Supplementary text

Supplementary text 1: Nucleosome occupancy and rotational positioning

Nucleosome occupancy is defined as the density of histone octamers. We examined 200-bp windows across the genome and compared nucleosome occupancy before and after genome activation using DiNuP (Fu et al., 2012). We found that only 0.08% of the genome showed a more than 2-fold occupancy change during the maternal-zygotic transition (Figure S2A). Thus, nucleosome occupancy does not change in the majority of genomic regions during zygotic genome activation. It has previously been reported that GC content can influence nucleosome occupancy (Hughes and Rando, 2009). We evaluated the role of DNA sequence preference in nucleosome occupancy and, indeed, observed a strong correlation between GC content and nucleosome occupancy both before and after genome activation (Figure S2B).

Rotational positioning is defined as the orientation of the DNA helix on the nucleosome surface. We determined the relative distance between nearby reads on the same strand and observed a 10-bp periodicity in both developmental stages (Figure S2C), which is a consequence of rotational positioning (Albert et al., 2007; Valouev et al., 2008). Like nucleosome occupancy, rotational positioning has been shown to be influenced by DNA sequence, especially the oscillating 10-base periodicity of AA/TT/AT dinucleotides (Satchwell et al., 1986; Segal et al., 2006; Zhang et al., 2009). To test the role of DNA sequence in our data, we aligned all non-redundant reads by their 5' ends and converted them to binary sequence according to whether AA/TT/AT was present at each dinucleotide position as described previously (Zhang et al., 2009). We found a clear 10-base AA/TT/AT periodicity in nucleosomal DNA both before and after zygotic genome activation (Figure S2D).

Since the genomic sequence of the zebrafish remains unchanged during genome activation, it is not surprising that neither nucleosome occupancy nor rotational positioning change dramatically during this developmental transition.

Supplementary text 2: Classification of transcribed and non-transcribed genes

To distinguish between transcribed and non-transcribed genes at dome stage, we used three different approaches. In the main text, we used a combination of H3K36me3 and Pol II ChIP-seq enrichment in gene bodies. We chose this approach because the presence of maternally provided transcripts complicates the use of expression data to determine which genes are transcribed at dome stage. For example, a gene that is transcribed both maternally and zygotically might have a similar expression profile as a gene that is transcribed only maternally and whose mRNA persists after the maternal-zygotic transition. Pol II ChIP-seq enrichment at gene bodies is a straightforward measure of transcription elongation, and previous studies have shown that H3K36me3 status is a reliable measure for active transcription (Guttman et al., 2009; Vastenhouw et al., 2010; Wagner and Carpenter, 2012). By using H3K36me3 status and Pol II status in gene bodies at dome stage, we assigned the terms "transcribed" only to genes undergoing transcription elongation, excluding genes in which transcription had been initiated but not elongated.

The detailed approach is as follows. If a gene's H3K36me3 status is high (containing more than 10 ChIP-seq reads in the last 2/3 region of its concatenated exons, and the average H3K36me3 signal in this region was above 4 RPKM), and Pol II gene body status is high (average signal in gene body was above 2 RPKM), the gene was regarded as transcribed. If the H3K36me3 status is low (average signal was less than 1 RPKM), and Pol II gene body status is low (average signal in gene body was less than 1 RPKM), the gene was regarded as non-transcribed. The transcription status for all other genes was defined as moderate. To ensure unambiguous classification between transcribed and non-transcribed genes, the genes with moderate status were excluded from further analysis.

Comparing our approach with standard gene expression analysis confirmed that our method reliably distinguishes transcribed from non-transcribed genes. We analyzed the expression levels of genes before and after the maternal-zygotic transition using gene expression profiling by RNA-Seq. For this analysis, we excluded maternally loaded mRNAs (genes with high maternal mRNA contribution, i.e., expressed above 1 RPKM at either 2/4 cell stage or 256-cell stage) and showed that there is a highly significant difference between 256-cell and dome stage for transcribed genes

(Wilcoxon rank sum test p-value 1.4×10^{-61}) but much weaker for non-transcribed genes (p-value 1.5×10^{-12}) (Figure S9).

In addition to the approach described above, to make the conclusion more solid, we applied two additional approaches to infer transcription status. First, we used a combination of H3K36me3 ChIP-seq status and Traveling Ratio (TR) of RNA Pol II. TR compares the ratio between Pol II density in promoter (from -30 bp to +300 bp relative to TSS) and in gene body (remaining part of gene), providing a measurement for Pol II elongation (Rahl et al., 2010; Reppas et al., 2006). If a gene's H3K36me3 status is high, and TR is smaller than 2, the gene was regarded as transcribed. If the H3K36me3 status is low, and TR is larger than 2, the gene was regarded as non-transcribed.

Third, we also used the classification of a recent study that defined transcription status by distinguishing between parental alleles (Harvey et al., 2013). In that study, "zygotic only" means only zygotically expressed genes, "maternal and zygotic" means genes with both maternal and zygotic expression, and "maternal only" means genes only maternally expressed. Genes defined as "zygotic only" or "maternal and zygotic" are "transcribed" genes at dome stage, while genes defined as "maternal only" are "non-transcribed" genes.

Supplementary text 3: Classification of elongating and non-elongating Pol II

For genes with high Pol II signal at promoters (i.e., with more than 10 ChIP-seq reads in the promoter, and average signal in the promoter larger than 2.5 RPKM), the combination of H3K36me3 status and Pol II TR value was used to distinguish between elongating and non-elongating Pol II. If the H3K36me3 status was high in gene bodies (i.e., with more than 10 reads in the last 2/3 region of concatenated exons, and the average signal in the region larger than 4 RPKM) and TR value was smaller than 2, Pol II status was regarded as elongating. For genes with a low H3K36me3 status (i.e., with the average signal in the last 2/3 region of concatenated exons less than 1 RPKM) and large TR value (> 2), Pol II statuses were regarded as non-elongating.

Supplementary methods

Nucleosome depletion at promoters

For each promoter (from 2 kb upstream to 2 kb downstream of the TSS), a nucleosome occupancy value was calculated for each 100-bp window. The lowest occupancy value was defined as the level of nucleosome depletion at this promoter.

Nucleosome organization in mouse ES cell

Mouse ES cells J1 (129S4/SvJae) were cultured in DMEM with 15% FBS and LIF. Collect 10ml cells in a 15ml BD tube, and fix in 1% formaldehyde for 1 minute at room temperature. Formaldehyde was quenched by adding 1ml 1.25M glycine. Cells were rinsed 3 times in ice-cold PBS, immediately resuspended in cell lysis buffer (10mM Tris-HCl pH7.5/10mM NaCl/0.5%NP40) and lysed for 5 min on ice. Nuclei were collected by centrifugation, washed with ice-cold PBS, collected by centrifugation again and resuspended in digestion buffer (50mM Tris-HCl pH7.5/1mM CaCl2/0.2%Triton X-100). Samples were divided into 200µl aliquots, prewarmed to 37°C and incubated with MNase (stored in MNase storage buffer (10mM HEPES pH7.5/100mM NaCl/1mM CaCl2/50% glycerol) and diluted in MNase dilution buffer (50mM Tris-HCl pH8/10mM NaCl/126mM CaCl2/5% glycerol)). Concentration of MNase and time of incubation were optimized to obtain 80% mononucleosomes as determined by Agilent 2100 BioAnalyzer analysis. Library were prepared using the Illumina sequencing library preparation protocol and sequenced on an Illumina HiSeq 2000. Totally 237 million mapped reads were generated.

Datasets and analysis for mouse ES cell

H3K4me3, H3K27me3, H3K36me3, and Pol II ChIP-seq datasets were derived from GEO with accession numbers: GSM307618, GSM307619, GSM307620, and GSM318444.

All sequenced reads were mapped back to the mouse genome (mm9 assembly). The following analysis procedures were identical to those used for zebrafish datasets:

nucleosome profiling, nucleosome array value calculations, and the calling of histone modification Pol II statuses.

RNA-Seq

RNA-Seq of PolyA-selected mRNA from 256-cell stage zebrafish embryos was performed essentially as described (Pauli et al., 2012).

Supplementary figure legends

Figure S1. Summary of experimental strategy and MNase-Seq datasets. A) Experimental strategy of probing nucleosome organization at 256-cell and dome stage. B) BioAnalyzer traces of MNase digestions to estimate fragment size for each sample. C) Nucleosome fragment length estimation for combined samples of each stage. The number of nucleosome pairs with a given start-to-end distance is indicated. D) Number of sequenced and mapped reads for each MNase-seq experiment. E) Pearson correlations between MNase-seq samples at regions with nucleosome arrays in any of the 4 samples. For each sample, mapped reads were extended to 147 bp towards their 3' ends, and middle 73 bp were taken and piled up to present the nucleosome profile. The mean values of all 147 bp windows in those regions were used to calculate the correlation coefficient.

Figure S2. Nucleosome occupancy and rotational positioning do not change significantly during genome activation. A) Pearson correlation of nucleosome occupancy between two stages (R=0.97). Mapped reads are extended to 147 bp towards their 3' ends and piled up to present the nucleosome occupancy. Median values of all 10-kb windows are used to calculate the correlation coefficient. B) Boxplots show the relationship between GC content and nucleosome occupancy at 256-cell (blue) and dome stages (green). For each 100-bp window, the GC content and the median of nucleosome occupancy value was calculated. All 100 bp windows are evenly grouped into 5 groups based on GC content from low to high. We observe a strong correlation between GC content and nucleosome occupancy at both stages. C) Nucleosome position relationship before (blue) and after (green) the maternal-

zygotic transition. The number of nucleosome pairs with a given start-to-start distance is indicated. We observe a clear 10-bp periodicity at both developmental stages. D) Fraction of AA/TT/TA dinucleotides in nucleosomal DNA when aligned by their 5' ends. We observe a clear 10-base AA/TT/AT periodicity in nucleosomal DNA at both developmental stages.

Figure S3. Formation of well-positioned nucleosome arrays during genome activation cannot be explained by DNA sequence preference. We applied a computational model that predicts nucleosome organization from DNA sequence (Kaplan et al., 2009) to the regions with well-positioned nucleosome arrays at dome stage. The predicted nucleosome positioning correlates weakly with the observed nucleosome organization at dome stage (R=0.12, panel A). For the same regions, the prediction correlates a bit better with the observed nucleosome organization at 256-cell stage (R=0.28, panel B). Nucleosome organization in both panels is represented by observed nucleosome score, as defined previously (Kaplan et al., 2009). C) Heatmaps of nucleosome organization, H3K4me3, H3K27me3, RNA Pol II signal and predicted nucleosome score around TSSs as well as H3K36me3 signal in last 2/3 region of concatenated exons. Genes are ranked by predicted nucleosome score at their promoter. Each horizontal line represents the average signal for 100 genes. Genes were evenly grouped into 20 bins based on predicted nucleosome score at their promoters, and the associated distribution of nucleosome array values at these promoters is given in the boxplot. In heatmaps, short genes (< 1kb) were excluded, and if one gene has multiple annotations, only the one with strongest H3K4me3 signal was kept. A total of 18,890 genes was used.

Figure S4. Genes with well-positioned nucleosome arrays at promoters at dome stage are functionally important. A) Gene Ontology (GO) categories that are significantly overrepresented in the genes with well-positioned nucleosome arrays at their promoters at dome stage. GO analysis was carried out with DAVID Functional Annotation Tools (http://david.abcc.ncifcrf.gov/). B) Percentage of genes with different levels of conservation. Conservation level of zebrafish genes is defined as the number of species with its orthologous gene. The 10 selected species are yeast, worm, fly, stickleback, medaka, fugu, tetraodon, frog, mouse, and human. To assess the enrichment of genes highly conserved across species (8-10 species) the p-value

was calculated with the exact binomial test. C) Percentage of housekeeping genes. Zebrafish housekeeping genes are defined as orthologous to human housekeeping genes (Zhu et al., 2008). P-values were calculated with the exact binomial test. In B and C, Ensembl IDs were used with a total number of 53,734 genes.

Figure S5. Two Pol II ChIP-seq datasets at dome stage with two different antibodies are highly correlated. A) Correlation between signals of two Pol II ChIP-seq datasets obtained with different antibodies. MACS was used to generate Pol II binding profiles (Zhang et al., 2008). Each dot represents the median of profile values of two experiments in a 1-kb region. B) Example profile of Pol II ChIP-seq data around *atp1a1* for the two used antibodies. In all analyses, we combined the data for the two antibodies. C) The correlation between Pol II and H3K4me3 density at promoters.

Figure S6. Co-occurrence of H3K4me3 and well-positioned nucleosome arrays at promoters at dome stage. A) Promoters with H3K4me3 and well-positioned nucleosome arrays show much stronger H3K4me3 enrichment than those with H3K4me3 but no nucleosome arrays. H3K4me3 density is calculated as the $\log_2$ transformed RPKM + 1 for each promoter. B) Venn diagram showing the overlap between well-positioned nucleosome arrays and H3K27me3 peaks at promoters. 100 bp overlap was used as a minimal required cutoff. C) Venn diagram showing the overlap between well-positioned nucleosome arrays and Pol II peaks at promoters. 100 bp overlap was used as a minimal required cutoff. D) Venn diagram showing the overlap between genes with well-positioned nucleosome arrays at promoters and high H3K36me3 status at gene bodies. E) The correlation between H3K27me3 density and nucleosome array value at promoters. H3K27me3 density is calculated as the $\log_2$ transformed RPKM + 1 for each promoter. F) The correlation between Pol II density and nucleosome array value at promoters. Pol II density is calculated as the $\log_2$ transformed RPKM + 1 for each promoter. G) The correlation between Pol II density on the the last 2/3 of gene bodies and nucleosome array value at promoters. Pol II density is calculated as the $\log_2$ transformed RPKM + 1 for the last 2/3 of the gene body. H) The correlation between H3K36me3 density at the last 2/3 of concatenated exons and nucleosome array value at promoters. H3K36me3 density is calculated as the $\log_2$ transformed RPKM + 1 for the last 2/3 of concatenated exons.

Figure S7. Heatmaps of nucleosome organization, H3K4me3, H3K27me3 and RNA Pol II signal around TSSs as well as H3K36me3 signal in the last 2/3 of concatenated exons. Genes are ranked by: A) H3K27me3 signal at promoters, B) H3K36me3 signal in gene bodies, C) Pol II signal at promoters, and D) Pol II signal in gene bodies. Each horizontal line represents the average signal for 100 genes. For each panel, genes were evenly grouped into 20 bins, and the associated distribution of nucleosome array values at these promoters is given in the boxplot. In heatmaps, short genes (< 1kb) were excluded, and if one gene has multiple annotations, only the one with the strongest H3K4me3 signal was kept. A total of 18,890 genes was used. Color represents RPKM value.

Figure S8. Co-occurrence of H3K4me3 and well-positioned nucleosome arrays at promoters in mouse ES cells. Heatmaps of nucleosome organization, H3K4me3, H3K27me3 and RNA Pol II signal around TSSs as well as H3K36me3 signal in the last 2/3 of concatenated exons. Genes are ranked by A) H3K4me3 signal at promoters, B) H3K27me3 signal at promoters, C) Pol II signal at promoters, and D) H3K36me3 signal in gene bodies. Each horizontal line represents the average signal for 100 genes. For each panel, genes were evenly grouped into 20 bins, and the associated distribution of nucleosome array values at these promoters is given in the boxplot. Color represents RPKM value.

Figure S9. Combination of H3K36me3 ChIP-seq enrichment in exons and Pol II enrichment in gene bodies can be used to assess transcription status of genes at dome stage. A) Boxplots of H3K36me3 ChIP-seq density for transcribed and non-transcribed genes at dome stage. Transcription status was inferred by combination of H3K36me3 and Pol II enrichment in gene bodies. H3K36me3 density is calculated as $\log_2$ transformed RPKM + 1 in the last 2/3 region of the concatenated exons of each gene. B) Boxplots of gene expression levels analyzed by RNA-Seq (Pauli et al., 2012) at dome stage. Transcription status was inferred by combination of H3K36me3 and Pol II enrichment in gene bodies. Quantile normalization was performed on RNA-seq datasets at 2/4 cell, 256-cell, and dome stage. To avoid interference by maternal mRNAs, we excluded genes with high maternal mRNA contribution from the analysis. Normalized expression level ($\log_2$ transformed RPKM + 1) of the remaining genes at 256-cell stage is indicated in blue. Expression of the same set of remaining genes at

dome stage is indicated in green. We observe a highly significant difference between the two time-points for the transcribed genes (Wilcoxon rank sum test p-value 1.4×10^{-61}) but much weaker for non-transcribed genes (1.5×10^{-12}).

Figure S10. Accompanying nucleosome profiles of Figure 3 in normalized version at dome stage and un-normalized version at 256-cell stage. A-C) Average promoter nucleosome organization at dome stage in normalized version for each class of genes in Figure 3. For A-C, all reads were extended to 147 bp; the middle 73 bp were taken, piled up, and normalized by the area under curve divided by the range of x-axis. D-F) Average promoter nucleosome organization at 256-cell stage for each class of genes in Figure 3. For D-F, all reads were extended to 147 bp; the middle 73 bp were taken, piled up, and normalized by sequencing coverage. Number of genes used to draw average profile is indicated for each class. Short genes (< 1kb) were excluded. Transcription status was inferred from the combination of H3K36me3 ChIP-seq enrichment in exons and Pol II enrichment in gene bodies.

Figure S11. Classification of genes in Figure 3 is reliable. A-C) Average promoter H3K4me3 profiles at dome stage for each class of genes in Figure 3. D-F) Average promoter H3K36me3 profiles at dome stage for each class of genes in Figure 3. G-I) Average promoter Pol II profiles at dome stage for each class of genes in Figure 3. In each panel, genomic background was indicated as black dashed line. Number of genes used to draw average profile was indicated for each class. Short genes (< 1kb) were excluded. Transcription status was inferred from the combination of H3K36me3 ChIP-seq enrichment in exons and Pol II enrichment in gene bodies.

Figure S12. Canonical nucleosome organization at promoters in the absence of transcriptional elongation. Transcription status was inferred using two alternative approaches. For A-D, transcription status was defined based on the combination of H3K36me3 enrichment and Pol II traveling ratio. A) Average promoter nucleosome organization for groups of genes. B) Average promoter H3K4me3 profile for each group of genes. C) Average H3K36me3 profile for each group of genes. D) Average promoter Pol II profile for each group of genes. For E-H, transcription status was

defined based on parental origin of transcripts (Harvey et al., 2013). “Zygotic only” means only zygotically expressed genes, “maternal and zygotic” means genes with both maternal and zygotic expression, and “maternal only” means genes only maternally expressed. Genes defined as “maternal only” are “non-transcribed” genes. E) Average promoter nucleosome organization for groups of genes. F) Average promoter H3K4me3 profile for each group of genes. G) Average H3K36me3 profile for each group of genes. H) Average promoter Pol II profile for each group of genes. For E-H, Ensembl IDs were used. If one Ensembl ID contained multiple transcripts, all transcripts were treated as different genes for this figure.

Figure S13. Genes with strong Pol II signal in their promoter show weaker nucleosome arrays. A) Example profile of nucleosome organization (green), Pol II (red), and H3K36me3 (black) around *wnt11*, a gene with elongating Pol II. B) Example profile of nucleosome organization (green), Pol II (red), and H3K36me3 (black) around *slc25a29*, a gene with non-elongating Pol II. C) Average promoter nucleosome organization for three groups of genes with elongating Pol II: top 10% (red), 10-50% (blue), and 50-100% (black). Groups refer to Pol II density at promoters from high to low. Genes with top 10% Pol II density show weaker nucleosome arrays. D) Average promoter nucleosome organization for three groups of genes with non-elongating Pol II: top 10% (red), 10-50% (blue), and 50-100% (black). Groups refer to Pol II density at promoters from high to low. Genes with top 10% Pol II density show weaker nucleosome arrays. E) Boxplots of nucleosome array value at promoters for three groups with elongating Pol II. Genes with higher Pol II density at their promoter tend to have lower nucleosome array values. F) Boxplots of nucleosome array value at promoters for three groups with non-elongating Pol II. Genes with higher Pol II density at their promoter tend to have lower nucleosome array values.

Figure S14. Promoter nucleosome organization in the absence of robust Pol II binding. A) Average promoter nucleosome organization for genes marked by H3K4me3 at promoters, but without Pol II (using very stringent cutoff, i.e., Pol II signals at promoters below the genomic average). B) Average promoter Pol II profiles for those genes.

Figure S15. Canonical nucleosome organization at promoters in the absence of transcription elongation in mouse ES cells. A) Average promoter nucleosome organization for transcribed (red) and non-transcribed (blue) genes. Transcription status is inferred from the combination of H3K36me3 ChIP-seq enrichment in exons and Pol II enrichment in gene body. B) Average promoter nucleosome organization for three types of non-transcribed genes: bivalent (marked by both H3K4me3 and H3K27me3 at promoters; red), monovalent (marked by H3K4me3, but not by H3K27me3 at promoters; blue), and non-marked (marked neither by H3K4me3 nor by H3K27me3 at promoters; black) genes. C) Average promoter nucleosome organization for three types of genes: with elongating Pol II (red), with non-elongating Pol II (blue), and marked by H3K4me3 but without Pol II (black). For A-C, all reads were extended to 147 bp; the middle 73 bp were taken, piled up, and normalized by sequencing coverage.

Figure S16. Classification of genes based on their promoter nucleosome organization at 256-cell stage. A) Genes are grouped evenly into five groups based on the promoter nucleosome occupancy at 256-cell stage. G1 represents the lowest nucleosome occupancy, G5 the highest. B) Genes are grouped evenly into five groups based on promoter nucleosome array value at 256-cell stage. G1 represents the lowest nucleosome array value, G5 the highest.

Figure S17. Nucleosome depletion at promoters does not predict H3K4me3 association and transcription. Genes are group evenly into five groups based on promoter nucleosome depletion level at 256-cell stage. G1 represents the lowest nucleosome occupancy, G5 the highest. The sum of the fractions of 5 groups is 1. (Figure S16). A) The fraction of genes marked by H3K4me3 at promoters at oblong (black) and dome stage (grey) in each group. B) The fraction of genes marked by H3K36me3 in exons at dome stage in each group.

Figure S18. Nucleosome organization at promoters predicts transcription. Genes are grouped evenly into five bins based on promoter nucleosome array value at 256-cell stage. G1 represents the lowest nucleosome array value, G5 the highest. The sum of the fractions of 5 groups is 1. A) The fraction of genes transcribed at dome stage in each group. Transcription status was defined based on the combination of H3K36me3

enrichment and Pol II traveling ratio. B) The fraction of genes transcribed at dome stage in each group. Transcription status was defined based on parental origin of transcripts (Harvey et al., 2013), and transcribed genes at dome stage included both "zygotic only" and "maternal and zygotic" genes. Likelihood ratio test of linear correlation is used to calculate p-value.

Figure S19. H3K4me3 is largely absent in promoters at 256-cell stage. Genes were grouped evenly into five groups based on promoter nucleosome array value at 256-cell stage. G1 represents the lowest nucleosome array value, G5 the highest (Figure S16). A) Boxplots of H3K4me3 density from ChIP-chip at preMBT (256-cell stage; blue) and MBT (oblong stage; green). ChIP-chip data are from an independent study (Lindeman et al., 2011). B) Boxplots of H3K4me3 density from ChIP-seq at 256-cell stage (blue) and oblong stage (red). H3K4me3 density is calculated as $\log_2$ transformed RPKM + 1 value for each promoter. H3K4me3 is largely absent in 256-cell stage in all 5 groups in both datasets, while the median of H3K4me3 density at oblong stage consistently increases from G1 to G5.

Supplementary tables

Table S1. Summary of histone modifications and Pol II ChIP-seq datasets. Number of sequenced and mapped reads for each ChIP-seq sample are shown.

ChIP-seq sample	# of sequenced reads	# of mappable reads
H3K4me3 (dome)	65.2 M	25.7 M (39.5%)
H3K27me3 (dome)	67.4 M	14.8 M (21.9%)
H3K36me3 (dome)	56.7 M	22.7 M (39.9%)
Pol II (dome; 4H8)	79.3 M	41.2 M (51.9%)
Pol II (dome; 8WG16)	113.4 M	58.2 M (51.3%)
H3K4me3 (oblong)	61.0 M	13.0 M (21.4%)
H3K4me3 (256-cell)	95.1 M	1.2 M (1.3%)

Supplementary References

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

Fu, K., Tang, Q., Feng, J., Liu, X.S., and Zhang, Y. (2012). DiNuP: a systematic approach to identify regions of differential nucleosome positioning. Bioinformatics *28*, 1965-1971.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P.*, et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Harvey, S.A., Sealy, I., Kettleborough, R., Fenyes, F., White, R., Stemple, D., and Smith, J.C. (2013). Identification of the zebrafish maternal and paternal transcriptomes. Development *140*, 2703-2710.

Hughes, A., and Rando, O. (2009). Chromatin 'programming' by sequence--is there more to the nucleosome code than %GC? Journal of biology *8*, 96.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Leproust, E., Hughes, T., Lieb, J., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Lindeman, L.C., Andersen, I.S., Reiner, A.H., Li, N., Aanes, H., Ostrup, O., Winata, C., Mathavan, S., Muller, F., Alestrom, P.*, et al.* (2011). Prepatterning of developmental gene expression by modified histones before zygotic genome activation. Developmental Cell *21*, 993-1004.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A.*, et al.* (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Research *22*, 577-591.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Reppas, N.B., Wade, J.T., Church, G.M., and Struhl, K. (2006). The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Molecular Cell *24*, 747-757.

Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). Sequence periodicities in chicken nucleosome core DNA. Journal of Molecular Biology *191*, 659-675.

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

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K.*, et al.* (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Research *18*, 1051-1063.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922-926.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol *13*, 115-126.

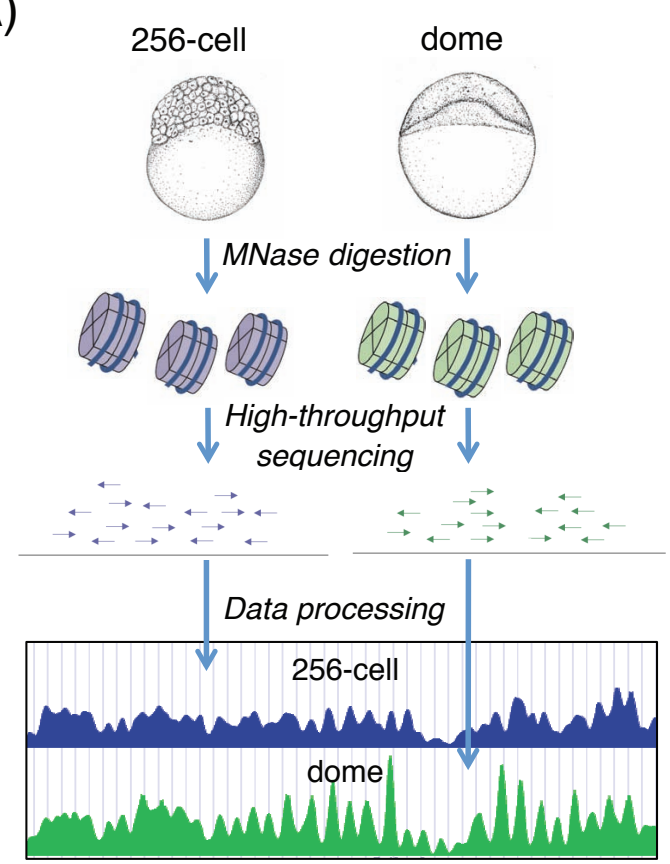
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biology *9*, R137.

Zhang, Y., Moqtaderi, Z., Rattner, B., Euskirchen, G., Snyder, M., Kadonaga, J., Liu, X., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nature Structural & Molecular Biology *16*, 847-852.

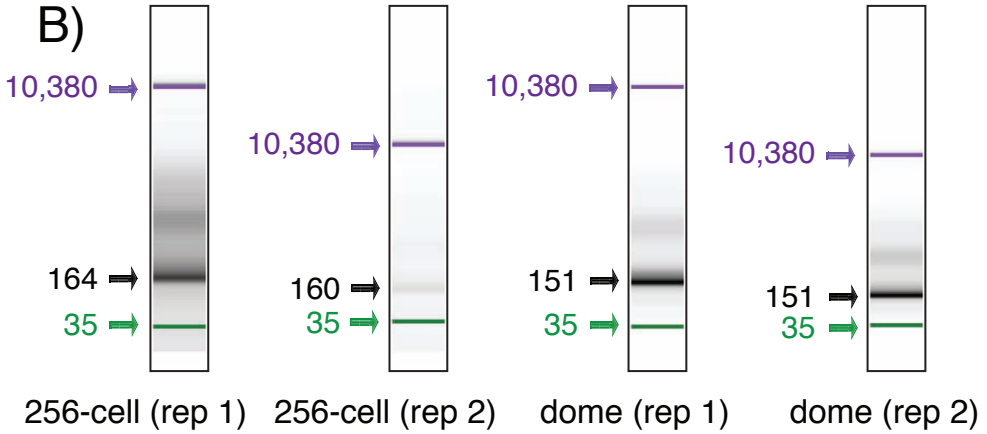
Zhu, J., He, F., Song, S., Wang, J., and Yu, J. (2008). How many human genes can be defined as housekeeping with current expression data? BMC Genomics *9*, 172.

Figure S1

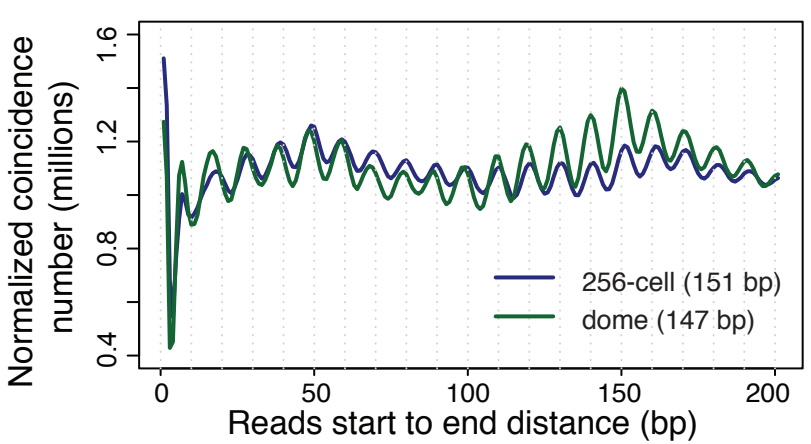
A)



B)



C)



D)

Sample	# of sequenced reads	# of mappable reads
256-cell (rep 1)	206 M	121 M
256-cell (rep 2)	691 M	355 M
dome (rep 1)	215 M	131 M
dome (rep 2)	720 M	388 M

E)

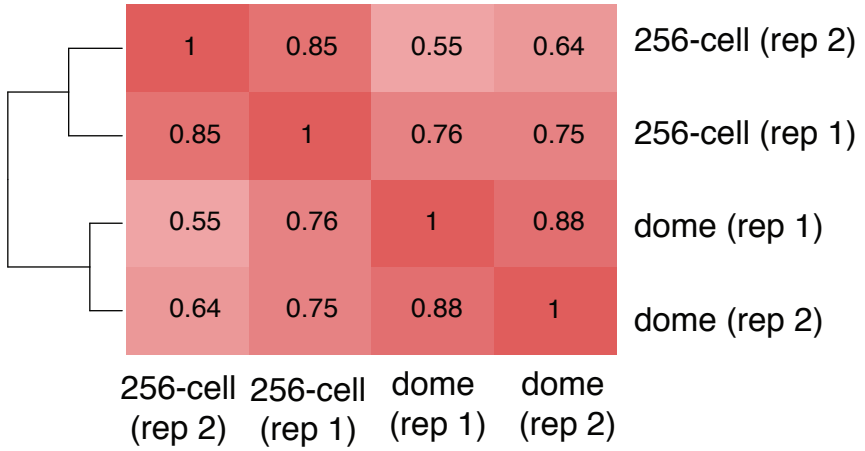


Figure S2

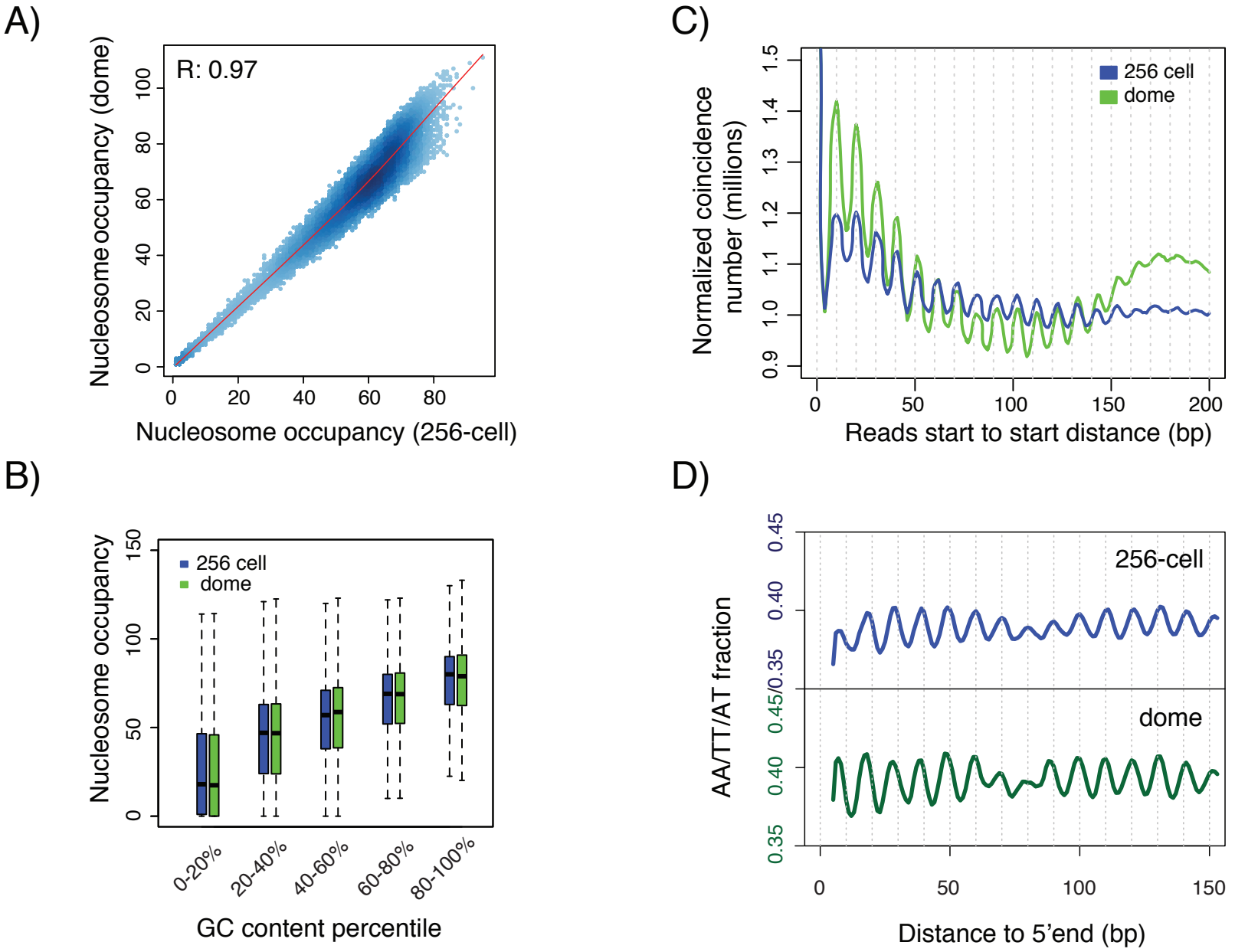
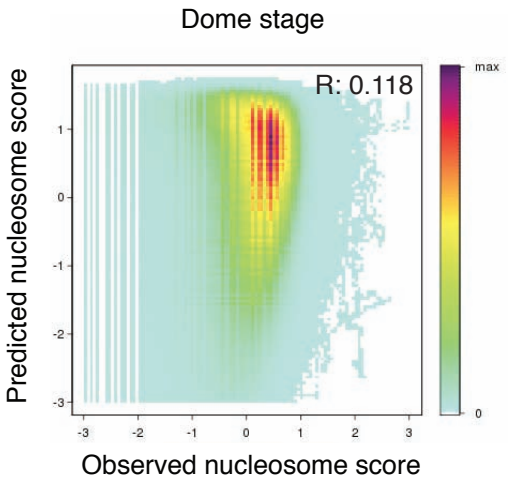
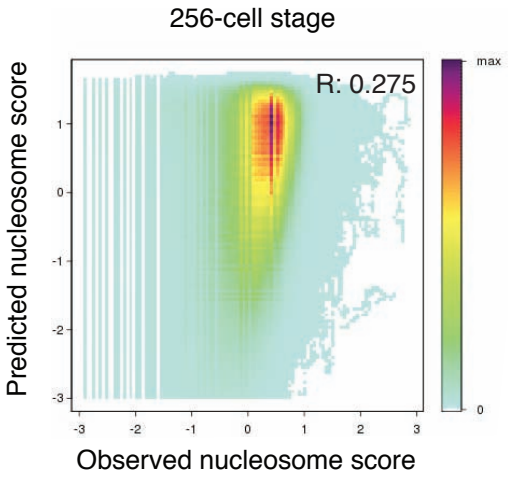

A)
R: 0.97
Nucleosome occupancy (dome)
Nucleosome occupancy (256-cell)
0
20
40
60
80
100
B)
256 cell
dome
Nucleosome occupancy
0
50
100
150
0-20%
20-40%
40-60%
60-80%
80-100%
GC content percentile
C)
256 cell
dome
Normalized coincidence number (millions)
0.9
1.0
1.1
1.2
1.3
1.4
1.5
0
50
100
150
200
Reads start to start distance (bp)
D)
256-cell
dome
AA/TT/AT fraction
0.35
0.40
0.45/0.35
0.40
0.45
0
50
100
150
Distance to 5'end (bp)

Figure S3

A)



B)



C)

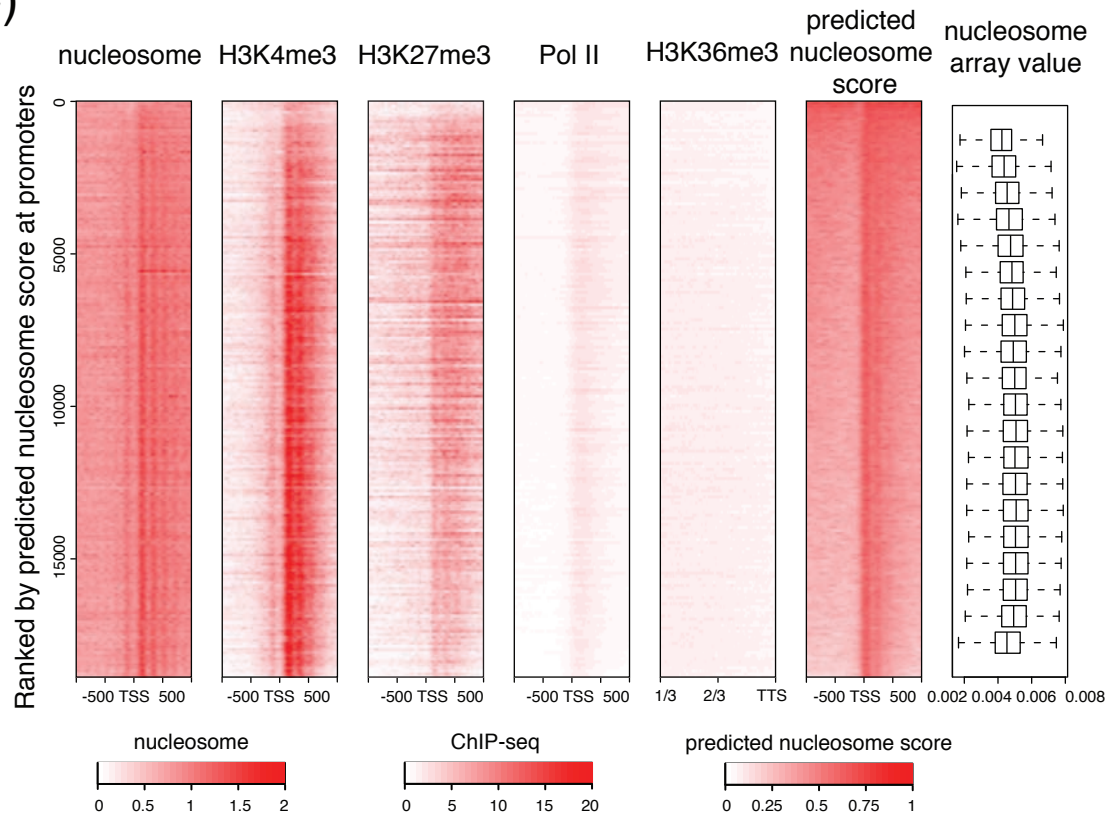


Figure S4

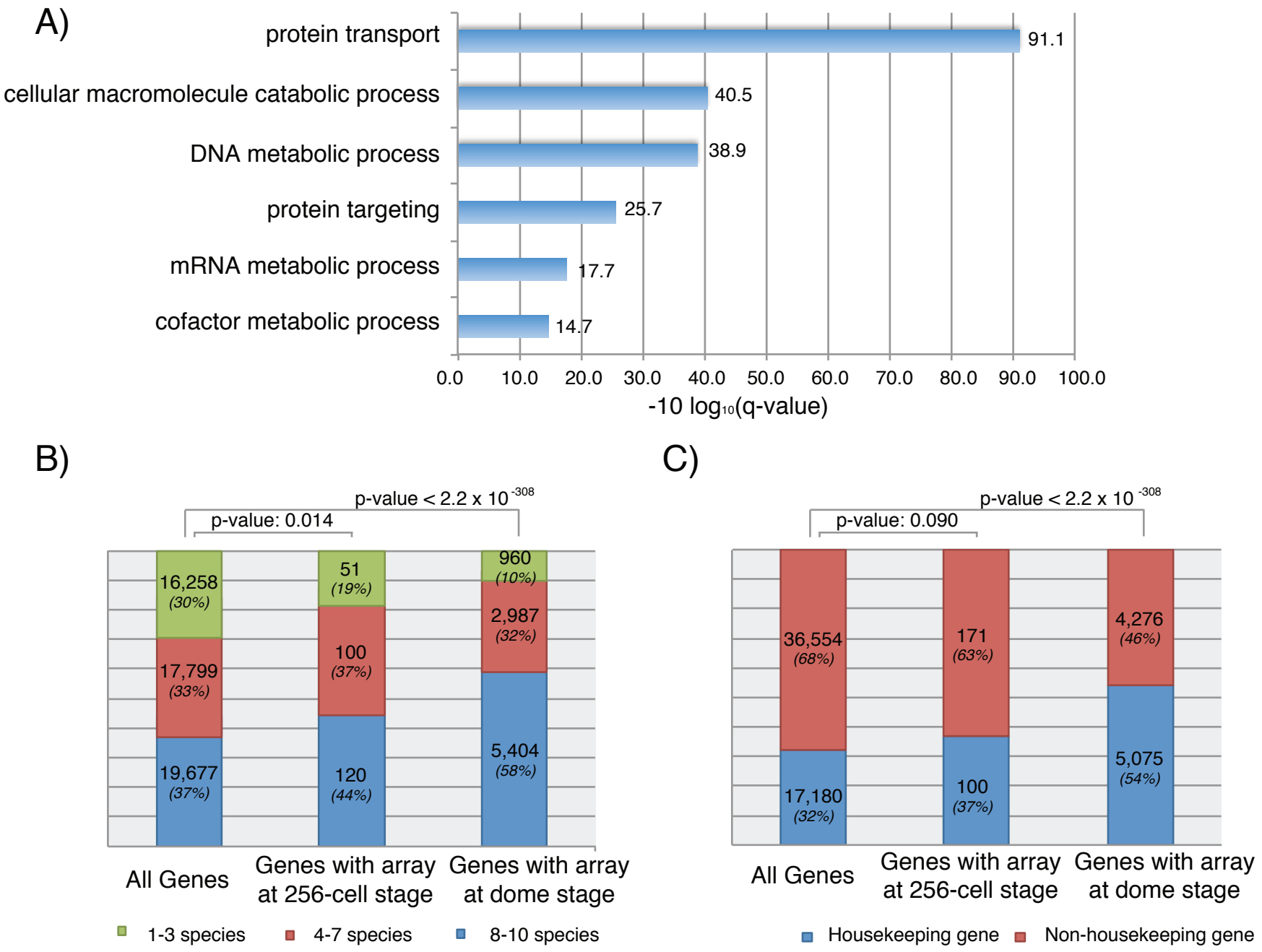
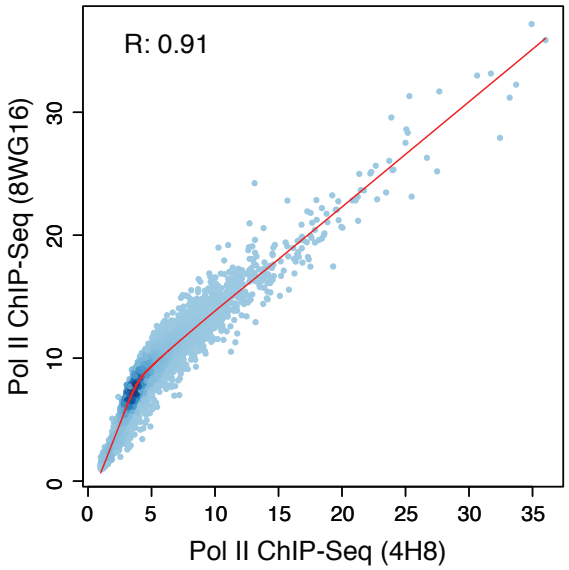
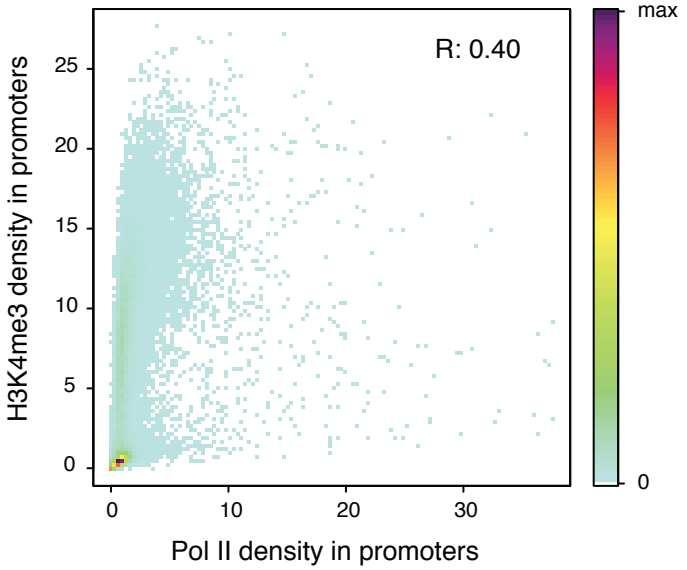

A)
protein transport
cellular macromolecule catabolic process
DNA metabolic process
protein targeting
mRNA metabolic process
cofactor metabolic process
91.1
40.5
38.9
25.7
17.7
14.7
0.0 10.0 20.0 30.0 40.0 50.0 60.0 70.0 80.0 90.0 100.0
-10 log10(q-value)
B)
p-value < 2.2 x 10 -308
p-value: 0.014
16,258 (30%)
17,799 (33%)
19,677 (37%)
51 (19%)
100 (37%)
120 (44%)
960 (10%)
2,987 (32%)
5,404 (58%)
All Genes
Genes with array at 256-cell stage
Genes with array at dome stage
1-3 species
4-7 species
8-10 species
C)
p-value < 2.2 x 10 -308
p-value: 0.090
36,554 (68%)
17,180 (32%)
171 (63%)
100 (37%)
4,276 (46%)
5,075 (54%)
All Genes
Genes with array at 256-cell stage
Genes with array at dome stage
Housekeeping gene
Non-housekeeping gene

Figure S5

A)



C)



B)

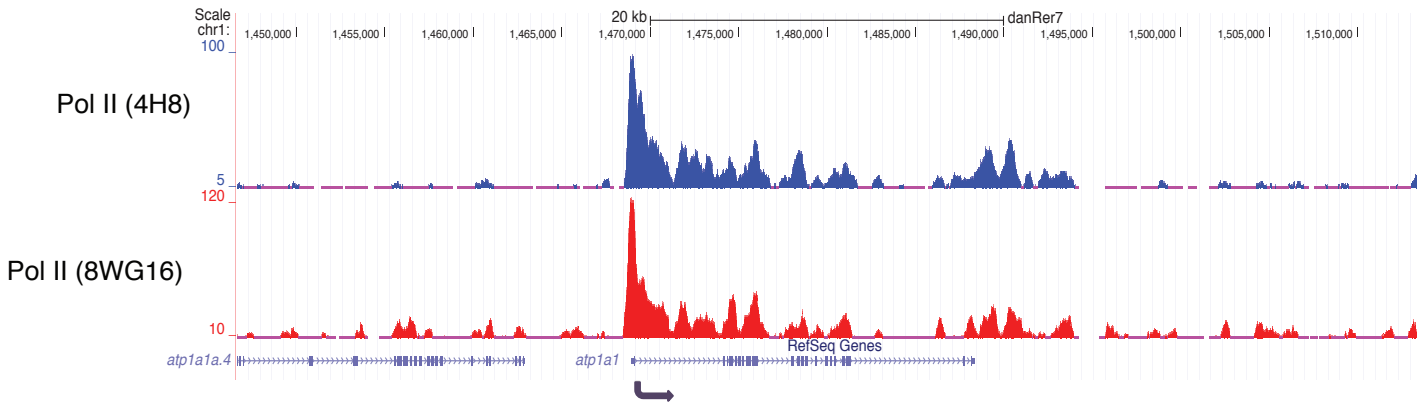


Figure S6

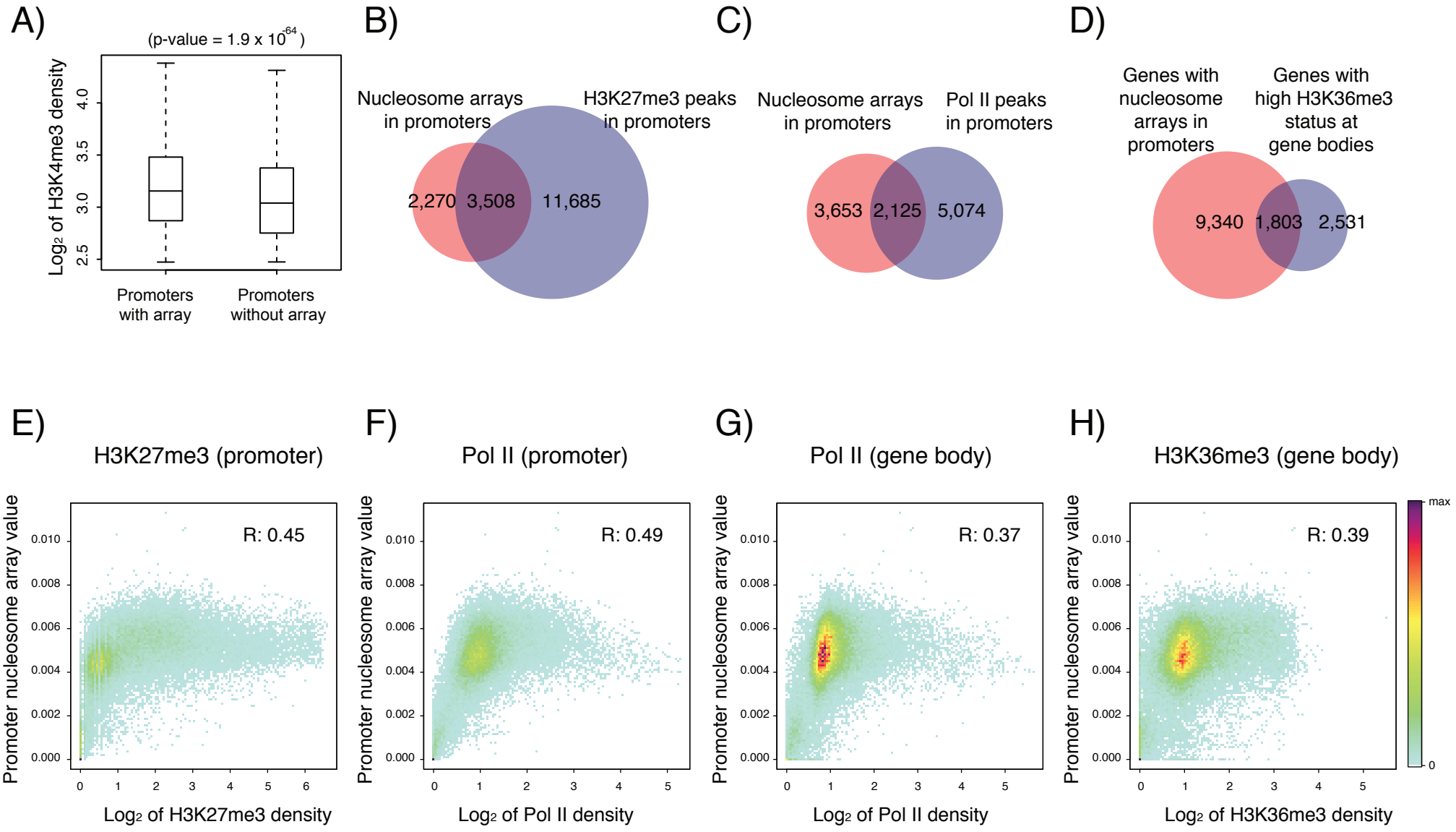

A)
(p-value = 1.9 x 10^-64)
Log2 of H3K4me3 density
2.5
3.0
3.5
4.0
Promoters with array
Promoters without array
B)
Nucleosome arrays in promoters
H3K27me3 peaks in promoters
2,270
3,508
11,685
C)
Nucleosome arrays in promoters
Pol II peaks in promoters
3,653
2,125
5,074
D)
Genes with nucleosome arrays in promoters
Genes with high H3K36me3 status at gene bodies
9,340
1,803
2,531
E)
H3K27me3 (promoter)
R: 0.45
Promoter nucleosome array value
Log2 of H3K27me3 density
F)
Pol II (promoter)
R: 0.49
Promoter nucleosome array value
Log2 of Pol II density
G)
Pol II (gene body)
R: 0.37
Promoter nucleosome array value
Log2 of Pol II density
H)
H3K36me3 (gene body)
R: 0.39
max
0
Promoter nucleosome array value
Log2 of H3K36me3 density

Figure S7

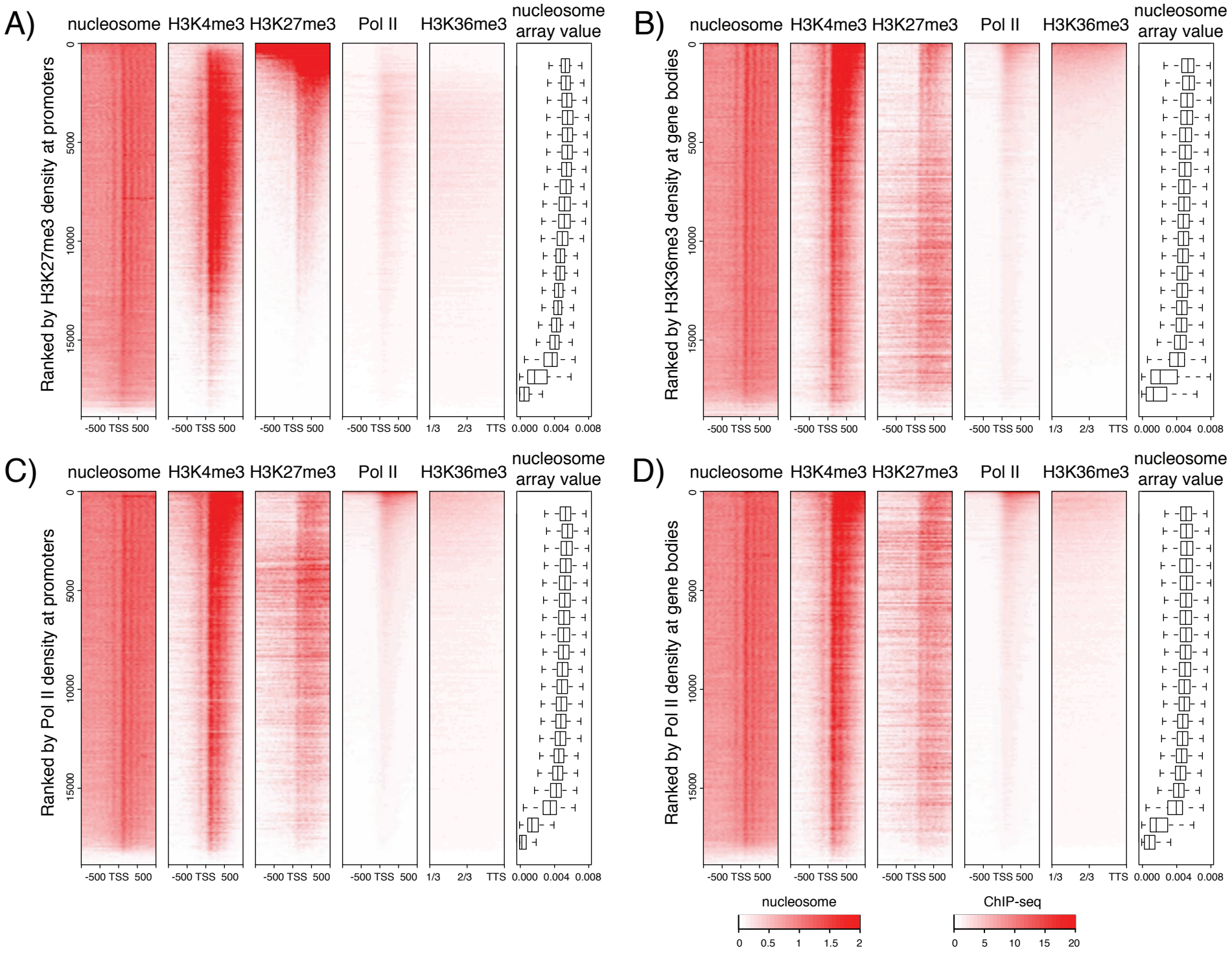

A)
nucleosome
H3K4me3
H3K27me3
Pol II
H3K36me3
nucleosome array value
Ranked by H3K27me3 density at promoters
0
5000
10000
15000
-500 TSS 500
1/3
2/3
TTS
0.000
0.004
0.008
B)
Ranked by H3K36me3 density at gene bodies
C)
Ranked by Pol II density at promoters
D)
Ranked by Pol II density at gene bodies
nucleosome
0
0.5
1
1.5
2
ChIP-seq
0
5
10
15
20

Figure S8

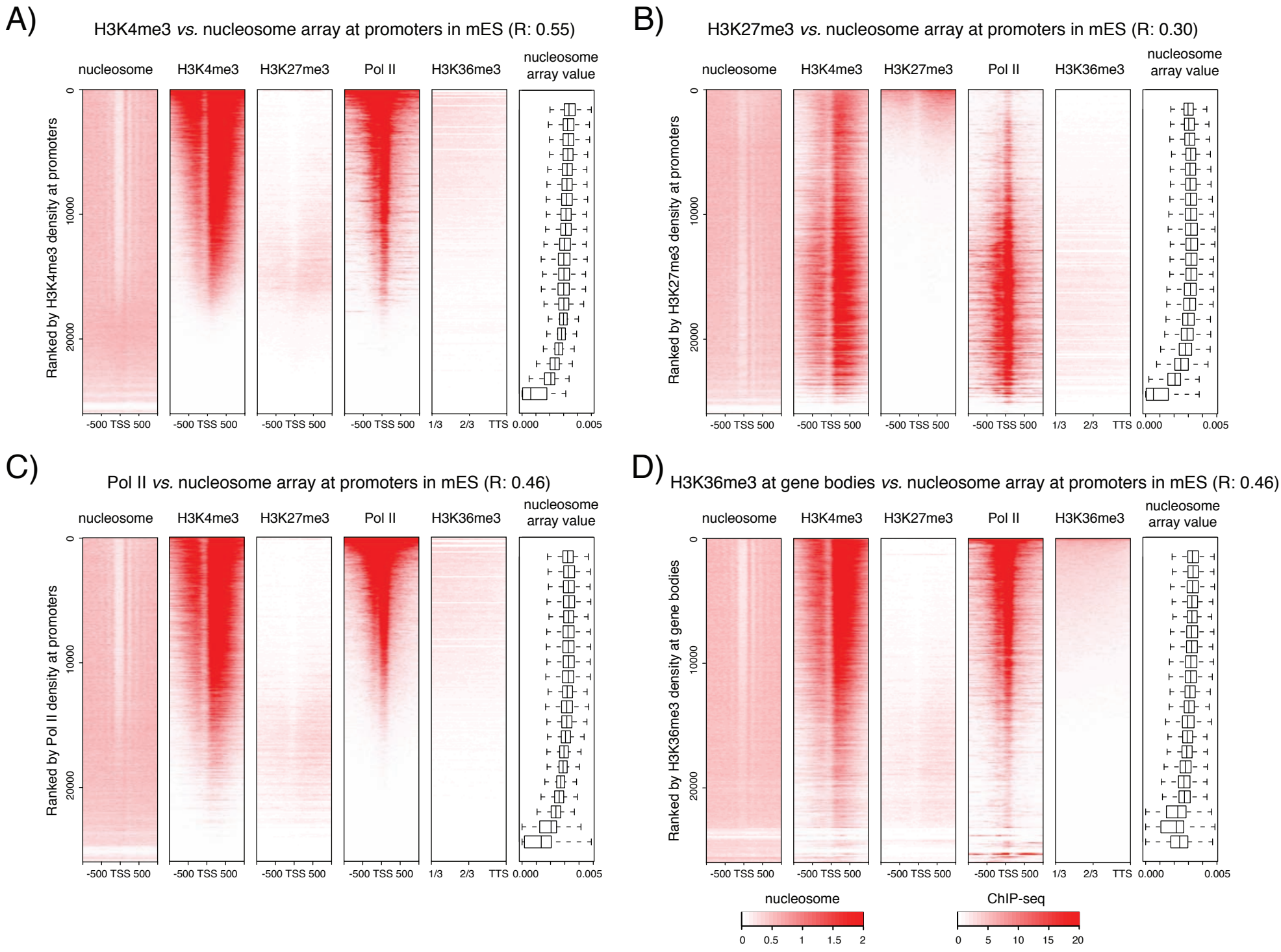

A)
H3K4me3 vs. nucleosome array at promoters in mES (R: 0.55)
nucleosome
H3K4me3
H3K27me3
Pol II
H3K36me3
nucleosome array value
Ranked by H3K4me3 density at promoters
0
10000
20000
-500 TSS 500
1/3
2/3
TTS
0.000
0.005
B)
H3K27me3 vs. nucleosome array at promoters in mES (R: 0.30)
Ranked by H3K27me3 density at promoters
C)
Pol II vs. nucleosome array at promoters in mES (R: 0.46)
Ranked by Pol II density at promoters
D)
H3K36me3 at gene bodies vs. nucleosome array at promoters in mES (R: 0.46)
Ranked by H3K36me3 density at gene bodies
nucleosome
0
0.5
1
1.5
2
ChIP-seq
0
5
10
15
20

Figure S9

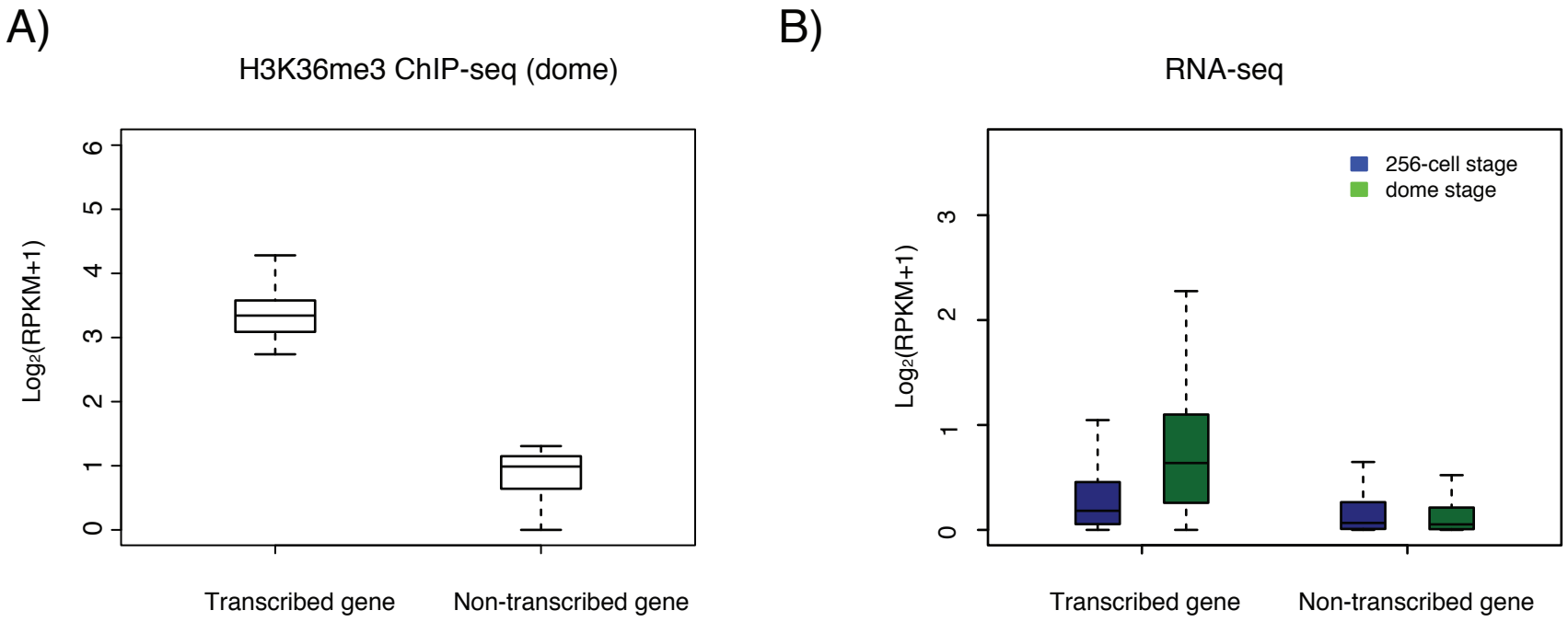

A)
H3K36me3 ChIP-seq (dome)
Log2(RPKM+1)
0
1
2
3
4
5
6
Transcribed gene
Non-transcribed gene
B)
RNA-seq
256-cell stage
dome stage
Log2(RPKM+1)
0
1
2
3
Transcribed gene
Non-transcribed gene

Figure S10

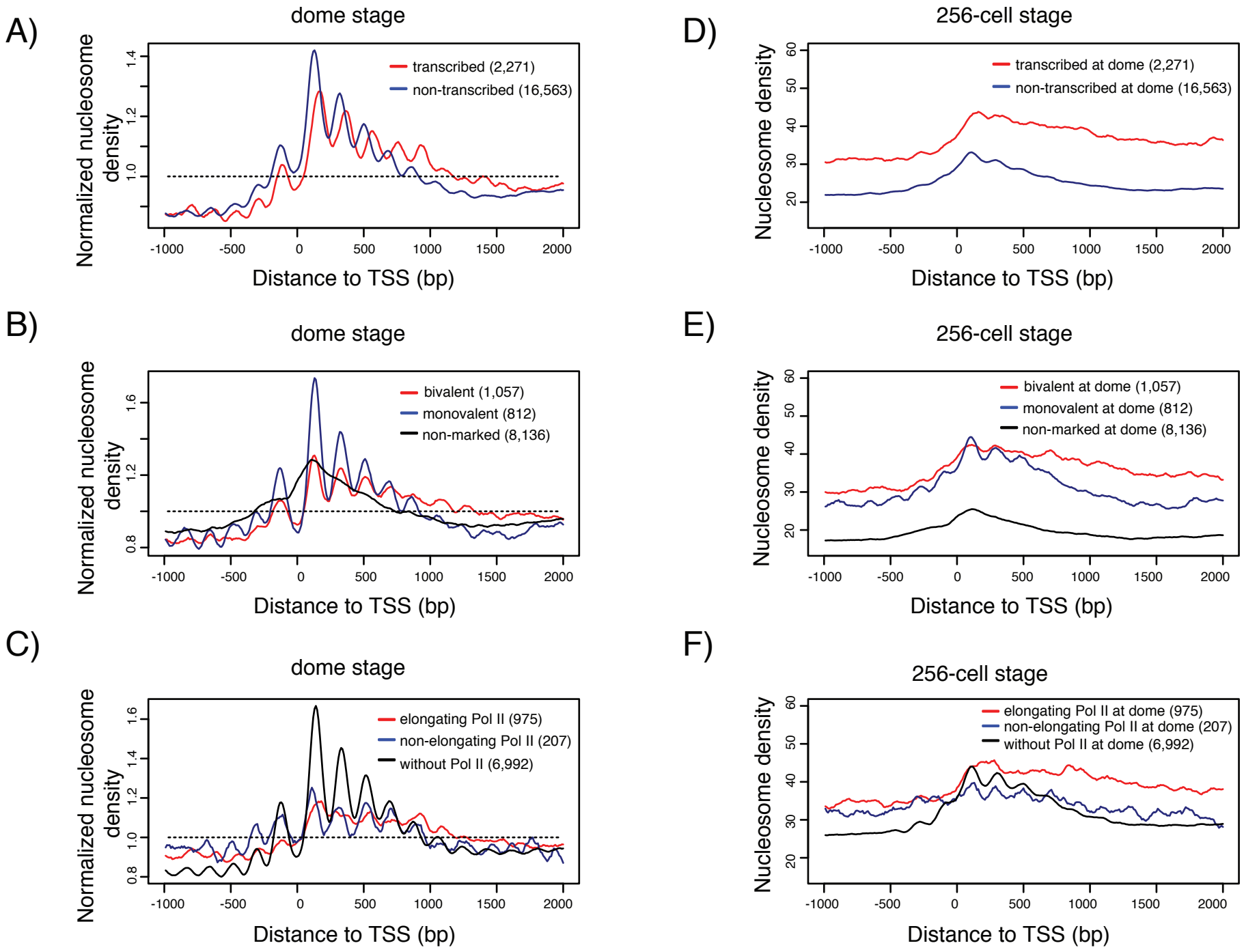

A)
dome stage
Normalized nucleosome density
transcribed (2,271)
non-transcribed (16,563)
Distance to TSS (bp)
B)
dome stage
Normalized nucleosome density
bivalent (1,057)
monovalent (812)
non-marked (8,136)
Distance to TSS (bp)
C)
dome stage
Normalized nucleosome density
elongating Pol II (975)
non-elongating Pol II (207)
without Pol II (6,992)
Distance to TSS (bp)
D)
256-cell stage
Nucleosome density
transcribed at dome (2,271)
non-transcribed at dome (16,563)
Distance to TSS (bp)
E)
256-cell stage
Nucleosome density
bivalent at dome (1,057)
monovalent at dome (812)
non-marked at dome (8,136)
Distance to TSS (bp)
F)
256-cell stage
Nucleosome density
elongating Pol II at dome (975)
non-elongating Pol II at dome (207)
without Pol II at dome (6,992)
Distance to TSS (bp)

Figure S11

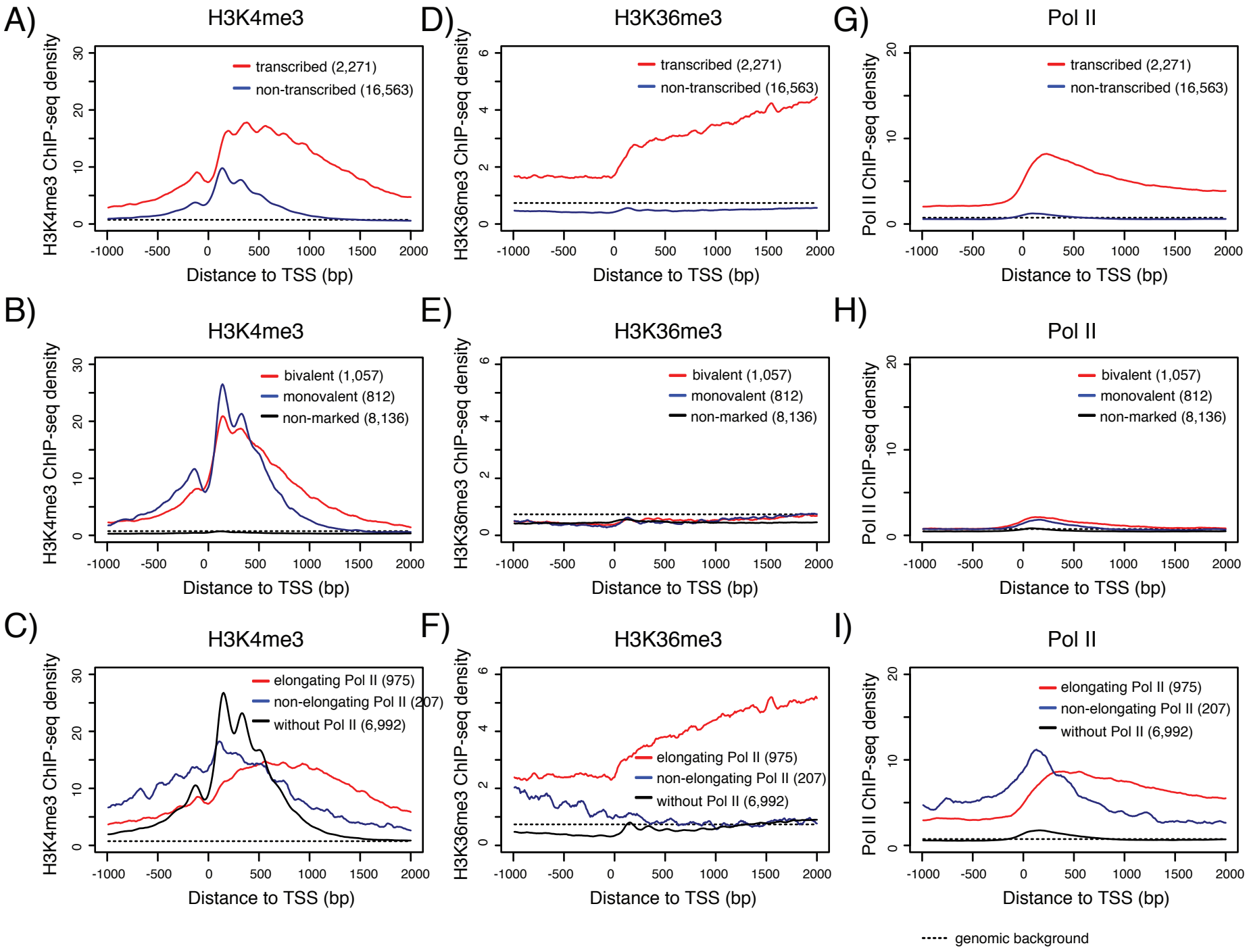
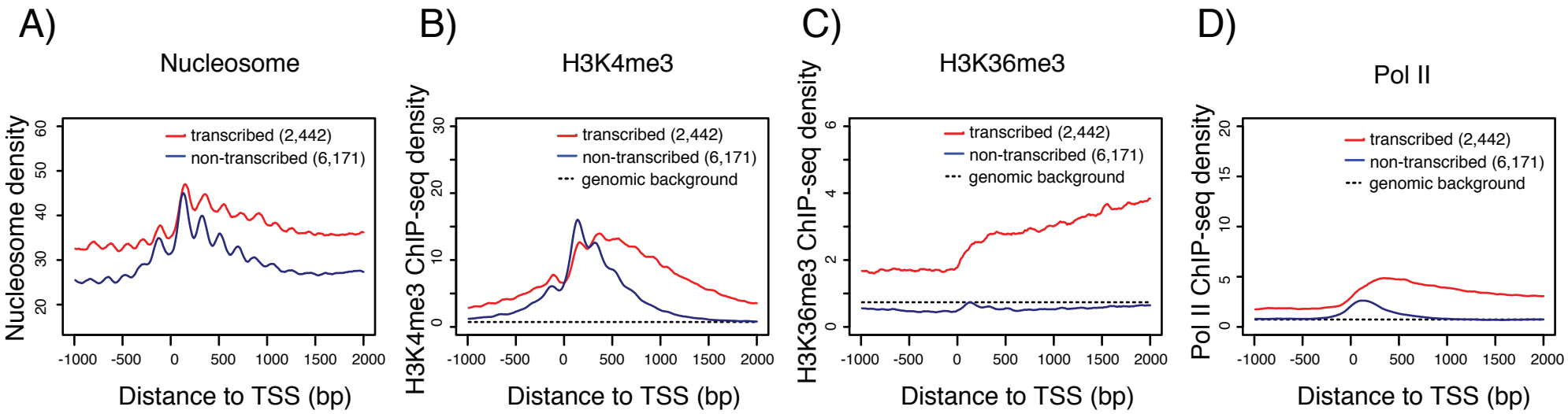
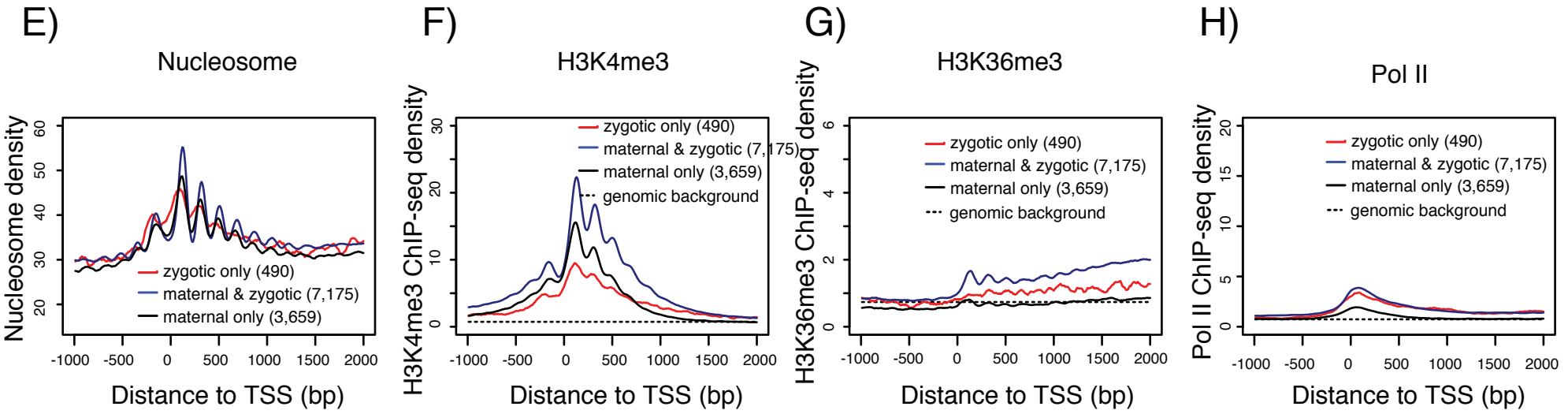

A)
H3K4me3
H3K4me3 ChIP-seq density
transcribed (2,271)
non-transcribed (16,563)
Distance to TSS (bp)
D)
H3K36me3
H3K36me3 ChIP-seq density
transcribed (2,271)
non-transcribed (16,563)
Distance to TSS (bp)
G)
Pol II
Pol II ChIP-seq density
transcribed (2,271)
non-transcribed (16,563)
Distance to TSS (bp)
B)
H3K4me3
H3K4me3 ChIP-seq density
bivalent (1,057)
monovalent (812)
non-marked (8,136)
Distance to TSS (bp)
E)
H3K36me3
H3K36me3 ChIP-seq density
bivalent (1,057)
monovalent (812)
non-marked (8,136)
Distance to TSS (bp)
H)
Pol II
Pol II ChIP-seq density
bivalent (1,057)
monovalent (812)
non-marked (8,136)
Distance to TSS (bp)
C)
H3K4me3
H3K4me3 ChIP-seq density
elongating Pol II (975)
non-elongating Pol II (207)
without Pol II (6,992)
Distance to TSS (bp)
F)
H3K36me3
H3K36me3 ChIP-seq density
elongating Pol II (975)
non-elongating Pol II (207)
without Pol II (6,992)
Distance to TSS (bp)
I)
Pol II
Pol II ChIP-seq density
elongating Pol II (975)
non-elongating Pol II (207)
without Pol II (6,992)
Distance to TSS (bp)
genomic background

Figure S12



Panel A-D: transcription status was defined based on the combination of H3K36me3 enrichment and Pol II traveling ratio.



Panel E-H: transcription status was defined based on parental origin of transcripts (Harvey *et al.*, 2013).

Figure S13

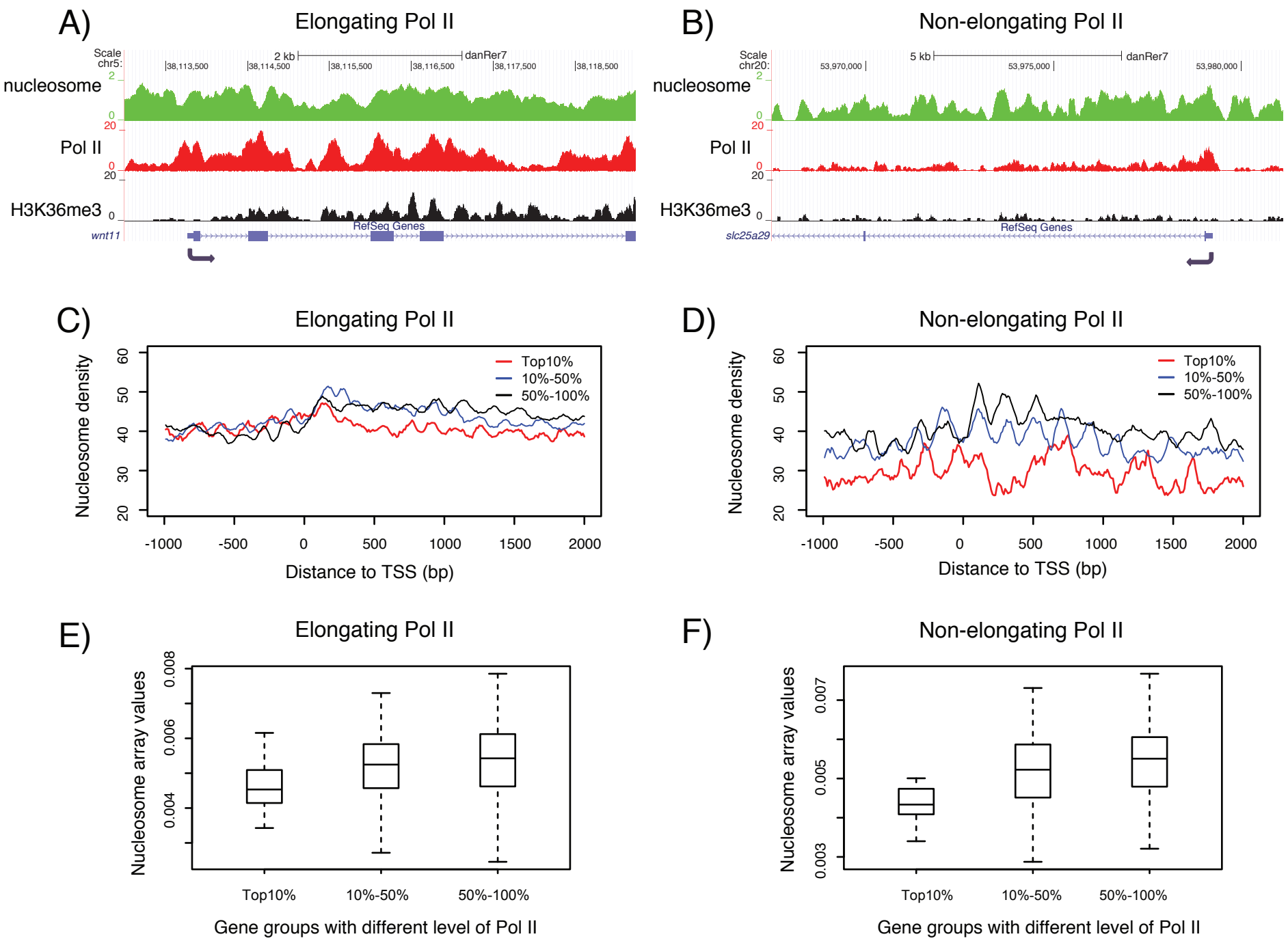

A)
Elongating Pol II
Scale
chr5:
2 kb
danRer7
38,113,500
38,114,500
38,115,500
38,116,500
38,117,500
38,118,500
nucleosome
Pol II
H3K36me3
RefSeq Genes
wnt11
B)
Non-elongating Pol II
Scale
chr20:
5 kb
danRer7
53,970,000
53,975,000
53,980,000
nucleosome
Pol II
H3K36me3
RefSeq Genes
slc25a29
C)
Elongating Pol II
Nucleosome density
Top10%
10%-50%
50%-100%
Distance to TSS (bp)
D)
Non-elongating Pol II
Nucleosome density
Top10%
10%-50%
50%-100%
Distance to TSS (bp)
E)
Elongating Pol II
Nucleosome array values
Top10%
10%-50%
50%-100%
Gene groups with different level of Pol II
F)
Non-elongating Pol II
Nucleosome array values
Top10%
10%-50%
50%-100%
Gene groups with different level of Pol II

Figure S14

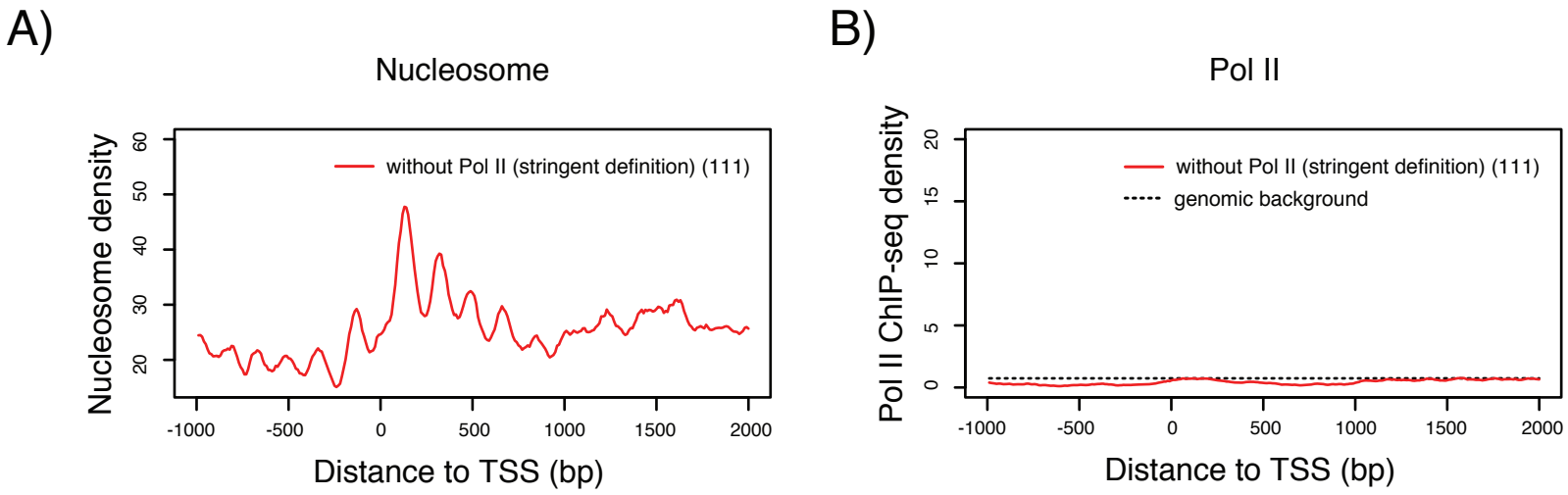
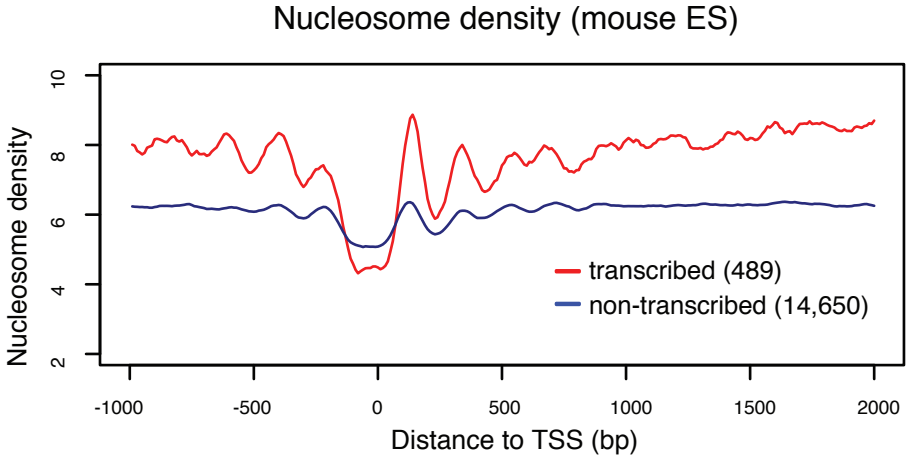
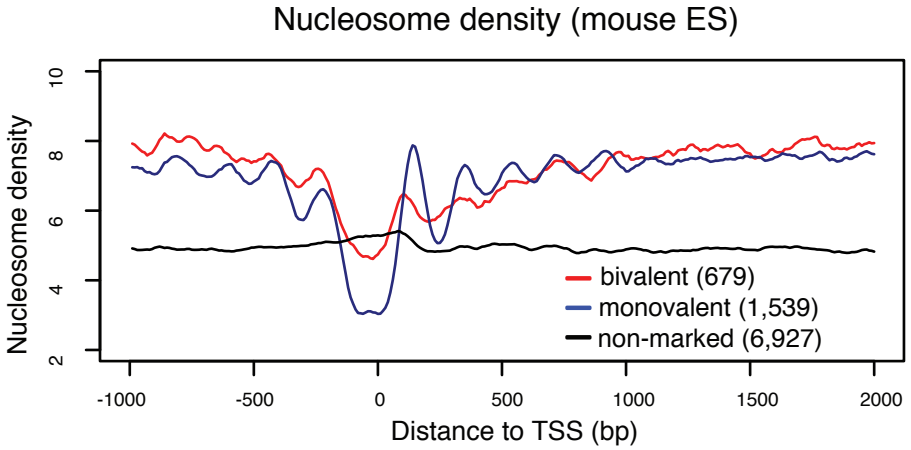

A)
Nucleosome
Nucleosome density
20
30
40
50
60
without Pol II (stringent definition) (111)
-1000
-500
0
500
1000
1500
2000
Distance to TSS (bp)
B)
Pol II
Pol II ChIP-seq density
0
5
10
15
20
without Pol II (stringent definition) (111)
genomic background
-1000
-500
0
500
1000
1500
2000
Distance to TSS (bp)

Figure S15

A)



B)



C)

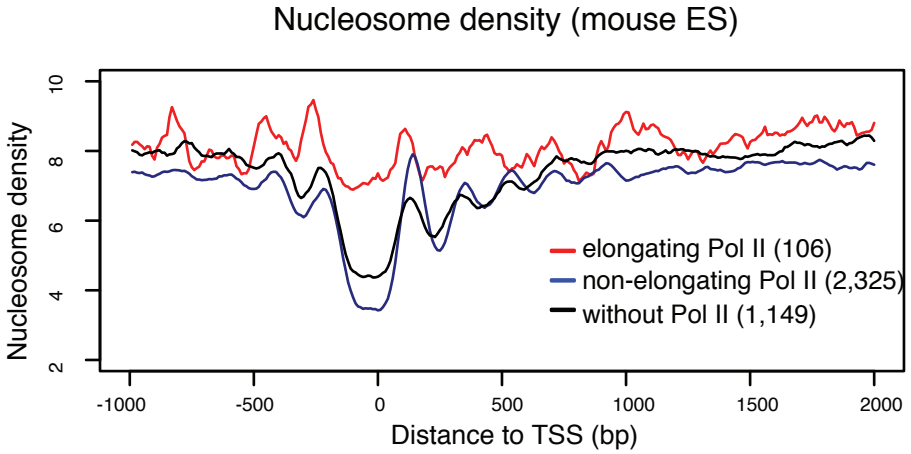
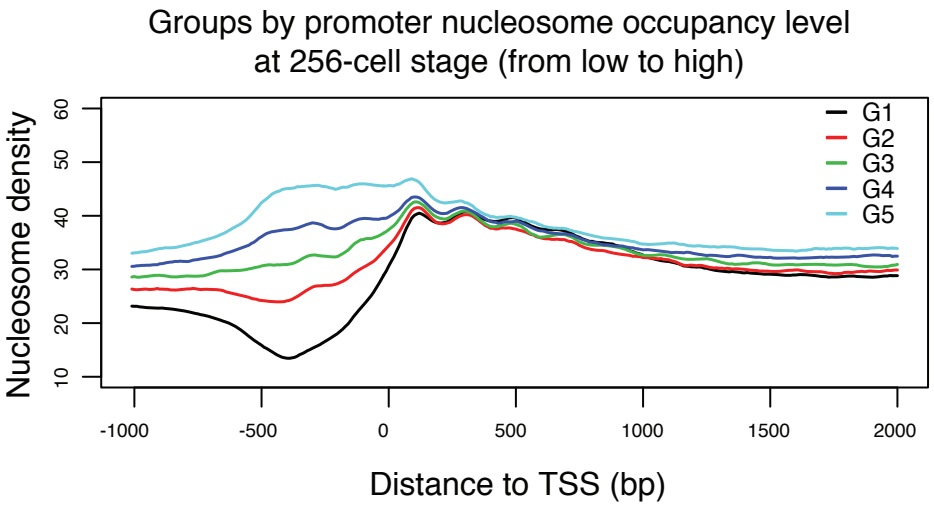


Figure S16

A)



B)

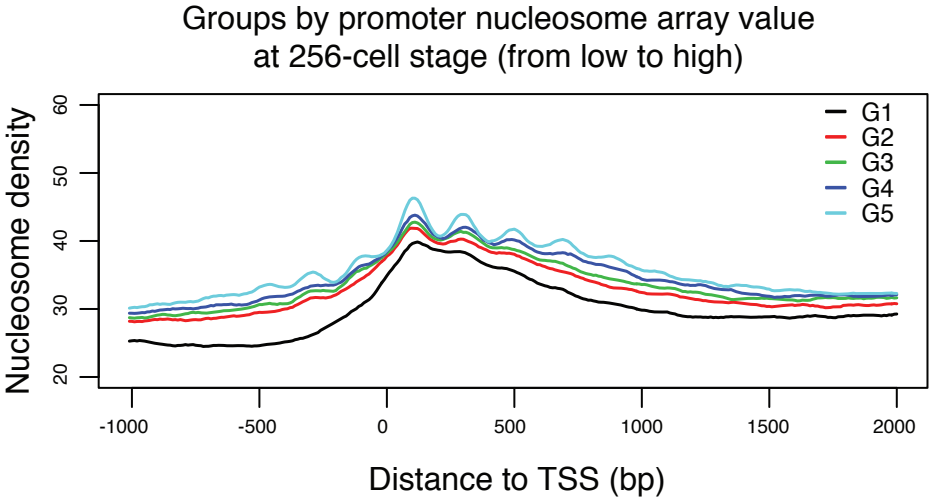


Figure S17

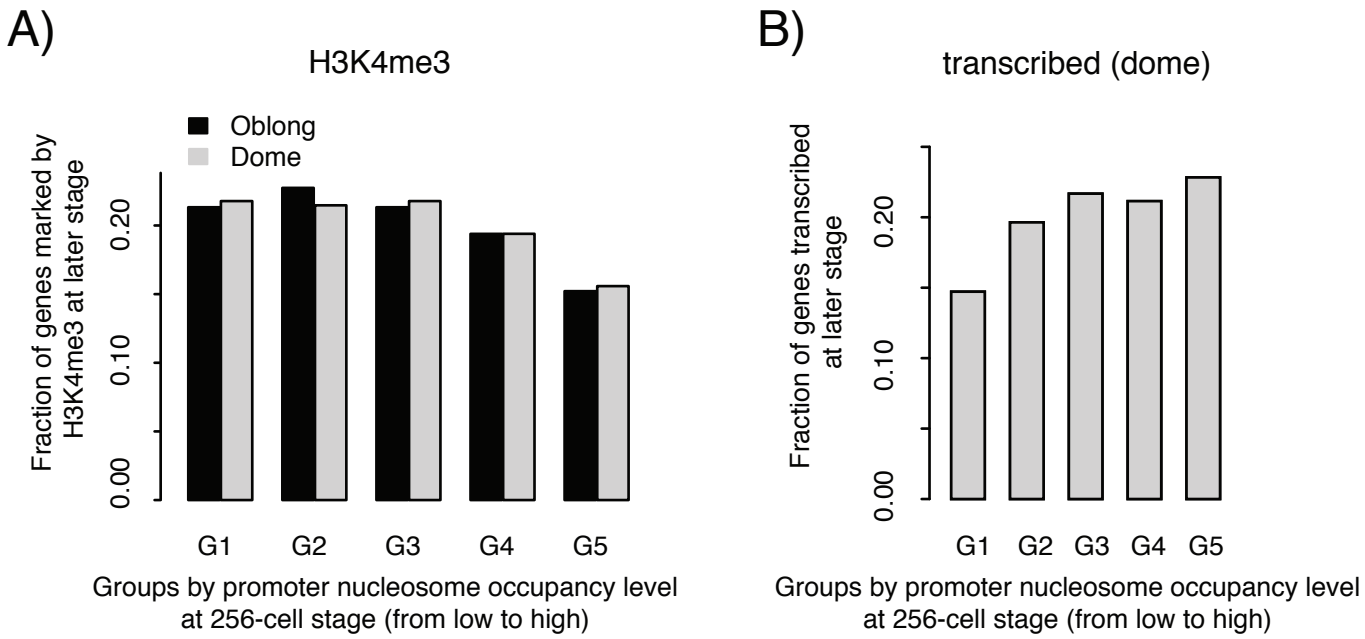
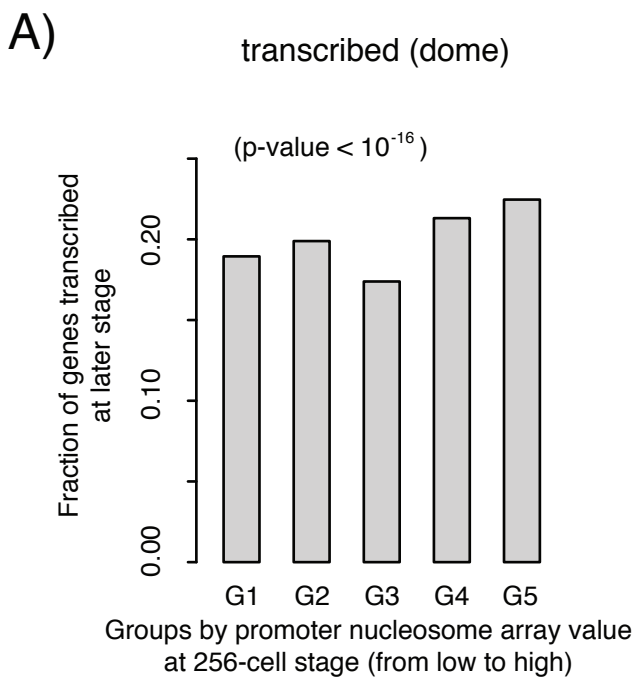
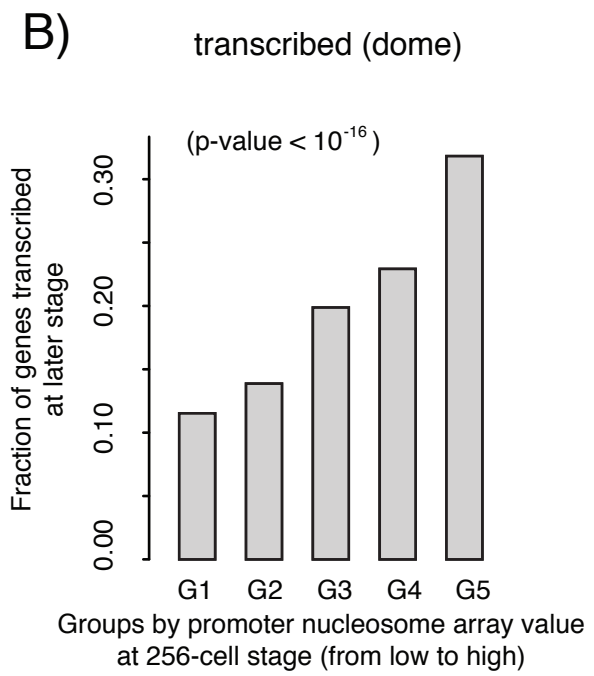

A)
H3K4me3
Oblong
Dome
Fraction of genes marked by H3K4me3 at later stage
0.00
0.10
0.20
G1
G2
G3
G4
G5
Groups by promoter nucleosome occupancy level at 256-cell stage (from low to high)
B)
transcribed (dome)
Fraction of genes transcribed at later stage
0.00
0.10
0.20
G1
G2
G3
G4
G5
Groups by promoter nucleosome occupancy level at 256-cell stage (from low to high)

Figure S18

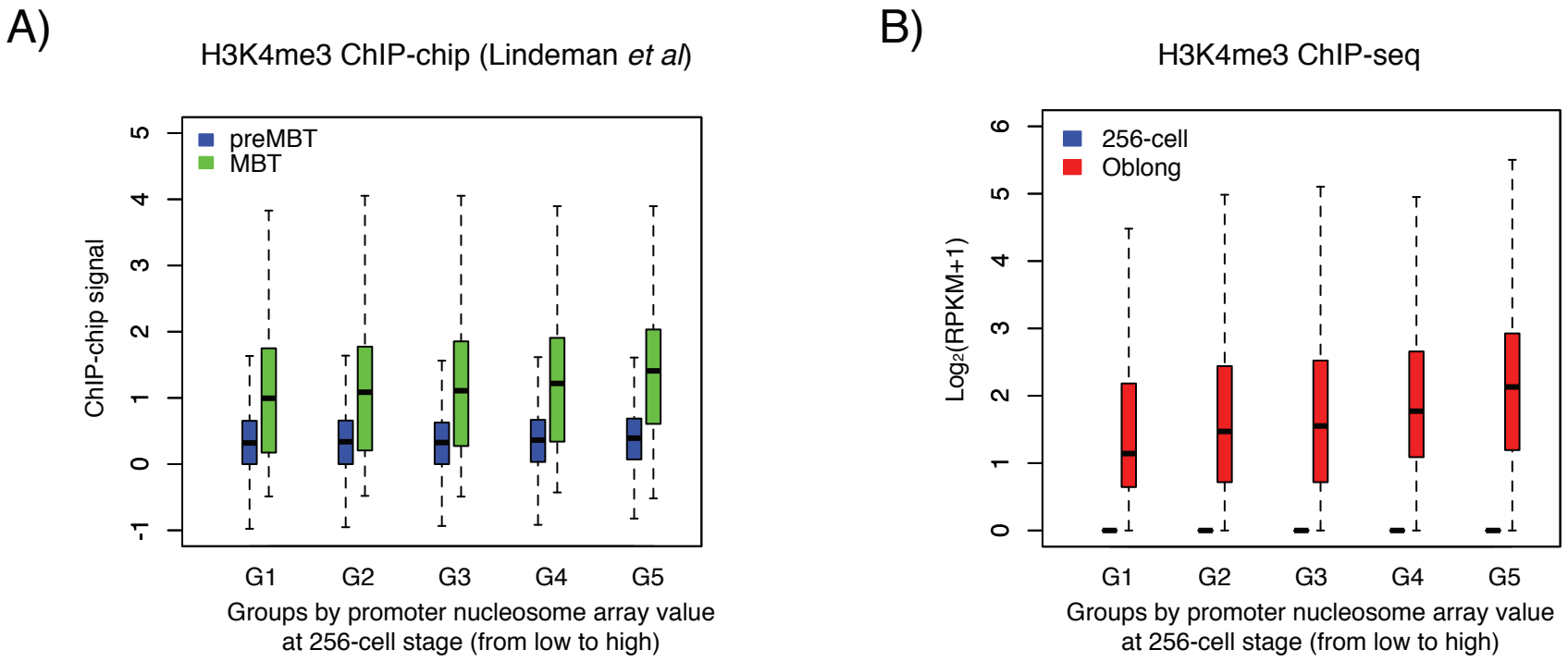


A) Transcription status was defined based on the combination of H3K36me3 enrichment and Pol II traveling ratio.



B) Transcription status was defined based on parental origin of transcripts (Harvey *et al.*, 2013).

Figure S19


A)
H3K4me3 ChIP-chip (Lindeman *et al*)
preMBT
MBT
ChIP-chip signal
-1
0
1
2
3
4
5
G1
G2
G3
G4
G5
Groups by promoter nucleosome array value
at 256-cell stage (from low to high)
B)
H3K4me3 ChIP-seq
256-cell
Oblong
Log2(RPKM+1)
0
1
2
3
4
5
6
G1
G2
G3
G4
G5
Groups by promoter nucleosome array value
at 256-cell stage (from low to high)