# Canonical nucleosome organization at promoters forms during genome activation

Yong Zhang, Nadine L Vastenhouw, Jianxing Feng, et al.

| | |
|---|---|
| **Supplemental Material** | http://genome.cshlp.org/content/suppl/2013/12/05/gr.157750.113.DC1.html |
| **P<P** | Published online November 27, 2013 in advance of the print journal. |
| **Accepted Preprint** | Peer-reviewed and accepted for publication but not copyedited or typeset; preprint is likely to differ from the final, published version. |
| **Open Access** | Freely available online through the *Genome Research* Open Access option. |
| **Creative Commons License** | This manuscript is Open Access.This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at http://creativecommons.org/licenses/by-nc/3.0/. |
| **Email Alerting Service** | Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.** |

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

# Canonical Nucleosome Organization at Promoters Forms During Genome Activation

Yong Zhang[1*$], Nadine L. Vastenhouw[2*$], Jianxing Feng[1#], Kai Fu[1#], Chenfei Wang[1],

Ying Ge[1], Paul van Hummelen[3], Alexander F. Schier[4$], and X. Shirley Liu[1,5$]

[1] School of Life Science and Technology, Tongji University, 1239 Siping Road, Shanghai 200092, China.

[2] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, D-01307 Dresden, Germany.

[3] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

[4] Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

[5] Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, Massachusetts 02115, USA.

[*] These authors contributed equally to this work.

[#] These authors contributed equally to this work.

[$] Correspondence should be addressed to X.S.L. (xsliu@jimmy.harvard.edu), Y.Z. (yzhang@tongji.edu.cn), N.L.V. (vastenhouw@mpi-cbg.de), or A.F.S (schier@fas.harvard.edu).

**Running title**: Nucleosome organization during genome activation

1

The organization of nucleosomes influences transcriptional activity by controlling accessibility of DNA binding proteins to the genome. Genome-wide nucleosome binding profiles have identified a canonical nucleosome organization at gene promoters, where arrays of well-positioned nucleosomes emanate from nucleosome-depleted regions. The mechanisms of formation and the function of canonical promoter nucleosome organization remain unclear. Here we analyze the genome-wide location of nucleosomes during zebrafish embryogenesis and show that well-positioned nucleosome arrays appear on thousands of promoters during the activation of the zygotic genome. The formation of canonical promoter nucleosome organization is independent of DNA sequence preference, transcriptional elongation, and robust RNA polymerase II (Pol II) binding. Instead, canonical promoter nucleosome organization correlates with the presence of Histone H3 Lysine 4 trimethylation (H3K4me3) and affects future transcriptional activation. These findings reveal that genome activation is central to the organization of nucleosome arrays during early embryogenesis.

2

# Introduction

Nucleosome organization affects transcriptional activity by influencing the access of DNA-binding proteins to the genome (Jiang and Pugh, 2009b). Recent studies in various model organisms have shown that nucleosome organization is not random in many functionally important regions of the genome (Albert et al., 2007; Jiang and Pugh, 2009b; Lee et al., 2007; Mavrich et al., 2008a; Mavrich et al., 2008b; Schones et al., 2008; Shivaswamy et al., 2008; Teif et al., 2012; Valouev et al., 2008; Valouev et al., 2011; Yuan et al., 2005). In particular, nucleosome-depleted regions have been observed near transcription start sites (TSS) of many genes. Moreover, arrays of positioned nucleosomes often emanate from nucleosome-depleted regions (Albert et al., 2007; Jiang and Pugh, 2009b; Valouev et al., 2008; Valouev et al., 2011). Studies in yeast have indicated that nucleosome patterns are influenced (Kaplan et al., 2009; Segal et al., 2006), but not determined (Stein et al., 2010; Weiner et al., 2010; Zhang et al., 2009), by DNA sequence. ATP-dependent chromatin remodelers (Gkikopoulos et al., 2011; Zhang et al., 2011) and transcription initiation and elongation (Hughes et al., 2012; Weiner et al., 2010; Zhang et al., 2009) have also been implicated in the organization of nucleosomes *in vivo*.

Although nucleosome organization profiles have been generated for vertebrate organisms (Gaffney et al., 2012; Hu et al., 2011; Li et al., 2011; Nekrasov et al., 2012; Teif et al., 2012; Valouev et al., 2011), it remains unclear how and when canonical promoter nucleosome structures form (Bai and Morozov, 2010). Here, we address these questions by analyzing changes in nucleosome organization during the maternal-zygotic transition in zebrafish embryos. In zebrafish, the bulk of zygotic transcription begins three hours

3

post fertilization (hpf), when a period of rapid synchronous cleavages ends and cell cycles lengthen (Kane and Kimmel, 1993; Schier, 2007; Tadros and Lipshitz, 2009). Specific histone modifications appear during this transition (Lindeman et al., 2011; Vastenhouw et al., 2010), resulting in a chromatin signature characteristic of embryonic pluripotency (Vastenhouw et al., 2010). Therefore, the maternal-zygotic transition provides an ideal *in vivo* system to study the relationship between transcriptional activity, specific histone methylation marks, and nucleosome organization *in vivo*.

## Results

**Emergence of well-positioned nucleosome arrays during genome activation**

To analyze nucleosome organization during genome activation, we combined micrococcal nuclease (MNase) digestion of chromatin with high-throughput sequencing (MNase-seq, Figure S1A-C). MNase preferentially cuts within linker DNA, so the sequenced fragments indicate the positions of nucleosomes when they are mapped back to the genome. We generated two biological replicates for two developmental stages, resulting in a total of 487 and 519 million mapped reads for 256-cell (2.5 hpf; before the maternal-zygotic transition) and dome/30% epiboly stage (4.5 hpf; after the maternal-zygotic transition) (Figure S1D, E), respectively. This represents over 50-fold genome coverage, one of the deepest nucleosome coverages of a vertebrate genome investigated to date.

We analyzed three major aspects of nucleosome organization: occupancy (nucleosome density), rotational positioning (orientation of the DNA helix on the nucleosome surface),

4

and translational positioning (relative location of nucleosomes along the DNA). While few changes were detected for nucleosome occupancy or rotational positioning during the maternal-zygotic transition (details in Figure S2, Supplementary text 1), changes in translational positioning of nucleosomes were widespread. By calculating positioning degree as previously defined (Zhang et al., 2009), we found that during genome activation arrays of well-positioned nucleosomes appear at many genomic loci (example shown in Figure 1A). Indeed, the number of arrays with five or more well-positioned nucleosomes increased over 10-fold from 256-cell stage (996 regions) to dome stage (10,523 regions). By analyzing the relative locations of all nucleosomes at both developmental stages, we found that nucleosomes are on average more regularly positioned at dome stage than at 256-cell stage (Figure 1B).

Nucleosome organization *in vivo* is known to be affected by DNA sequence preference (Kaplan et al., 2009; Segal et al., 2006; Struhl and Segal, 2013; Supplementary text 1). To examine the role of sequence preference on the appearance of nucleosome arrays, 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 (after genome activation). The observed and predicted nucleosome organization show very weak correlation at dome stage (Figure S3A, C), while for the same genomic regions, the sequence-based prediction performs slightly better at 256-cell stage (before genome activation, Figure S3B). This raises the possibility that the formation of nucleosome arrays during genome activation works against the DNA sequence preference of nucleosomes.

5

To investigate the potential functional relevance of well-positioned nucleosome arrays at dome stage, we examined their genomic distribution. Well-positioned nucleosome arrays are highly enriched at promoters (Figure 1C). Indeed, the average promoter nucleosome profile across all genes shows a clear nucleosome-depleted region and positioned nucleosome arrays emanating from the TSS (Figure 1D). This type of canonical promoter nucleosome organization has been reported in other organisms (Jiang and Pugh, 2009a; Radman-Livaja and Rando, 2010). Gene Ontology (GO) analysis revealed that several GO categories related to basic cellular functions are significantly overrepresented in the genes with well-positioned nucleosome arrays at their promoters (Figure S4A). Furthermore, genes with these arrays at their promoters are more conserved across eukaryotic species (Figure S4B) and enriched for housekeeping genes (Figure S4C). Analysis of the nucleosome arrays observed at 256-cell stage revealed a statistically significant but weaker enrichment at promoters (Figure 1C). The average promoter profile at the 256-cell stage also shows weaker canonical nucleosome organization (Figure 1D), and no GO categories were significantly enriched. Taken together, our results reveal that well-positioned nucleosome arrays emerge at promoters of housekeeping genes during genome activation.

**Co-occurrence of well-positioned nucleosome arrays with H3K4me3 marks**
Because well-positioned nucleosome arrays at dome stage are highly enriched at promoters, we investigated the potential relationship between well-positioned nucleosome arrays and histone marks associated with transcription. We performed

6

nucleosome-resolution ChIP-seq of H3K4me3, H3K27me3, H3K36me3, and ChIP-seq of RNA Pol II with two antibodies at dome stage (Table S1, Figure S5A, B), when these marks are well established (Lindeman et al., 2011; Vastenhouw et al., 2010). Interestingly, 91% of nucleosome arrays at promoters were marked by H3K4me3 (Figure 2A, B), and H3K4me3 enrichment was stronger in promoters with well-positioned nucleosome arrays than in those without (Figure S6A). These results suggest a potential quantitative relationship between H3K4me3 and nucleosome arrays at promoters. Indeed, when we ranked genes by their promoter H3K4me3 signals, those with better rank had stronger canonical promoter nucleosome organization (Figure 2C). To more accurately assess the quantitative relationship between the level of nucleosome arrays and different marks, we assigned a "nucleosome array value" to each promoter (for details see Online Methods). Overall, nucleosome array values at promoters showed stronger correlation with H3K4me3 enrichment at promoters (Figure 2D) than with either H3K27me3 enrichment at promoters (Figure S6B, E, S7A), Pol II signals at promoters and gene bodies (Figure S6C, F, G, S7C, D), or H3K36me3 enrichment at gene bodies (Figure S6D, H, S7B). To determine whether there is also a correlation of positioned nucleosome arrays with promoter H3K4me3 in mammalian pluripotent cells, we performed the same analyses in mouse embryonic stem cells and found similar relationships, but the differences among marks were moderate (Figure S8).

**Formation of canonical nucleosome organization is independent of transcriptional elongation**

7

Nucleosome arrays have been reported to result from transcriptional activity (Hughes et al., 2012; Weiner et al., 2010). We therefore analyzed whether well-positioned nucleosome arrays at promoters are specifically associated with genes that undergo transcriptional elongation. The early zebrafish embryo is loaded with maternal transcripts, precluding the identification of transcription status from standard RNA-seq or expression microarray data. We used three approaches to determine the transcription elongation status of genes. First, we used a combination of H3K36me3 ChIP-seq enrichment in exons and Pol II ChIP-seq enrichment in gene bodies to distinguish transcribed from non-transcribed genes at dome stage (details in Supplementary text 2, Figure S9). Surprisingly, transcribed and non-transcribed genes show similar canonical promoter nucleosome organization at dome stage using this approach (Figure 3A, S11A, D, G). In a second approach, we used a combination of H3K36me3 ChIP-seq enrichment in exons and Pol II traveling ratio (Rahl et al., 2010; Reppas et al., 2006) to distinguish transcribed from non-transcribed genes (details in Supplementary text 2). This approach also revealed clear canonical promoter nucleosome organization in non-transcribed genes at dome stage (Figure S12A-D). In a third approach, we used the classification of a recent study that defined transcription status by distinguishing between parental alleles (Harvey et al., 2013). Using this approach, we also found clear canonical promoter nucleosome organization in non-transcribed genes at dome stage (Figure S12E-H).

In a previous study, we classified three types of non-transcribed genes at dome stage based on histone marks in promoters: bivalent (H3K4me3/H3K27me3), monovalent (H3K4me3), and non-marked genes (neither H3K4me3 nor H3K27me3) (Vastenhouw et

8

al., 2010). We analyzed the nucleosome organization patterns for these classes of genes and found the canonical nucleosome organization to be present at the promoters of bivalent and monovalent genes, but absent from non-marked genes (Figure 3B). Together, these findings suggest that the relationship between the establishment of well-positioned nucleosome arrays and H3K4me3 during the maternal-zygotic transition is largely independent of transcriptional elongation.

**Promoter nucleosome organization in the absence of robust Pol II binding**

Pol II has been suggested to contribute to the formation of well-positioned nucleosome arrays in promoters (Hughes et al., 2012). To assess the effect of Pol II on nucleosome positioning, we classified Pol II at promoters into elongating and non-elongating based on the combination of promoter Pol II enrichment, gene body H3K36me3 enrichment and Pol II traveling ratio (details in Supplementary text 3, Figure S13A, B). Consistent with previous findings (Yin et al., 2011), the location of the +1 nucleosome is shifted downstream in promoters with elongating Pol II compared to non-elongating Pol II (Figure 3C, red curve versus blue curve). Moreover, genes with the strongest Pol II signal in their promoter have weaker nucleosome arrays (Figure S13C-F), which has been observed previously (Gilchrist et al., 2010) and is potentially a reflection of rapid nucleosome turnover.

Nucleosome array values at promoters are better correlated with H3K4me3 than with Pol II (Figure 2D, S6F) and, consistent with our previous study (Vastenhouw et al., 2010), many promoters with high H3K4me3 have undetectable levels of Pol II (Figure S5C).

9

This motivated us to examine nucleosome organization at promoters that are marked by H3K4me3 but lack Pol II signals. Interestingly, those promoters still formed well-positioned nucleosome arrays (Figure 3C, black curve). We examined the possibility that some genes without Pol II are bound by Pol II very weakly. As the average Pol II signal shows a tiny hump at the TSS of genes in this group (Figure S11I), we selected a subset of promoters in this group for which the Pol II signal is below the genomic average and analyzed their nucleosome profile. This more stringently selected sub-group of promoters still has well-positioned nucleosome arrays (Figure S14). Importantly, we observed a similar pattern of well-positioned nucleosome arrays at promoters in mouse embryonic stem cells that are marked by H3K4me3 but lack detectable Pol II binding (Figure S15C). Although very low, transient or undetectable Pol II binding might be involved in nucleosome positioning, our data reveals that robust binding by Pol II is not required for the establishment of well-positioned nucleosome arrays at promoters during the maternal-zygotic transition.

**Correlation of early promoter nucleosome organization with later transcription**

Nucleosome organization in promoters is proposed to facilitate recognition by the transcriptional machinery (Hughes et al., 2012; Yen et al., 2012). Since canonical nucleosome organization at promoters consists of nucleosome-depleted regions and arrays of positioned nucleosomes, we investigated whether these features have any predictive value for gene activation. When classifying genes into five groups based on their level of promoter nucleosome depletion at 256-cell stage (Figure S16A, details in supplementary methods), we did not observe a significant relationship between promoter

nucleosome depletion and H3K4me3 at promoters or gene transcription at later stages (Figure S17). In contrast, when genes are classified based on nucleosome array value at the 256-cell stage (details in Online Methods; Figure S16B), those with higher promoter nucleosome array values at the 256-cell stage have higher levels of promoter H3K4me3 (Figure 4A) and are more likely to be transcribed at dome stage (Figure 4B, Figure S18). To substantiate this finding, we analyzed an additional developmental stage by generating ChIP-seq for H3K4me3 at an earlier developmental stage (oblong; 3.5 hpf; Table S1). Indeed, better nucleosome array values at the 256-cell stage were associated with higher H3K4me3 marks at oblong (Figure 4A). Thus, the presence of nucleosome arrays at promoters before genome activation correlates with subsequent H3K4me3 acquisition and gene expression. To exclude the possibility that these associations are caused by histone modifications rather than nucleosome arrays at the 256-cell stage, we also generated ChIP-seq profiles for H3K4me3 at the 256-cell stage (Table S1). Consistent with our previous results (Vastenhouw et al., 2010) and corroborated by analyzing data from an independent study (Lindeman et al., 2011) (Figure S19A), signals of H3K4me3 were largely absent in promoters at the 256-cell stage and less correlated with later gene activation (Figure S19B). Taken together, the early presence of nucleosome arrays is a better predictor of future gene activity than nucleosome-depletion.

## Discussion

We have exploited the transition from transcriptional silence to widespread gene expression during early zebrafish embryogenesis to determine how nucleosomes are organized at promoters. Our study provides three major insights. First, most well-

11

positioned nucleosome arrays appear on promoters only during genome activation. The emergence of these well-positioned nucleosome arrays cannot be explained by DNA sequence preference. As nucleosome organization is likely to be determined by a combination of multiple factors, more studies will be needed to identify the DNA binding proteins and chromatin remodelers that are involved in the formation of nucleosome arrays during genome activation.

Second, canonical nucleosome organization can form independently of robust Pol II binding and active transcription, but correlates well with the presence of H3K4me3 marks. We favor a model wherein ATP-dependent activities are driving nucleosome array formation at 5' ends of genes, as described in studies of *in vitro* nucleosome positioning in yeast (Zhang et al., 2011). Our results indicate that H3K4me3 is a better predictor of nucleosome array formation at promoters than Pol II binding. Indeed, promoters marked by high H3K4me3 but no detectable Pol II binding can form well-positioned nucleosome arrays. Although it is impossible to exclude the possibility that very low, transient and undetectable levels of Pol II binding contribute to nucleosome positioning, our analysis of data in mouse embryonic stem cells also suggests that H3K4me3 may be a better predictor of nucleosome array formation than Pol II binding. We suggest that genome activation but not transcription is central to the organization of nucleosome arrays during early embryogenesis.

Third, promoter nucleosome organization before genome activation has predictive value for future gene activation. This observation is reminiscent of the proposed roles of

12

monovalent (H3K4me3) and bivalent (H3K4me3 and H3K27me3) domains in poising genes for activation (Bernstein et al., 2006; Vastenhouw and Schier, 2012). It is tempting to speculate that pioneer factors, chromatin remodelers, and histone modifiers establish nucleosome arrays and H3K4me3 marks at promoters before and during genome activation and prepare genes for subsequent transcriptional activation.

## Online methods

### Zebrafish

Zebrafish were maintained and raised under standard conditions. Wild-type embryos were collected at the 1-cell stage, synchronized, and allowed to develop to the desired stage at 28°C.

### MNase digestion

MNase digestion was performed essentially as described (He et al., 2010). Briefly, between 250 and 650 embryos were carefully staged, dechorionated, and fixed in 1% formaldehyde for 3 min at room temperature. Formaldehyde was quenched by adding glycine to a final concentration of 0.125 M. Embryos were rinsed 3 times in ice-cold PBS, immediately resuspended in cell lysis buffer (10 mM Tris-HCl pH7.5/10 mM NaCl/0.5% NP40), and lysed for 3 to 4 min on ice. Nuclei were collected by centrifugation, washed with ice-cold PBS, collected by centrifugation again, and resuspended in digestion buffer (50 mM Tris-HCl pH7.5/1 mM $CaCl_2$/0.2% Triton X-100). Samples were divided into 200 µL aliquots, pre-warmed to 25°C, and incubated with MNase [stored in MNase storage buffer (10 mM Hepes pH7.5/100 mM NaCl/1 mM $CaCl_2$/50% glycerol) and

13

diluted in MNase dilution buffer (50 mM Tris-HCl pH8/10 mM NaCl/126 mM $CaCl_2$/5% glycerol)]. Concentration of MNase and time of incubation were optimized to obtain 80% mononucleosomes as determined by BioAnalyzer analysis (Figure S1B).

If nucleosomes were directly sequenced, reactions were stopped by adding the samples to 200 μL MNase STOP and de-crosslinking buffer (1% SDS/100 mM $NaHCO_3$/25 mM EDTA) in a 15-mL tube. Samples were then sonicated and centrifuged for 10 min at 4°C, and the supernatant was incubated at 65°C for 90 min to reverse crosslinks. Upon precipitation, DNA was purified by QIAquick PCR purification kit (QIAGEN).

If nucleosomes were used for ChIP, reactions were stopped by adding samples to 200 μL MNase STOP buffer (10 mM Tris-HCl pH7.6/20 mM EDTA) in a 15-mL tube. Samples were then sonicated and dialyzed against ChIP buffer (27.8 mM Tris-HCl pH7.7/111 mM NaCl/4.1 mM EDTA/0.34% SDS/0.8% Triton X-100) at 4°C for 2 h. After centrifugation, the supernatant (chromatin) was used in ChIPs as described below.

**Antibodies**

Antibodies used were H3K4me3 (Millipore 07-473), H3K27Me3 (Millipore 07-449), H3K36me3 (Abcam ab9050), and RNA Pol II (8WG16/Covance MMS-126R and ab5408).

**ChIP**

14

RNA Pol II ChIPs were performed essentially as described (Vastenhouw et al., 2010) with some minor modifications. Briefly, between 400 and 800 embryos were staged, dechorionated, and fixed in 1.85% formaldehyde and 12.5% DMSO for 15 min at room temperature. Formaldehyde was quenched by adding glycine to a final concentration of 0.125 M. Embryos were rinsed 3 times in ice-cold PBS, immediately resuspended in cell lysis buffer (10 mM Tris-HCl pH7.5/10 mM NaCl/0.5% NP40), and lysed for 15 min on ice. Nuclei were collected by centrifugation, resuspended in nuclear lysis buffer (50 mM Tris-HCl pH7.5/10 mM EDTA/1% SDS), and lysed for 10 min on ice. Samples were diluted 3 times in IP dilution buffer (16.7 mM Tris-HCl pH7.5/167 mM NaCl/1.2 mM EDTA/0.01% SDS), and sonicated to obtain fragments of ~400 bp. Triton X-100 was added to a final concentration of 0.75%. Lysate was incubated overnight, with rotating, at 4°C with 25 μL of protein G magnetic Dynabeads (Invitrogen) that had been pre-bound to an excess amount of antibody. Bound complexes were extensively washed with RIPA (50 mM HEPES pH7.6/1 mM EDTA/0.7% DOC/1% Igepal/0.5 M LiCl) and TBS, then eluted from the beads with elution buffer (50 mM $NaHCO_3$/1% SDS). Cross-links were reversed overnight at 65°C, and DNA was purified by QIAquick PCR purification kit (QIAGEN). For MNase/ChIP, between 300 and 700 embryos were used per condition and, after dialysis, chromatin was added to the bead/antibody complex as described above.

**Library preparation and sequencing**

Libraries were prepared using the Illumina sequencing library preparation protocol and sequenced on an Illumina HiSeq 2000.

15

**Read mapping and nucleosome profiles**

All sequenced reads were mapped back to the zebrafish genome (zv9 assembly) using Bowtie. Uniquely mapped reads with a maximum of two mismatches were kept for further analysis. The formula used to calculate the genomic coverage of nucleosomal sequence reads: Number of reads $\times$ 147 bp / Genome size. To create nucleosome profiles, all mapped reads were extended to 147 bp in their 3' direction, and the middle 73 bp were piled up, as described previously (Zhang et al., 2008).

**Nucleosome array value and well-positioned nucleosome arrays**

The 73rd bp downstream the 5' end of a nucleosomal sequence read is termed the mid-point of nucleosomal DNA. To capture the features of well-positioned nucleosome arrays, we displayed the signal of nucleosome mid-points in a two-dimensional coordinate system with genomic positions on the X-axis and the number of nucleosome centers on the Y-axis. Signal processing was performed as follows. First, Gaussian smoothing was applied on the signals with a 147-bp sliding window and 30 bp as standard deviation. The difference of the signals on adjacent base pairs (i.e., the first-order derivative of the signal) was taken, then the absolute value of the difference on every base pair was calculated. If the value at a certain base pair was larger than any value within the [-73 bp, +73 bp] window, it was defined as a local maximum. All adjacent maxima were connected to create a curve spanning the whole genome. The height of the curve at each base pair is the nucleosome array value. The nucleosome array value for each promoter was then calculated by taking the average value of all base pairs in the promoter region.

16

Based on nucleosome array values, well-positioned nucleosome arrays were identified as follows. The average nucleosome array value of the whole genome was calculated and used as background, and the nucleosome array value at each base pair was transformed to a fold enrichment value over the background. This creates two types of segments: over background (fold enrichment larger than 1) and under background (fold enrichment larger than 1). Well-positioned nucleosome arrays are defined as over background segments with length larger than 1000 bp and average fold enrichment larger than 1.75.

**Histone modification and Pol II status of genes**

For this analysis, we used a set of 48,373 previously annotated genes in the zebrafish genome (Pauli et al., 2012). To ensure unambiguous classification between high and low statuses, genes with moderate enrichment for histone modifications and Pol II were excluded from average profiles analyses. For H3K4me3, if a gene contained more than 10 ChIP-seq reads in its promoter region (defined as region from 2 kb upstream of the TSS to 2 kb downstream of the TSS), and the average signal in its promoter was above 4 RPKM (Reads Per Kilo bases per Million reads), the H3K4me3 status of the gene was considered high. If the average signal in its promoter was less than 1 RPKM, the H3K4me3 status of the gene was considered low. The H3K4me3 status for all other genes was considered moderate. The H3K27me3 status was defined with the same criteria. For H3K36me3, if a gene contained more than 10 reads in the last 2/3 region of concatenated exons, and the average signal in the same region was above 4 RPKM, the H3K36me3 status of the gene was defined as high. If the number of reads in the average signal was less than 1 RPKM, the H3K36me3 status of the gene was low. The

17

H3K36me3 status for all other genes was defined as moderate. For Pol II in promoters, if a gene contained more than 10 ChIP-seq reads in its promoter region, and the average signal in its promoter was above 2.5 RPKM, the Pol II promoter status of the gene was considered high. If the average signal in its promoter was less than 1 RPKM, the Pol II promoter status of the gene was considered low. The Pol II promoter status for all other genes was considered moderate. For Pol II in gene bodies, if the average signal in gene body was above 2 RPKM, the Pol II gene body status of the gene was considered high. If the average signal in gene body was less than 1 RPKM, the Pol II gene body status of the gene was considered low. The Pol II gene body status for all other genes was considered moderate.

**Accession codes**

The DNA sequencing data of zebrafish from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE44269. The DNA sequencing data of mouse from this study have been submitted to NCBI GEO under accession number GSE51766.

# Acknowledgments

## Figure legends

**Figure 1**. Well-positioned nucleosome arrays are established at promoters during genome activation. A) Sample profile of nucleosome organization around the promoter of *supt6h* at 256-cell (blue, before the maternal-zygotic transition) and dome (green, after the maternal-zygotic transition). Gaussian smoothing of midpoints of nucleosomal DNA was used to create the nucleosome profile. B) Relationship between the position of adjacent nucleosomes before (blue) and after (green) the maternal-zygotic transition. The total number of nucleosomal sequencing reads with a start-to-start distance equal to the value on the X-axis is shown. C) Genomic distribution of well-positioned nucleosome arrays at 256-cell and dome stage. Promoters: regions from 2 kb upstream of the TSS to 2 kb downstream of the TSS. Gene bodies: regions from 2 kb downstream of the TSS to the transcription termination sites (TTS). Intergenic regions: all other genomic regions. D) Average promoter nucleosome organization across all genes before (blue) and after (green) the maternal-zygotic transition. All reads were extended to 147 bp; the middle 73 bp were taken, piled up, and normalized by sequencing coverage. The +1, +2, +3 and -1 nucleosomes around TSS at dome stage are indicated.

**Figure 2**. Well-positioned nucleosome arrays and H3K4me3 co-occur at promoters at dome stage. A) Sample profile of nucleosome organization (green) and H3K4me3 (red) around *mrpl16* and *mrpl39*. Gaussian smoothing of the mid-points of nucleosomal DNA was used to represent the nucleosome profile. Identified well-positioned nucleosome arrays are indicated by black bars. B) Venn diagram showing the overlap of well-

19

positioned nucleosome arrays and H3K4me3 peaks at promoters. 100 bp overlap was used as a minimal required cutoff. C) Heatmaps of nucleosome organization, H3K4me3, H3K27me3 and RNA Pol II signal around TSSs as well as H3K36me3 signal in last 2/3 regions of concatenated exons. Genes are ranked by H3K4me3 signal at their promoter. Each horizontal line represents the average signal for 100 genes. Color represents RPKM value. Genes were evenly grouped into 20 bins based on H3K4me3 density 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 the strongest H3K4me3 signal was kept. A total of 18,890 genes was used. D) The correlation between H3K4me3 density and nucleosome array value at promoters. H3K4me3 density is calculated as the $\log_2$ transformed RPKM + 1 for each promoter.

**Figure 3**. Canonical nucleosome organization at promoters is independent to transcriptional elongation. A) Average promoter nucleosome organization for transcribed (red) and non-transcribed (blue) genes. Transcription status is inferred from H3K36me3 ChIP-seq enrichment in exons and RNA Pol II ChIP-seq enrichment in gene bodies. 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), with H3K4me3 at promoters but without Pol II (black).

20

For A-C, 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 was indicated for each class. Short genes (< 1kb) were excluded. The accompanying nucleosome profiles for each class of genes in normalized version and at 256-cell stage are shown in Figure S10.

**Figure 4**. Nucleosome organization predicts H3K4me3 and 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 marked by H3K4me3 at promoters at oblong (black) and dome stage (grey) for each group. B) The fraction of genes transcribed at dome stage in each group. Transcription status is inferred from H3K36me3 ChIP-seq enrichment in exons and RNA Pol II ChIP-seq enrichment in gene bodies.

## 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.

Bai, L., and Morozov, A.V. (2010). Gene regulation by nucleosome positioning. Trends in genetics *26*, 476-483.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Gaffney, D., McVicker, G., Pai, A., Fondufe-Mittendorf, Y., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J. (2012). Controls of nucleosome positioning in the human genome. PLoS genetics *8*, e1003036.

Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540-551.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science *333*, 1758-1760.

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.

He, H.H., Meyer, C., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M.*, et al.* (2010). Nucleosome dynamics define transcriptional enhancers. Nature Genetics *42*, 343-347.

Hu, G., Schones, D.E., Cui, K., Ybarra, R., Northrup, D., Tang, Q., Gattinoni, L., Restifo, N.P., Huang, S., and Zhao, K. (2011). Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. Genome Research *21*, 1650-1658.

Hughes, A.L., Jin, Y., Rando, O., and Struhl, K. (2012). A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. Molecular Cell *48*, 5-15.

Jiang, C., and Pugh, B.F. (2009a). A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. Genome Biology *10*, R109.

Jiang, C., and Pugh, B.F. (2009b). Nucleosome positioning and gene regulation: advances through genomics. Nature Reviews Genetics *10*, 161-172.

Kane, D.A., and Kimmel, C.B. (1993). The zebrafish midblastula transition. Development *119*, 447-456.

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.

Lee, W., Tillo, D., Bray, N., Morse, R., Davis, R., Hughes, T., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nature Genetics *39*, 1235-1244.

Li, Z., Schug, J., Tuteja, G., White, P., and Kaestner, K.H. (2011). The nucleosome map of the mammalian liver. Nature Structural & Molecular Biology *18*, 742-746.

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.

Mavrich, T., Ioshikhes, I., Venters, B., Jiang, C., Tomsho, L., Qi, J., Schuster, S., Albert, I., and Pugh, B.F. (2008a). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Research *18*, 1073-1083.

23

Mavrich, T., Jiang, C., Ioshikhes, I., Li, X., Venters, B., Zanton, S., Tomsho, L., Qi, J.,

Glaser, R., Schuster, S.*, et al.* (2008b). Nucleosome organization in the Drosophila

genome. Nature *453*, 358-362.

Nekrasov, M., Amrichova, J., Parker, B.J., Soboleva, T.A., Jack, C., Williams, R., Huttley,

G.A., and Tremethick, D.J. (2012). Histone H2A.Z inheritance during the cell cycle

and its impact on promoter organization and dynamics. Nature Structural &

Molecular Biology *19*, 1076-1083.

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.

Radman-Livaja, M., and Rando, O. (2010). Nucleosome positioning: how is it

established, and why does it matter? Developmental Biology *339*, 258-266.

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.

Schier, A.F. (2007). The maternal-zygotic transition: death and birth of RNAs.

Science *316*, 406-407.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K.

(2008). Dynamic regulation of nucleosome positioning in the human genome. Cell

*132*, 887-898.

24

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.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biology *6*, e65.

Stein, A., Takasuka, T.E., and Collings, C.K. (2010). Are nucleosome positions in vivo primarily determined by histone-DNA sequence preferences? Nucleic acids research *38*, 709-719.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nature Structural & Molecular Biology *20*, 267-273.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033-3042.

Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J., Marth, C., Hofer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nature Structural & Molecular Biology *19*, 1185-1192.

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.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature *474*, 516-520.

Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. Curr Opin Cell Biol *24*, 374-386.

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.

Weiner, A., Hughes, A., Yassour, M., Rando, O., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Research *20*, 90-100.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide Nucleosome Specificity and Directionality of Chromatin Remodelers. Cell *149*, 1461-1473.

Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2011). A high-resolution whole-genome map of key chromatin modifications in the adult Drosophila melanogaster. PLoS genetics *7*, e1002380.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626-630.

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.

Zhang, Y., Shin, H., Song, J.S., Lei, Y., and Liu, X.S. (2008). Identifying Positioned Nucleosomes with Epigenetic Marks in Human from ChIP-Seq. BMC Genomics *9*, 537.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science *332*, 977-980.
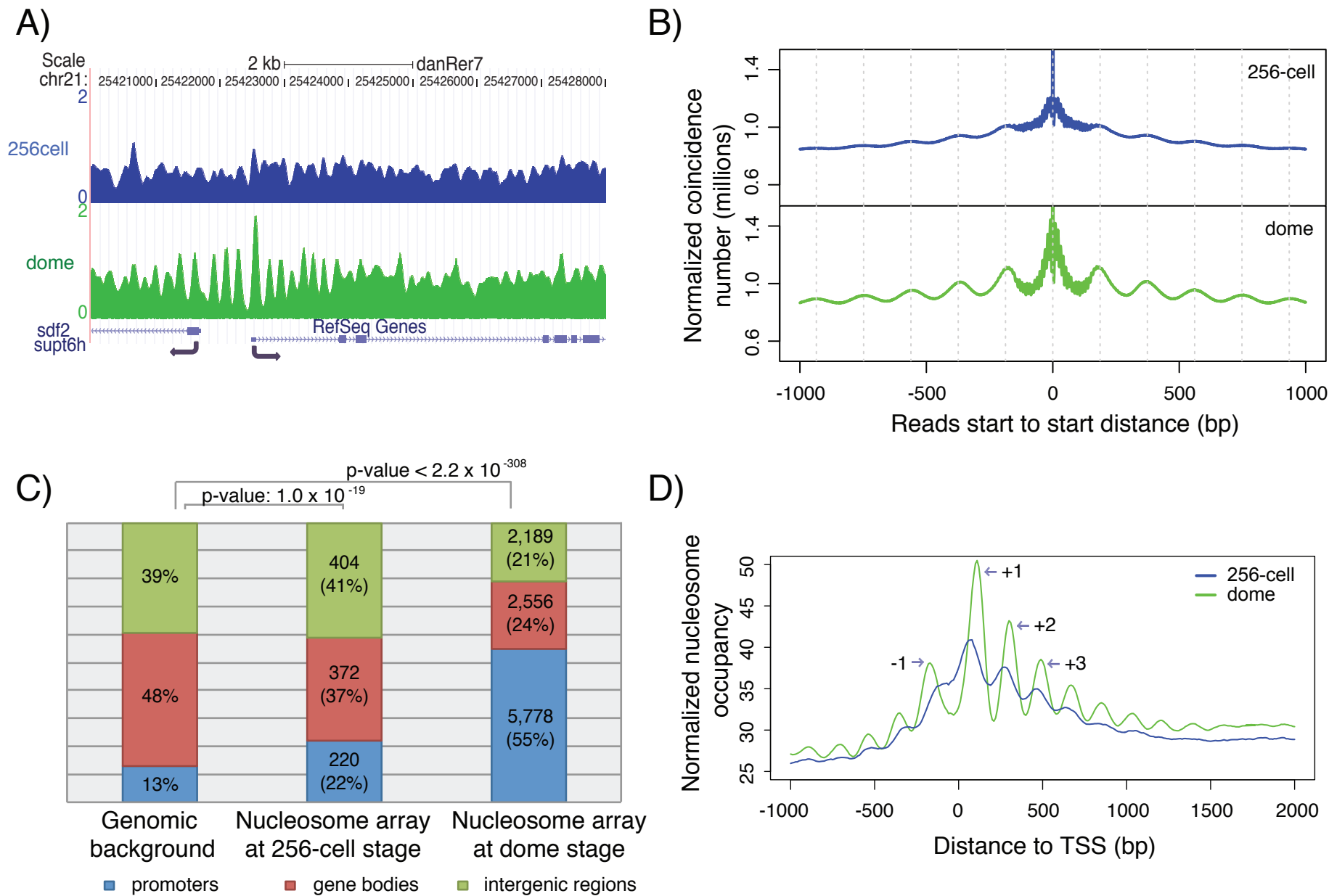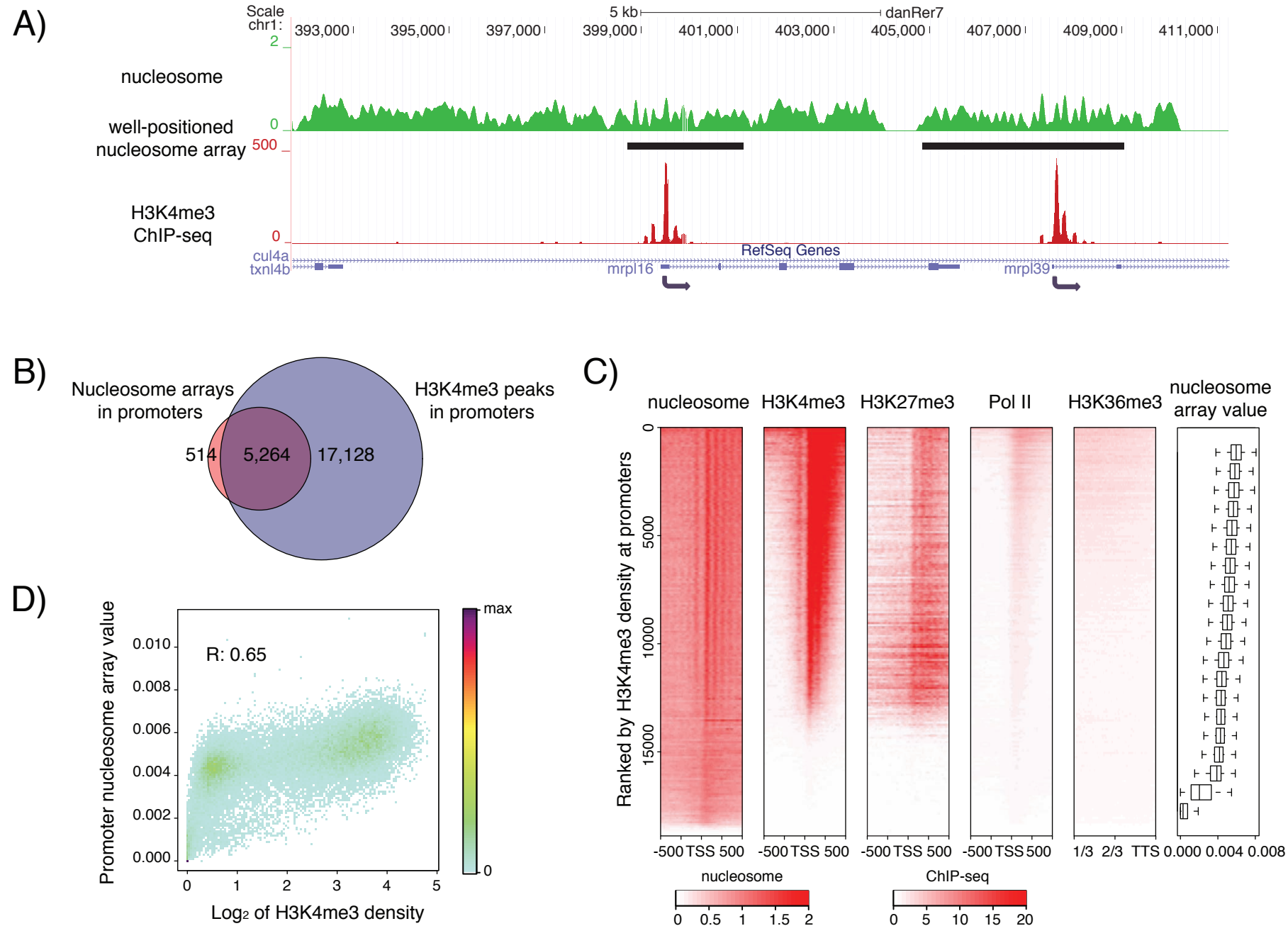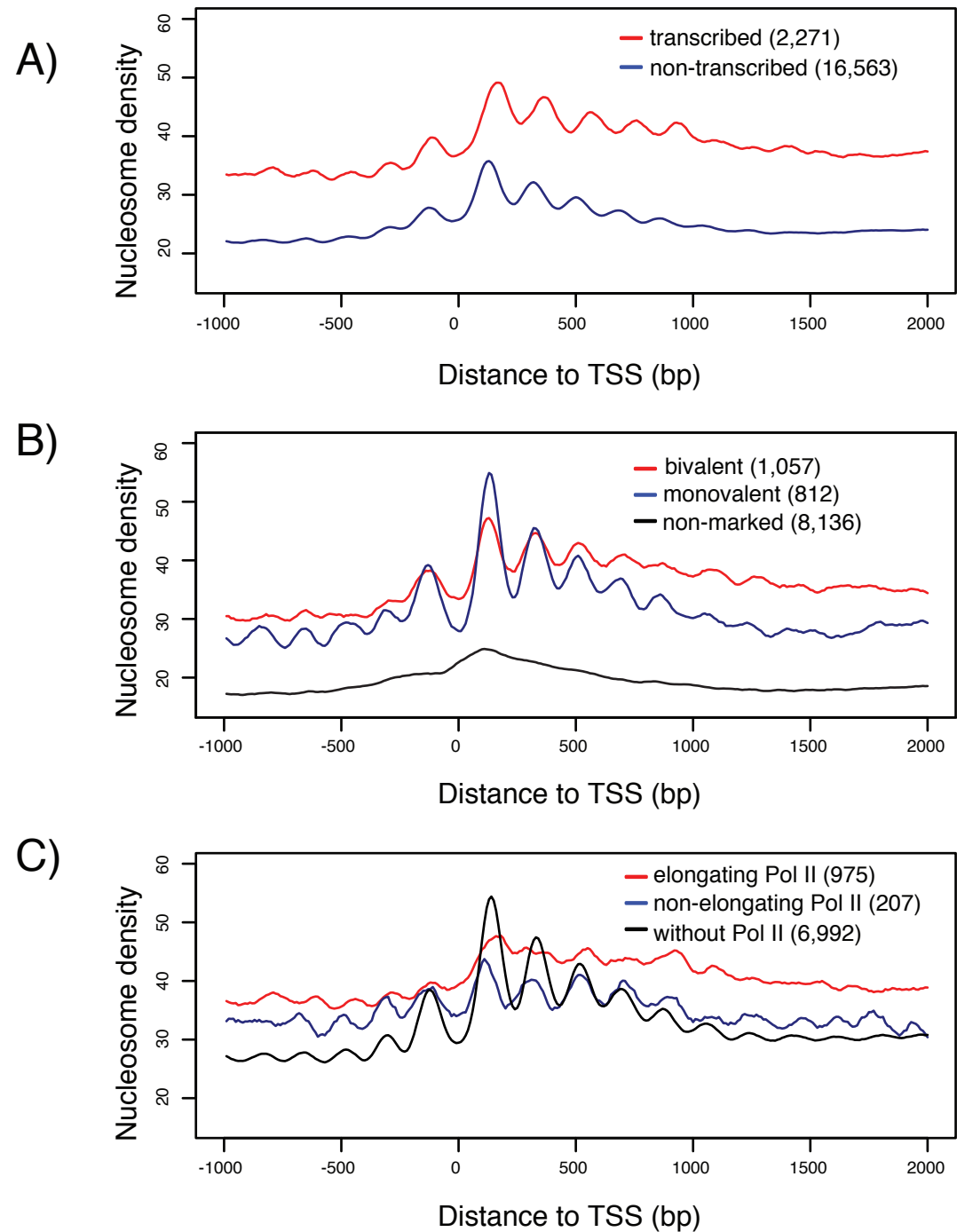
# Figure 1

# Figure 2

Figure 3

# Figure 4

A)

**H3K4me3**



Fraction of genes marked by H3K4me3 at later stage

■ Oblong (p-value < $10^{-16}$ )
□ Dome   (p-value < $10^{-16}$ )

Groups by promoter nucleosome array value
at 256-cell stage (from low to high)

B)

**transcribed (dome)**



Fraction of genes transcribed at later stage

(p-value < $10^{-16}$ )

Groups by promoter nucleosome array value
at 256-cell stage (from low to high)