## Supplementary Methods

### Arabidopsis nucleosome preparation

Arabidopsis nuclei were prepared as described previously (Bernatavichute et al. 2008 PLoS ONE 3: e3156) with the following modifications. After spinning the nuclei pellet at 3000 rpm for 5 min (SS-34, Sorvall), the pellet was resuspended in 5 ml of HBB buffer (25 mM Tris-Cl pH 7.6, 0.44 M sucrose, 10 mM MgCl2, 0.1% Triton-X, 10 mM beta-mercaptoethanol), applied to a 40/60% Percoll (GE Healthcare) gradient in HBB and spun 2000 rpm for 20 min (SS-34, Sorvall). Isolated nuclei were washed two times in HBB buffer and flash frozen in liquid nitrogen in HBC buffer (25 mM Tris-Cl pH 7.6, 25 mM Tris-Cl pH 7.6, 440 mM sucrose, 10 mM MgCl2 and 0.1% Triton-X, 10 mM beta-mercaptoethanol, 20% glycerol). Nuclei from 1/5 of each preparation were treated with four ul of RNAse A, 10 ug/ul, (Qiagen) and digested with 0.02 U/ul (final concentration) of Micrococcal Nuclease (Takara), 3 min. Digestion was done in digestion buffer (16 mM Tris-Cl, pH 7.6, 50 mM NaCl, 2.5 mM CaCl2, 0.01 mM PMSF and EDTA-free protease inhibitor cocktail (Roche)) and stopped with 10 mM EDTA and treated with Proteinase K (Roche). DNA was purified either with phenol:chlorophorm extraction and precipitated with salts and ethanol or with ChIP DNA Clean and Concentrator kit (Zymo Research). Purified DNA was run on 2% agarose gel and bands corresponding to ~150 bp were cut and purified with a Gel Purification kit (Qiagen). We cloned and performed traditional DNA sequencing of digested fragments to check the quality of the library. The average length of 55 fragments was 146 nucleotides.

### RNA-Seq expression

Plants were grown in a climate-controlled growth chamber with a photoperiod of 11 h light, 13 h dark, at a photon efflux density of 145 µmol $m^{-2}$ $s^{-1}$ during the day, and day and night

temperatures of 20 ºC and 18 ºC, respectively. 6-week-old plants were harvested for RNA extraction.

To generate mRNA expression profiles, total RNA was isolated from 100-mg subsamples of tissues. RNase-free DNase set for on-column DNase digest and RNeasy spin columns (Qiagen, Hilden, Germany) were used according to the manufacturer's instructions for subsequent digestion of genomic DNA and RNA purification.

Ten µg of total RNA per sample were used to generate the cDNA Colony Template Libraries (CTL) for high-throughput DNA sequencing using Illumina technology according to the Whole Transcriptome Shotgun protocol (Fasteris SA, Genome Analyzer Service, Switzerland). CTLs were sequenced to obtain a library of 36 base-pair long reads.

**H3-Chip-seq**

Chromatin was immunoprecipitated using an antibody against histone H3 as previously described (Bernatavichute et al. 2008 PLoS ONE 3: e3156). Libraries were constructed and DNA was sequenced according to manufacturer instructions (Illumina).

**Nucleosome predictions**

Nucleosome predictions were calculated using a previously described algorithm (Kaplan et al. 2009 Nature 458: 362-366). The predictions were made on the TAIR7 version of the *Arabidopsis* genome. The probability scores for nucleosome start sites were used for further analysis.

**Human DNA methylation**

The HSF1 hESC line was cultured as described previously (Xiao et al. 1999 Embo J 18: 5943-5952). BS-Seq DNA library construction, high-throughput Illumina sequencing by and analysis were performed similar to that previously described (Cokus et al. 2008 Nature 452: 215-219). 684 million 47 nucleotide reads were used in the analysis of DNA methylation of nucleosomal DNA.

**Data analysis**

Nucleosome-seq and Chip-seq and were mapped using Bowtie (Langmead et al. 2009 Genome Biol 10: R25) with a tolerance of 3 mismatches. mRNA-Seq sequence reads were aligned against the set of transcript sequences from the TAIR v7.0 database using SOAP (Li et al. 2008 Bioinformatics 24: 713-714). A tolerance of up to two-mismatches was employed and reads with more than one match to the reference sequences were eliminated. Genes with more than one isoform were collapsed into a single gene. The total number of alignment hits was used as an estimate for gene expression levels. Because cDNA synthesis was performed using oligo dT priming before random fragmentation, we chose not to normalize transcript count estimates by the gene length. Instead, genes under 1000 bases were discarded due to disproportionate representation in the expression data.

## Supplementary Figures

### Supplementary Figure 1 | Chromosomal views of nucleosome read distributions.

**a-d,** A chromosomal view of nucleosome read counts and ChIP-seq read counts in 100kb bins along chromosome 2-5 shows nucleosome enrichment in pericentromeric regions.

### Supplementary Figure 2 | Relationship between forward and reverse stranded reads, and between forward stranded reads and other forward stranded reads.

**a,** The starts of nucleosome reads on the reverse strand peaked 148 bases away from nucleosome reads on the forward strand. **b,** A zoomed in version of **Figure 1c** is provided to demonstrate a ten base periodicity. The 10 base periodicities in both panels is likely due to the rotational setting tendencies of the histone/DNA interface (Albert et al. 2007 Nature 446: 572-576). Because of a 10 base periodicity in some dinucleotides (**Supplementary Figure 3**), anything but a 10 base translational shift would result in a change in the bases that are accessible to histones and would therefore be disfavored.

### Supplementary Figure 3 | Dinucleotide profiles across the nucleosome.

**a-f,** The fraction of each dinucleotide is plotted as a function of the distance from the start of nucleosome reads. AA **(a)**, CT **(b)**, CC **(c)**, AC **(d)**, TT **(e)**, and CG **(f)** dinucleotide plots are shown. For clarity, 10 bases around the beginning of reads have been omitted since these may reflect sequence biases of the micrococcal nuclease enzyme.

### Supplementary Figure 4 | FFTs of dinucleotide profiles

**a-f,** FFTs are calculated for each of the dinucleotide profiles in Supplementary Figure 3. AA **(a)**, CT **(b)**, CC **(c)**, AC **(d)**, TT **(e)**, CG **(f)**, all show a period of 10 bases over the nucleosome-bound region.

### Supplementary Figure 5 | AA/TA/TT dinucleotide frequency in Arabidopsis vs. in Saccharomyces

Arabidopsis AA/TA/TT dinucleotide frequency is compared to that in Saccharomyces (Segal et al. 2006 Nature 442: 772-778).

### Supplementary Figure 6 | CHG and CHH methylation occurs in phase with AA dinucleotides.

**a,b,** As with CG methylation (Figure 1D), CHG **(a)** and CHH **(b)** methylation profiles are also in phase with AA dinucleotides and out of phase with GC dinucleotides. Lines were drawn to indicate peaks in AA dinucleotides. **c-e,** For additional reference, CC dinucleotides are plotted along with CG methylation **(c)**, CHG methylation **(d)**, and CHH methylation **(e)**.

### Supplementary Figure 7 | CG, CHG and CHH methylation vs CG, CHG, and CHH sequences.

CG methylation **(a)**, CHG **(b)** and CHH **(c)** methylation profiles are plotted along with CG dinucleotides, CHG trinucleotides, and CHH trinucleotides. Lines were drawn to indicate peaks in AA dinucleotides.

**Supplementary Figure 8 | Methylation profiles at H3 ChIP-Seq reads.**

As an independent confirmation of the enrichment of methylation on nucleosomes, chromatin immunoprecipitation was performed using an antibody against unmodified histone H3 and the resulting fragments were sequenced and mapped. **a-c,** The weighted average percent CG **(a)**, CHG **(b)**, and CHH **(c)** DNA methylation was calculated and plotted at each distance from ChIP-seq read start sites (0).

**Supplementary Figure 9 | DNA methylation profiles of nucleosome-bound DNA in Arabidopsis promoters.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in promoter regions - defined as 5 kb upstream of TSSs - were considered for this analysis.

**Supplementary Figure 10 | DNA methylation profiles of nucleosome-bound DNA in Arabidopsis genes.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in genes were considered for this analysis.

**Supplementary Figure 11 | DNA methylation profiles of nucleosome-bound DNA in Arabidopsis repeats.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in repetitive regions were considered for this analysis. Tandem repeats, inverted repeats, dispersed repeats, and telomeric regions were included.

**Supplementary Figure 12 | DNA methylation profiles of nucleosome-bound DNA in Arabidopsis euchromatic regions.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in euchromatic regions were considered for this analysis.

**Supplementary Figure 13 | DNA methylation profiles of nucleosome-bound DNA in Arabidopsis centromeric regions.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in centromeric regions were considered for this analysis.

**Supplementary Figure 14 | DNA methylation profiles of nucleosome-bound DNA in Human promoters.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in promoter regions - defined as 5 kb upstream of TSSs - were considered for this analysis.

**Supplementary Figure 15 | DNA methylation profiles of nucleosome-bound DNA in Human genes.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in genes were considered for this analysis.

**Supplementary Figure 16 | DNA methylation profiles of nucleosome-bound DNA in Human repeats.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes in repeats were considered for this analysis.

**Supplementary Figure 17 | DNA methylation profiles of nucleosome-bound DNA in Human CpG islands.**

The weighted average percent DNA methylation was calculated and plotted at each distance from nucleosome start sites (0). **a, b, c,** CG methylation **(a)**, CHG methylation **(b)**, and CHH methylation **(c)** each show a 10 base periodicity over nucleosome-bound DNA (1-147 bases). **d, e, f,** The FFTs calculated over the region of the nucleosome demonstrate this periodicity in CG **(d)**, CHG **(e)**, and CHH **(f)** contexts. Only nucleosomes within 1 kb of CpG islands were considered for this analysis.

**Supplementary Figure 18 | Nucleosome and Pol II levels in Exons.**

We performed chromatin immunoprecipitation with an antibody against RNA polymerase II (Pol II), and hybridized the resulting DNA to a whole genome Affymetrix microarray. We normalized these data to randomly sheared genomic DNA to control for probe efficiencies. Nucleosome midpoints and Pol II levels are plotted over the intron-exon boundary.

**Supplementary Figure 19 | Nucleosome enrichment over exons is GC independent and not due to consensus site sequences.**

**a-c,** Exons were categorized based upon different GC content constraints to address the possibility that GC content of the DNA alone could explain nucleosomal positioning at the intron-exon boundary, as exons are known to have a higher GC content than introns. **a,** Exons were scored based on their GC content and separated into groups - low GC (bottom 33%) and high GC (top 33%). **b,** Exons were scored based on the GC content of the immediately preceding intron and similarly segregated into low and high GC groups. **c,** Finally, exons were scored by the change in GC content over the intron-exon boundary, and similarly segregated. **a-c,** In each panel, the mean GC content in introns of the low and high groups is indicated on the left, and the mean GC content in exons of the low and high groups is indicated on the right. In each case, the nucleosome profile at the boundary retains its general shape, suggesting that GC content alone cannot explain the nucleosome patterning. **d,** A nucleosome prediction algorithm was used to profile the likelihood of nucleosomes starting across the intron-exon boundary. The profiles for unmodified exon boundaries and boundaries where the consensus site is randomized are very similar, suggesting that the consensus splice site is not necessary to position nucleosomes. **e,** The consensus site is shown, with the arrow representing the start of the exon. **f,** The same loci are shown, but where the nucleotides were randomized.

**Supplementary Figure 20 | Microarrray confirmation of nucleosome profile at intron-exon boundaries.**

**a-c,** We also sought to determine whether the enrichment of nucleosomes in exons relative to introns could be an artifact of our technique of isolating mononucleosomes after MNase digestion. We analyzed our previously published dataset in which we immunoprecipitated formaldehyde crosslinked and sonicated chromatin with antibodies against the unmodified C terminus of histone H3 **(a)**, H3K4me1 **(b)**, or H3K27me3 **(**c**)**, and hybridized the resulting DNA to a whole genome Affymetrix tiling array (Zhang et al. 2007 PLoS Biol 5: e129; Zhang et al. 2009

Genome Biol 10: R62). We normalized this data with control DNA from crosslinked and sonicated input chromatin that had not undergone immunoprecipitation, to control for probe hybridization efficiencies. H3 **(a)**, H3K4me1 **(b)**, and H3K27me3 **(c)** ChIP-chip signals over the intron-exon boundary confirm results presented in Figure 4a.

**Supplementary Figure 21 | Expression independence of nucleosome profile at intron-exon boundary.**

**a,** To test if active transcription or splicing is causing the nucleosomes to become phased on exons, we categorized exons into two groups: highly and poorly transcribed groups. The highly transcribed group refers to those top 1/3rd exons that were derived from genes with the highest RNA-seq transcript counts. The poorly transcribed group refers to those exons that were derived from genes with no RNA-seq transcript counts. The two profiles are very similar, suggesting that nucleosome positioning signals at the exon are not strongly dependent on transcription or splicing and are instead likely to be hard coded in the exonic DNA sequences.

**Supplementary Figure 22 | Predicted nucleosomes at the intron-exon boundary**

**a-c,** Exons were categorized in the manner described in Supplementary Figure 9 and predicted nucleosome start sites were plotted around exons in each of these categories. **d-h** Exons were size selected as described in Figure 4; predicted nucleosome start sites and MNase nucleosome start sites were plotted over these exons.

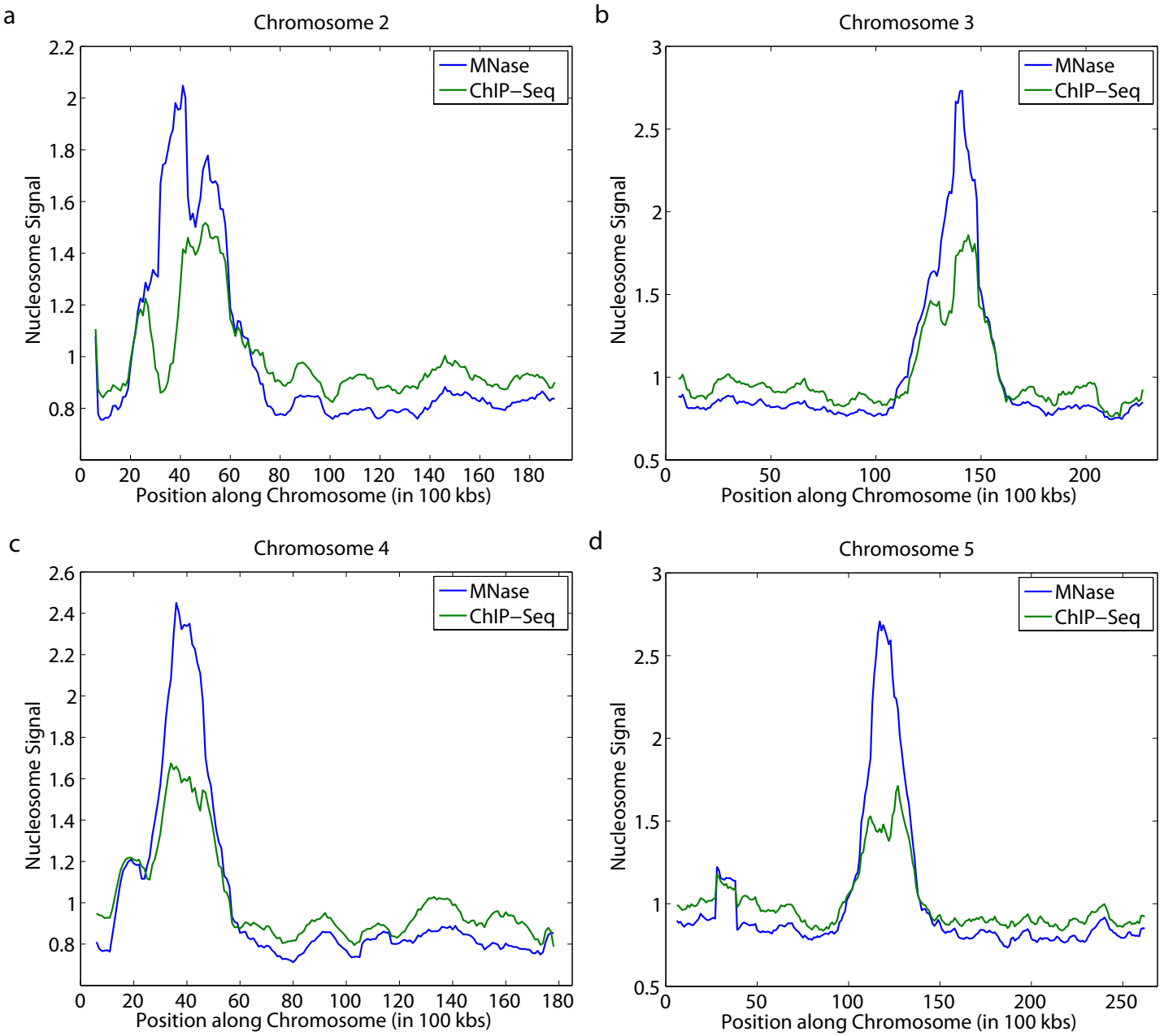
**Supplementary Figure 23 | Methylation profiles of the highest scoring and lowest scoring predicted nucleosome start sites.**

**a-f**, Predicted nucleosomes were separated into high and low scoring subsets. CG **(a, d)**, CHG **(b, e)**, and CHH **(c, f)** methylation was profiled across these sites. These results confirm our findings from MNase data. Moreover, high scoring nucleosomes have higher CG methylation over nucleosomes and low scoring nucleosomes have lower CG methylation over nucleosomes.

**Supplementary Figure 24 | Chromosomal view of predicted nucleosome start sites**

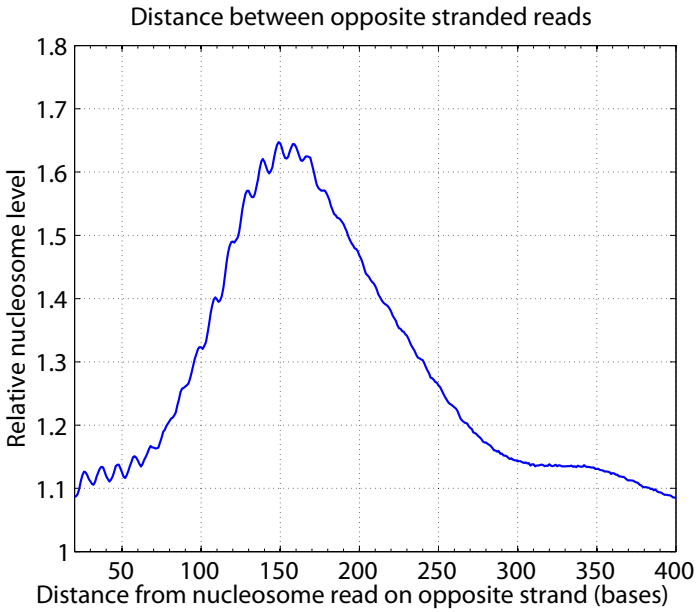
**a-e** Predicted nucleosome start sites in 100 kb bins along chromosomes 1 **(a)**, 2 **(b)**, 3 **(c)**, 4 **(d)**, and 5 **(e)** are shown.

Supplementary Figure 1.

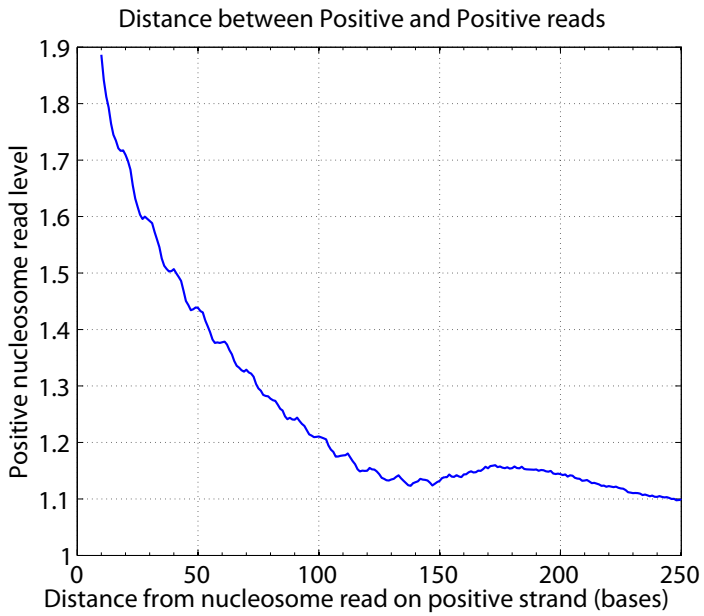


Supplementary Figure 2.

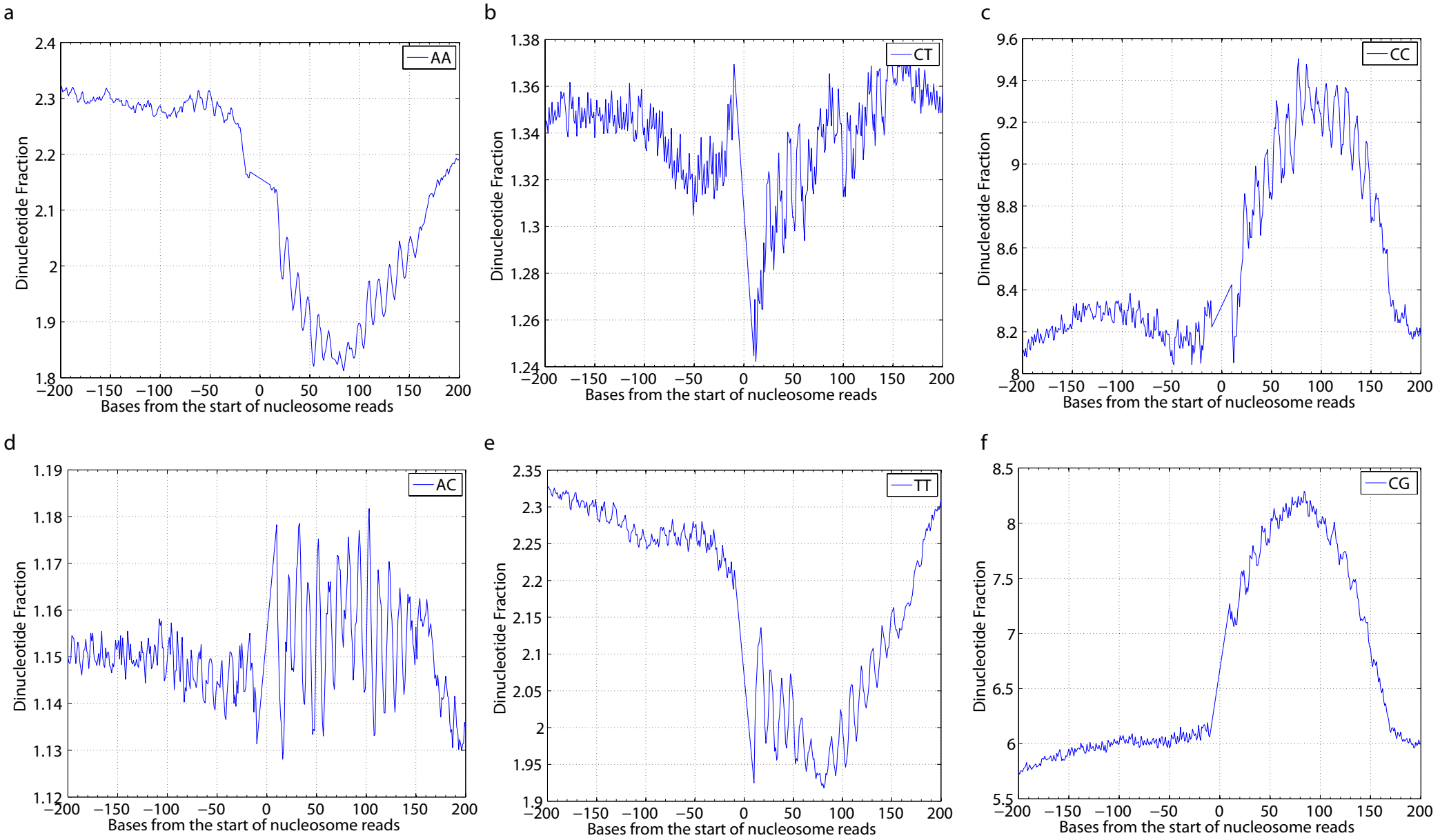
a



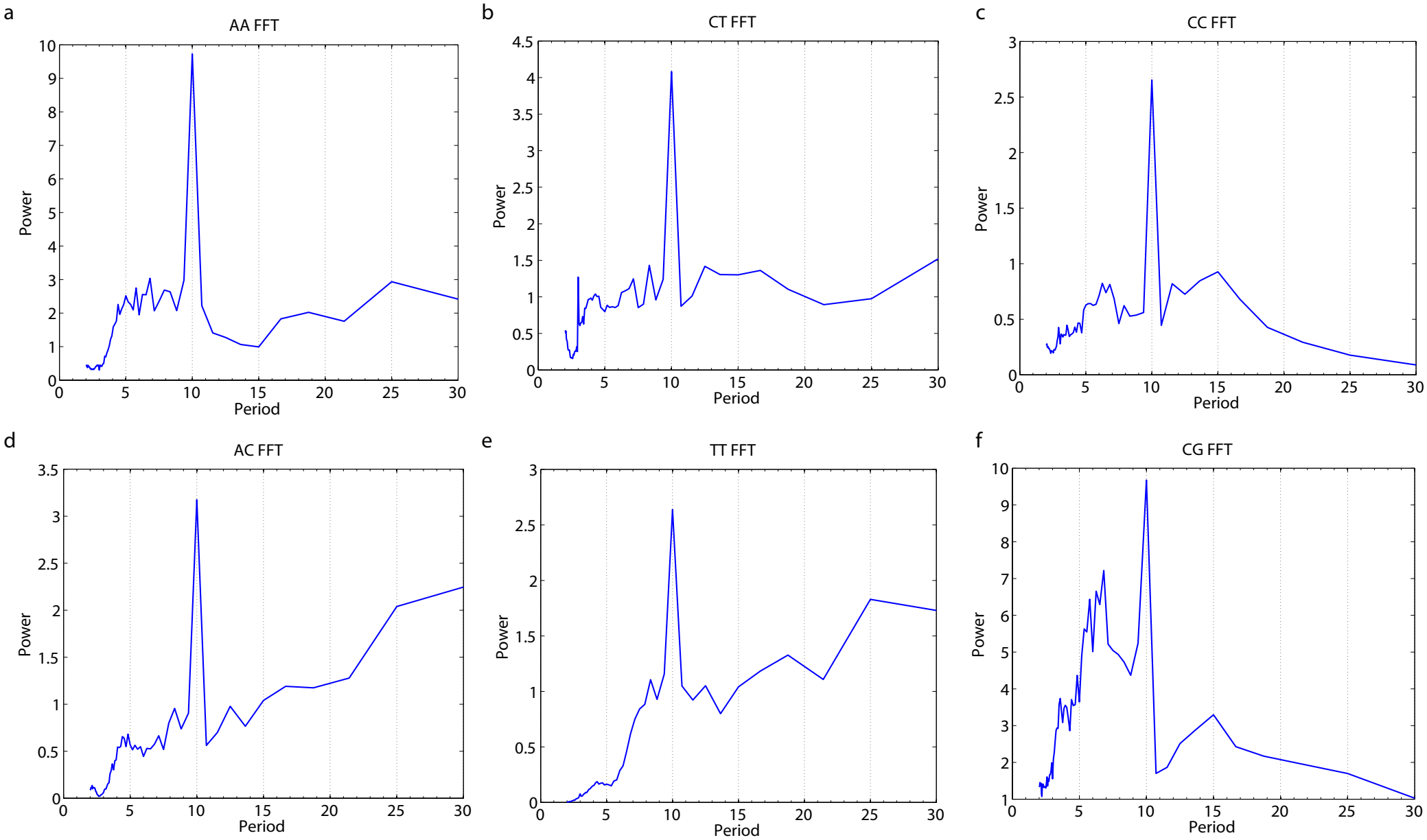

b



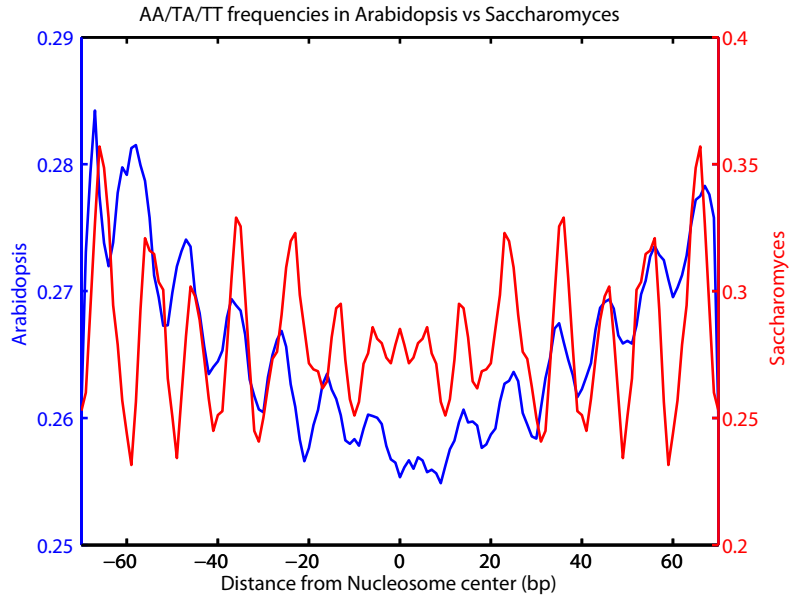
Supplementary Figure 3.



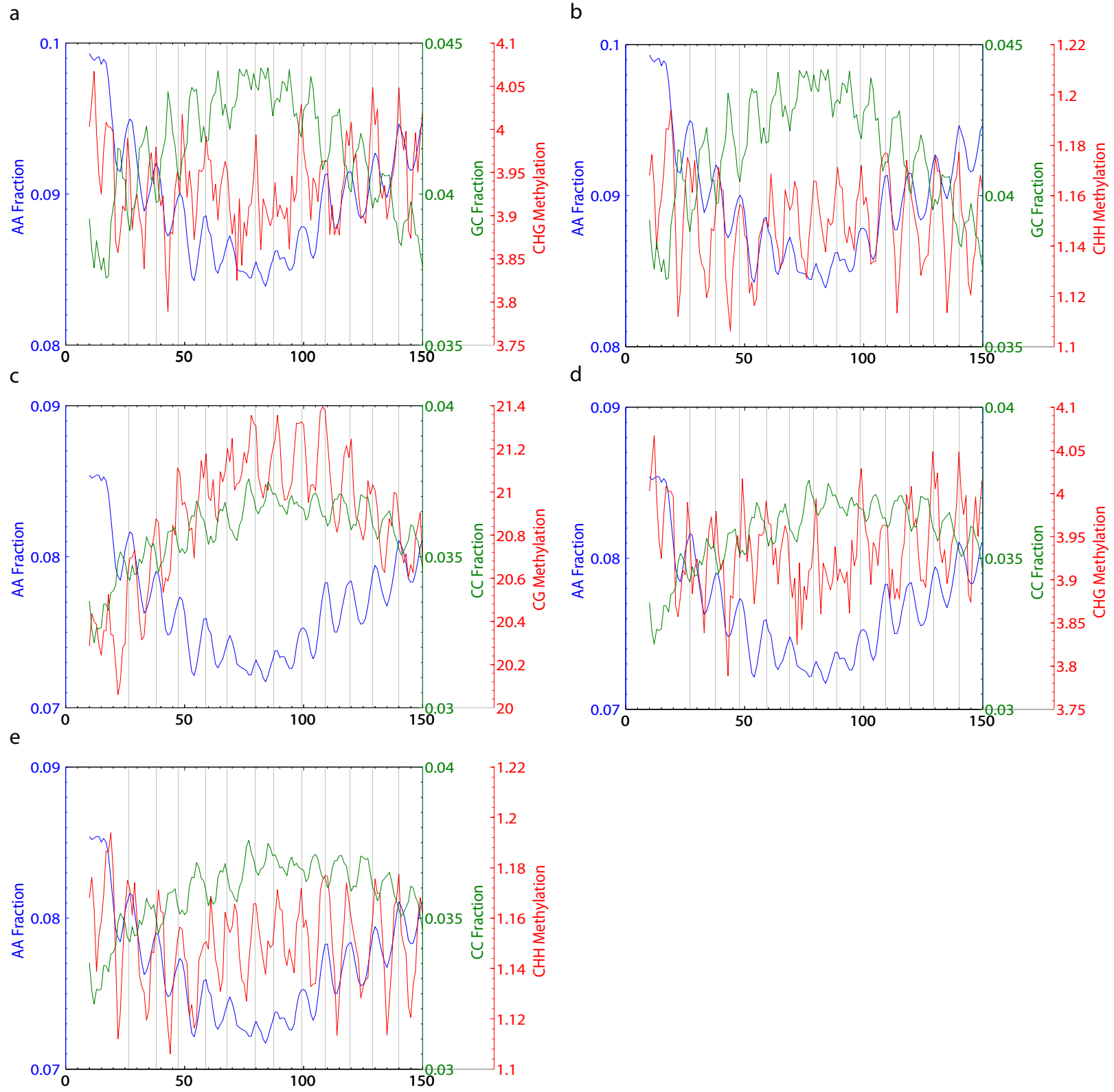
Supplementary Figure 4.



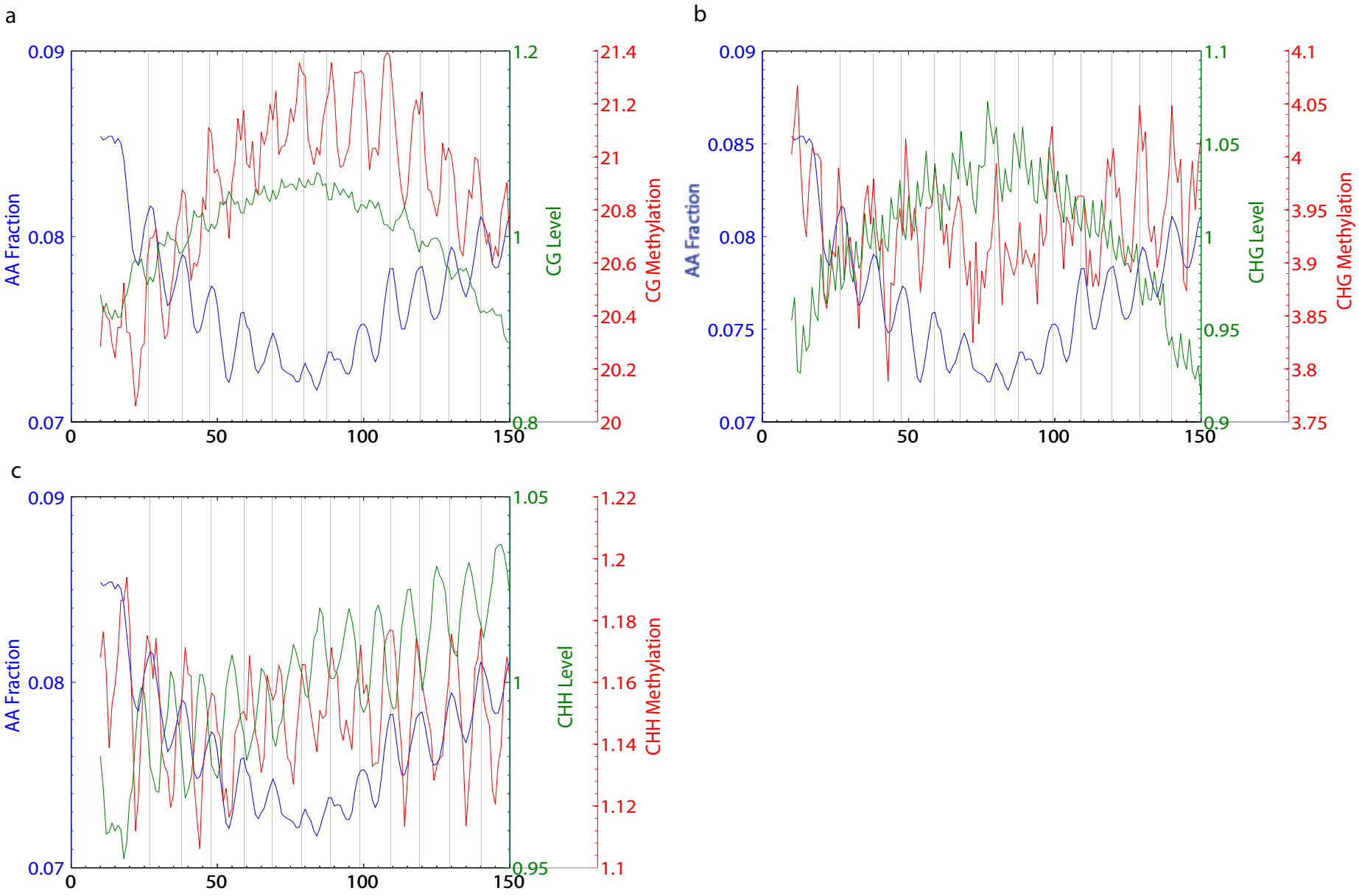
Supplementary Figure 5.



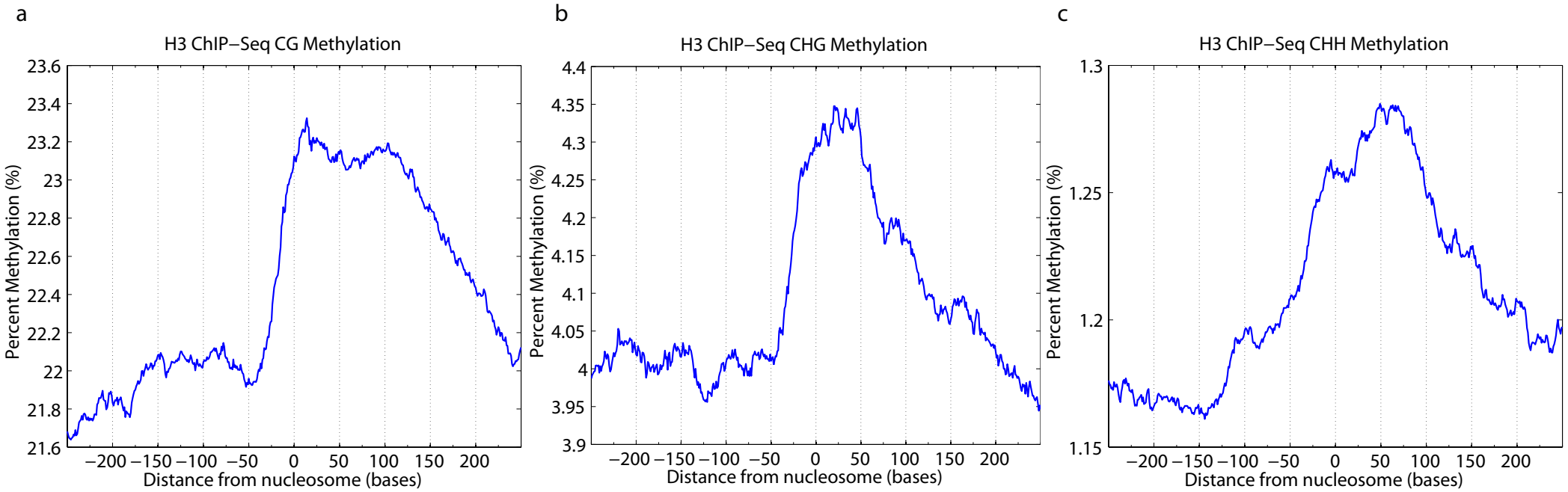
Supplementary Figure 6.



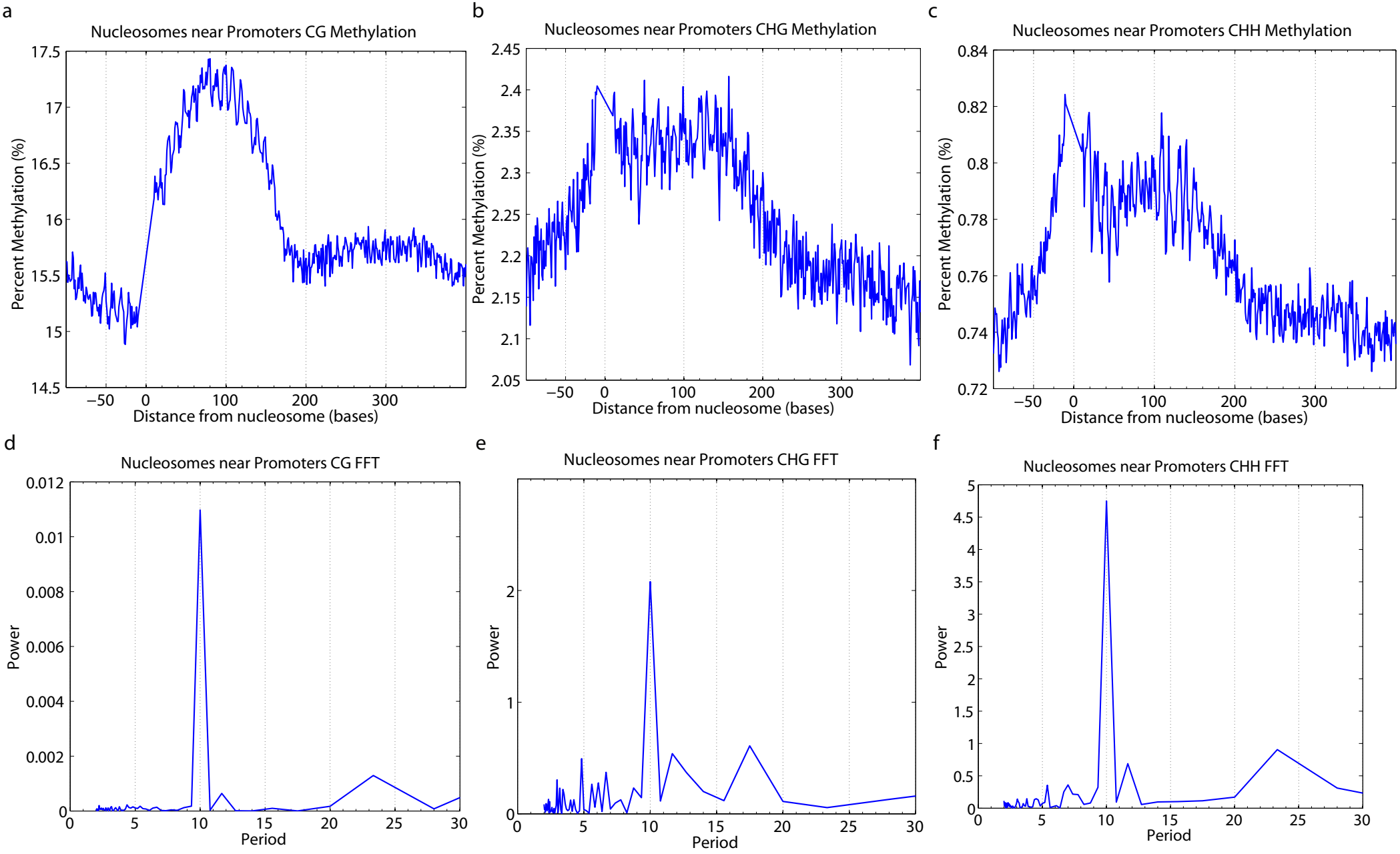
Supplementary Figure 7.



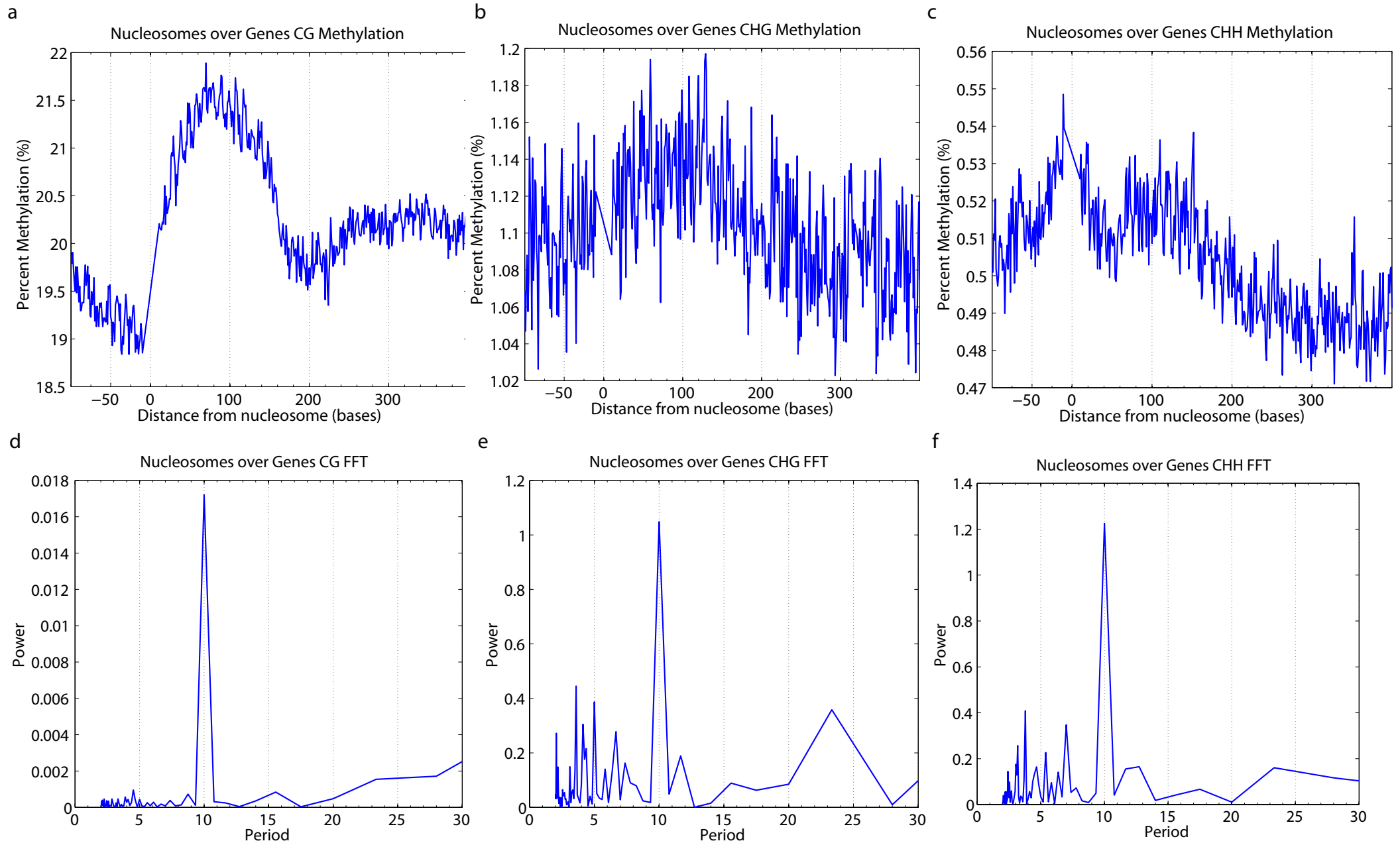
Supplementary Figure 8.



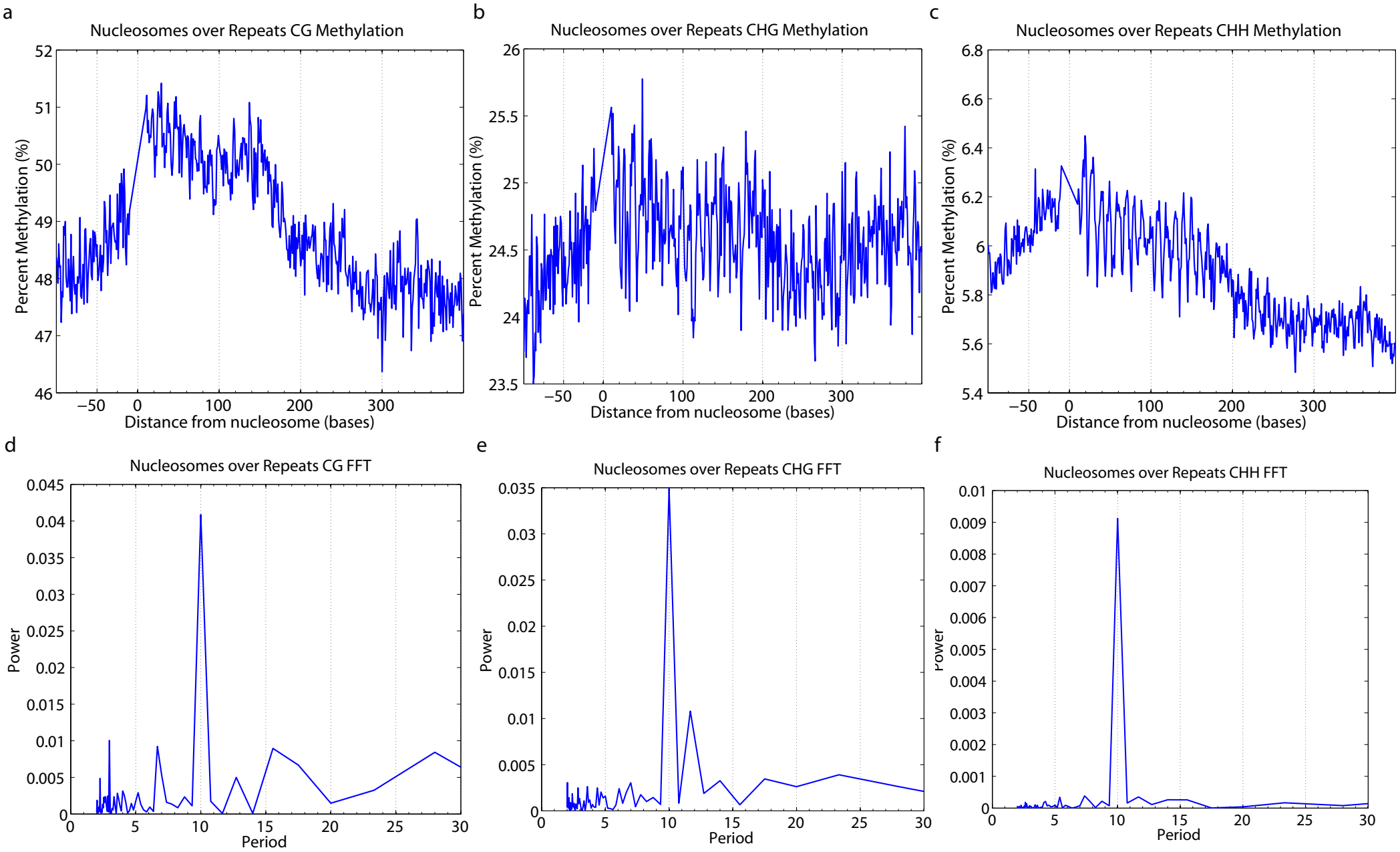
Supplementary Figure 9.



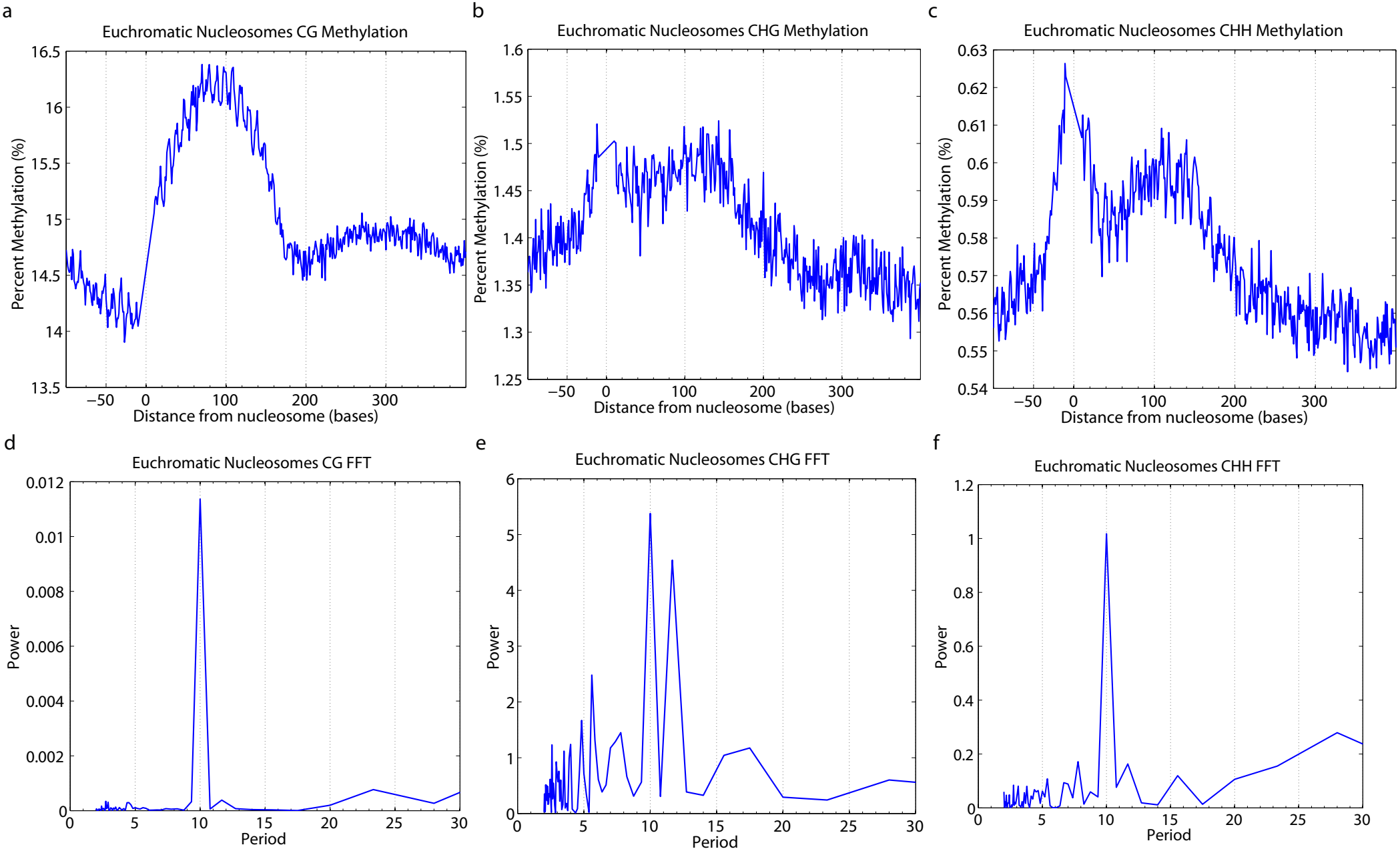
Supplemental Figure 10.



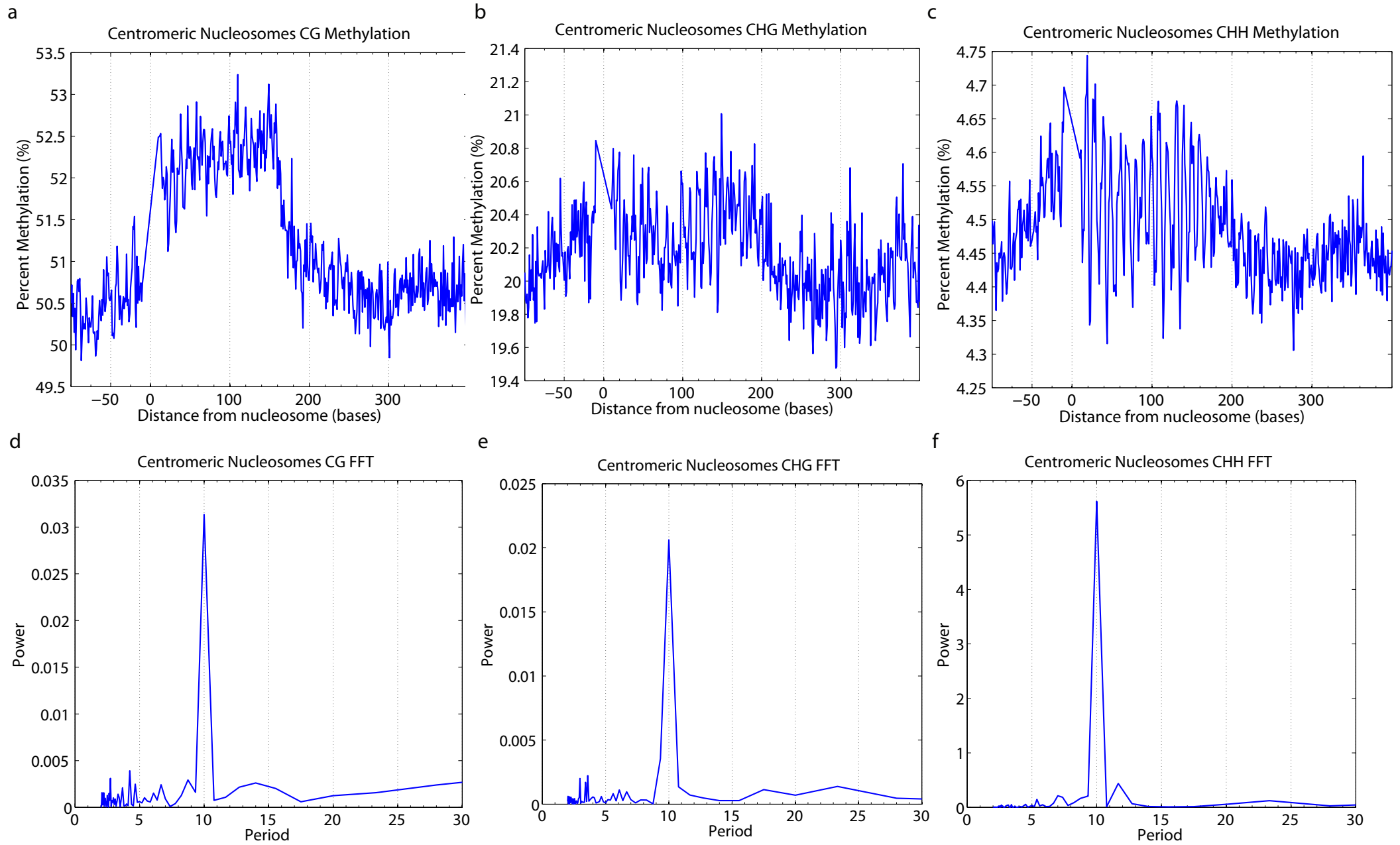
Supplemental Figure 11.



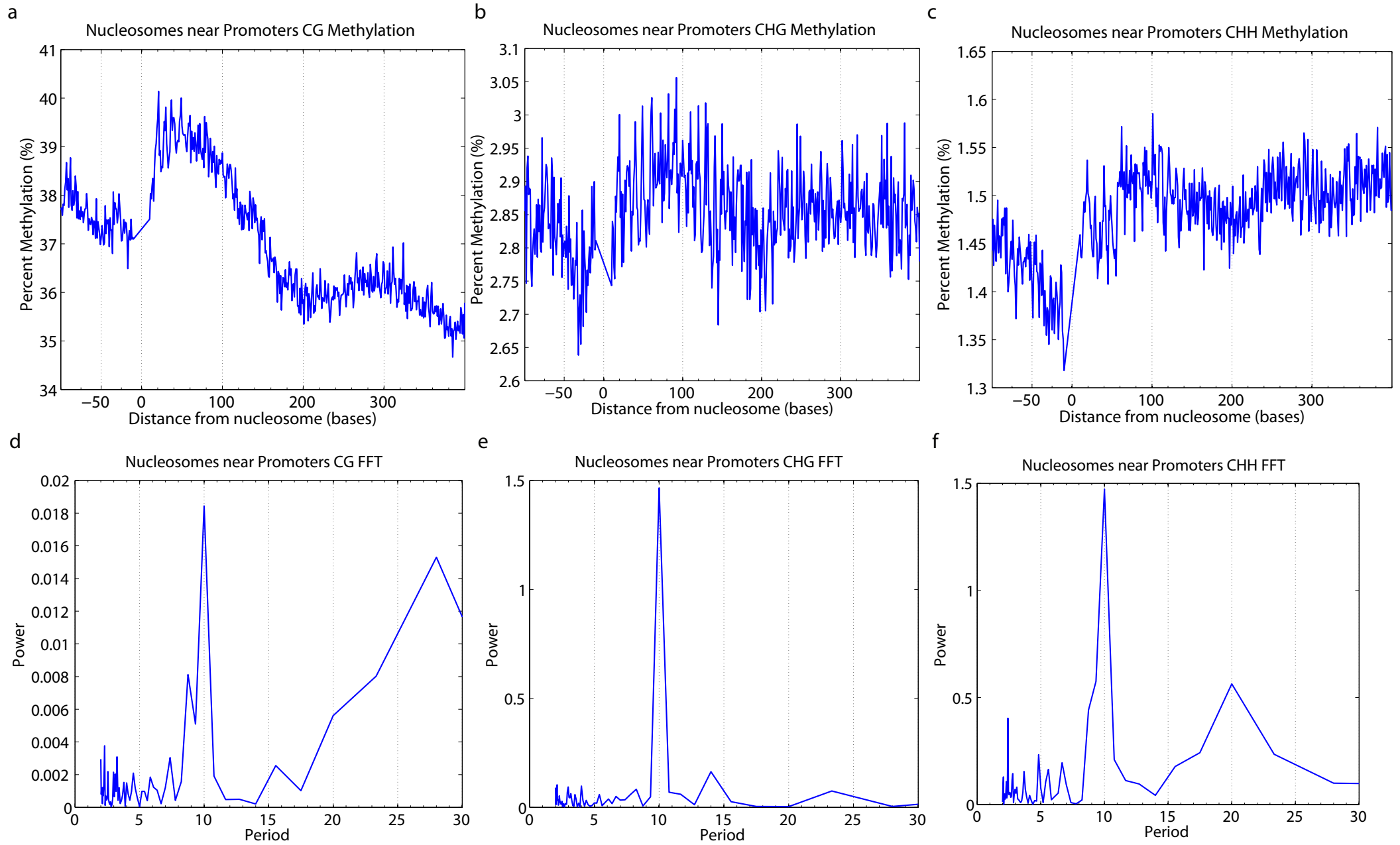
Supplemental Figure 12.



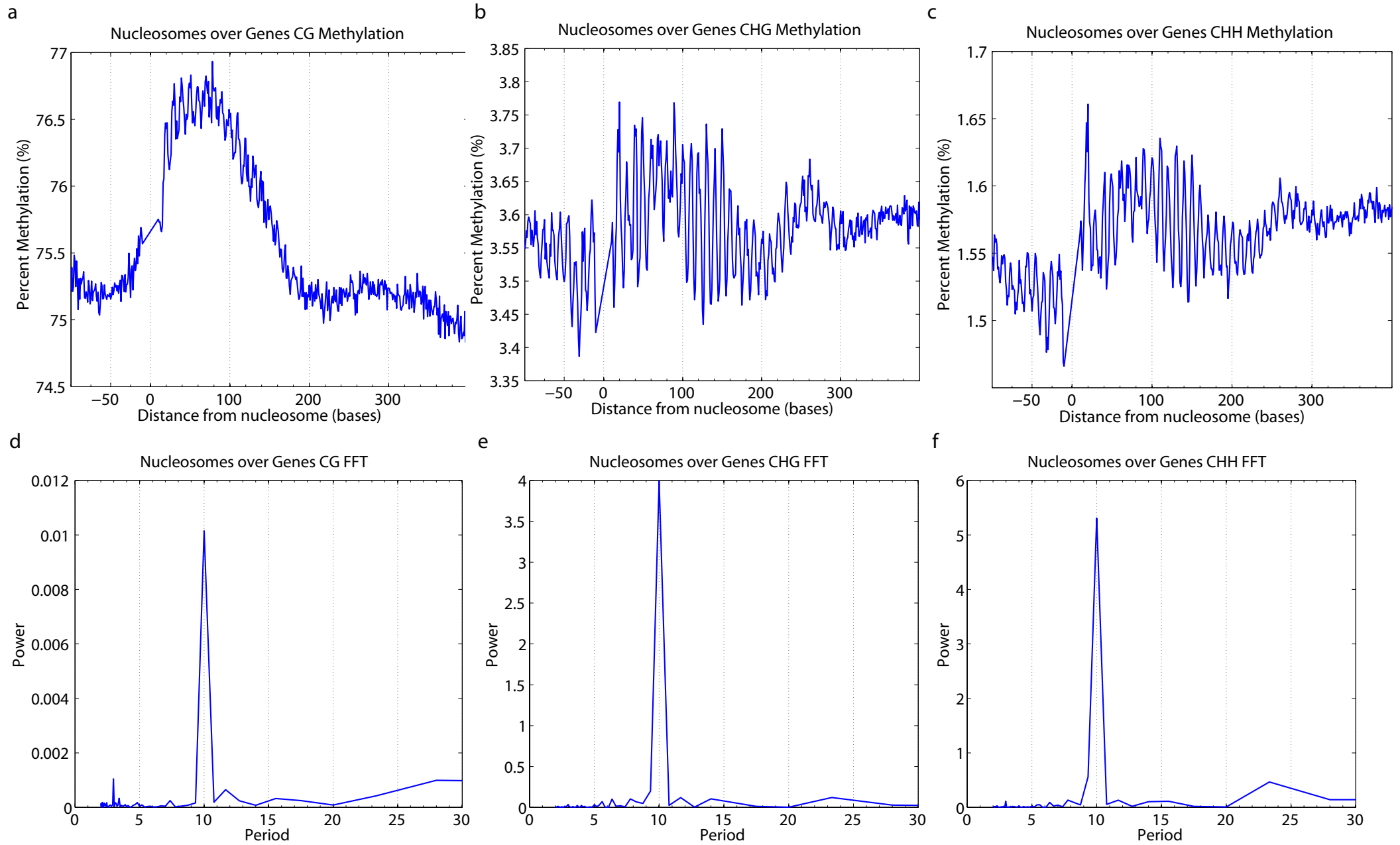
Supplemental Figure 13.



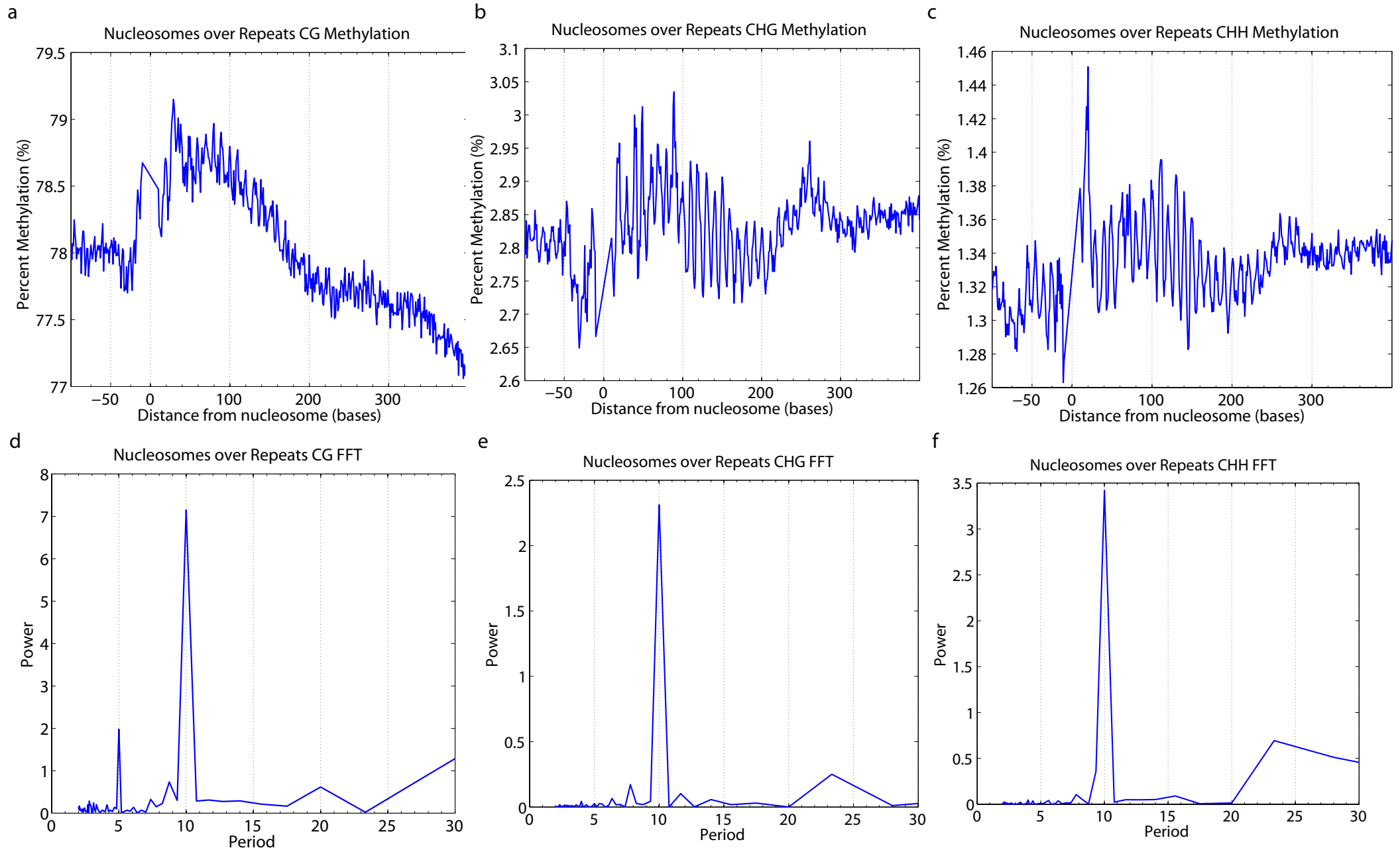
Supplemental Figure 14.



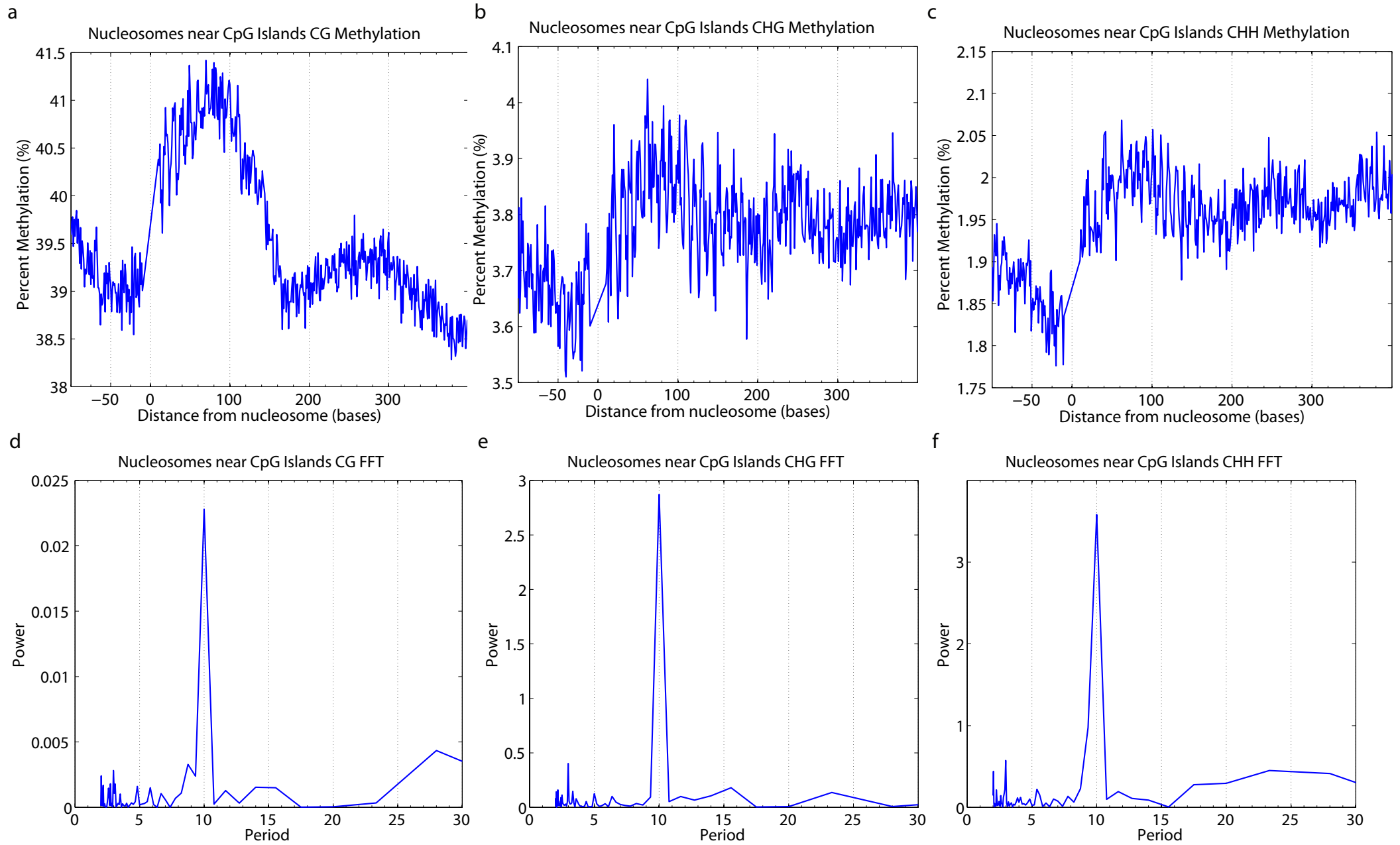
Supplemental Figure 15.



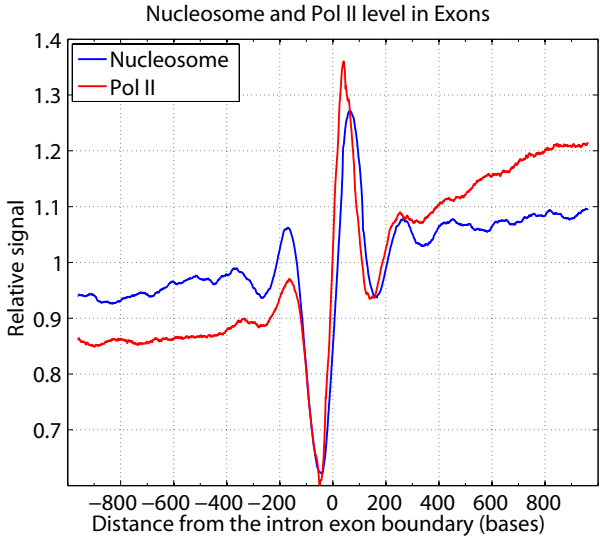
Supplemental Figure 16.



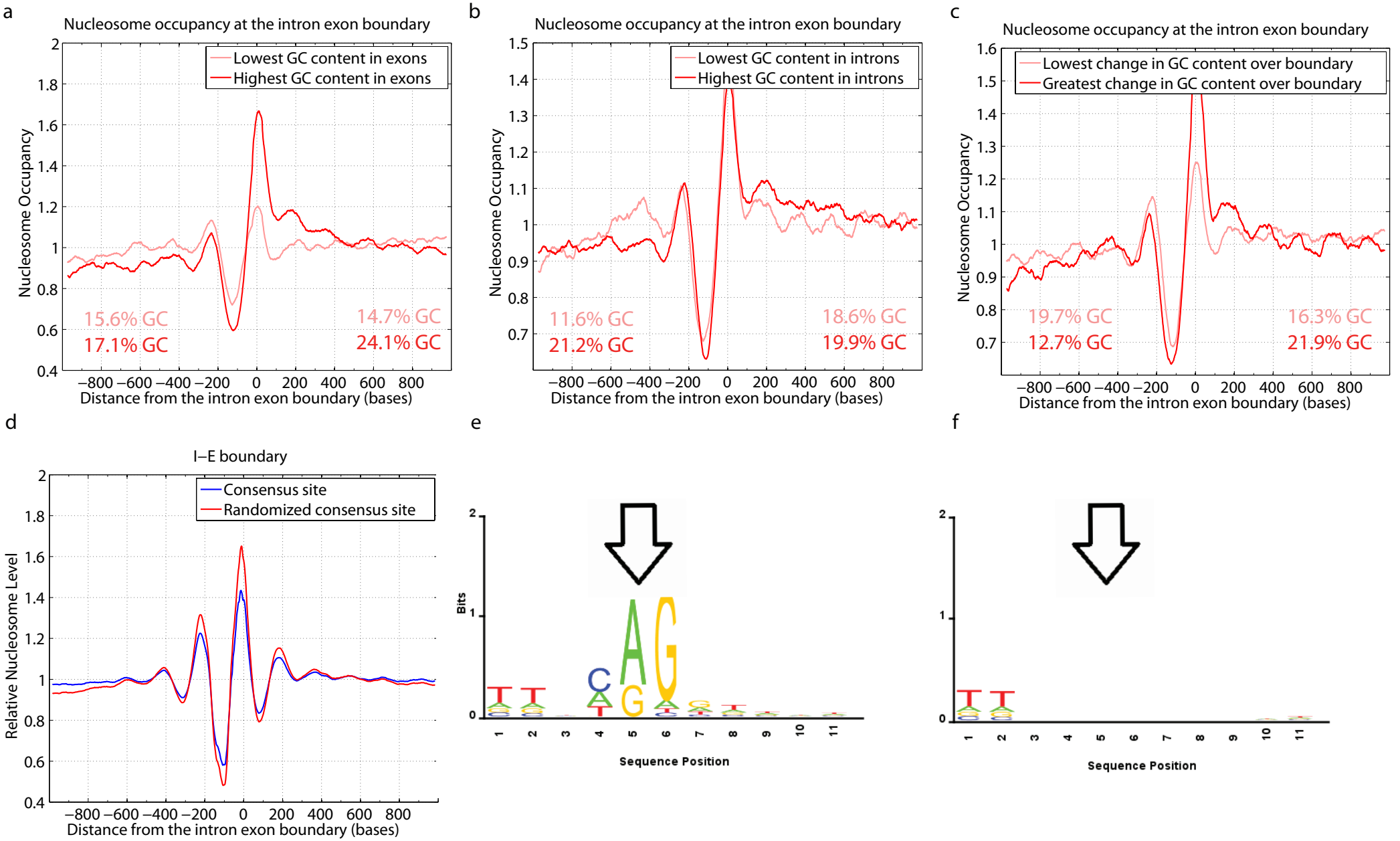
Supplemental Figure 17.



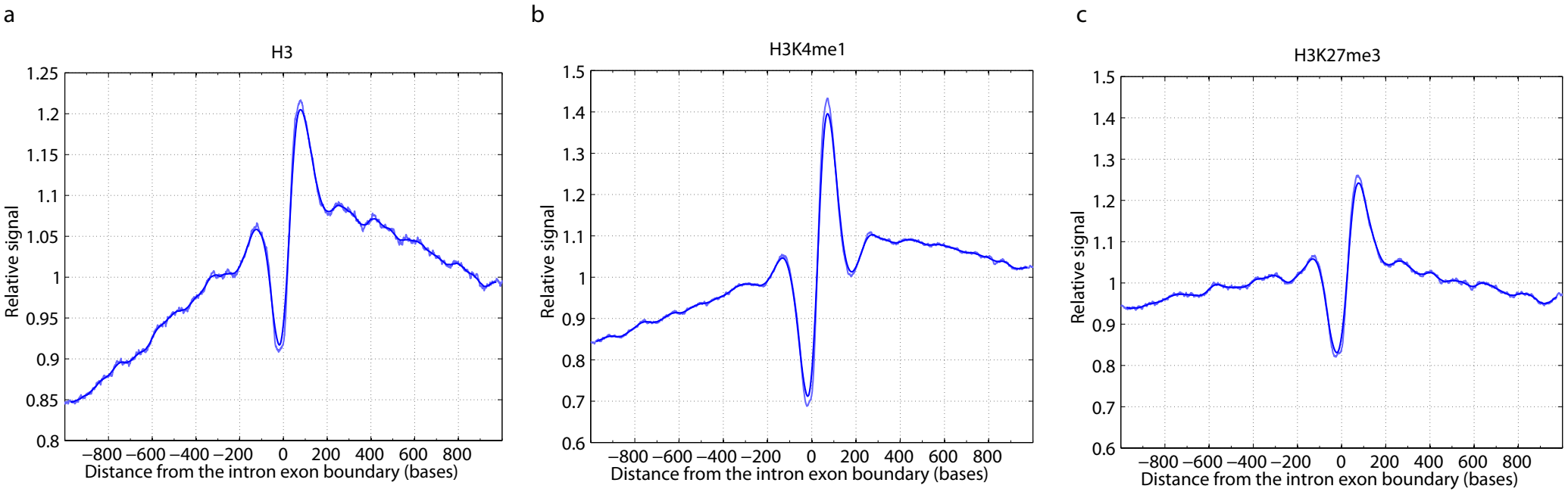
Supplementary Figure 18.



Supplementary Figure 19.

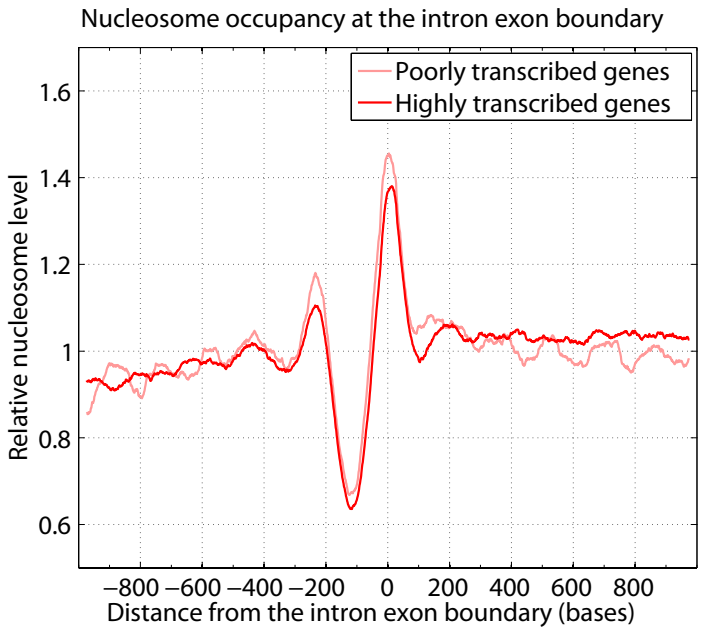


Supplementary Figure 20.

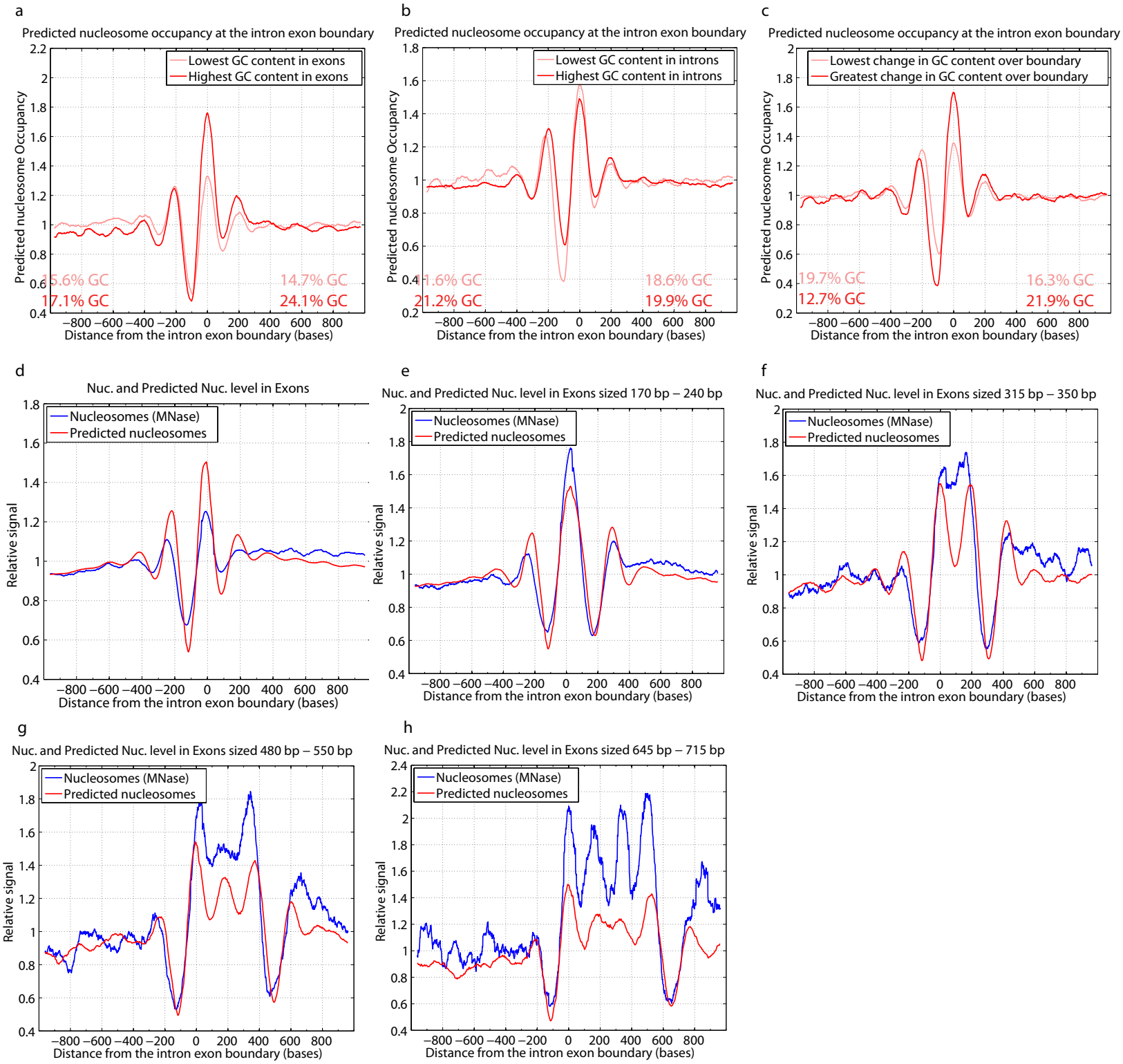


Supplementary Figure 21.

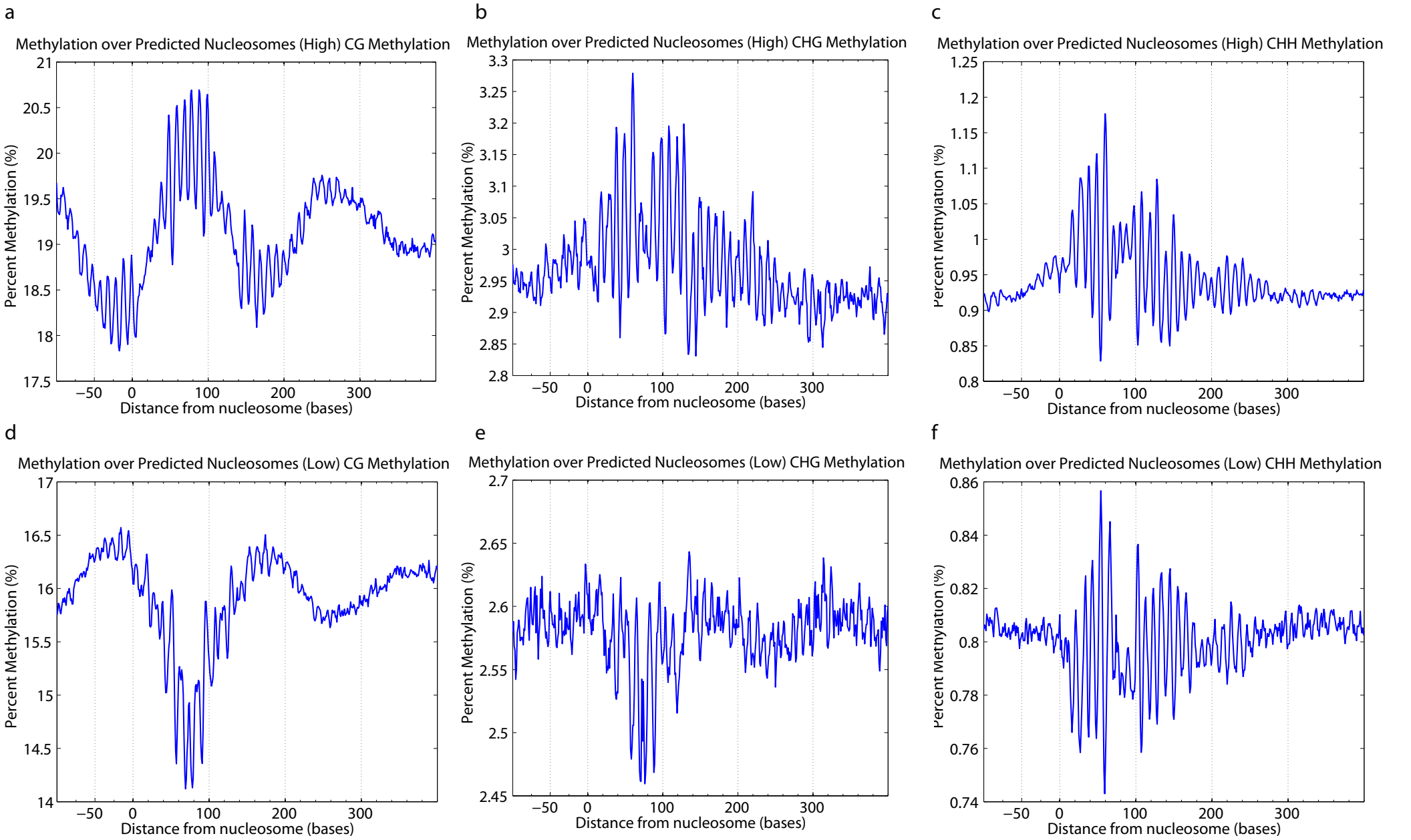
a



Supplementary Figure 22.

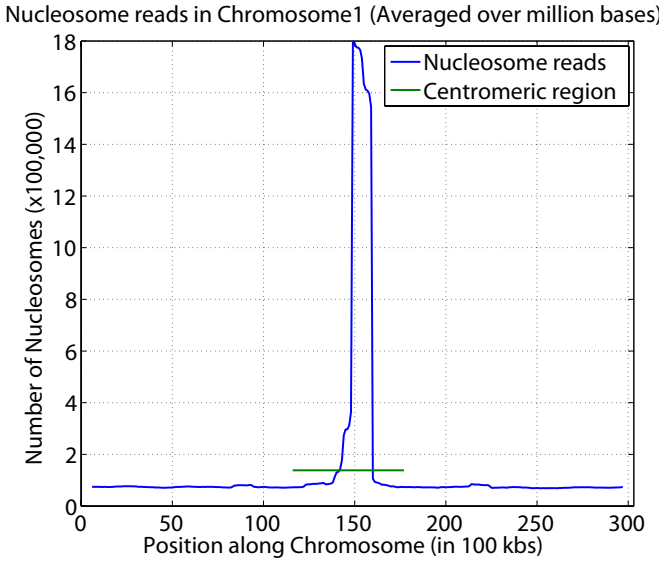


Supplementary Figure 23.

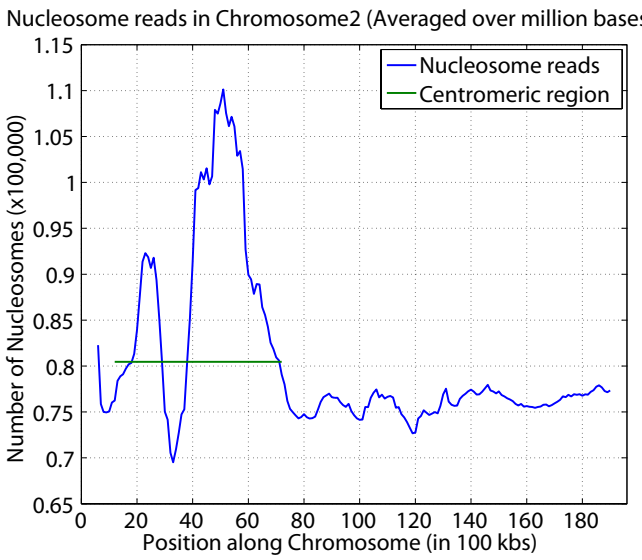


Supplementary Figure 24.

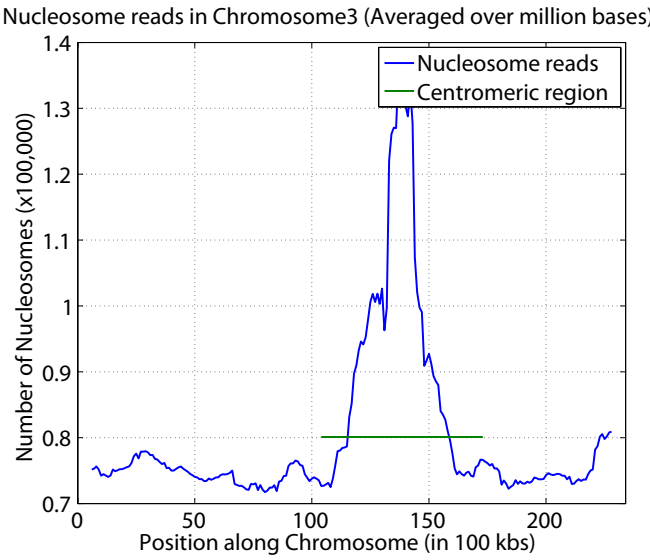
a



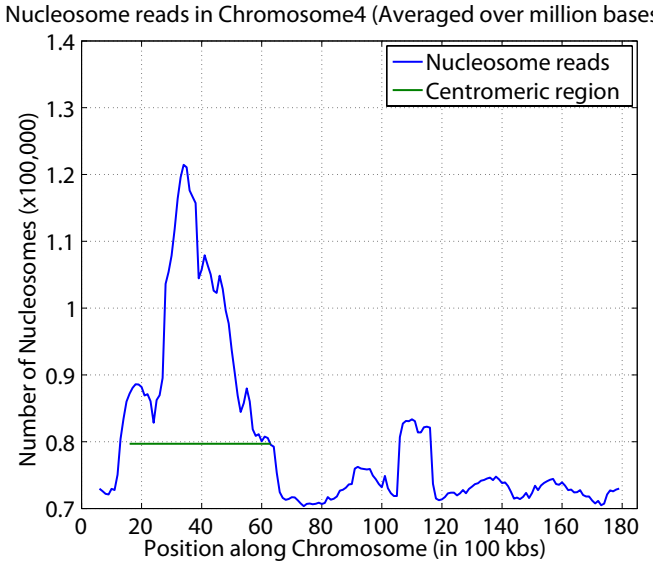

b




c




d




e