Supplementary Information

Conserved and unique organization of nucleosomes and putative regulatory elements in the *Drosophila* genome

Mavrich et al.

Supplementary Methods

Egg Collection, Dechorionation, Crosslink and Nuclei Preparation.

Eight grams of Oregon R embryos 0-12 hrs old were collected and crosslinked at a time, similarly to what has been previously described[1, 2]. The embryos were dechorionated for 90 seconds, washed, and equally divided into eight 50 ml tubes containing 10mls ChIP-FIX (50mM HEPES (pH = 7.6), 100mM NaCl, 0.1mM EDTA, 0.5mM EGTA, 2% formaldehyde) and 30 mls heptane. The tubes were vigorously shaken for 15min at 25°C and then centrifuged at 1500 x g, 1 min, 25°C, followed by the removal of the aqueous layer. The embryos were washed as follows: once with PBS + 0.01% TritonX-100 + 0.125M glycine, and twice with PBS + 0.01% TritonX-100. For each wash, 10 mls of the indicated buffer was added, the tubes were shaken for 1 min and spun down as before, and the aqueous layer was removed (except after the last wash, in which the heptane was also removed). The embryos were washed once more with 20 mls PBS + 0.01% TritonX-100.

The crosslinked embryos were resuspended in 40 mls Homogenization Buffer (10mM HEPES (pH = 7.6), 0.3M Sucrose, 10mM KCl, 1.5mM $MgCl_2$, 0.5mM EGTA, 1mM DTT, 1mM NaBisulfite, 0.2mM PMSF) in a dounce with 10 strokes of the loose pestle and 15 strokes of the tight pestle, while on ice. The sample was spun at 2000 x g, 10 min, 4°C, the supernatant was removed, and the pellet was washed once in 10 mls NPS Buffer. After a spin at 16,000 x g, 5 min, 4°C the supernatant was removed and the pellet was frozen in liquid nitrogen and stored at -80°C until ready for chromatin digestion.

Nucleosome preparations.

See Fig. S2 for a schematic of the procedure. MNase digestion of chromatin was carried out similarly to what has been previously described[3]. Fifteen grams of embryos were thawed and resuspended, using a dounce, in a total volume of 36 mls in NPS Buffer + 1mM β-Mercaptoethanol. 40kU Micrococcal Nuclease was added to the sample and incubated for 2 hrs, 25°C, on a rototorque. The sample was chilled on ice for 10 min and EDTA was added to 10mM concentration to quench the digestion. The sample was then spun at 15,000 x g, 10 min, 4°C and the supernatant was discarded. The pellet was washed twice with 36 mls FA Lysis Buffer + SDS and spun down as before, discarding the supernatant each time.

To solubilize digested chromatin, the sample was briefly sonicated. The sample was resuspended in 18 mls FA Lysis Buffer + SDS and equally divided into fifteen 15 ml tubes. Two samples were sonicated at a time in the Diagenode Bioruptor™ on medium power for five sessions (each session consisting of 30 seconds ON and 30 seconds OFF). Samples were then transferred to 1.7 ml tubes and spun at 16,000 x g, 10 min, 4°C. The pellets were discarded and the supernatants were frozen at -20°C until ready for chIP.

The supernatants were thawed on ice and spun as before to remove any debris. Solubilized digested chromatin from 15 grams embryos were combined and the volume was increased four-fold with FA Lysis Buffer to dilute out the SDS. Then the chromatin was equally divided into five 15 ml tubes. To each tube 170 μl α-H2A.Z antibody[4] was added and samples were incubated for 14 hrs, 4°C, on a rototorque. Chromatin was pre-cleared by incubating with 115 μl bed volume Sepharose 4B (Amersham, 17-0120-01) for 15 min, 4°C, on a rototorque. The resin was spun down at 2000 x g, 3 min, 25°C and each supernatant was transferred to a new tube containing 115 μl bed volume Protein A Sepharose (Amersham, 17-0780-01). Samples were incubated for 1.5 hrs, 4°C, on a rototorque.

Afterwards, the resin was spun down at 1000 x g, 2 min, 25°C, the supernatants were removed, and all resin was combined into two tubes and washed with various buffers: twice with FA Lysis Buffer, twice with High Salt Wash Buffer, twice with FA Wash 2 Buffer, twice with FA Wash 3 Buffer, and once with TE (10mM Tris-Cl (pH = 8), 1mM EDTA). For each wash, 15 mls of the indicated buffer was added, the sample was incubated for 5 min, 25°C, on a rototorque, the resin was spun down at 1000 x g, 2 min, 25°C, and the supernatant was discarded. Additionally, the resin was transferred to a new tube during the second FA Lysis and FA Wash 3 washes to reduce background. Finally, the resin in each tube was resuspended in 5.85 mls ChIP Elution Buffer (25mM Trizma, 2mM EDTA, 0.2M NaCl, 0.5% SDS) and incubated for 15 min at 65°C. The resin was spun down and all eluates were combined in a single tube.

The crosslinks were reversed and the chIP DNA was isolated and gel-purified as previously described[3] (see Fig. S3), with the exception that during gel-purification DNA fragments ranging in size from 100-200bp were excised. Gel-purified DNA was subject to pyrosequencing using the Roche GS20/FLX in accordance with the manufacturer's instructions. Raw sequencing reads can be accessed through NCBI Trace Archives TI xxxxxx, Sequencing Center = "CCGB". Bulk downloads or specific queries of nucleosome positions can be accessed from http://drosophila.atlas.bx.psu.edu/ or Table S2.

Nucleosome mapping.

Sequencing reads were aligned to the FlyBase *D. melanogaster* reference genome (release 5.2) using BLAST. Sequences were aligned to all regions having >90% identity. Aligned regions were denoted on the *Drosophila* genome browser (http://drosophila.atlas.bx.psu.edu/) by the coordinate of each read midpoint. Since the entire nucleosomal DNA was sequenced, both nucleosome borders were identified, thereby allowing nucleosome positions to be defined relative to each border: 73 bp interior to each end. The clustered distribution of reads (bar graph) were smoothed at a coarse-grain level as described previously[3], using 20 as the smoothing parameters, respectively (see Fig. S2). The smoothing included adjustments of bar heights (read locations) in proportion to empirical determinations of MNase bias at each cut site (Fig. S4). Nucleosome positions were defined as the closest genomic coordinate to the peak of the smoothed distribution. Three coarse-grain determinations were made using only the "W" strand, only the "C" strand, and both strands. The error between the W and C calls is presented in Fig. S6.

Analysis was performed on only those 112,750 nucleosomes having been defined by 3 or more reads, although virtually identical patterns (and conclusions) were achieved when all H2A.Z nucleosomes (defined by 1 or more reads) were included (data not shown). Nucleosome fuzziness (Fig. S16) was calculated as the standard deviation for the set of reads whose midpoint locations (defined above) resided within 73 bp of the coarse-grain nucleosomal midpoint.

Nucleosome distribution profiles.

The annotation of all *Drosophila* features (TSS included) were downloaded from FlyBase release 5.2 (ftp://ftp.flybase.net/genomes/dmel/dmel_r5.2_FB2007_01/fasta). The TSSs are annotated in the feature transcripts, and represent both experimental and computationally derived determinations. Only 3,419 (~24% of total) mRNA genes have alternative transcripts, many of which have the same TSS but with variant exon/intron structure. To remove the redundancy, only the 5'-most TSS was used in this study for genes with alternative transcripts. This eliminates TSSs located internal to genes. However, indistinguishable results were obtained when we used the 3'-most TSS for each gene (data not shown). The fact that yeast[5, 6] and *Drosophila* TSS annotations were each derived from multiple lab sources, include both experimental and computational predictions, are internally consistent (yielding robust nucleosome landscape patterns), and are widely used, would indicate that there is unlikely to be a level of random or systematic error in the annotations that would lead to the offset of the nucleosome landscape reported in Fig. 1b.

Nucleosome distances from the reference feature (e.g. TSS in Fig. 1b, or ORF end in Fig. 4b) were binned in 10 bp intervals for all genes that contain H2A.Z nucleosomes within 1 kb of the reference point (11,675 genes). Regions located internal to or within 300 bp of a nearby gene were removed from the analysis, except where indicated, so as to minimize potential influence from nearby genes. A minimum of 300 bp in either direction of the reference feature was analyzed. None of these filters substantially affected the distribution of the data or conclusions. Equivalent filters were applied to the other data sets (described below).

Bin data were normalized to number of regions represented in each bin, and smoothed using a moving average. The size of the moving average (typically between 3 and 8 bins) was set to provide the optimal balance between signal and noise for purposes of visual display. The size of moving average within the chosen range had no effect on the conclusions drawn.

For Fig. 5b, 1,868 genes contained both a paused Pol II (see Fig. 5a) and H2A.Z nucleosomes within -/+1kb. Data were binned (10 bp bins) and plotted as a 3-bin moving average. 9,610 genes were used in the "not-paused" class.

DNA motif distribution

The 194 motifs were obtained by combining the novel motifs and known TF motifs from a previous report[7]. Three motifs were excluded due to rare occurrences (*twin of eyeless* TGGAGGDGGWAHTMATBVRTGWDDDRKKMW*, glass* CAATGCACTTCTGGGGCTTCCAC*, and abnormal chemosensory jump 6* TGCATAATTAATTAC). Four short or highly degenerate motifs (*prospero* CWYBDCY, *apterous* TAAT, *mitochondrial transcription factor A* TTATS and *bric a brac 1* WHWWWWWWWWKK) caused computational failure due to over-abundance. Their locations were therefore restricted to within -/+ 1 kb of the TSS

Based on the motif consensus sequences, all occurrences of each motif were identified on both strands across the entire fly genome. Palindromic motifs were only searched on one strand. Motif midpoint distances to the TSS were calculated and binned (size = 10 bp). Bin counts were smoothed using a 7 bin moving average and plotted in Figs. 2b and S12. Each motif was classified into one of four classes based upon visual inspection of the profiles (Table S4), as shown in Fig. 2. To derive the values in Fig. 2b bin counts were normalized to the average bin count from -/+1 kb of the TSS, then $\log_2$ transformed.

The frequency of each motif distance from an H2A.Z nucleosome midpoints located throughout the entire genome was measured in the same manner. Several negative controls were performed. First, motifs located within 1 kb of the TSS were excluded from the analysis. Second, motifs >1kb from the TSS were excluded from the analysis. Third, motifs <1kb from the TSS had their location inverted (e.g. a site at +200 became -200), and all motif-nucleosome distance recalculated, or just those that had inverted locations.

GAGAG site identification.

The *D. melanogaster* genome (r5.2) was scanned for GAGAG motifs that resided within 20 bp of another GAGAG motif (Table S3), as they tend to co-occur[8, 9]. Such clusters were binned in 20 bp intervals based upon distance from the TSS. Binned data were smoothed using an 5-bin moving average. Clusters that were within 300 bp of an adjacent gene were removed from the analysis (which had little effect on the distribution).

GAF binding.

ChIP-chip for GAF was carried out using Schneider 2 cells, and hybridized to Affymetrix *Drosophila* Tiling 1.0R Arrays. Signal analysis, interval analysis, and peak calling were performed using Model-based Analysis of Tiling-arrays software[10]. The bandwidth and max gap parameters had a value of 150 bp, and peaks were called using a 1% false-discovery rate threshold. The distance from transcription start site was calculated for the 2,907 GAF peaks above a 1% false-discovery rate threshold. The GAF composite distribution profile was generated by binning these distance values in 20 bp intervals for all genes that contain a GAF peak within 1 kb of the transcription start site. The data was smoothed using an 8-bin sliding window average in 1 bin steps.

Dinucleotide frequencies.

To survey for all 16 dinucleotides across nucleosomal DNA we used all 7,600 nucleosomal DNA sequences that were exactly 147 bp long. Dinucleotide counts were performed in the 5'-3' direction on both strands and summed. Data were smoothed use a 3 bp moving average, except at position -/+73 and -/+72 where 1 and 2 bp moving averages, respectively, were used.

Nucleosome positioning sequences.

Correlation profiles were calculated as described elsewhere[11], using the AA/TT dinucleotide pattern obtained previously[12] or from a *Drosophila*-specific pattern in which the AA/TT dinucleotide pattern was derived from

top 1000 most well positioned H2A.Z nucleosomes, representing <1% of all nucleosomes. The *Drosophila* pattern represents AA/TT dinucleotide positional frequency profiles dyad symmetrized as described previously[12]. The CC/GG pattern was derived from the same nucleosome set used to derive the AA/TT pattern. This pattern is shown in Fig. S15. Note, in all cases the number of nucleosomes used to define the search pattern represents <1% of all nucleosome locations, and thus are expected to contribute <1% to resulting genome-wide correlation profiles. Regions within 300 bp of an adjacent gene were excluded from the analysis to avoid potential complication associated with Pol II binding to nearby genes. The data were normalized to the number of regions present in each bin so as to account for the removal of the excluded regions.

Genome-wide mapping of RNA polymerase II.

The genome-wide distribution of Pol II was determined as described elsewhere[13]. Briefly, stage 14 embryos were crosslinked with formaldehyde; chromatin was isolated by $CsCl_2$ gradient ultracentrifugation and sonicated; and Pol II immunoprecipitated using 8WG16 antibody (Covance) against the CTD domain of the Pol II large subunit. The Pol II ChIP and control ChIP DNA samples, along with DNA purified from input chromatin samples, were amplified and hybridized to Affymetrix Drosophila Genomics Tiling Arrays, each of which contains about 3 million oligo probes that cover the euchromatic portion of the genome at an average resolution of about one per 36 bp. The ChIP-chip data was analyzed using TiMAT, a software developed at Berkeley Drosophila Transcription Network Project[Li, 2008 #910]. 4,286 crosslinked regions of Pol II were identified genome-wide at a 1% FDR based on the symmetric null distribution method of TiMAT. These regions were ranked based on the peak window score of each region. Within these regions, peaks of Pol II were located using the Mpeak algorithm[14]. All parameters for Mpeak were set at the default except for the "Largest search range" which was reduced to 1.0 kb. 1,956 genes were found to contain at least one Pol II Mpeak within 1 kb of a TSS, and no Pol II Mpeak between +1 kb and the end of the gene. These genes are defined as putative Pol II-paused genes. Pol II locations in Fig. 5a were binned in 20 bp intervals. 8 bin moving averages were generated and plotted as a filled line graph. 8,736 genes lacked a Pol II Mpeak anywhere in the transcription unit, of which 3,742 contained Pol II crosslinked regions somewhere in the body of the gene, and 4,994 lacked significant Pol II binding.

Genome-wide mapping of nucleosomes bound to RNA polymerase II

Nucleosomal DNA from 3 grams of Oregon R embryos 0-12 hrs old was prepared similarly to H2A.Z nucleosomal DNA as described above, with the following changes. After sonication, solubilized DNA was treated with an equal volume of 6M urea, then dialyzed for 20-24 hrs against 100-fold volume of FA Lysis Buffer + 0.05% SDS at 4°C (SpectraPor membrane, 6k-8k MWCO, #632650). This treatment enhanced the recovery of Pol II during immunoprecipitation. The sample was then incubated for 16-18 hrs at 4°C with 54 μl α-Rpb3 antibodies (raised in rabbit) and immunoprecipitated as described above. Immunoprecipitated DNA was not detected in the absence of antibody or in the absence of crosslinking (not shown). Eluate DNA was gel-purified in the size range of 75-200 bp and 30 ng DNA was LM-PCR amplified (adaptor sequence: GCGGTGACCCGGGAGATCTGAATTC). Two biological replicates were prepared for hybridization with the GeneChip WT Double-Stranded DNA Terminal Labeling Kit (Affymetrix #900812) using the manufacturer's recommended protocol, and hybridized to Affymetrix *Drosophila* Tiling 1.0R Arrays.

We used MAT program[10] to detect nucleosome-Pol II bound regions with the following settings: bandwith 70 bp, minimum probe count 3, MaxGap 147 bp by considering the size of nucleosomal DNA fragments (~150 bp), 25-mer probe and 36-bp average probe separation. Interval cutoff was set to p-value of 1e-2. We developed a peak detection algorithm that iteratively selects the maximal, non-overlapping peaks based on the score MAT assigned to the probe. The script is available upon request. A total genome-wide count of 55,029 Pol II-crosslinked nucleosomes were obtained by this criteria. The distribution of these nucleosomes relative to the TSS is displayed in Fig. 5c, and represents the average of two biological replicates. The nucleosome count data were binned in 10 bp and further smoothed using a 3 bin moving average.

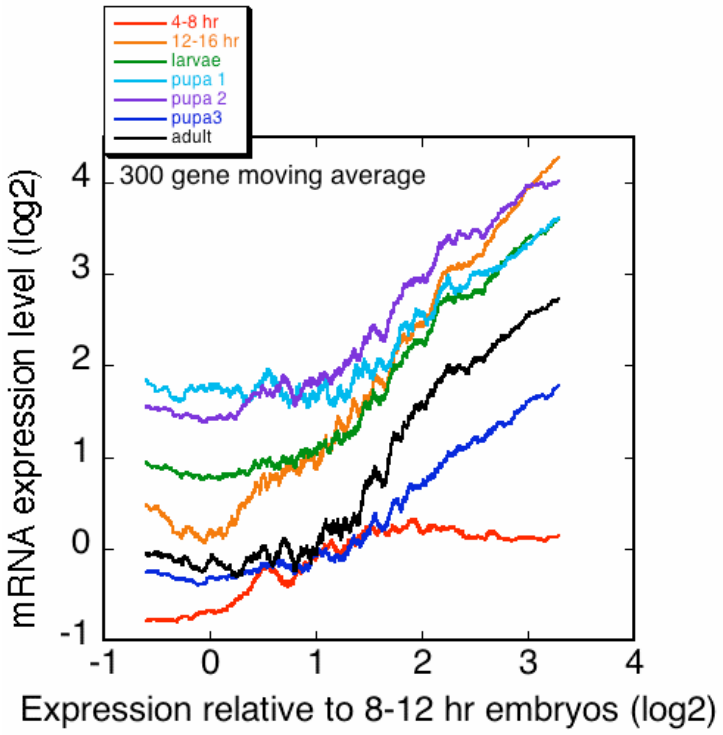


Figure S1. Moving average of mRNA expression levels at different stages of *Drosophila* development are plotted against mRNA levels found in 8-12 hr embryos. All data are from published data sets[15], and have been first normalized to expression levels from 0-4 hr embryos.

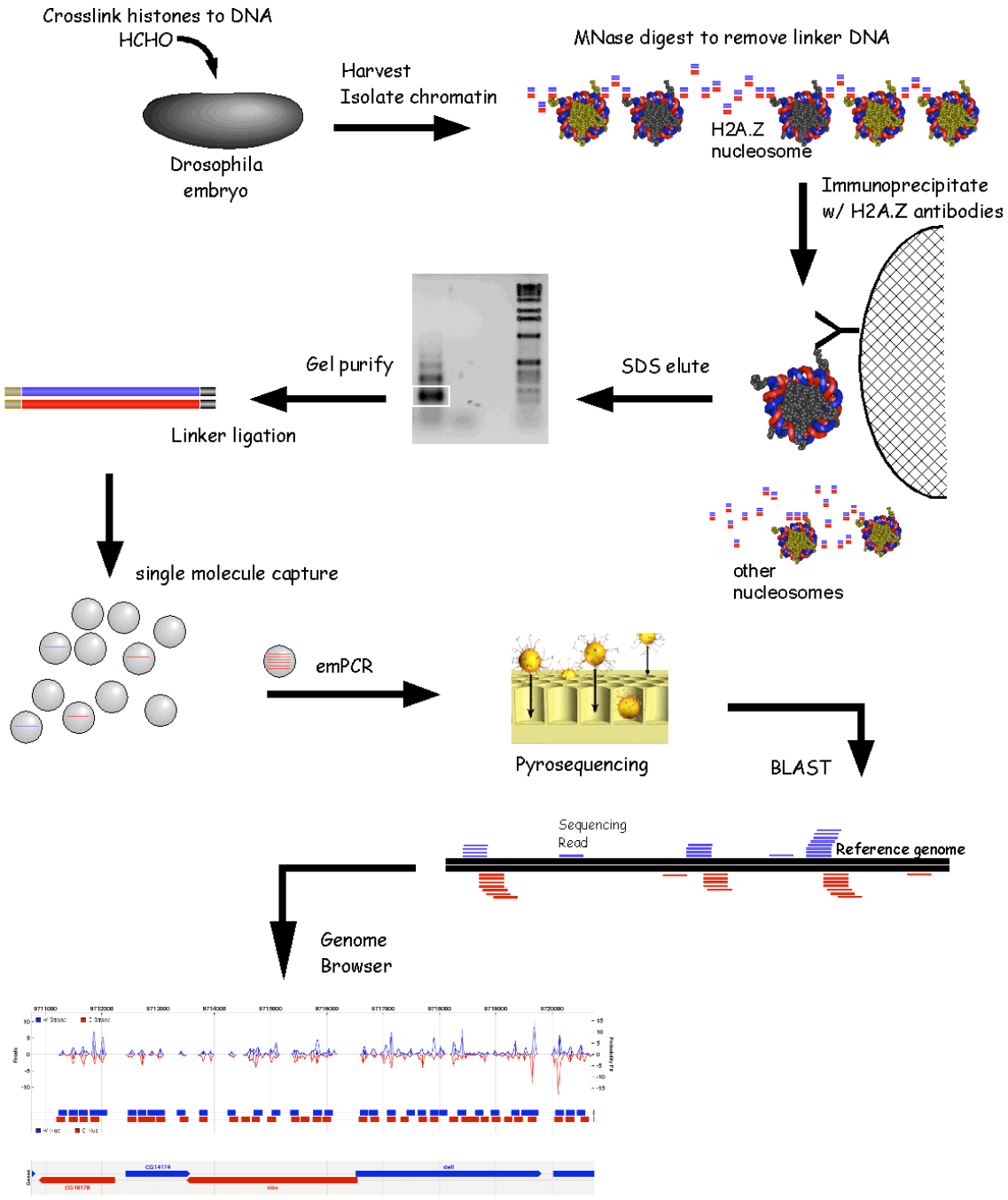

Crosslink histones to DNA
HCHO
Drosophila embryo
Harvest
Isolate chromatin
MNase digest to remove linker DNA
H2A.Z nucleosome
Immunoprecipitate w/ H2A.Z antibodies
other nucleosomes
SDS elute
Gel purify
Linker ligation
single molecule capture
emPCR
Pyrosequencing
BLAST
Sequencing Read
Reference genome
Genome Browser

Figure S2a

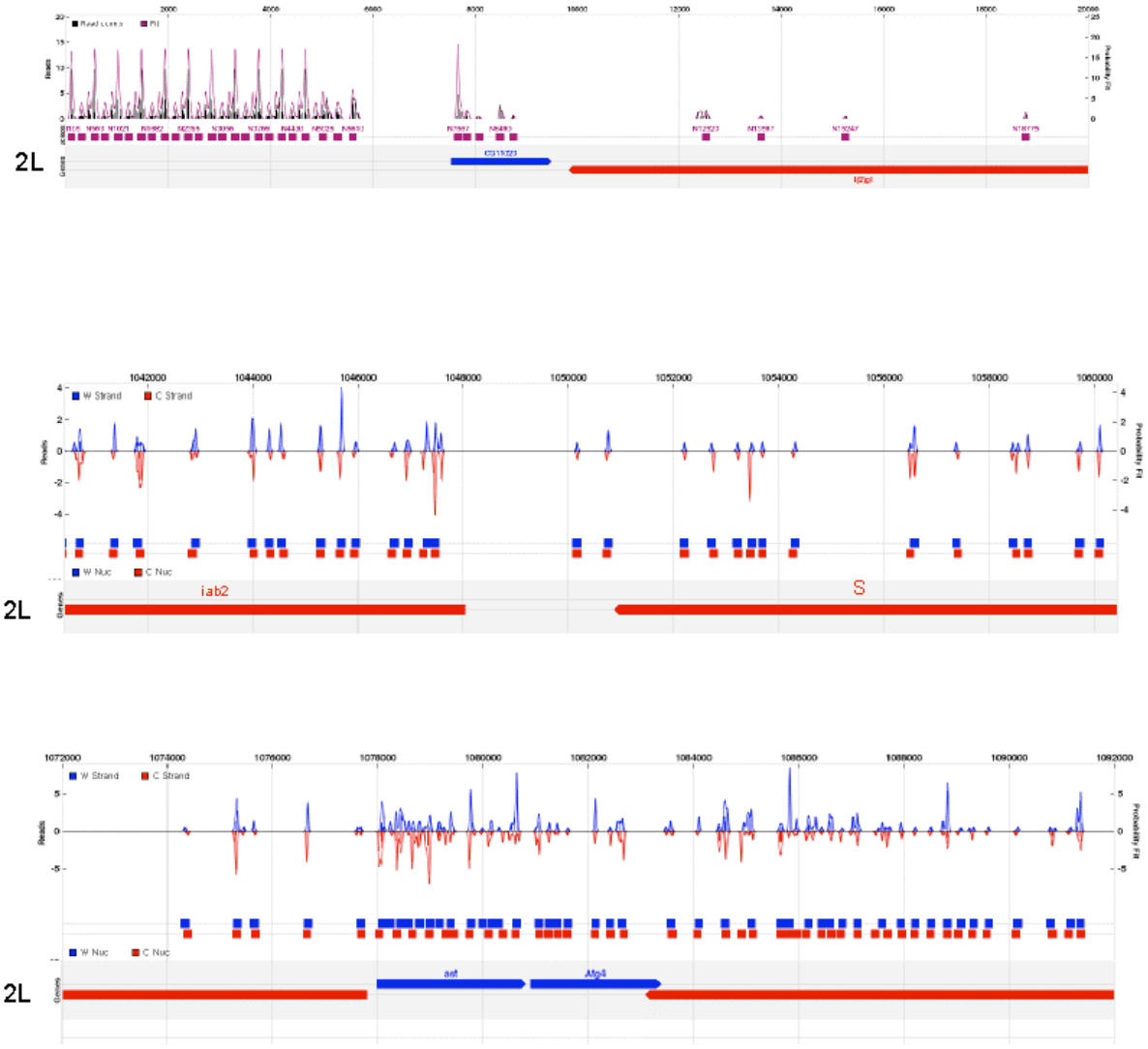


Figure S2b

Figure S2. Outline of *Drosophila* embryo nucleosomal DNA isolation, DNA sequencing, and browser presentation. **a**, Outline of nucleosome preparation and sequencing methods. See Methods for details. **b**, Browser shots at selected locations along chromosome 2L, displaying coarse-grain smoothing. Upper panel represents the telomeric region with W- and C-strand data sets combined and smoothed at a coarse-grain level. The two lower panels have the W and C data separated out. The nucleosome tracks are 147 bp in length and are centered to the smoothed distribution peaks. Below the nucleosome tracks in all panels are the gene tracks.

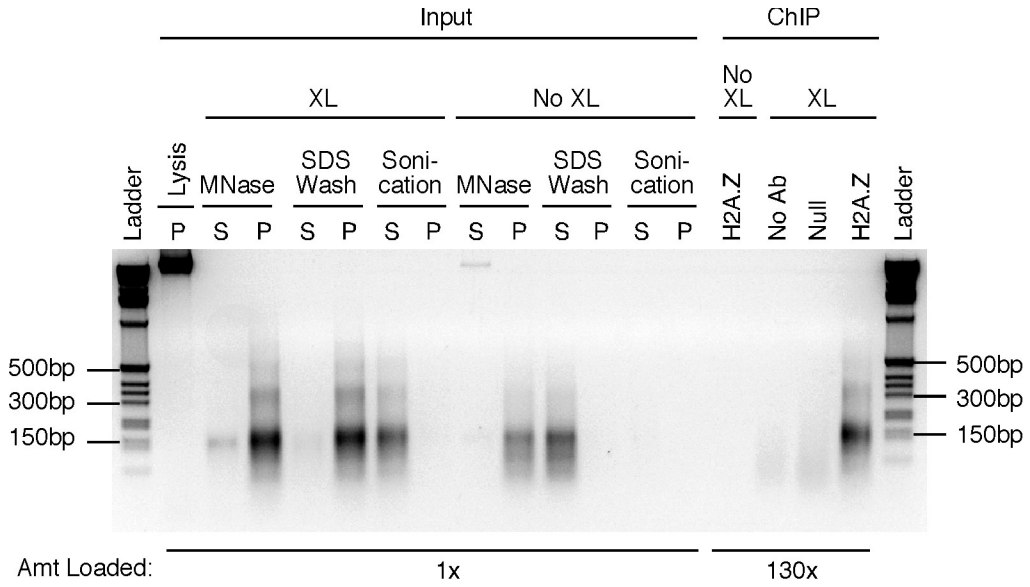


Figure S3. Isolation of H2A.Z nucleosome core particle DNA from *Drosophila* embryos. Ethidium bromide stained agarose gel of samples at various purification stages. P and S denote pellet and supernate, respectively. XL denotes formaldehyde crosslinking. Brief sonication after MNase digestion facilitated solubilization of the crosslinked material. Null denotes use of a nonspecific antibody. H2A.Z denotes use of *Drosophila* H2A.Z/H2Av antibodies[4].

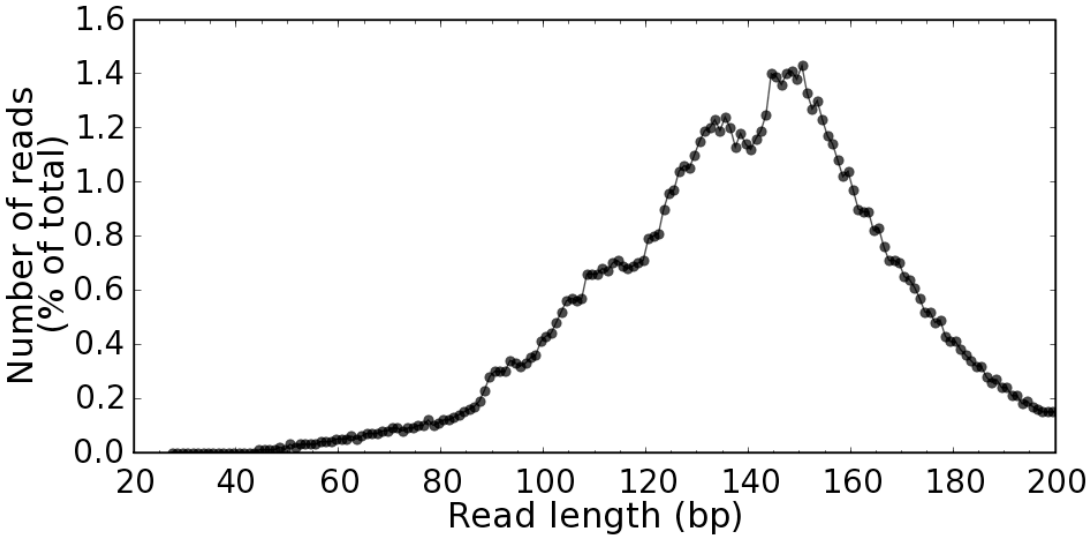


Figure S4. Distribution of H2A.Z nucleosomal DNA read lengths. The minor peaks and shoulders occur at 20 bp intervals, reflecting over-digestion by MNase in intervals of 10 bp from both ends of the nucleosome.

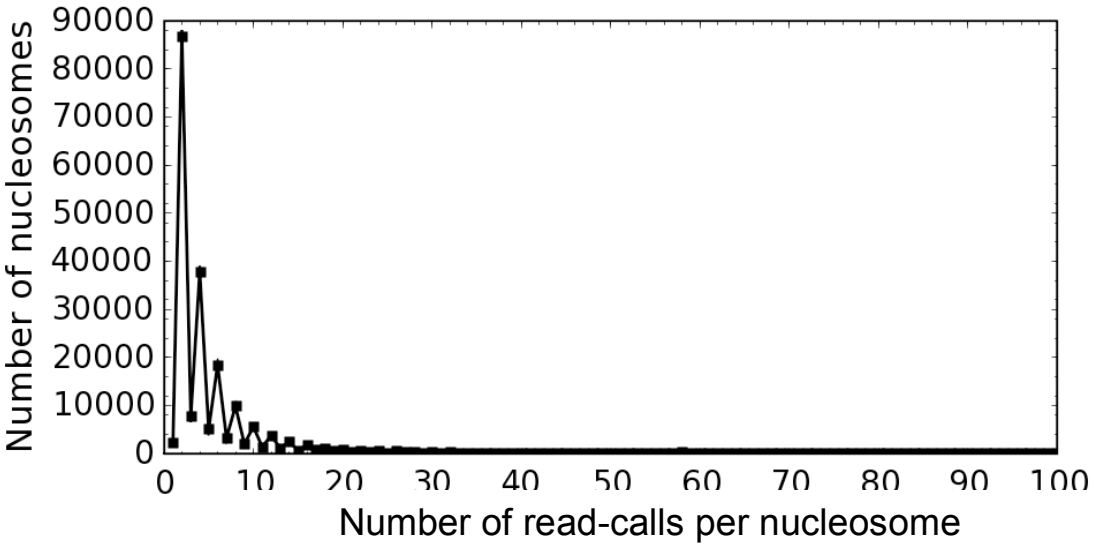


Figure S5. Number of "read calls" per nucleosome. Nearly all sequencing reads encompassed the entire nucleosome, allowing two calls to be made per read (corresponding to the location of the 5' border and the 3' border). In rare cases of reads <100 bp in length, only the 5' end of the read was used to make a call, hence the zig-zag pattern in the graph. The 100 bp cut-off was imposed because nucleosomal DNA was size selected by agarose gel electrophoresis to be well above this size. Those below 100 bp are likely to be artifacts of premature termination during DNA sequencing.

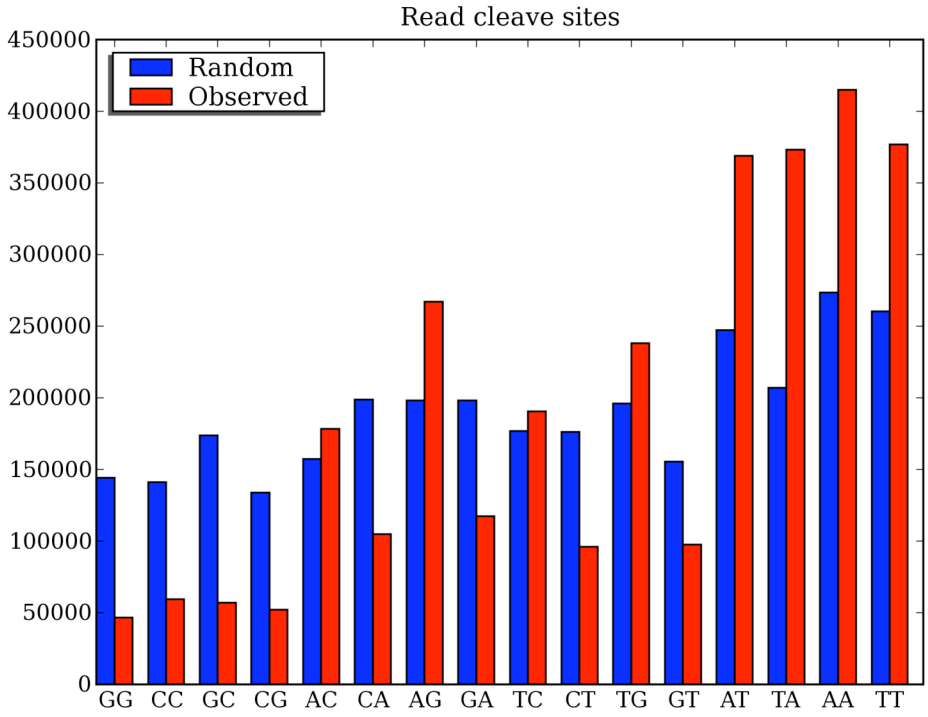


Figure S6. MNase bias correction. Each blue bar displays the frequency of dinucleotides that bracket each measured MNase cut site. An MNase cut site is defined as the 5'-most (or 3'-most) nucleotide in each sequenced read, and its next upstream (or downstream) nucleotide, respectively, in the *D. melanogaster* genome reference map (r5.2). The frequency of randomly selected dinucleotides in a 20 bp window centered over each cut site is shown in red and provides a measure of local dinucleotide sequence bias inherent in the genome. Larger windows gave the same result. The raw MNase cutting frequency was normalized to intrinsic genomic bias prior to applying it as a correction factor.

It is important to note that our studies involve over-digestion with MNase, and this diminishes the severity of MNase sequence bias[16, 17]. The correction applied here further diminishes its effect but does not completely eliminate the bias.

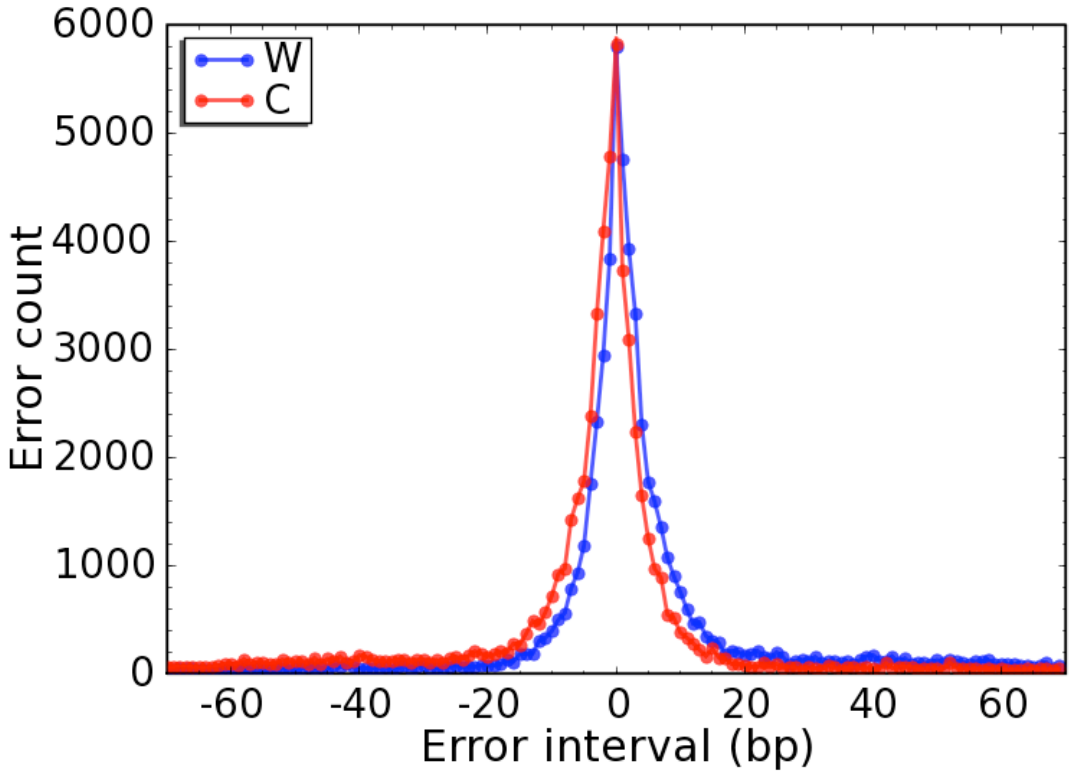


Figure S7. Error analysis of *Drosophila* H2A.Z nucleosome maps derived separately from the Watson and Crick strands. The distance between each coarse-grain call made separately on the W and C strand and each call made from the combined data set was binned in 1 bp intervals and plotted as a smoothed frequency distribution. Only those nucleosomes having three or more reads were used.

Note that each read, by virtue of spanning the entire nucleosome length, simultaneously identified both borders of a nucleosome, which greatly reduced the number of reads needed to achieve a high level of accuracy.

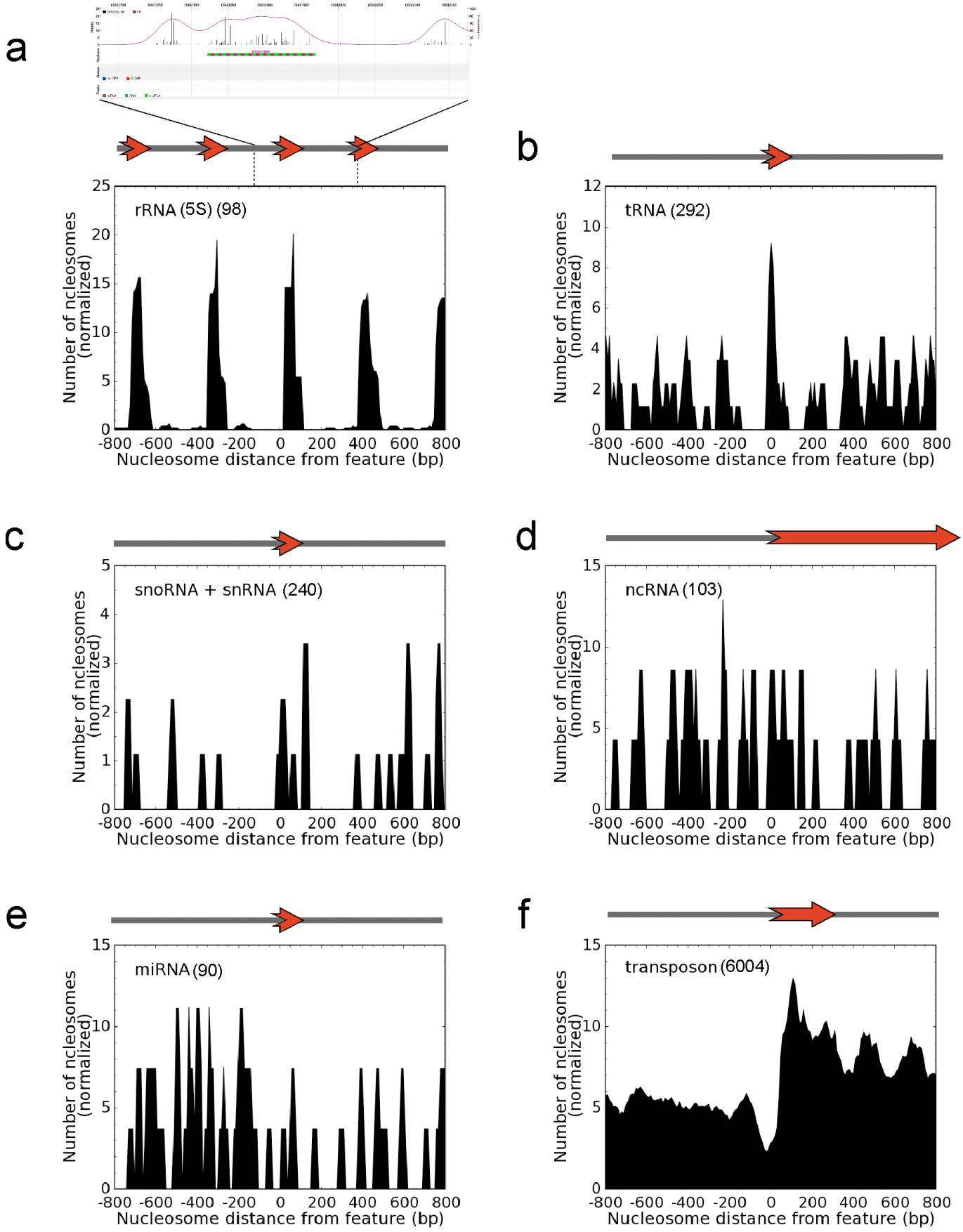

a
rRNA (5S) (98)
Number of ncleosomes (normalized)
Nucleosome distance from feature (bp)
b
tRNA (292)
Number of ncleosomes (normalized)
Nucleosome distance from feature (bp)
c
snoRNA + snRNA (240)
Number of ncleosomes (normalized)
Nucleosome distance from feature (bp)
d
ncRNA (103)
Number of ncleosomes (normalized)
Nucleosome distance from feature (bp)
e
miRNA (90)
Number of ncleosomes (normalized)
Nucleosome distance from feature (bp)
f
transposon (6004)
Number of ncleosomes (normalized)
Nucleosome distance from feature (bp)

Figure S8. Distribution of H2A.Z nucleosomes around non-coding RNA genes. a, Composite distribution of sequencing reads around 5S rRNA genes. b-f, Distribution of nucleosomes around tRNA (b), snRNA and snoRNA (c), ncRNA (d), miRNA (e), and transposable elements (f). All reference points are transcription start sites.

The *Drosophila* genome is composed of a wide variety of noncoding transcription units including rRNA, tRNA, snRNA, snoRNA, miRNA, and a variety of transposable elements. The 5S rRNA gene is transcribed by Pol III and is tandemly repeated nearly 100 times in 373 bp intervals[18]. A single H2A.Z-containing nucleosome is contained within each repeat that is centered 120 bp downstream of the TSS (panel a), which is a similar location to that seen for Pol II-transcribed genes. Intriguingly, this creates a 250 bp region that is free of H2A.Z nucleosomes. tRNA, snRNA, and snoRNA genes and their surrounding regions were relatively depleted of H2A.Z nucleosomes (panels b,c), which may reflect a general nucleosome depletion due to high expression levels. The few that did possess a nucleosome had it centered over the TSS, and thus might be selectively repressed in embryos. ncRNA and miRNA genes displayed modest levels of H2A.Z with no notable uniformity of organization (panels d,e). Transposable elements, as a whole, appeared to have an H2A.Z nucleosomal array located 70-75 bp downstream of the TSS, which places the start site on the nucleosome edge (panel f), as is seen in *Saccharomyces*[3].

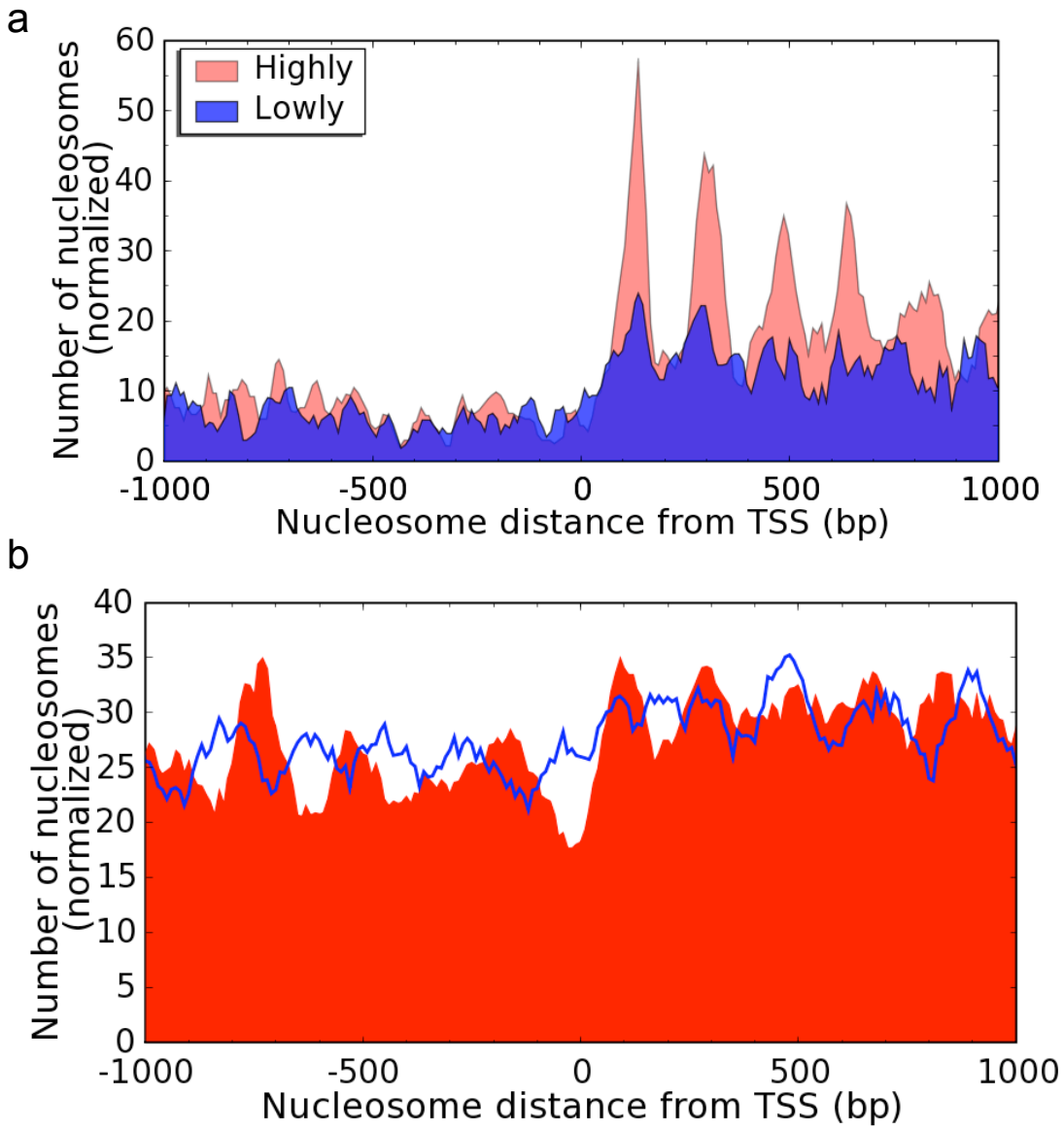


Figure S9. **a**, Distribution of H2A.Z nucleosomes for genes that are highly (top 10%) or lowly (bottom 10%) expressed in *Drosophila* embryos[19]. **b**, Distribution of bulk nucleosomes for genes that are highly (red fill plot) or lowly (blue trace) expressed in *Drosophila* embryos[19].

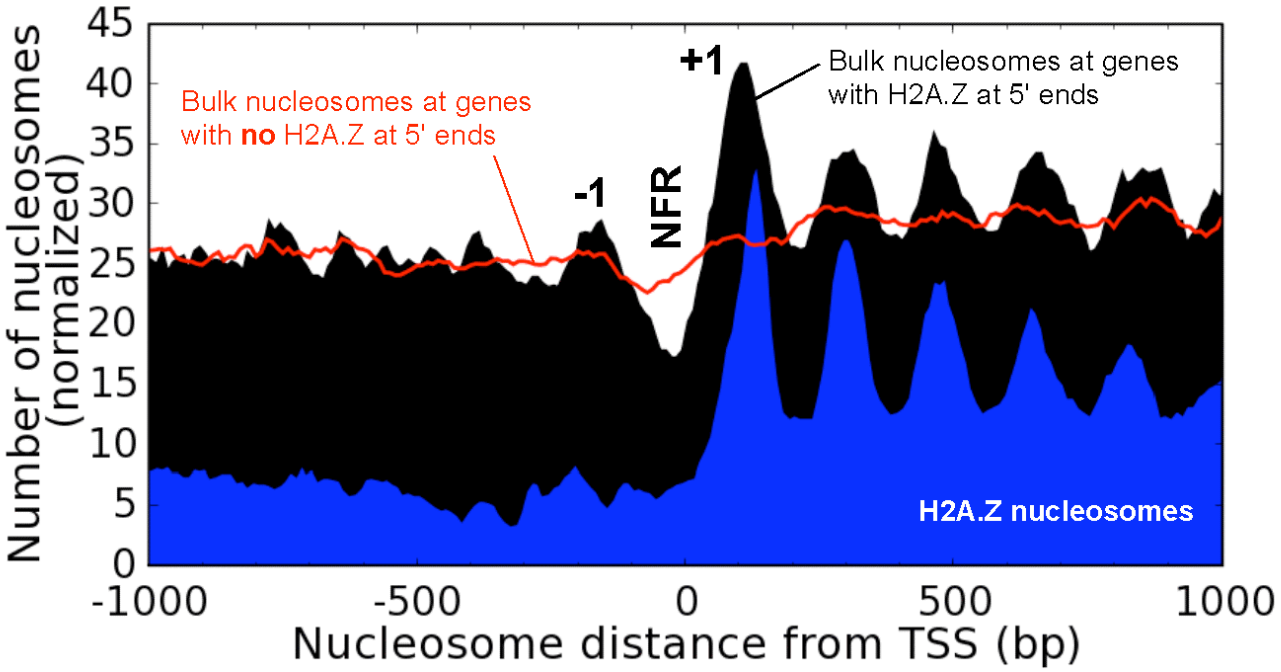


Figure S10. Composite distribution of all nucleosomes relative to the TSS. The black plot shows the nucleosome distribution (5 bin moving average) for 5,701 genes that contain an H2A.Z nucleosome at the +1 position (between +60 and +180 relative to the TSS, as defined by ChIP-seq; the overall H2A.Z nucleosome distribution is shown in blue). The red trace shows the nucleosome distribution for all remaining genes (8,442). Bulk nucleosomes were detected by Affymetrix tiling arrays having 36 bp average probe spacing. The H2A.Z profile from Fig. 1b is shown in blue. The shallower peaks and valleys for all nucleosomes vs H2A.Z nucleosomes at the same set of genes might, in part, be a reflection of the lower resolution of these tiling arrays (~36 bp), compared to detection by sequencing.

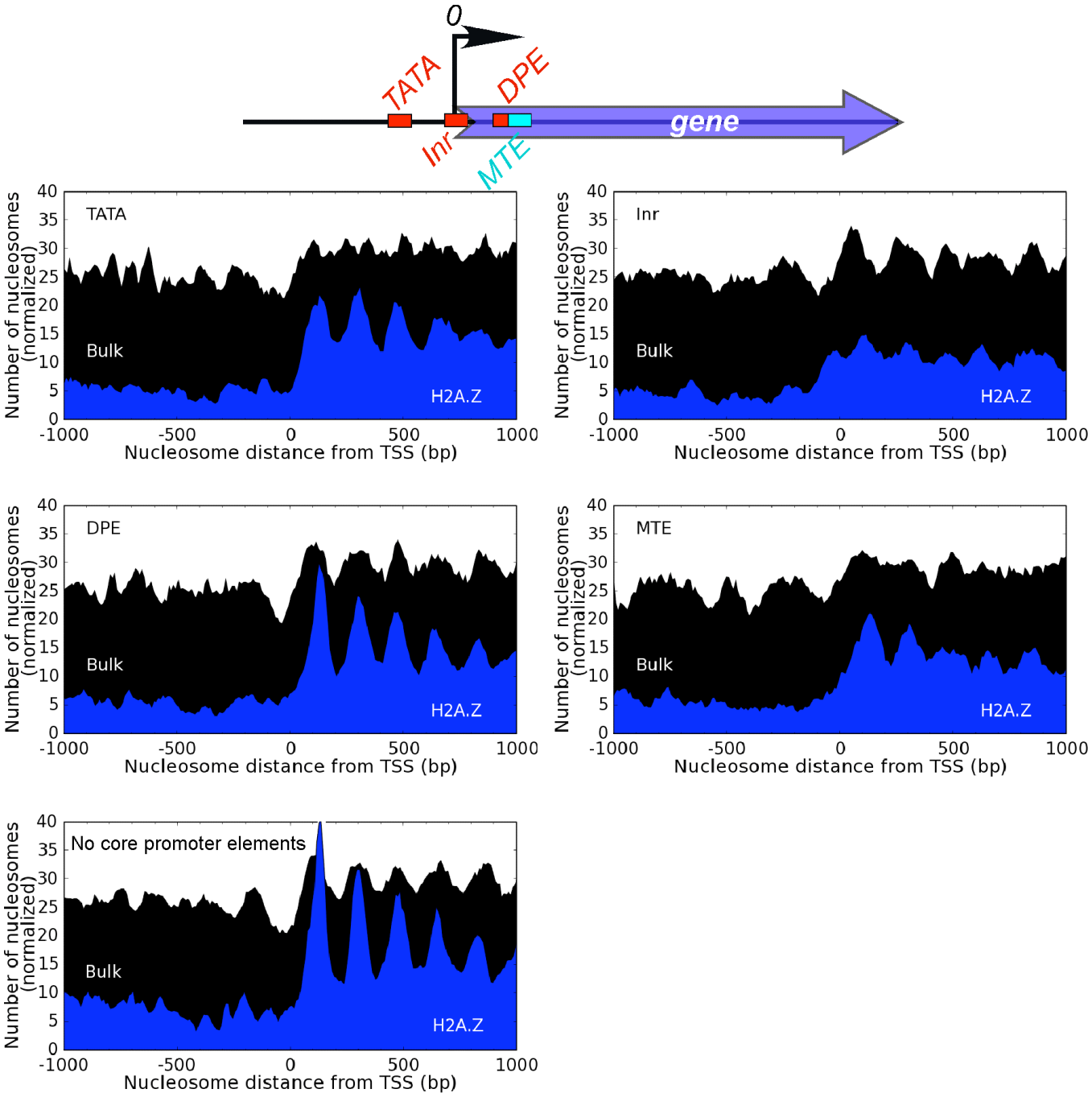
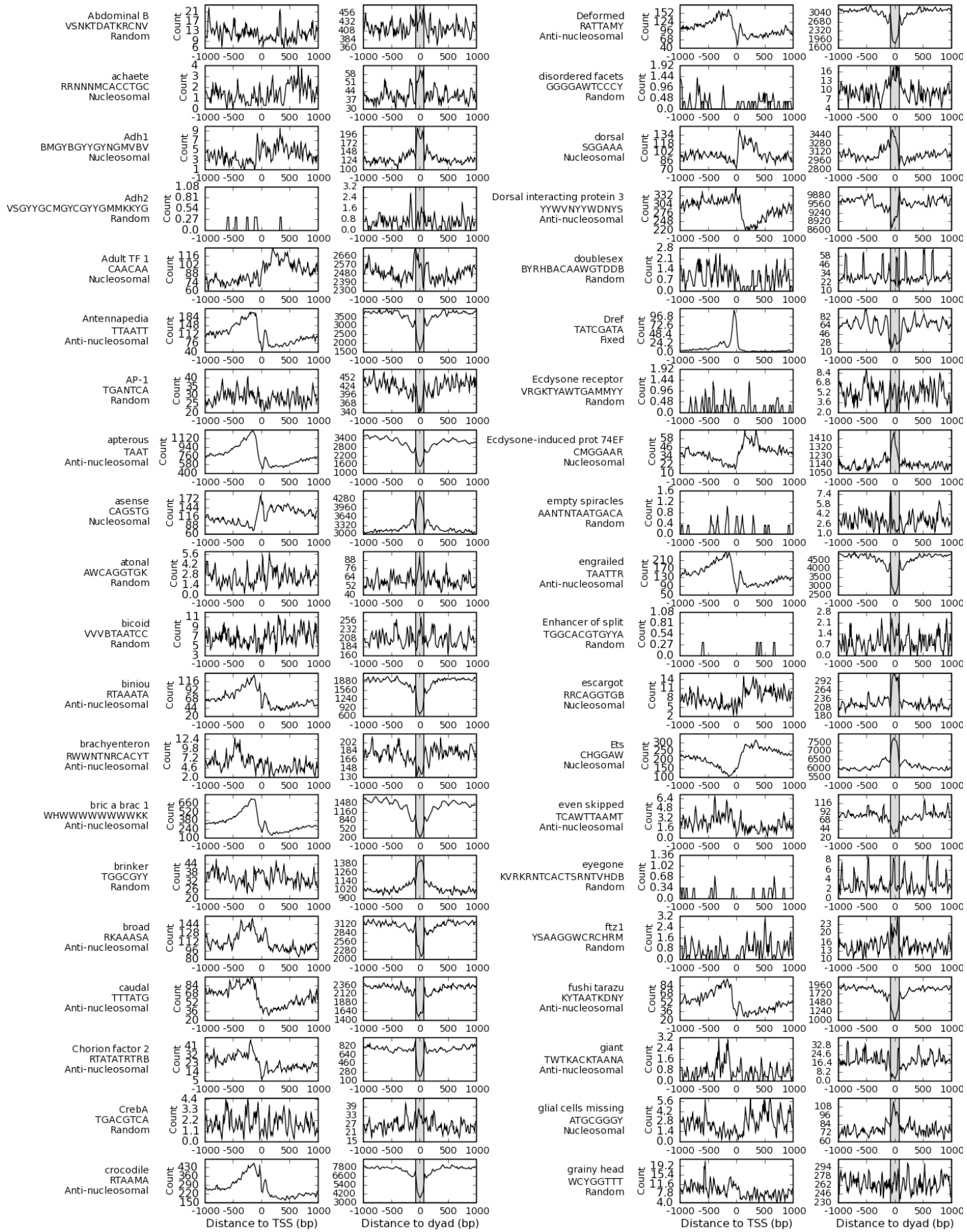
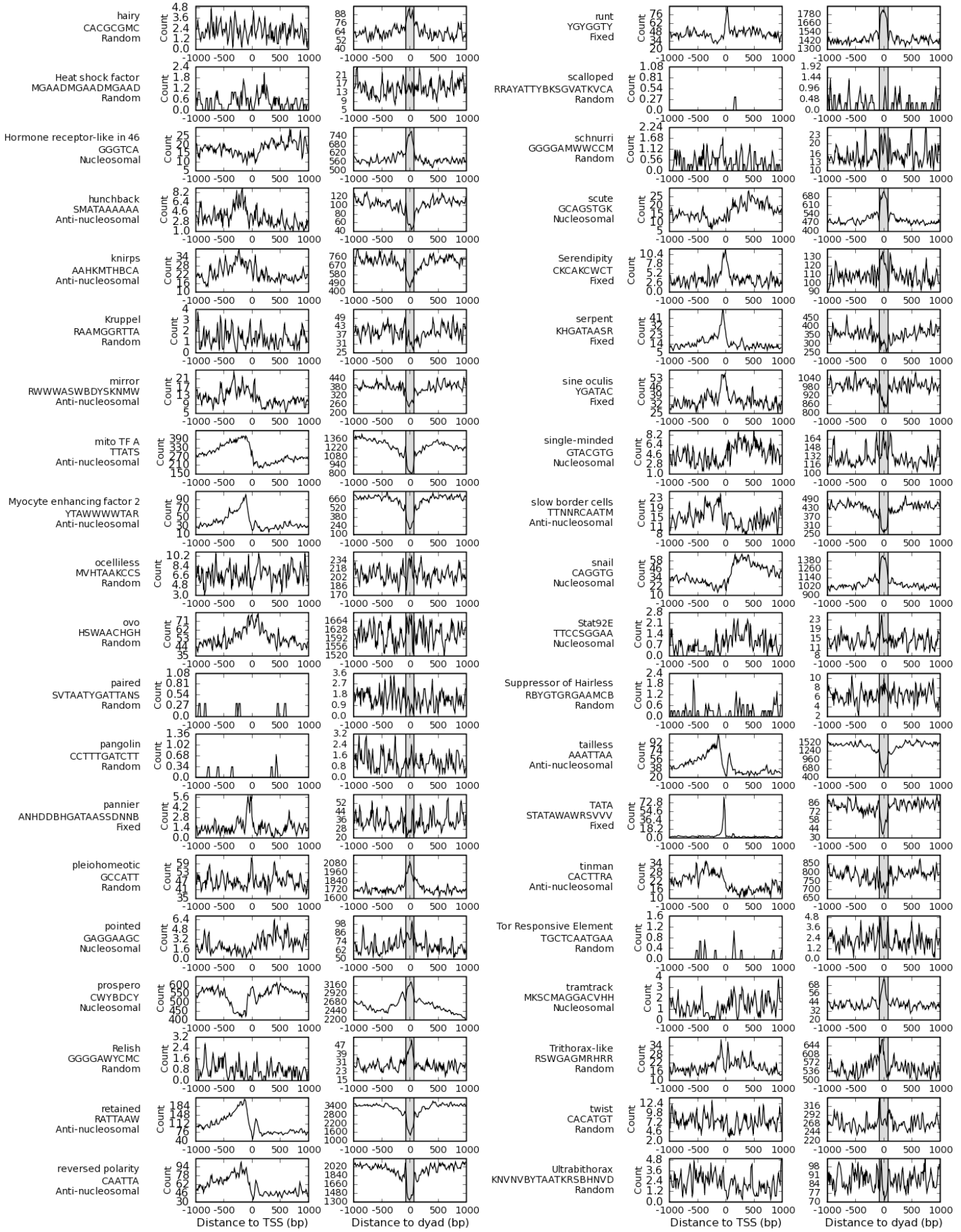
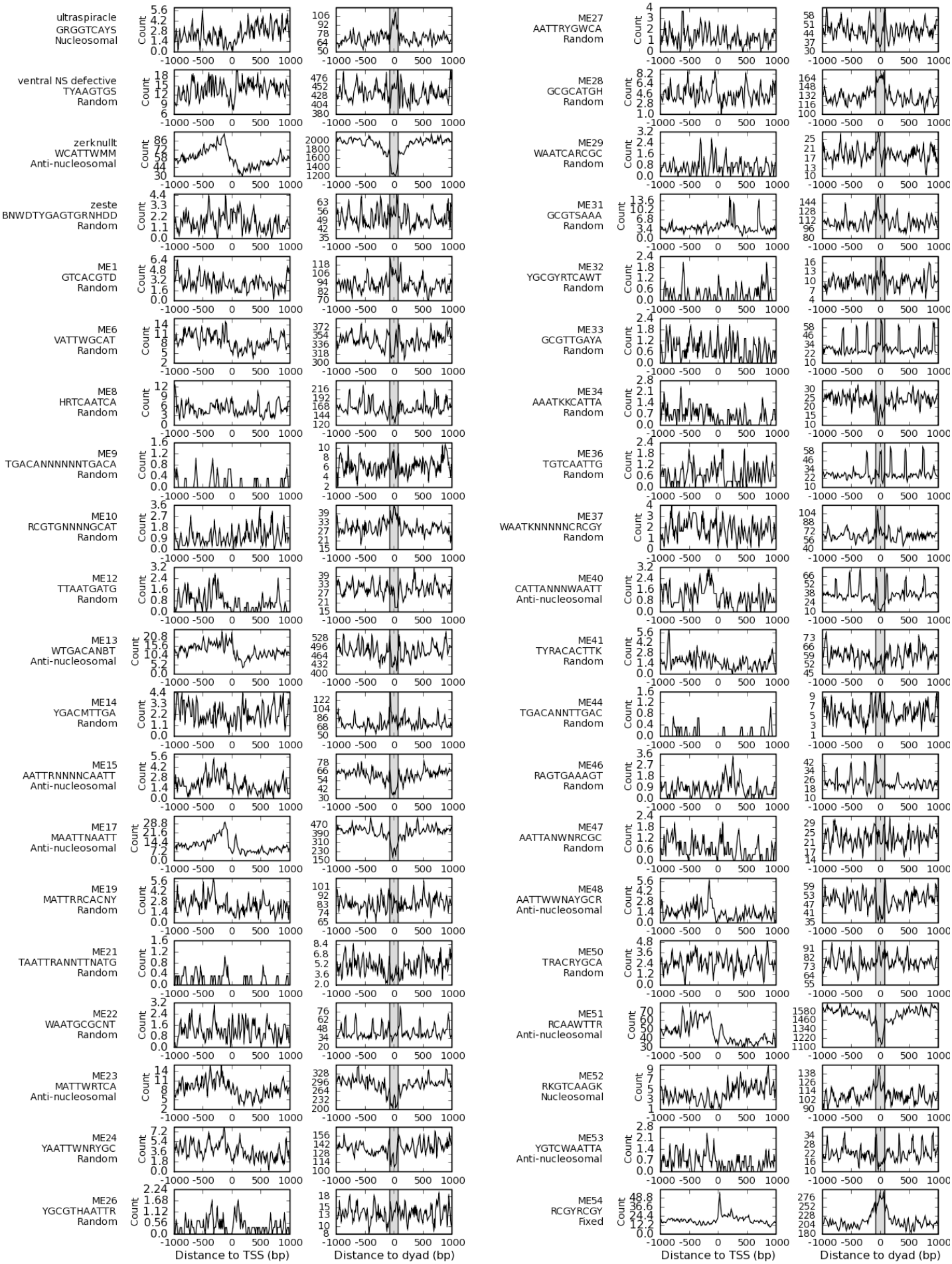


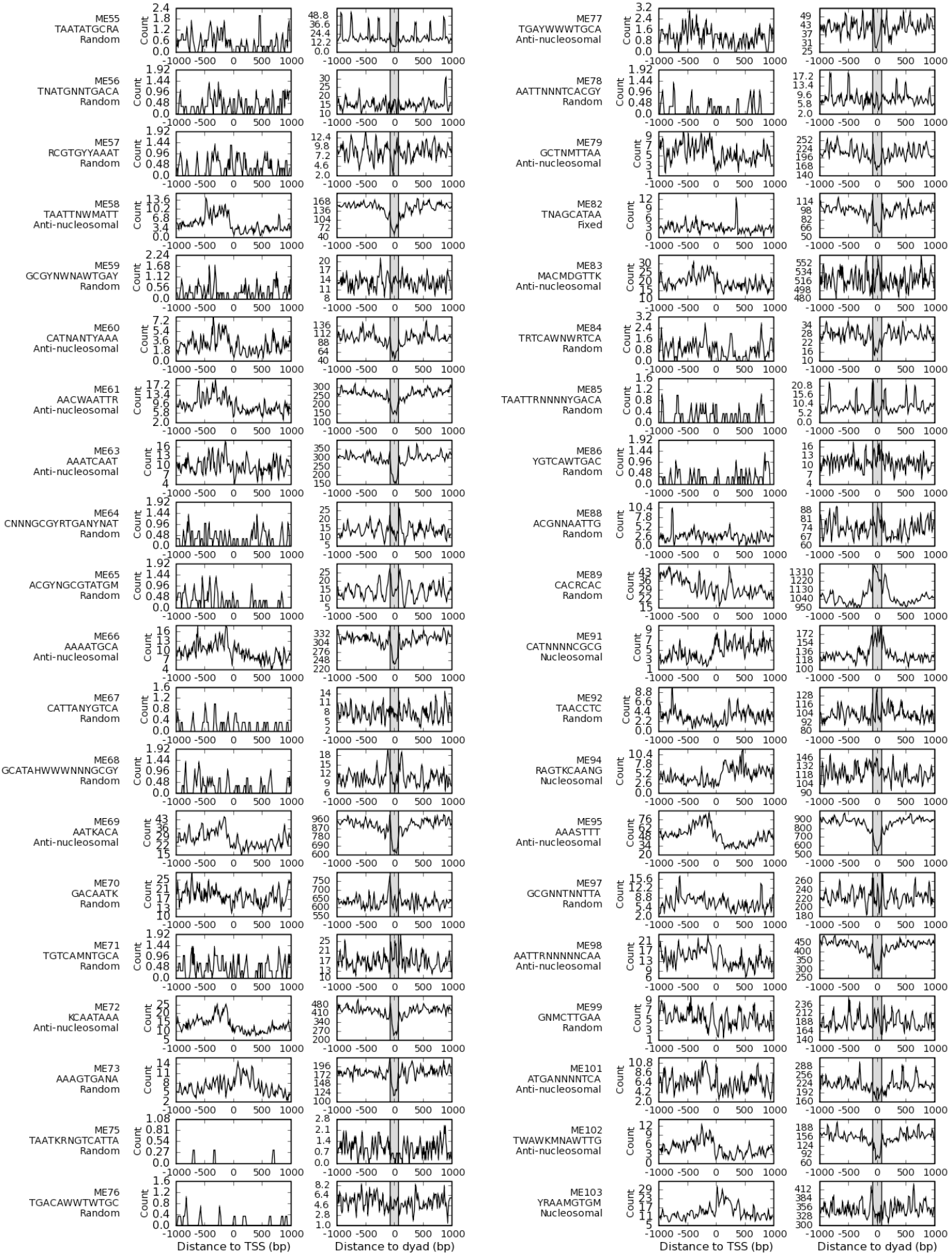
Figure S11. Composite distribution of all nucleosome (black) and H2A.Z nucleosomes (blue) on various classes of promoters. Nucleosome midpoint distances from the transcriptional start site (TSS) were binned (10 bp bin size) and plotted as a moving average.

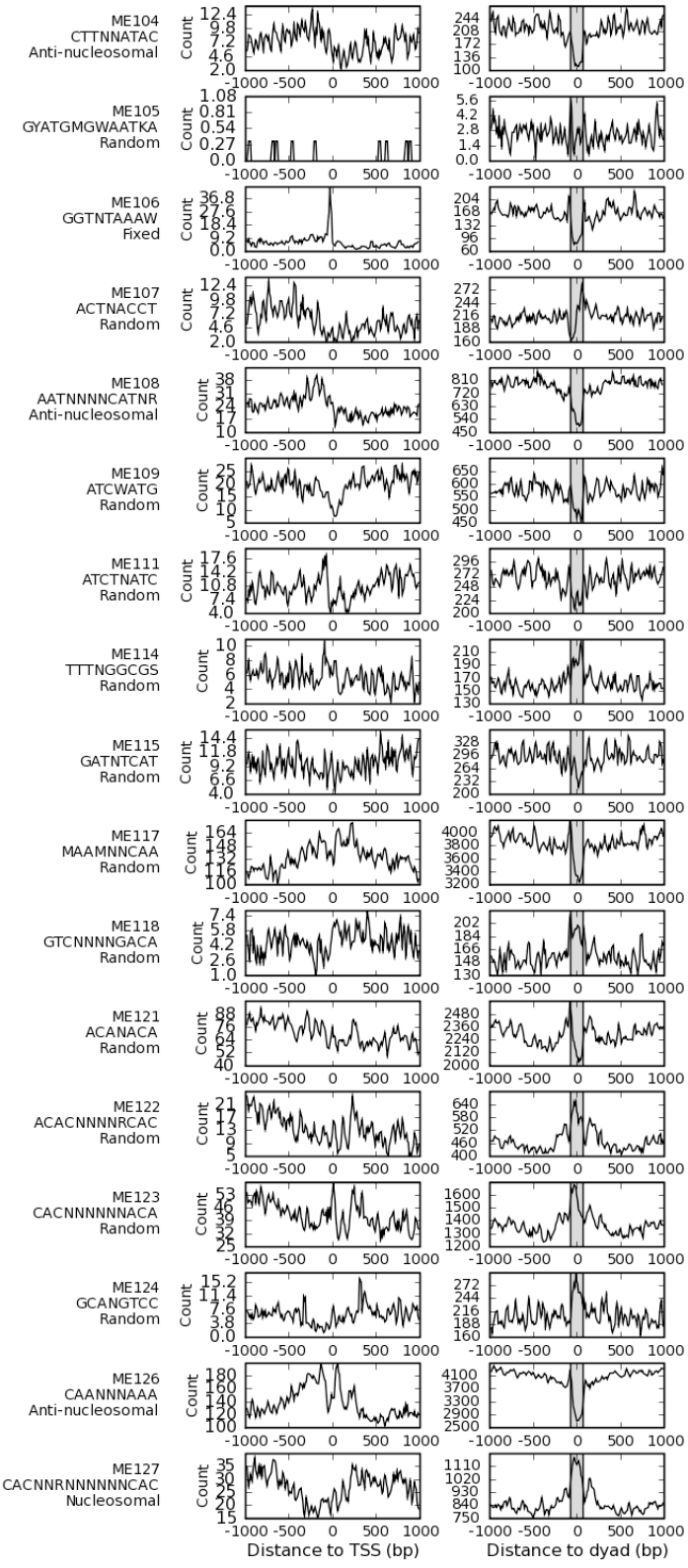
Drosophila core promoter consensus sequences were obtain from published studies[20-23]. TATA box containing genes were defined as those genes containing TATAAA in 5 out of 6 positions in a region from -45 to -15 relative to the TSS[21, 22] (Table S2). All genes were inspected at positions -10 ~ +10 to TSS for Inr, positions +28 ~ +33 for DPE, and positions +7 ~ +31 for MTE. From this, 3,463 TATA-containing genes, 1172 Inr-containing genes, 2,858 DPE-containing genes, 897 MTE-containing genes, and 7,536 genes with no core promoter elements were obtained.

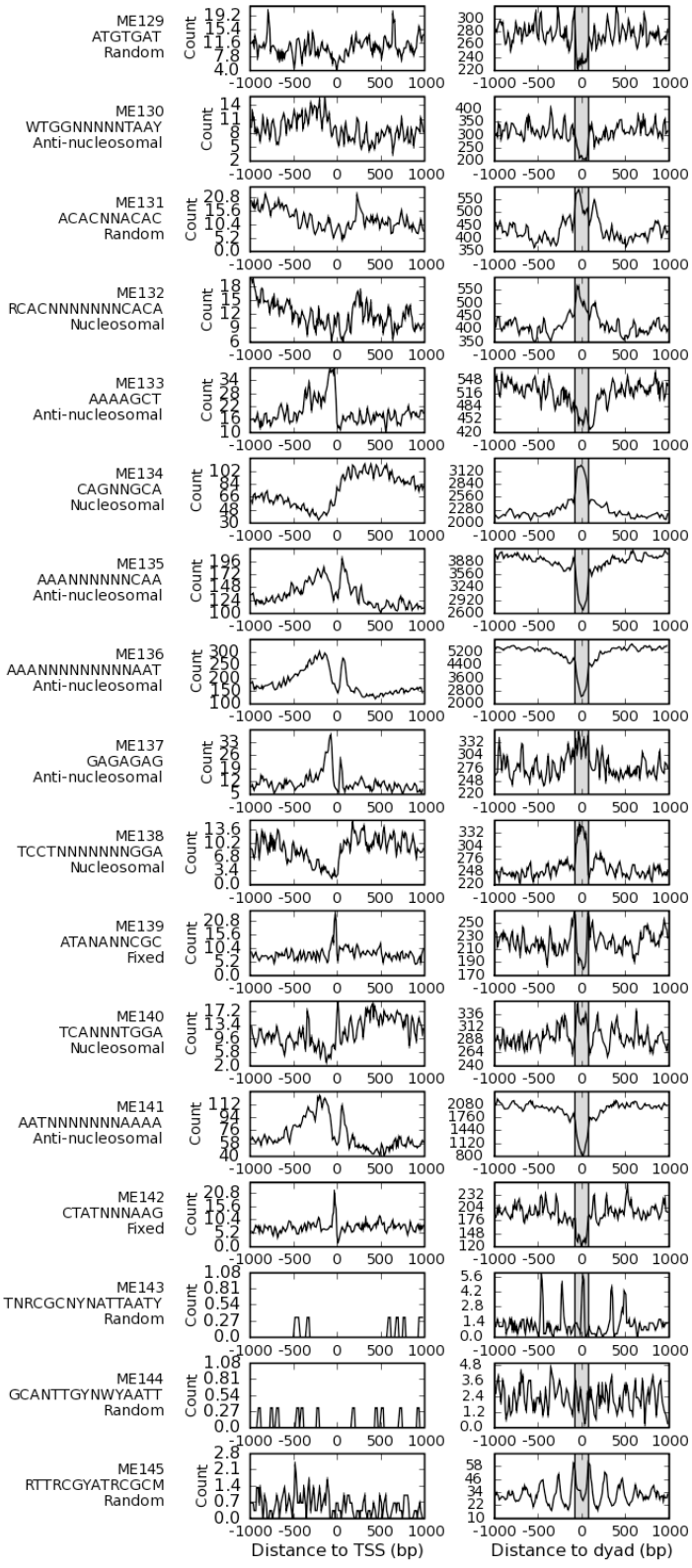

Abdominal B
VSNKTDATKRCNV
Random
achaete
RRNNNMCACCTGC
Nucleosomal
Adh1
BMGYBGYYGYNGMVBV
Nucleosomal
Adh2
VSGYYGCMGYCGYYGMMKKYG
Random
Adult TF 1
CAACAA
Nucleosomal
Antennapedia
TTAATT
Anti-nucleosomal
AP-1
TGANTCA
Random
apterous
TAAT
Anti-nucleosomal
asense
CAGSTG
Nucleosomal
atonal
AWCAGGTGK
Random
bicoid
VVVBTAATCC
Random
biniou
RTAAATA
Anti-nucleosomal
brachyenteron
RWWNTNRCACYT
Anti-nucleosomal
bric a brac 1
WHWWWWWWWWKK
Anti-nucleosomal
brinker
TGGCGYY
Random
broad
RKAAASA
Anti-nucleosomal
caudal
TTTATG
Anti-nucleosomal
Chorion factor 2
RTATATRTRB
Anti-nucleosomal
CrebA
TGACGTCA
Random
crocodile
RTAAMA
Anti-nucleosomal
Deformed
RATTAMY
Anti-nucleosomal
disordered facets
GGGGAWTCCCY
Random
dorsal
SGGAAA
Nucleosomal
Dorsal interacting protein 3
YYWVNYYWDNYS
Anti-nucleosomal
doublesex
BYRHBACAAWGTDDB
Random
Dref
TATCGATA
Fixed
Ecdysone receptor
VRGKTYAWTGAMMYY
Random
Ecdysone-induced prot 74EF
CMGGAAR
Nucleosomal
empty spiracles
AANTNTAATGACA
Random
engrailed
TAATTR
Anti-nucleosomal
Enhancer of split
TGGCACGTGYYA
Random
escargot
RRCAGGTGB
Nucleosomal
Ets
CHGGAW
Nucleosomal
even skipped
TCAWTTAAMT
Anti-nucleosomal
eyegone
KVRKRNTCACTSRNTVHDB
Random
ftz1
YSAAGGWCRCHRM
Random
fushi tarazu
KYTAATKDNY
Anti-nucleosomal
giant
TWTKACKTAANA
Anti-nucleosomal
glial cells missing
ATGCGGGY
Nucleosomal
grainy head
WCYGGTTT
Random
Count
Distance to TSS (bp)
Distance to dyad (bp)


hairy
CACGCGMC
Random
Heat shock factor
MGAADMGAADMGAAD
Random
Hormone receptor-like in 46
GGGTCA
Nucleosomal
hunchback
SMATAAAAAA
Anti-nucleosomal
knirps
AAHKMTHBCA
Anti-nucleosomal
Kruppel
RAAMGGRTTA
Random
mirror
RWWWASWBDYSKNMW
Anti-nucleosomal
mito TF A
TTATS
Anti-nucleosomal
Myocyte enhancing factor 2
YTAWWWWTAR
Anti-nucleosomal
ocelliless
MVHTAAKCCS
Random
ovo
HSWAACHGH
Random
paired
SVTAATYGATTANS
Random
pangolin
CCTTTGATCTT
Random
pannier
ANHDDBHGATAASSDNNB
Fixed
pleiohomeotic
GCCATT
Random
pointed
GAGGAAGC
Nucleosomal
prospero
CWYBDCY
Nucleosomal
Relish
GGGGAWYCMC
Random
retained
RATTAAW
Anti-nucleosomal
reversed polarity
CAATTA
Anti-nucleosomal
runt
YGYGGTY
Fixed
scalloped
RRAYATTYBKSGVATKVCA
Random
schnurri
GGGGAMWWCCM
Random
scute
GCAGSTGK
Nucleosomal
Serendipity
CKCAKCWCT
Fixed
serpent
KHGATAASR
Fixed
sine oculis
YGATAC
Fixed
single-minded
GTACGTG
Nucleosomal
slow border cells
TTNNRCAATM
Anti-nucleosomal
snail
CAGGTG
Nucleosomal
Stat92E
TTCCSGGAA
Nucleosomal
Suppressor of Hairless
RBYGTGRGAAMCB
Random
tailless
AAATTAA
Anti-nucleosomal
TATA
STATAWAWRSVVV
Fixed
tinman
CACTTRA
Anti-nucleosomal
Tor Responsive Element
TGCTCAATGAA
Random
tramtrack
MKSCMAGGACVHH
Nucleosomal
Trithorax-like
RSWGAGMRHRR
Random
twist
CACATGT
Random
Ultrabithorax
KNVNVBYTAATKRSBHNVD
Random
Count
Distance to TSS (bp)
Distance to dyad (bp)


ultraspiracle
GRGGTCAYS
Nucleosomal
ventral NS defective
TYAAGTGS
Random
zerknullt
WCATTWMM
Anti-nucleosomal
zeste
BNWDTYGAGTGRNHDD
Random
ME1
GTCACGTD
Random
ME6
VATTWGCAT
Random
ME8
HRTCAATCA
Random
ME9
TGACANNNNNNTGACA
Random
ME10
RCGTGNNNNGCAT
Random
ME12
TTAATGATG
Random
ME13
WTGACANBT
Anti-nucleosomal
ME14
YGACMTTGA
Random
ME15
AATTRNNNNCAATT
Anti-nucleosomal
ME17
MAATTNAATT
Anti-nucleosomal
ME19
MATTRRCACNY
Random
ME21
TAATTRANNTTNATG
Random
ME22
WAATGCGCNT
Random
ME23
MATTWRTCA
Anti-nucleosomal
ME24
YAATTWNRYGC
Random
ME26
YGCGTHAATTR
Random
ME27
AATTRYGWCA
Random
ME28
GCGCATGH
Random
ME29
WAATCARCGC
Random
ME31
GCGTSAAA
Random
ME32
YGCGYRTCAWT
Random
ME33
GCGTTGAYA
Random
ME34
AAATKKCATTA
Random
ME36
TGTCAATTG
Random
ME37
WAATKNNNNNCRCGY
Random
ME40
CATTANNNWAATT
Anti-nucleosomal
ME41
TYRACACTTK
Random
ME44
TGACANNTTGAC
Random
ME46
RAGTGAAAGT
Random
ME47
AATTANWNRCGC
Random
ME48
AATTWWNAYGCR
Anti-nucleosomal
ME50
TRACRYGCA
Random
ME51
RCAAWTTR
Anti-nucleosomal
ME52
RKGTCAAGK
Nucleosomal
ME53
YGTCWAATTA
Anti-nucleosomal
ME54
RCGYRCGY
Fixed
Count
Distance to TSS (bp)
Distance to dyad (bp)


ME55
TAATATGCRA
Random
ME56
TNATGNNTGACA
Random
ME57
RCGTGYYAAAT
Random
ME58
TAATTNWMATT
Anti-nucleosomal
ME59
GCGYNWNAWTGAY
Random
ME60
CATNANTYAAA
Anti-nucleosomal
ME61
AACWAATTR
Anti-nucleosomal
ME63
AAATCAAT
Anti-nucleosomal
ME64
CNNNGCGYRTGANYNAT
Random
ME65
ACGYNGCGTATGM
Random
ME66
AAAATGCA
Anti-nucleosomal
ME67
CATTANYGTCA
Random
ME68
GCATAHWWWNNNGCGY
Random
ME69
AATKACA
Anti-nucleosomal
ME70
GACAATK
Random
ME71
TGTCAMNTGCA
Random
ME72
KCAATAAA
Anti-nucleosomal
ME73
AAAGTGANA
Random
ME75
TAATKRNGTCATTA
Random
ME76
TGACAWWTWTGC
Random
ME77
TGAYWWWTGCA
Anti-nucleosomal
ME78
AATTNNNTCACGY
Random
ME79
GCTNMTTAA
Anti-nucleosomal
ME82
TNAGCATAA
Fixed
ME83
MACMDGTTK
Anti-nucleosomal
ME84
TRTCAWNWRTCA
Random
ME85
TAATTRNNNNYGACA
Random
ME86
YGTCAWTGAC
Random
ME88
ACGNNAATTG
Random
ME89
CACRCAC
Random
ME91
CATNNNNCGCG
Nucleosomal
ME92
TAACCTC
Random
ME94
RAGTKCAANG
Nucleosomal
ME95
AAASTTT
Anti-nucleosomal
ME97
GCGNNTNNTTA
Random
ME98
AATTRNNNNNCAA
Anti-nucleosomal
ME99
GNMCTTGAA
Random
ME101
ATGANNNNTCA
Anti-nucleosomal
ME102
TWAWKMNAWTTG
Anti-nucleosomal
ME103
YRAAMGTGM
Nucleosomal
Count
Distance to TSS (bp)
Distance to dyad (bp)


ME104 CTTNNATAC Anti-nucleosomal
ME105 GYATGMGWAATKA Random
ME106 GGTNTAAAW Fixed
ME107 ACTNACCT Random
ME108 AATNNNNCATNR Anti-nucleosomal
ME109 ATCWATG Random
ME111 ATCTNATC Random
ME114 TTTNGGCGS Random
ME115 GATNTCAT Random
ME117 MAAMNNCAA Random
ME118 GTCNNNNGACA Random
ME121 ACANACA Random
ME122 ACACNNNNRCAC Random
ME123 CACNNNNNNACA Random
ME124 GCANGTCC Random
ME126 CAANNNAAA Anti-nucleosomal
ME127 CACNNRNNNNNNCAC Nucleosomal
Count
Distance to TSS (bp)
Distance to dyad (bp)


ME129 ATGTGAT Random
ME130 WTGGNNNNNTAAY Anti-nucleosomal
ME131 ACACNNACAC Random
ME132 RCACNNNNNNNCACA Nucleosomal
ME133 AAAAGCT Anti-nucleosomal
ME134 CAGNNGCA Nucleosomal
ME135 AAANNNNNNCAA Anti-nucleosomal
ME136 AAANNNNNNNNNAAT Anti-nucleosomal
ME137 GAGAGAG Anti-nucleosomal
ME138 TCCTNNNNNNNGGA Nucleosomal
ME139 ATANANNCGC Fixed
ME140 TCANNNTGGA Nucleosomal
ME141 AATNNNNNNNAAAA Anti-nucleosomal
ME142 CTATNNNAAG Fixed
ME143 TNRCGCNYNATTAATY Random
ME144 GCANTTGYNWYAATT Random
ME145 RTTRCGYATRCGCM Random
Count
Distance to TSS (bp)
Distance to dyad (bp)

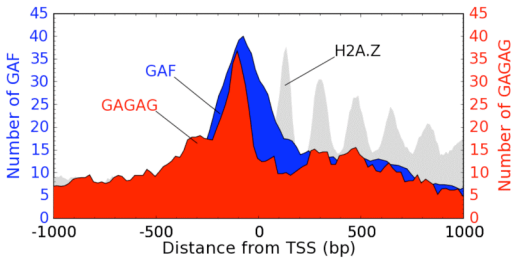

Number of GAF
Number of GAGAG
GAF
GAGAG
H2A.Z
Distance from TSS (bp)

Figure S12. Composite distribution of individual conserved motifs around transcriptional start sites (left panel) and nucleosome dyads. Each motif was binned in 10 bp bins according to distance from the TSS or nucleosome dyad. Plots show a moving average of 3 bins.

To verify that the distribution of at least one motif matched the genome-wide distribution of its cognate DNA binding factor, ChIP-chip was performed with the GAGA associated factor GAF that uses motif ME137: GAGAGAG, and is one of the best characterized *Drosophila* proteins that disrupts nucleosomes[9, 24]. Shown in blue is the distribution of GAF, and in red the distribution of double GAGAG sites, since double sites tend to co-occur[8, 9]. The distribution of H2A.Z nucleosomes from Fig. 1b is shown as a gray backdrop.

Controls for motif organization around nucleosomes. To address the possibility that "nucleosomal" motifs were enriched on H2A.Z nucleosomes only because they were enriched in regions of high nucleosome abundance, and that "anti-nucleosomal" motifs were depleted at H2A.Z nucleosomes because they were generally found in nucleosome-free regions, we artificially rearranged motif locations (Fig. S13). All motifs within 1 kb of the TSS were either 1) removed from the analysis, 2) analyzed separately, or 3-4) had their location relative to the TSS flipped (e.g. a +200 site changed to -200). The latter was analyzed either 3) with or 4) without motifs >1 kb from the TSS included. Except in case 4, all motif rearrangements failed to alter the organizational relationships of motifs with H2A.Z nucleosomes, reflecting the fact that motifs within 1 kb of the TSS represent a small fraction of the total genome-wide motif count. In case 4 (lower right panel), "nucleosomal" motifs were artificially placed in the 5' NFR, and "anti-nucleosomal" motifs were place in nucleosomal regions. When their relationship with resident nucleosomes was examined (and all other motifs excluded), the arrangement was random. These controls indicate that the organization of motifs around H2A.Z nucleosomes is not an indirect consequence of motifs and nucleosomes being organized around the TSS.

We also performed additional controls whereby each motif was scrambled, keeping the same overall base composition. Except for the 'fixed' motifs, distribution patterns were largely unchanged, which indicates that motif organization and biased base composition around nucleosomes are highly intertwined. This has significant implications with regard to the dominant sequence space in which motifs evolve. Thus, for example, nucleosomal motifs would tend to evolve on nucleosomal DNA because nucleosomal DNA is more enriched with the bases that comprise the motif. The same relationship would be true for anti-nucleosomal motifs and linker regions.

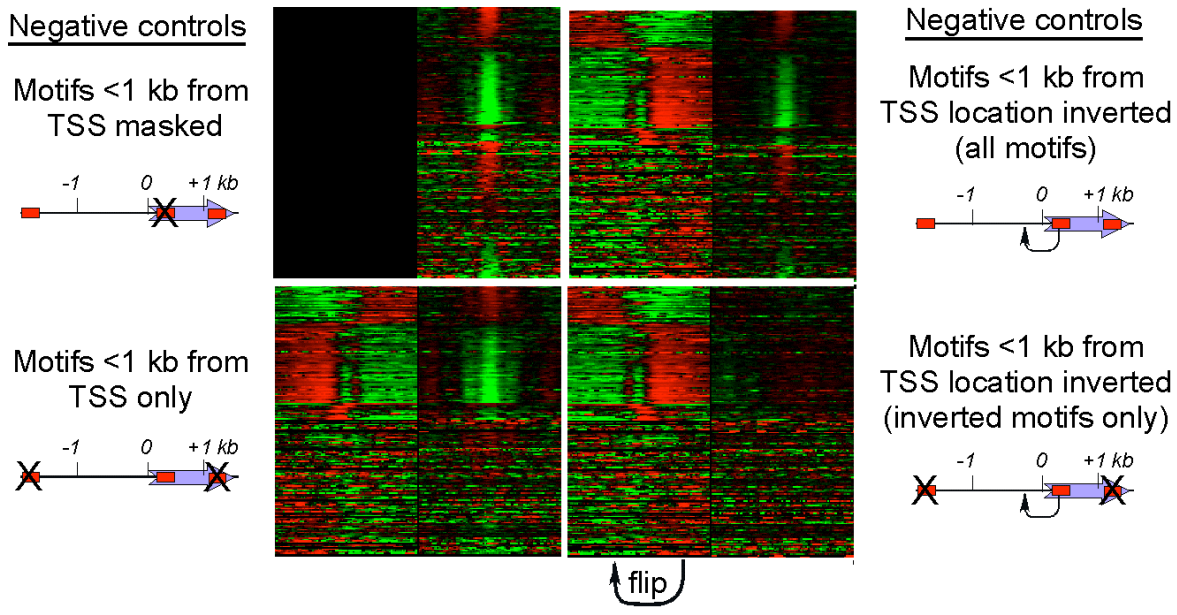


Figure S13. The four panels represent negative controls for Fig. 2b of the main text. Motifs within 1 kb of the TSS were either removed (upper left) or analyzed separately (lower left), or had their relative locations to the TSS inverted (right panels). For the latter, motifs >1 kb from the TSS were either included (upper right) or excluded (lower right) from the analysis. Data within the control panels are organized as in Fig. 2b (data available as Table S4).

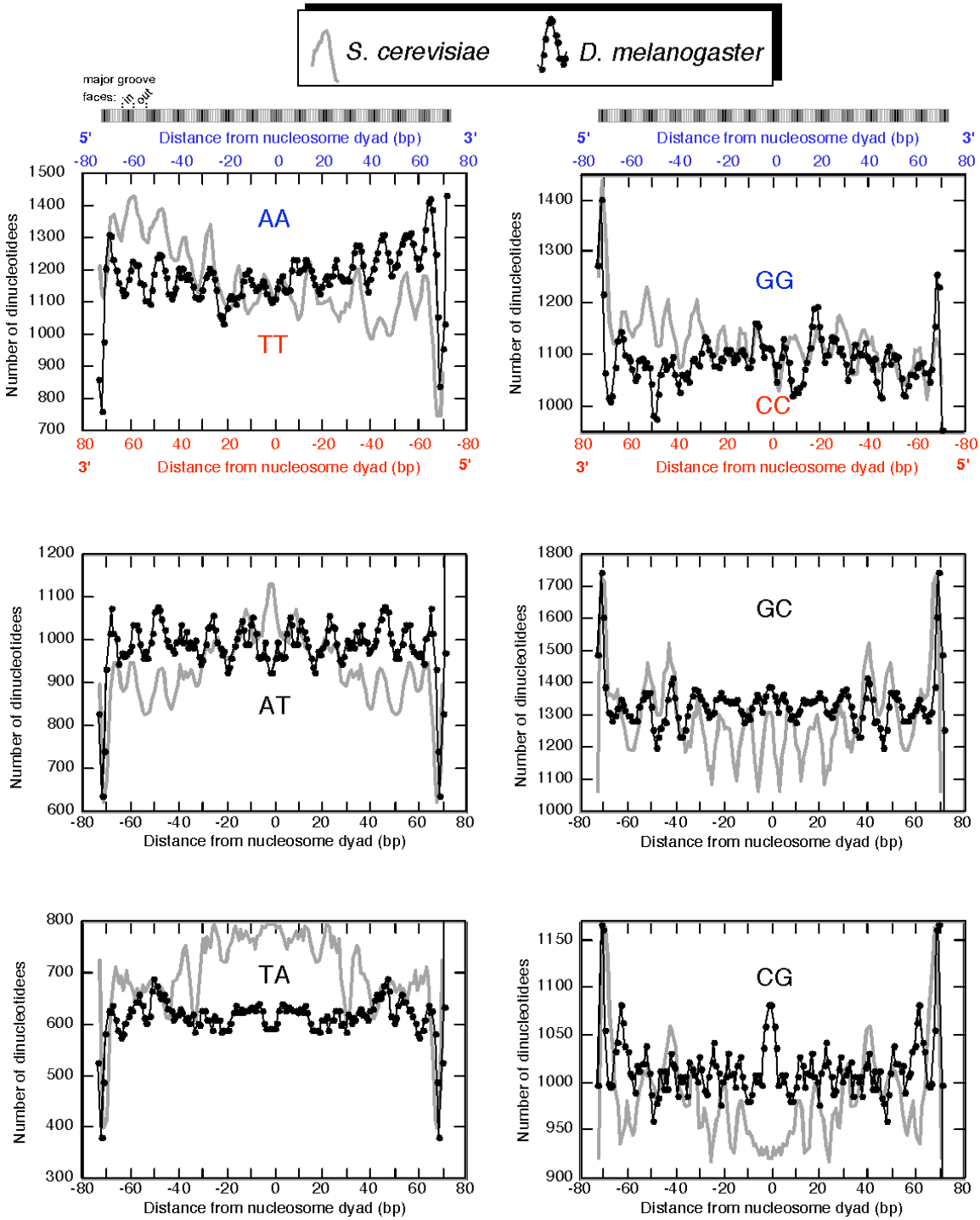


Figure S14 (cont.)

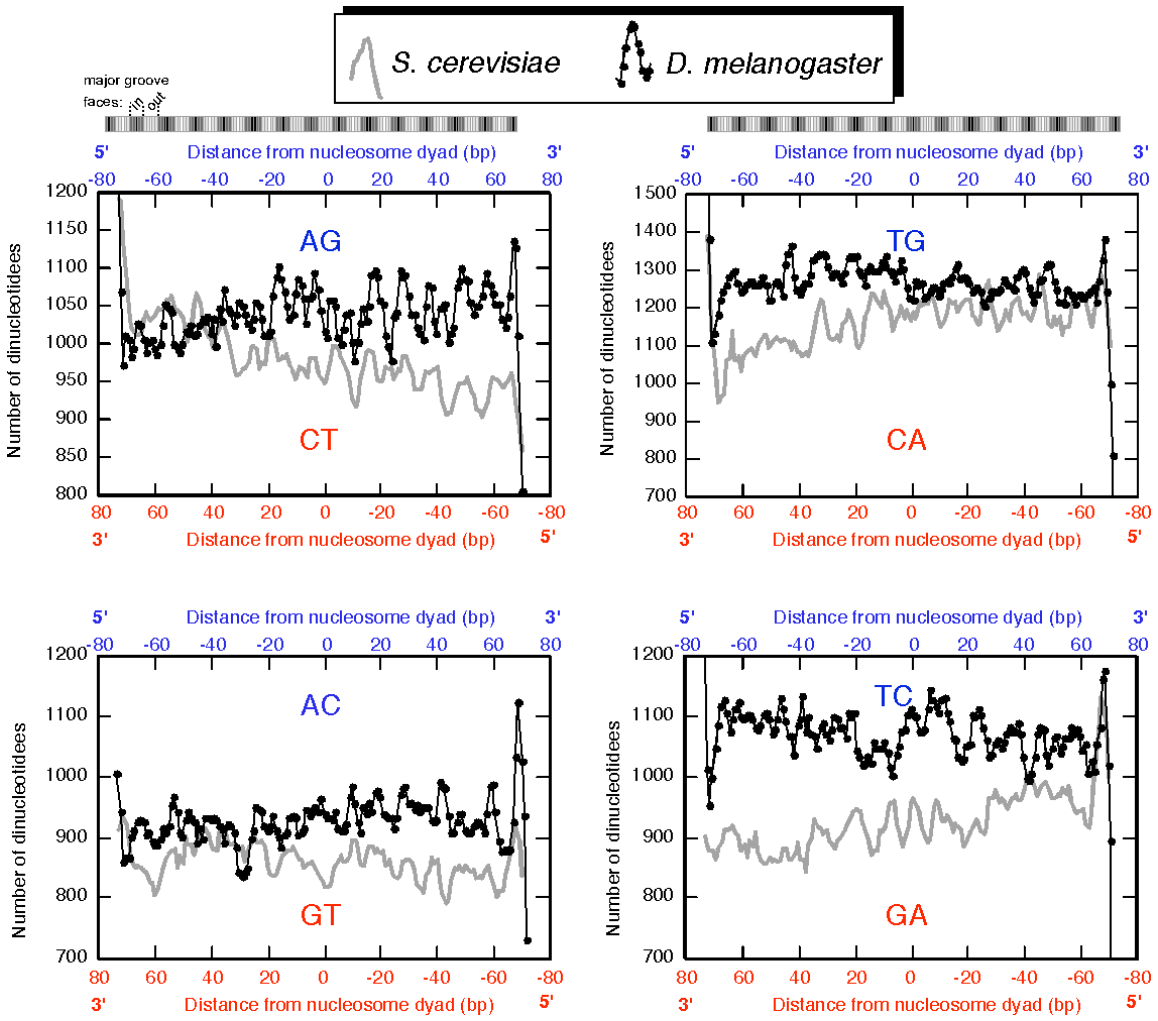


Figure S14. Frequency distribution of the indicated dinucleotide pairs at every position along the 147 bp H2A.Z nucleosomal DNA.

Since each nucleosome contains two DNA strands, the dinucleotide count was determined for both strands, reading in the 5'-3' direction. As a result, the pattern of any dinucleotide (e.g. 5'-GT-3'), read from right to left in red is the same as its reverse complement (e.g. 5'-AC-3'), when the pattern is read from left to right in blue. Therefore, the x-axis on the top of each relevant panel is for one dinucleotide and is reversed at the bottom for the reverse complement dinucleotide (as color coded). Dinucleotide counts at each location were smoothed as described in the methods. The gray trace reports the dinucleotide distribution from *S. cerevisiae*[3].

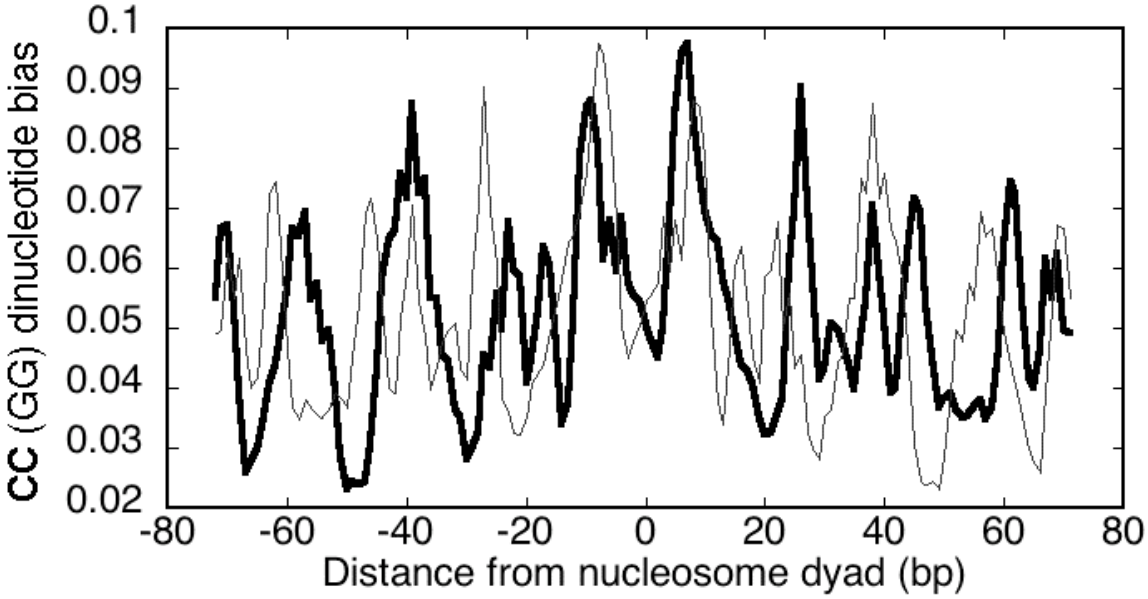


Figure S15. Nucleosome positioning sequence (NPS) patterns used to measure NPS correlations in *D. melanogaster*. Frequency distributions of CC (thick line) and GG (thin line) dinucleotides across selected nucleosomes are shown. Panels are smoothed using 3 bp moving averages.

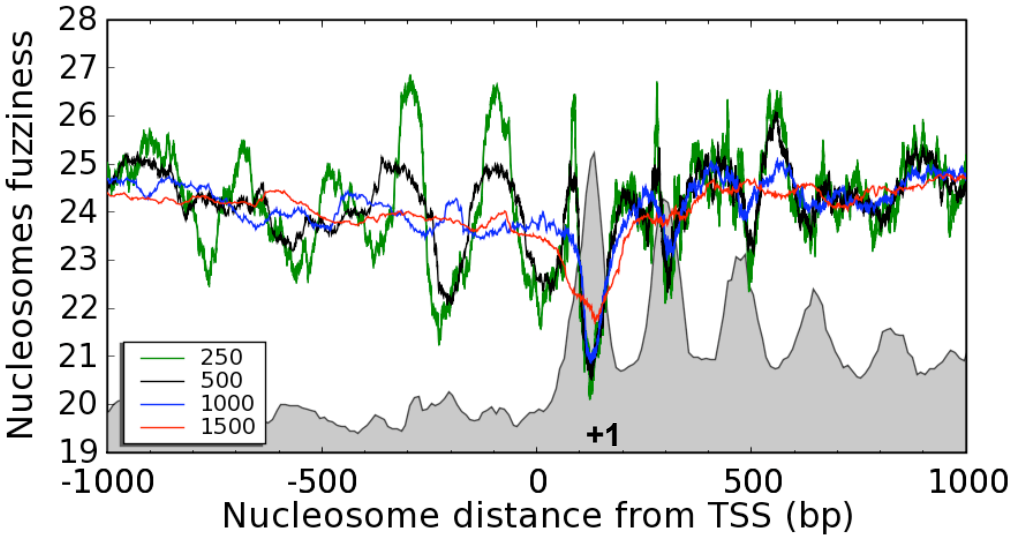


Figure S16. Moving average of H2A.Z nucleosome fuzziness relative to TSSs. The traces represent moving averages of the indicated number of nucleosomes (not bp). Fuzziness is measured as the standard deviation of sequence tag (reads) locations for individual H2A.Z nucleosomes that were defined by at least 7 reads. This minimal read count represents a trade-off between sufficient reads to accurately measure fuzziness and total number of nucleosomes being monitored (49,444 out of 207,025 or 24% of nucleosomes in the region met the criteria). The distribution of H2A.Z nucleosomes from Fig. 1b is shown as a gray backdrop. The low fuzziness of the +1 nucleosome indicates that it is well positioned.

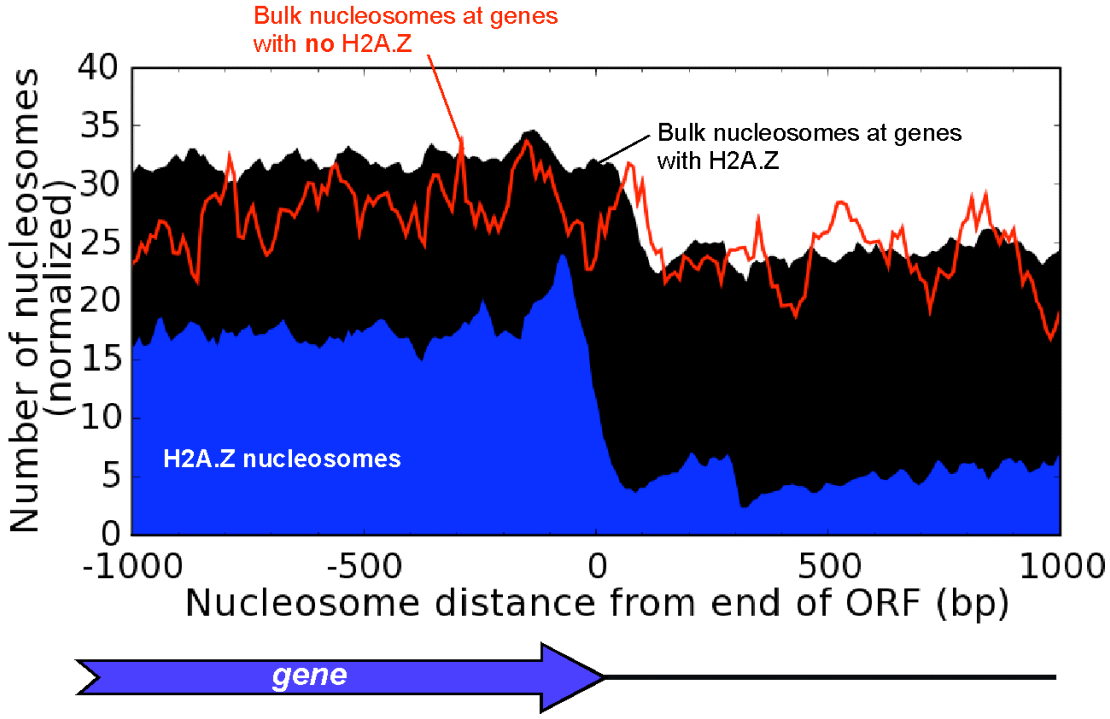


Figure S17. Composite distribution of bulk nucleosomes relative to the ORF termination site. Bulk nucleosome profile for genes containing H2A.Z is shown as a black filled plot (moving average of five 10-bp bins); bulk nucleosome profile for genes lacking H2A.Z is shown as a red trace (moving average of eight 10-bp bins). The distribution of H2A.Z nucleosomes from Fig. 4b is shown in blue. Bulk nucleosomes were detected by Affymetrix tiling arrays having 36 bp average probe spacing.

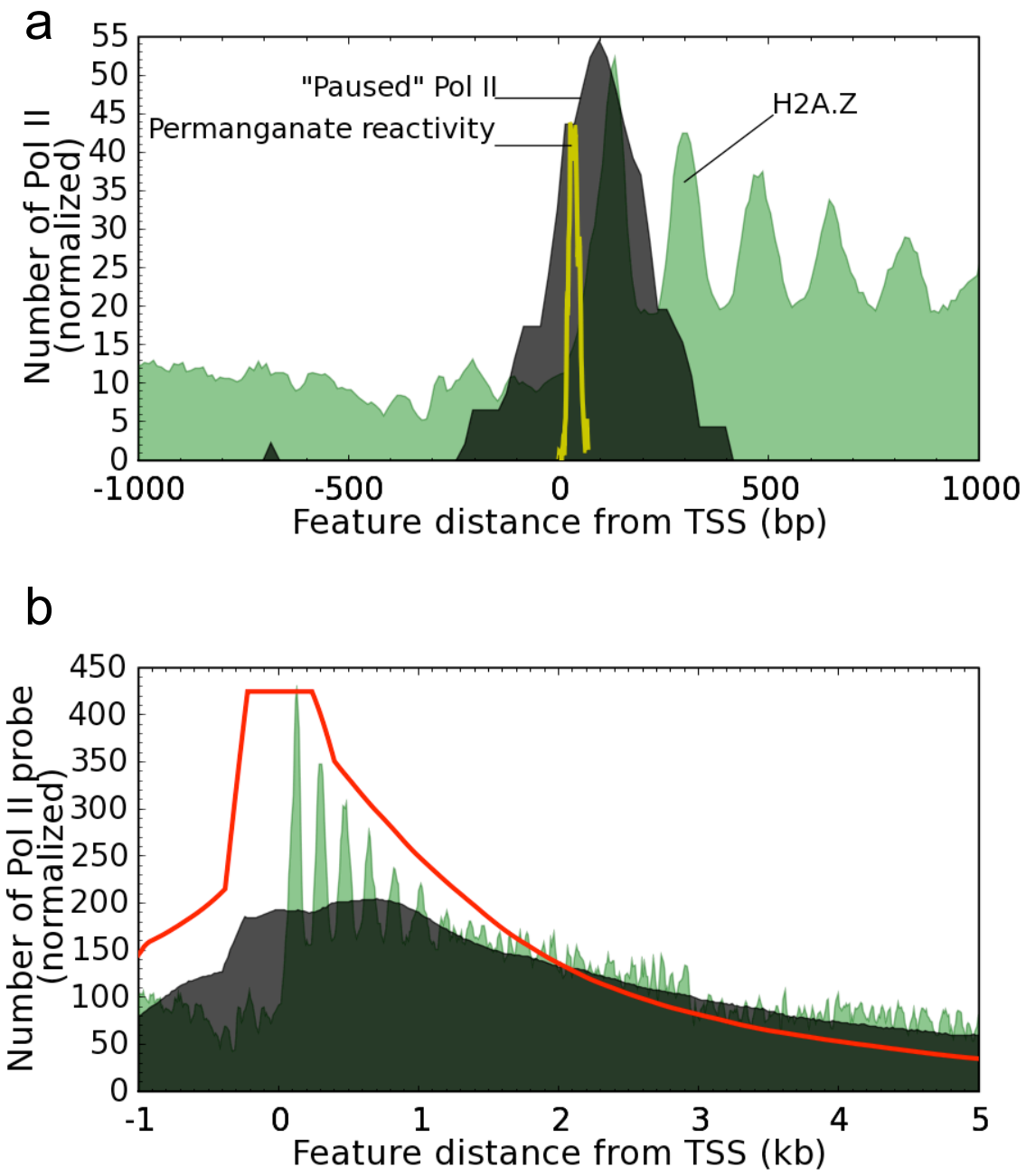


Figure S18. a, Distribution of Pol II (black), as measured by genome-wide ChIP-chip (Affymetrix, 36 bp spacing), displayed for the 50 genes in which pausing was detected by permanganate footprinting. The distribution of permanganate-detected Pol II and H2A.Z nucleosomes is the same as in Fig. 5a.
b, Density of Pol II-bound probes at 8,736 genes that are not in the "Paused" class. The number of Pol II bound probes were binned in 20 bp intervals (black filled plot), and normalized to the number of coding regions present in each bin (red trace). A 300 bp minimum search region was imposed.

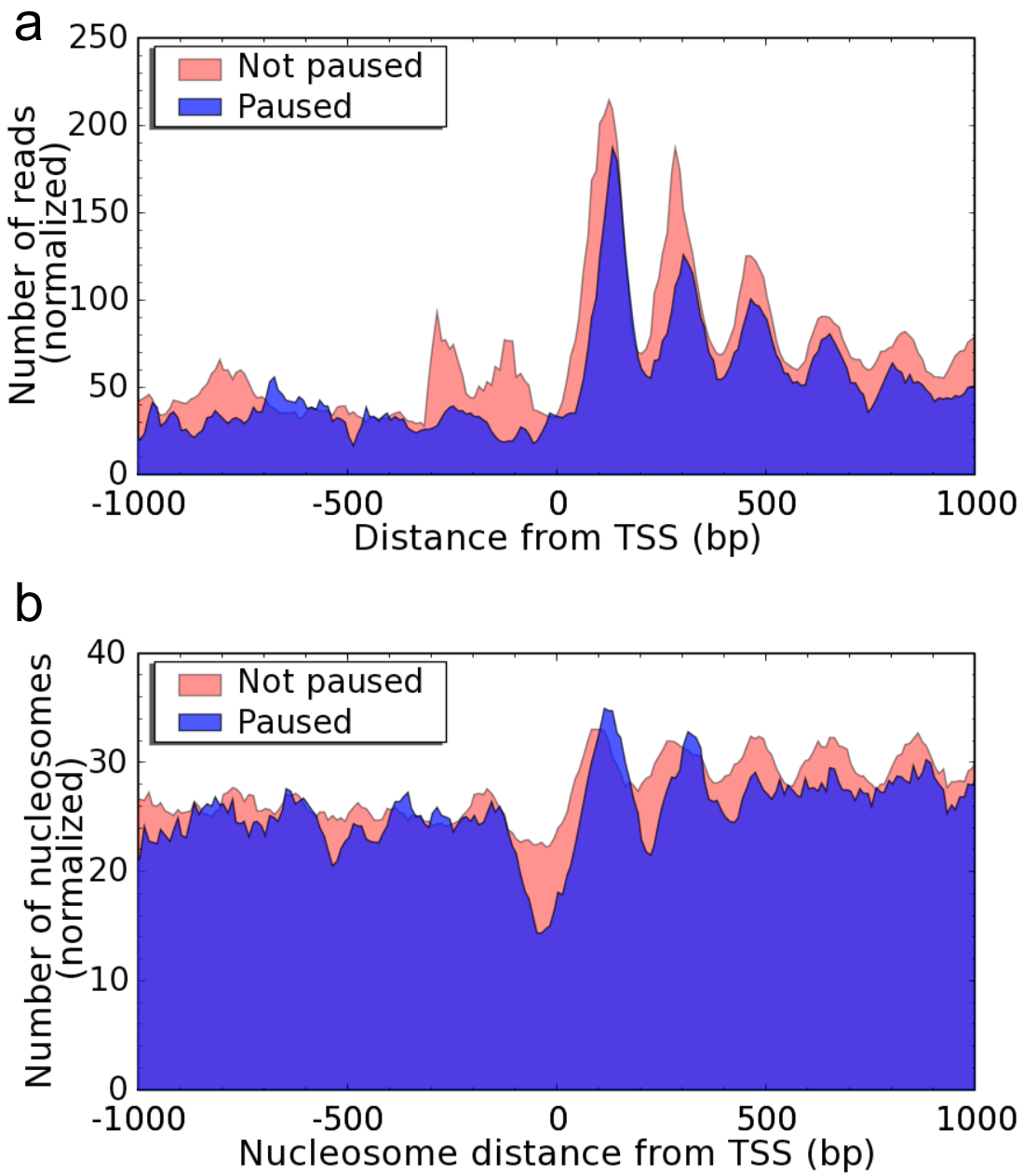


Fig. S19 Distribution of H2A.Z sequencing reads for "Paused" and "Not paused" genes. The plots and set of genes analyzed is similar to Fig. 5b, except in panel **a** the x-axis measures the TSS distance to each nucleosome defined by an individual read, rather than a consensus location defined by a cluster of reads (as shown in Fig. 5b). In panel b, the distribution of bulk nucleosomes is measured using Affymetrix tiling arrays (36 bp average probe spacing).

Supplementary References

1. Fryberg, E. & Goldstein, L. Drosophila melanogaster: Practical Uses in Cell and Molecular Biology. *Meth Cell Biol* **44** (1994).
2. Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J. & Paro, R. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. *Embo J* **17**, 5141-50 (1998).
3. Albert, I. et al. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446**, 572-6 (2007).
4. Leach, T. J. et al. Histone H2A.Z is widely but nonrandomly distributed in chromosomes of Drosophila melanogaster. *J Biol Chem* **275**, 23267-72 (2000).
5. David, L. et al. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* **103**, 5320-5 (2006).
6. Zhang, Z. & Dietrich, F. S. Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. *Nucleic Acids Res* **33**, 2838-51 (2005).
7. Stark, A. et al. Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* **450**, 219-32 (2007).
8. van Steensel, B., Delrow, J. & Bussemaker, H. J. Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding. *Proc Natl Acad Sci U S A* **100**, 2580-5 (2003).
9. Mito, Y., Henikoff, J. G. & Henikoff, S. Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**, 1408-11 (2007).
10. Johnson, W. E. et al. Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci U S A* **103**, 12457-62 (2006).
11. Ioshikhes, I. P., Albert, I., Zanton, S. J. & Pugh, B. F. Nucleosome positions predicted through comparative genomics. *Nat. Genet.* **38**, 1210-5 (2006).
12. Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. & Trifonov, E. N. Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J Mol Biol* **262**, 129-39 (1996).
13. Li, X. et al. Transcription factors bind thousands of active and inactive regions in teh Drosophila genome. *PLoS Biol* **6**, xxx-xxx (2008).
14. Kim, T. H. et al. A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-80 (2005).
15. Hild, M. et al. An integrated gene annotation and transcriptional profiling approach towards the full gene content of the Drosophila genome. *Genome Biol* **5**, R3 (2003).
16. Horz, W. & Altenburger, W. Sequence-specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Res* **9** (1981).
17. Dingwall, C., Lomonossoff, G. P. & Laskey, R. A. High sequence specificity of micrococcal nuclease. *Nucleic Acids Res* **9**, 2659-73 (1981).
18. Louis, C., Schedl, P., Samal, B. & Worcel, A. Chromatin structure of the 5S RNA genes of D. melanogaster. *Cell* **22**, 387-92 (1980).
19. Arbeitman, M. N. et al. Gene expression during the life cycle of Drosophila melanogaster. *Science* **297**, 2270-5 (2002).
20. Gershenzon, N. I., Trifonov, E. N. & Ioshikhes, I. P. The features of Drosophila core promoters revealed by statistical analysis. *BMC Genomics* **7**, 161 (2006).

21. Kutach, A. K. & Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol* **20**, 4754-64 (2000).
22. Ohler, U., Liao, G. C., Niemann, H. & Rubin, G. M. Computational analysis of core promoters in the Drosophila genome. *Genome Biol* **3**, RESEARCH0087 (2002).
23. Celniker, S. E. et al. Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. *Genome Biol* **3**, RESEARCH0079 (2002).
24. Lehmann, M. Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure. *Trends Genet* **20**, 15-22 (2004).

Table 1 (page 1 out of 2676 pages)

name	chrom	strand	start	end	peakheight	readcount	stddev
N104	2L	W	31	177	4.557773055	8.137301817	23.51654264
N296	2L	W	223	369	1.728022429	2.980227848	34.69870315
N560	2L	W	487	633	4.743203455	12.20245634	35.25747784
N754	2L	W	681	827	1.728022429	2.980227848	34.69870315
N1018	2L	W	945	1091	4.743203455	12.20245634	35.25747784
N1212	2L	W	1139	1285	1.728022429	2.980227848	34.69870315
N1476	2L	W	1403	1549	4.743203455	12.20245634	35.25747784
N1670	2L	W	1597	1743	1.728022429	2.980227848	34.69870315
N1934	2L	W	1861	2007	4.743203455	12.20245634	35.25747784
N2128	2L	W	2055	2201	1.728022429	2.980227848	34.69870315
N2392	2L	W	2319	2465	4.743203455	12.20245634	35.25747784
N2586	2L	W	2513	2659	1.728022429	2.980227848	34.69870315
N2850	2L	W	2777	2923	4.743203455	12.20245634	35.25747784
N3044	2L	W	2971	3117	1.728022429	2.980227848	34.69870315
N3308	2L	W	3235	3381	4.743203455	12.20245634	35.25747784
N3502	2L	W	3429	3575	1.728022429	2.980227848	34.69870315
N3766	2L	W	3693	3839	4.743203455	12.20245634	35.25747784
N3960	2L	W	3887	4033	1.728022429	2.980227848	34.69870315
N4224	2L	W	4151	4297	4.743203455	12.20245634	35.25747784
N4418	2L	W	4345	4491	1.728022429	2.980227848	34.69870315
N4682	2L	W	4609	4755	4.743203455	12.20245634	35.25747784
N4876	2L	W	4803	4949	1.728022429	2.980227848	34.69870315
N5099	2L	W	5026	5172	4.065829177	5.763716665	35.11504901
N5297	2L	W	5224	5370	1.901523693	3.238831304	32.00390601
N5615	2L	W	5542	5688	4.95920251	5.457324553	15.2894299
N7681	2L	W	7608	7754	4.93812753	7.417145993	31.90611227
N7840	2L	W	7767	7913	1.836700671	1.881330831	0
N8092	2L	W	8019	8165	0.424640534	0.424640534	0
N8488	2L	W	8415	8561	2.697807877	3.12222464	0
N8758	2L	W	8685	8831	0.559121734	0.559121734	0
N12365	2L	W	12292	12438	1.348836339	1.348792053	0
N12518	2L	W	12445	12591	1.826764304	2.061966826	0
N13606	2L	W	13533	13679	0.542888956	0.542888956	0
N15259	2L	W	15186	15332	0.424640534	0.424640534	0
N18784	2L	W	18711	18857	0.849281069	0.849281069	0
N22167	2L	W	22094	22240	3.886716537	6.21194889	24.26834112
N22544	2L	W	22471	22617	1.283188984	1.283188984	0
N22829	2L	W	22756	22902	0.559121734	0.559121734	0
N23065	2L	W	22992	23138	2.025439235	2.788212316	28.86751346
N23608	2L	W	23535	23681	1.827815847	2.363571782	28.86751346
N23846	2L	W	23773	23919	2.793643496	2.80548235	0
N26482	2L	W	26409	26555	1.456690297	1.456690297	0
N26642	2L	W	26569	26715	0.559121734	0.559121734	0
N28042	2L	W	27969	28115	0.585693307	0.983762268	0
N28733	2L	W	28660	28806	0.605276529	0.605276529	0
N28908	2L	W	28835	28981	1.894035265	2.378709116	0
N30419	2L	W	30346	30492	0.849281069	0.849281069	0

Table 2 (page 1 out of 562 pages)

#UID	feature_name	type	chr	arm	strand	start	end	gene_name
FBtr0089256	CG11023-RA	mRNA	2	L	W	7529	9491	FBgn0031208
FBtr0111298	CG17683-RA	mRNA	2	R	W	18442	20468	FBgn0040002
FBtr0078166	l(2)gl-RB	mRNA	2	L	C	21372	9836	FBgn0002121
FBtr0112919	RhoGAP1A-RA	mRNA	X		W	23878	36728	FBgn0025836
FBtr0078965	auxillin-RA	mRNA	3	R	W	37505	53244	FBgn0037218
FBtr0111224	CG41459-RA	mRNA	3	R	W	41136	41534	FBgn0084021
FBtr0078968	Gel-RB	mRNA	3	R	W	60868	66780	FBgn0010225
FBtr0072463	Lsp1gamma-RA	mRNA	3	L	W	66439	69003	FBgn0002564
FBtr0089687	CG14642-RB	mRNA	3	R	C	68508	66855	FBgn0037222
FBtr0078101	galectin-RA	mRNA	2	L	W	72388	76203	FBgn0031213
FBtr0100135	SP71-RA	mRNA	X		W	72458	96056	FBgn0029128
FBtr0078975	CG14639-RA	mRNA	3	R	W	72744	74040	FBgn0037224
FBtr0072541	CG13405-RA	mRNA	3	L	C	75266	74470	FBgn0035097
FBtr0089178	ci-RA	mRNA	4		C	77667	68334	FBgn0004859
FBtr0078981	CG14644-RA	mRNA	3	R	C	78765	77614	FBgn0037226
FBtr0089175	RpS3A-RA	mRNA	4		C	87863	86745	FBgn0017545
FBtr0072540	CG12483-RA	mRNA	3	L	C	98371	97619	FBgn0040688
FBtr0078979	CG9766-RB	mRNA	3	R	C	103515	94943	FBgn0037229
FBtr0089437	M(2)21AB-RE	mRNA	2	L	W	106903	114432	FBgn0005278
FBtr0111268	CG17665-RA	mRNA	2	R	C	130444	123268	FBgn0039997
FBtr0078150	CG31974-RA	mRNA	2	L	C	142977	141074	FBgn0051974
FBtr0089171	Ank-RC	mRNA	4		C	150378	137015	FBgn0011747
FBtr0072538	CG7051-RB	mRNA	3	L	C	152589	149675	FBgn0035100
FBtr0089445	CG32000-RE	mRNA	4		W	152850	160709	FBgn0052000
FBtr0113120	p130CAS-RC	mRNA	3	L	W	153071	174528	FBgn0035101
FBtr0078148	CG11620-RA	mRNA	2	L	C	153106	151103	FBgn0031226
FBtr0078119	CG3436-RA	mRNA	2	L	W	156029	157666	FBgn0031229
FBtr0091612	CG33635-RA	mRNA	2	L	C	158603	157944	FBgn0053635
FBtr0100649	CG13376-RA	mRNA	X		C	160859	160278	FBgn0029518
FBtr0078951	hkb-RA	mRNA	3	R	C	172372	170728	FBgn0001204
FBtr0072537	CG7049-RA	mRNA	3	L	C	176226	175108	FBgn0035102
FBtr0100021	CG33978-RA	mRNA	4		C	196605	179844	FBgn0053978
FBtr0089153	CG2219-RA	mRNA	4		C	199980	198058	FBgn0039889
FBtr0072480	mri-RB	mRNA	3	L	W	200180	202165	FBgn0035107
FBtr0078913	CG11739-RC	mRNA	3	R	W	204386	206932	FBgn0037239
FBtr0070070	Or1a-RA	mRNA	X		W	209156	210590	FBgn0029521
FBtr0072528	CG7028-RA	mRNA	3	L	C	211462	207884	FBgn0027587
FBtr0089146	CG2316-RD	mRNA	4		C	214448	204842	FBgn0039890
FBtr0072496	CG16971-RB	mRNA	3	L	W	224105	226169	FBgn0035114
FBtr0089112	Crk-RC	mRNA	4		W	230882	233794	FBgn0024811
FBtr0112760	CG34454-RA	mRNA	3	L	C	232302	231737	FBgn0085483
FBtr0112759	CG34453-RA	mRNA	3	L	C	233248	232657	FBgn0085482
FBtr0078923	CG31528-RA	mRNA	3	R	W	242153	243389	FBgn0051528
FBtr0089144	CG31999-RA	mRNA	4		C	246580	235635	FBgn0051999
FBtr0089115	yellow-h-RA	mRNA	4		W	248550	251058	FBgn0039896

Table 3 (page 1 out of 2262 pages)

type	chr	arm	strand	start	end	gene_name	GAGA_start	GAGA_end
mRNA	2	R	W	18024494	18060339	FBgn0000008	18024393	18024397
mRNA	2	R	W	18024494	18060339	FBgn0000008	18024431	18024435
mRNA	2	R	W	18024494	18060339	FBgn0000008	18024716	18024720
mRNA	2	R	W	18024494	18060339	FBgn0000008	18026194	18026198
mRNA	2	R	W	18024494	18060339	FBgn0000008	18027428	18027432
mRNA	2	L	W	11210681	11261081	FBgn0000011	11207691	11207695
mRNA	2	L	W	11210681	11261081	FBgn0000011	11210221	11210225
mRNA	2	L	W	11210681	11261081	FBgn0000011	11212951	11212955
mRNA	2	L	W	11210681	11261081	FBgn0000011	11213668	11213672
mRNA	3	R	C	12655769	12633344	FBgn0000014	12658178	12658174
mRNA	3	R	C	12655769	12633344	FBgn0000014	12656484	12656480
mRNA	3	R	C	12655769	12633344	FBgn0000014	12656369	12656365
mRNA	3	R	C	12655769	12633344	FBgn0000014	12656034	12656030
mRNA	3	R	C	12655769	12633344	FBgn0000014	12655978	12655974
mRNA	3	R	C	12655769	12633344	FBgn0000014	12655964	12655960
mRNA	3	R	C	12655769	12633344	FBgn0000014	12653743	12653739
mRNA	3	R	C	12797958	12752932	FBgn0000015	12800934	12800930
mRNA	3	R	C	12797958	12752932	FBgn0000015	12800351	12800347
mRNA	3	R	C	12797958	12752932	FBgn0000015	12799643	12799639
mRNA	3	R	C	12797958	12752932	FBgn0000015	12799346	12799342
mRNA	3	L	C	16641674	16615461	FBgn0000017	16643243	16643239
mRNA	3	L	C	16641674	16615461	FBgn0000017	16643039	16643035
mRNA	3	L	C	16641674	16615461	FBgn0000017	16641719	16641715
mRNA	3	L	C	16641674	16615461	FBgn0000017	16641685	16641681
mRNA	3	L	C	16641674	16615461	FBgn0000017	16640785	16640781
mRNA	3	L	C	16641674	16615461	FBgn0000017	16640411	16640407
mRNA	3	L	C	16641674	16615461	FBgn0000017	16640319	16640315
mRNA	3	L	C	16641674	16615461	FBgn0000017	16640312	16640308
mRNA	3	L	C	16641674	16615461	FBgn0000017	16638921	16638917
mRNA	3	L	C	16641674	16615461	FBgn0000017	16638886	16638882
mRNA	3	L	C	16641674	16615461	FBgn0000017	16638872	16638868
mRNA	2	L	C	10975273	10973442	FBgn0000018	10974192	10974188
mRNA	2	L	C	10975273	10973442	FBgn0000018	10972283	10972279
mRNA	X		W	264064	265024	FBgn0000022	263233	263237
mRNA	X		W	264064	265024	FBgn0000022	263882	263886
mRNA	X		W	264064	265024	FBgn0000022	263890	263894
mRNA	X		W	264064	265024	FBgn0000022	265430	265434
mRNA	3	R	C	9084590	9053962	FBgn0000024	9085955	9085951
mRNA	3	R	C	9084590	9053962	FBgn0000024	9085013	9085009
mRNA	3	R	C	9084590	9053962	FBgn0000024	9084993	9084989
mRNA	3	R	C	9084590	9053962	FBgn0000024	9084983	9084979
mRNA	3	R	C	9084590	9053962	FBgn0000024	9084977	9084973
mRNA	3	R	C	9084590	9053962	FBgn0000024	9084654	9084650
mRNA	3	R	C	9084590	9053962	FBgn0000024	9084648	9084644

Table 4 (page 1 out of 140 pages)

#Factor	Name	Class	-505	-495	Cont. →
EWEIGHT					
tramtrack	MKSCMAGGACVHH	nucleosomal	-0.33335628	-0.140711202	-0.555748702
Adh1	BMGYBGYYGYNGMVBV	nucleosomal	-0.310110643	-0.189816409	-0.133232881
ME134	CAGNNGCA	nucleosomal	-0.479629913	-0.448603017	-0.460160048
scute	GCAGSTGK	nucleosomal	-0.195146309	-0.29876894	-0.345563152
achaete	RRNNNMCACCTGC	nucleosomal	-0.648805406	-1.648805406	-0.911839812
escargot	RRCAGGTGB	nucleosomal	-0.355230163	-0.289641821	-0.196532417
ultraspiracle	GRGGTCAYS	nucleosomal	0.334018782	0.401132978	-0.151408045
glial cells missing	ATGCGGGY	nucleosomal	0.055592326	0.133594838	-0.114332675
ME127	CACNNRNNNNNNCAC	nucleosomal	0.005517621	-0.058258289	-0.066432221
snail	CAGGTG	nucleosomal	-0.390347225	-0.404219399	-0.41120583
ME138	TCCTNNNNNNNGGA	nucleosomal	-0.103372795	-0.077837703	-0.003837122
asense	CAGSTG	nucleosomal	-0.27481088	-0.261910391	-0.323124817
ME91	CATNNNNCGCG	nucleosomal	-0.708198616	-0.860201709	-0.506564754
Hormone receptor-like in 46	GGGTCA	nucleosomal	0.074877816	0.074877816	0.019595381
pointed	GAGGAAGC	nucleosomal	-0.413033955	-0.413033955	-0.654042054
ME132	RCACNNNNNNNCACA	nucleosomal	0.099334204	0.013942713	-0.003759289
Ets	CHGGAW	nucleosomal	-0.238546864	-0.223354309	-0.21639596
prospero	CWYBDCY	nucleosomal	0.039913579	0.018743521	0.023226264
Ecdysone-induced protein 74EF	CMGGAAR	nucleosomal	-0.158696416	-0.196334074	-0.228466694
ME52	RKGTCAAGK	nucleosomal	-0.35824438	-0.232713497	-0.294114042
single-minded	GTACGTG	nucleosomal	-0.541639012	-0.421344778	-0.310313466
Stat92E	TTCCSGGAA	nucleosomal	-0.141554144	-0.878519738	-0.463482239
ME140	TCANNNTGGA	nucleosomal	-0.325474097	-0.446780393	-0.446780393
dorsal	SGGAAA	nucleosomal	-0.041644302	-0.023996436	0.017071813
ME94	RAGTKCAANG	nucleosomal	-0.641446145	-0.641446145	-0.278876066
Adult TF 1	CAACAA	nucleosomal	-0.231152604	-0.250555136	-0.23390847
ME103	YRAAMGTGM	nucleosomal	-0.315123819	-0.260676035	-0.076964081
ME137	GAGAGAG	anti-nucleosomal	-0.18378701	-0.335790103	-0.335790103
ME83	MACMDGTTK	anti-nucleosomal	0.072603406	0.191247903	0.237941138
Dorsal interacting protein 3	YYWVNYYWDNYS	anti-nucleosomal	0.117147898	0.126586123	0.138630957
ME13	WTGACANBT	anti-nucleosomal	0.389963904	0.426184092	0.36530185
tinman	CACTTRA	anti-nucleosomal	0.310523812	0.326205434	0.125740851
ME133	AAAAGCT	anti-nucleosomal	-0.154578385	-0.154578385	-0.106484097