SUPPLEMENTAL MATERIAL

Comparative analysis of H2A.Z nucleosome organization in human and yeast genome

Michael Y. Tolstorukov[1,2,*], Peter V. Kharchenko[1,2,3,*], Joseph A. Goldman[4], Robert E. Kingston[4] and Peter J. Park[1,2,3,§]

[1]Center for Biomedical Informatics, Harvard Medical School, 10 Shattuck St., Boston, MA 02115 USA
[2]Department of Medicine, Brigham and Women's Hospital, 77 Avenue Louis Pasteur, Boston, MA 02115 USA
[3]Children's Hospital Informatics Program, 300 Longwood Ave., Boston, MA 02115 USA
[4]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA.

*These authors contributed equally to this work.
[§]The correspondence should be addressed to PJP (peter_park@harvard.edu)

Autocorrelation analysis of human nucleosome sequences

Since the periodic pattern in yeast was most pronounced for the strongly-positioned H2A.Z sequences (Albert et al., 2007), we confirmed the absence of the periodic patterns in the 10% of human nucleosome sequences associated with highest numbers of sequenced tags (Figure S18). The sequence organization in the vicinity of the TSS differs in human and yeast (Lander et al., 2001; Oliver et al., 1992) and we verified that the 10-bp periodicity is not exhibited by the nucleosomes associated with the subsets of genes stratified by the presence of CpG islands, TATA-box, INR motif or by overall GC content in the TSS proximity (Figure S19-21).

Transcription machinery may shift nucleosomes from the thermodynamically favorable locations, hindering detection of the 10-bp pattern in nucleosome sequences. Using the availability of expression data for resting CD4+ T cell where 'bulk' nucleosomes were detected (Schones et al., 2008) we confirmed that similar results can

be obtained for all the detected 'bulk' nucleosomes and for the subsets associated with transcriptionally active and silent genes (Figures S14, S21I,J).

Validation of inner variation in GC-profile for human H2A.Z nucleosomes

In addition to the validation described in the main text, the GC-content and tag density profiles were calculated for human H2A.Z sequences that contain inner tags at only one side of the distinctive 147-bp fragment (Figure S24) and for the sequences that contain inner tags at both ends of the 147-bp fragment (Figures S25). Only one pronounced variation in GC-profile is seen in the first case and two symmetrical variations are seen in the second case. To exclude possible bias due to nucleosome detection procedure we calculated GC-profile for all sequenced H2A.Z tags extended to 200 bp towards 3'-end (Figure S26). The corresponding variation in GC-content is clearly seen for this set of sequences 120-130 bp downstream from the cut site. No complementary sequences were added in this case therefore the GC-content variation cannot be an artifact of accounting for both direct and reverse sequences for each H2A.Z nucleosome fragment.

The 10-bp periodicity in tag density and average dinucleotide profiles

The emergence of the 10-bp periodicity in the distribution of the sequenced tags without the 10-bp periodicity in the underlying DNA sequence may be connected to the formation of higher order chromatin structures (Thoma, 1992). For example, the shift of the histone core on the DNA sequence by 5 bp would result in 180-degree rotation of the nucleosome in 3D space. Such a rotation has to be compensated by adding torsion stress to the DNA molecule in the linker regions in order to maintain the nucleosome packaging into a fiber. This mechanism does not require presence of any periodicity in the underlying DNA sequence.

Another possible mechanism, which can also be responsible for emergence of 10-bp periodicity in average dinucleotide profiles, is connected to the presence of a strong positioning signal that does not exhibit 10-bp periodicity in the DNA sequence. Such a signal may appear as a single motif of about 10 bp in length or as several shorter motifs forming a specific pattern along the 147-bp path of nucleosomal DNA (Figure S28). In

the case of the rest of the nucleosomal DNA being 'neutral' in terms of nucleosome positioning, presence of even a sparse positioning signal may determine rotational setting of a nucleosome. An example of such a 'distributed' positioning signal is presented in Figure S28B-E. Although there is no apparent periodicity in the DNA sequence, the clear rotation preference is seen in the distribution of the sequenced tags and in the deformation energy profile around the detected nucleosome position. We compared the deformation energy score at each dinucleotide step of nucleosomal DNA calculated for the genomic sequence and for a DNA model without sequence-dependence of the structural properties, so called mixed-sequence DNA (Tolstorukov et al., 2007). As expected for the positioning sequence, the energy score is noticeably lower for the genomic DNA, by more than 15%. Many of the steps that provide the largest contribution to the energy score difference arrear at the sites of extreme DNA deformations associated with DNA bending into the minor groove (Tolstorukov et al., 2007). Analysis of the dinucleotide step distribution in this sequence shows that although in some cases the same steps appear to be separated by ~10 bp (green oval) in most cases the same steps are separated by other distances (red oval). In many cases different steps occupy the critical positions separated by ~10 (blue oval) effectively diminishing the deformation energy, which also does not contribute to the formation of the 10-periodicity in the sequence.

To check if it is possible to obtain the pronounced 10-bp pattern in the sequences that do not have 10-bp periodicity by shifting the histone core several base pairs to select the preferable rotational setting we performed the following simulation. The set of random sequences was aligned in such a way that the occurrences of AA dinucleotides were maximized at the locations separated by 10 bp intervals (Figure S29). The maximal allowed sequence shift during alignment was 3 bp. The average dinucleotide profile calculated for the aligned sequences shows clear 10-bp periodicity not only for AA dimers but also for GG dimers in counter-phase.

The described mechanisms are not exclusive and may be both operative in the human genome. Obviously, the presence of the sequences, exhibiting 10-bp periodicity may be even more effective for restricting nucleosome rotational setting than the mechanisms described above; such sequences, however, were not detected in noticeable numbers in our analysis of the human chromatin.

REFERENCES

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. Journal of molecular biology *319*, 1097-1113.

Johnson, S.M., Tan, F.J., McCullough, H.L., Riordan, D.P., and Fire, A.Z. (2006). Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. Genome research *16*, 1505-1516.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Oliver, S.G., van der Aart, Q.J., Agostoni-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P., Benit, P., and et al. (1992). The complete DNA sequence of yeast chromosome III. Nature *357*, 38-46.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Thoma, F. (1992). Nucleosome positioning. Biochimica et biophysica acta *1130*, 1-19.

Tolstorukov, M.Y., Colasanti, A.V., McCandlish, D.M., Olson, W.K., and Zhurkin, V.B. (2007). A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. Journal of molecular biology *371*, 725-738.

SUPPLEMENTAL FIGURE CAPTIONS:

Figure S1. MNase digestion of human chromatin from HeLa cells.

(A) Chromatin samples enriched for H2A and H2A.Z nucleosomes were treated with 280 U/ml MNase for one hour at room temperature to verify the differences between the digestion of the canonical H2A nucleosomes and the nucleosomes containing H2A.Z variant. After deproteinization DNA fragments were run for a longer time on 7% acrylamide gel than in the case shown in Figure 2 in the main text to display the differences between the digestion of the H2A and H2A.Z nucleosomes more clearly. The 1-kb ladder is shown on both sides of the gel.

(B) Chromatin samples were treated with increasing amounts of MNase (140, 280, and 700 U/ml) for one hour at room temperature. Results for the input and for the samples enriched for H2A and H2A.Z nucleosomes are shown for each MNase concentration tested.
(C) MNase digestion of the *in vitro* assembled recombinant H2A and H2A.Z nucleosomes. Nucleosomes were reconstituted on the TPT positioning sequence and treated with 280 U/ml MNase for increasing periods of time, 25s, 1.5m, 3m, 4.5m (indicated on the plot). PAGE was used to check the lengths of the digestion products.

Figure S2. Tag density distribution around H2A.Z nucleosome predictions based on 117-bp inter-strand peak distance. Dark red and blue lines represent tag density on the positive and negative strands respectively. Dotted orange lines mark the possible positions of secondary peaks expected for the alternative orientation of the shortened 117-bp fragments relative to the complete 147-bp sequence in the cell population. Upper diagram: the cyan line gives a 117-bp reference and the gray lines give 147-bp reference lengths.

Figure S3. Secondary nucleosome positions tend to be separated by 10 bp. Many nucleosomes exhibit tag patterns suggesting multiple overlapping preferred positions. The plot shows distribution of distances between secondary positions belonging to the same coarse-grained nucleosome call. Both H3K4me3 (black) and H2A.Z (red) nucleosomes show preferred distance of 10 bp between the secondary positions.

Figure S4. The autocorrelation in the H2A.Z tag density shown for the tags located in the 2-kb proximity of the TSS of annotated genes (black line) and for the tags distant from the TSS (red line).

Figure S5. The nucleosome positioning patterns around TSS of the genes characterized by presence of (A) CpG island, (B) TATA-box, and (C) INR motif.

Figure S6. Comparison of the WW and SS dinucleotide profiles for the yeast H2A.Z nucleosome sequences determined by Albert et al., 2007 (cyan and pink lines) and using

the nucleosome detection algorithm developed in this study (blue and red lines). The set from Albert et al. comprises ~8,000 nucleosome sequences and the set determined in this study comprises ~ 6,000 nucleosome sequence.

Figure S7. Dinucleotide profiles for the sequences around all MNase cut sites determined in ChiP-Seq experiment for human and yeast H2A.Z nucleosomes. Sequences are aligned by the cut site at positions zero. The sequences only in direct orientation were taken into consideration. (A,B) Profiles for WW and SS dinucleotides and (C,D) profiles for AA, TT, CC, and GG dinucleotides for human and yeast H2A.Z sequences respectively.

Figure S8. Dinucleotide profiles for the nucleosome sequences and the sequences around all MNase cut sites for human H3K4me3-enriched and 'bulk' nucleosomes. (A,B) Average profiles for WW and SS dinucleotides calculated for nucleosome sequences in H3K4me3 and 'bulk' nucleosomes. Profiles for human sequences are shown with red and blue lines. Corresponding profiles for yeast H2A.Z sequences are shown with pink and cyan lines for comparison. (C,D) Profiles WW and SS dinucleotides and (E,F) profiles for AA, TT, CC, and GG dinucleotides for sequence around experimental cut sites in human H3K4me3 and 'bulk' nucleosome sets respectively. The sequences are aligned by the cut site at positions zero. Sequences only in the direct orientation are taken into consideration.

Figure S9. Dinucleotide autocorrelation function for human and yeast nucleosome sequences. Data are shown for WW, SS and 10 individual dinucleotides. The complementary dinucleotides were considered separately (AA≠TT). Since both direct and reverse orientations were used for calculation of autocorrelation function the profiles for complementary dinucleotides are identical and only one profile is shown in each case.

Figure S10. Dinucleotide autocorrelation function for human H3K4me3 nucleosome sequences. Data are shown for WW, SS and 10 individual dinucleotides.

Figure S11. Dinucleotide autocorrelation function for human 'bulk' nucleosome sequences. Data are shown for WW, SS and 10 individual dinucleotides.

Figure S12. Periodograms representing the spectral density calculated for the autocorrelations of 10 individual dinucleotides in human and yeast H2A.Z nucleosome sequences. The complementary dinucleotides were considered separately (AA≠TT). The red lines represent the power spectral density for nucleosomal sequences and solid and dashed blue lines represent the statistical significance levels P = 0.001 and P = 0.05 respectively (not corrected for multiple hypothesis testing). Randomizations for individual dinucleotides were repeated 10 times (independently in each of three 41-bp non-overlapping windows per sequence) and the random sequence with the dimer composition closest to that of genomic sequence was taken for further analysis.
Note that the power density frequently reaches the level of significance P = 0.05 on the periodograms, which is expected for such a relaxed criterion. We are not focusing on such observations for three reasons: (i) the periods corresponding to such instances are not consistent between the plots, (ii) most of these periods cannot be directly related to the structural organization of nucleosomal DNA, and (iii) substantial number of points are expected to cross P = 0.05 level by chance, given the number of points in each plot and number of subsets examined.

Figure S13. Periodograms representing the spectral density calculated for autocorrelations of the WW, SS, and 10 individual dinucleotides in the human H3K4me3 nucleosomes.

Figure S14. Periodograms representing the spectral density calculated for autocorrelations of the WW, SS, and 10 individual dinucleotides in the broad class of human nucleosomes not associated with any specific histone variant or modification. Many of the nucleosomes in this set are located in DNA regions containing considerable fraction of repetitive sequences. Since presence of the repeats in nucleosome sequences can interfere with the periodicity analysis, increasing the level of noise in the power spectra, the nucleosomes containing the repetitive sequences were filtered out.

Figure S15. Dinucleotide autocorrelation function for the sets of human and yeast H2A.Z sequences located near all experimentally determined MNase cuts. Each cut site was

treated as 5'-end of nucleosomal fragment. The sequences only in direct orientation are included in the analysis. Data are shown for WW, SS and 16 individual dinucleotides. The complementary dinucleotides were considered separately (AA≠TT).

Figure S16. Dinucleotide autocorrelation function for human H3K4me3 sequences located near all experimentally determined MNase cuts. Data are shown for WW, SS and 16 individual dinucleotides.

Figure S17. Dinucleotide autocorrelation function for human 'bulk' nucleosome sequences located near all experimentally determined MNase cuts. Data are shown for WW, SS and 16 individual dinucleotides.

Figure S18. Periodograms representing the spectral density calculated for the autocorrelations of WW (=AA, TT, AT, TA) and SS (=GG, CC, GC, CG) in the 10% of nucleosome sequences that are associated with highest numbers of sequenced tags. (A) human H2A.Z, (B) human H3K4me3, and (C) human 'bulk' nucleosomes

Figure S19. Periodograms representing the spectral density calculated for the WW and SS dinucleotides in the subsets of human H2A.Z nucleosomes. (A) Nucleosome sequences that have GC-content higher than 60% and (B) GC-content lower than 50%; (C-H) nucleosome positions located in the proximity of TSS (±2 kb) characterized by the presence (C,E,G) or absence (D,F,H) of CpG island, TATA-box, and INR motif, respectively.

Figure S20. Periodograms representing the spectral density calculated for the WW and SS dinucleotides in the subsets of human H3K4me3 nucleosomes. (A) Nucleosome sequences that have GC-content higher than 60% and (B) GC-content lower than 50%; (C-H) nucleosome positions located in the proximity of TSS (±2 kb) characterized by the presence (C,E,G) or absence (D,F,H) of CpG island, TATA-box, and INR motif, respectively.

Figure S21. Periodograms representing the spectral density calculated for the autocorrelations of WW and SS dinucleotides in the subsets of human 'bulk' nucleosomes. (A) Nucleosome sequences that have GC-content higher than 60% and (B) sequences that have GC-content lower than 50%; (C-H) nucleosome positions located in the proximity of TSS (±2 kb) characterized by the presence (C,E,G) or absence (D,F,H) of CpG island, TATA-box, and INR motif, respectively, and (I,J) nucleosome positions within the boundaries of no more than 2 kb upstream of 5'-ends of the expressed and silent genes respectively.

Figure S22. MNase bias reflected in the average nucleotide profile of the DNA fragments around 5'-ends of all sequence tags in the Solexa data. Profiles for the individual nucleotides and GC-profile are shown. Zero position on the x-axis corresponds to the MNase cut.

Figure S23. Tag density (upper panels) and average GC-profiles (lower panels) calculated for H2A.Z nucleosomes located in the proximity of the H3K4me3 nucleosomes. Results are shown for the H2A.Z nucleosomes shifted by no more than 5 bp (A) and by 30-50 bp (B) relative to the H3K4me3 nucleosomes.

Figure S24. Tag density (A) and average GC-content profile (B) for the subsets of H2A.Z nucleosomes containing only one inner-cut site inside 147-bp fragments. Dark red and blue lines in (A) represent tag density on DNA positive and negative strands respectively. The sequences in the subset were oriented in such a way that cuts occur at 3'-end side relative to the "plus" strand. Thick line in (B) represents 11-bp smoothing of the initial data shown with thin line.

Figure S25. Tag density (A) and average GC-content profile (B) for the subsets of H2A.Z nucleosomes containing two inner-cut sites inside 147-bp fragments. The sequences in the subset are oriented in such a way that cuts occur at 3'-end side relative to the 'plus' strand. The sequences only in direct orientation were included in the analysis (i.e., no complementary sequences were added).

Figure S26. GC-content profiles for the sequences around all MNase cut sites for human H2A.Z sample. Sequences are aligned by the cut site at position zero, so that the nucleosome fragments protected from the digestion are located on the right side. The 147-bp distance from the cut site is marked with red line. The sequences only in direct orientation are taken into consideration.

Figure S27. Periodograms representing the spectral density for the WW and SS dinucleotide autocorrelation functions calculated in the 2-kb regions around transcription starts in human (A,B) and yeast (C) genomes. Results are shown for the human genes where at least one stable H2A.Z (A) or H3K4me3 (B) nucleosome position was detected and for all yeast genes (C).

Figure S28. Restricted rotational setting of a nucleosome as a result of the presence of a positioning signal not exhibiting the 10-bp periodicity.

(A) Schematic illustration of the underlying mechanism. Thick color bars represent the sequence motifs favorable for the sites where the DNA minor groove faces inward (blue) and outward (red) the histone surface (gray), or in between such sites (green). Red arrow indicated relative orientation of the histone core and DNA. Since similar structural organization is imposed by the histone core on nucleosomal DNA with a periodicity about 10 bp (Luger et al., 1997) such a sequence motif would be favorable for the all the settings of the histone core on the DNA separated by 10-bp distance.

(B) Example of the detected nucleosome position exhibiting clear 10-bp rotational positioning. The H3K4me3 nucleosome from the promoter region of the gene TIMM17A is shown. Red and blue bars represent the tag counts on the plus and minus strands respectively. The smoothed tag density is shown with red and blue lines.

(C) The 147-bp sequence corresponding to the nucleosome position shown in (A). The sites of the minor and major groove bending are colored blue and red respectively; the 10-bp periodicity is assumed in locations of these sites, which correspond to the observed sequence periodicity in location of WW and SS dimers in the published profiles for yeast and worm sequences (Albert et al., 2007; Johnson et al., 2006; Segal et al., 2006). There is no apparent 10-bp periodicity in the human sequence.

(D) Deformation energy profile calculated around the nucleosome position shown in (A). The deformation energy was calculated as described in (Tolstorukov et al., 2007), using central 129 bp of the high-resolution nucleosome structure symmetrized relative to the dyad (Davey et al., 2002). The main minimum in the profile corresponds to the experimentally detected nucleosome position (position zero) with 2-bp accuracy. The local minima indicated with numbers correspond to the same rotational setting of the nucleosome.

(E) The eleven dinucleotide steps making the largest contributions to the difference between the deformation energy score of the genomic sequence and the averaged mixed-sequence DNA, which represents the duplex with no sequence dependence of the structural properties. The positions of the steps relative to the center of the predicted nucleosome are indicated above the sequence and the contributions of each step to the energy difference are indicated blow the sequence. The total energy score of the genomic nucleosome sequence is 128 au less than that of the mixed-sequence DNA, which is indicative of the positioning sequence.

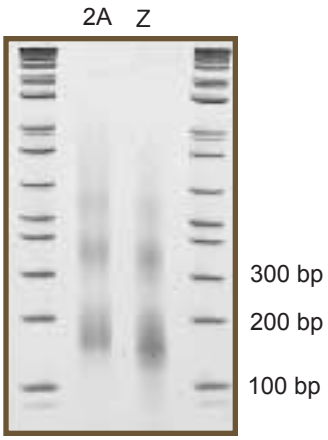
Figure S29. Alignment of random sequences can result in pronounced periodicity in average dinucleotide profile. (A) Schematic illustration of the alignment. Random sequences are shifted so that the frequencies of a specific dinucleotide (AA) are maximized at given locations in the average profile. The enrichments at the locations separated by 10 bp (red rectangles) originate from different sequences. (B) Average profiles for AA, TT, GG, and CC dinucleotides in a set of random sequences of the same size and dinucleotide composition as those of the set of human H2A.Z nucleosomes. (C) Average profiles for the random set after alignment aimed to maximize the frequencies of AA dinucleotides at base pairs +/-4,5,6, 14,15,16,...,64,65,66 relative to sequence center. The maximal sequence shift allowed in the alignment was 3 bp. Both direct and reverse sequence orientations were accounted for and smoothing in 3-bp running window was applied to the profiles in (B) and (C).

Figure S30. GC-profile for the human nucleosomes from resting CD4+ T cells. Both direct and reverse sequence orientations were considered.

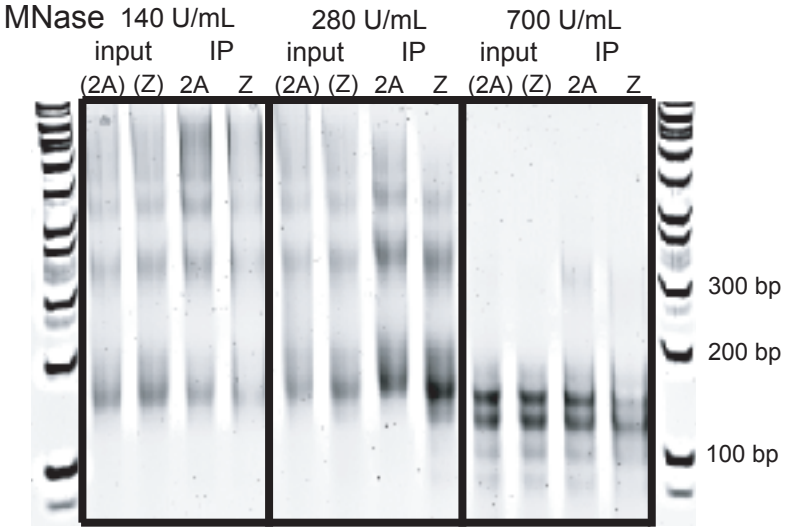
Figure S31. Histogram of the distances between 5'-5' and 3'-3' ends of H2A.Z nucleosome predictions based on the 117-bp and 147-bp distance between peaks in the tag density.

Figure S32. Results obtained for the tag contributions normalized by the relative occurrence of the 5-mers at the sites of the MNase cuts and genome-wide. (A) Stable nucleosome pattern around TSS; (B) periodograms of the WW and SS dinucleotide autocorrelations calculated for the H2A.Z and K4me3 nucleosome predictions; (C) GC-profiles for the H2A.Z and K4me3-associated sequences.

A



B



C

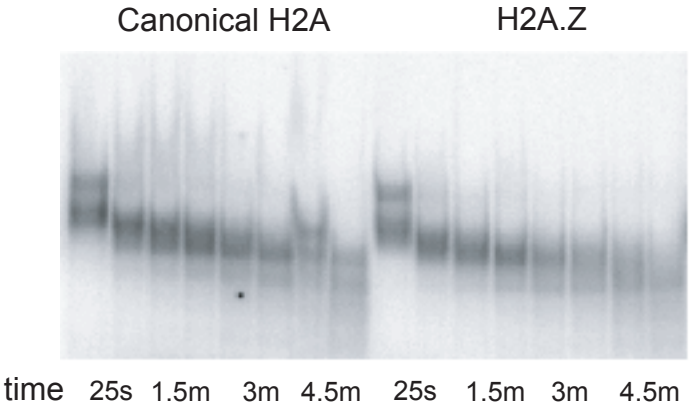


Figure S1

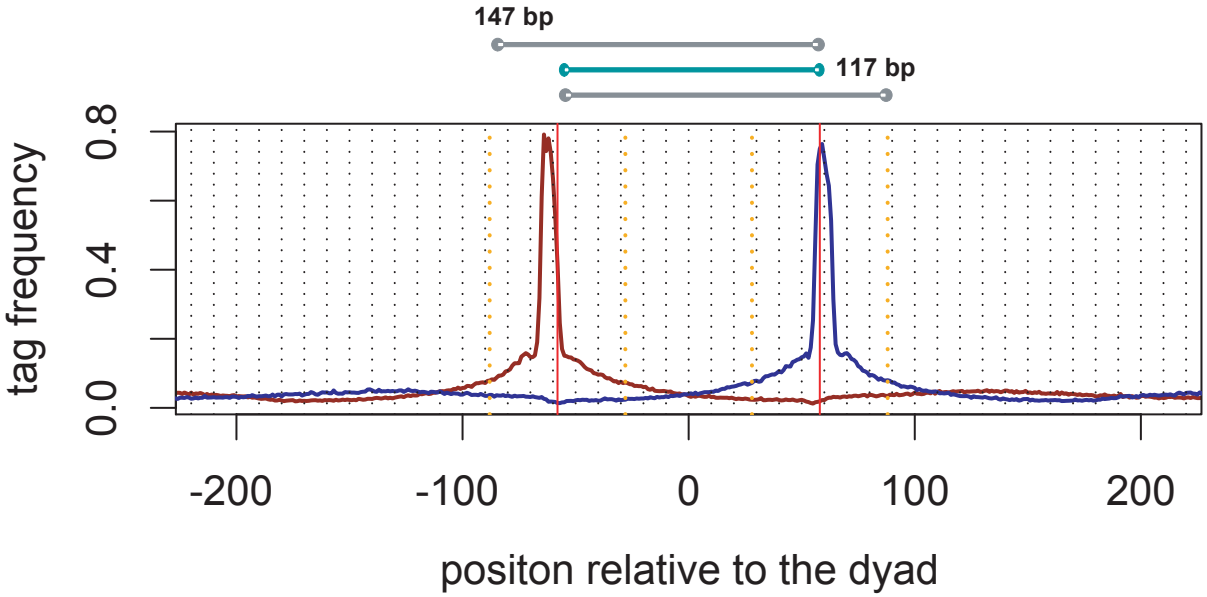

147 bp
117 bp
tag frequency
0.0
0.4
0.8
-200
-100
0
100
200
positon relative to the dyad

Figure S2

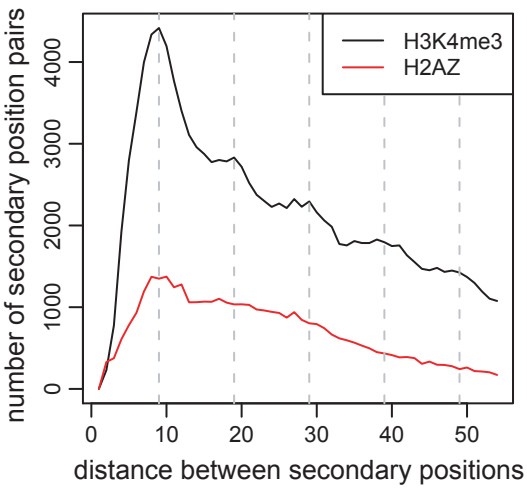


Figure S3

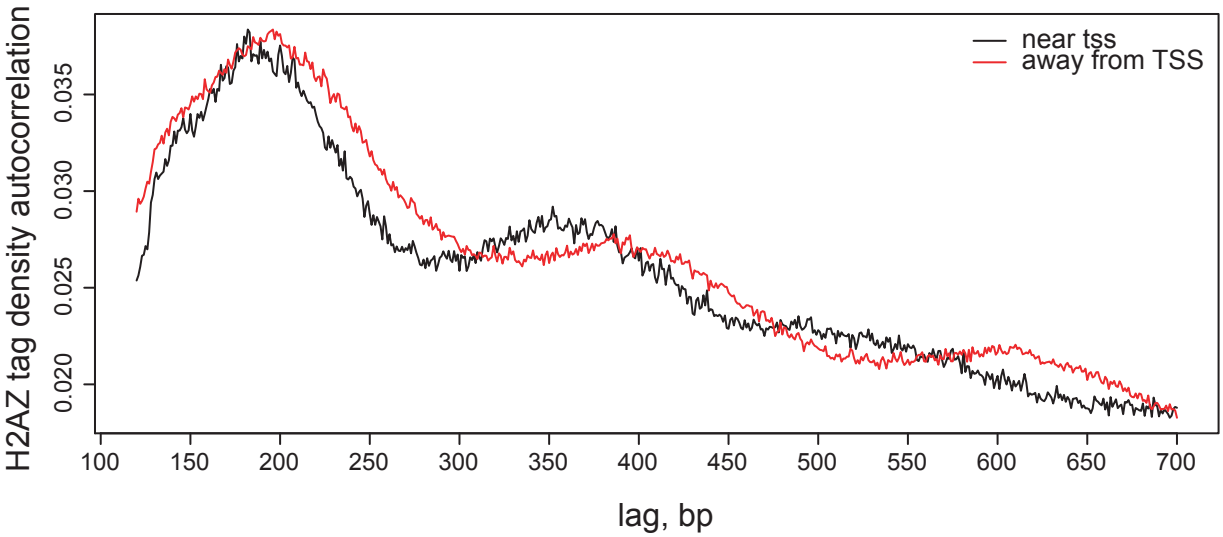
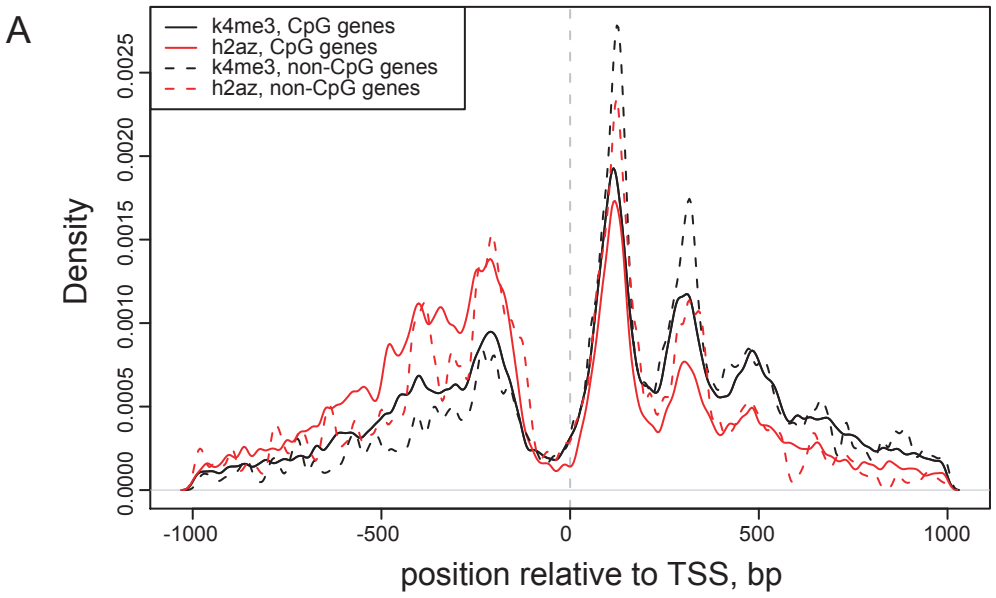
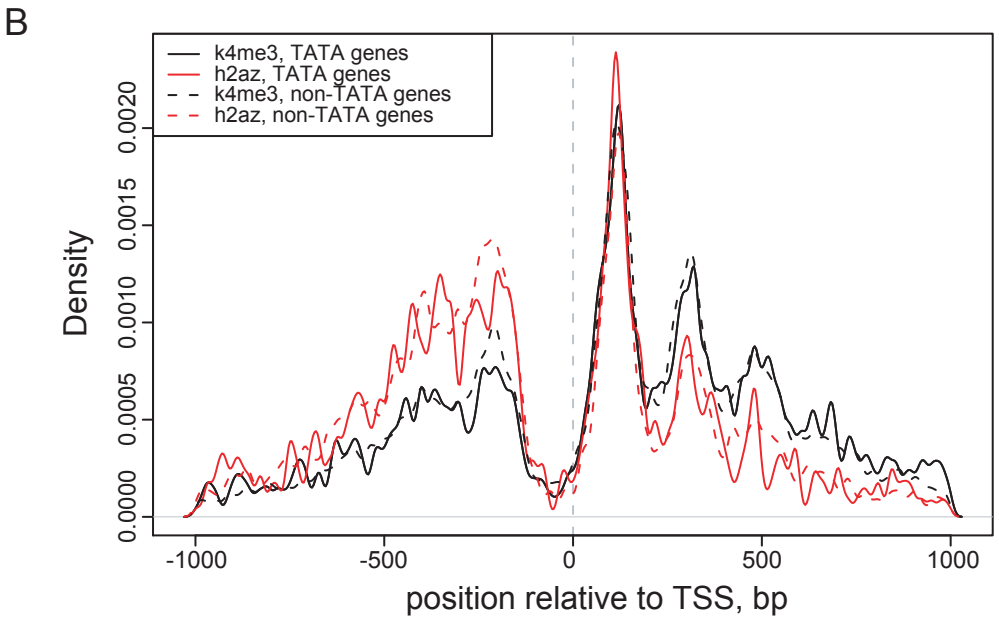


Figure S4


A
k4me3, CpG genes
h2az, CpG genes
k4me3, non-CpG genes
h2az, non-CpG genes
Density
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
-1000
-500
0
500
1000
position relative to TSS, bp


B
k4me3, TATA genes
h2az, TATA genes
k4me3, non-TATA genes
h2az, non-TATA genes
Density
0.0000
0.0005
0.0010
0.0015
0.0020
-1000
-500
0
500
1000
position relative to TSS, bp

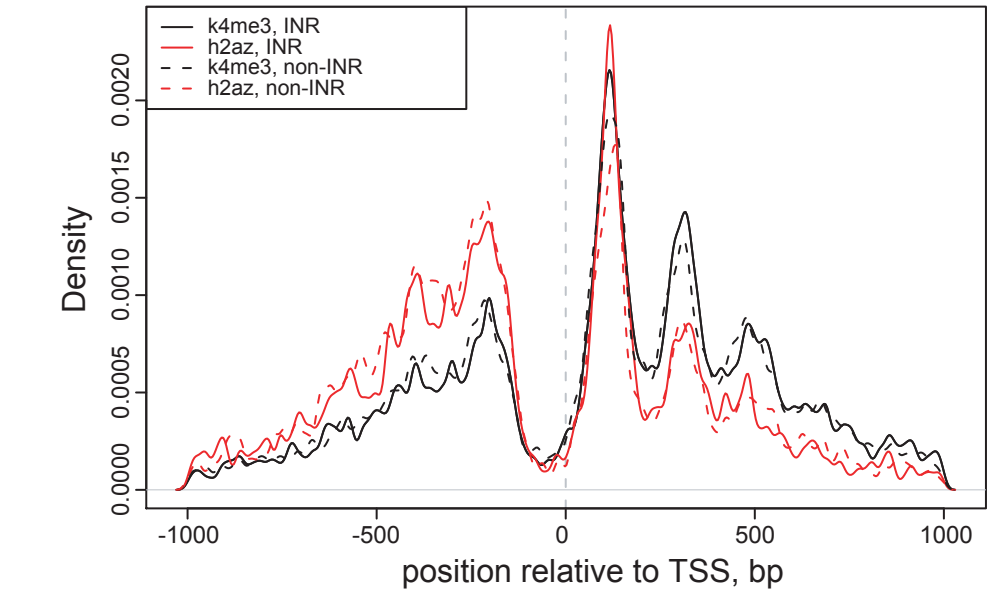

C
k4me3, INR
h2az, INR
k4me3, non-INR
h2az, non-INR
Density
0.0000
0.0005
0.0010
0.0015
0.0020
-1000
-500
0
500
1000
position relative to TSS, bp

Figure S5

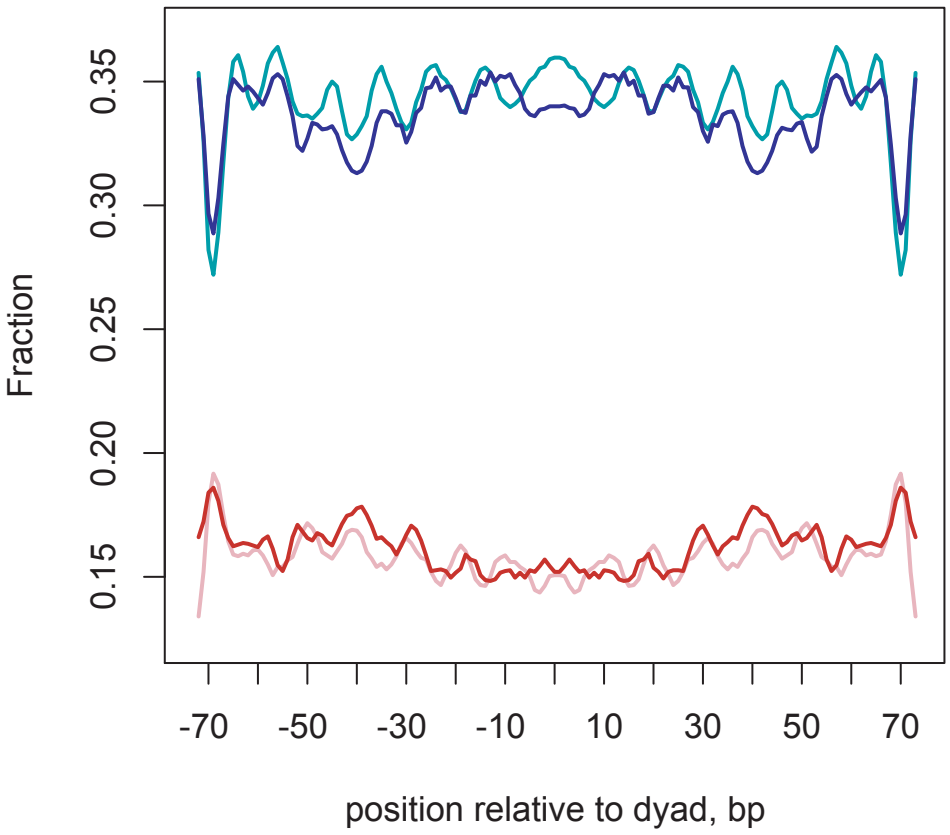
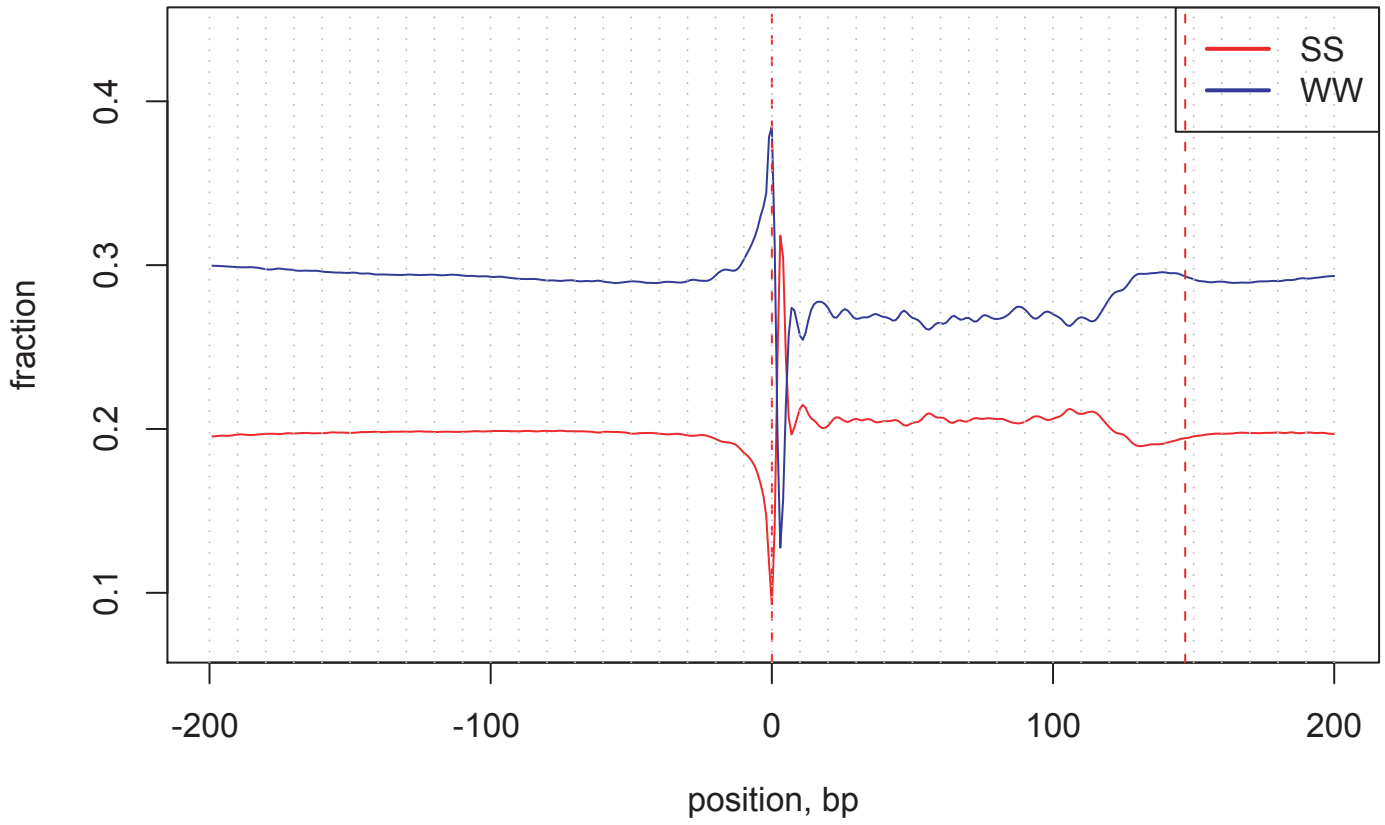

Fraction
0.15
0.20
0.25
0.30
0.35
-70
-50
-30
-10
10
30
50
70
position relative to dyad, bp

Figure S6

A


human H2A.Z
SS
WW
fraction
0.1
0.2
0.3
0.4
-200
-100
0
100
200
position, bp

B

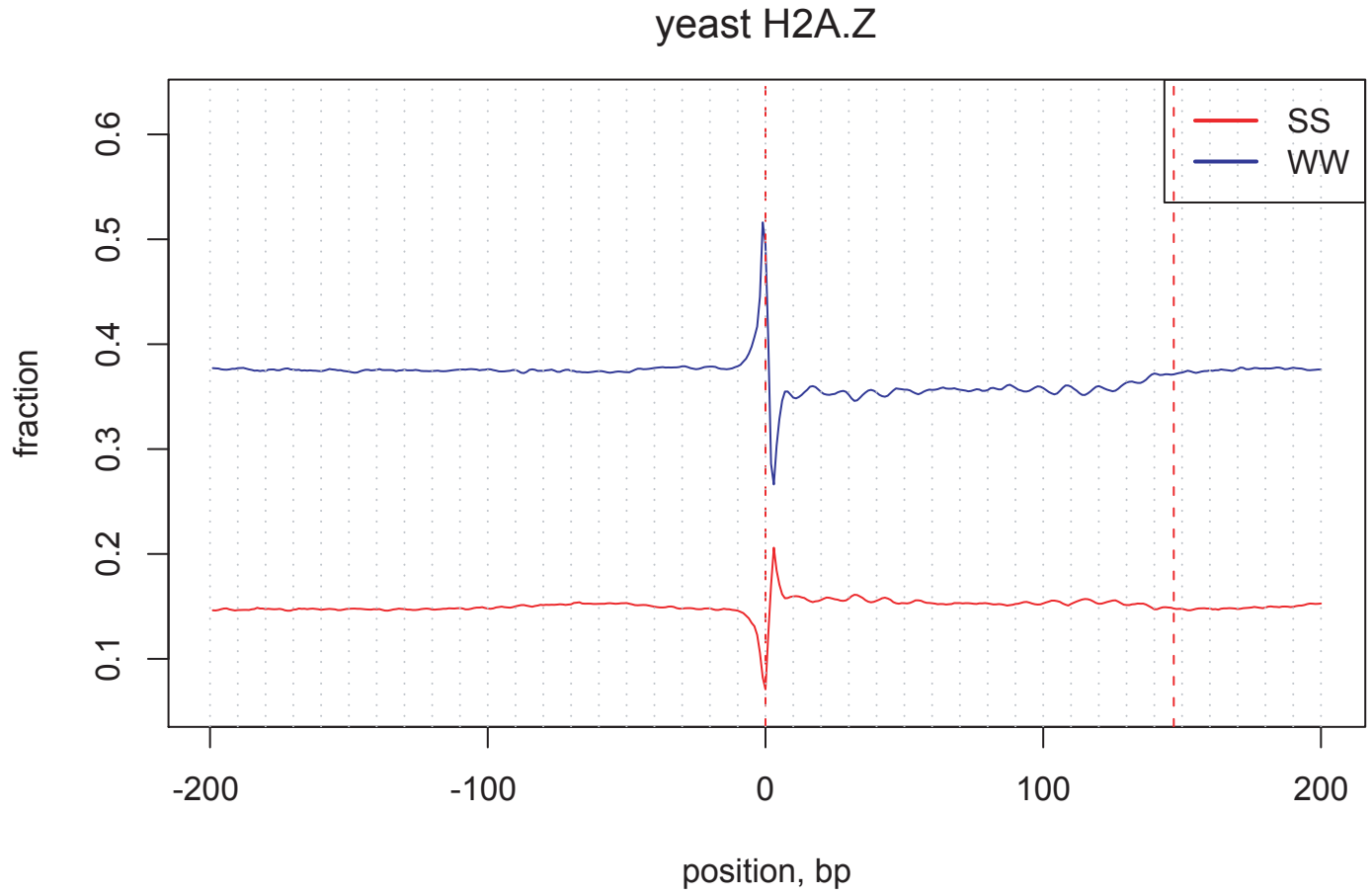
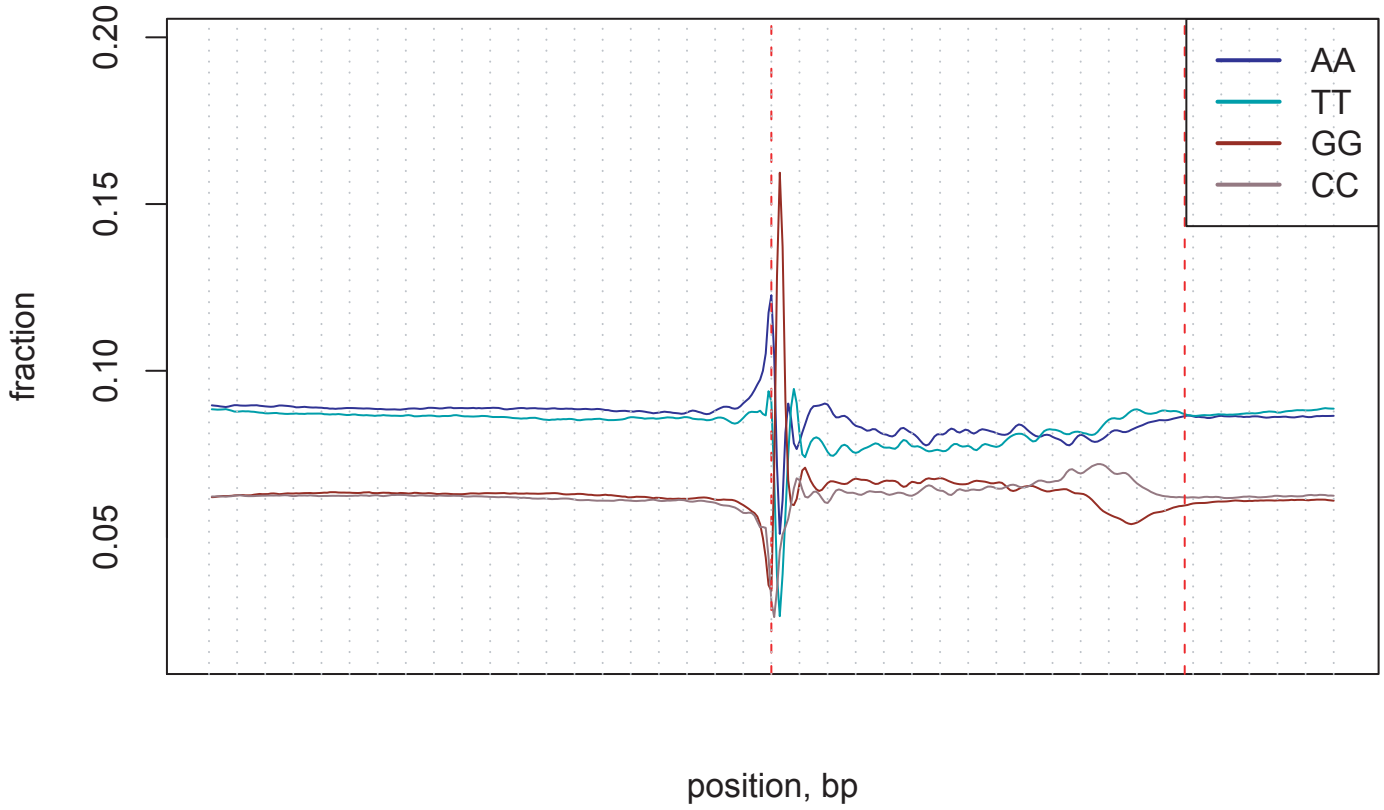

yeast H2A.Z
SS
WW
fraction
0.1
0.2
0.3
0.4
0.5
0.6
-200
-100
0
100
200
position, bp

Figure S7

C



D

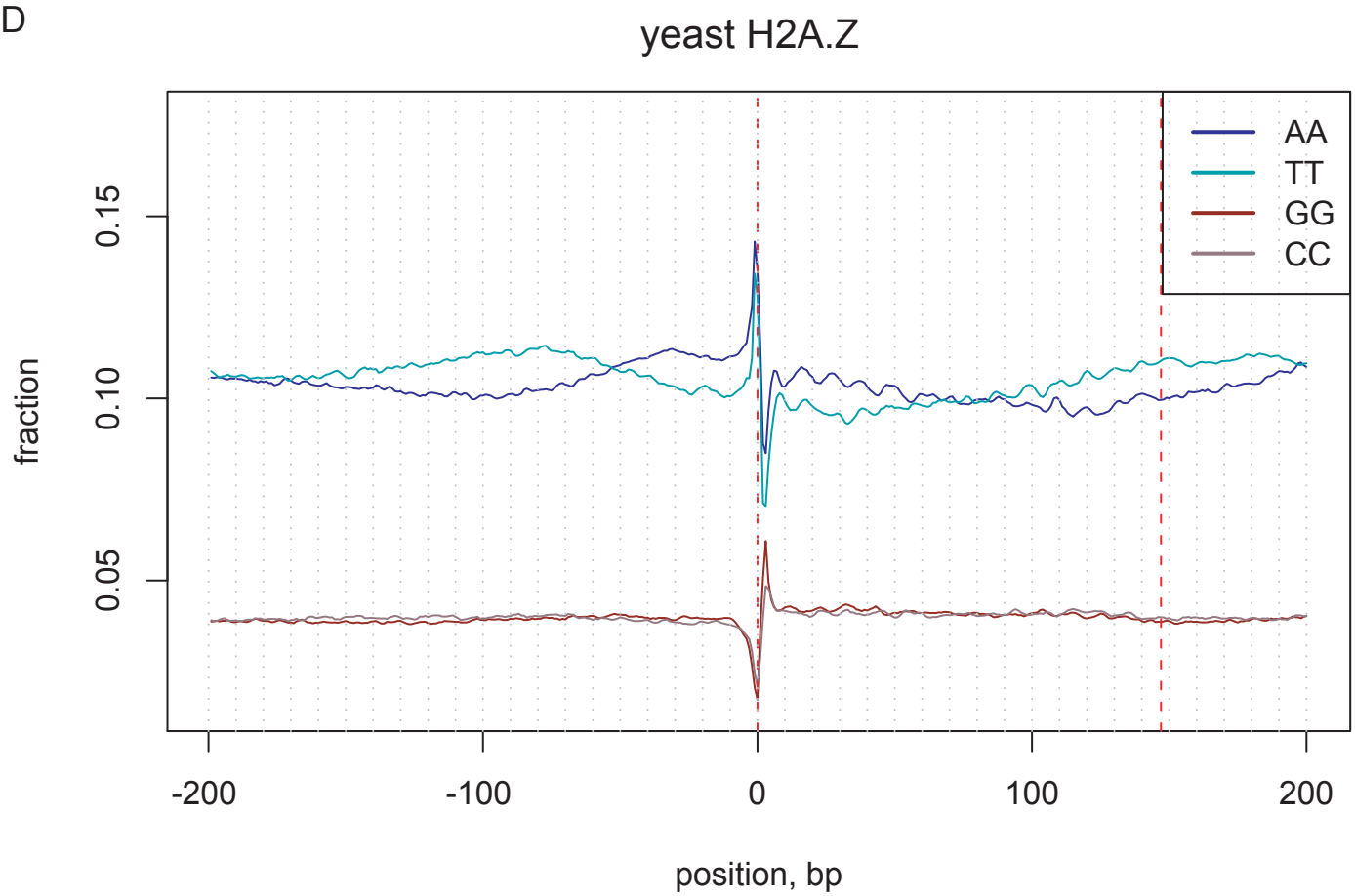
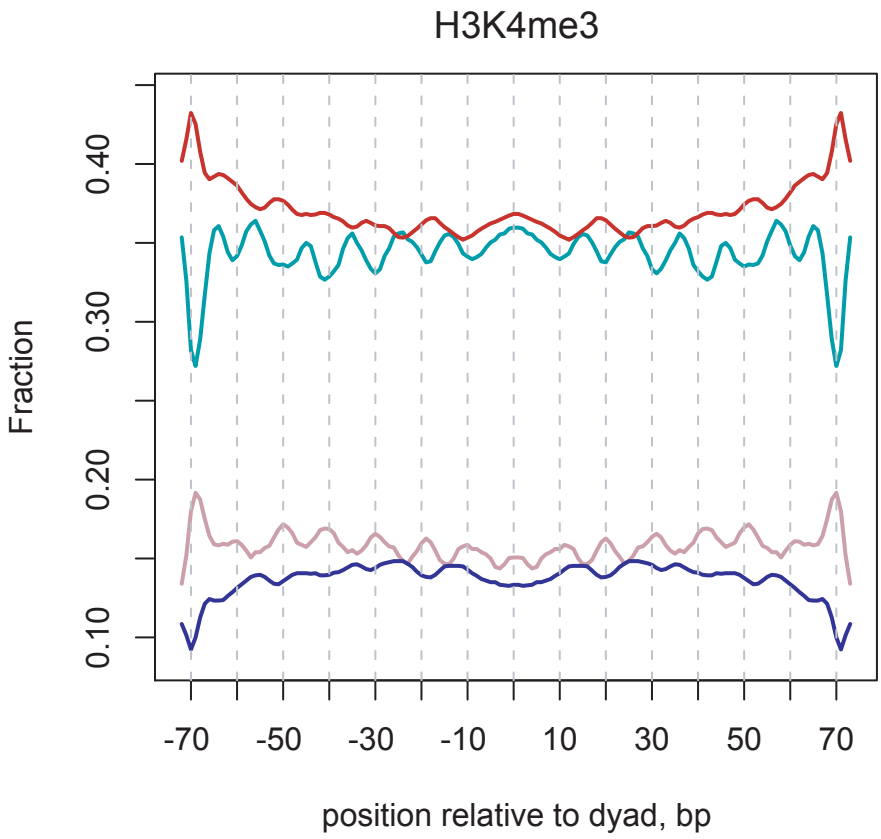


Figure S7 continued

A


H3K4me3
Fraction
0.10
0.20
0.30
0.40
-70
-50
-30
-10
10
30
50
70
position relative to dyad, bp

B

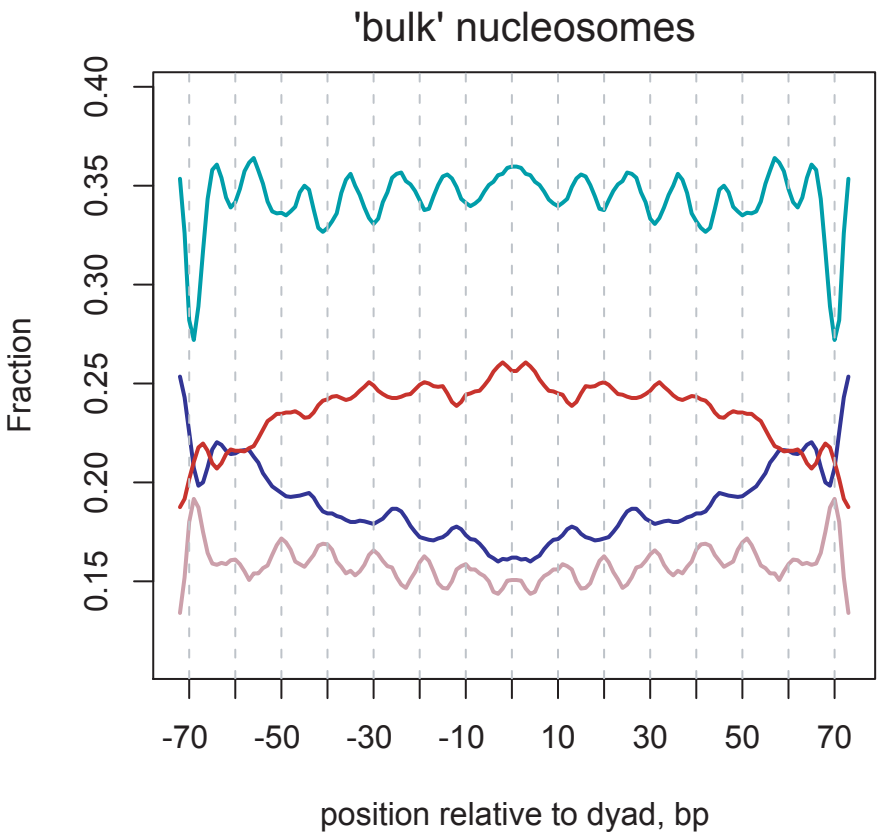
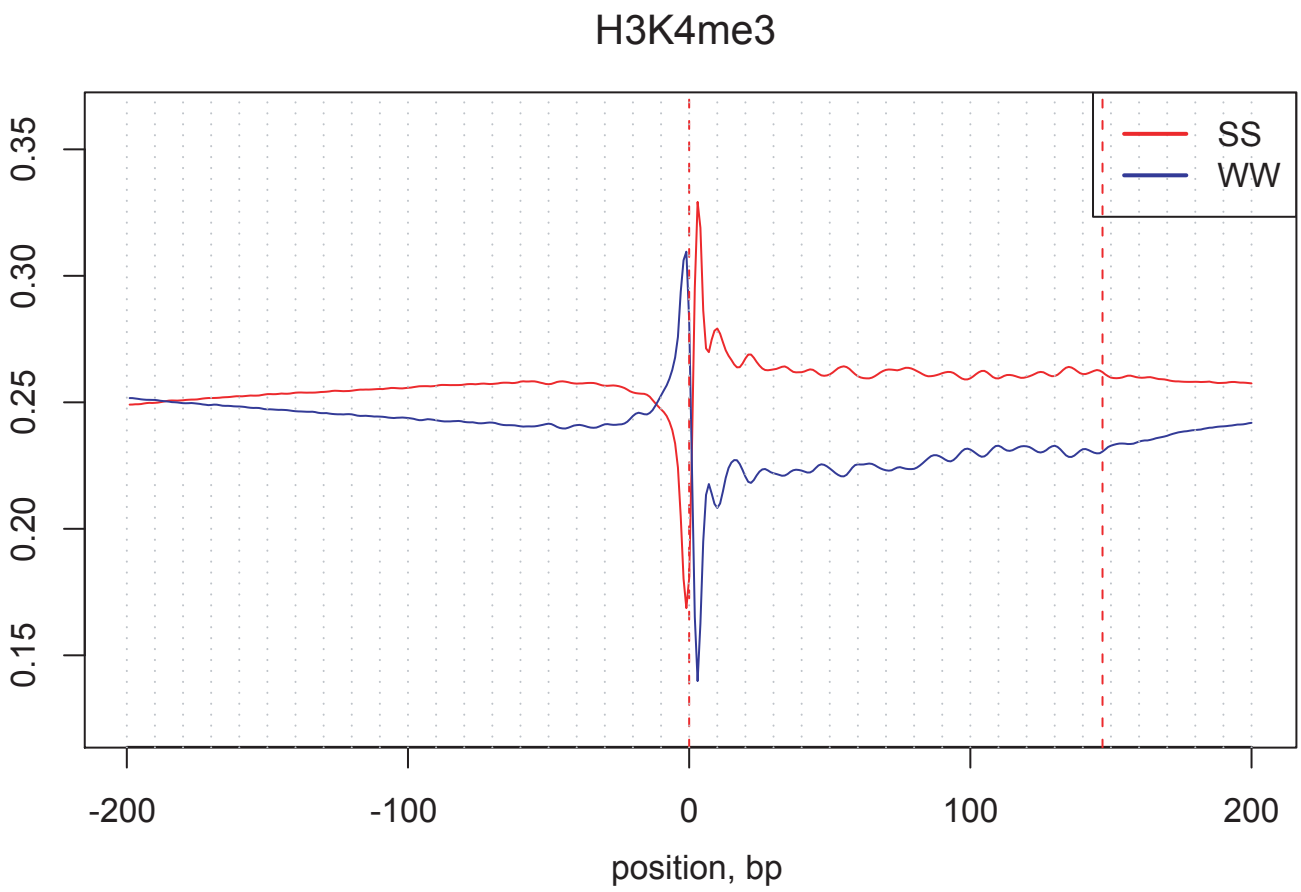

'bulk' nucleosomes
Fraction
0.15
0.20
0.25
0.30
0.35
0.40
-70
-50
-30
-10
10
30
50
70
position relative to dyad, bp

Figure S8

C



D

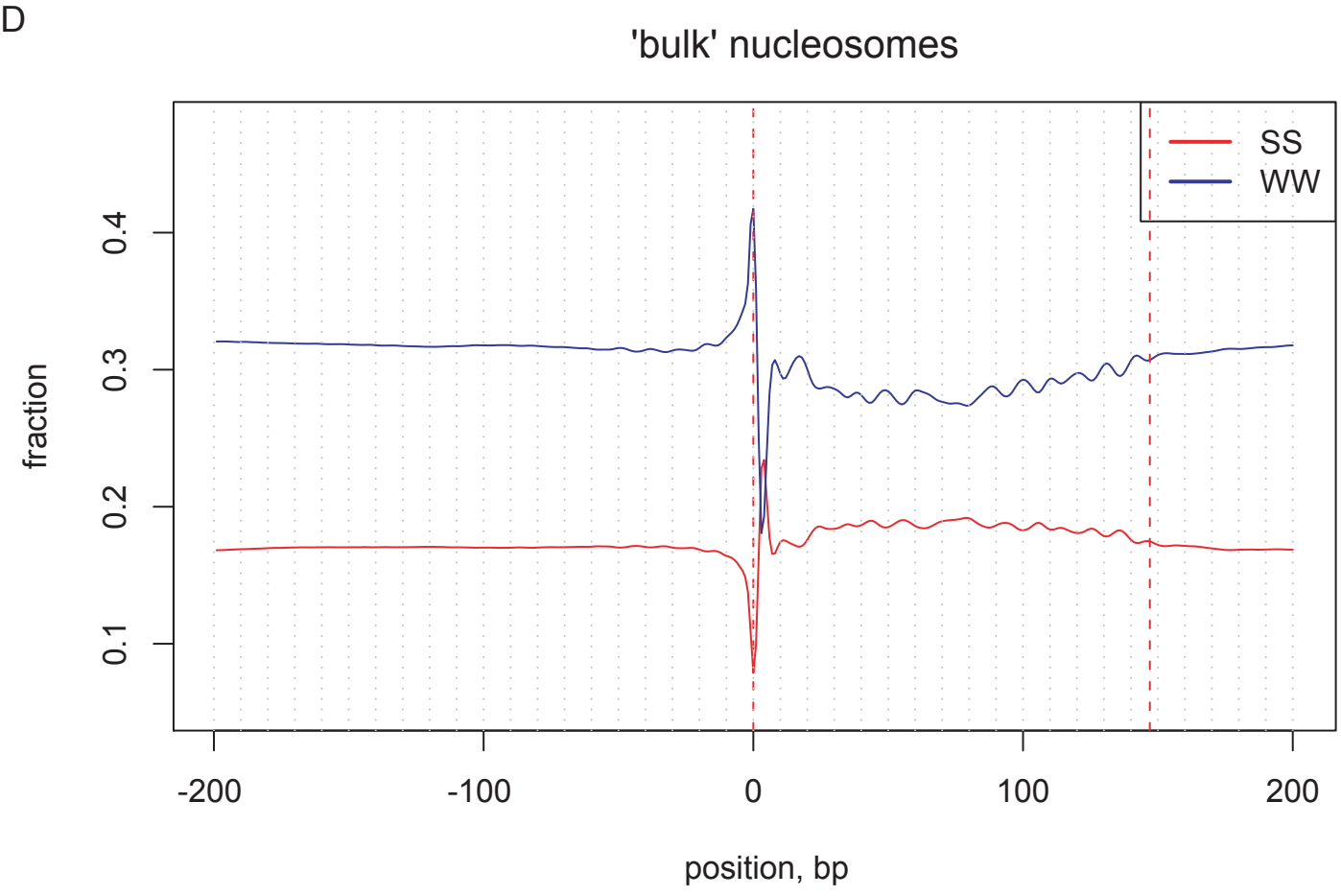
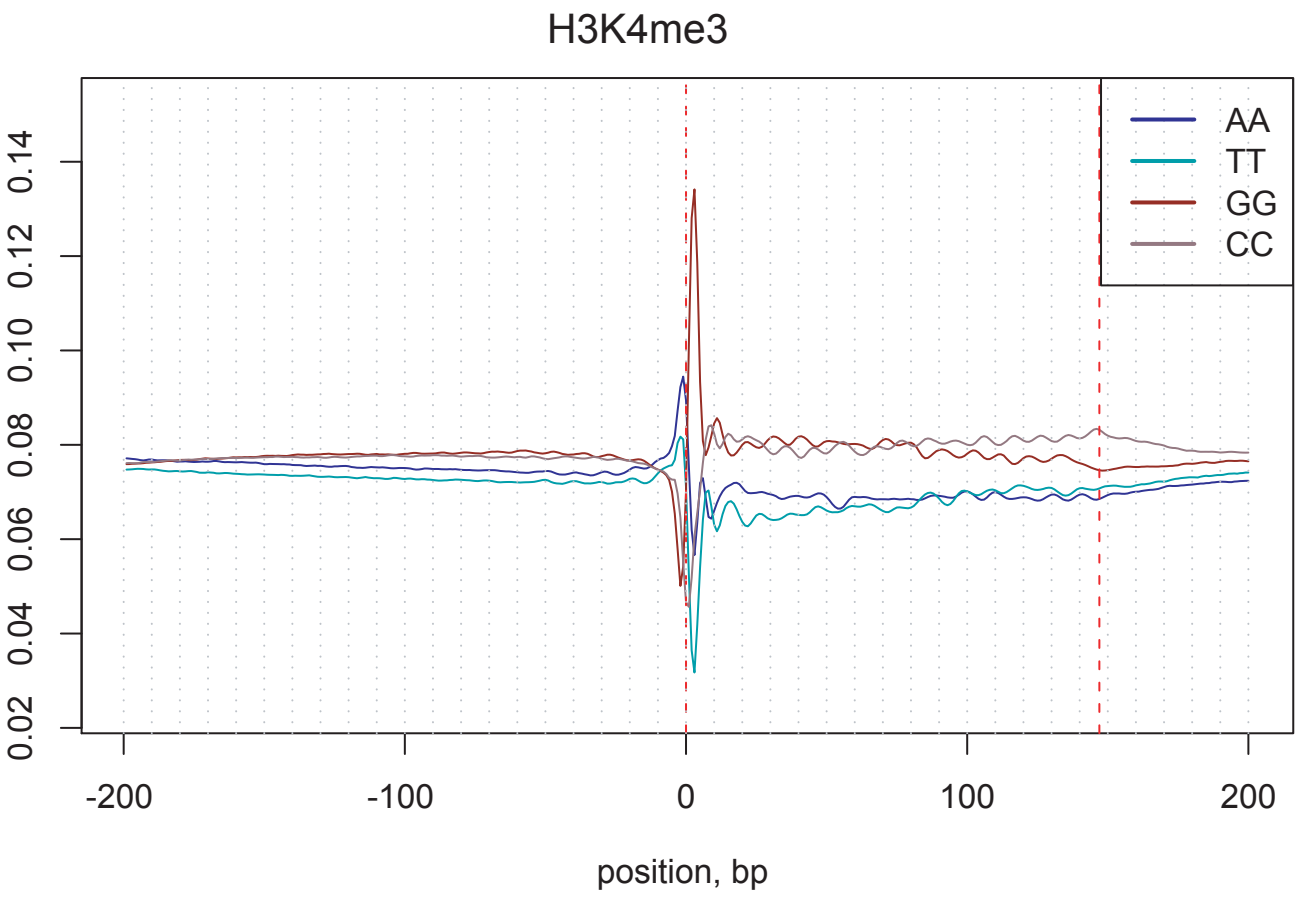


Figure S8 continued

E



F

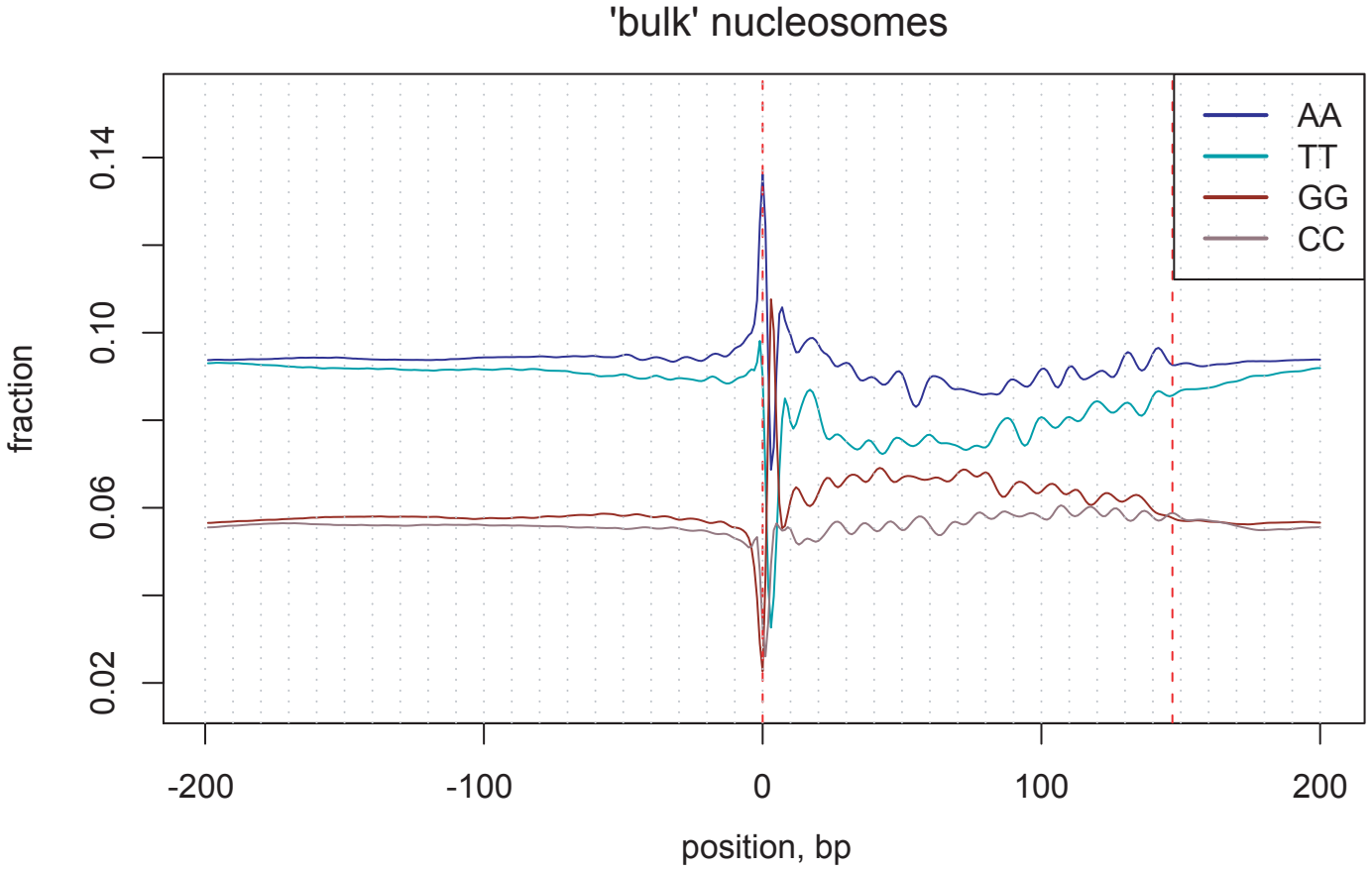


Figure S8 continued

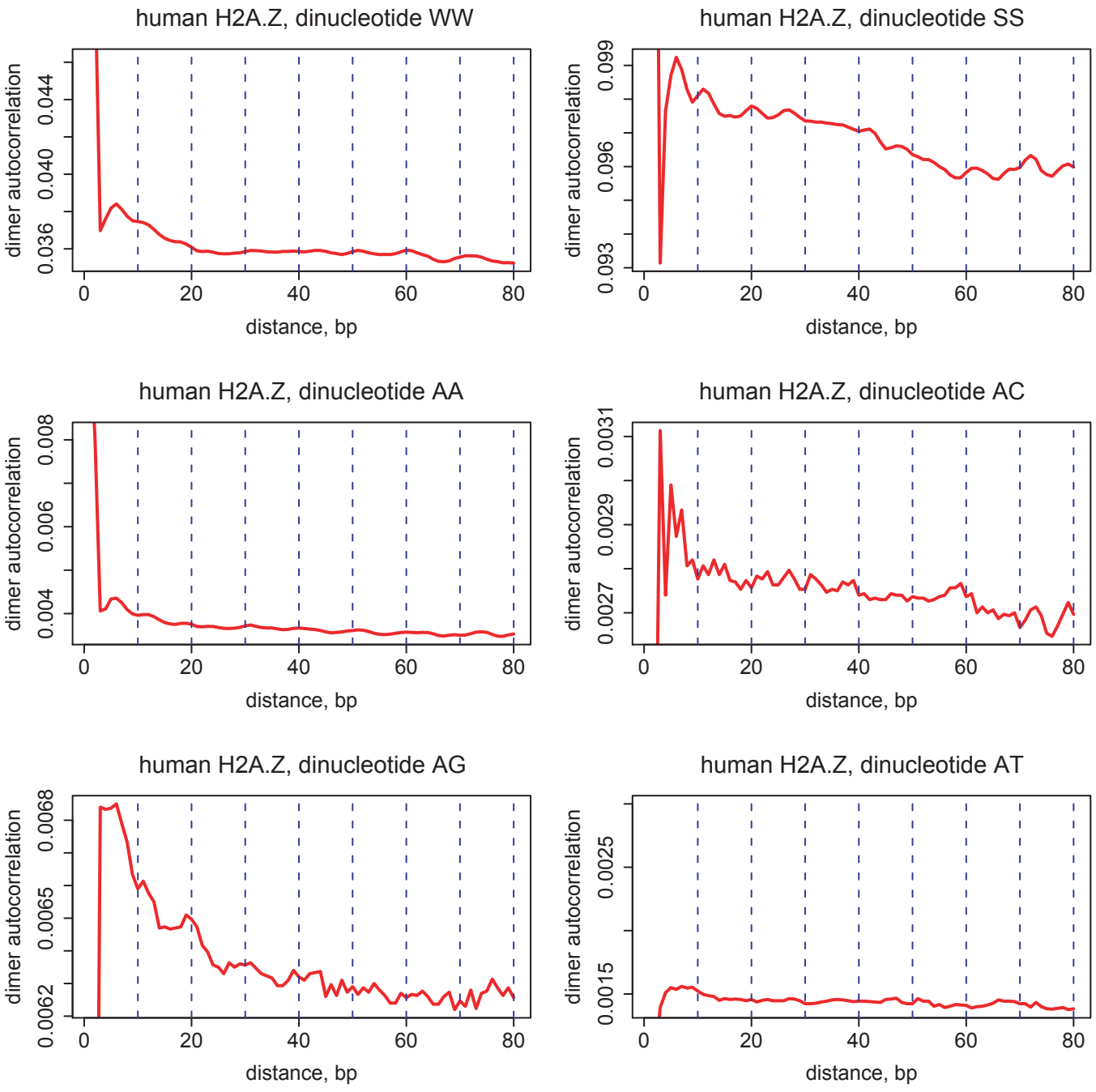

human H2A.Z, dinucleotide WW
dimer autocorrelation
0.036
0.040
0.044
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide SS
dimer autocorrelation
0.093
0.096
0.099
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide AA
dimer autocorrelation
0.004
0.006
0.008
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide AC
dimer autocorrelation
0.0027
0.0029
0.0031
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide AG
dimer autocorrelation
0.0062
0.0065
0.0068
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide AT
dimer autocorrelation
0.0015
0.0025
0
20
40
60
80
distance, bp

Figure S9

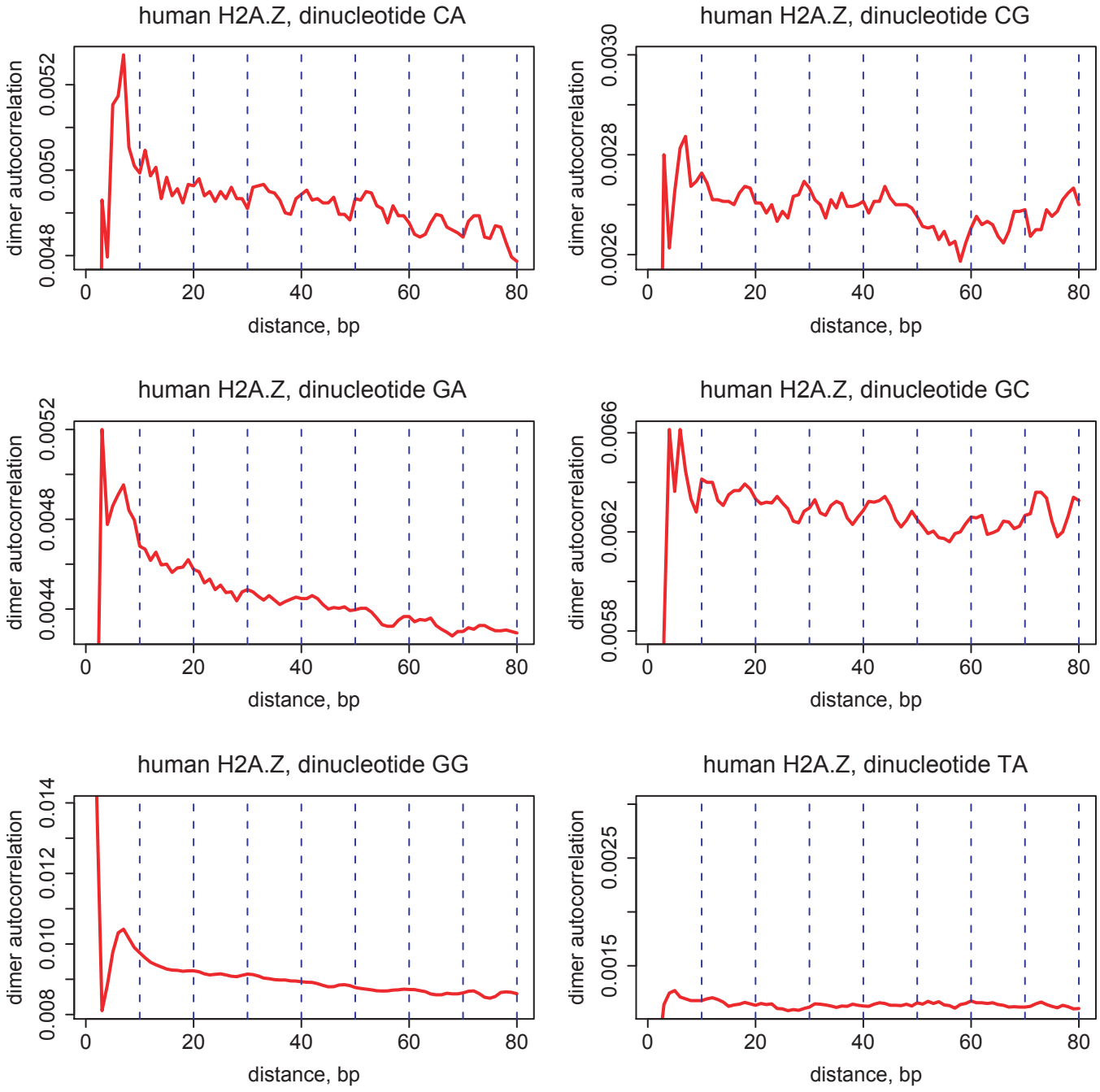


Figure S9 continued

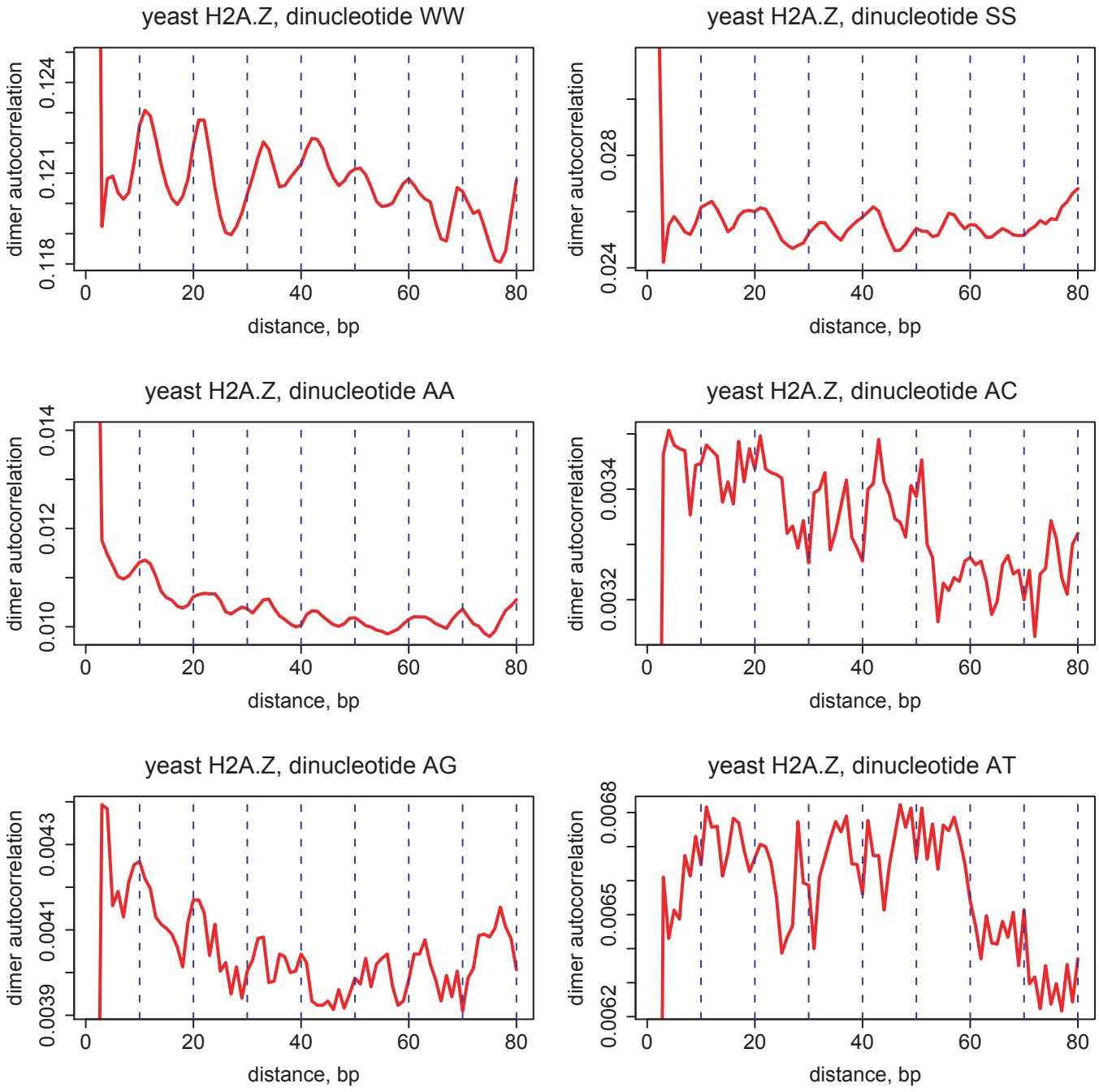


Figure S9 continued

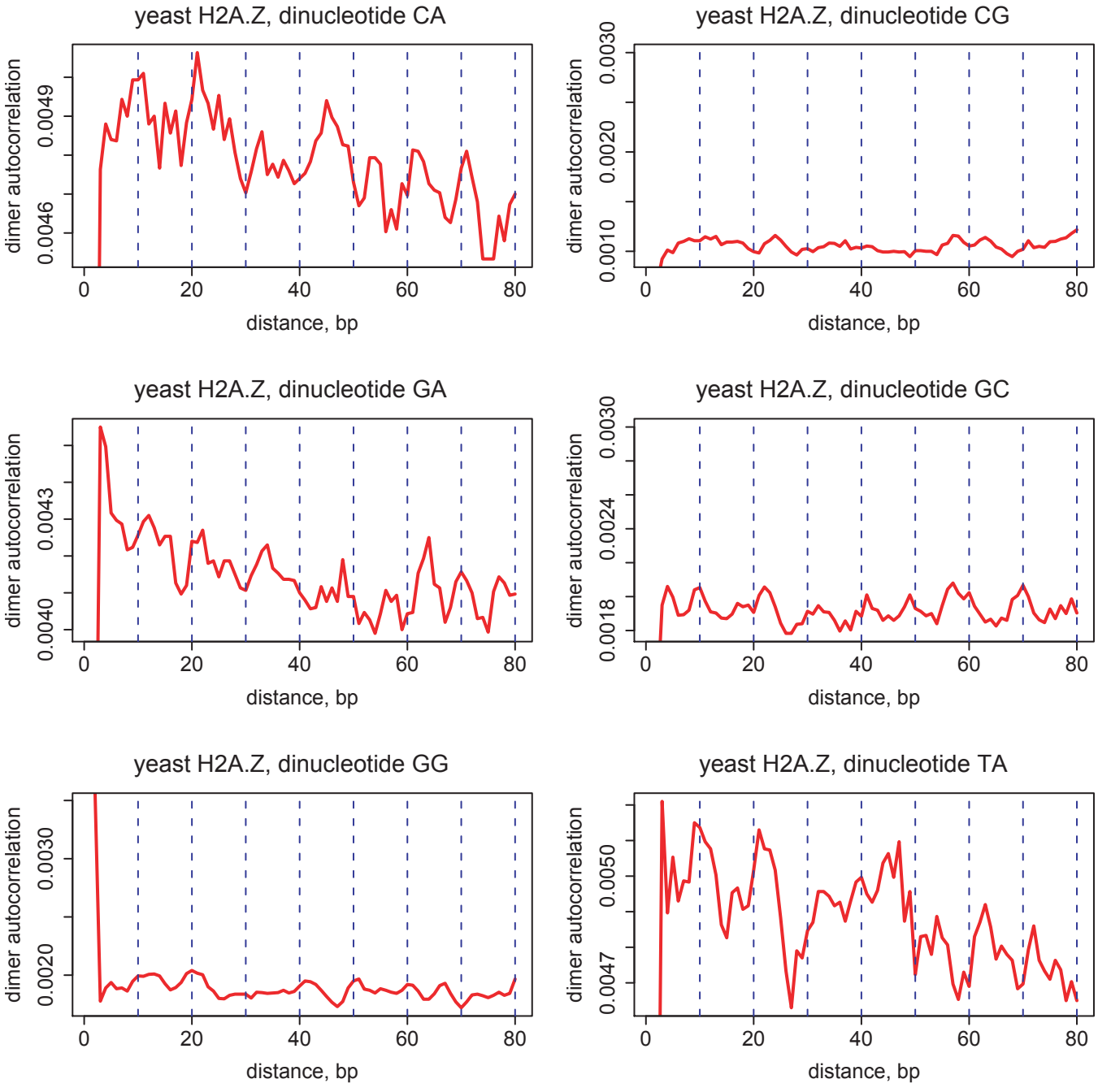


Figure S9 continued

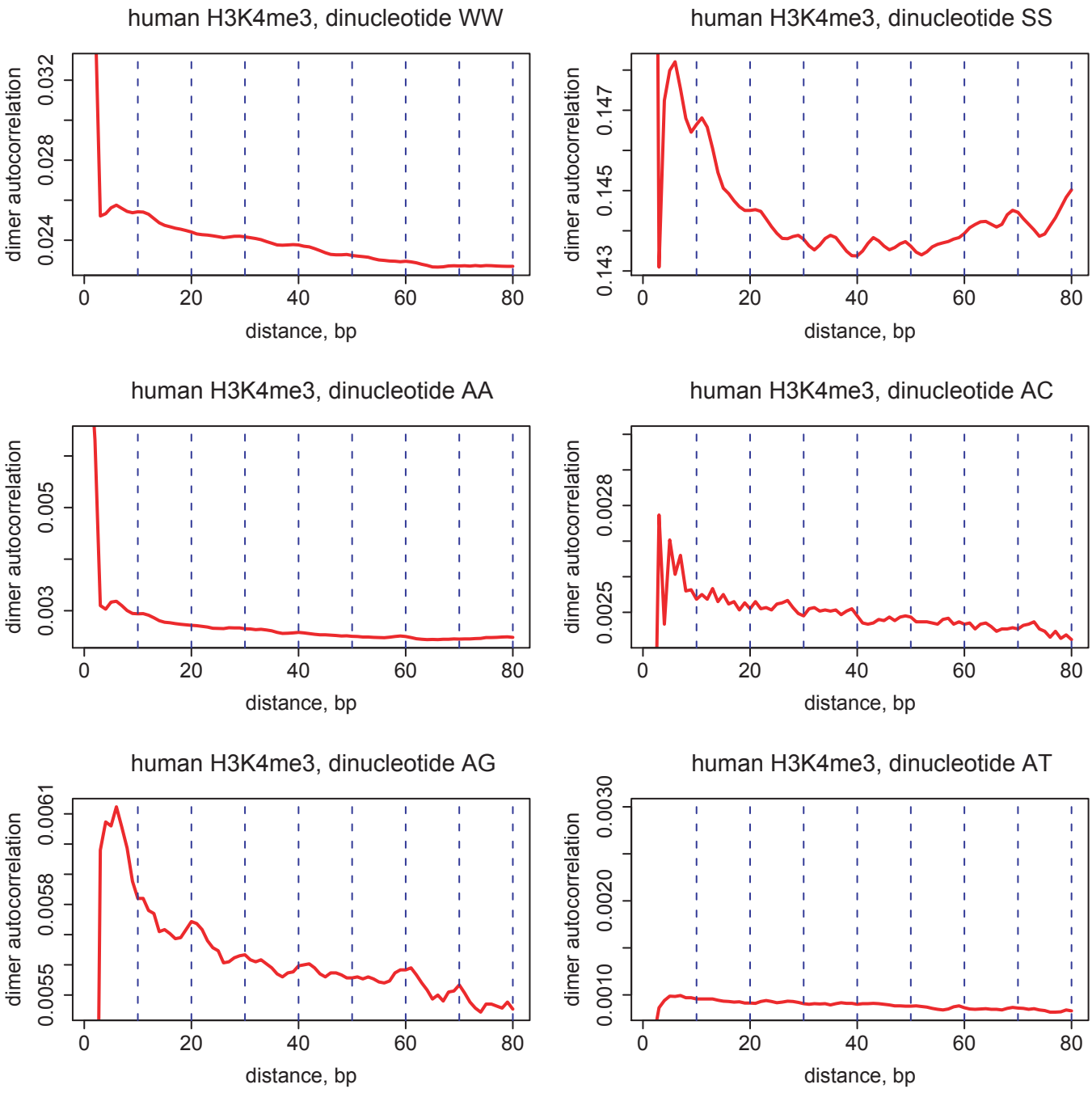

human H3K4me3, dinucleotide WW
dimer autocorrelation
0.024
0.028
0.032
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide SS
dimer autocorrelation
0.143
0.145
0.147
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide AA
dimer autocorrelation
0.003
0.005
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide AC
dimer autocorrelation
0.0025
0.0028
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide AG
dimer autocorrelation
0.0055
0.0058
0.0061
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide AT
dimer autocorrelation
0.0010
0.0020
0.0030
0
20
40
60
80
distance, bp

Figure S10

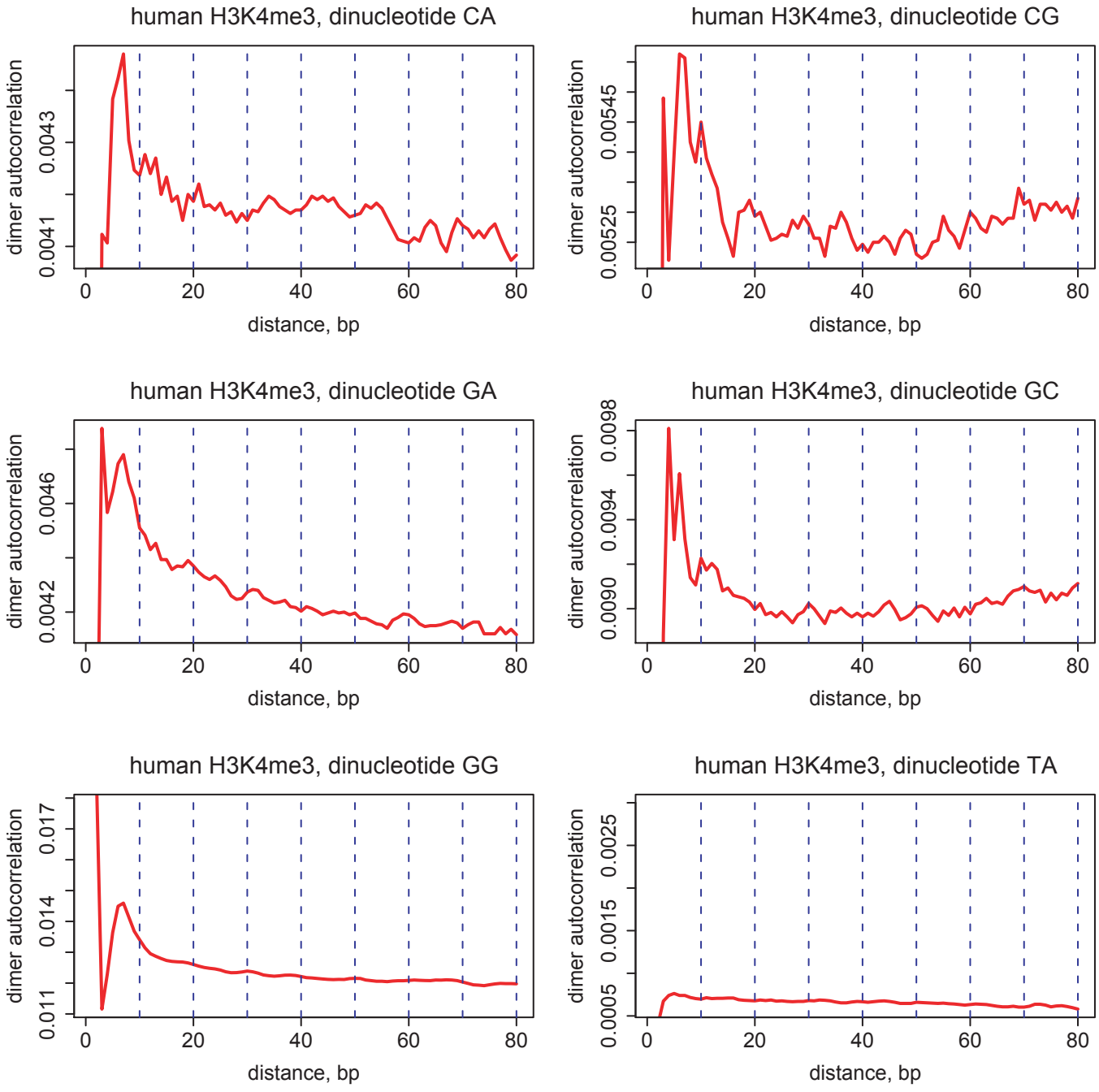


Figure S10 continued

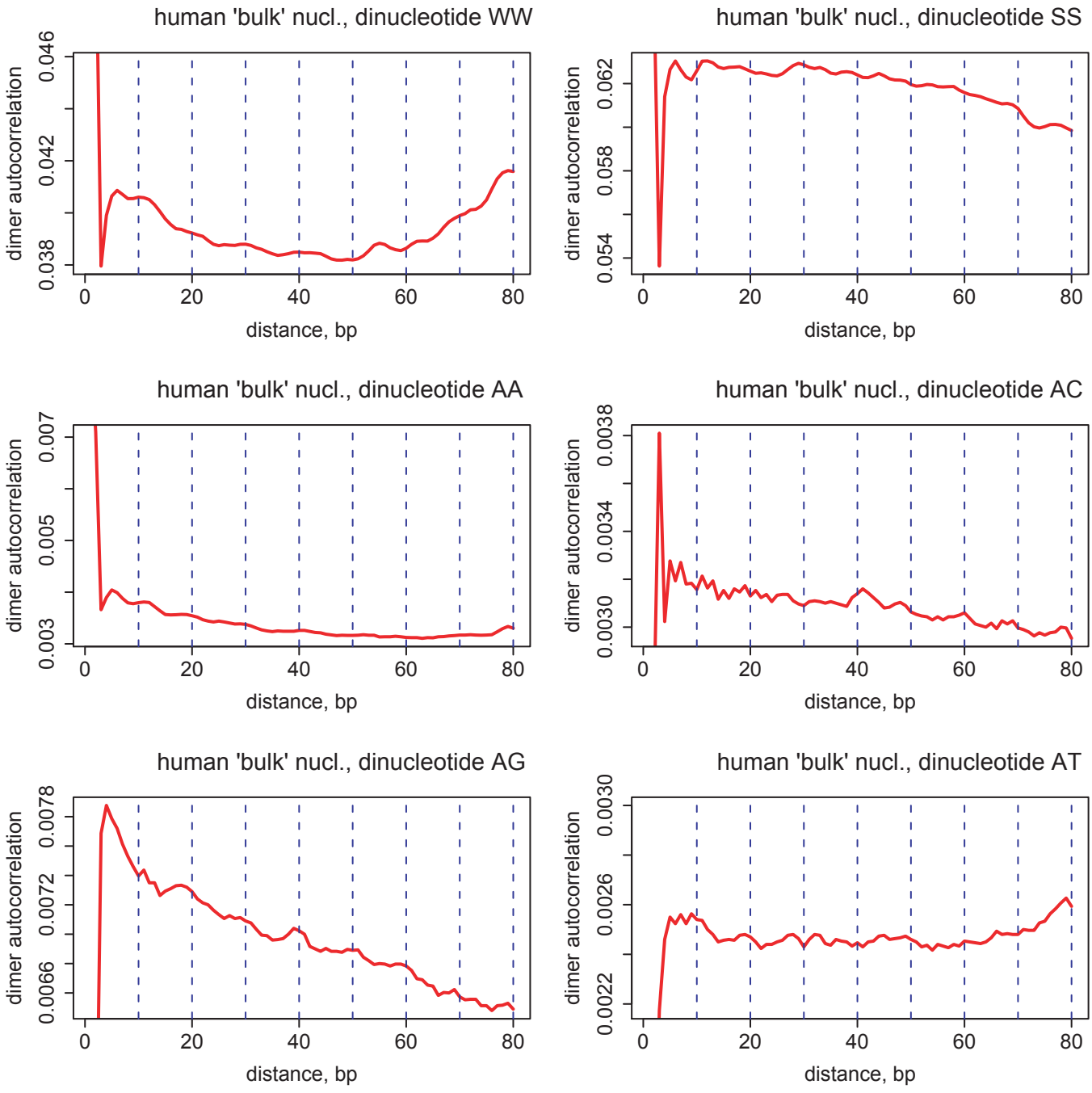

human 'bulk' nucl., dinucleotide WW
dimer autocorrelation
0.038
0.042
0.046
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide SS
dimer autocorrelation
0.054
0.058
0.062
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide AA
dimer autocorrelation
0.003
0.005
0.007
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide AC
dimer autocorrelation
0.0030
0.0034
0.0038
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide AG
dimer autocorrelation
0.0066
0.0072
0.0078
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide AT
dimer autocorrelation
0.0022
0.0026
0.0030
0
20
40
60
80
distance, bp

Figure S11

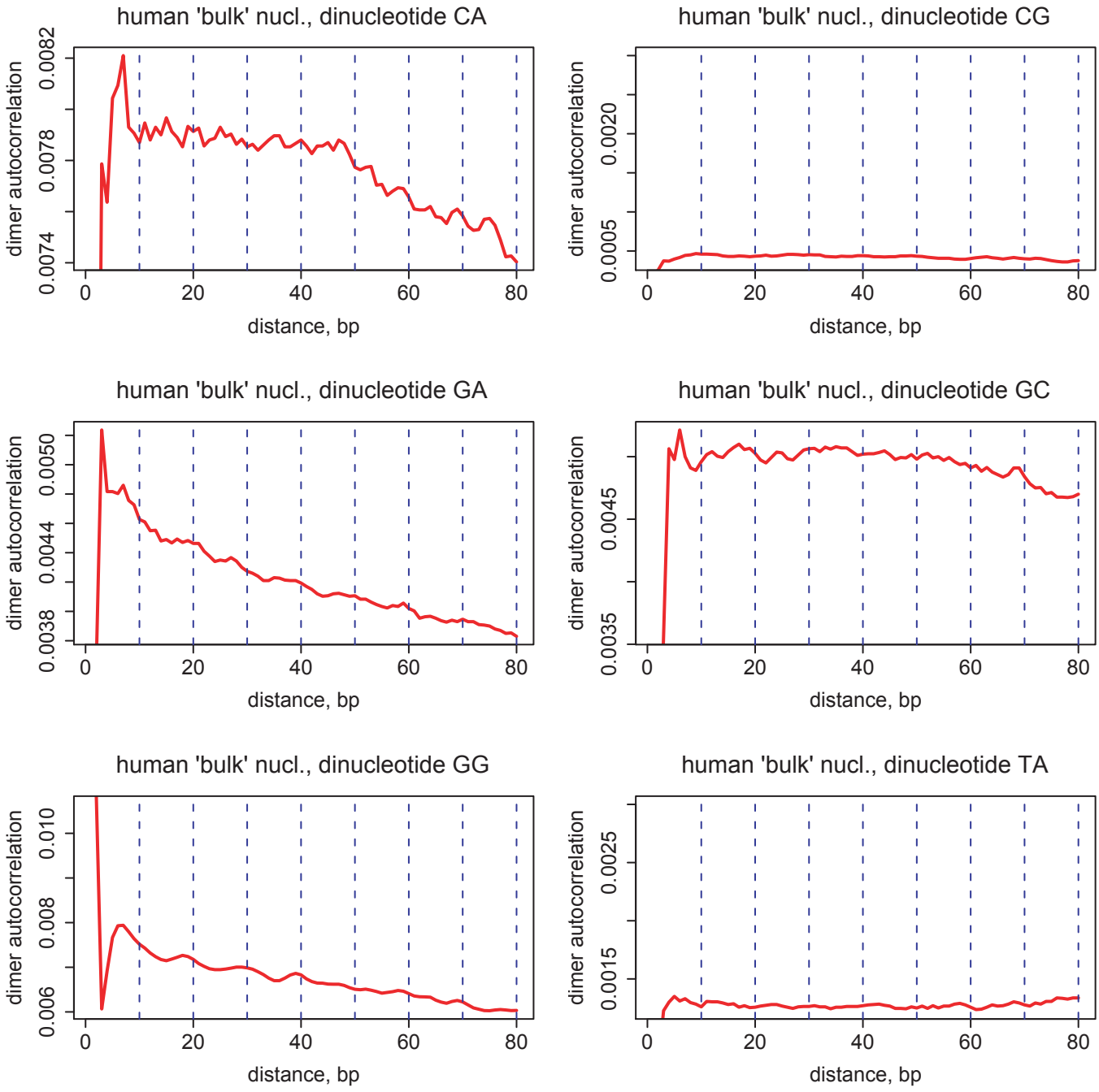


Figure S11 continued

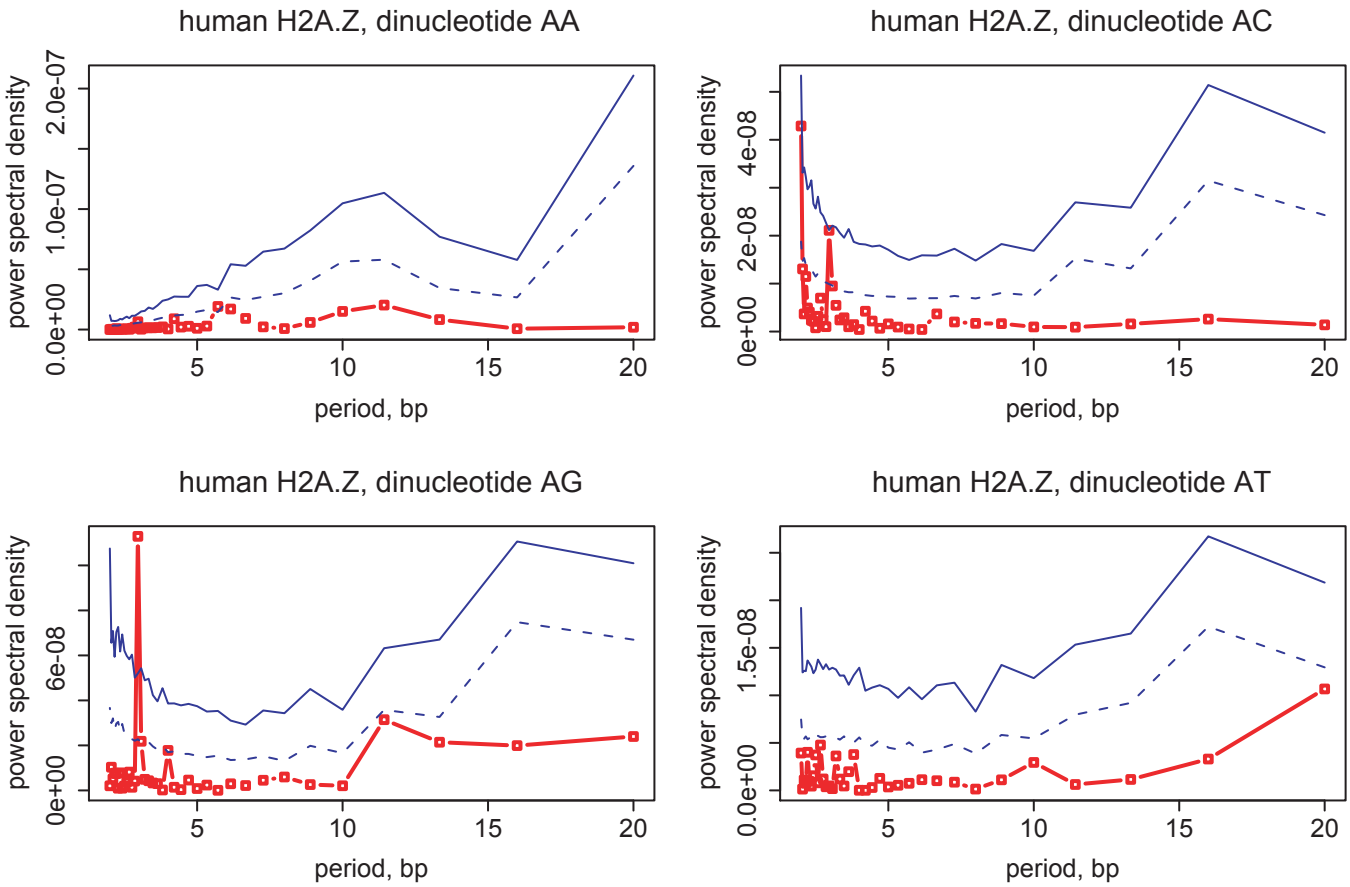

human H2A.Z, dinucleotide AA
power spectral density
0.0e+00
1.0e-07
2.0e-07
5
10
15
20
period, bp
human H2A.Z, dinucleotide AC
power spectral density
0e+00
2e-08
4e-08
5
10
15
20
period, bp
human H2A.Z, dinucleotide AG
power spectral density
0e+00
6e-08
5
10
15
20
period, bp
human H2A.Z, dinucleotide AT
power spectral density
0.0e+00
1.5e-08
5
10
15
20
period, bp

Figure S12

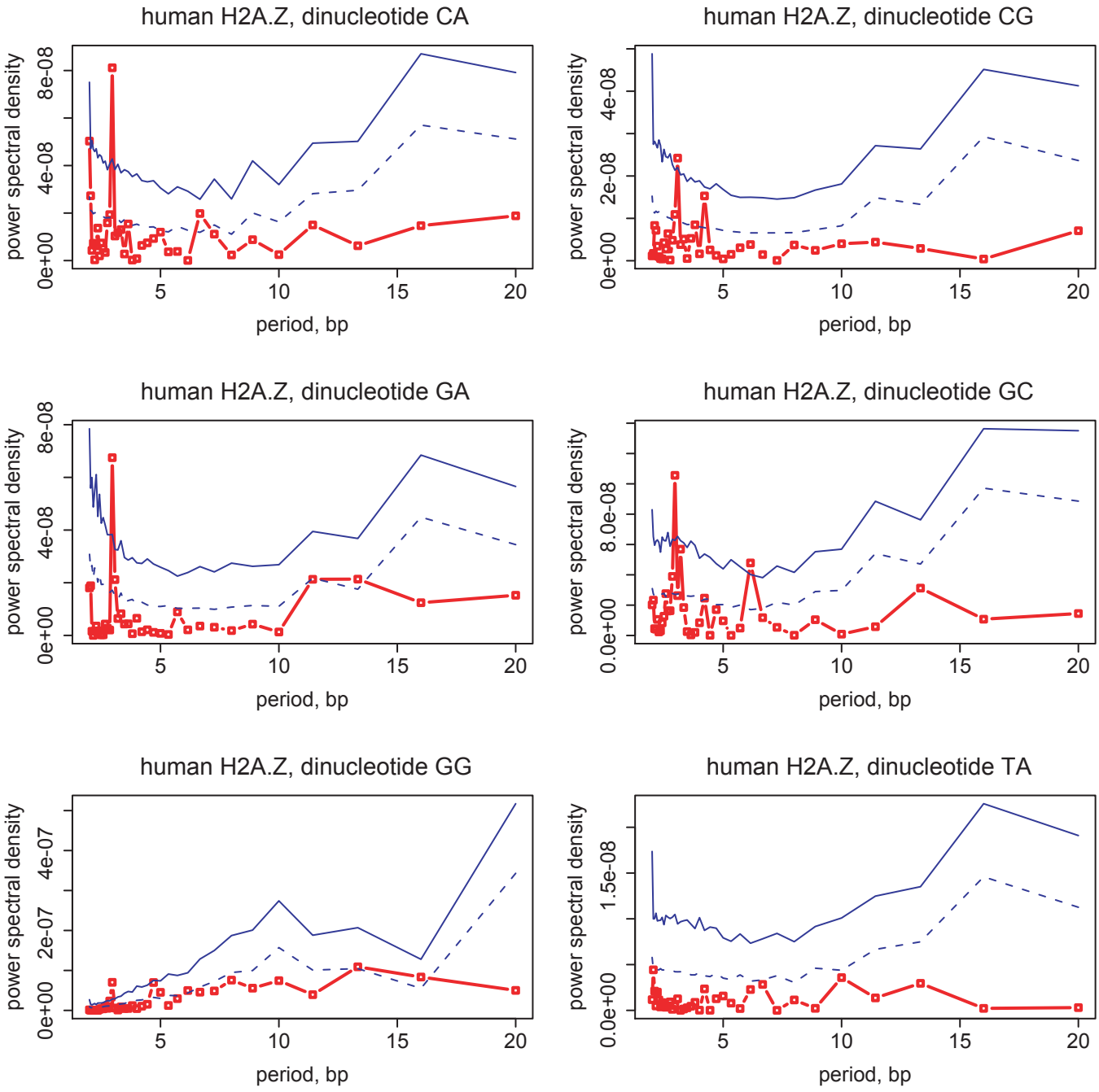


Figure S12 continued

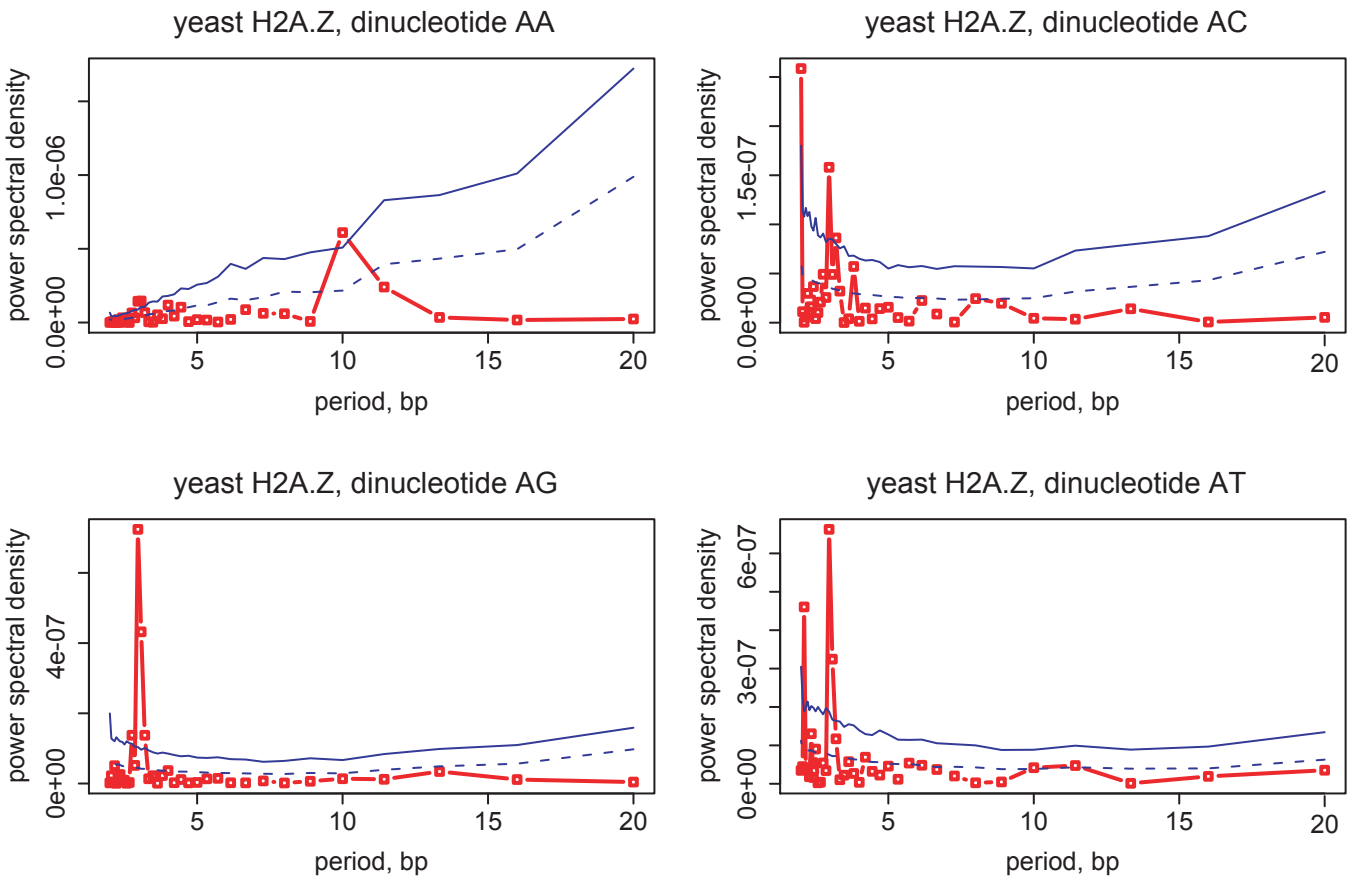


Figure S12 continued

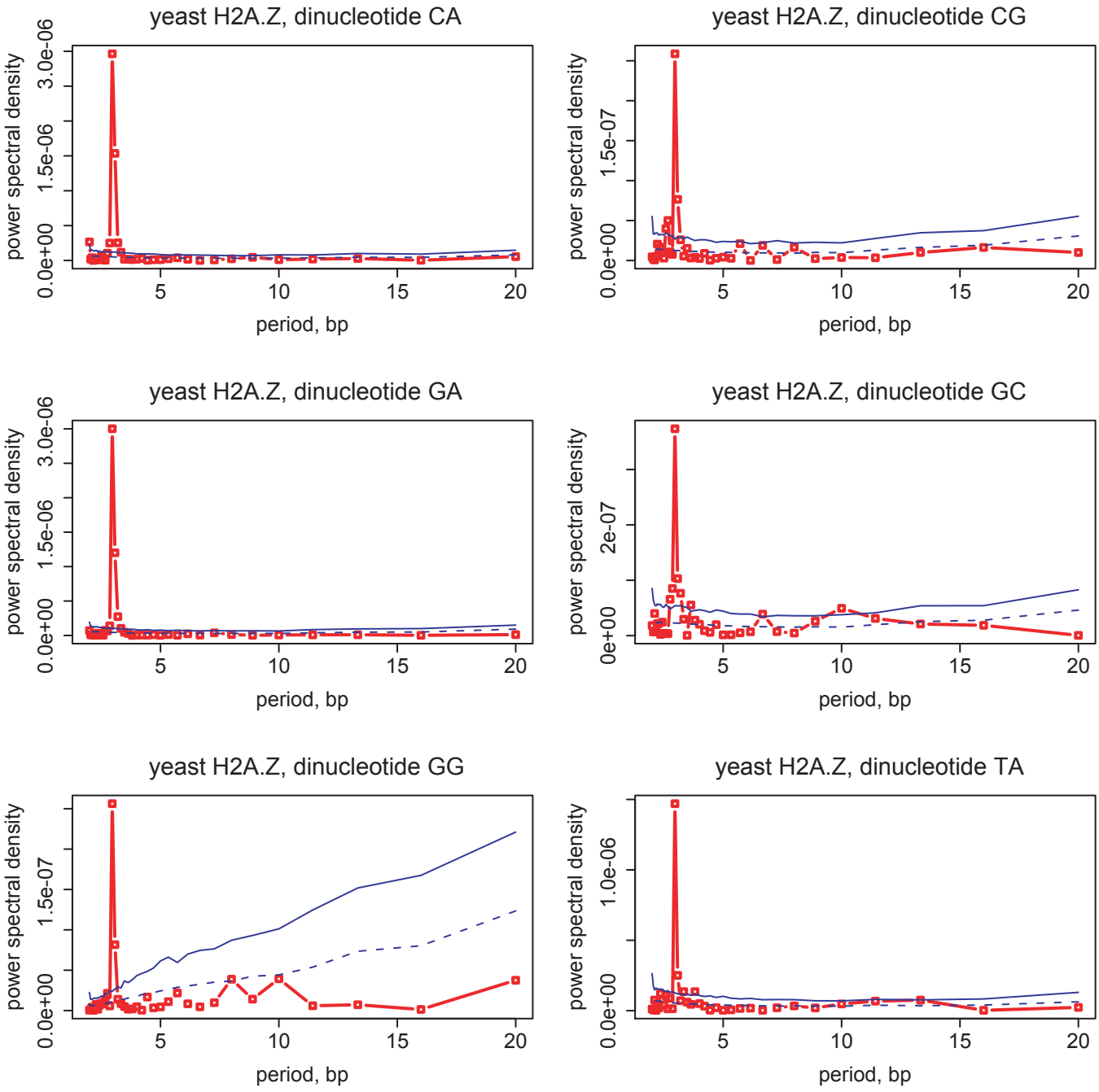


Figure S12 continued

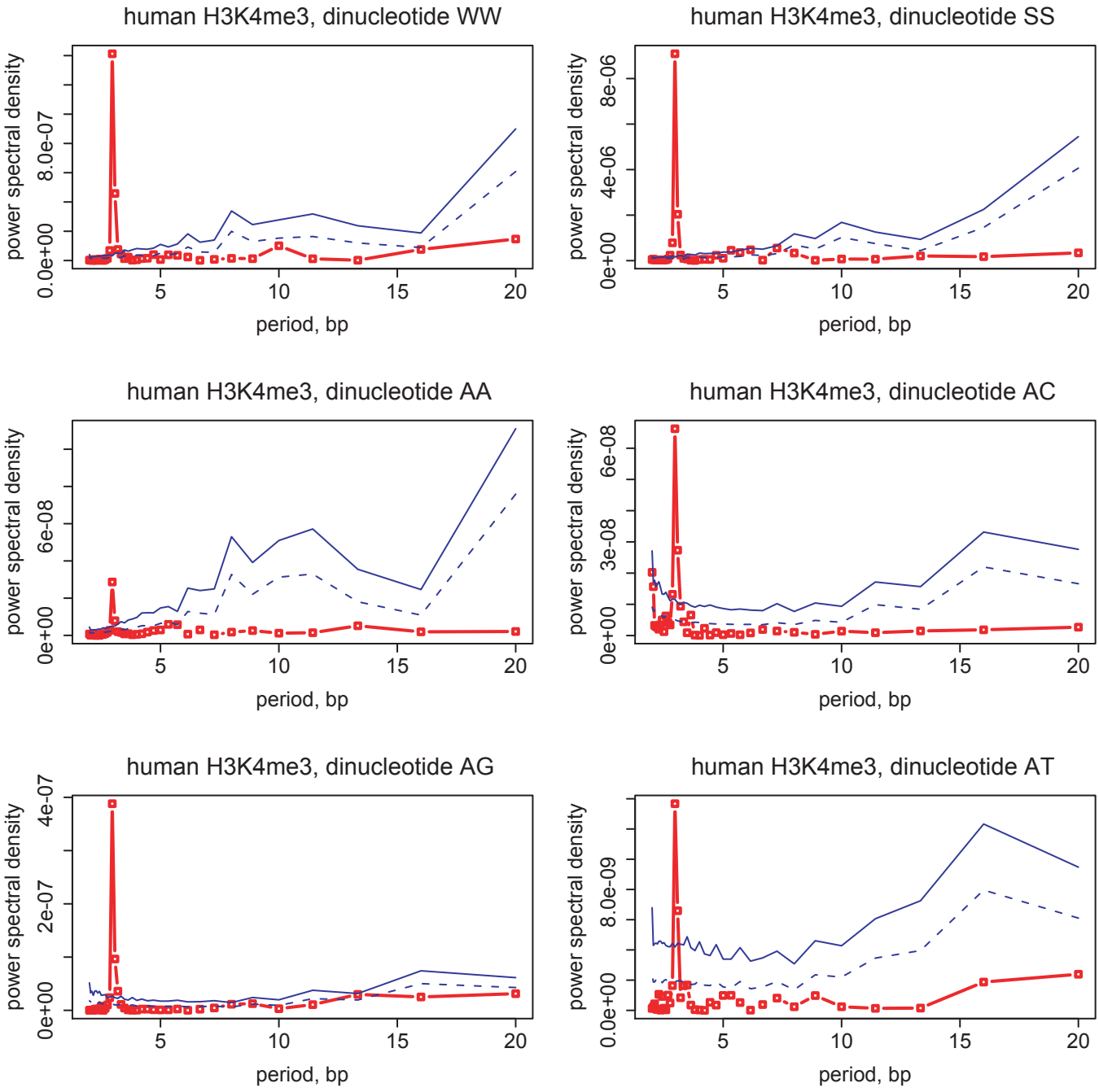

human H3K4me3, dinucleotide WW
power spectral density
0.0e+00
8.0e-07
5
10
15
20
period, bp
human H3K4me3, dinucleotide SS
power spectral density
0e+00
4e-06
8e-06
5
10
15
20
period, bp
human H3K4me3, dinucleotide AA
power spectral density
0e+00
6e-08
5
10
15
20
period, bp
human H3K4me3, dinucleotide AC
power spectral density
0e+00
3e-08
6e-08
5
10
15
20
period, bp
human H3K4me3, dinucleotide AG
power spectral density
0e+00
2e-07
4e-07
5
10
15
20
period, bp
human H3K4me3, dinucleotide AT
power spectral density
0.0e+00
8.0e-09
5
10
15
20
period, bp

Figure S13

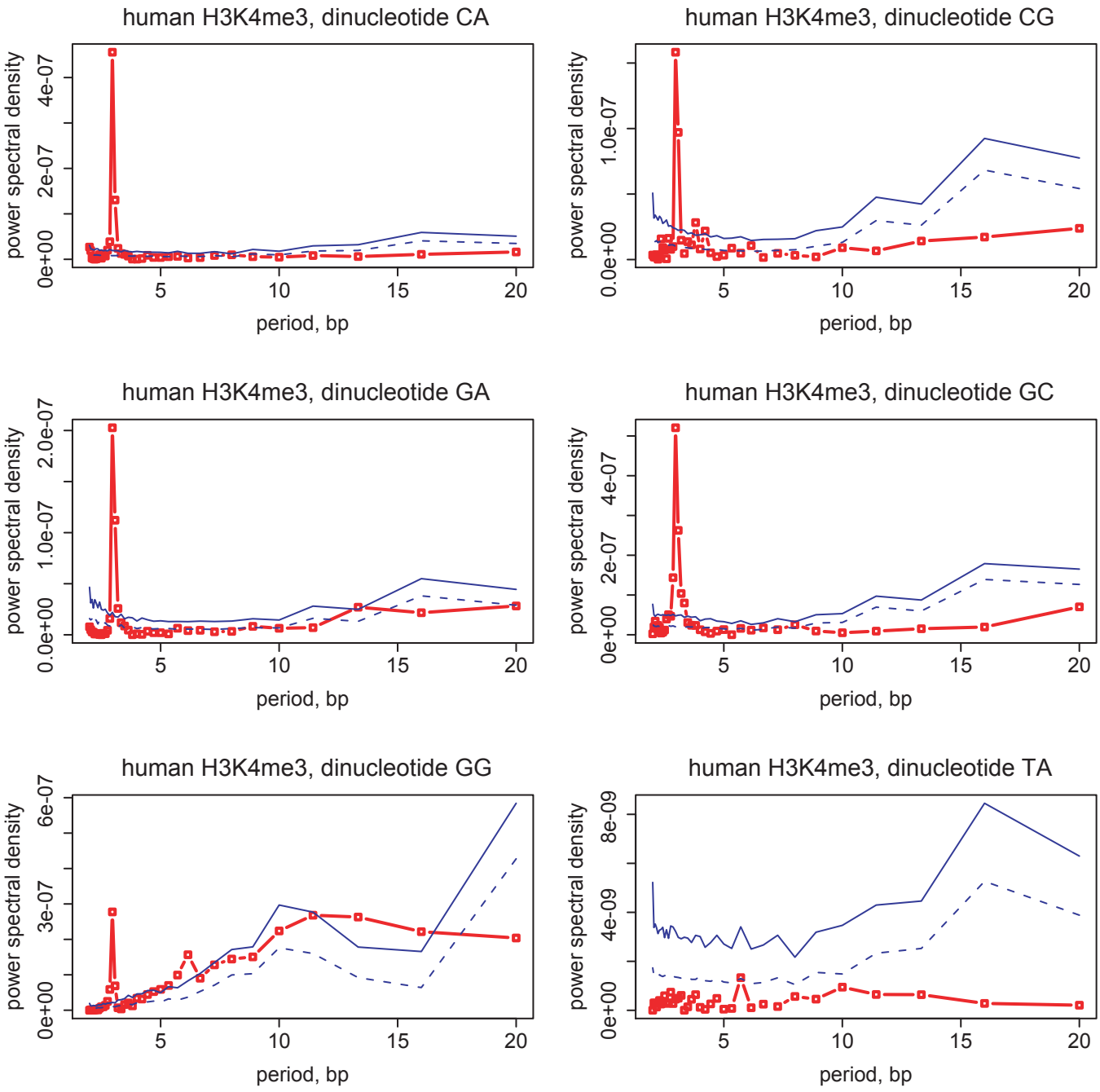


Figure S13 continued

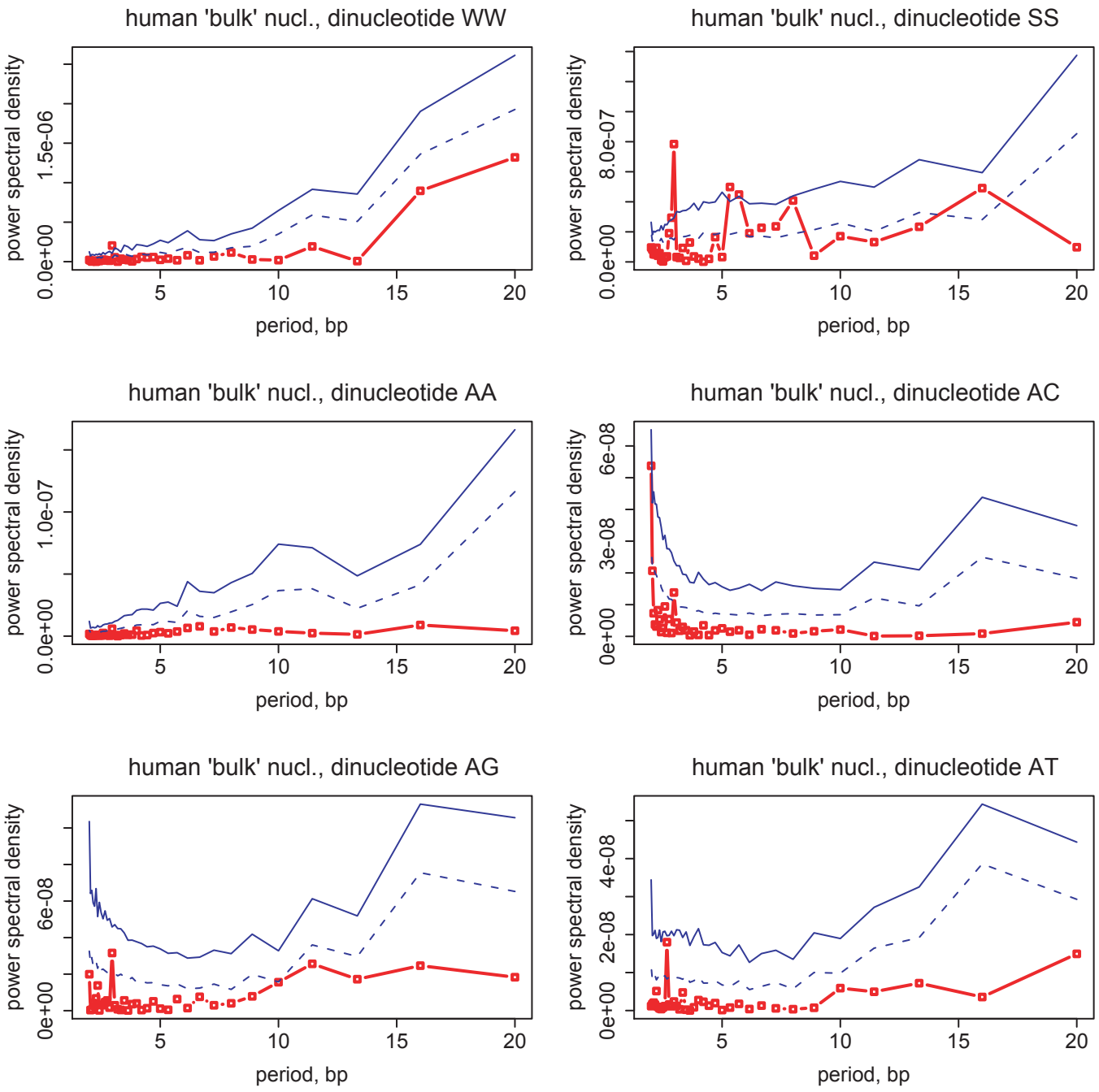

human 'bulk' nucl., dinucleotide WW
power spectral density
0.0e+00
1.5e-06
5
10
15
20
period, bp
human 'bulk' nucl., dinucleotide SS
power spectral density
0.0e+00
8.0e-07
5
10
15
20
period, bp
human 'bulk' nucl., dinucleotide AA
power spectral density
0.0e+00
1.0e-07
5
10
15
20
period, bp
human 'bulk' nucl., dinucleotide AC
power spectral density
0e+00
3e-08
6e-08
5
10
15
20
period, bp
human 'bulk' nucl., dinucleotide AG
power spectral density
0e+00
6e-08
5
10
15
20
period, bp
human 'bulk' nucl., dinucleotide AT
power spectral density
0e+00
2e-08
4e-08
5
10
15
20
period, bp

Figure S14

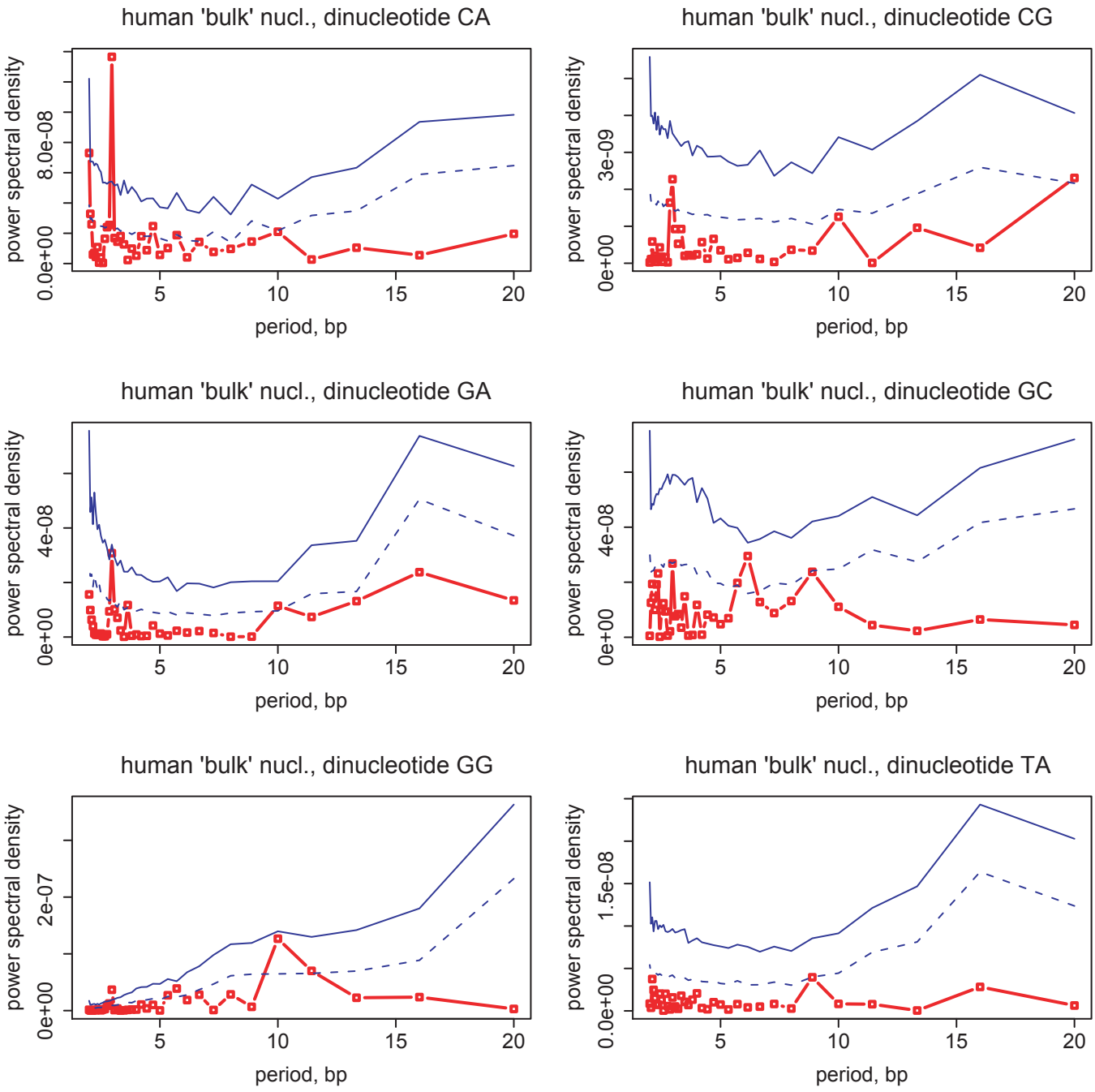


Figure S14 continued

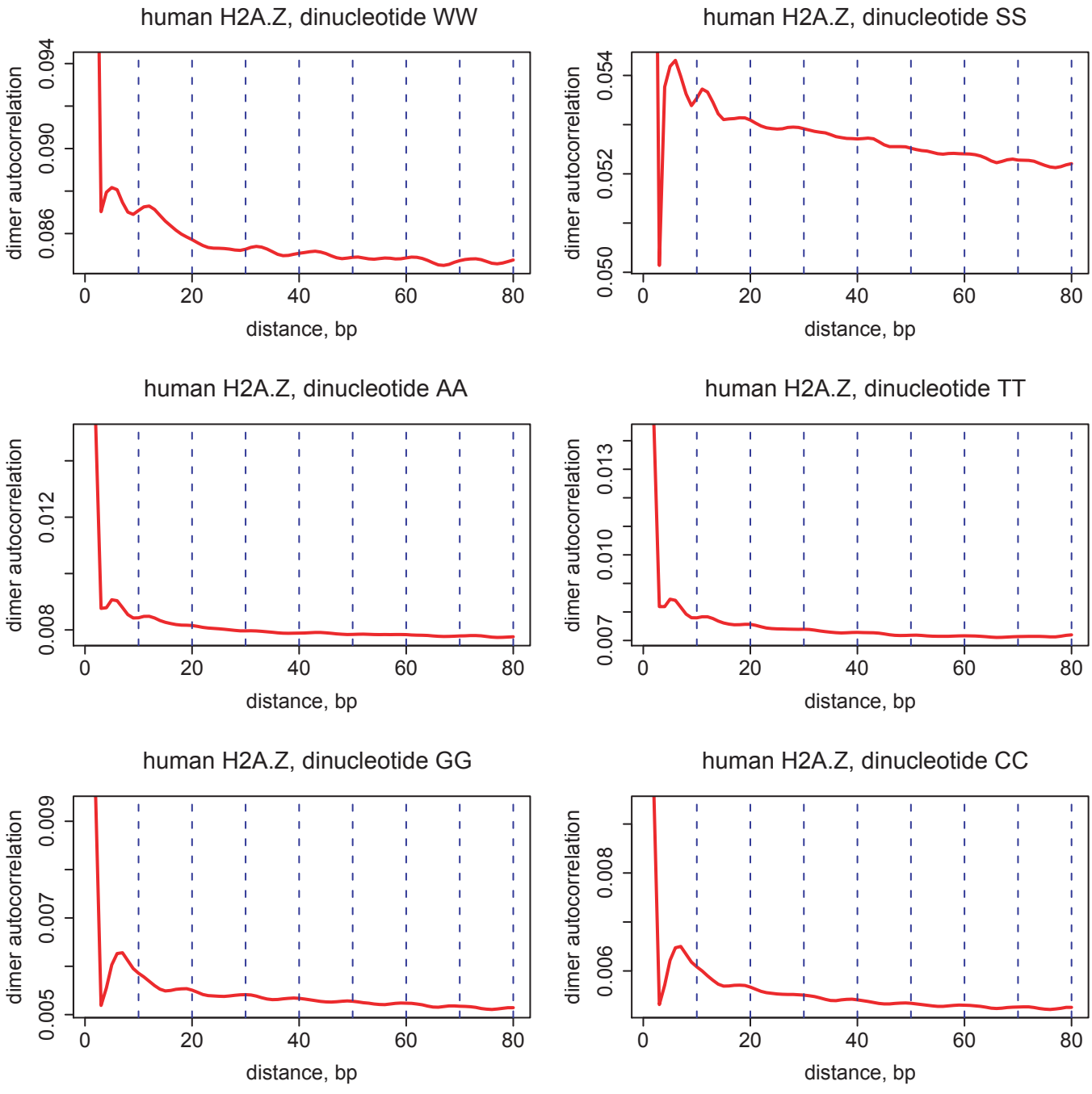

human H2A.Z, dinucleotide WW
dimer autocorrelation
0.086
0.090
0.094
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide SS
dimer autocorrelation
0.050
0.052
0.054
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide AA
dimer autocorrelation
0.008
0.012
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide TT
dimer autocorrelation
0.007
0.010
0.013
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide GG
dimer autocorrelation
0.005
0.007
0.009
0
20
40
60
80
distance, bp
human H2A.Z, dinucleotide CC
dimer autocorrelation
0.006
0.008
0
20
40
60
80
distance, bp

Figure S15

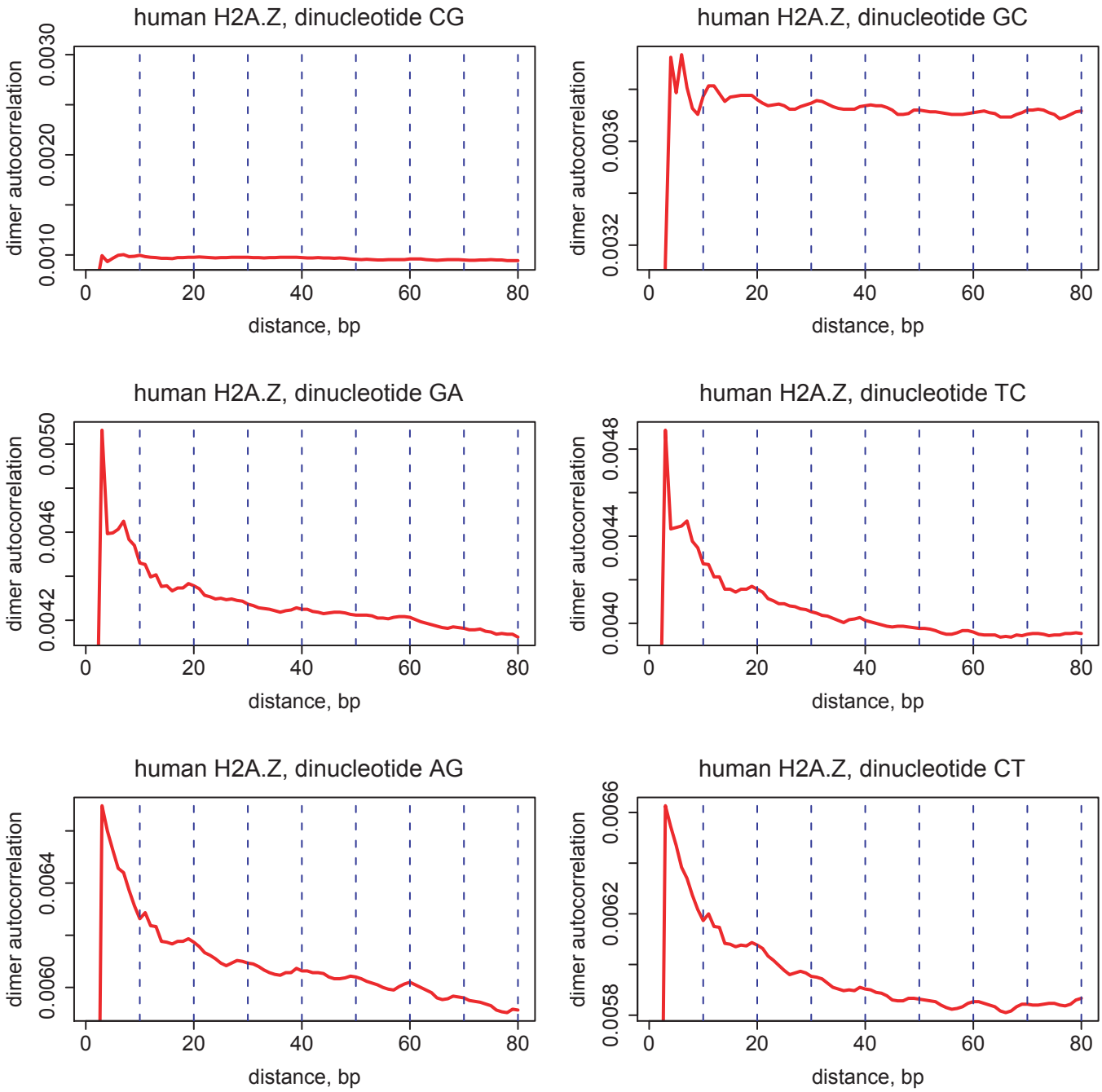


Figure S15 continued

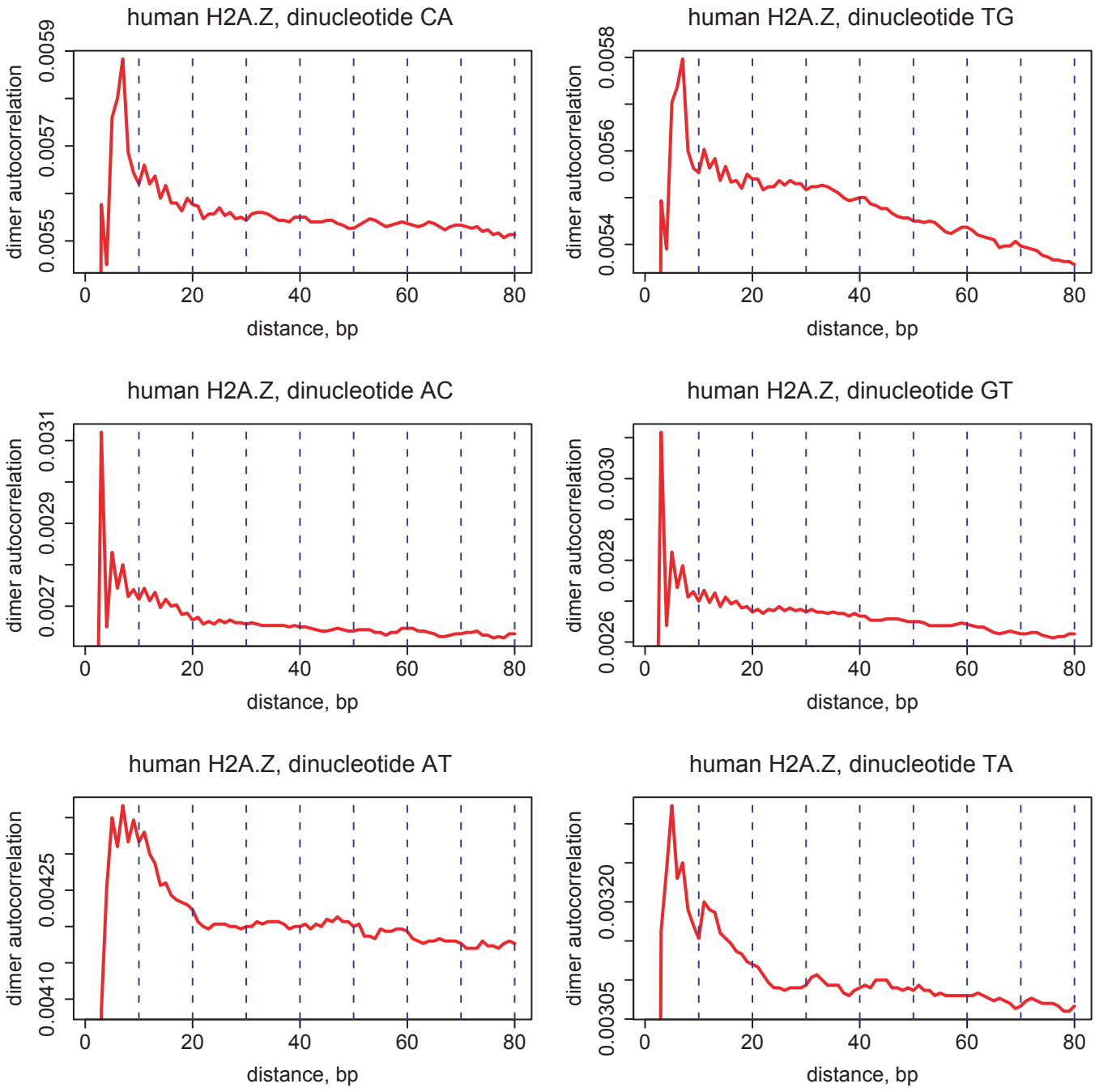


Figure S15 continued

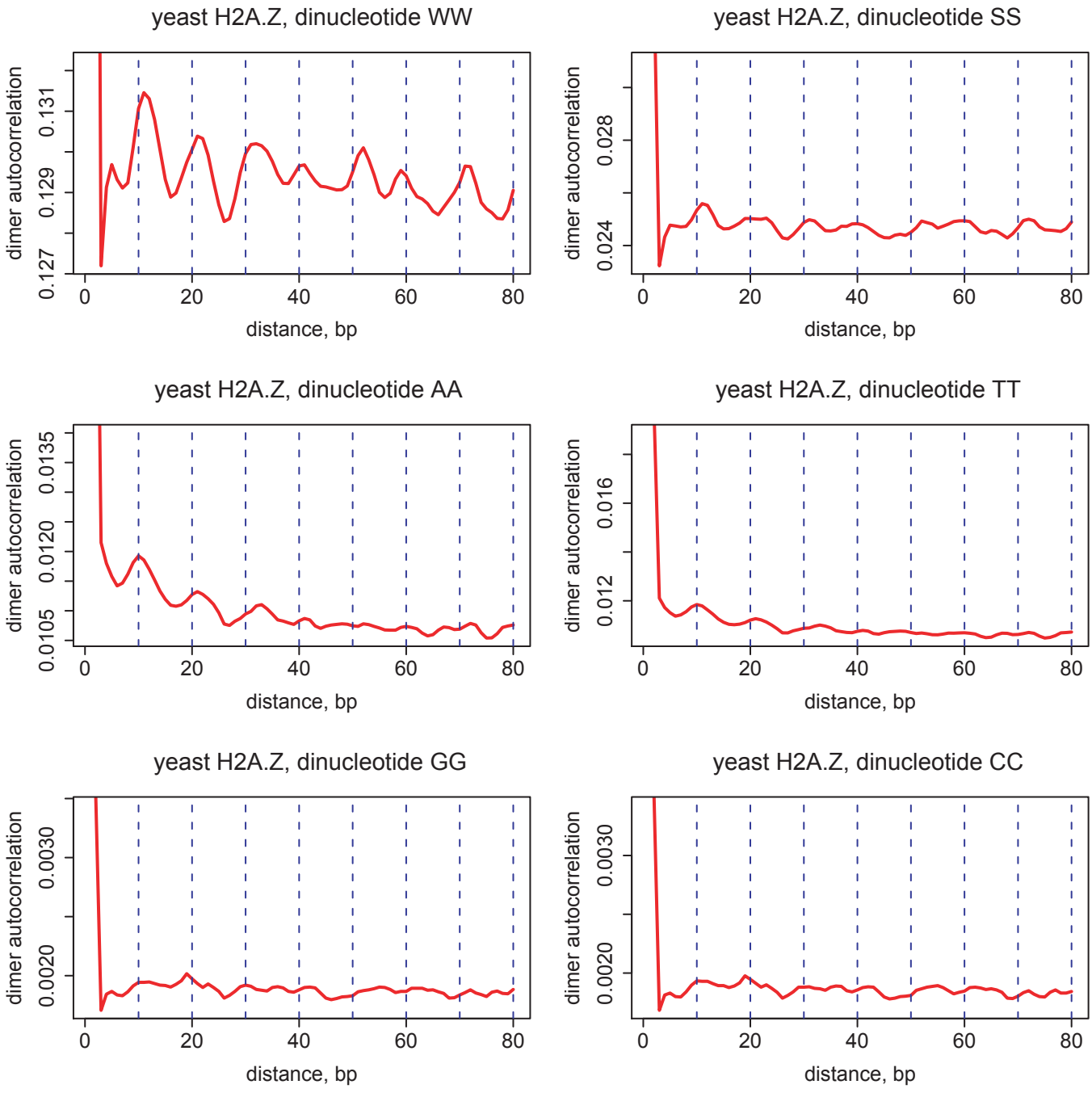


Figure S15 continued

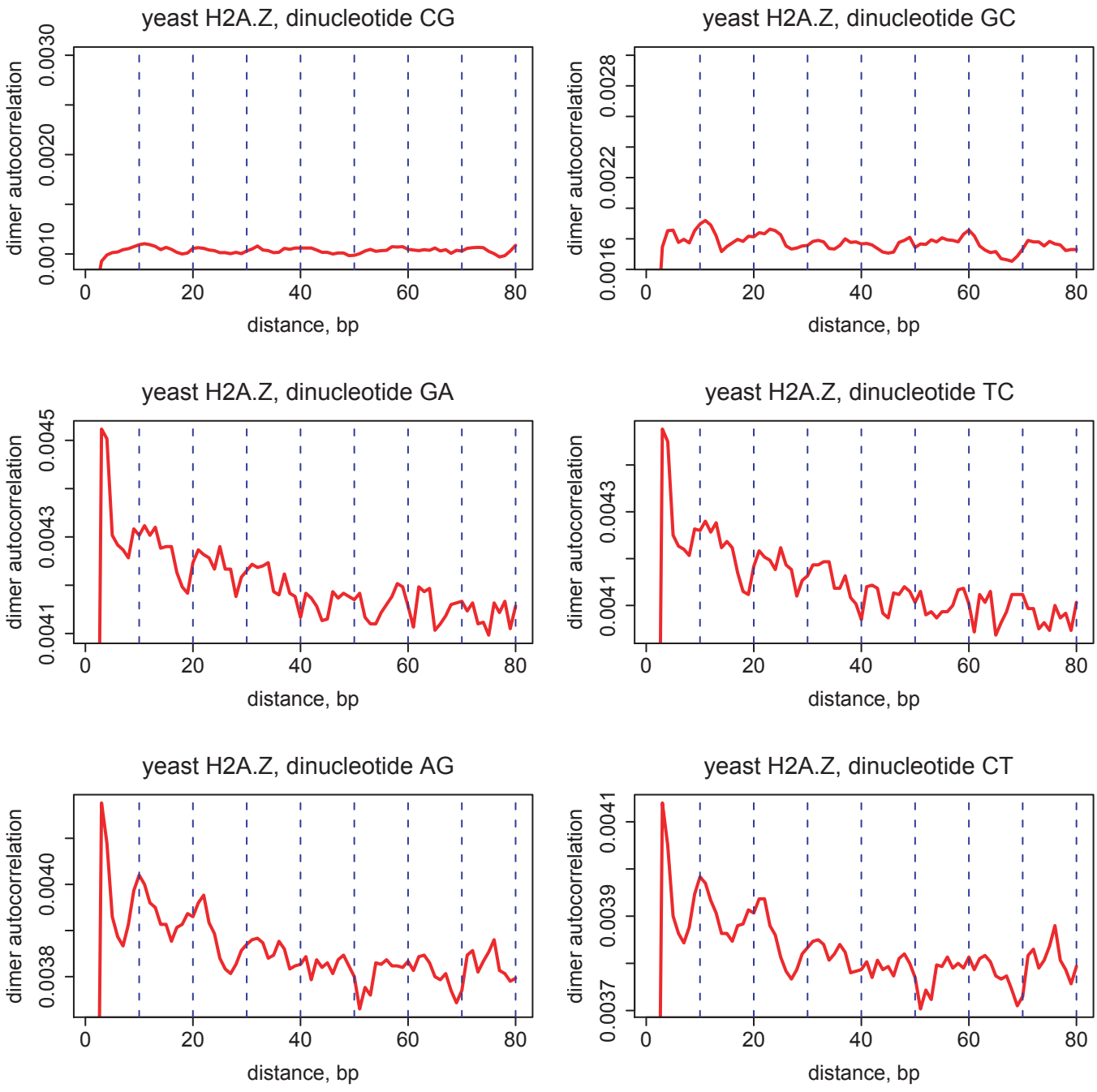


Figure S15 continued

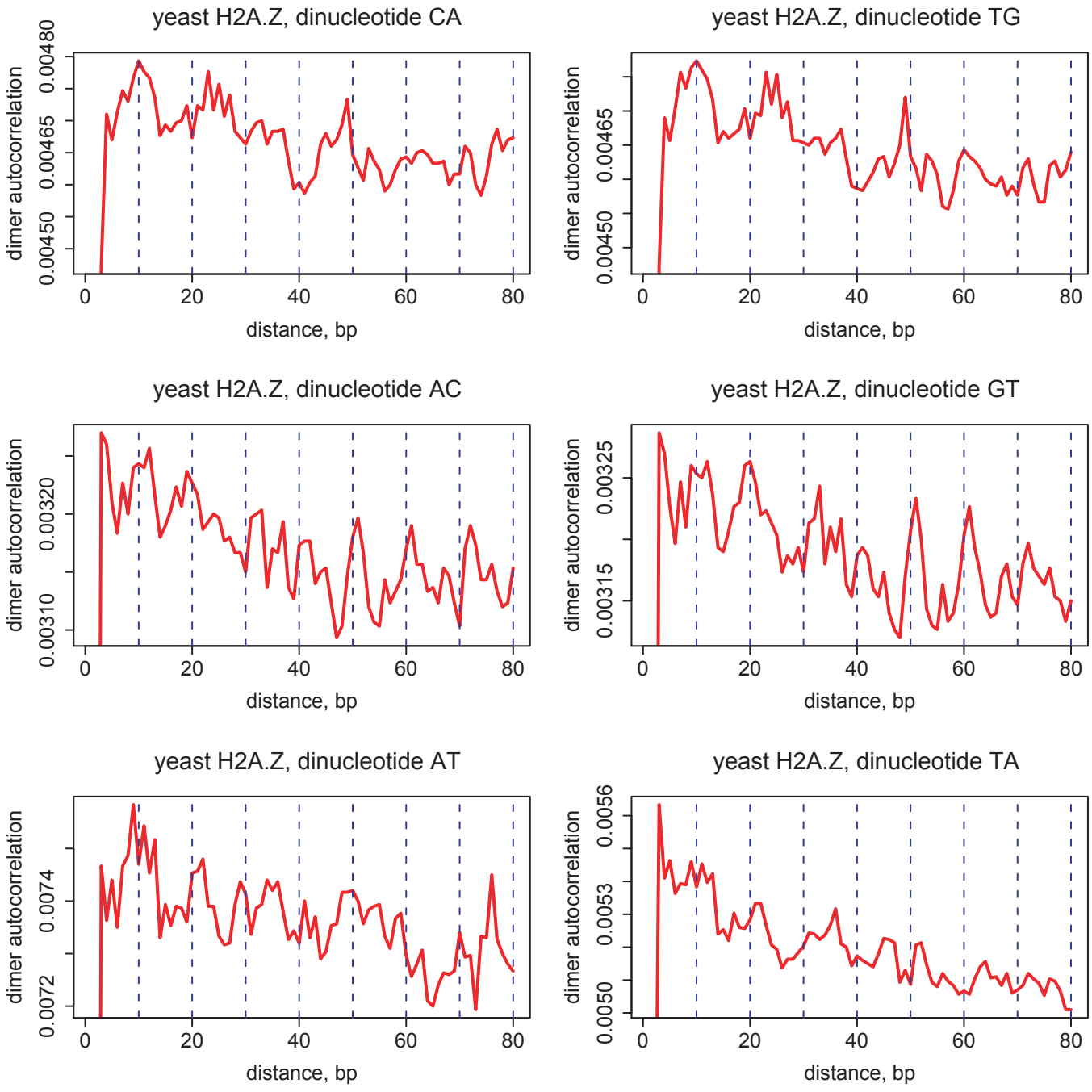


Figure S15 continued

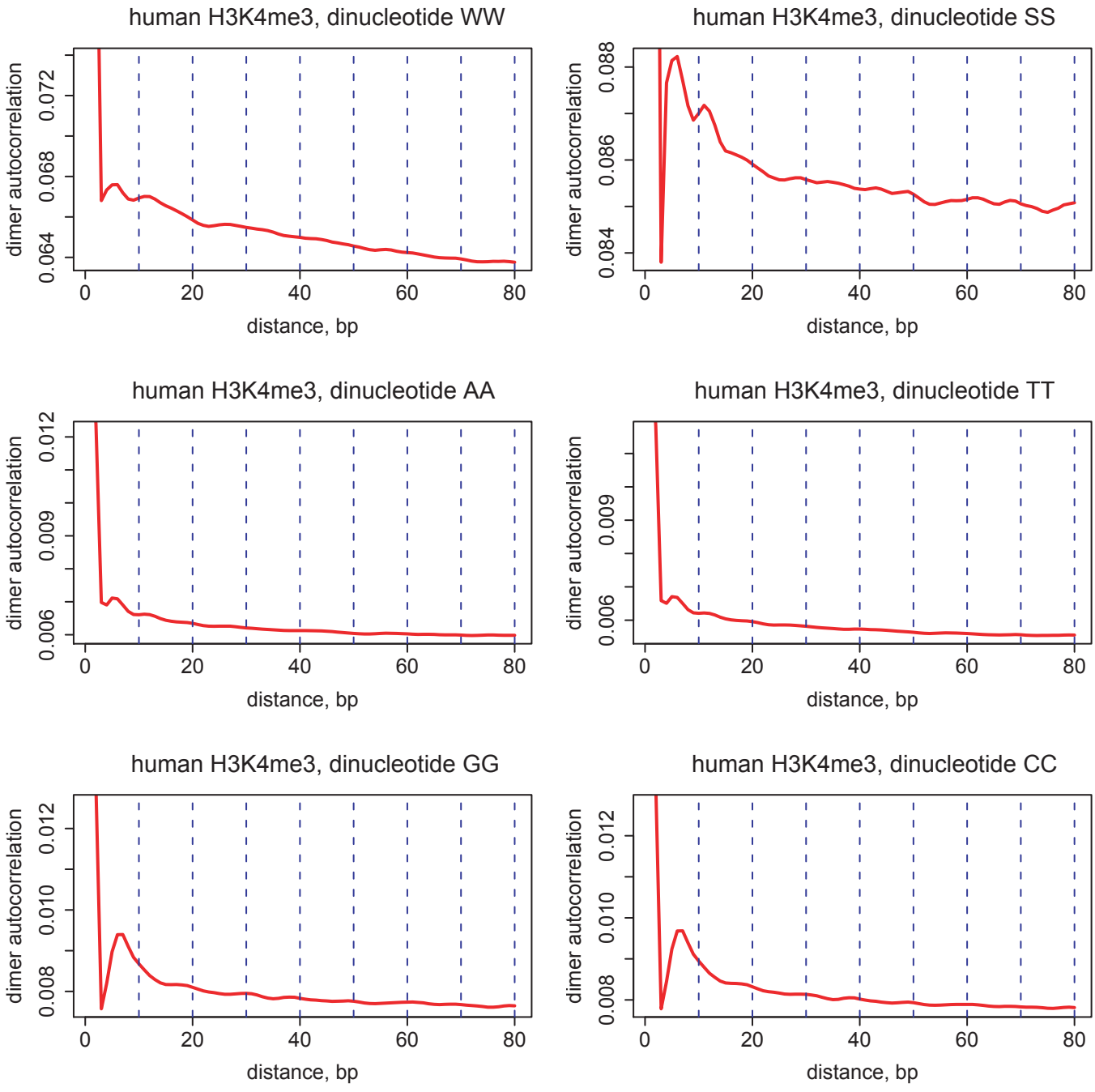

human H3K4me3, dinucleotide WW
dimer autocorrelation
0.064
0.068
0.072
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide SS
dimer autocorrelation
0.084
0.086
0.088
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide AA
dimer autocorrelation
0.006
0.009
0.012
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide TT
dimer autocorrelation
0.006
0.009
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide GG
dimer autocorrelation
0.008
0.010
0.012
0
20
40
60
80
distance, bp
human H3K4me3, dinucleotide CC
dimer autocorrelation
0.008
0.010
0.012
0
20
40
60
80
distance, bp

Figure S16

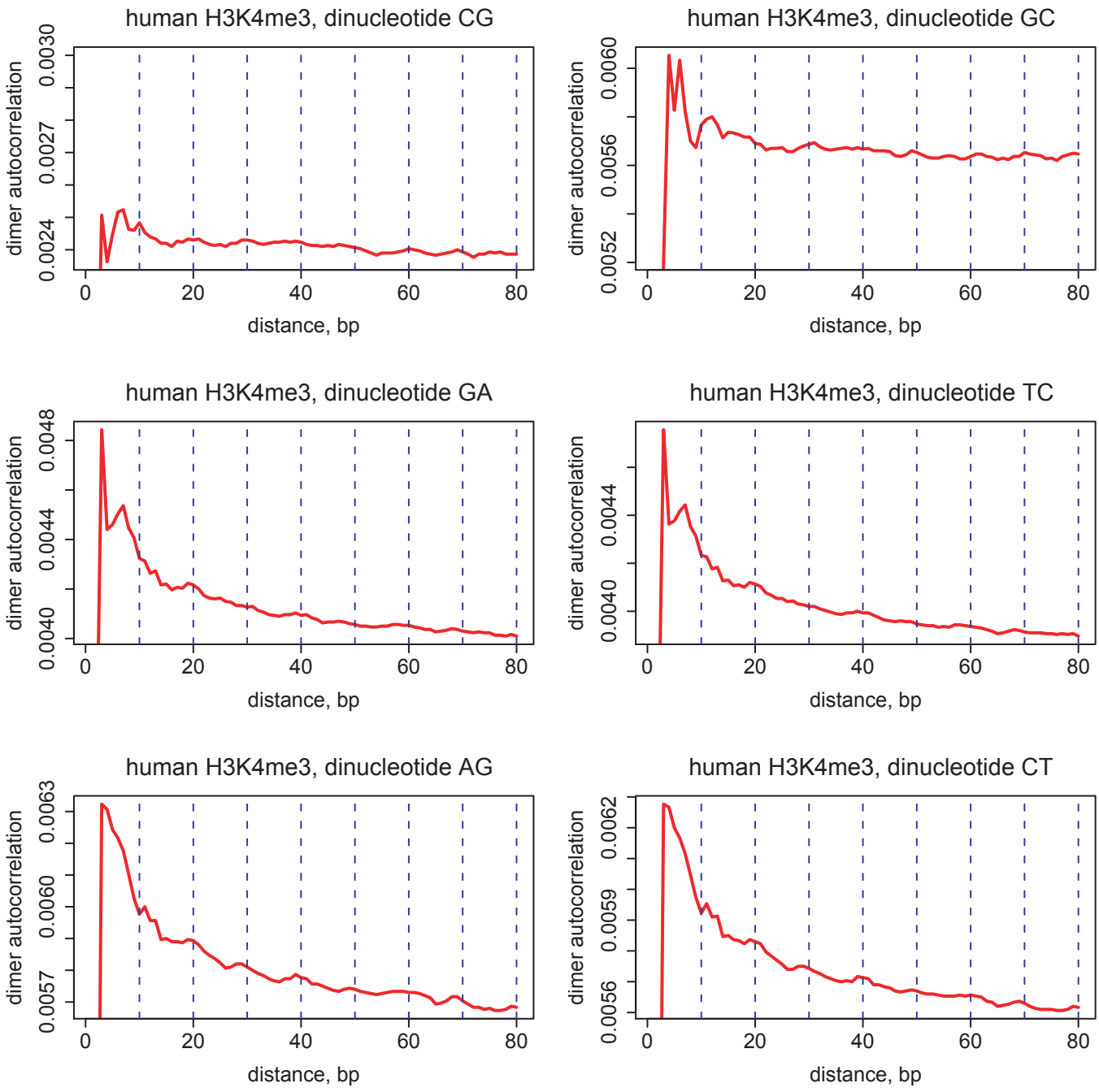


Figure S16 continued

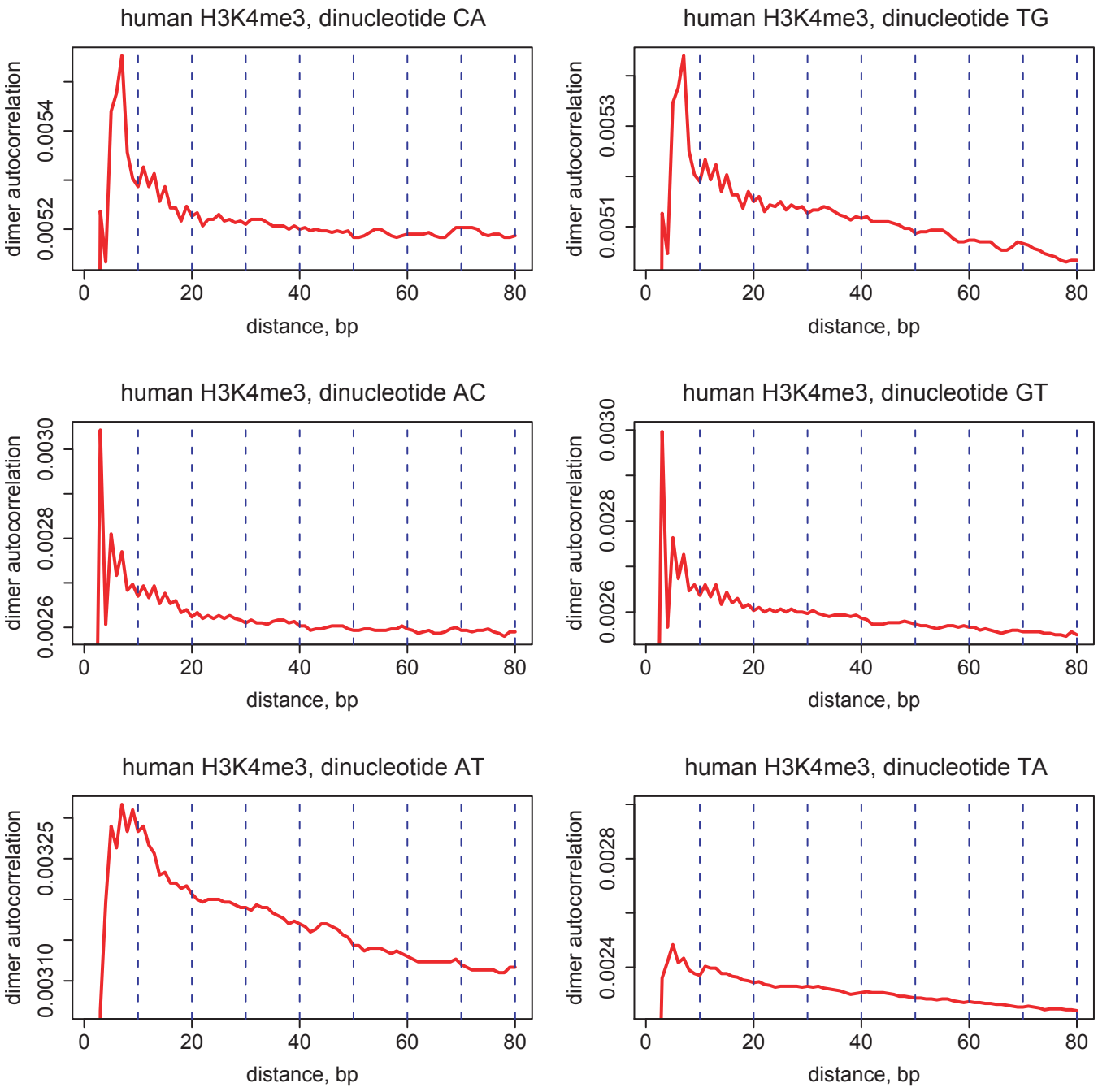


Figure S16 continued

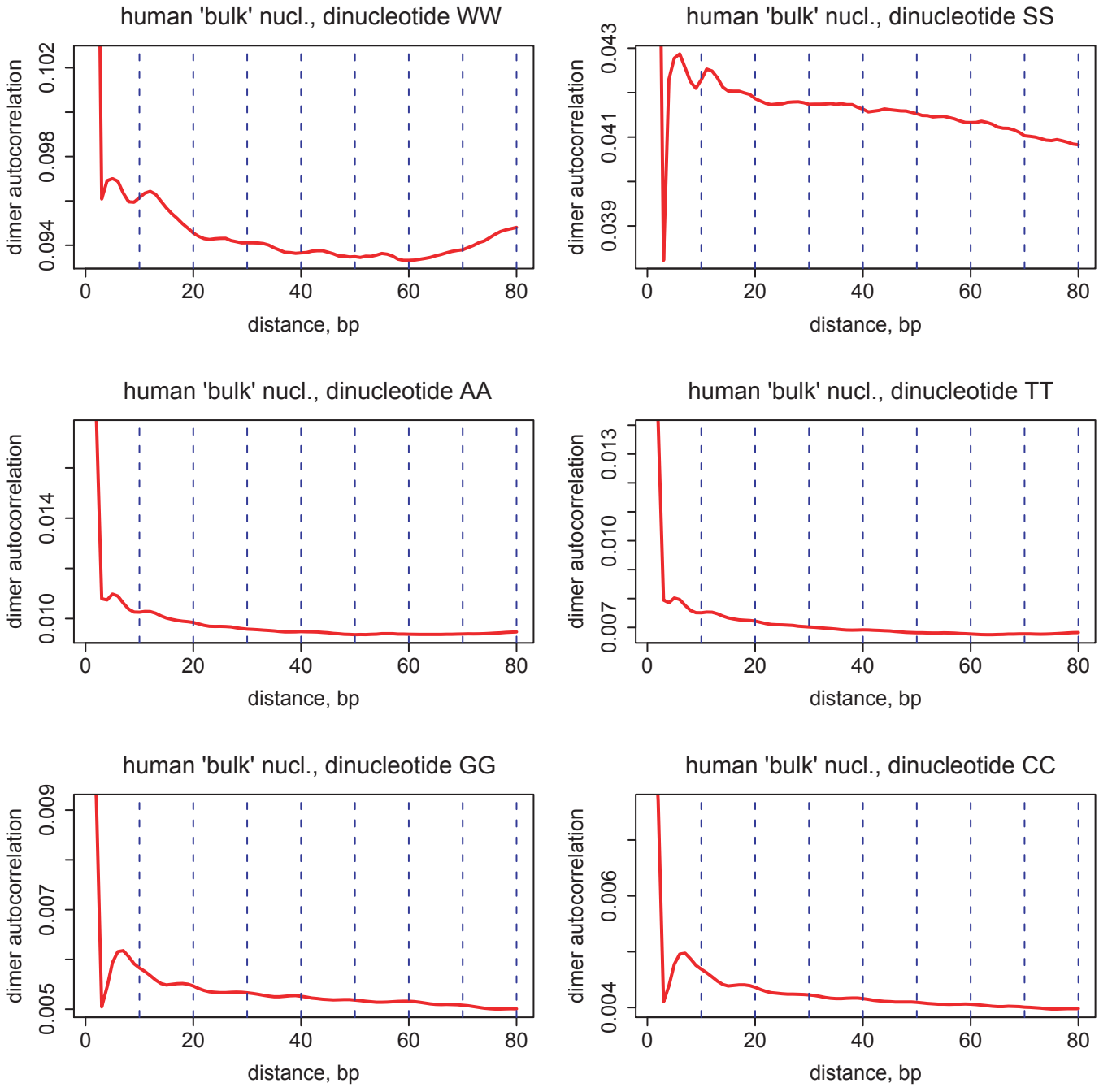

human 'bulk' nucl., dinucleotide WW
dimer autocorrelation
0.094
0.098
0.102
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide SS
dimer autocorrelation
0.039
0.041
0.043
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide AA
dimer autocorrelation
0.010
0.014
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide TT
dimer autocorrelation
0.007
0.010
0.013
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide GG
dimer autocorrelation
0.005
0.007
0.009
0
20
40
60
80
distance, bp
human 'bulk' nucl., dinucleotide CC
dimer autocorrelation
0.004
0.006
0
20
40
60
80
distance, bp

Figure S17

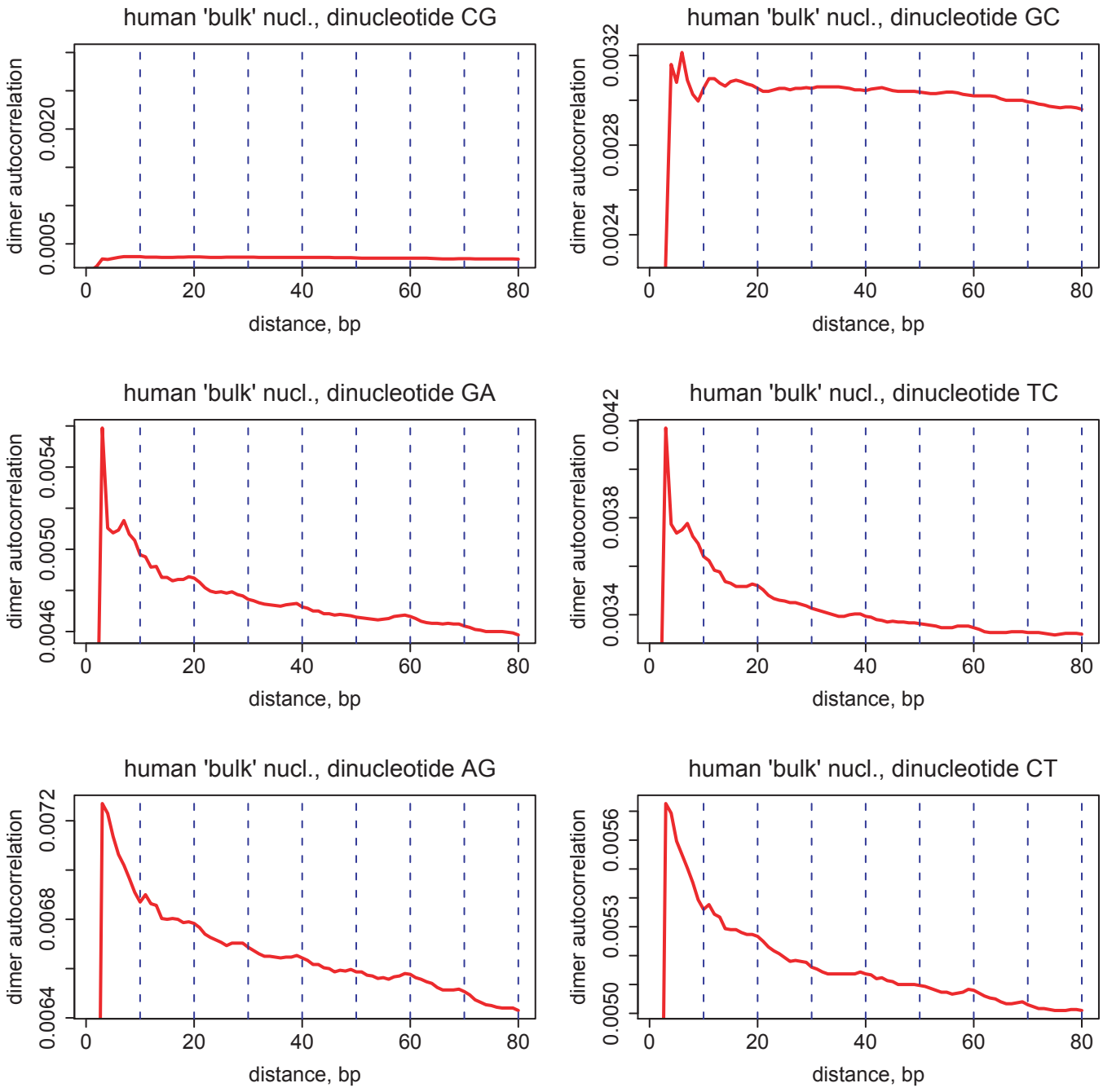


Figure S17 continued

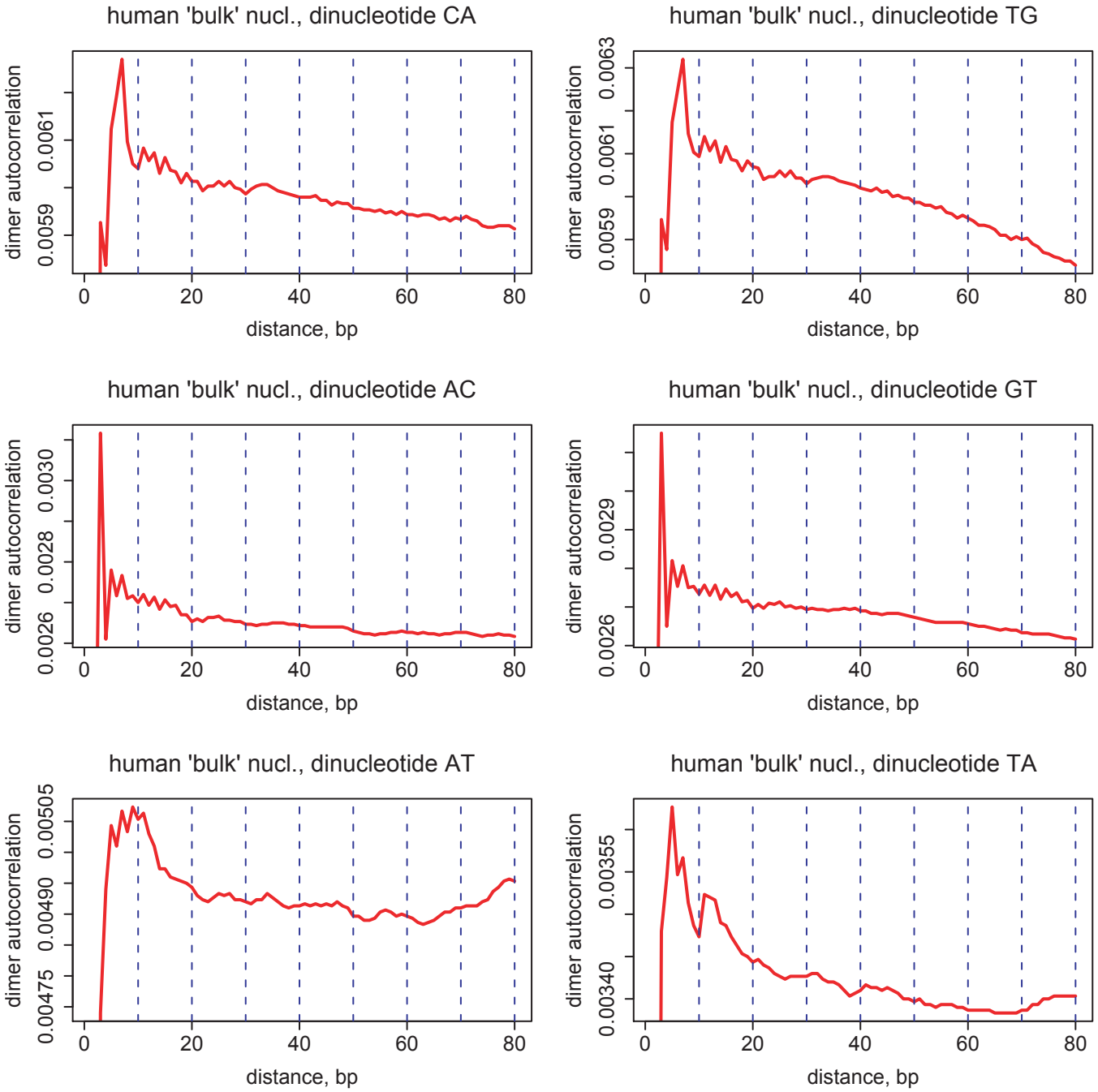


Figure S17 continued

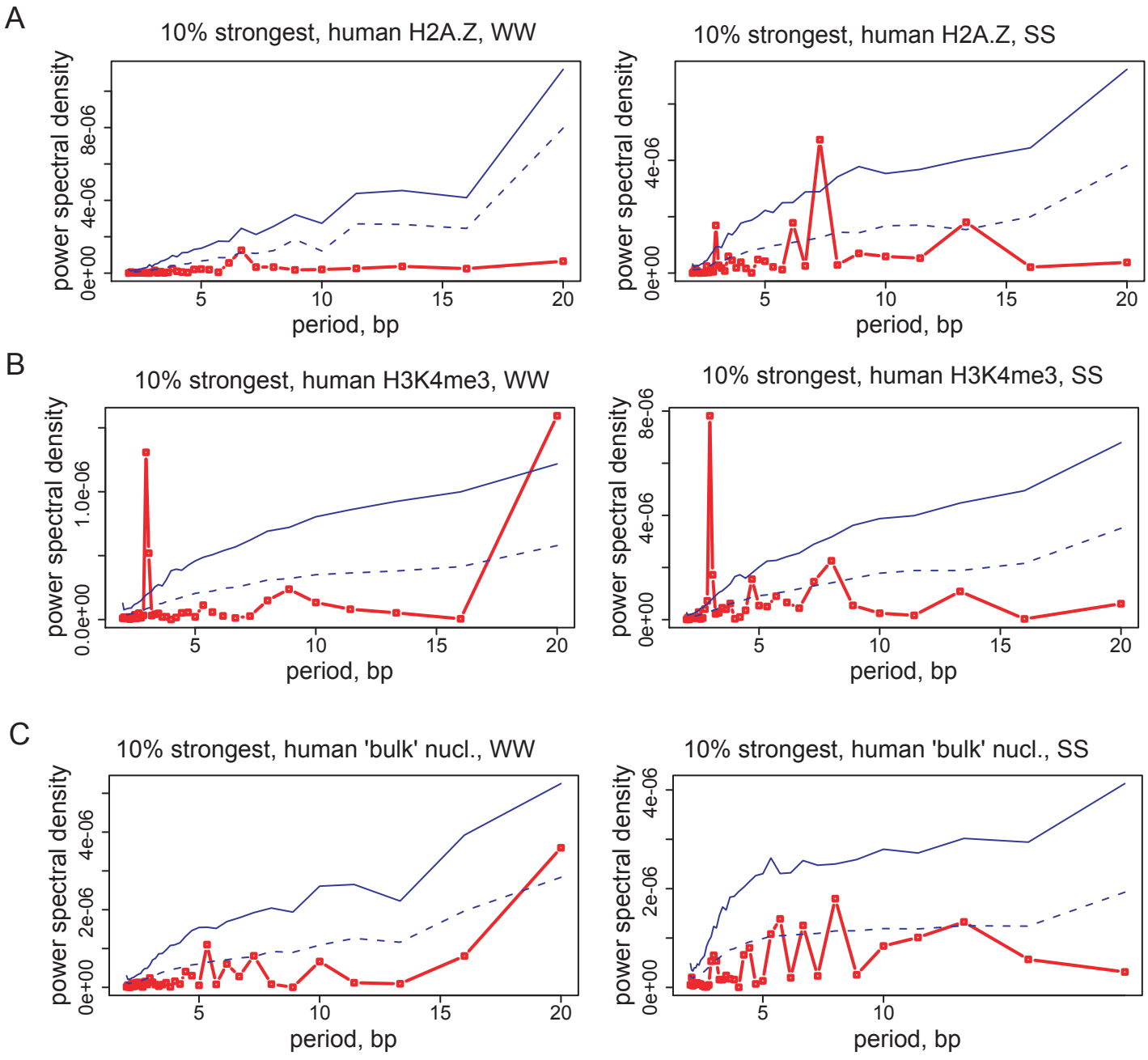

A
10% strongest, human H2A.Z, WW
power spectral density
0e+00
4e-06
8e-06
5
10
15
20
period, bp
10% strongest, human H2A.Z, SS
power spectral density
0e+00
4e-06
5
10
15
20
period, bp
B
10% strongest, human H3K4me3, WW
power spectral density
0.0e+00
1.0e-06
5
10
15
20
period, bp
10% strongest, human H3K4me3, SS
power spectral density
0e+00
4e-06
8e-06
5
10
15
20
period, bp
C
10% strongest, human 'bulk' nucl., WW
power spectral density
0e+00
2e-06
4e-06
5
10
15
20
period, bp
10% strongest, human 'bulk' nucl., SS
power spectral density
0e+00
2e-06
4e-06
5
10
period, bp

Figure S18

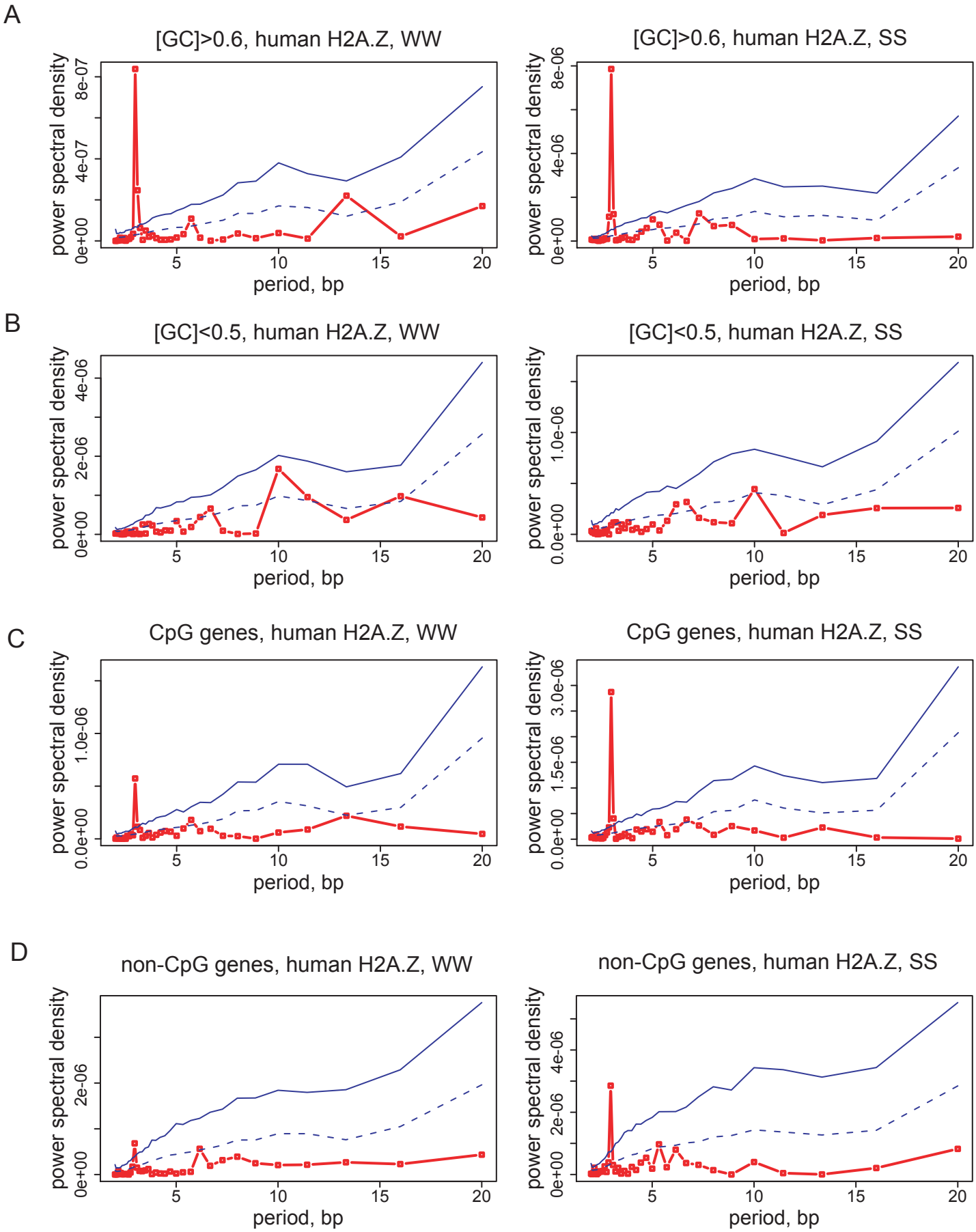

A
[GC]>0.6, human H2A.Z, WW
power spectral density
0e+00
4e-07
8e-07
5
10
15
20
period, bp
[GC]>0.6, human H2A.Z, SS
power spectral density
0e+00
4e-06
8e-06
5
10
15
20
period, bp
B
[GC]<0.5, human H2A.Z, WW
power spectral density
0e+00
2e-06
4e-06
5
10
15
20
period, bp
[GC]<0.5, human H2A.Z, SS
power spectral density
0.0e+00
1.0e-06
5
10
15
20
period, bp
C
CpG genes, human H2A.Z, WW
power spectral density
0.0e+00
1.0e-06
5
10
15
20
period, bp
CpG genes, human H2A.Z, SS
power spectral density
0.0e+00
1.5e-06
3.0e-06
5
10
15
20
period, bp
D
non-CpG genes, human H2A.Z, WW
power spectral density
0e+00
2e-06
5
10
15
20
period, bp
non-CpG genes, human H2A.Z, SS
power spectral density
0e+00
2e-06
4e-06
5
10
15
20
period, bp

Figure S19

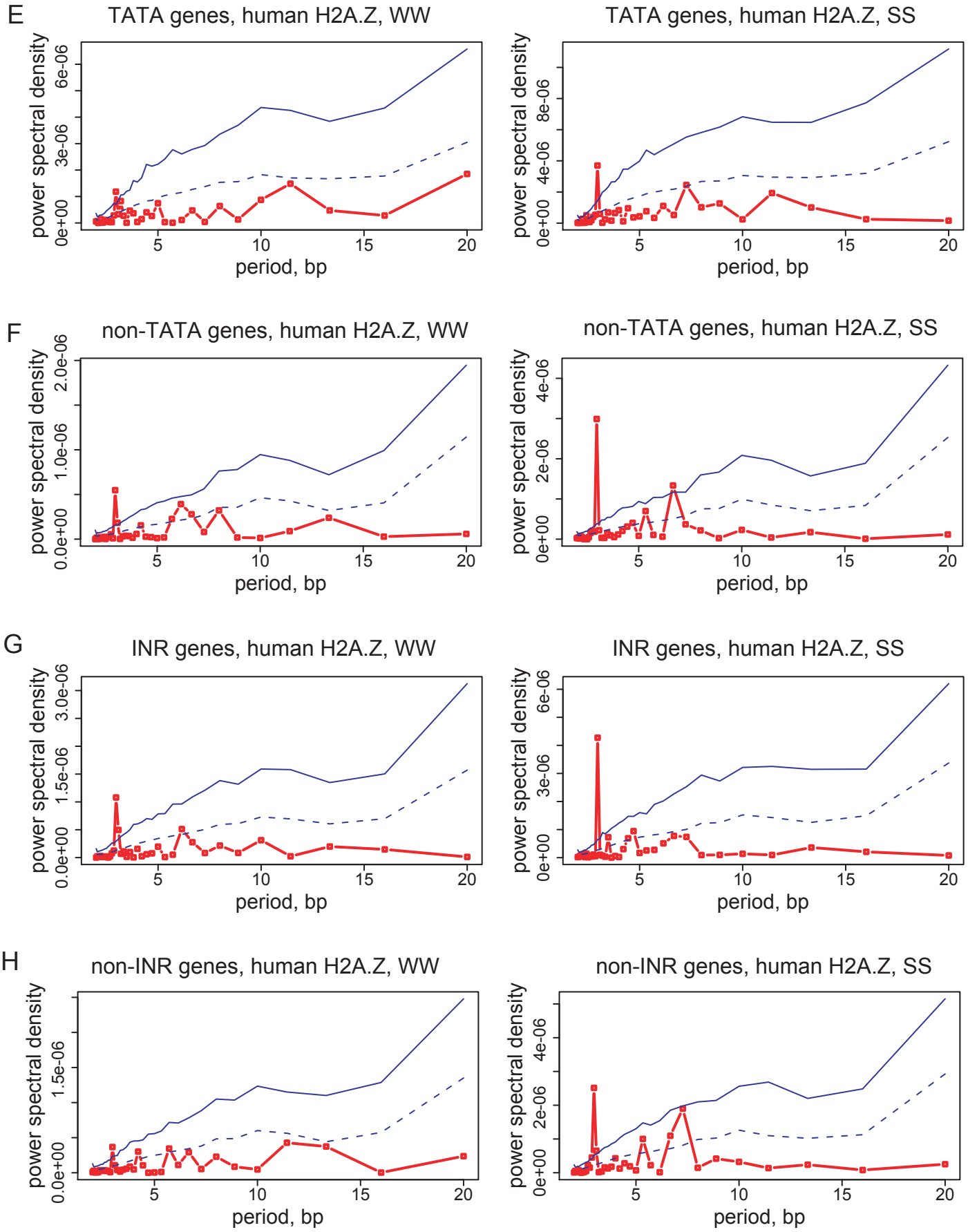


Figure S19 continued

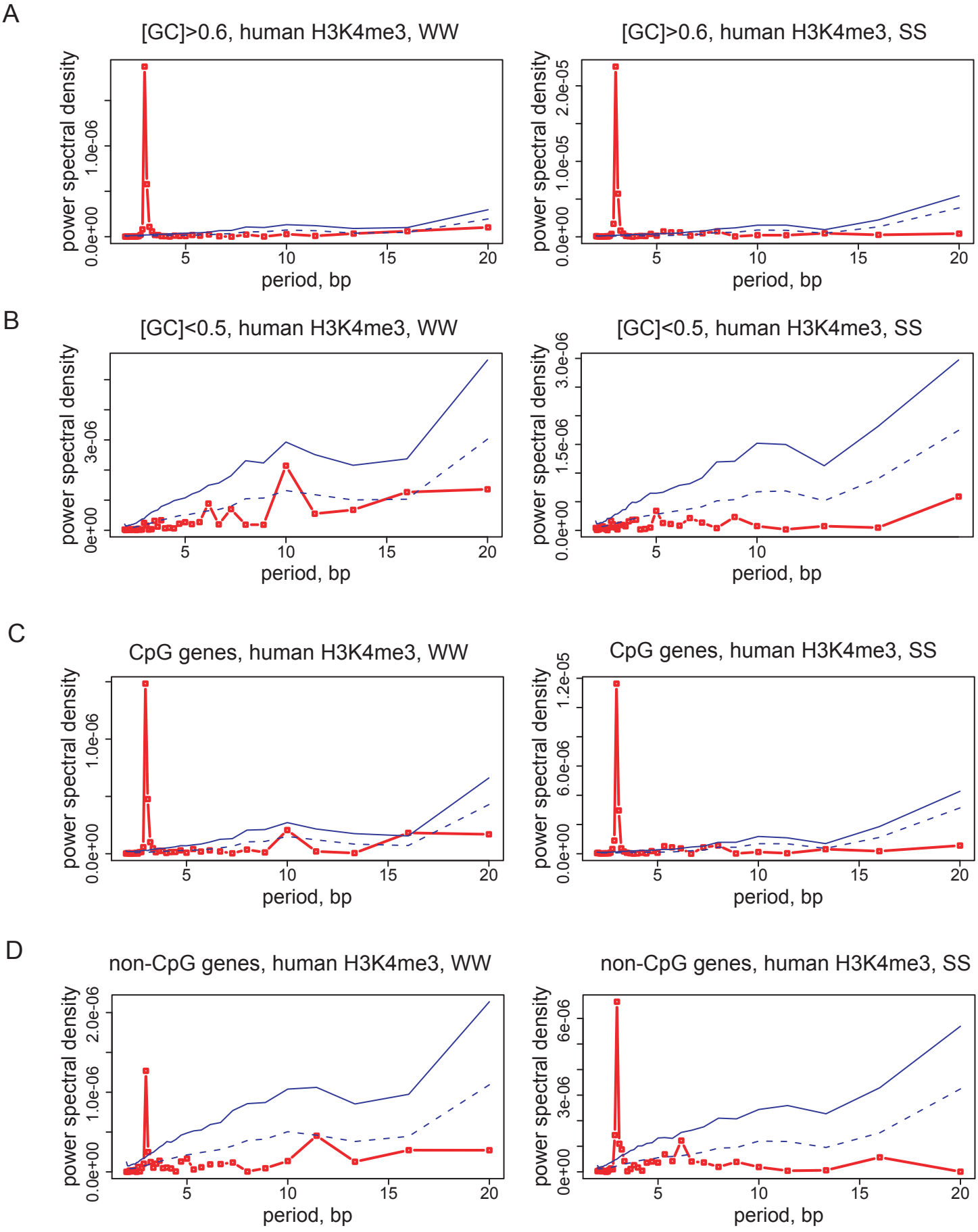
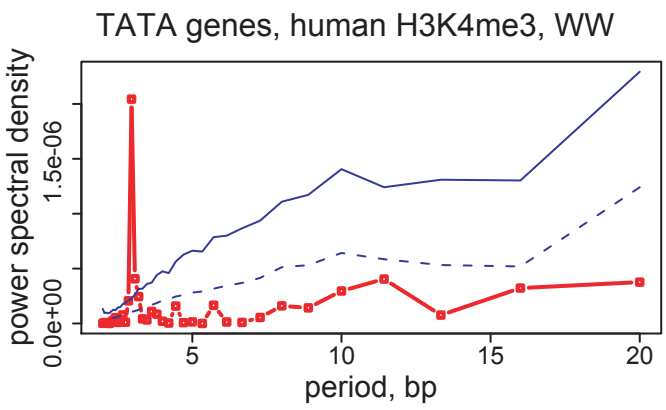
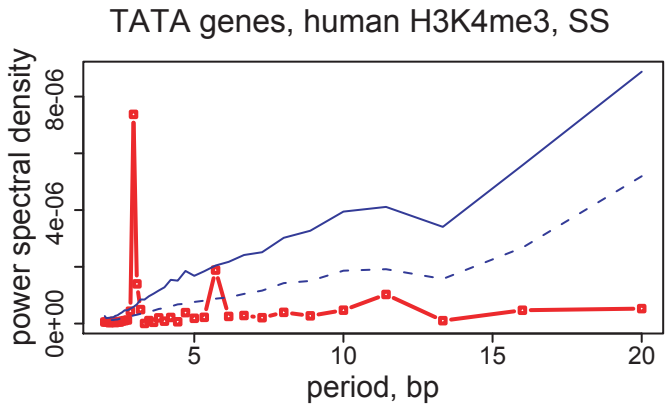

A
[GC]>0.6, human H3K4me3, WW
[GC]>0.6, human H3K4me3, SS
B
[GC]<0.5, human H3K4me3, WW
[GC]<0.5, human H3K4me3, SS
C
CpG genes, human H3K4me3, WW
CpG genes, human H3K4me3, SS
D
non-CpG genes, human H3K4me3, WW
non-CpG genes, human H3K4me3, SS
power spectral density
period, bp

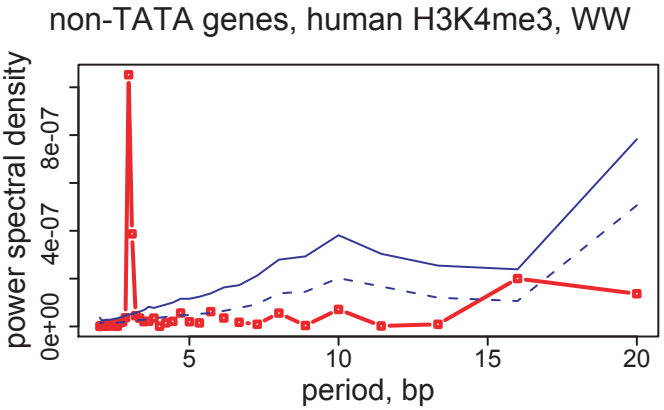
Figure S20

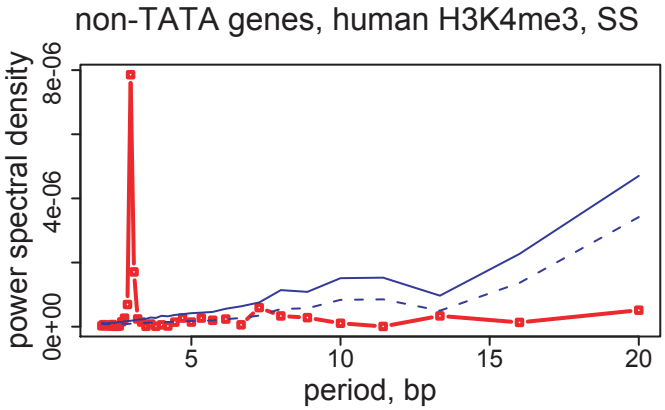
E



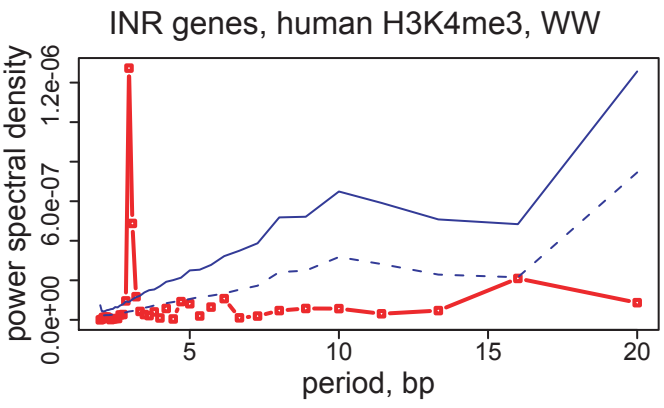


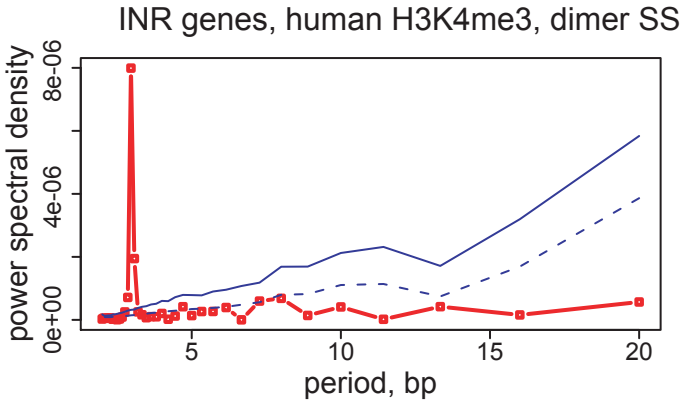
F



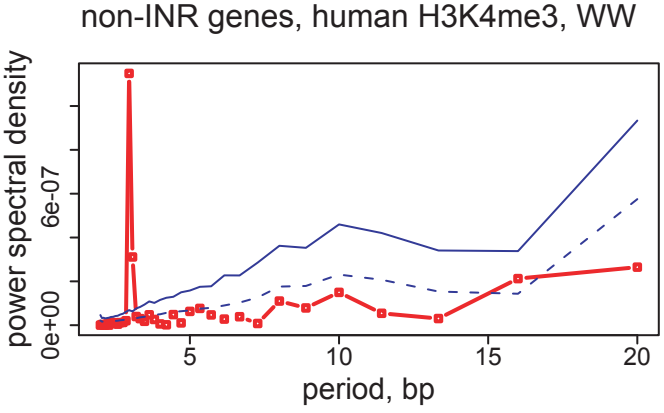


G





H



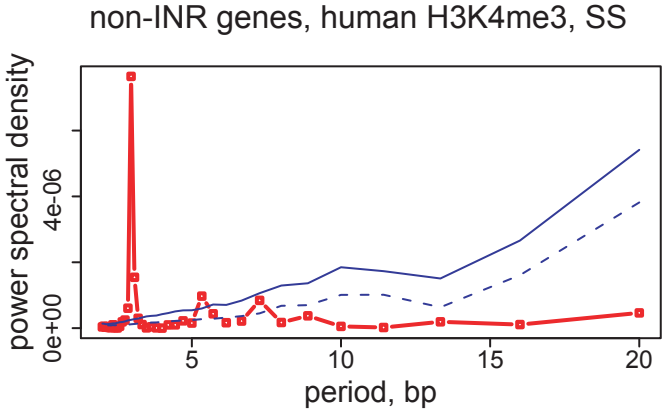


Figure S20 continued

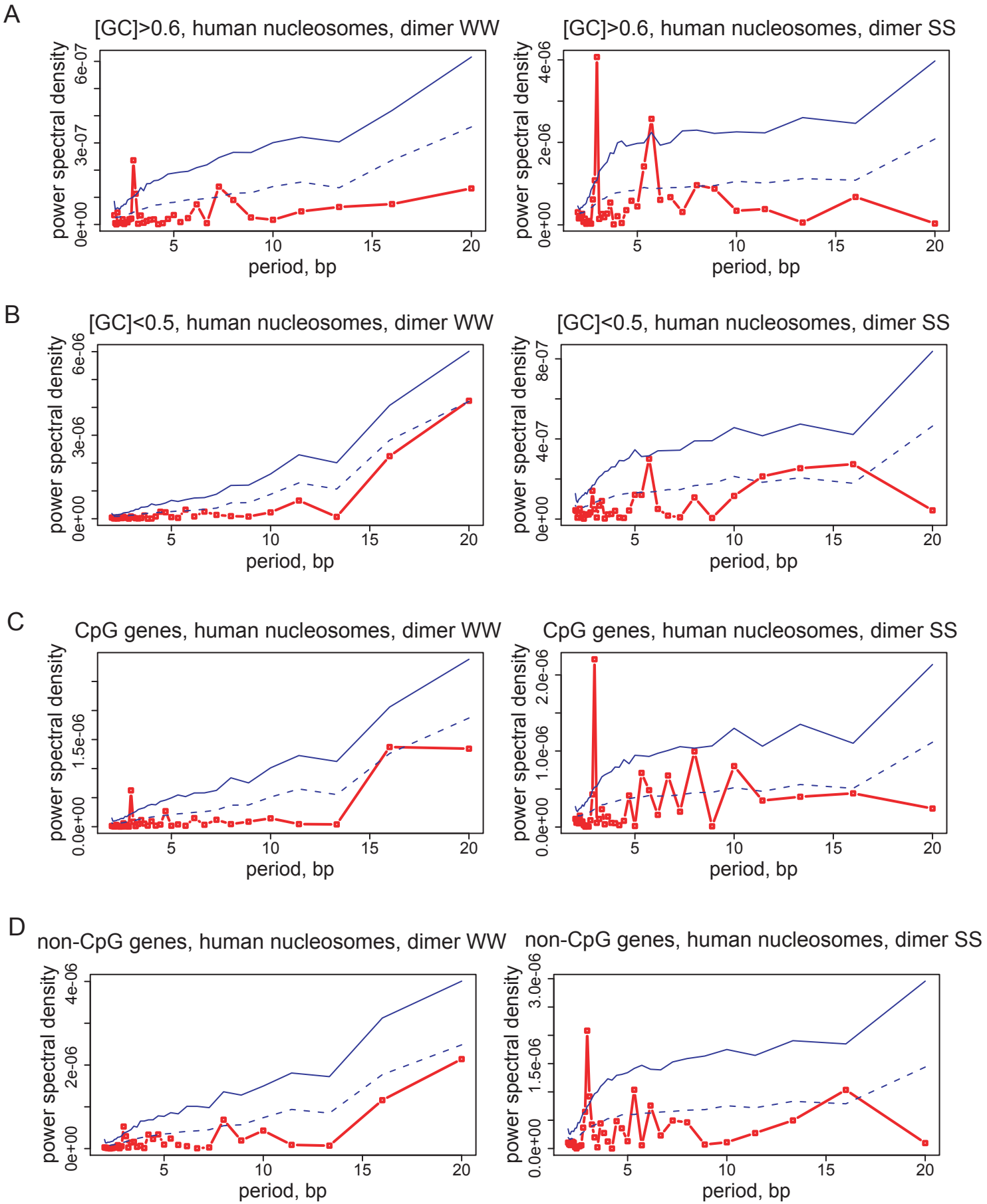
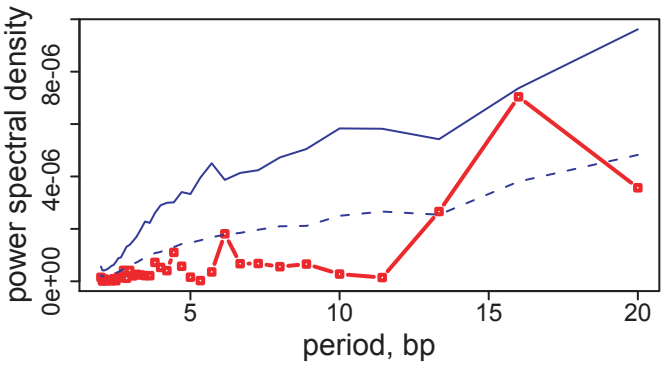
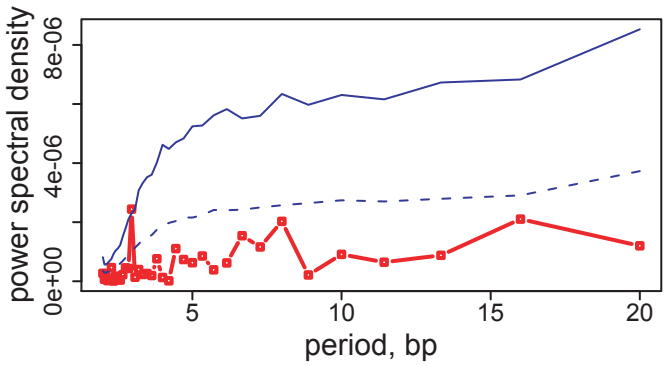

A
[GC]>0.6, human nucleosomes, dimer WW
power spectral density
0e+00
3e-07
6e-07
5
10
15
20
period, bp
[GC]>0.6, human nucleosomes, dimer SS
power spectral density
0e+00
2e-06
4e-06
5
10
15
20
period, bp
B
[GC]<0.5, human nucleosomes, dimer WW
power spectral density
0e+00
3e-06
6e-06
5
10
15
20
period, bp
[GC]<0.5, human nucleosomes, dimer SS
power spectral density
0e+00
4e-07
8e-07
5
10
15
20
period, bp
C
CpG genes, human nucleosomes, dimer WW
power spectral density
0.0e+00
1.5e-06
5
10
15
20
period, bp
CpG genes, human nucleosomes, dimer SS
power spectral density
0.0e+00
1.0e-06
2.0e-06
5
10
15
20
period, bp
D
non-CpG genes, human nucleosomes, dimer WW
power spectral density
0e+00
2e-06
4e-06
5
10
15
20
period, bp
non-CpG genes, human nucleosomes, dimer SS
power spectral density
0.0e+00
1.5e-06
3.0e-06
5
10
15
20
period, bp

Figure S21

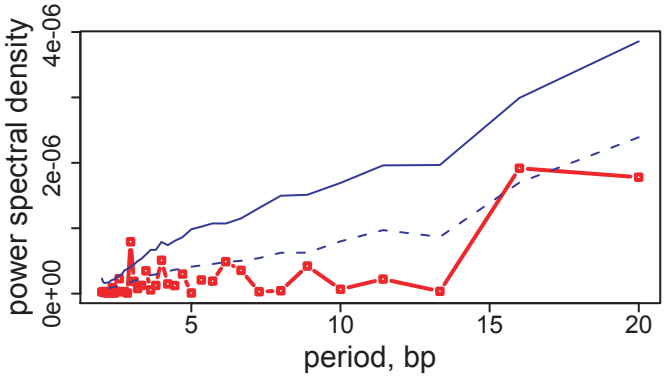
E TATA genes, human nucleosomes, dimer WW



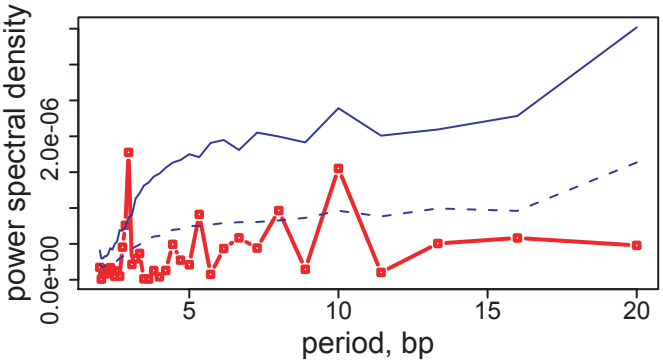
TATA genes, human nucleosomes, dimer SS



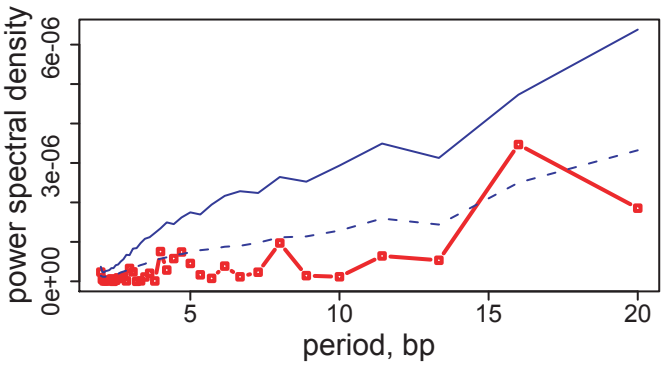
F non-TATA genes, human nucleosomes, dimer WW



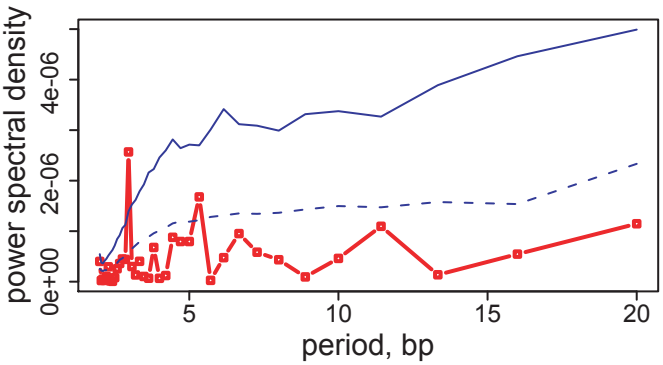
non-TATA genes, human nucleosomes, dimer SS



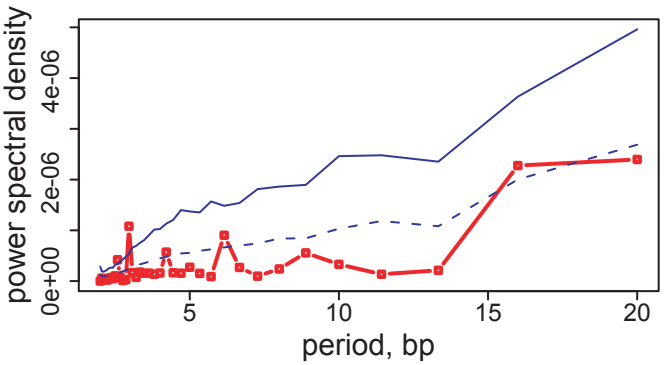
G INR genes, human nucleosomes, dimer WW



INR genes, human nucleosomes, dimer SS



H non-INR genes, human nucleosomes, dimer WW



non-INR genes, human nucleosomes, dimer SS

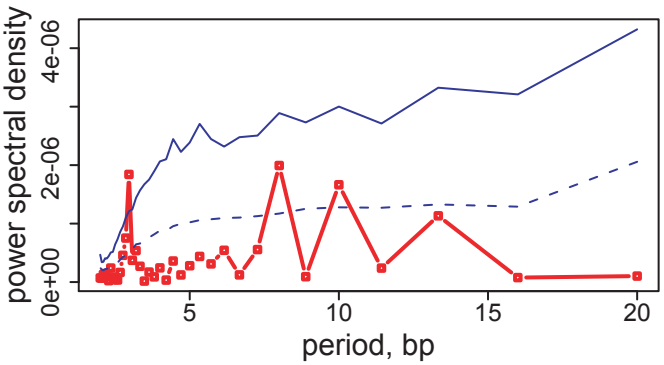
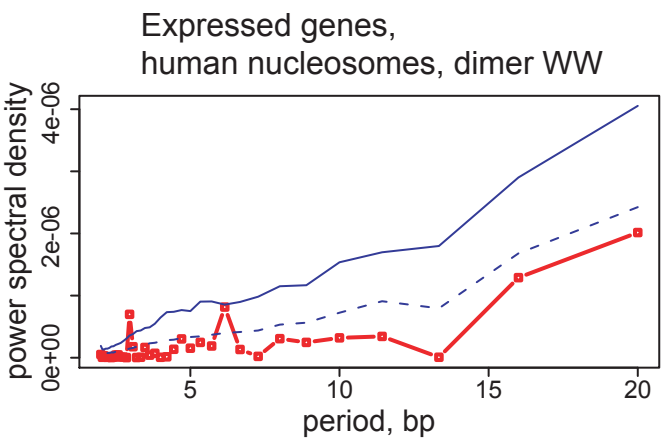
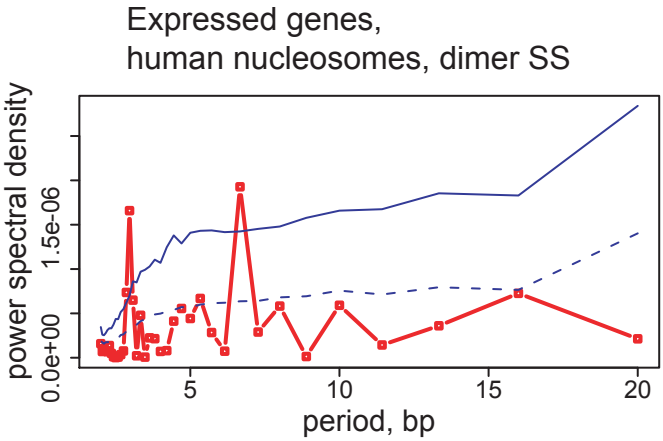


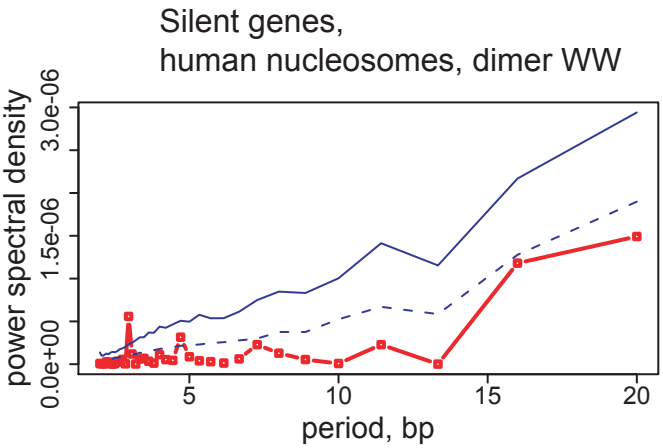
Figure S21 continued

I





J



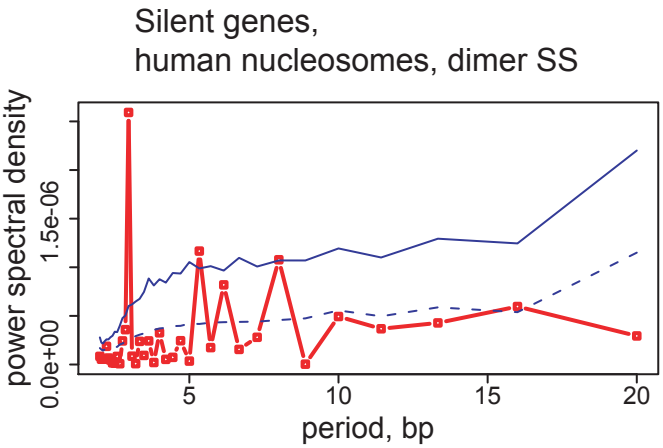
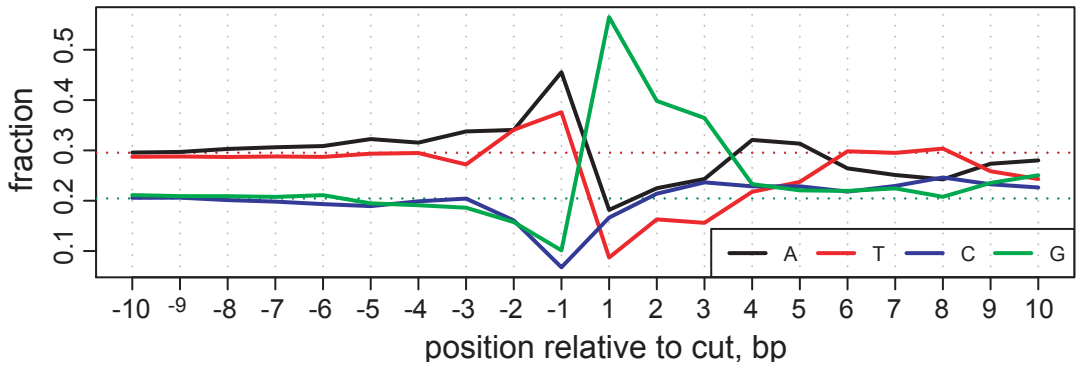


Figure S21 continued


fraction
0.1
0.2
0.3
0.4
0.5
A
T
C
G
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10
position relative to cut, bp

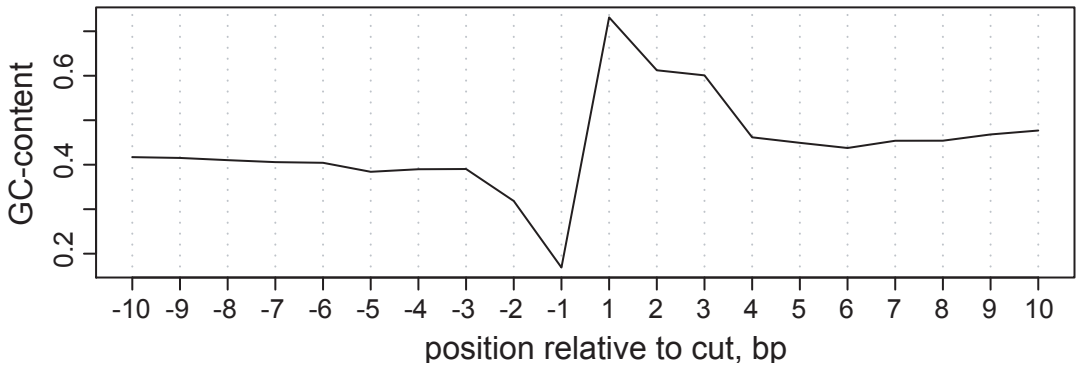

GC-content
0.2
0.4
0.6
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10
position relative to cut, bp

Figure S22

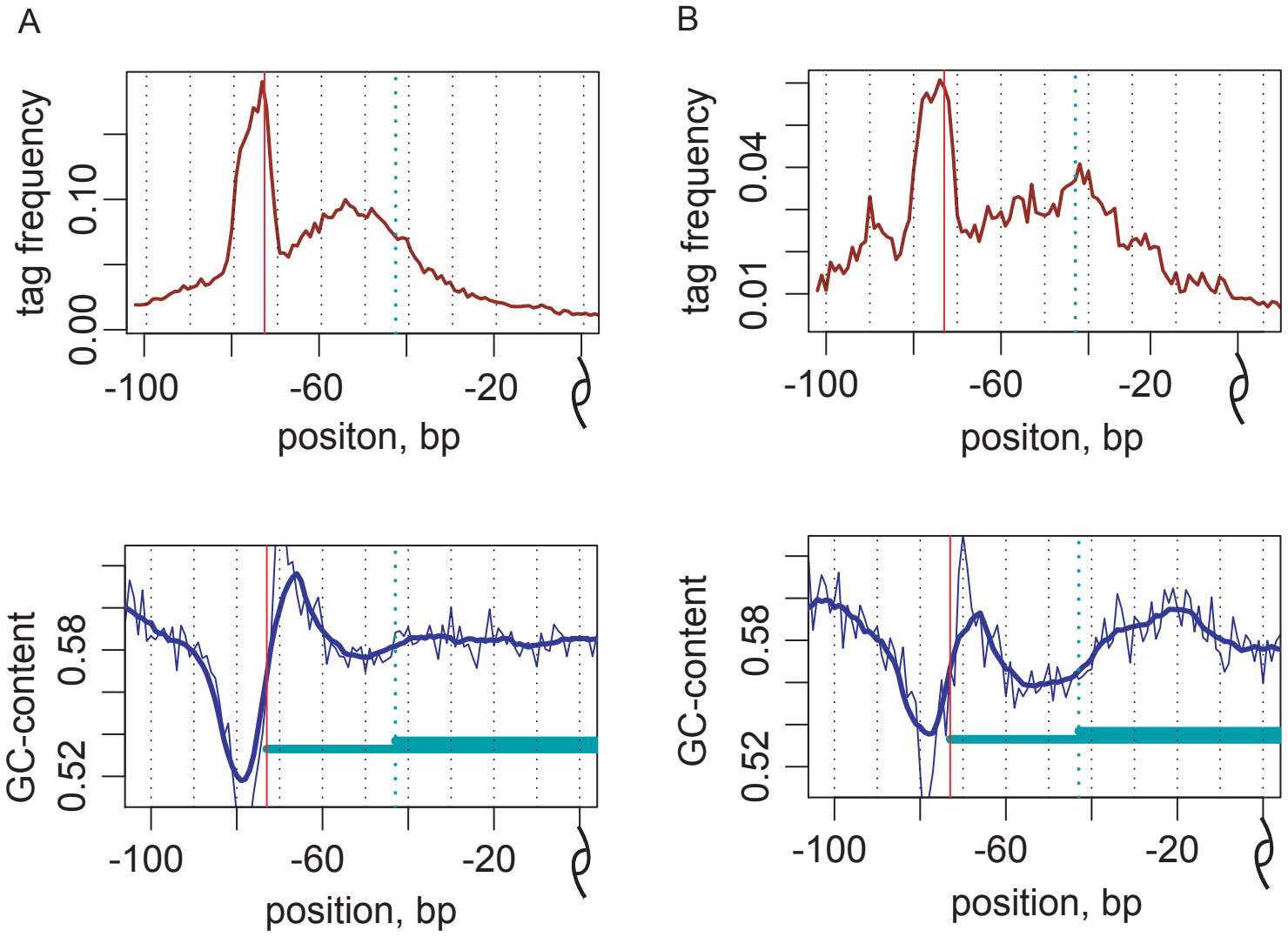
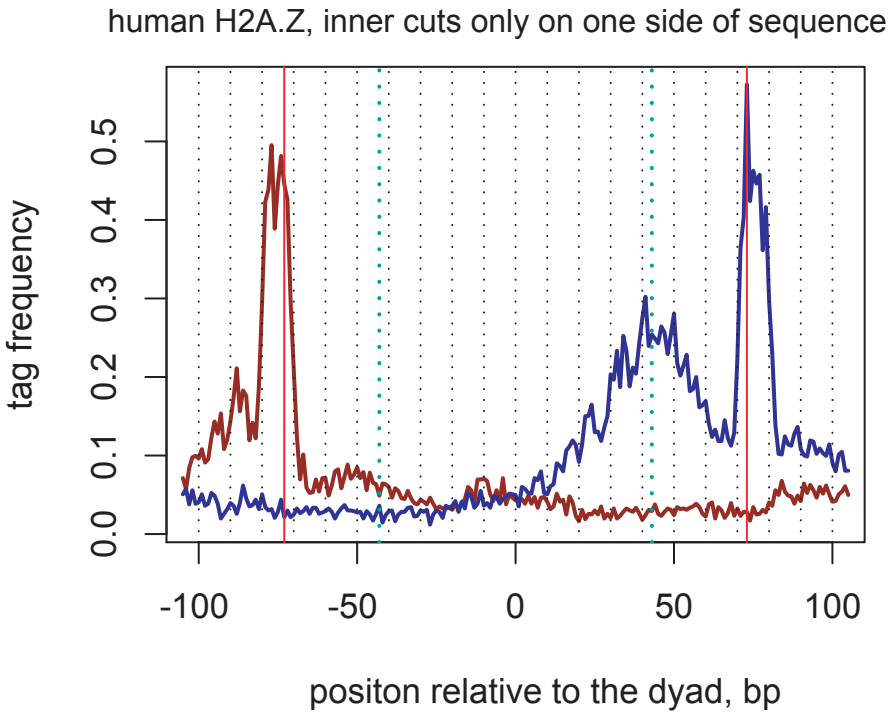

A
B
tag frequency
0.00
0.10
0.01
0.04
-100
-60
-20
positon, bp
GC-content
0.52
0.58
position, bp

Figure S23


human H2A.Z, inner cuts only on one side of sequence
tag frequency
0.0
0.1
0.2
0.3
0.4
0.5
-100
-50
0
50
100
positon relative to the dyad, bp

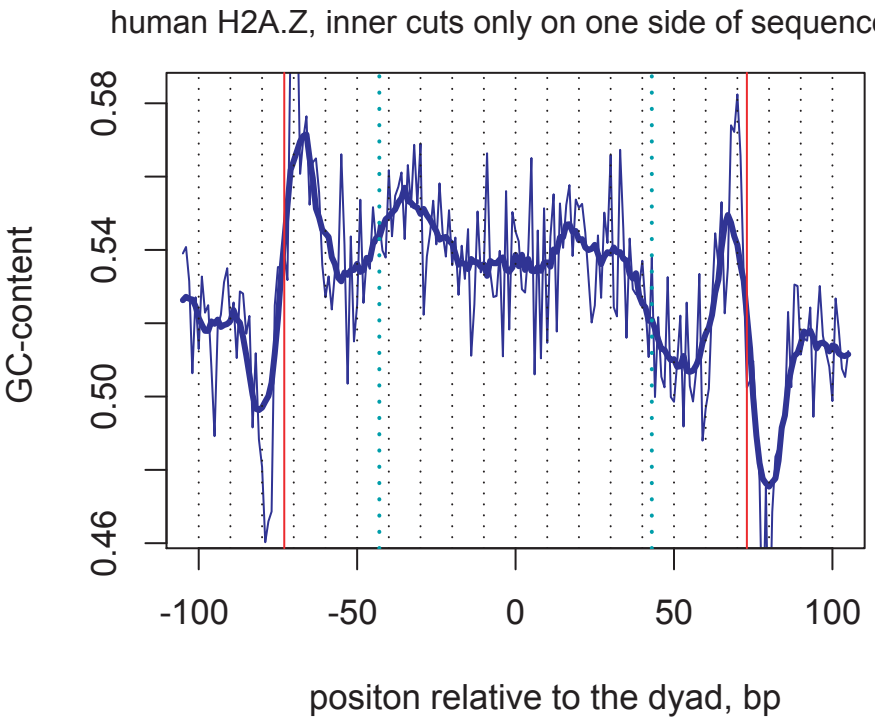
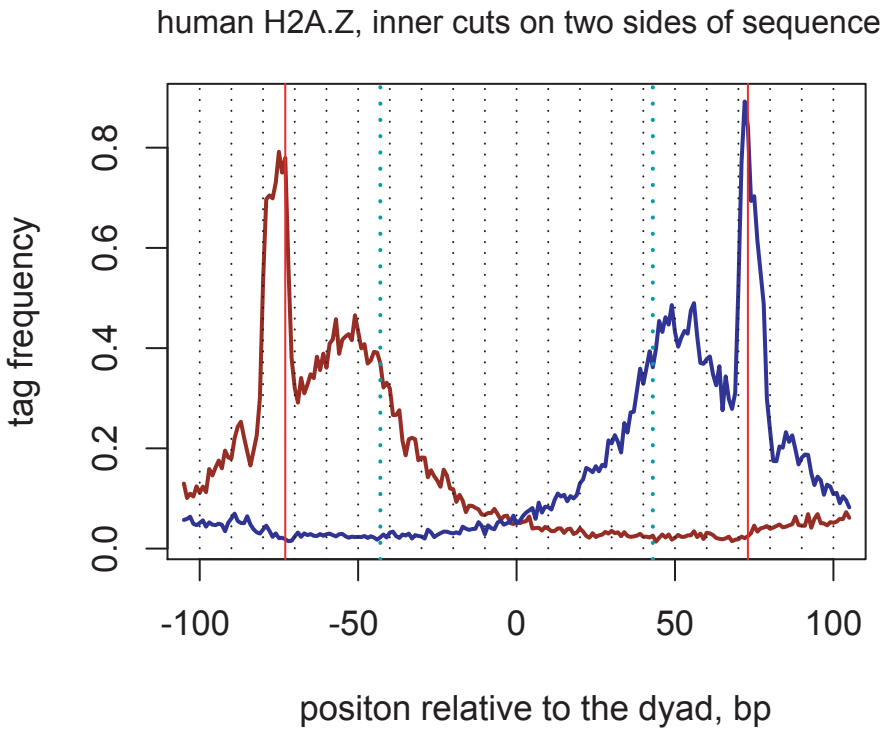

human H2A.Z, inner cuts only on one side of sequence
GC-content
0.46
0.50
0.54
0.58
-100
-50
0
50
100
positon relative to the dyad, bp

Figure S24


human H2A.Z, inner cuts on two sides of sequence
tag frequency
0.0
0.2
0.4
0.6
0.8
-100
-50
0
50
100
positon relative to the dyad, bp

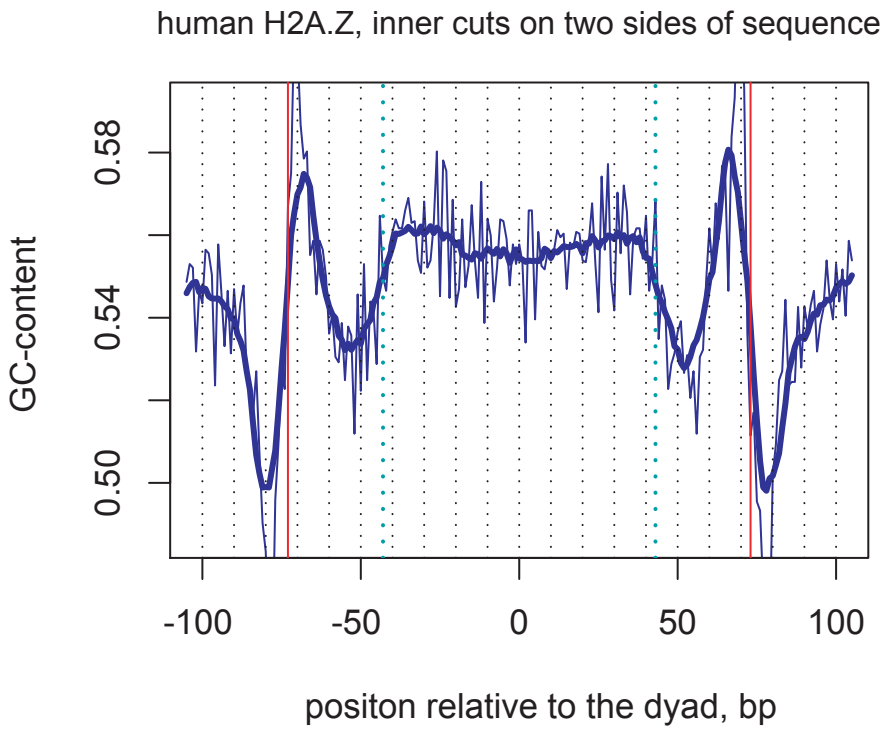

human H2A.Z, inner cuts on two sides of sequence
GC-content
0.50
0.54
0.58
-100
-50
0
50
100
positon relative to the dyad, bp

Figure S25

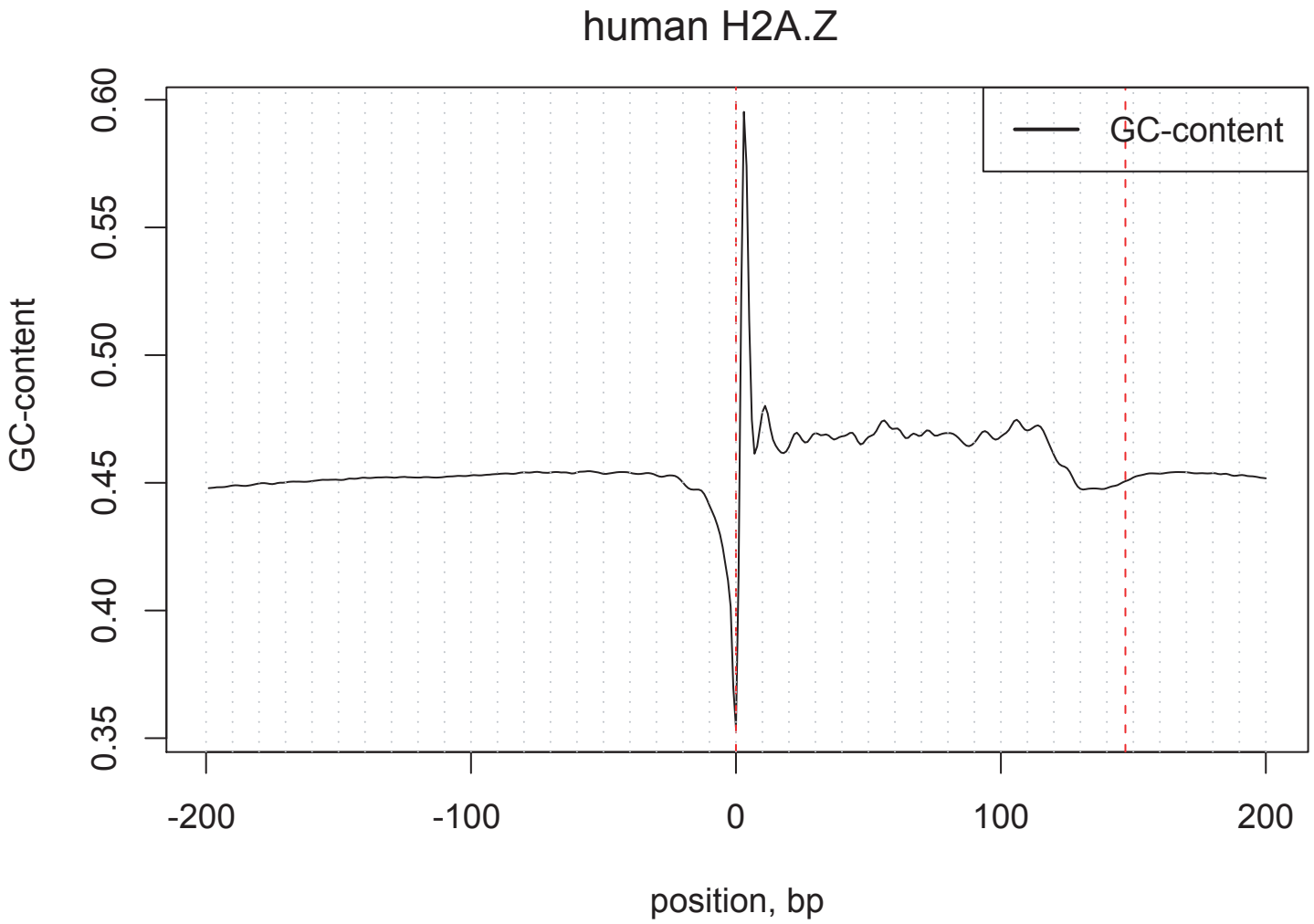

human H2A.Z
GC-content
0.60
0.55
0.50
0.45
0.40
0.35
GC-content
-200
-100
0
100
200
position, bp

Figure S26

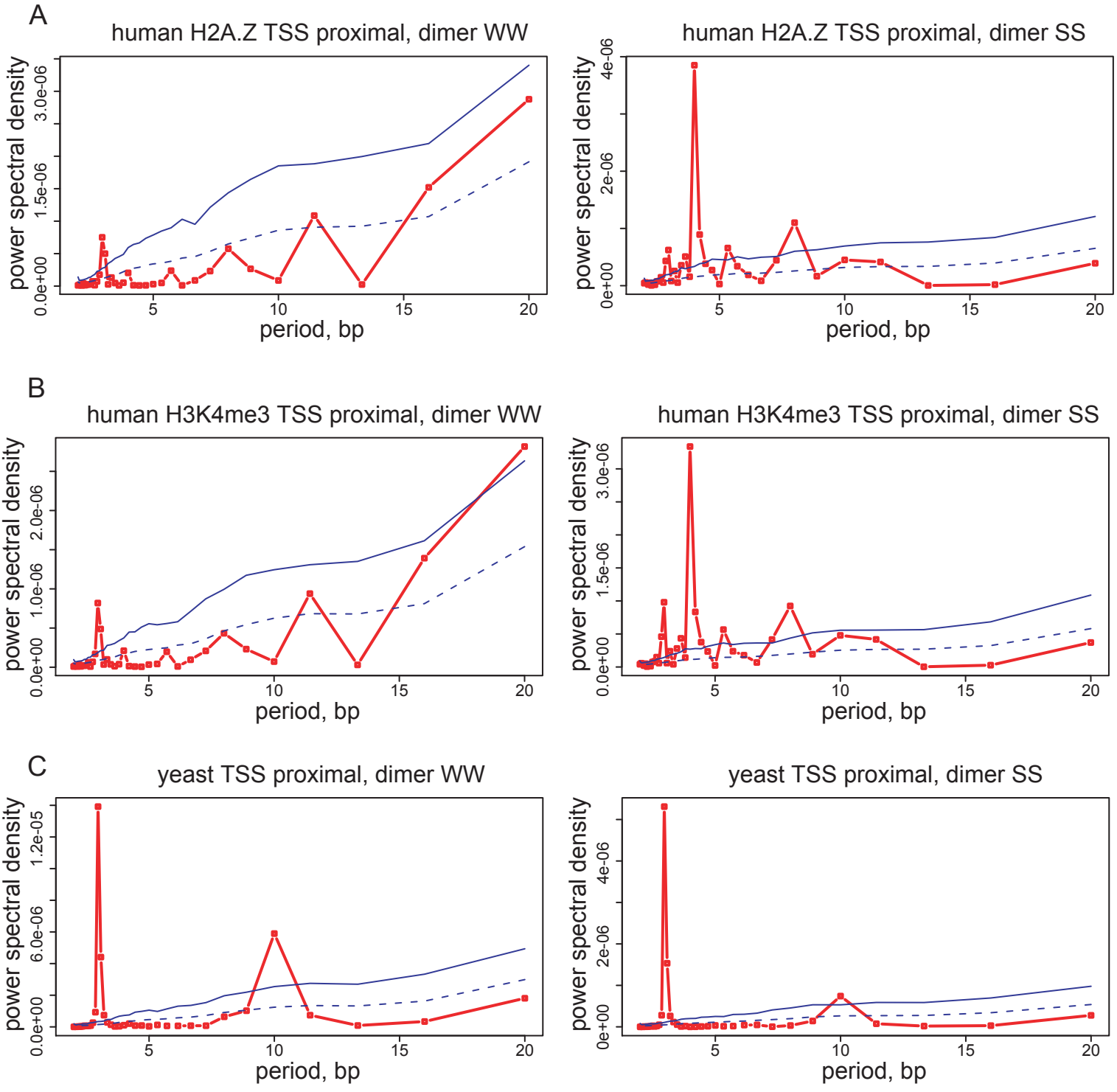
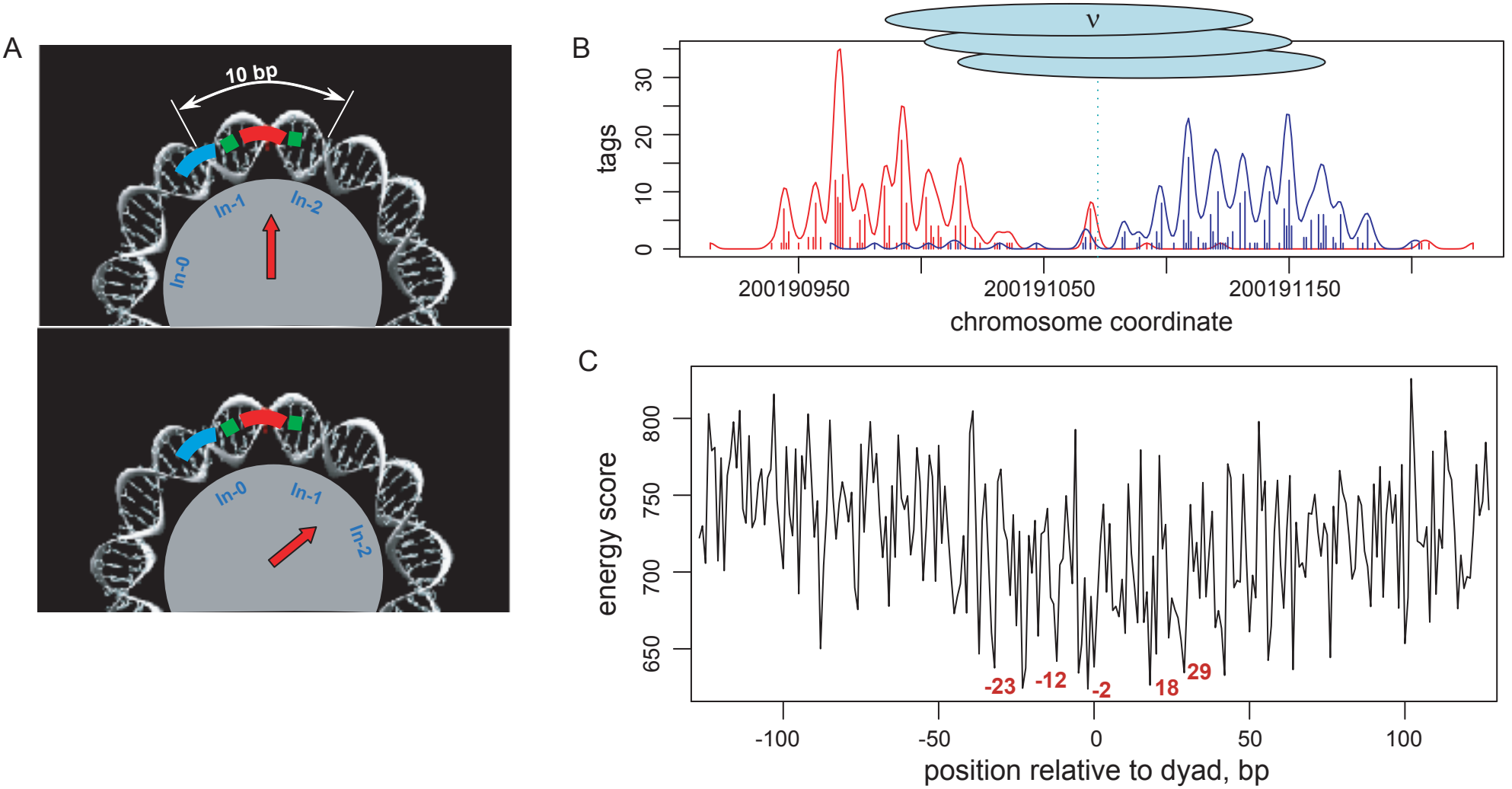

A
human H2A.Z TSS proximal, dimer WW
human H2A.Z TSS proximal, dimer SS
power spectral density
period, bp
B
human H3K4me3 TSS proximal, dimer WW
human H3K4me3 TSS proximal, dimer SS
power spectral density
period, bp
C
yeast TSS proximal, dimer WW
yeast TSS proximal, dimer SS
power spectral density
period, bp

Figure S27



D

AACCTCTGGTGACCGGGTAGTGTGACTCCGTCATTGGTAGCCCCCCGCCCACACAGCCGGGACTCGCGGCCTTTT
GAGGACGTGGGCAGGGCCAGTCTCGCCACGCCGCGGCCAATGCCCCGACCACACGCTGGCCACAGGAGTCCTG

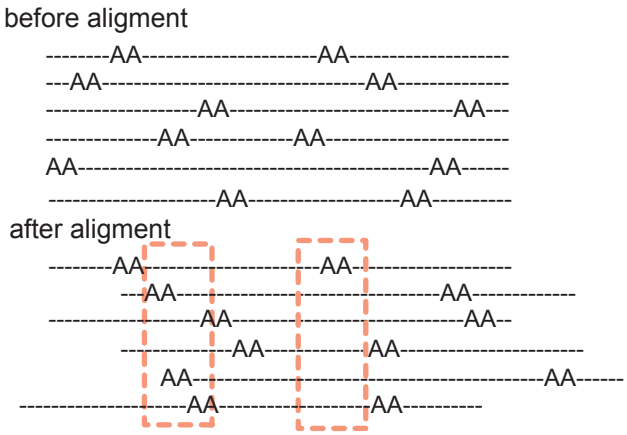
E

-60	-58	-36	-16	-7	7	17	30	36	48	58
CG	**GG**	**TA**	**CG**	**CG**	**CG**	**GC**	**CG**	**GG**	**GA**	**TG**
-6.69	-7.5	-19.4	-8.56	-7.19	-7.19	-5.3	-5.73	-9.15	-5.5	-18.6

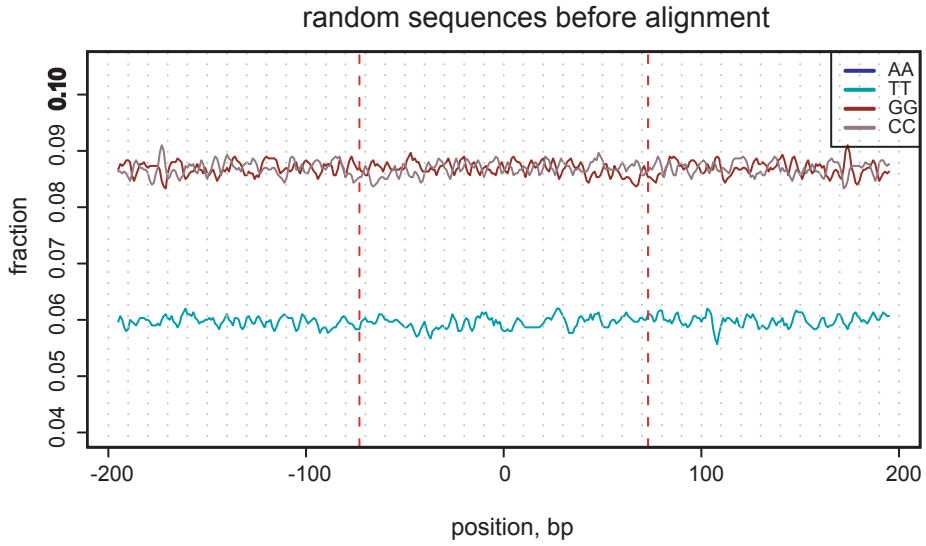
Combined energy gain at 11 steps ~ 100 au

Figure S28

A



B



C

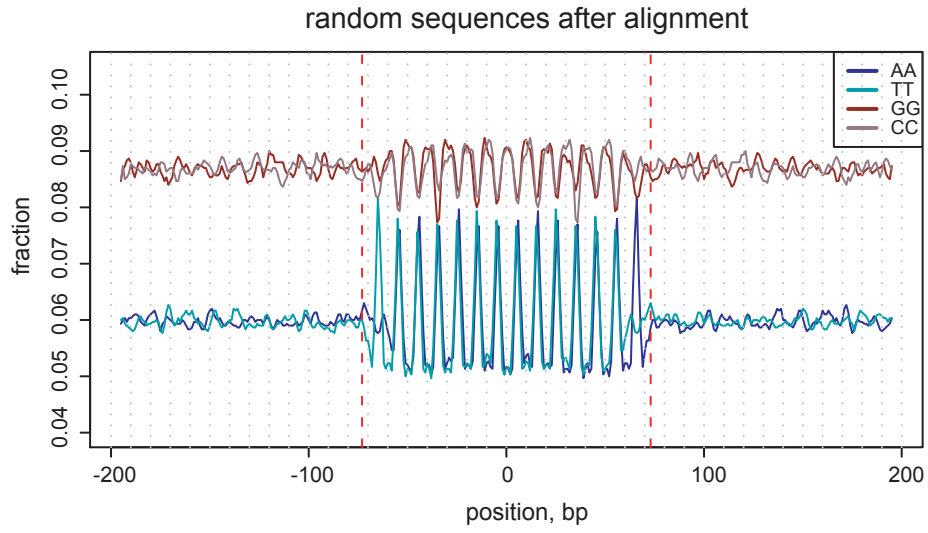


Figure S29

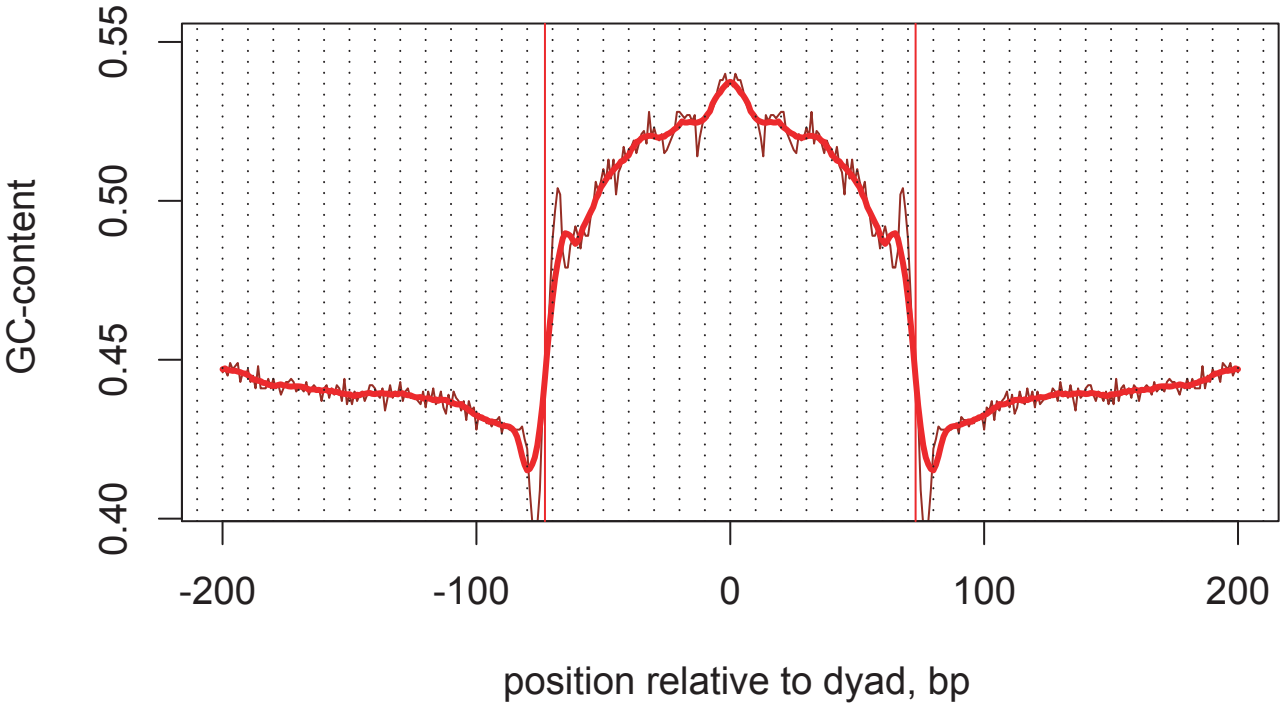

GC-content
0.40
0.45
0.50
0.55
-200
-100
0
100
200
position relative to dyad, bp

Figure S30

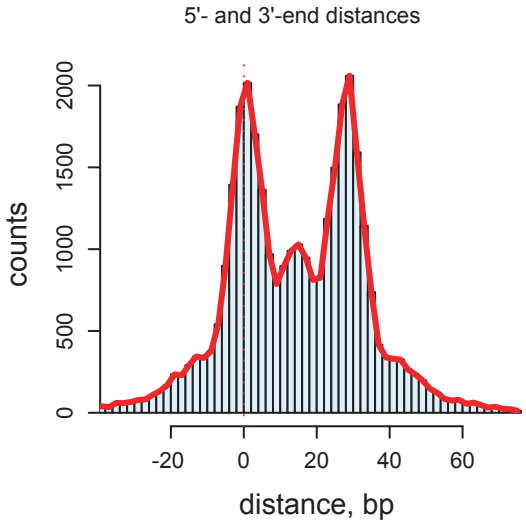
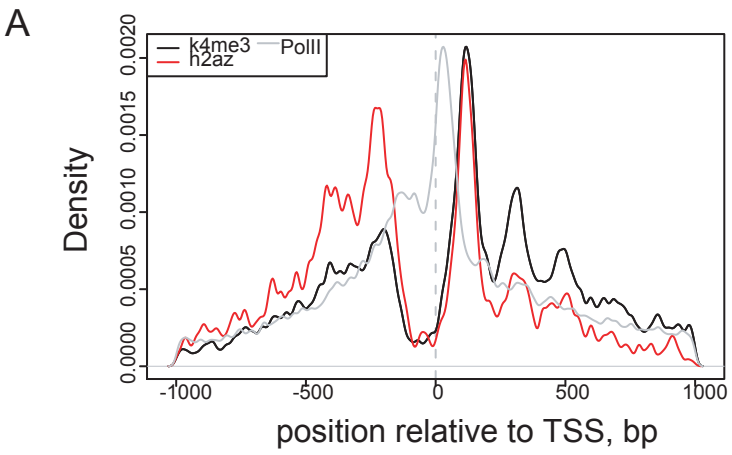
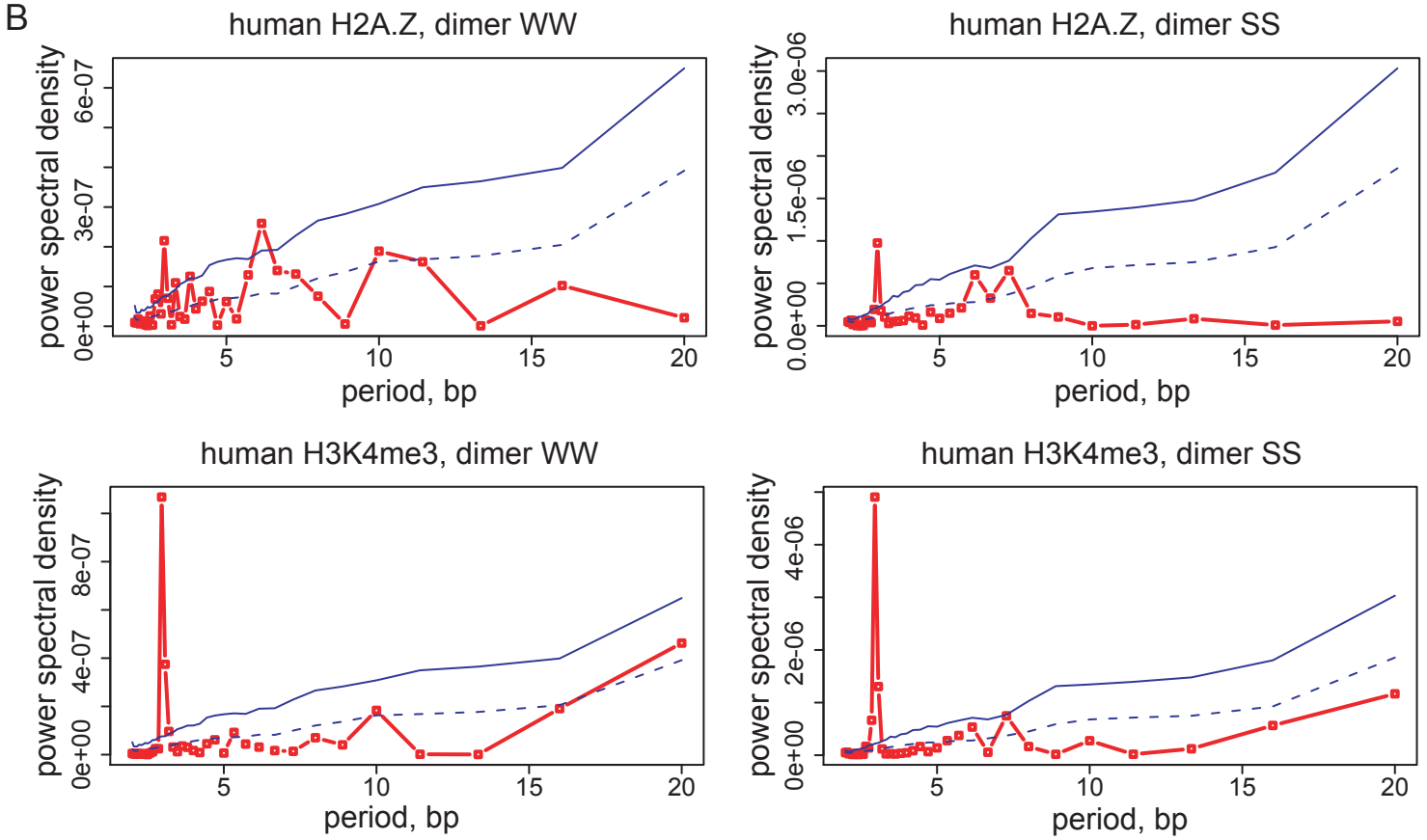

5'- and 3'-end distances
counts
0
500
1000
1500
2000
-20
0
20
40
60
distance, bp

Figure S31


A
k4me3
h2az
PolII
Density
0.0000
0.0005
0.0010
0.0015
0.0020
-1000
-500
0
500
1000
position relative to TSS, bp


B
human H2A.Z, dimer WW
power spectral density
0e+00
3e-07
6e-07
5
10
15
20
period, bp
human H2A.Z, dimer SS
0.0e+00
1.5e-06
3.0e-06
human H3K4me3, dimer WW
0e+00
4e-07
8e-07
human H3K4me3, dimer SS
0e+00
2e-06
4e-06

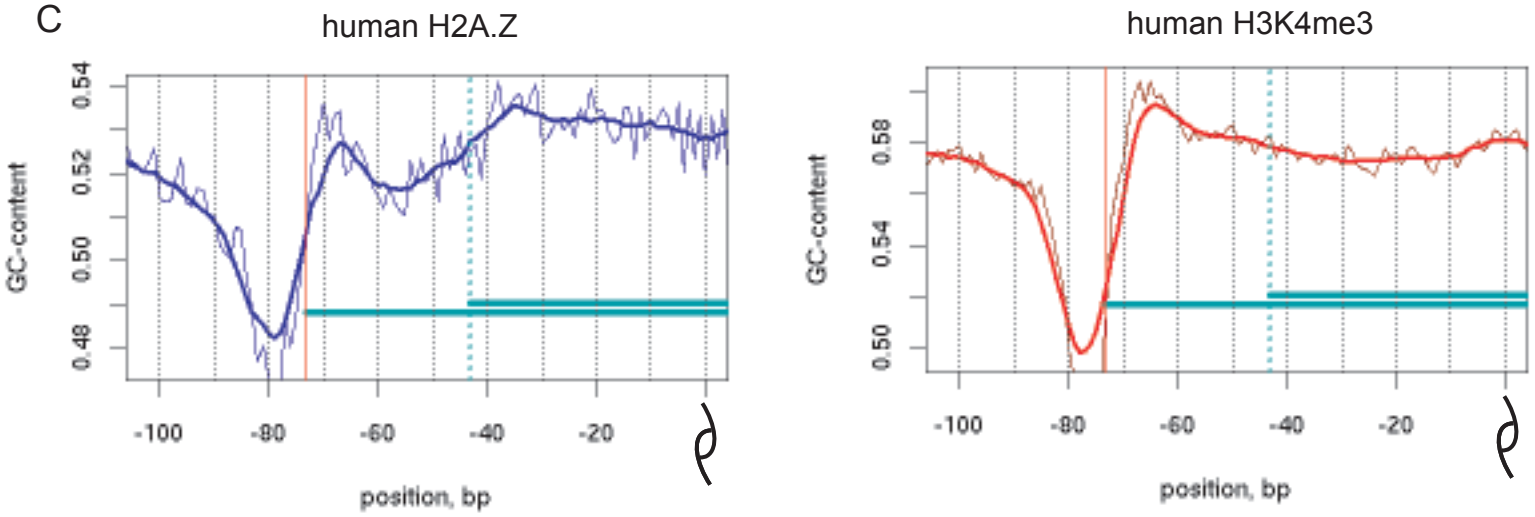

C
human H2A.Z
GC-content
0.48
0.50
0.52
0.54
-100
-80
-60
-40
-20
position, bp
human H3K4me3
0.50
0.54
0.58

Figure S32