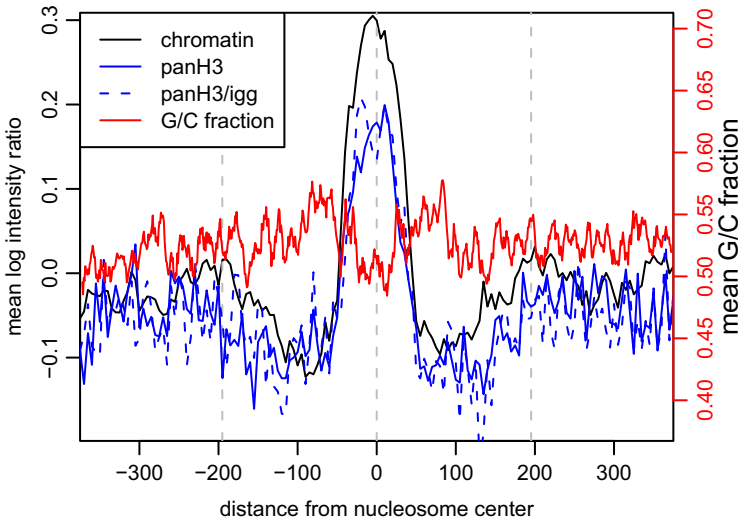
Dinucleotide ratios in predicted fixed nucleosome sequences. **a.** Fraction of each dinucleotide is shown for positions relative to the predicted nucleosome dyad positions. The sequences were symmetrized. **b.** The same fractions are shown relative to the mean fraction value for each dinucleotide.

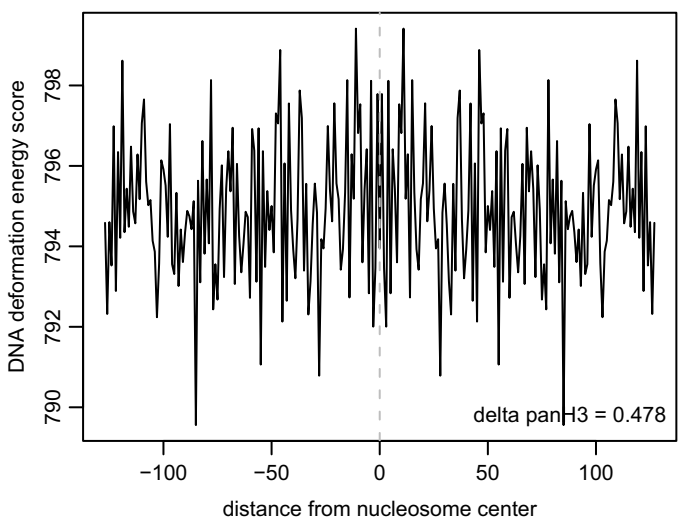


Nucleosomes with increased and decreased core G/C content. Delta GC measures the difference in the G/C content between predicted nucleosome centers and positions 100bp away from the center. **a.** Distribution of delta GC values for predicted fixed nucleosome positions. Number and fraction of nucleosomes with positive (red) and negative (blue) delta GC is given. Subsequent plots show average profiles and energy scores (as in Figure 3 of the main text) for nucleosomes with positive delta GC (**b,c**) and negative delta GC (**d,e**). Delta panH3 (lower right corner of the c. and e. plots) is the average difference in the pan-H3 ChIP/input levels between the predicted nucleosome core and linker regions.

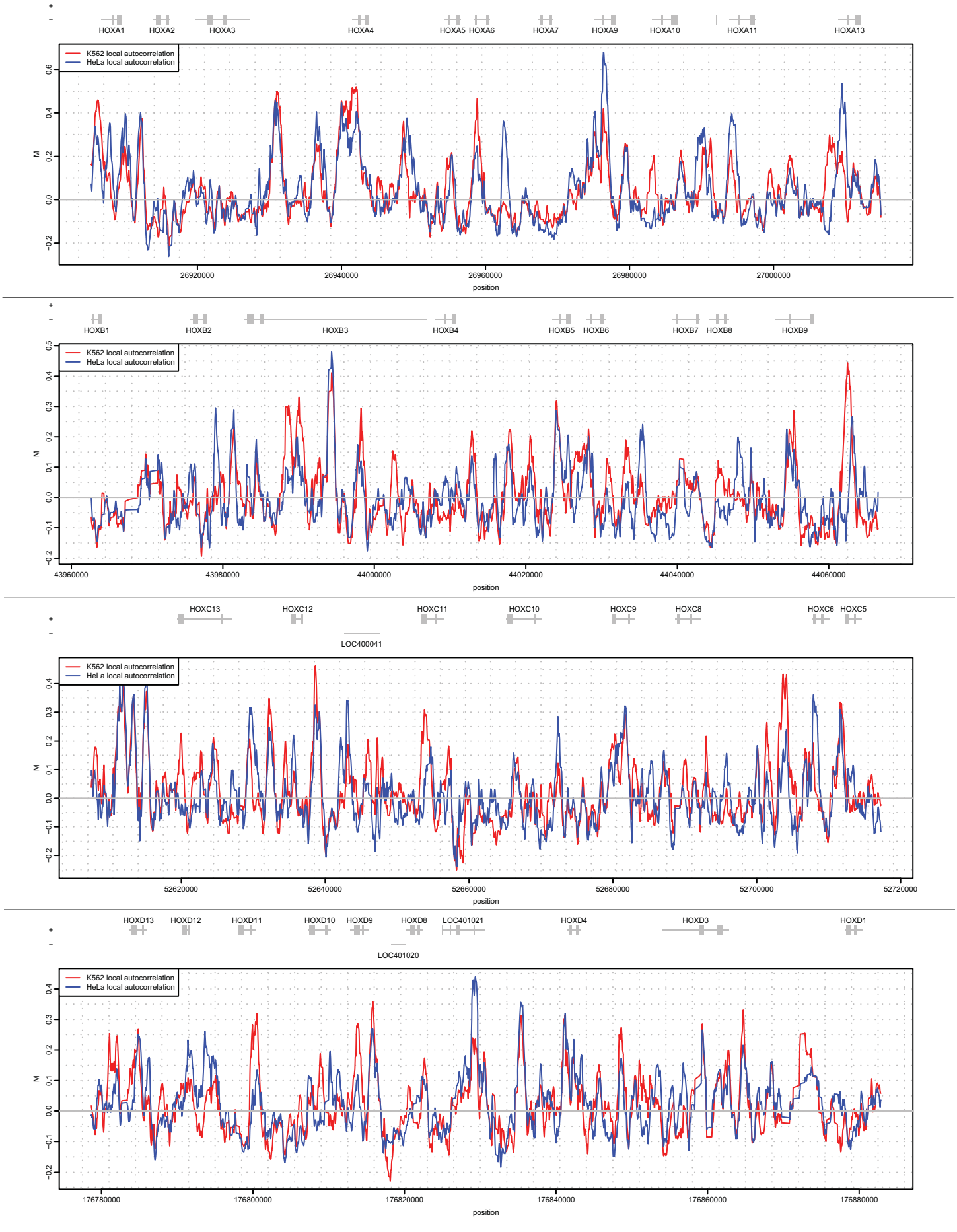








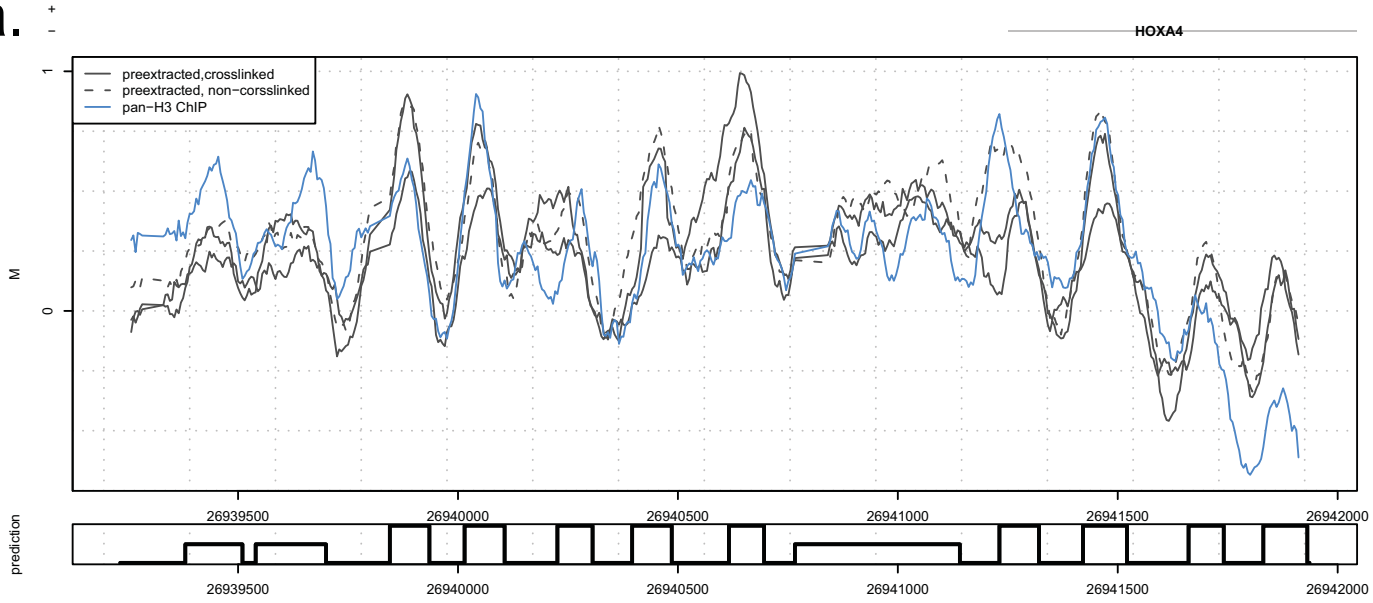
Positions in HOX clusters exhibiting 195bp periodicity. Delta 195bp periodicity scores were calculated using local autocorrelation (see Methods). The plots below show periodicity scores for each of the four HOX clusters. Scores are shown for both K562 (red) and HeLa (blue) chromatin profiles.



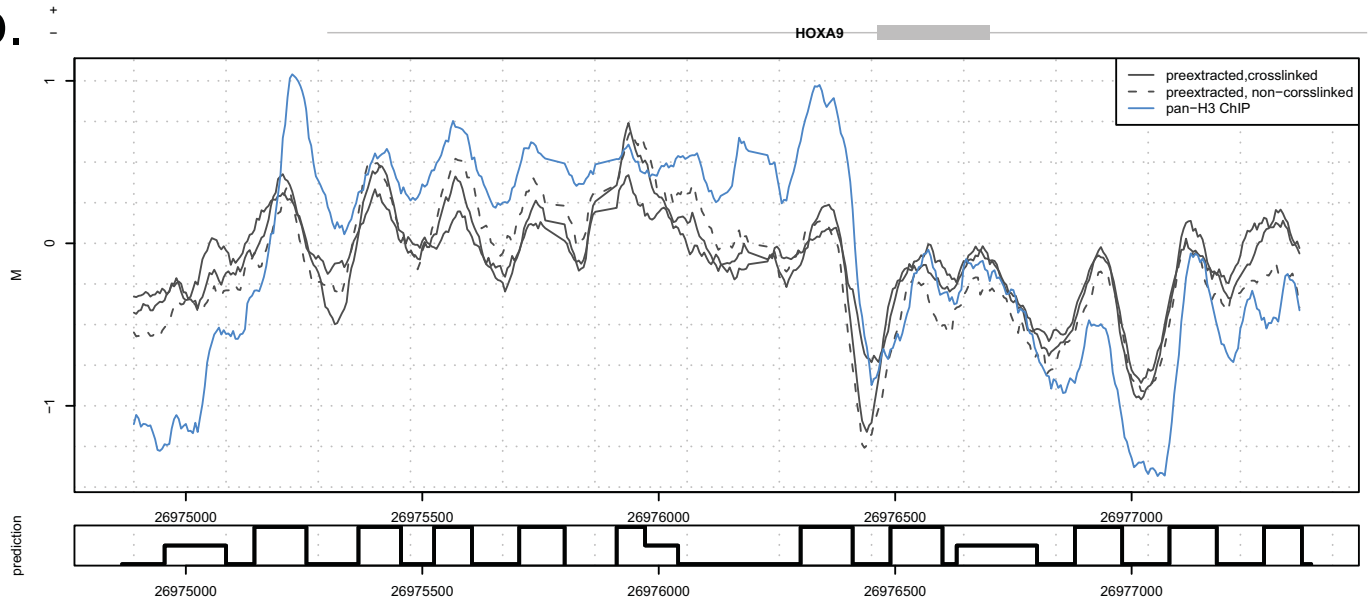
Examples of regular nucleosome arrays. Figures below provide examples of regular nucleosome grids with 195bp periodicity (**a-c**) and 230bp periodicity (**e**). The grid (dotted lines) gives 195bp (or 230bp) nucleosome spacing.

Figure 6

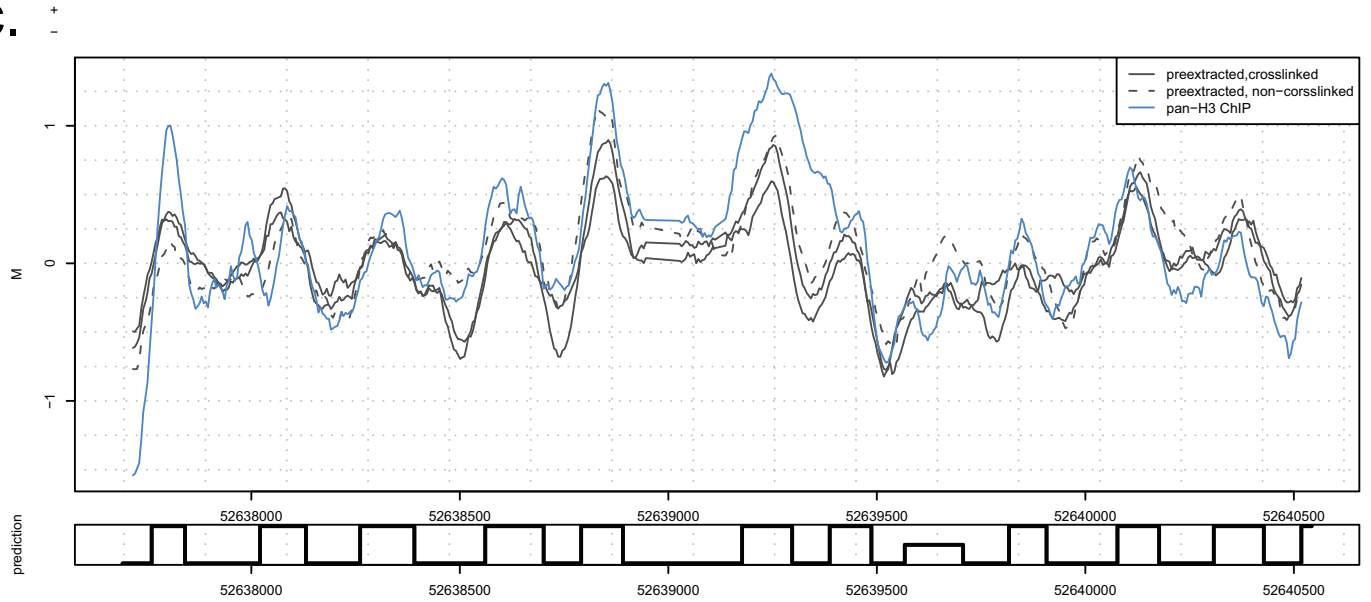
a.



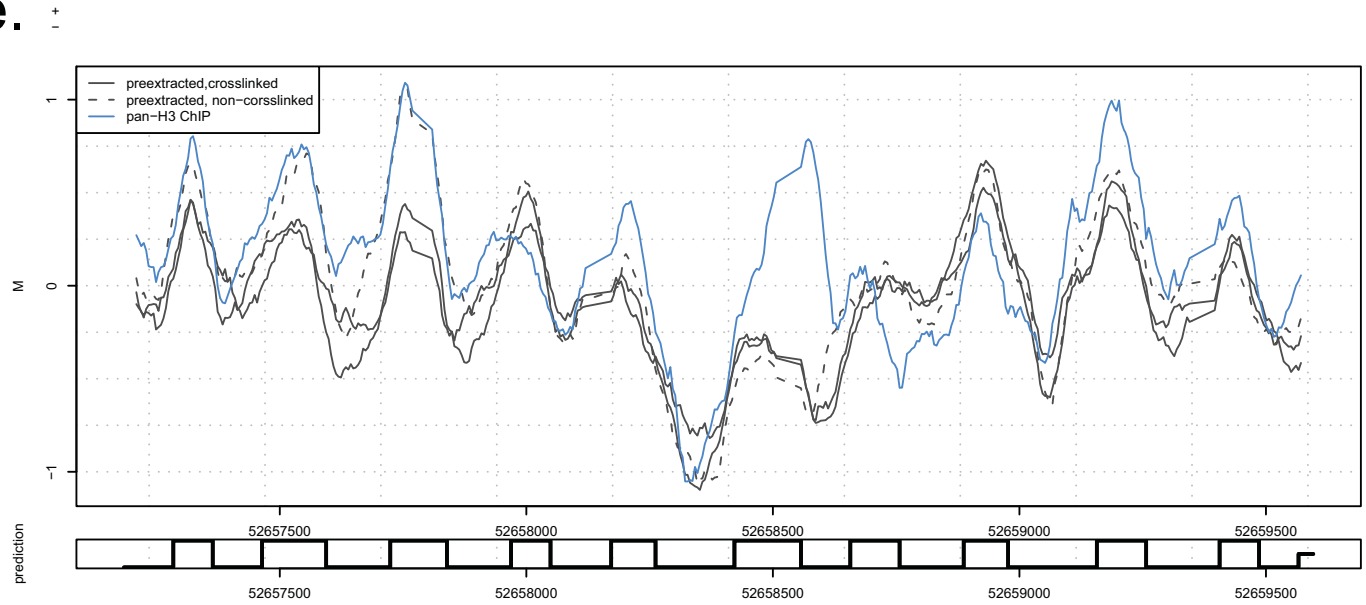
b.

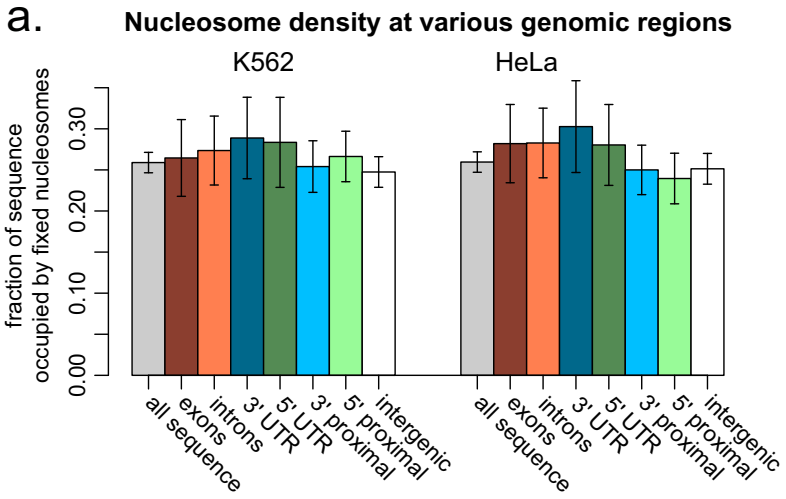


c.



e.


preextracted,crosslinked
preextracted, non−corsslinked
pan−H3 ChIP
M
1
0
−1
prediction
52657500
52658000
52658500
52659000
52659500



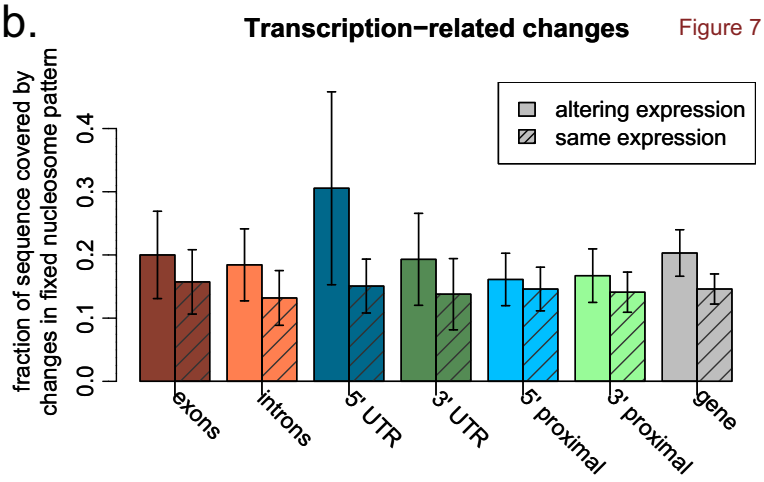
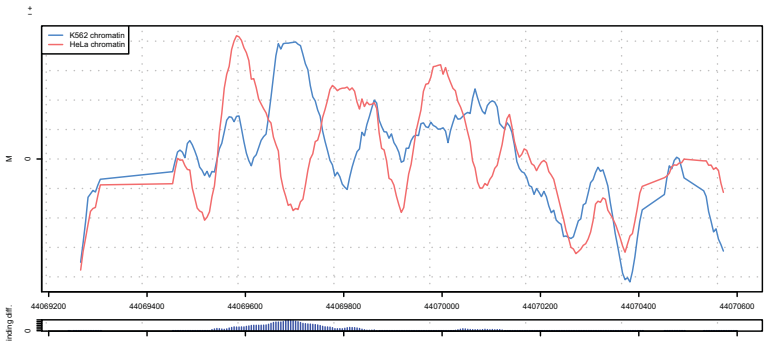


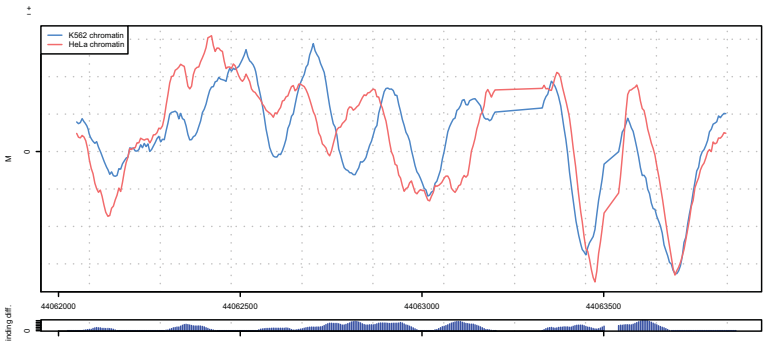
Figure 7

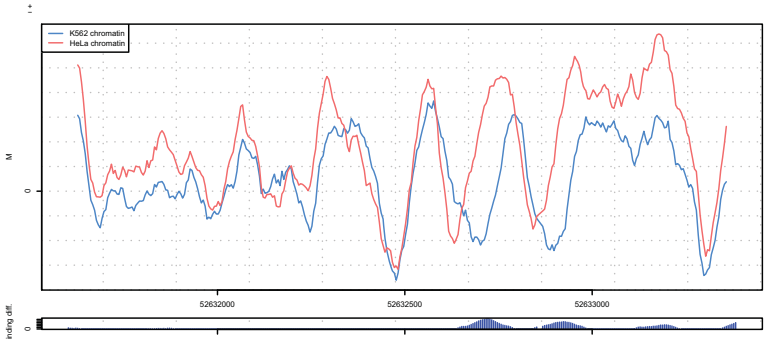
Nucleosome density and transcription-related changes. a. Fraction of sequence covered by predicted fixed nucleosome positions for different elements of gene structure in HeLa and K562 cells. Gene regions exhibit consistently higher density of fixed nucleosome positions. b. Fraction of sequence affected by differences in predicted fixed nucleosome regions between HeLa and K562 cell lines. Genes that later their expression state between the two cell lines exhibit more changes in the predicted fixed nucleosome position pattern.

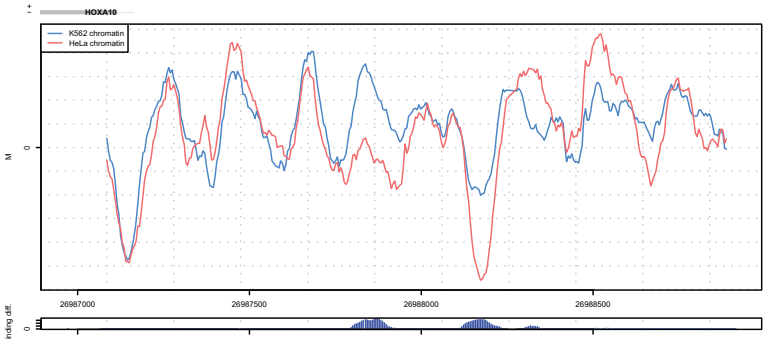
Examples of nucleosome differences between cell lines. Figures below provide examples of differences in nucleosome placement between K562 and HeLa cell lines. The figures show average pre-extracted chromatin profiles from all measurements for a given cell line. Binding difference scores are shown below each profile plot.

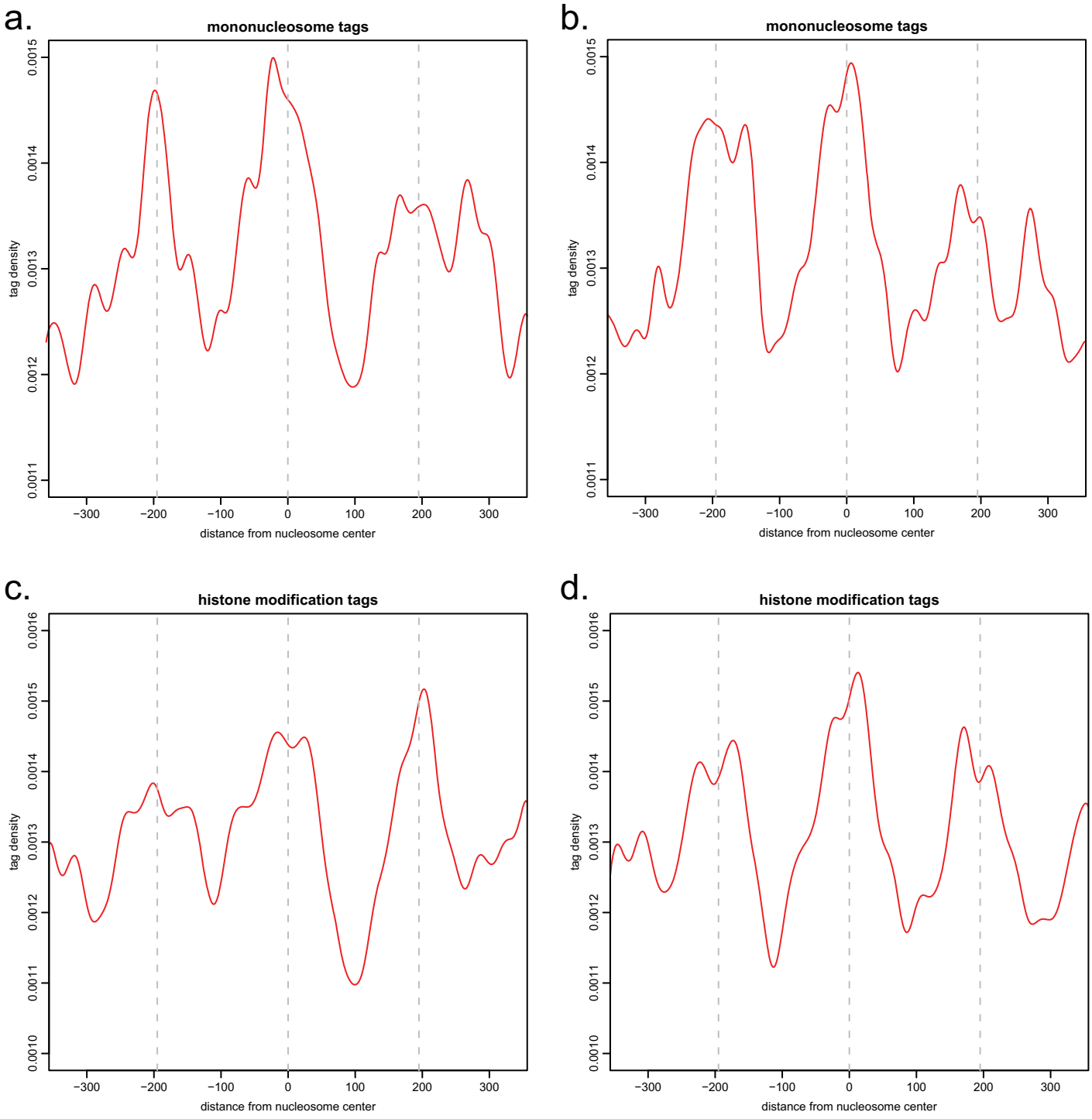
Figure 8









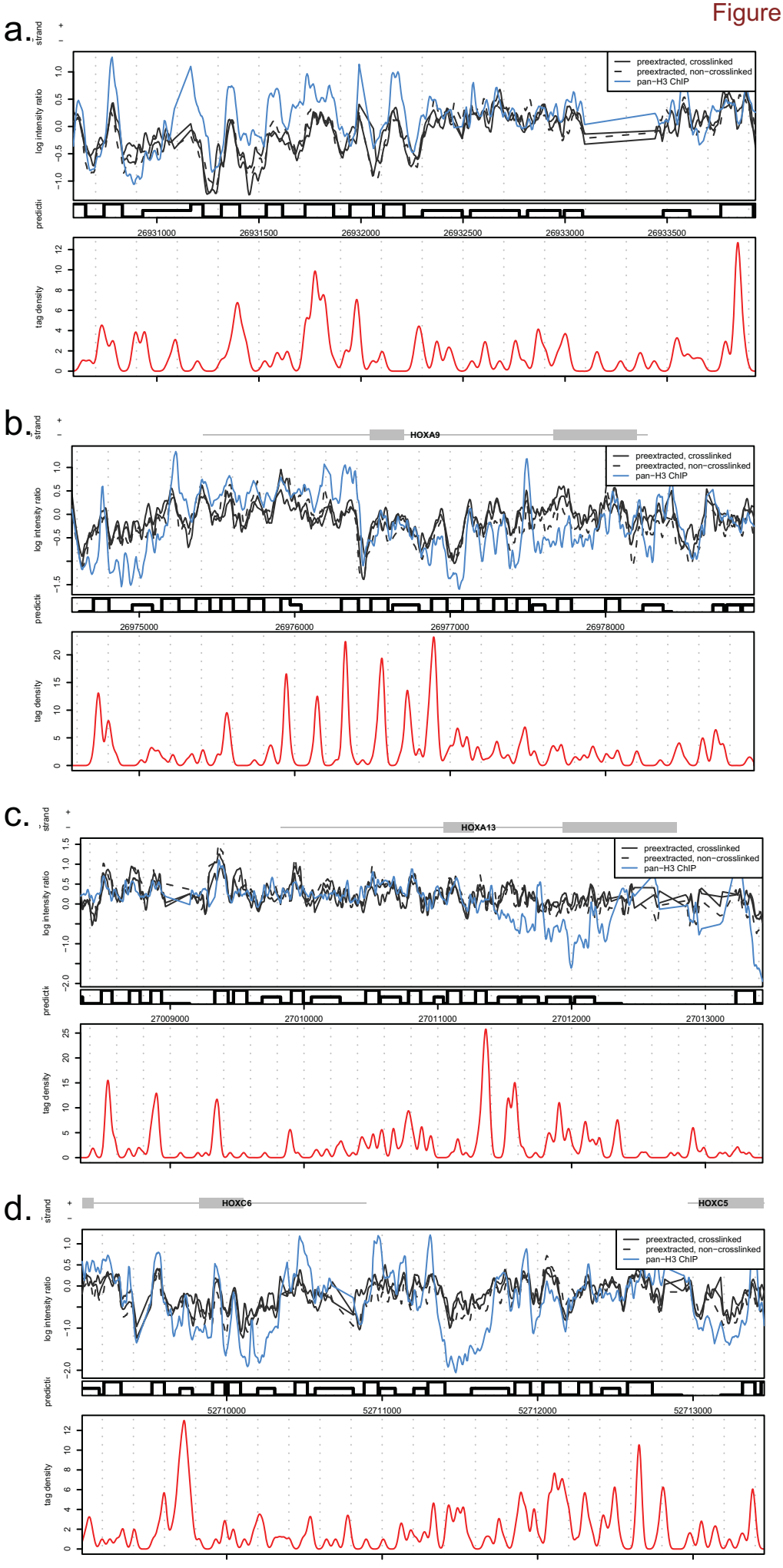


Average mononucleosome tag density profiles around fixed nucleosome positions. a. Using data from Schones *et al.* we identified positions corresponding to the 5′ ends of all mononucleosome sequence tags occurring within 350bp arond fixed nucleosome positions found in our analysis of K562 cell microarrays. The average density profiles were then calculated using a Gaussian smoothing kernel with a bandwidth of 10bp. The plots show peaks corresponding to the selected nucleosome (center) and adjacent nucleosomes at distances of approximately 200bp. **b.** Analogous profiles for fixed nucleosome positions identified in HeLa cells. **c,d.** Analogous profiles for K562 and HeLa cells, showing combined average density of sequence tags associated with histone methylations and H2A.Z variant taken from Barski *et al.*

Examples of agreement between identified fixed nucleosome positions and mononucleosome tag data by Schones *et al*.

The K562 chromatin profiles and nucleosome predictions are shown on the top track. The bottom track shows mononucleosome tag density as measured by Schones *et al*. Prominent (>6 tags) mononucleosome tag density peaks typically correspond to stable nucleosome positions.

The tag density profiles were calculated using a Gaussian kernel with 15bp bandwidth, scaling the kernel so that the height of the peak corresponds to the number of tags mapping to a given position. The density was calculated based on the 5′ ends of sequence tags matching to positive or negative strands. The 5′ end positions of the positive strand peaks were shifted prior to density calculation by 72bp to the right, and negative strand peaks were shifted 72bp to the left in order to account for the 147bp mononucleosome fragment length.



Examples of disagreement between identified fixed nucleosome positions and mononucleosome tag data by Schones *et al*.

The plots are analogous those shown in the Supplementary Figure 10, with regions illustrating low seque tag density at many of the fixed nucleosome positions identified in this study.

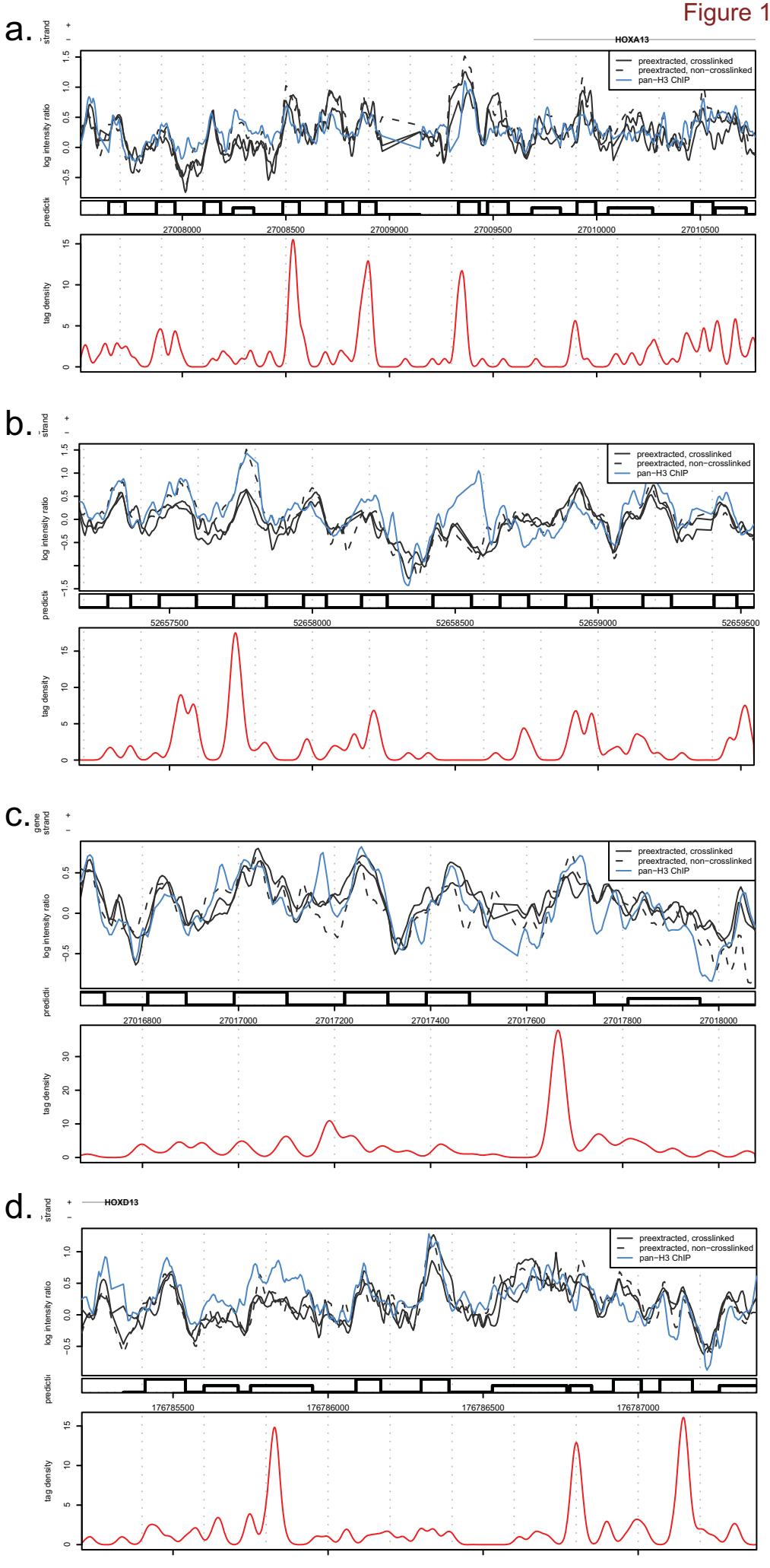
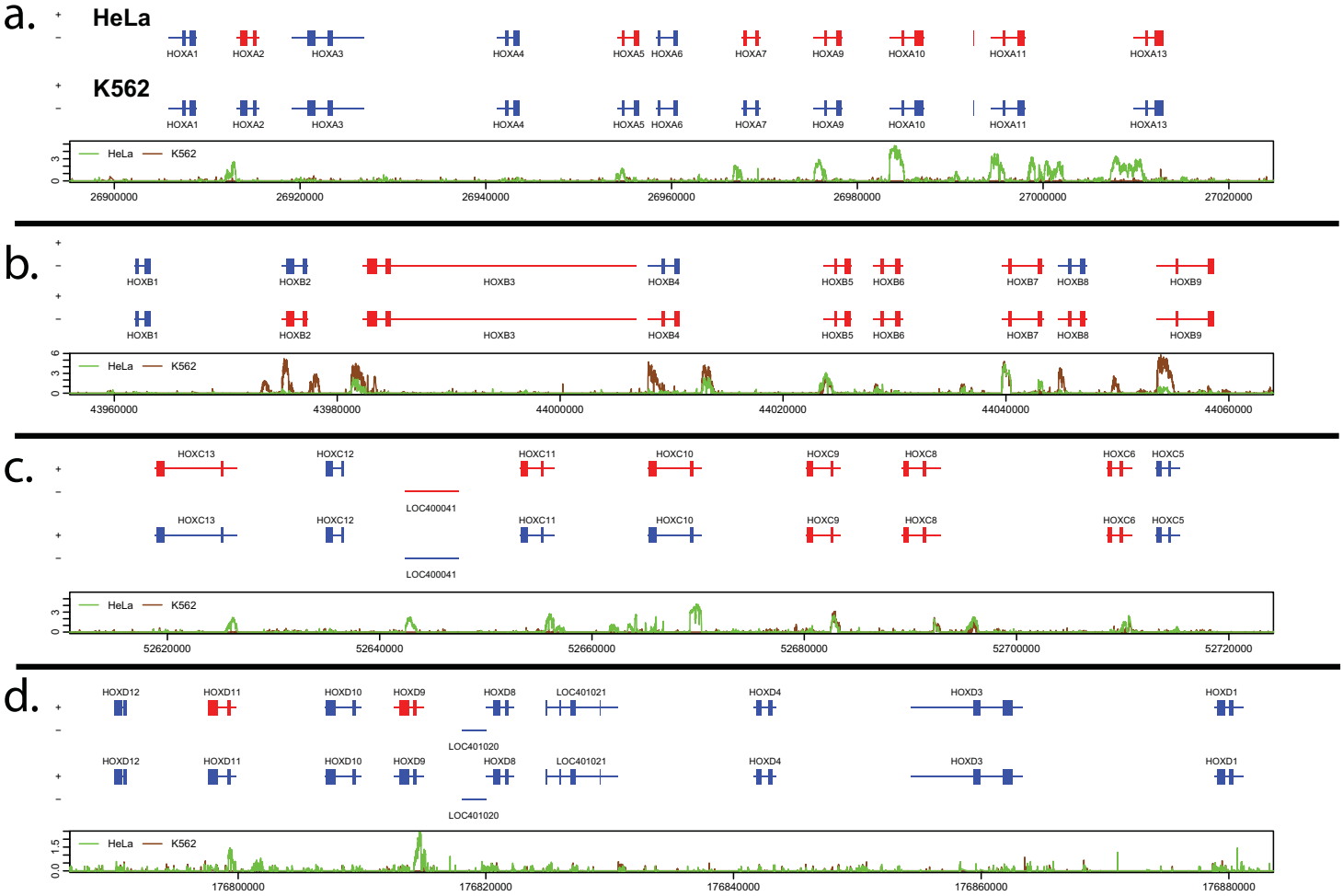
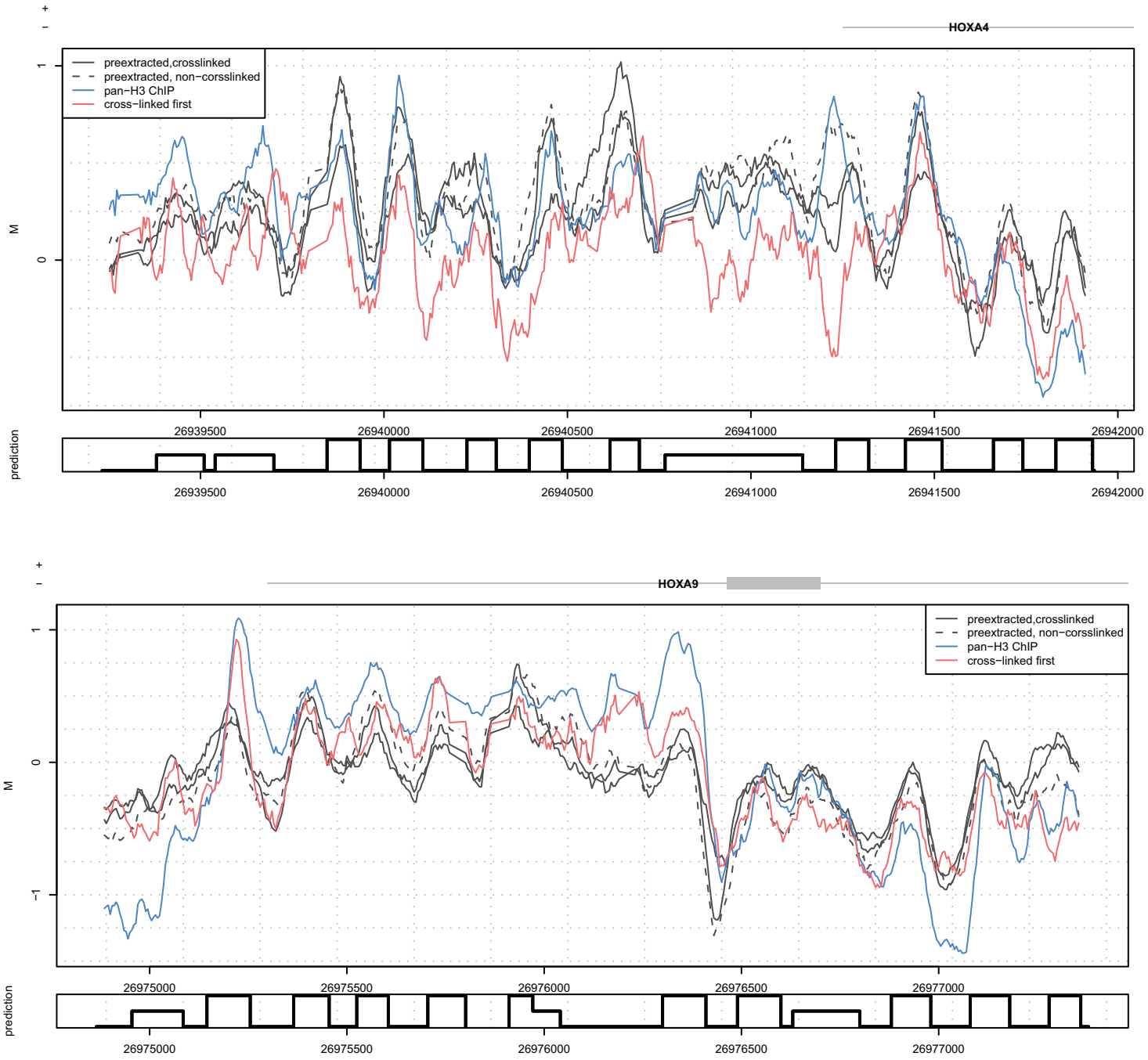


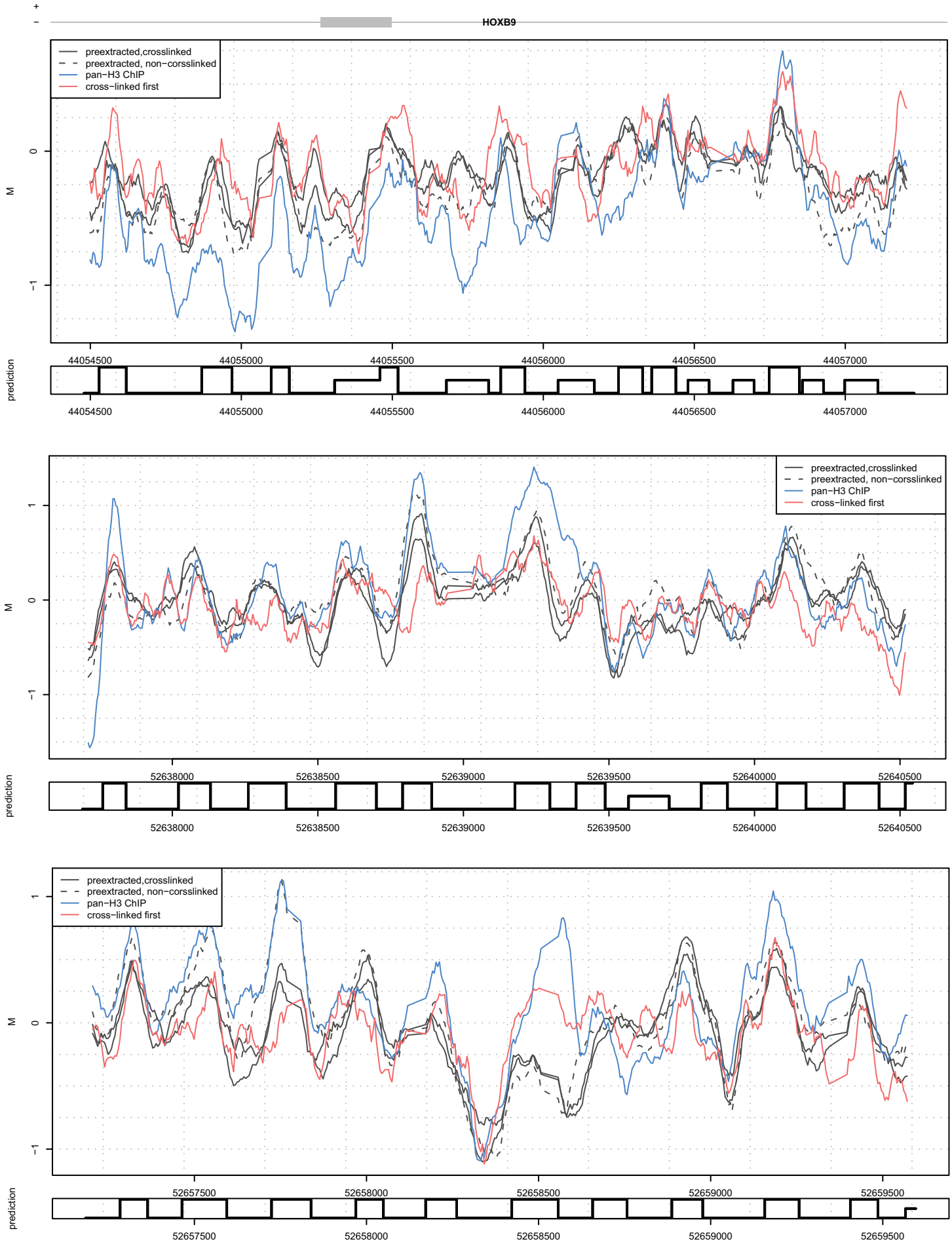
Figure 12

Expression states of HOX genes in K562 and HeLa cells. UFor each of the four HOX clusters (a-d), the plots show expression states of protein coding genes (red - expressed; blue - silent) in both cell lines. The bottom track shows log intensity ratio resulting from hybridization of cDNA to the tiling microarrays (see Methods). HeLa profiles are shown in green, K562 profiles in brown. In such profiles, the expressed genes are expected to exhibit peaks near the 3′ end.



Validating chromatin profiles without pre-extraction. The plots illusrtate agreement between K562 chromatin profiles obtained with (black lines) and without (red line) pre-extraction step. The pan-H3 ChIP profile is shown in blue. The same regions as in Supp. Figure 8 are shown. The profile obtained without pre-extraction exhibits peaks at most fixed nucleosome positions found in the pre-extracted data. However this profile is noisier and exhibits additional, narrowly-spaced peaks that are not characteristic of fixed nucleosome positions, and are more likely to reflect other




+
-
HOXB9
preextracted,crosslinked
preextracted, non−corsslinked
pan−H3 ChIP
cross−linked first
M
0
−1
44054500
44055000
44055500
44056000
44056500
44057000
prediction
preextracted,crosslinked
preextracted, non−corsslinked
pan−H3 ChIP
cross−linked first
M
1
0
−1
52638000
52638500
52639000
52639500
52640000
52640500
prediction
preextracted,crosslinked
preextracted, non−corsslinked
pan−H3 ChIP
cross−linked first
M
1
0
−1
52657500
52658000
52658500
52659000
52659500
prediction