

Transcriptional landscape of the human cell cycle

Yin Liu^{a,b,c,1}, Sujun Chen^{a,b,d,e,1}, Su Wang^{b,c,f}, Fraser Soares^d, Martin Fischer^g, Feilong Meng^h, Zhou Du^{b,c,i}, Charles Lin^j, Clifford Meyer^{c,k}, James A. DeCaprio^g, Myles Brown^{c,g,2}, X. Shirley Liu^{c,k,2}, and Housheng Hansen He^{d,e,2}

^aClinical Translational Research Center, Shanghai Pulmonary Hospital, Tongji University, Shanghai 200433, China; ^bDepartment of Bioinformatics, School of Life Sciences, Tongji University, Shanghai 200092, China; ^cCenter for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215; ^dPrincess Margaret Cancer Center, University Health Network, Toronto, ON M5G1L7, Canada; ^eDepartment of Medical Biophysics, University of Toronto, Toronto, ON M5G1L7, Canada; ^fDepartment of Biomedical Informatics, Harvard Medical School, Boston, MA 02215; ^gDepartment of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215; ^hState Key Laboratory of Molecular Biology, Chinese Academy of Sciences Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China; ⁱProgram in Cellular and Molecular Medicine, Boston Children's Hospital, Howard Hughes Medical Institute, Boston, MA 02115; ^jDepartment of Molecular and Human Genetics, Baylor College of Medicine Houston, TX 77030; and ^kDepartment of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, MA 02215

Contributed by Myles Brown, January 18, 2017 (sent for review October 25, 2016; reviewed by Karen Knudsen and John T. Lis)

Steady-state gene expression across the cell cycle has been studied extensively. However, transcriptional gene regulation and the dynamics of histone modification at different cell-cycle stages are largely unknown. By applying a combination of global nuclear run-on sequencing (GRO-seq), RNA sequencing (RNA-seq), and histone-modification ChIP sequencing (ChIP-seq), we depicted a comprehensive transcriptional landscape at the G0/G1, G1/S, and M phases of breast cancer MCF-7 cells. Importantly, GRO-seq and RNA-seq analysis identified different cell-cycle-regulated genes, suggesting a lag between transcription and steady-state expression during the cell cycle. Interestingly, we identified genes actively transcribed at early M phase that are longer in length and have low expression and are accompanied by a global increase in active histone 3 lysine 4 methylation (H3K4me2) and histone 3 lysine 27 acetylation (H3K27ac) modifications. In addition, we identified 2,440 cell-cycle-regulated enhancer RNAs (eRNAs) that are strongly associated with differential active transcription but not with stable expression levels across the cell cycle. Motif analysis of dynamic eRNAs predicted Kruppel-like factor 4 (KLF4) as a key regulator of G1/S transition, and this identification was validated experimentally. Taken together, our combined analysis characterized the transcriptional and histone-modification profile of the human cell cycle and identified dynamic transcriptional signatures across the cell cycle.

GRO-seq | nascent RNA | transcriptional regulation | epigenetics | cell cycle

The process of cell division is vital to the growth and development of an organism as a single fertilized cell develops into a mature organism and organs undergo cellular renewal or repair (1–3). Tight control of molecular events during the cell cycle guarantees fidelity in preserving genetic information and the prevention of unwarranted cell division. The progression of the cell cycle involves well-orchestrated transcriptional and epigenetic controls (4–7). Dysregulation of this process can lead to various diseases, including cancer (8–10). Because the proportion of actively dividing cells is considerably higher in cancers than in normal tissues, targeting the cell cycle is an attractive therapeutic option for cancer treatment (11, 12).

Many proteins that carry out important functions during the cell cycle display a cyclic expression pattern that is often regulated on the transcriptional level (13, 14). Because it has been shown that cell-cycle gene expression serves as a tumor signature (15), extensive efforts have been devoted to identify periodically expressed genes across the cell cycle using microarray platforms (4, 16–20). In addition, identification of the cell-cycle-regulated genes and follow-up mechanistic studies of individual genes have greatly advanced our understanding of cell-cycle progression and the development of diseases. For instance, the transcriptional regulation of cell-cycle genes was found to be governed by the RB-E2F, DREAM, and MMB-FOXO1 complexes (21–23). However, analyses using microarray or RNA sequencing (RNA-seq) can identify only the accumulated, steady-state gene-expression level;

the temporal regulation of transcriptional dynamics during the cell cycle remains elusive.

The nuclear run-on assay has long been used to investigate RNA polymerase activity and nascent RNA transcription for individual genes (24). The global nuclear run-on followed by RNA sequencing (GRO-seq) assay enables the investigation of temporal transcription at a genome-wide scale (25–28). In addition, GRO-seq also can be used to analyze transcription from active regulatory sequences such as enhancers (29). Recent studies have found that the loading of RNA polymerase II (Pol II) at enhancer regions can lead to widespread active transcription and production of enhancer RNAs (eRNAs) (30, 31). Since then, eRNAs have been demonstrated to play important roles in regulating enhancer-promoter interactions and target gene transcription, rather than merely being transcriptional noises from enhancers (32–35). GRO-seq has been applied widely to study transcriptional regulation in a variety of biological systems but has not been explored in the context of the cell cycle (26, 29, 36).

In this study, we used a combination of GRO-seq, RNA-seq, and histone-modification ChIP sequencing (ChIP-seq) to investigate transcriptional and epigenetic dynamics across the cell cycle. The multilevel data of nascent transcription, steady-state expression level, and chromatin status provide insights not only into genes periodically regulated during the cell cycle but also into the underlying regulatory mechanisms. Our data depict a comprehensive transcriptional and epigenetic landscape of the human cell cycle and will be a valuable resource for cell-cycle studies.

Significance

Our study provided a comprehensive view of the transcriptional landscape across the cell cycle. We revealed lag between transcription and steady-state RNA expression at the cell-cycle level and characterized a large amount of active transcription during early mitosis. In addition, our analysis identified thousands of enhancer RNAs and related transcription factors that are highly correlated with cell-cycle-regulated transcription but not with steady-state expression, thus highlighting the importance of transcriptional and epigenetic dynamics during cell-cycle progression.

Author contributions: Y.L., S.C., M.B., X.S.L., and H.H.H. designed research; Y.L., C.L., and H.H.H. performed research; Y.L., S.C., S.W., M.F., Z.D., C.M., and J.A.D. analyzed data; and Y.L., S.C., F.S., M.F., F.M., M.B., X.S.L., and H.H.H. wrote the paper.

Reviewers: K.K., Thomas Jefferson University; and J.T.L., Cornell University.

The authors declare no conflict of interest.

Data deposition: The MCF-7 GRO-seq, RNA-seq, ChIP-seq, and DNase-seq raw sequence tags and processed bed files reported in this paper have been submitted to the National Center for Biotechnology Gene Expression Omnibus (GEO) database (accession no. GSE94479).

¹Y.L. and S.C. contributed equally to this work.

²To whom correspondence may be addressed. Email: myles_brown@dfci.harvard.edu, xshliu@jimmy.harvard.edu, or hansenhe@uhnresearch.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617636114/-DCSupplemental.

Results

Mapping the Transcriptional and Epigenetic Landscape Across the Cell Cycle. To investigate dynamic transcriptional and epigenetic gene regulation across the cell cycle, we performed GRO-seq (25), RNA-seq, and ChIP-seq of histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 4 methylation (H3K4me2), two histone modifications that mark promoters and enhancers, in the synchronized human breast cancer cell line MCF-7 (Fig. 1A). The cells were synchronized to G0/G1 with hormone starvation, to G1/S with double thymidine treatment, and to early M phase with thymidine-nocodazole treatment (*Materials and Methods*). The degree of synchrony at each cell-cycle stage was monitored by flow cytometry analysis of DNA content using propidium iodide staining (*SI Appendix, Fig. S1*) (37). Nuclei were isolated from two biological replicates of the synchronized cells and were subjected to the GRO-seq procedure (*Materials and Methods*) for nascent RNA analysis. Total RNA from the same batch of synchronized cells was subjected to RNA-seq analysis to investigate steady-state expression levels. Chromatin profiling was conducted using replicated H3K27ac and H3K4me2 ChIP-seq at all three cell-cycle stages.

Approximately 30–40 million reads were uniquely mapped to the human genome for each GRO-seq sample, and the biological replicates for each cell-cycle stage were highly correlated (*SI Appendix, Table S1*). Most reads (70%) align on the coding strand within the boundaries of annotated National Center for Biotechnology Information (NCBI)

Reference Sequence (RefSeq) genes, and the typical bidirectional transcription at the transcriptional start site (TSS) was observed (*SI Appendix, Fig. S2* and *Table S1*) (37). Thirty million reads were uniquely mapped for each RNA-seq sample, more than 50% of which were mapped to the annotated RefSeq gene exons and UTRs (*SI Appendix, Table S2*). Fifteen to twenty million reads were uniquely mapped for each ChIP-seq sample, and the correlation between biological replicates was more than 0.96 (*SI Appendix, Table S3*).

GRO-Seq and RNA-Seq Identify Different Cell-Cycle-Regulated Genes.

To investigate the correlation between nascent transcription and steady-state expression levels, we compared the GRO-seq and RNA-seq read counts for all the RefSeq annotated genes (*Materials and Methods*). Although the correlations between GRO-seq samples at different cell-cycle stages ranged from 0.89–0.99, and those of the RNA-seq samples ranged from 0.78–0.90, the correlations between GRO-seq and RNA-seq data were below 0.5 (*SI Appendix, Fig. S3*), indicating significant difference between newly synthesized and accumulated RNA levels. As an example, we plotted the histogram of GRO-seq and RNA-seq read counts for Centromere-associated protein E precursor (*CENPE*) (Fig. 1B), a centrosome-associated mitotic kinesin (38). Although the RNA-seq signal for *CENPE* is three- to fourfold higher at M phase than at G0/G1 and G1/S, the GRO-seq signal plateaus at G1/S (Fig. 1B), suggesting a lag of RNA-seq in reflecting transcription at cell-cycle stages. This lagging effect was observed for most of the mitotic genes we curated from publicly available datasets profiling cell-cycle expression (Fig. 1C) (16–20).

To assess the dynamics of transcriptional regulation at different cell-cycle stages, we identified genes differentially transcribed across the cell cycle. GRO-seq read counts for RefSeq genes at different cell-cycle stages were normalized using spiked-ins as described previously (39), and the normalized read counts were subjected to DESeq2 analysis (*SI Appendix, Fig. S4*) (40). Gene ontology (GO) analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (41) was conducted on genes with significant [false-discovery rate (FDR) <0.01] differential transcription among cell-cycle stages (Fig. 1D, *SI Appendix, Fig. S5*, and *Dataset S1*). Genes with higher transcription in G1/S than in G0/G1 and M phases were enriched for the GO term “M phase,” suggesting that mitotic genes are actively transcribed in G1/S and accumulated at M phase. Importantly, most of the mitotic genes we curated (*Materials and Methods*) overlapped with genes in the M phase GO term and had the highest steady-state expression at M phase (Fig. 1C). In addition, genes highly transcribed at G0/G1 were enriched in GO terms of “ribonucleotide binding” (*SI Appendix, Fig. S5*), suggesting that genes required for DNA synthesis at S phase are, at least in part, actively transcribed at G0/G1. Finally, genes with decreased transcription at G0/G1 compared with the G1/S and M phases were enriched in GO terms of “cytosolic ribosomes” and “organelles” (Fig. 1D and *SI Appendix, Fig. S5*), and the duplication of organelles and cytosolic components are the major activities at the G1 phase. Taking these results together, we observed a prevailing lag of mRNA abundance compared with gene transcription, suggesting that the transcriptional activation of genes precedes the accumulation of their transcribed products at the cell-cycle level. These results demonstrate well-orchestrated transcriptional dynamics during the cell cycle.

Active Transcription at Early Mitosis. To investigate further the dynamic pattern of transcription at different cell-cycle stages, we performed unsupervised *k*-means clustering of all the differentially transcribed genes in GRO-seq samples (Fig. 2A). Among the six dynamic transcription patterns, cluster 3 contains genes with highest transcription at M phase (Fig. 2A). The transcription complexes are in general inactivated and disassembled from chromatin during chromosome condensation in the prophase, and the mitotic phase is known for silence of transcription (42). Although a fraction of transcription factors, including mixed-lineage leukemia (MLL), bromo-domain containing 4 (BRD4), Forkhead box A1 (FOX A1), and GATA-binding protein 1 (GATA1), are retained on mitotic chromosomes, they are thought to facilitate rapid gene reactivation post mitosis (43–46). Our thymidine-nocodazole blocking followed by a shake-off method enriches cells at the early mitotic phase. The

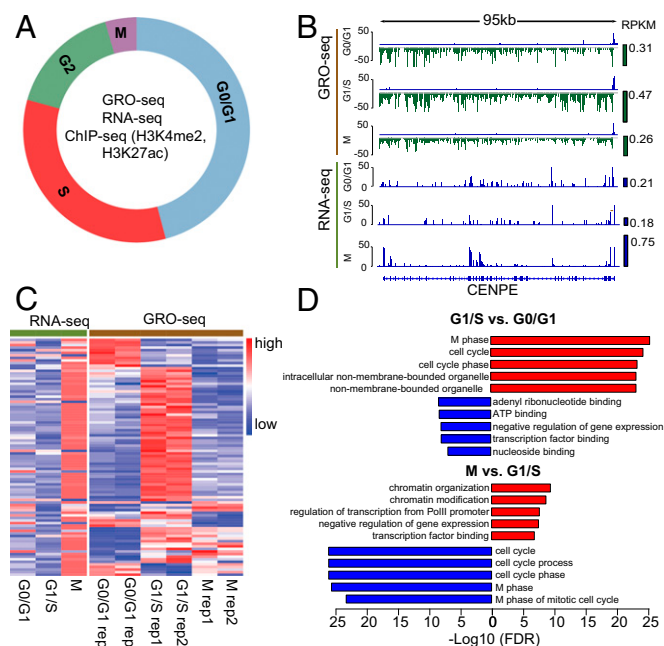


Fig. 1. GRO-seq and RNA-seq identify different cell-cycle-regulated genes. (A) Illustration of transcriptional dynamics analysis across the cell-cycle stages in MCF-7 cells. GRO-seq and ChIP-seq experiments were performed in two biological replicates, and RNA-seq was performed without replicates. (B) Transcription and expression of *CENPE* as measured by GRO-seq and RNA-seq at different cell-cycle stages. Green and blue bars on the right side of the signal tracks represent the *CENPE* transcription and expression levels as measured by reads per kilobase per million mapped reads (RPKM). (C) Transcription (GRO-seq) and expression (RNA-seq) of curated mitotic genes. The genes specifically up-regulated at G2/M were curated from published datasets (*Materials and Methods*). Read counts of each gene were normalized among the three cell-cycle stages so that their mean equals 0 and the SD equals 1, with red representing higher signal and blue representing lower signal. (D) GO analysis of cell-cycle-stage-specific genes identified by GRO-seq analysis. Bar length represents the $-\log_{10}$ FDR. Red bars indicate terms enriched for up-regulated genes; blue bars indicate terms enriched for down-regulated genes. The top five enriched terms are shown for each comparison.

identification of genes with peak transcription at early M phase using GRO-seq was intriguing. Analysis of RNA-seq data revealed that this group of genes had the highest expression level at G0/G1 (*SI Appendix, Figs. S6 and S7*), consistent with a lag of RNA-seq in reflecting the transcription observed for the mitotic genes (Fig. 1C).

Further inspection of the top differentially transcribed genes in cluster 3, such as *TNS3* and *LDLRAD4*, found them to be of larger size (Fig. 2B). We thus analyzed the length distribution of the genes in the six clusters. The one-tailed Wilcoxon rank sum test between each cluster pair showed that cluster 3 is significantly enriched for longer genes (Fig. 2C and *SI Appendix, Fig. S8A*). We then binned all differentially transcribed genes into 10 length intervals (with interval 1 containing the shortest genes and interval 10 containing the longest genes) and plotted the proportion of genes in each interval for the six clusters. More than 20% of the cluster 3 genes were in interval 10, compared with 6–11% of the genes in the other clusters (*SI Appendix, Fig. S8B*). The longer genes had relatively lower transcription levels than shorter ones (Fig. 2D).

To compare the transcription pattern of these long genes during the cell cycle, we plotted the GRO-seq signal along the gene body for the cluster 3 genes in the longest gene length interval (*SI Appendix, Fig. S8C*). Genes were equally divided into 50 bins from the TSS to cleavage/polyadenylation (CPA) site, and the average GRO-seq signal within each bin was summarized. Many of these genes had a strong GRO-seq signal at the TSS and a weak signal in the gene body at G0/G1, indicating paused Pol II. Interestingly, as the cell cycle progressed the signal in the gene body became stronger, especially at M phase. To determine whether the strong GRO-seq signal observed at M phase is from paused or actively elongating Pol II, we reanalyzed publicly available Pol II ChIP-seq data in mitotic HeLa cells treated with or without flavopiridol, which inhibits the transcription elongation factor P-TEFb and the release of promoter-proximal paused Pol II (47). Flavopiridol treatment resulted in a significantly stronger Pol II signal at the promoter regions (*SI Appendix, Fig. S9*), suggesting that these Pol II were actively released in early mitotic cells without flavopiridol treatment. Collectively, our analysis identified genes

actively transcribed at early M phase, which tend to be longer genes with lower expression.

H3K27ac and H3K4me2 Signals Increase Globally at Mitosis. Given the importance of changes in chromatin structure during the cell cycle (48–50), we performed histone-modification ChIP-seq of H3K4me2 and H3K27ac to investigate chromatin dynamics across the cell cycle. When normalized to same sequencing depth, the peak numbers at different cell-cycle stages were very similar (*SI Appendix, Table S3*). In addition, the correlations of the ChIP-seq signal among different cell-cycle stages were very high (*SI Appendix, Figs. S10 and S11*), indicating that the local histone-modification states were remarkably stable despite the dramatic changes in chromosome organization across the cell cycle (5).

We then grouped genes into high, medium, and low categories based on the expression level calculated from RNA-seq data and plotted the normalized tag counts from the histone marks ChIP-seq data at promoter regions (TSS $\pm$ 1 kb). Consistent with a previous report (51), both H3K4me2 and H3K27ac were positively correlated with gene-expression level (*SI Appendix, Fig. S12 A and B*). A similar trend was observed when genes were grouped based on the transcription level calculated from GRO-seq data (*SI Appendix, Fig. S12 A and B*). We then sought to determine whether the differential transcription across the cell cycle was correlated with changes in local histone-modification states. Intriguingly, when normalized to the same sequencing depth, the H3K4me2 and H3K27ac signals were much stronger at M phase for all the differentially transcribed genes, regardless of their transcription patterns across the cell-cycle stages (Fig. 3A and B and *SI Appendix, Fig. S8 C and D*).

To investigate whether this correlation results from the global increase of the histone modifications at M phase, we plotted the ChIP-seq signal at the union of peaks identified from all the samples and indeed found both the H3K4me2 and the H3K27ac signals were significantly higher at M phase (Fig. 3C). We then performed Western blot analysis of H3K4me2 and H3K27ac at different cell-cycle stages. When normalized to total H3 or tubulin, the protein levels of modified H3 were higher at M phase than at the G0/G1 and G1/S phases (Fig. 3D), as was consistent with the ChIP-seq signals.

Identification of Cell-Cycle-Regulated eRNAs. Pol II and many other transcription factors associate with a large number of enhancers marked by H3K4me1 and H3K27ac and produce noncoding eRNAs that have been demonstrated to play an important role in transcriptional regulation (30, 31). Therefore we sought to determine whether there were cell-cycle-regulated eRNAs and to identify their role in cell-cycle progression. To this end, we developed a computational pipeline to characterize cell-cycle-regulated eRNAs (Fig. 4A). H3K27ac ChIP-seq peaks at different cell-cycle stages were merged, and promoter regions were filtered out, leaving $\sim$ 50,000 peaks as potential active enhancers. Because eRNAs are often transcribed bidirectionally from enhancer regions (31), we applied a sliding window approach to identify enhancer regions with bidirectional transcription (Fig. 4B). This analysis identified a total of 4,922 eRNAs, 2,440 of which were differentially transcribed at different cell-cycle stages (Fig. 4C).

To investigate the role of eRNAs in transcriptional regulation and cell-cycle progression, we analyzed the correlation between differentially transcribed eRNAs and genes. Binding and Expression Target Analysis (BETA) software (52) was used to reveal a relationship between eRNAs and genes. Accordingly, a strong correlation was identified between eRNAs and differentially transcribed genes as determined by GRO-seq (Fig. 4D and *SI Appendix, Fig. S13*). Similar analysis performed with RNA-seq data found much weaker correlation between the eRNAs and differentially expressed genes (*SI Appendix, Fig. S14*), further emphasizing the advantage of analyzing temporal transcription regulation from GRO-seq rather than analyzing steady-state gene expression from RNA-seq.

Analysis of eRNAs Identifies Kruppel-Like Factor 4 KLF4 as a Key Regulator of G1/S Transition. After confirming the correlation between cell-cycle-regulated eRNAs and genes, we sought to identify

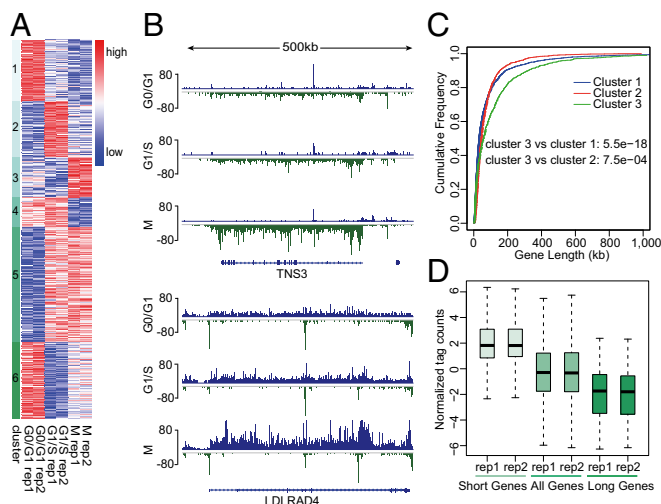


Fig. 2. Active transcription at early M phase. (A) Clustering of differentially transcribed genes at different cell-cycle stages identified by GRO-seq analysis. Genes are classified into six clusters using unsupervised k -means clustering. (B) Representative long genes up-regulated at M phase. Blue and green bars represent the GRO-seq signal from the plus and minus strands, respectively. Signals were normalized to a total of 10 million reads, and replicates were combined. (C) Empirical cumulative distribution of gene length. Blue, red, and green traces represent the groups of genes with higher GRO-seq signal at G0/G1, G1/S, and M phases, respectively. (D) GRO-seq signal of short (top 25% shortest, two far-left boxes), all (middle two boxes), and long (top 25% longest, two far-right boxes) genes that are highly expressed at M phase (red trace in C).

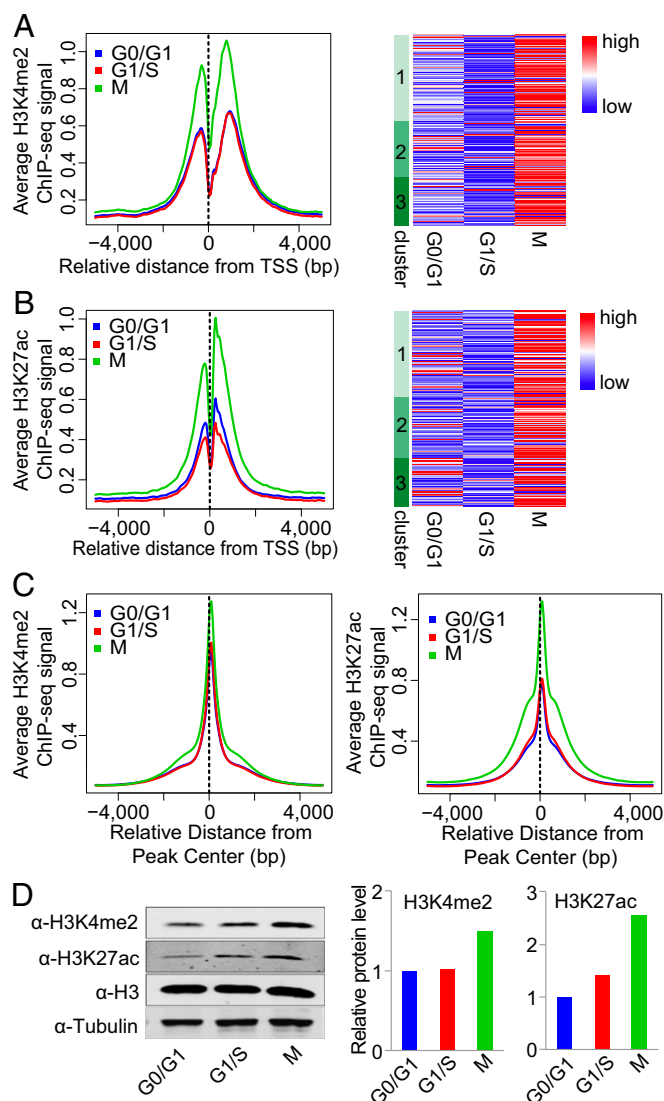


Fig. 3. Global increase in H3K4me2 and H3K27ac signals at early M phase. (A and B) H3K4me2 (A) and H3K27ac (B) ChIP-seq signal at promoter regions across different cell-cycle stages. (Left) Average H3K4me2 (A) and H3K27ac (B) signal (RPKM) of genes in clusters 1, 2, and 3. (Right) Heatmap of H3K4me2 (A) and H3K27ac (B) signals of genes in clusters 1, 2, and 3. (C, Left) Average H3K4me2 ChIP-seq signal at all H3K4me2 peak regions. (Right) Average H3K27ac ChIP-seq signal at all H3K27ac peak regions. (D, Left) Western blot of H3K4me2 and H3K27ac at different cell-cycle stages. H3 and tubulin were used as loading controls. (Right) Bar plots represent the quantitative value (normalized to H3) of H3K4me2 and H3K27ac protein levels from the Western blot.

the transcription factors that regulate these eRNAs. Transcription factors usually bind to open chromatin regions. To improve the spatial resolution of the transcription factor-binding sites in eRNA regions, we performed DNase sequencing (DNase-seq) in MCF-7 cells to identify open chromatin regions that were hypersensitive to DNase I digestion. Subsequently, we performed motif analysis at DNase I hypersensitivity sites (DHS) in the eRNA regions (SI Appendix, Fig. S15). This analysis identified motifs of multiple transcription factors including Kruppel-like factor 4 (KLF4), which ranked top in the eRNA regions that are highly transcribed at G0/G1 (Fig. 5A and SI Appendix, Table S4). KLF4 is important in cell-cycle control, cellular differentiation, and carcinogenesis (53–55). We then compared published KLF4 ChIP-seq peaks (56) with eRNAs and found

that KLF4 binding was enriched in eRNA regions with peaked transcription at G0/G1 (Fig. 5B and SI Appendix, Fig. S16).

In addition, GRO-seq revealed that the KLF4 transcription level was high at G0/G1 and decreased at the G1/S and M phases (Fig. 5C). A more dramatic decrease was observed at the protein level through Western blotting (Fig. 5D). To evaluate the role of KLF4 in cell-cycle progression, we silenced KLF4 by siRNAs in MCF-7 cells. FACS analysis showed a significant increase in the proportion of cells at S phase (Fig. 5E). Moreover, cell-proliferation analysis showed that silencing KLF4 promotes MCF-7 cell growth (Fig. 5F). Together the results indicate that KLF4 has an important role in blocking G1/S transition. Because transcription factors function by regulating the transcription of specific gene targets, we aimed to identify a KLF4-dependent gene signature that modulates this G1/S transition through the regulation of eRNAs. BETA analysis identified 10 potential KLF4 direct target genes, six of which were significantly down-regulated upon silencing of KLF4 (Fig. 5G). Among the six genes down-regulated by KLF4 silencing, *KRT19* has been reported to suppress cell proliferation, and silencing of *KRT19* leads to an increased proportion of cells at S phase (57). On the other hand, *CCND1*, a gene with decreased expression at S phase, is a well-characterized cell-cycle regulator that promotes the G1/S transition (58), and *NEAT1* and *HSPB1* have been reported to promote the proliferation of breast cancer cells (59, 60). Thus the exact mechanism by which KLF4 controls G1/S transition is unclear and warrants further investigation.

Discussion

In this study, we systematically investigated transcriptional and epigenetic dynamics during the cell cycle by analyzing GRO-seq, RNA-seq, and histone marks ChIP-seq data at G0/G1, G1/S, and M phases in the MCF-7 breast cancer cell line. Our study revealed (i) a lag between transcription and steady-state RNA expression at the cell-cycle level; (ii) a large amount of active transcription during early mitosis; (iii) a global increase in active histone modifications at mitosis; (iv) thousands of cell-cycle-regulated eRNAs; and (v) dynamic eRNAs bound by transcription factors such as KLF4 that regulate cell-cycle progression.

Steady-state mRNA abundance is influenced by a few factors, including transcription, RNA processing, maturation, and degradation. Therefore, measuring steady-state mRNA levels by microarray or RNA-seq techniques may not accurately reflect active transcription.

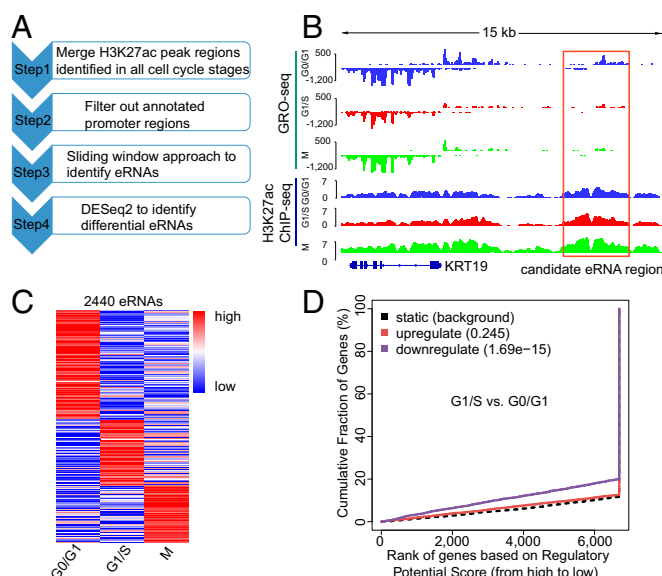


Fig. 4. Identification of cell-cycle-regulated eRNAs. (A) Workflow for identification of cell-cycle-stage-specific eRNAs. (B) Example of a candidate eRNA near gene *KRT19*. (C) Heatmap of 2,440 differential eRNAs. Color shows the relative GRO-seq signal at each eRNA region at three different stages. (D) Correlation between cell-cycle-regulated eRNAs and differentially transcribed genes.

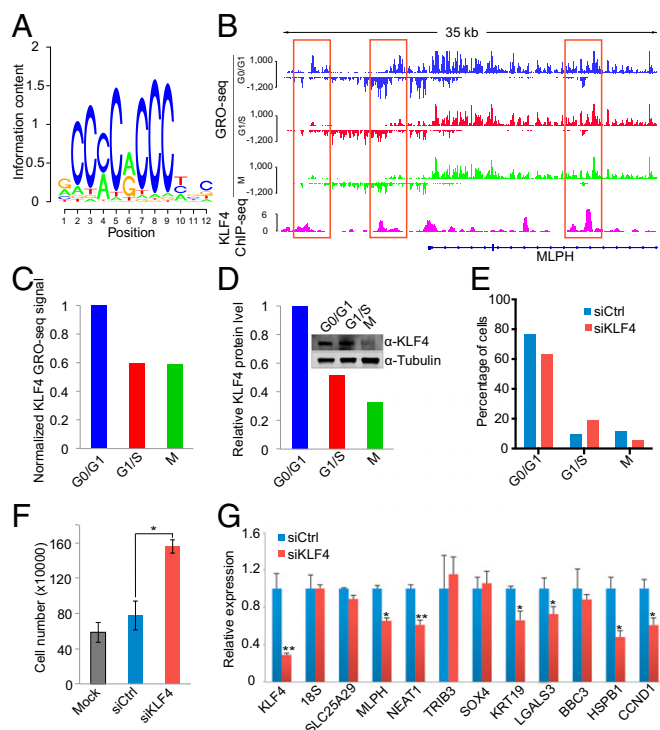


Fig. 5. KLF4 regulates eRNAs and target genes to control G1/S transition. (A) The KLF4 motif was identified as the most significant one enriched in G0/G1-specific eRNA regions. (B) Example of KLF4 eRNAs (marked with red boxes) near the target gene *MLPH*. (C and D) KLF4 RNA (C) and protein levels (D) across different cell-cycle stages. Tubulin was used as the loading control for Western blot analysis. (E) Flow cytometry cell-cycle analysis of MCF-7 cells in which KLF4 was silenced with siRNAs. (F) Cell growth analysis of MCF-7 cells upon silencing of KLF4. Mock, without transfection. (G) Validation of candidate KLF4 target genes identified through eRNA analysis. RT-PCR was performed for candidate KLF4 target genes upon silencing of KLF4. Error bars indicate the SEM; * $P < 0.05$; ** $P < 0.01$ (t test).

Indeed, GRO-seq and 4-thiouridine metabolic labeling followed by sequencing (4sU-seq) analyses that measure nascent transcription have revealed a broad inconsistency between transcription rate and mRNA levels (25, 28, 61, 62). Specifically, there is a delay in steady-state expression reflecting the transcription and mass production of rapidly degraded transcripts that are not detectable at the steady-state expression level. Most of the previous nascent transcription analyses were performed with unsynchronized cells or with synchronized cells within a short time window that was insufficient to cover multiple cell-cycle stages (26, 28, 29, 32, 35, 36, 62). Importantly, our GRO-seq and RNA-seq analysis at different cell-cycle stages revealed a lag between active transcription and steady-state expression during the cell cycle. The RNA degradation rate has been considered the most prominent measurable factor that contributes to the lag between transcription and accumulated RNA levels. Recent studies demonstrated that the half-lives of mammalian genes range from less than 1 min to more than 3 h (61, 62). In agreement with these observations, our data showed that mitotic genes are most highly transcribed at G1/S, and the genes most highly transcribed at M phase are more abundant at G0/G1, suggesting that these genes have an extremely long half-life.

Mitotic chromatin is transcriptionally inactive in general, and even ongoing transcriptions are aborted to ensure the integrity of the separating chromosomes (63). However, exceptions have been found in which the promoter of the cyclin B1 gene maintains an open chromatin configuration, and the gene is actively transcribed during mitosis (64). Recently, additional large-scale studies have revealed that part of the mitotic chromatin remains accessible to Pol II and transcription factors such as MLL, BRD4, GATA1, FOXA1, and AR (43–46, 65). Our GRO-seq data showed that although *CCNB1* transcription peaks at G1/S,

strong nascent transcription was observed at M phase. More interestingly, we identified a group of genes with a transcription peak at M phase. The observation that this group was enriched for unusually long genes made us hypothesize that the GRO-seq signal was from the incomplete transcription from previous stages (66). We therefore compared the GRO-seq signal along the gene body to identify the longest quarter of genes with the highest GRO-seq signal at M phase. If the hypothesis is correct, we should be able to observe a GRO-seq signal pattern shifted from the TSS toward the CPA site during the cell-cycle progression from G0/G1 to M phase. Our analysis revealed a uniform distribution of signal along the gene body for most genes. In addition, reanalysis of publicly available Pol II ChIP-seq data in early mitotic cells pretreated with and without flavopiridol (47) confirmed that Pol II is actively engaged at the TSS of these genes. Together, the results suggested that the high GRO-seq signal of these genes arose from active transcription at early M phase rather than from incomplete transcription at the G0/G1 and G1/S phases. Importantly, Liang et al. (47) recently reported mitotic transcriptional activation as a mechanism to clear actively engaged Pol II from mitotic chromatin; this mechanism is consistent with our observation of active transcription at early mitotic cells.

In support of active transcription at M phase, we observed extremely stable chromatin states marked by active histone modifications H3K4me2 and H3K27ac across different cell-cycle stages. In addition, the total H3K4me2 and H3K27ac levels increased significantly at M phase. Previous studies have identified mitotic-specific H4K20 methylation and the dynamics of H3K36 and H3K27 methylations across the cell cycle (67–69). The functional role of post-transcriptional histone modifications in the cell cycle is still largely unknown and warrants further analysis. It is worth noting that these observations were made in cancer cells with uncontrolled cell division; these cells may differ from normal cells with more stringent cell-cycle regulation (70). Future studies are warranted to explore the mechanisms underlying the active transcription during mitosis in normal and cancer cells.

Taken together, our analyses identified thousands of eRNAs and related transcription factors that are highly correlated with cell-cycle-regulated transcription but not with steady-state expression, thus highlighting the importance of transcriptional and epigenetic dynamics during cell-cycle progression. Overall, our study provides a comprehensive view of transcriptional landscape across the cell cycle and deepens our understanding of transcriptional dynamics during cell cycle. Future studies combining transcription, expression, and proteomics data at more detailed time courses are warranted to provide a more comprehensive view of cell-cycle regulation.

Materials and Methods

Cell Culture and Synchronization. The MCF-7 cells were obtained from ATCC and were cultured in DMEM medium supplemented with 10% (vol/vol) FBS, 1% penicillin-streptomycin, and 1% glutamine in a 5% (vol/vol) CO₂ humidified incubator. Cells were synchronized to G0/G1, G1/S, and M phase with hormone starvation, thymidine double treatment, and thymidine-nocodazole treatment, respectively. For detailed operations, see *SI Appendix, Supplemental Methods*.

GRO-Seq Library Construction. Nuclear run-on experiments (*SI Appendix, Supplemental Methods*) were performed as described previously (25). The resultant RNA was purified further with the TURBO DNA-free kit (AM1907; Life Technologies) to remove residue DNA contamination. Libraries then were constructed with the Encore Complete RNA-Seq DR Multiplex System 1–8 (0333-32; NuGEN Technologies, Inc.) and were sequenced to 50 bp with an Illumina HiSeq machine.

Additional experimental procedures and methods are described in *SI Appendix, Supplemental Methods*.

Availability of Data and Material. MCF-7 GRO-seq, RNA-seq, ChIP-seq, and DNase-seq raw sequence tags and processed bed files have been submitted to the National Center for Biotechnology Gene Expression Omnibus (GEO) database under accession no GSE94479.

ACKNOWLEDGMENTS. This work was supported by National Natural Science Foundation of China Grant 31329003 (to X.S.L.), the Princess Margaret Cancer

Foundation (H.H.H.), Canada Foundation for Innovation and Ontario Research Fund Grant CF132372 (to H.H.H.), Natural Sciences and Engineering Research Council of Canada Discovery Grant 498706 (to H.H.H.), NIH Grant 1R01GM099409 (to X.S.L.), US Public Health Service Grants R01CA63113 and R01CA173023 (to J.A.D.), and a German National Academy of Sciences Leopoldina

Fellowship (to M.F.). H.H.H. holds an Institute Community Support (ICS)-Institute of Genetics Maud Menten New Principal Investigator Prize (ICS-145381); an Office of Management, Information, and Research (OMIR) Early Researcher Award; a Terry Fox New Investigator Award; and a Canadian Institutes of Health Research (CIHR) New Investigator Salary Award.

- Okayama H, et al. (1996) Cell cycle control in fission yeast and mammals: Identification of new regulatory mechanisms. *Adv Cancer Res* 69:17–62.
- McGill CJ, Brooks G (1995) Cell cycle control mechanisms and their role in cardiac growth. *Cardiovasc Res* 30(4):557–569.
- Piatelli M, Tanguay D, Rothstein T, Chiles T (2003) Cell cycle control mechanisms in B-1 and B-2 lymphoid subsets. *Immunol Res* 27(1):31–52.
- Fraser HB (2013) Cell-cycle regulated transcription associates with DNA replication timing in yeast and human. *Genome Biol* 14(10):R111.
- Schafer KA (1998) The cell cycle: A review. *Vet Pathol* 35(6):461–478.
- Zhang L, Ma H, Pugh BF (2011) Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Res* 21(6):875–884.
- Ramachandran S, Henikoff S (2016) Transcriptional Regulators Compete with Nucleosomes Post-replication. *Cell* 165(3):580–592.
- Bello MJ, Rey JA (2006) The p53/Mdm2/p14ARF cell cycle control pathway genes may be inactivated by genetic and epigenetic mechanisms in gliomas. *Cancer Genet Cytogenet* 164(2):172–173.
- Dash BC, El-Deiry WS (2004) Cell cycle checkpoint control mechanisms that can be disrupted in cancer. *Methods Mol Biol* 280:99–161.
- Malumbres M, Barbacid M (2009) Cell cycle, CDKs and cancer: A changing paradigm. *Nat Rev Cancer* 9(3):153–166.
- Wäsch R (2011) Targeting mitotic exit for cancer treatment. *Expert Opin Ther Targets* 15(7):785–788.
- Dominguez-Brauer C, et al. (2015) Targeting mitosis in cancer: Emerging strategies. *Mol Cell* 60(4):524–536.
- Dynlacht BD (1997) Regulation of transcription by proteins that control the cell cycle. *Nature* 389(6647):149–152.
- Brandeis M, Hunt T (1996) The proteolysis of mitotic cyclins in mammalian cells persists from the end of mitosis until the onset of S phase. *EMBO J* 15(19):5280–5289.
- Nevins JR, Potti A (2007) Mining gene expression profiles: Expression signatures as cancer phenotypes. *Nat Rev Genet* 8(8):601–609.
- Whitfield ML, et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 13(6):1977–2000.
- Bar-Joseph Z, et al. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc Natl Acad Sci USA* 105(3):955–960.
- Peña-Díaz J, et al. (2013) Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication. *Nucleic Acids Res* 41(5):2846–2856.
- Grant GD, et al. (2013) Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Mol Biol Cell* 24(23):3634–3650.
- Sadasivam S, Duan S, DeCaprio JA (2012) The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes Dev* 26(5):474–489.
- Sadasivam S, DeCaprio JA (2013) The DREAM complex: Master coordinator of cell cycle-dependent gene expression. *Nat Rev Cancer* 13(8):585–595.
- Bertoli C, Klier S, McGowan C, Wittenberg C, de Bruin RA (2013) Chk1 inhibits E2F6 repressor function in response to replication stress to maintain cell-cycle transcription. *Curr Biol* 23(17):1629–1637.
- Fischer M, Grossmann P, Padi M, DeCaprio JA (2016) Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res* 44(13):6070–6086.
- Gariglio P, Bellard M, Chambon P (1981) Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Res* 9(11):2589–2598.
- Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322(5909):1845–1848.
- Step SE, et al. (2014) Anti-diabetic rosiglitazone remodels the adipocyte transcriptome by redistributing transcription to PPAR γ -driven enhancers. *Genes Dev* 28(9):1018–1028.
- Jonkers I, Kwak H, Lis JT (2014) Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 3:e02407.
- Hah N, et al. (2011) A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* 145(4):622–634.
- Hah N, Murakami S, Nagari A, Danko CG, Kraus WL (2013) Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* 23(8):1210–1223.
- De Santa F, et al. (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* 8(5):e1000384.
- Kim TK, et al. (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465(7295):182–187.
- Lam MT, Li W, Rosenfeld MG, Glass CK (2014) Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39(4):170–182.
- Hsieh CL, et al. (2014) Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci USA* 111(20):7319–7324.
- Yang Y, et al. (2016) Enhancer RNA-driven looping enhances the transcription of the long noncoding RNA DHR54-AS1, a controller of the DHR54 gene cluster. *Sci Rep* 6:20961.
- Li W, et al. (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498(7455):516–520.
- Wang D, et al. (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474(7351):390–394.
- Min IM, et al. (2011) Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* 25(7):742–754.
- Barisic M, et al. (2015) Mitosis. Microtubule deetyrosination guides chromosomes during mitosis. *Science* 348(6236):799–803.
- Lovén J, et al. (2012) Revisiting global gene expression analysis. *Cell* 151(3):476–482.
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
- Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
- Segil N, Guermah M, Hoffmann A, Roeder RG, Heintz N (1996) Mitotic regulation of TFIID: Inhibition of activator-dependent transcription and changes in subcellular localization. *Genes Dev* 10(19):2389–2400.
- Zhao R, Nakamura T, Fu Y, Lazar Z, Spector DL (2011) Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nat Cell Biol* 13(11):1295–1304.
- Caravaca JM, et al. (2013) Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev* 27(3):251–260.
- Kadauke S, et al. (2012) Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* 150(4):725–737.
- Blöbel GA, et al. (2009) A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Mol Cell* 36(6):970–983.
- Liang K, et al. (2015) Mitotic transcriptional activation: Clearance of actively engaged Pol II via transcriptional elongation control in mitosis. *Mol Cell* 60(3):435–445.
- Pederson T (1972) Chromatin structure and the cell cycle. *Proc Natl Acad Sci USA* 69(8):2224–2228.
- Nair VD, et al. (2012) Involvement of histone demethylase LSD1 in short-time-scale gene expression changes during cell cycle progression in embryonic stem cells. *Mol Cell Biol* 32(23):4861–4876.
- Singh AM, et al. (2015) Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. *Stem Cell Rep* 5(3):323–336.
- Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823–837.
- Wang S, et al. (2013) Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat Protoc* 8(12):2502–2515.
- Hu D, et al. (2015) Interplay between arginine methylation and ubiquitylation regulates KLF4-mediated genome stability and carcinogenesis. *Nat Commun* 6:8419.
- Hu D, Zhou Z, Davidson NE, Huang Y, Wan Y (2012) Novel insight into KLF4 proteolytic regulation in estrogen receptor signaling and breast carcinogenesis. *J Biol Chem* 287(17):13584–13597.
- Gamper AM, et al. (2012) Regulation of KLF4 turnover reveals an unexpected tissue-specific role of pVHL in tumorigenesis. *Mol Cell* 45(2):233–243.
- Mohammed H, et al. (2013) Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Reports* 3(2):342–349.
- Ju JH, et al. (2013) Regulation of cell proliferation and migration by keratin19-induced nuclear import of early growth response-1 in breast cancer cells. *Clin Cancer Res* 19(16):4335–4346.
- Santra MK, Wajapeyee N, Green MR (2009) F-box protein FBXO31 mediates cyclin D1 degradation to induce G1 arrest after DNA damage. *Nature* 459(7247):722–725.
- Ke H, et al. (2016) NEAT1 is Required for Survival of Breast Cancer Cells Through FUS and miR-548. *Gene Regul Syst Bio* 10(Suppl 1):11–17.
- Cayado-Gutiérrez N, et al. (2013) Downregulation of Hsp27 (HSPB1) in MCF-7 human breast cancer cells induces upregulation of PTEN. *Cell Stress Chaperones* 18(2):243–249.
- Schwanhäusser B, et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473(7347):337–342.
- Rabani M, et al. (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* 29(5):436–442.
- Shermoen AW, O'Farrell PH (1991) Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* 67(2):303–310.
- Sciortino S, et al. (2001) The cyclin B1 gene is actively transcribed during mitosis in HeLa cells. *EMBO Rep* 2(11):1018–1023.
- McNair C, et al. (2016) Cell cycle-coupled expansion of AR activity promotes cancer progression. *Oncogene* 10(1038):1–14.
- Helmrich A, Ballarín M, Tora L (2011) Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* 44(6):966–977.
- Rice JC, et al. (2002) Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes Dev* 16(17):2225–2230.
- Zee BM, Britton LM, Wolle D, Haberman DM, Garcia BA (2012) Origins and formation of histone methylation across the human cell cycle. *Mol Cell Biol* 32(13):2503–2514.
- Ma Y, Kanakousaki K, Buttitta L (2015) How the cell cycle impacts chromatin architecture and influences cell fate. *Front Genet* 6:19.
- Shah MA, Denton EL, Arrowsmith CH, Lupien M, Schapira M (2014) A global assessment of cancer genomic alterations in epigenetic mechanisms. *Epigenetics Chromatin* 7(1):29.

Supporting materials

Transcriptional landscape of the human cell cycle

Yin Liu^{a,b,c,1}, Sujun Chen^{a,b,d,e,1}, Su Wang^{b,c,f}, Fraser Soares^d, Martin Fischer^g, Feilong Meng^h, Zhou Du^{b,c,i}, Charles Lin^j, Clifford Meyer^{c,k}, James A. DeCaprio^g, Myles Brown^{c,g,2}, X. Shirley Liu^{c,k,2}, and Housheng Hansen He^{d,e,2}

Supplemental Figures

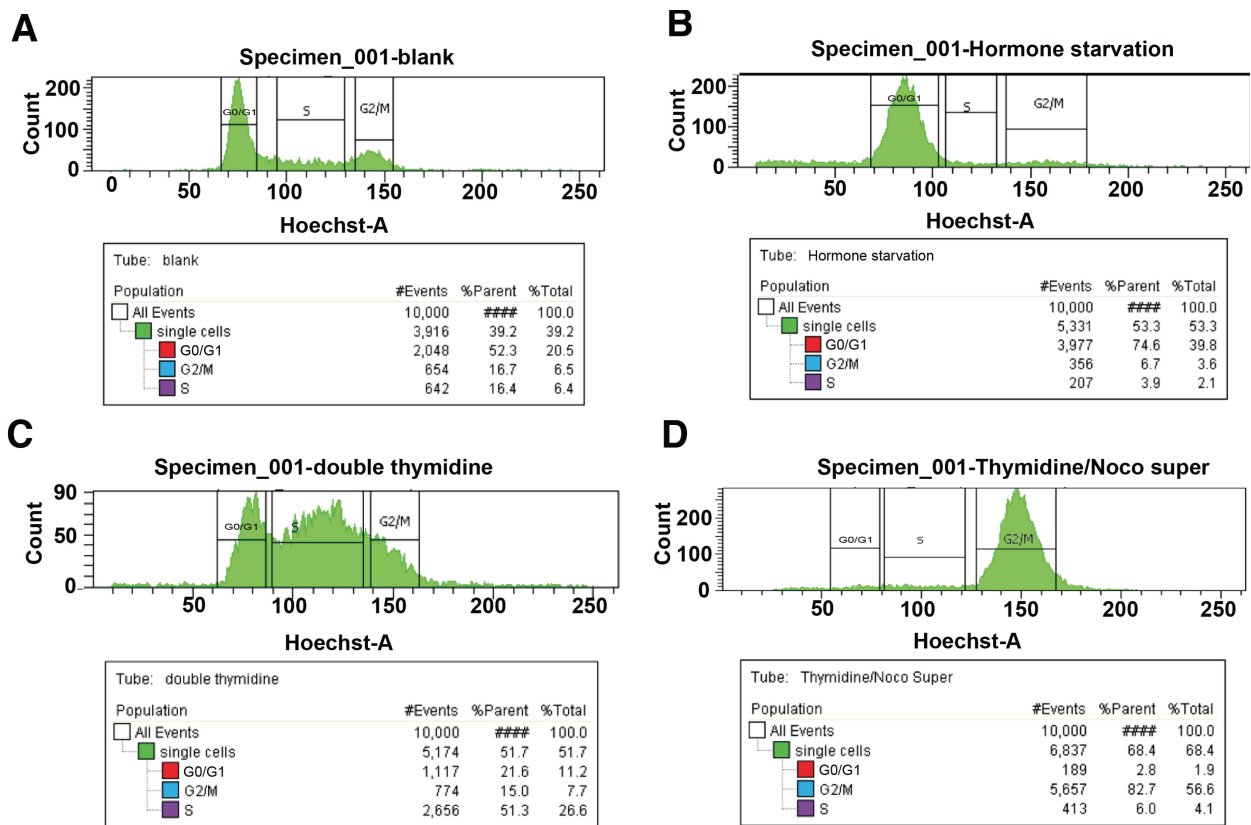


Figure S1. Synchronization of MCF-7 cells at different cell cycle stages. FACS showing DNA content of unsynchronized (A) and synchronized MCF-7 cells at (B) G0/G1, (C) G1/S and (D) M phases.

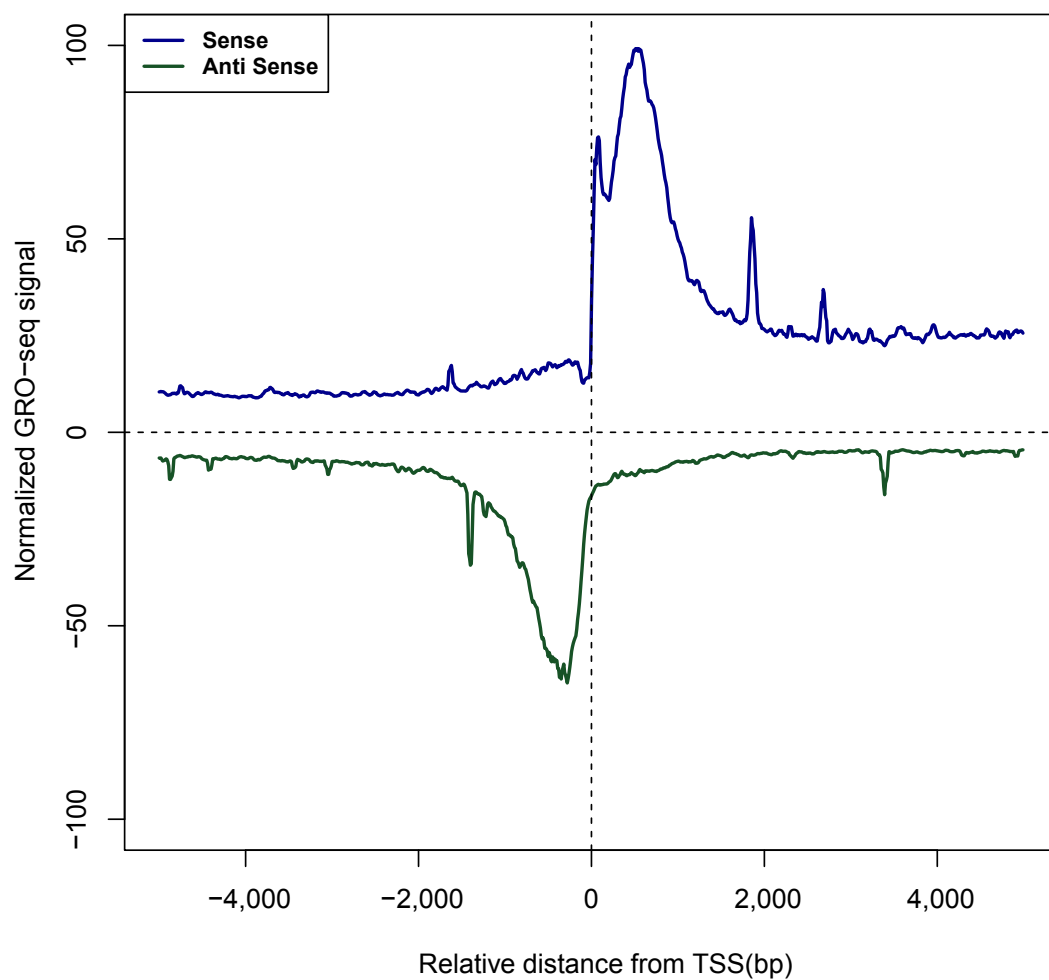


Figure S2. GRO-seq signal at the gene promoter regions. Divergent transcription at transcription start sites was observed. Sense: gene sense strand GRO-seq signal; Anti Sense: gene antisense strand GRO-seq signal. TSS: transcriptional start site.

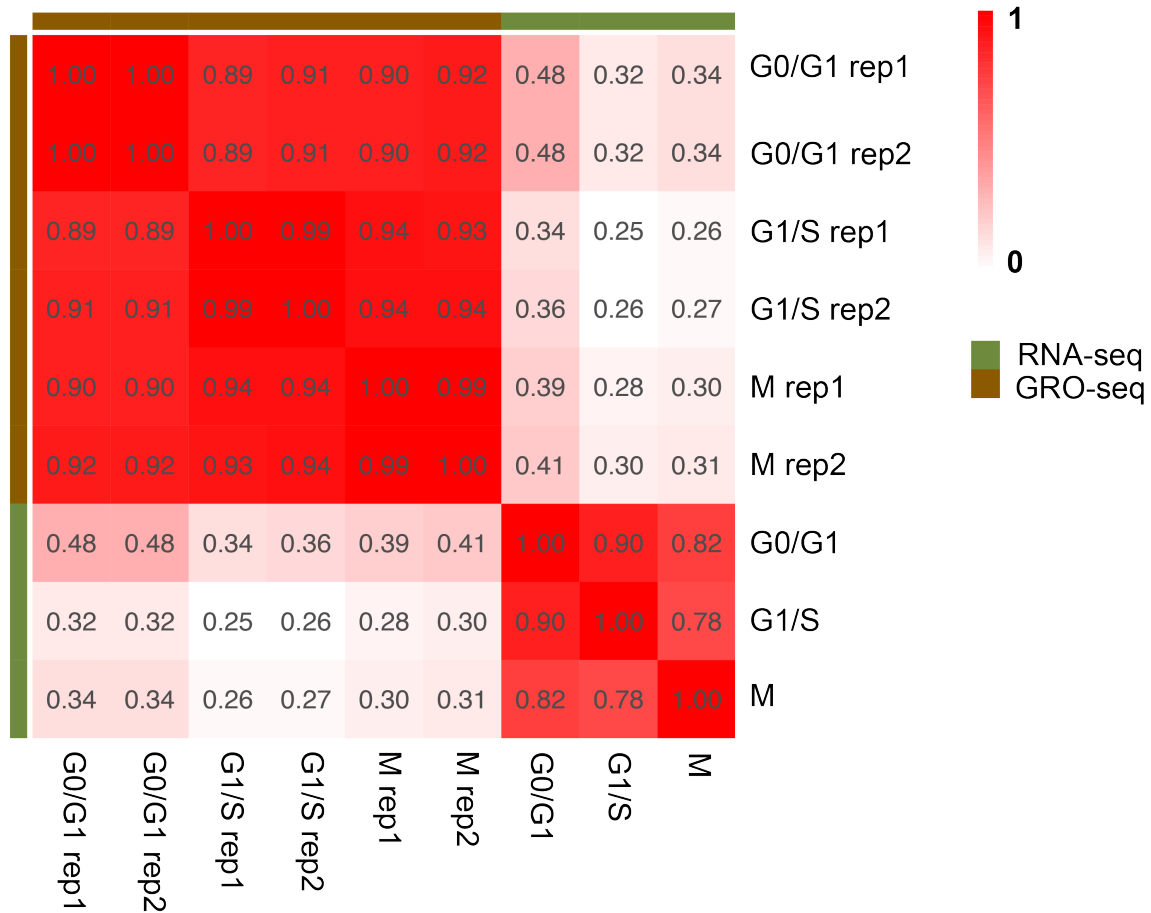


Figure S3. Correlation between GRO-seq and RNA-seq samples. GRO-seq and RNA-seq reads counts for all the RefSeq annotated genes were used to calculate pearson correlation between each sample.

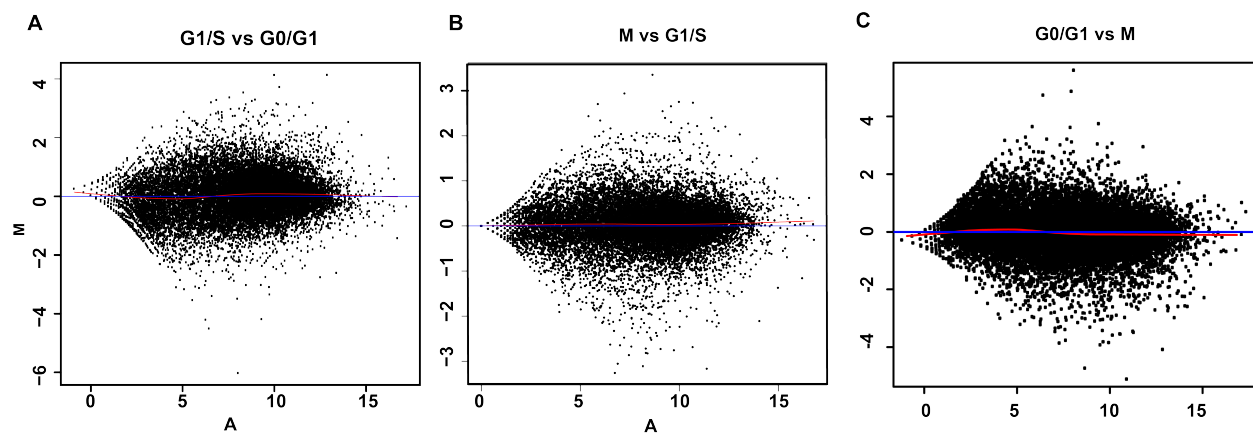


Figure S4. MAPlot of GRO-seq data. DESeq2 normalized GRO-seq signal was used to generated MAPlot for G1/S vs G0/G1(A), M vs G1/S (B) and G0/G1 vs M (C). Log2

transformed mean read counts with 1 as offset was used as ‘A’ while log2 transformed fold change was used as ‘M’.

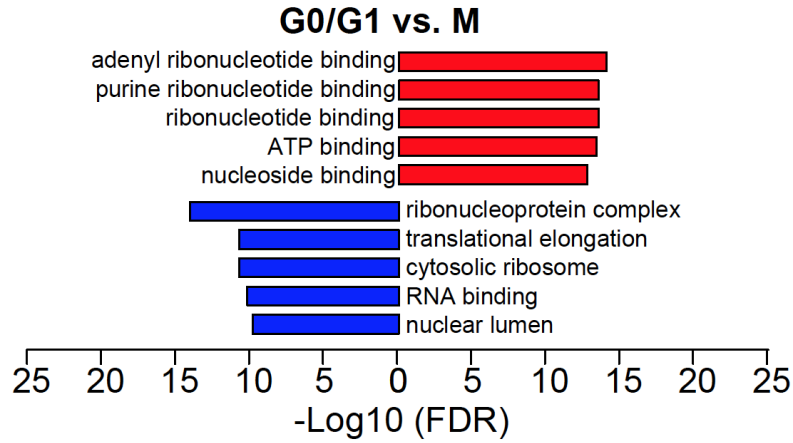


Figure S5. GO analysis of genes differentially transcribed between G0/G1 and M phases. Red color represents GO terms of genes with higher transcription level at G0/G1 and blue color represents GO terms of genes with lower transcription at G0/G1 compared with M phase. Top five enriched terms are shown for each comparison.

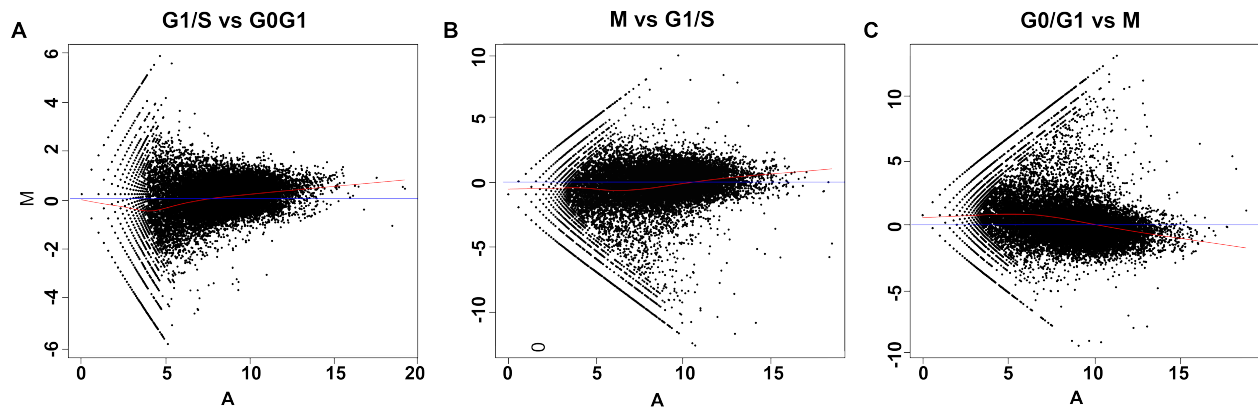


Figure S6. MAPlot for RNA-seq data. Gfold output is used to generated MAPlot for G1/S vs G0/G1(A), M vs S (B) and G0/G1 vs M (C). Log2 transformed read counts with 1 as offset is used as ‘A’ while log2 transformed fold change for ‘M’.

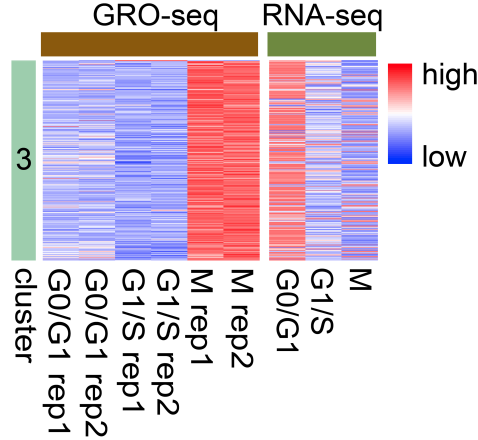


Figure S7. Normalized GRO-seq and RNA-seq signals for cluster 3 genes. Read counts for genes at different cell cycle stages are normalized so that their mean equals to 0 and standard deviation equals to 1, with red for higher signal and blue for lower signal.

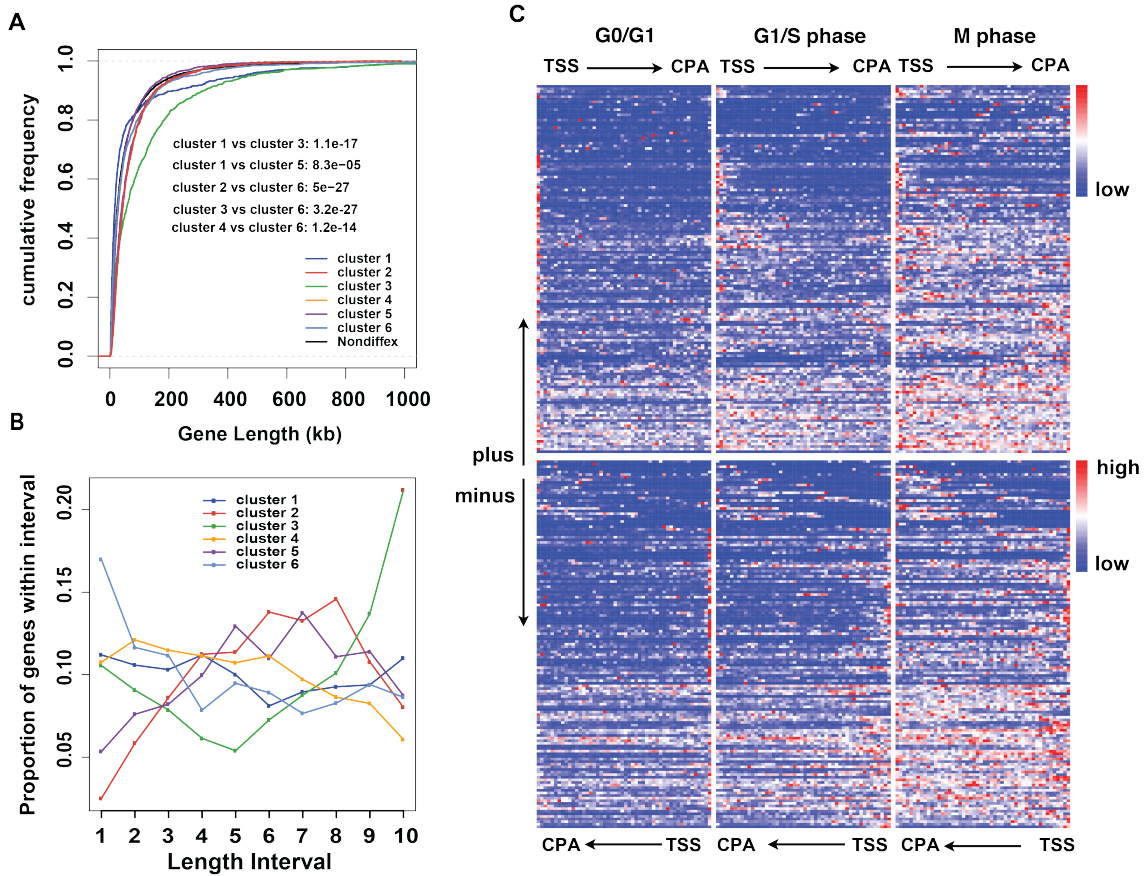


Figure S8. Genes upregulated at M phase tend to be long genes. (A) Empirical cumulative distribution of length of genes in the 6 clusters. (B) Proportion of genes in the 6 clusters in different length bins. All differential genes in the 6 clusters are divided into 10 bins according to their length, with 1 represents genes of shortest length and 10 represents genes of longest length. (C) GRO-seq signal across the gene body of M phase long genes (cluster three in figure panel A falls within the 10th interval of figure panel B) at G0/G1 (left panel), G1/S (middle) and M (right

panel) phases. Genes are divided into fifty bins and signal within each bin is summarized. The top and bottom three panels represent genes in plus and minus strands, respectively. CPA: cleavage/polyadenylation site.

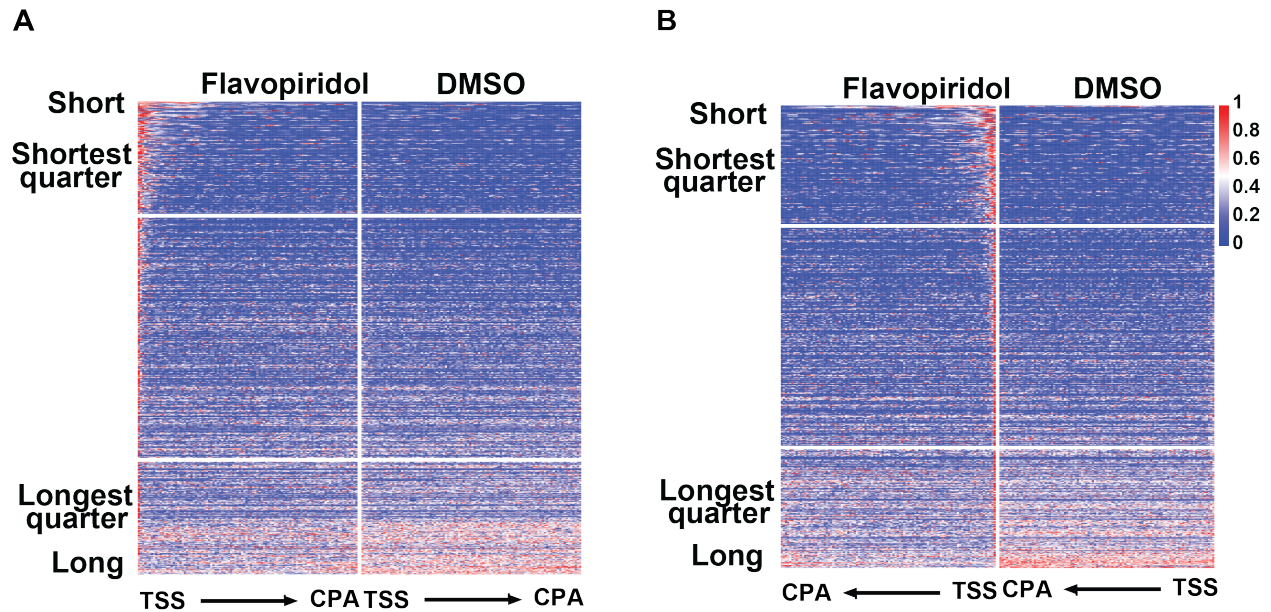


Figure S9. Active transcription of long genes at early M phase. Pol II signal (GSE71848) across the bodies of cluster 3 plus strand (A) and minus strand (B) genes at early M phase. Left shows signal with flavopiridol treatment while right for signal with DMSO treatment. Color represents normalized signal intensity with red indicates higher intensity and blue indicates lower intensity.

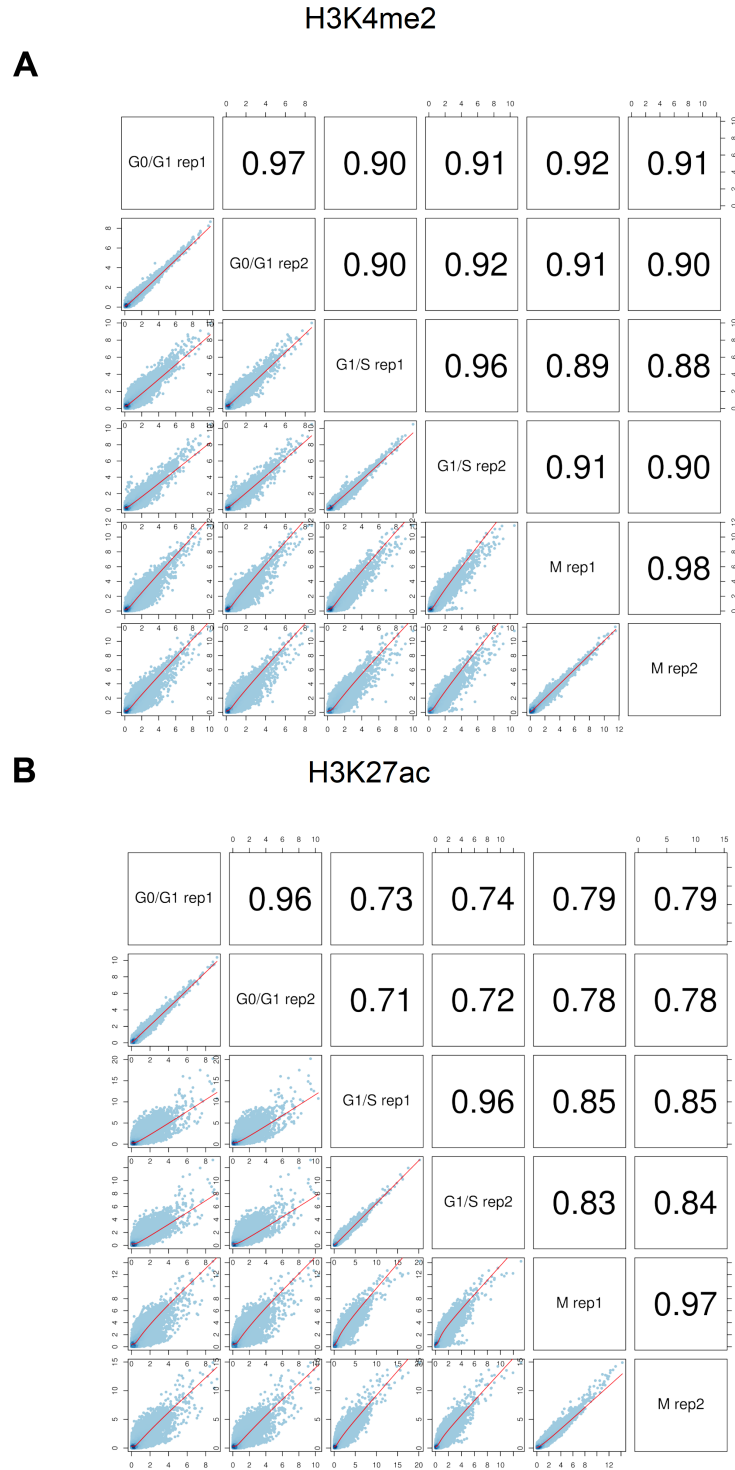


Figure S10. Correlation between ChIP-seq samples. Signals at H3K4me2 (A) and H3K27ac (B) peak regions were summarized and pearson correlation between samples was calculated.

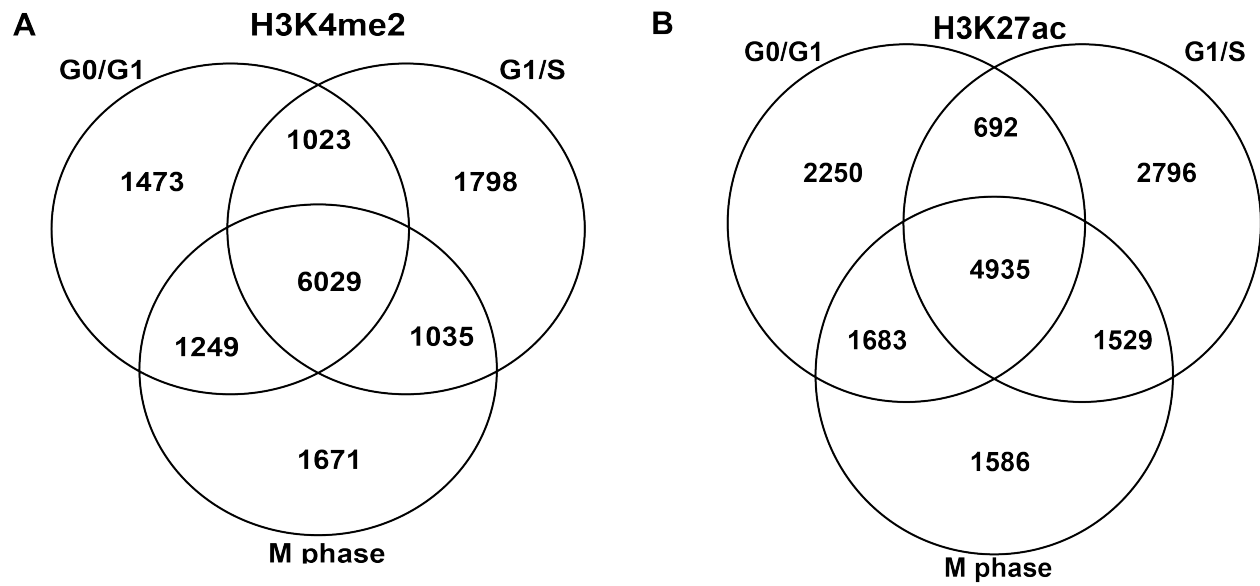


Figure S11. Overlap among ChIP-seq peaks from G0/G1, G1/S and M phase. Top 10000 peaks ranked by P value were used for H3K4me2 (A) and H3K27ac (B) ChIP-seq comparison.

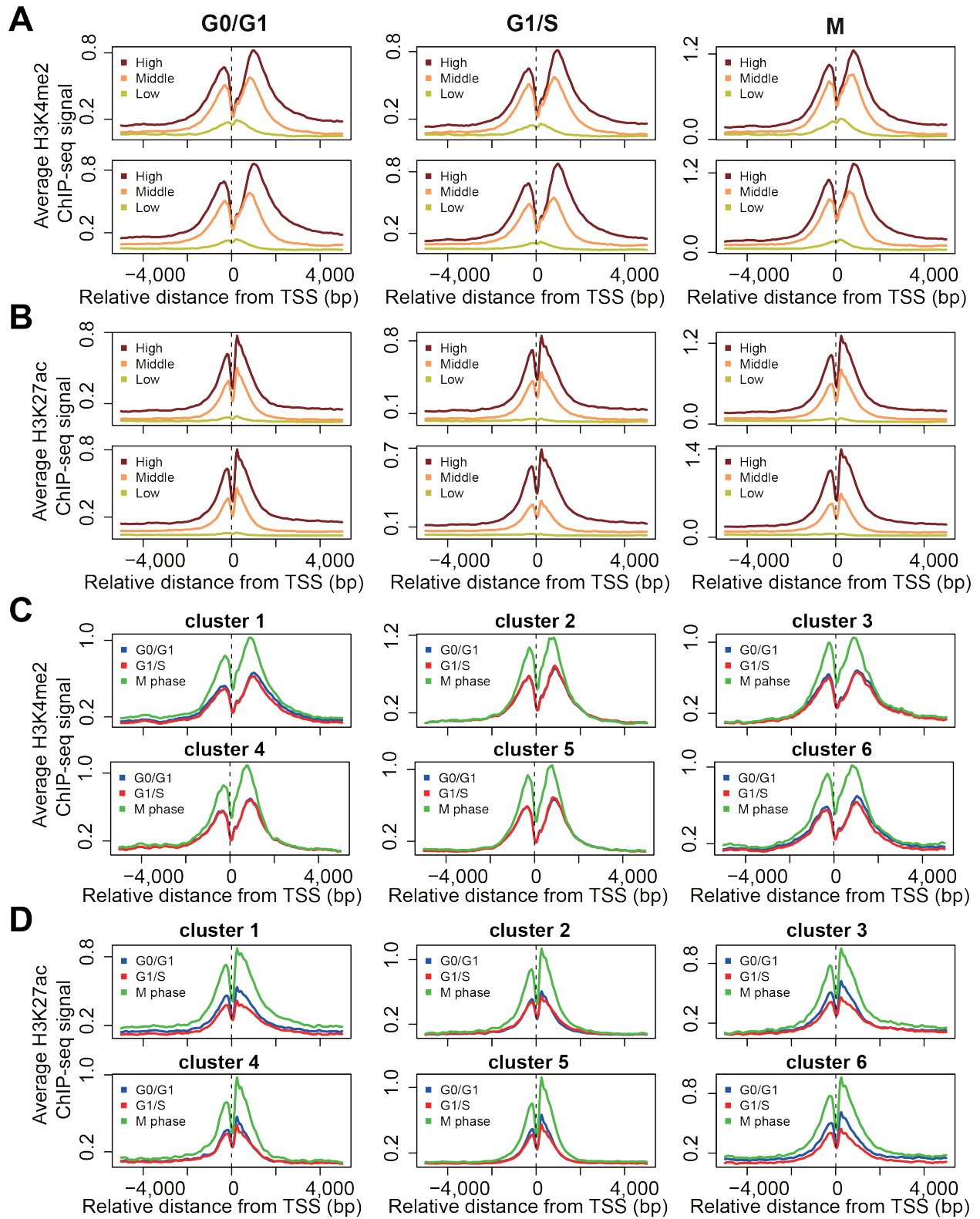


Figure S12. Correlation between RNA-seq, GRO-seq and histone modifications (H3K4me2 and H3K27ac) ChIP-seq signals. Average H3K4me2 (A) and H3K27ac (B) signals at G0/G1 (left panel), G1/S (middle panel) and M (right panel) phases at promoter regions. Genes are

divided into high, medium and low expression bins according to expression (RNA-seq; upper panel) and transcription (GRO-seq; bottom panel) levels. Box represents expression value of genes in each bin as measured by normalized GRO-seq read counts on gene body. Average (C) H3K4me2 and (D) H3K27ac ChIP-seq signal of genes in the 6 clusters.

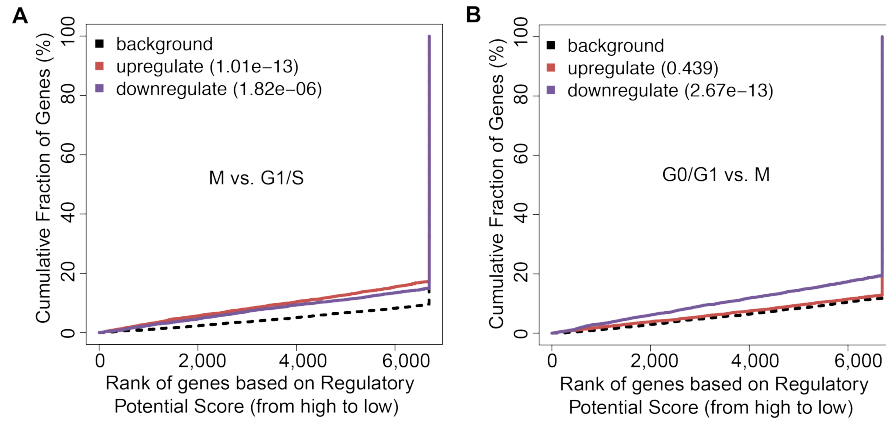


Figure S13. Correlation between cell cycle regulated eRNAs and differentially transcribed genes (calculated from GRO-seq). BETA analysis was performed for genes differentially transcribed between M and G1/S phases (A), G0/G1 and M phases (B).

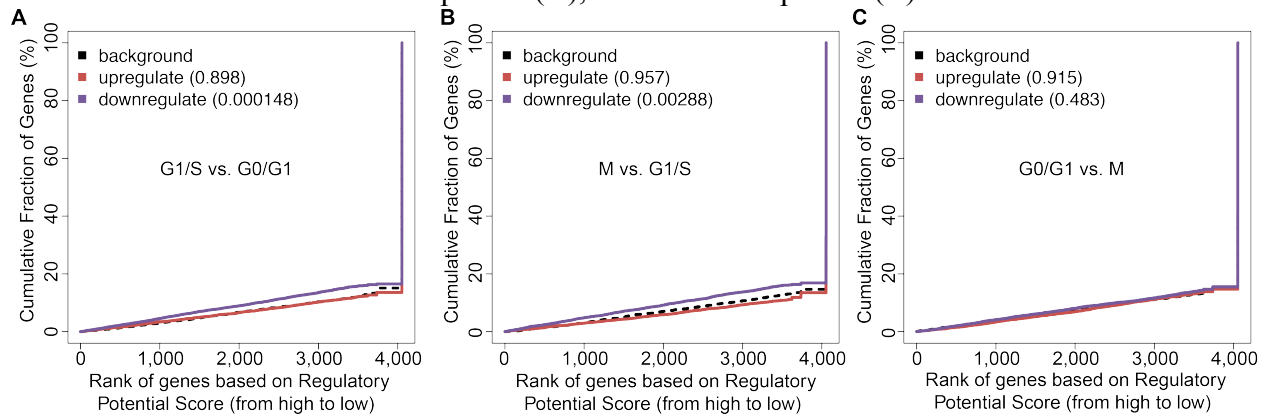


Figure S14. Correlation between cell cycle regulated eRNAs and differentially expressed genes (calculated from RNA-seq). BETA analysis was performed for genes differentially expressed between G1/S and G0/G1 (A), M and G1/S phases (B), G0/G1 and M phases (C).

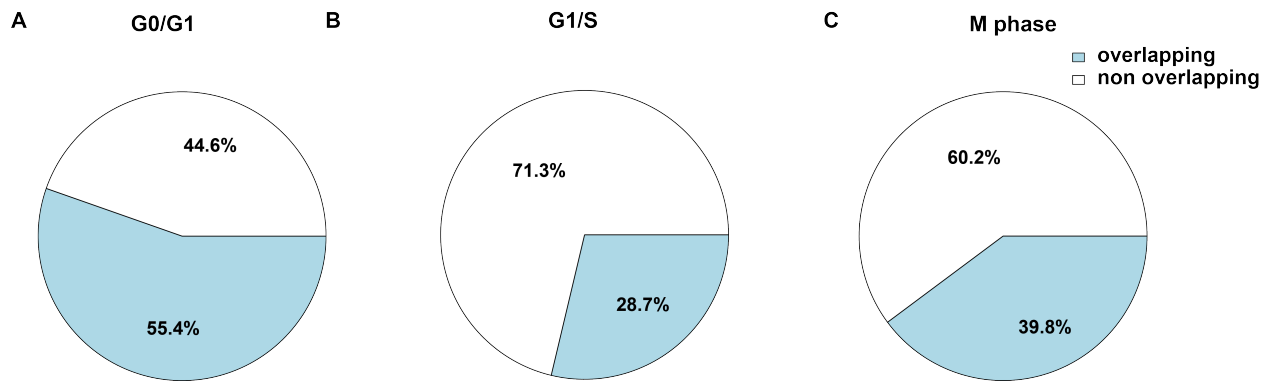


Figure S15. Overlap between eRNAs and DNase I hypersensitivity sites (DHSs). Fraction of identified eRNAs overlap with DHSs at G0/G1(A), G1/S phase (B) and M phase(C). Cornflower blue represents the proportion of eRNAs overlap with DHSs while white represents the non-overlapping proportion. Numbers indicate corresponding percentage.

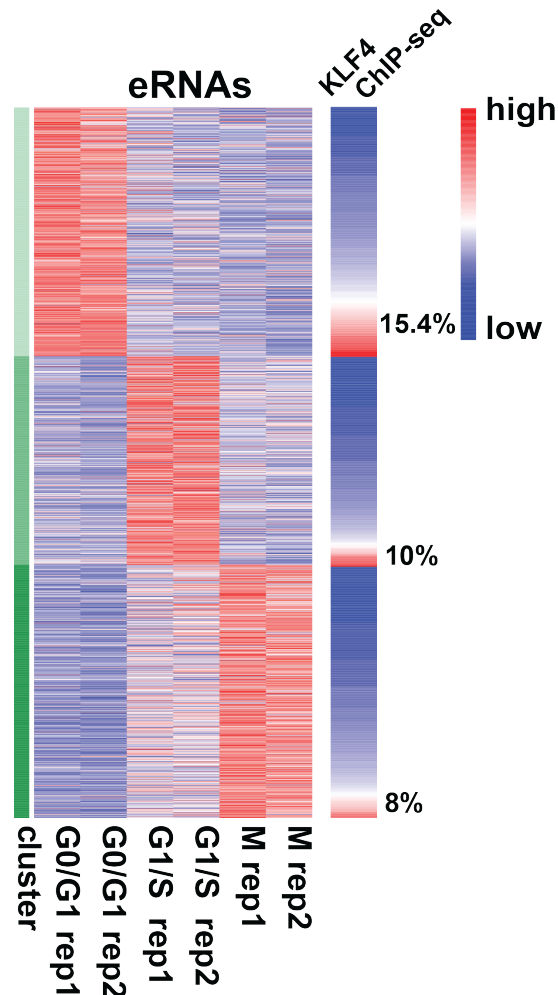


Figure S16. Correlation between eRNA and KLF4 binding. Left panel: clusters of cell cycle stage specific eRNAs. Read counts for eRNAs among different stages are normalized so that their mean equals to 0 and standard deviation equals to 1, with red for higher signal and blue for lower signal. Right panel: KLF4 ChIP-seq signal at eRNA regions. Numbers on the right side of

KLF4 ChIP-seq signal panel represent the percentage of eRNA regions overlapping with KLF4 ChIP-seq peaks in each cluster.

Supplemental Tables

Table S1: Summary of GRO-seq data.

Samples	Total Reads	Mapped Reads	Reads in Gene	Percentage of Reads within Genes	Correlation between Replicates
G0/G1 rep1	56,685,478	35,369,554	24,243,734	69	0.99
G0/G1 rep2	60,368,552	36,383,854	24,950,768	69	
G1/S rep1	52,774,862	32,345,172	22,763,712	70	0.99
G1/S rep2	59,272,432	38,132,400	27,121,920	71	
M rep1	71,371,192	44,652,594	31,706,331	71	0.99
M rep2	77,329,328	50,918,378	36,241,287	71	

Table 2: Summary of RNA-seq data.

Samples	Total Reads	Mapped Reads	Reads in Exon	Reads in UTRs (5'+3')	Percentage of Reads within Genes
G0/G1	42,323,301	32,883,868	8217,475	8,663,122	51
G1/S	38,550,123	30,261,889	8,708,944	9,456,981	60
M	30,128,558	30,128,558	6,020,357	9,670,172	52

Table S3: Summary of ChIP-seq data.

Histone Marks	Samples	Total Reads	Mapped Reads	Peak Number ^a	Peak Number Merged ^b	Correlation between Replicates
H3K4me2	G0/G1 rep1	26,388,765	20,667,934	98,035	105,653	0.97
H3K4me2	G0/G1 rep2	27,451,444	21,072,341	74,432		
H3K4me2	G1/S rep1	28,267,586	22,736,451	112,056	120,513	0.96
H3K4me2	G1/S rep2	21,415,085	17,336,836	97,834		
H3K4me2	M rep1	21,415,085	17,120,386	96,147	105,230	0.98
H3K4me2	M rep2	19,497,695	15,658,691	94,231		
H3K27ac	G0/G1 rep1	23,196,413	17,356,857	40,626	60,926	0.96
H3K27ac	G0/G1 rep2	19,363,495	14,506,659	41,703		
H3K27ac	G1/S rep1	13,966,918	10,941,444	54,484	64,664	0.96
H3K27ac	G1/S rep2	21,684,672	16,455,765	37,075		
H3K27ac	M rep1	26,372,329	21,106,214	66,859	76,359	0.97
H3K27ac	M rep2	20,837,808	15,523,175	49,207		

a: mappable reads down sampled to 15M for H3K4me2 and 10.9M for H3K27ac.

b: down sampled reads in the replicates merged.

Table S4: Motifs enriched in eRNA regions.

G0/G1_diff eRNA_down			G0/G1_diff eRNA_up		
motif id	factor	-10*LOG(pval)	motif id	factor	-10*LOG(pval)
M01588	KLF4 Klf4	690.776	M00915	AP-2 TFAP2A	533.247
M00931	SP1	657.94	M00470	TFAP2C	463.239
M00932	SP1	507.189	M01199	TRIM28	334.371
M00196	SP1	505.34	M00189	AP-2 TFAP2A	275.915
M00255	GC box	488.629	M01219	SP1 SP3	215.07
M00933	SP1	485.496	M00982	Egr1	205.468
M01273	SP4	459.837	M00800	TFAP2A	199.156
M01783	SP2 Sp2	436.219	M01045	AP-2alphaA	184.778
M01200	CTCF	372.797	M01047	AP-2alphaA	181.333
MA0079	SP1	362.492	M01122	ZNF219	135.999
M00803	E2F1 E2f1	331.231	M01643	NHP10	135.196
M01780	RAP1	272.713	M01068	UF1H3BE TA	132.855
M01643	NHP10	271.05	M01200	CTCF	129.899
M01057	ERF2	243.275	M00245	EGR3	113.516
M00958	ABI4	237.11	M00652	NFE2L1	106.16
M01175	ZBTB7B	234.49	M00244	EGR4	98.859
MA0139	CTCF	228.555	MA0139	CTCF	98.763
M00321	Muscle initiator	206.938	M00400	ABF1	97.79
M00695	TEAD2	204.242	M00698	HEB TCF1 2	91.205
M00323	Muscle initiator sequences-19	202.669	M00469	AP-2alpha	86.846
M00324	Muscle initiator sequence-20	201.675	MA0003	TFAP2A	86.846

M00008	SP1	179.615
M01199	TRIM28	172.098
M01118	WT1 Wt1	169.286
M00982	Egr1	169.193
M01219	SP1 SP3 Sp1	161.719
M01622	UGA3	161.423
M00778	AhR AHR	160.985
S_diff_eRNA_down		
motif id	factor	-10*LOG(pval)
M00915	AP-2 TFAP2A	425.29
M00189	AP-2 TFAP2A	257.064
M00982	Egr1	246.708
M01199	TRIM28	181.392
M00470	TFAP2C	180.17
M00652	NFE2L1	177.438
M00034	p53 TP53	162.067
M01122	ZNF219	158.356
M00442	ABF	149.596
M01196	CTF1	137.278
MA0119	TLX1 NFIC	137.278
M00469	AP-2alpha	137.203
MA0003	TFAP2A	137.203
M01045	AP-2alphaA	137.035
M00665	SP3	130.27
M00400	ABF1	114.277
M01643	NHP10	111.564
M01219	SP1	108.137
M00244	EGR4	102.563
M01068	UF1H3BET A	98.694
M01200	CTCF	85.807
MA0139	CTCF	72.845
M00245	EGR3	69.194

M00034	p53 TP53	82.658
MA0163	PLAG1	80.54

S_diff_eRNA_up		
motif id	factor	-10*LOG(pval)
M00697	HBP-1b	161.808
M00113	CREB	137.064
M00801	CREB	133.483
M01227	MAFB	128.706
M00694	E4F1	123.624
M00936	HBP-1a	118.992
M01168	SREBP	108.07
M00514	ATF4	107.082
M01034	Ebox	90.682
M00916	CREB	90.132
M00490	BACH2	88.707
M01580	Rtg3	86.518
M00944	CPRF-3	84.112
M00375	TGA1b	84.004
M00976	ARNT	83.527
M01586	TGA2	78.711
M00237	ARNT	72.091
M00778	AhR	69.65

M_diff_eRNA_down		
motif id	factor	-10*LOG(pval)
MA0099	JUN	690.776
M00199	AP-1	690.776
M01267	FOSL1	690.776
M00926	AP-1	690.776
M00490	BACH2	658.885
M00517	JUND	638.734
M01687	ARG RII	582.877
M00983	MAF	534.752
M00495	BACH1	528.145
M00204	GCN4	419.701
M00174	AP-1	404.155
M01795	Cap1	370.82
M00188	AP-1	365.739
M00037	NFE2	358.717
M00925	JUNB	305.676
M00924	AP-1	305.237
M00821	GABPA Gab pa	300.038
M01673	ARG RI	280.009
M00038	GCN4	262.864
M01555	GCN4	261.842
M00172	AP-1	210.159
M00514	ATF4	203.524
MA0150	NFE2L2	182.751

M_diff_eRNA_up		
motif id	factor	- 10*LOG(pval)
M00932	SP1	342.723
M00931	SP1	335.991
M00196	SP1	304.745
M00255	GC box	297.429
M00958	ABI4	254.201
M00982	Egr1	240.785
M01273	SP4	237.7
M00803	E2F1	213.235
M00189	AP-2	207.997
M01622	UGA3	200.453
M01199	TRIM28	200.382
M00933	SP1	195.873
M00807	ERG1	192.864
M01118	WT1	188.773
M01122	ZNF219	184.466
M01520	RSC30	177.751
M01175	ZBTB7B	169.432
M00244	EGR4	143.644
M00243	EGR1	143.359
M00246	EGR2	140.473
M01588	KLF4	138.616
MA0073	RREB1	135.543
M01518	Cha4	132.852
M00245	EGR3	129.238
M01057	ERF2	127.829
M01792	HMO1	126.081
M01783	SP2	126.027
MA0079	SP1	124.243

Dataset S1: file “Dataset S1.xls” contains complete lists of GO terms enriched in differentially transcribed genes at different cell cycle stages. Terms with FDR less than 0.01 are included for each comparison.

Supplemental methods

Cell synchronization

G0/G1: MCF-7 cells were cultured in white DMEM media supplemented with 10% (vol/vol) charcoal-stripped FBS for 72 hours.

G1/S: At 25-30% confluence, MCF-7 cells were wash twice with 1xPBS and cultured in red DMEM media supplemented with 10% (vol/vol) FBS, 1% Pen-Strep, 1% Glutamine and 2mM thymidine (Sigma-Aldrich, St. Louis, MO) for 18 hours (first block). After the first thymidine block, MCF-7 cells were with 1xPBS to remove thymidine and then cultured in fresh red DMEM supplemented with 10% (vol/vol) FBS, 1% Pen-Strep and 1% Glutamine for 9hours. After releasing, perform second round of thymidine block for 17 hours.

M phase: At 40% confluence, MCF-7 cells were cultured in red DMEM supplemented with 10% (vol/vol) FBS, 1% Pen-Strep, 1% Glutamine and 2mM thymidine for 24 hours (S-phase block). After thymidine block, wash with 1xPBS and culture in fresh red DMEM supplemented with 10% (vol/vol) FBS, 1% Pen-Strep and 1% Glutamine for 3 hours. After releasing, add 200ng/ml nocodazole (Sigma Aldrich) to the media for 12 hours (mitotic block). After mitotic block, gently shake off and collect the non-adhesion cells.

RNA-seq

Total RNA was extracted from synchronized MCF-7 cells with TRIZOL REAGENT (Life Technologies, 15596026) and fragmentized with sonication. Libraries were then constructed with Encore Complete RNA-Seq DR Multiplex System 1-8 (Nugen Technologies Inc, 0333-32) and sequenced to 50bp with Illumina HiSeq machine.

ChIP-seq

Chromatin from synchronized MCF-7 cells was digested with Nuclease micrococcal from Staphylococcus aureus (Sigma Aldrich, N3755-50UN) and dialyzed with Slide-A-Lyzer Dialysis Cassettes (Thermo Fisher Scientific, PI66380). Dynabeads Protein A (Life Technologies, 10002D) and G (Life Technologies, 10004D) were used for immunoprecipitation with antibodies against H3K4me2 (EMD Millipore, 07-030) and H3K27ac (Abcam, ab4729). ChIP-seq libraries were constructed with ThruPLEX DNA-seq kit (Rubicon) and sequenced to 50bp with Illumina HiSeq machine.

DNase-seq

DNase hypersensitivity mapping was performed as previously described (1). Nuclei from asynchronous MCF-7 cells were digested with DNase I (Roche Diagnostics, 11873580001) and 50 to 100 bp DNA fragments were gel purified. DNase-seq libraries were constructed with ThruPLEX DNA-seq kit (Rubicon) and sequenced to 50bp with Illumina HiSeq machine.

siRNA transfection

siRNAs targeting KLF4 and control siRNA were purchased from GE Dharmacon. Lipofectamine RNAiMAX transfection reagent (ThermoFisher, 13778150) was used for siRNA transfection following the manufacturer's instructions.

Nuclear run-on assay

Nuclei from 5×10^6 MCF-7 cells were isolated, run-on-transcribed with Br-UTP and other NTPs (Thermo Fisher Scientific, FERR0481), and base-hydrolyzed to yield nascent RNAs with an average size around 100nt. RNA fragmentation reagent (Life Technologies, AM8740) was used to fragment RNA. Br-UTP-incorporated nascent transcripts were purified with BrdU antibody (Santa Cruz Biotechnology; sc-32323). Micro Bio-Spin 30 Columns (Bio-Rad Laboratories, 7326250) was used for RNA clean up.

GRO-seq data analysis

GRO-seq reads were mapped to the human genome (hg19) using Bowtie (2) with default parameters. Replicates were highly correlated and were pooled for further analysis. Normalization between datasets was performed with Drosophila S2 spike-in cells as previously introduced (3). The hg19 RefSeq gene list was used for all transcription level analysis. Htseq (v0.5.4p3) was used to get gene level read counts number from Bowtie mapped bam files. The resultant gene read count table was subjected to DESeq2 (v1.8.2) for differential gene analysis and a cutoff of 0.01 for FDR was chosen to identify significant differential genes. For the visualization of GRO-seq data, we generated bedGraph files using genomeCoverageBed function in the BEDTools suite(4) with signal scaled to reads per 10 million for both plus and minus strands. The bedGraph files for replicates were pooled together before converting to bigwig format using the bedGraphToBigWig function in BLAT suite (5).

ChIP-seq and DNase-seq data analysis

ChIP-seq and DNase-seq reads were mapped to human genome (hg19) using Bowtie. MACS2 was used for peak calling with the parameter “--SPMR” on. Resultant bedgraph files were converted to big wiggle files with bedGraphToBigWig function.

RNA-seq data analysis

RNA-seq reads were aligned to human genome (hg19) with tophat (v2.1.1). GFOLD (v1.1.3) (6) was used to generate gene level read count and carry out differential expression analysis. Genes with GFOLD value not equal to zero were deemed differential.

Gene Ontology Analysis

Gene ontology analyses were performed using David (<http://david.abcc.ncifcrf.gov/home.jsp>; (7)). All expressed genes were used as a background for GO analysis. Terms with FDR less than 0.01 were considered significantly enriched. A complete list of terms enriched can be found in Dataset S1.

BETA analysis

Binding and Expression Target Analysis (BETA) analysis was performed between eRNAs and differentially transcribed or expressed genes separately. BETA analysis takes into account the distance between the eRNAs and genes nearby, the number of eRNAs within certain distance, and the differential transcription level of genes identified to calculate the correlation between eRNAs and nearby genes. GRO-seq data was used to identify differentially transcribed genes while differentially expressed genes were identified with RNA-seq data. Red lines represent genes that are up-regulated while purple lines represent genes that are down-regulated, respectively. Dashed lines represent non-differential background genes. P value indicates the significance of difference in each gene group compared to the background.

Curation of cell cycle regulated genes

High confidence cell cycle reference gene sets were created using publicly available cell cycle expression profiling datasets (8-12). We identified 259 genes as cell cycle genes in at least three of the five datasets (13). Among them, 104 genes are with peak expression at G2/M.

References for supporting materials:

1. He HH, *et al.* (2014) Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat Methods* 11(1):73-78.
2. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.
3. Loven J, *et al.* (2012) Revisiting global gene expression analysis. *Cell* 151(3):476-482.
4. Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.
5. Kent WJ, Zweig AS, Barber G, Hinrichs AS, & Karolchik D (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 26(17):2204-2207.
6. Feng J, *et al.* (2012) GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* 28(21):2782-2788.
7. Huang da W, Sherman BT, & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44-57.
8. Whitfield ML, *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular Biology of the Cell* 13(6):1977-2000.
9. Bar-Joseph Z, *et al.* (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* 105(3):955-960.
10. Sadasivam S, Duan S, & DeCaprio JA (2012) The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes Dev* 26(5):474-489.
11. Pena-Diaz J, *et al.* (2013) Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication. *Nucleic Acids Research* 41(5):2846-2856.
12. Grant GD, *et al.* (2013) Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Molecular Biology of the Cell* 24(23):3634-3650.
13. Fischer M, Grossmann P, Padi M, & DeCaprio JA (2016) Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic acids research*.