Supplementary Information for

A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition

June 16, 2016

Chris C.-S. Hsiung, Caroline Bartman, Peng Huang, Paul Ginart, Aaron J. Stonestrom, Cheryl A. Keller, Carolyne Face, Kristen S. Jahn, Perry Evans, Laavanya Sankaranarayanan, Belinda Giardine, Ross C. Hardison, Arjun Raj, Gerd A. Blobel

Supplemental Results

Supplemental Experimental Procedures

Supplemental Results

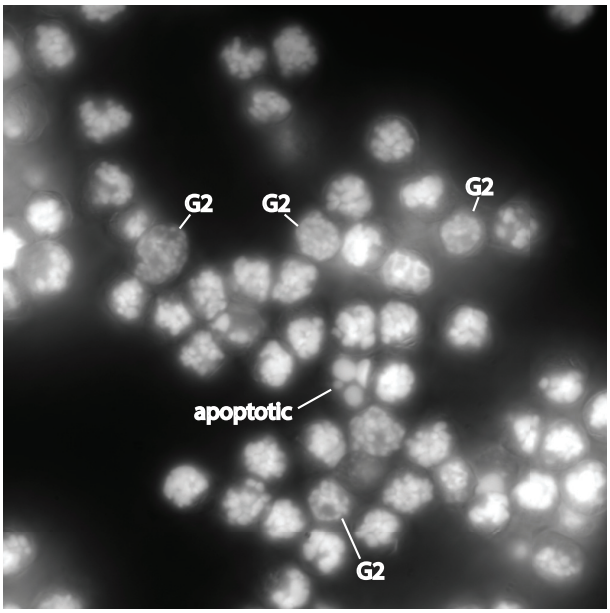


Figure S1 – Related to Fig. 1A: Estimate of interphase contamination in 0min sample. Cells were arrested with nocodazole and sorted for 4N, YFP-MD-high as described in Fig. 1A for the 0min time point. Shown is a representative field of cells in the DAPI channel by wide-field microscopy (single optical plane is shown). There is approximately 10% G2 contamination (27 out of 271 cells), as judged by microscopic appearance of DAPI staining, and minimal apoptotic cells.

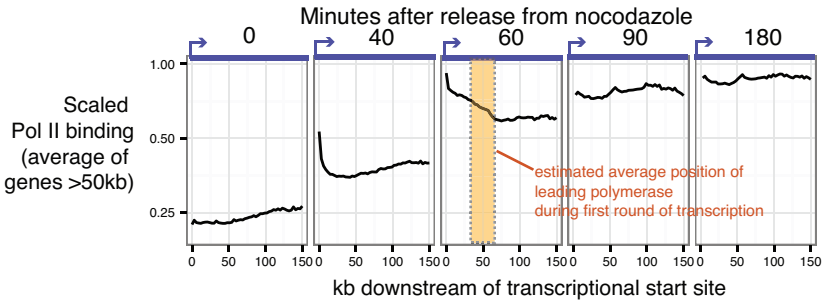
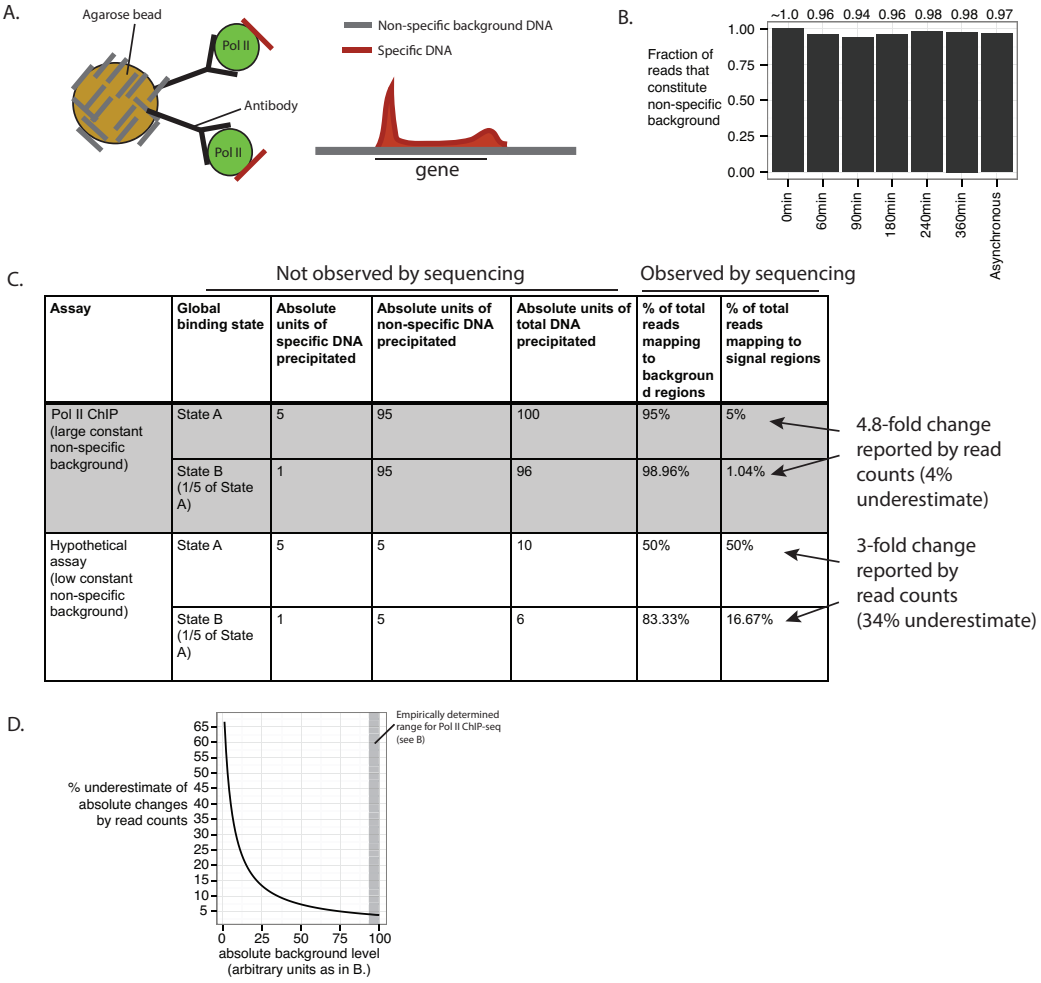


Figure S2 – Related to Fig. 1B: Genome-wide average position of Pol II leading edge. Average Pol II ChIP-seq read densities (scaled to the read density in the asynchronous sample of each gene) obtained from 3kb bins across all genes >50kb is plotted in the sense direction along genomic coordinates relative to the transcriptional start site. The estimated averaged position of the Pol II leading edge at the 1h time point, determined manually by eye, is indicated by the shaded region.



C.

			Not observed by sequencing		Observed by sequencing	
Assay	**Global binding state**	**Absolute units of specific DNA precipitated**	**Absolute units of non-specific DNA precipitated**	**Absolute units of total DNA precipitated**	**% of total reads mapping to background regions**	**% of total reads mapping to signal regions**
Pol II ChIP (large constant non-specific background)	State A	5	95	100	95%	5%
	State B (1/5 of State A)	1	95	96	98.96%	1.04%
Hypothetical assay (low constant non-specific background)	State A	5	5	10	50%	50%
	State B (1/5 of State A)	1	5	6	83.33%	16.67%

4.8-fold change reported by read counts (4% underestimate)

3-fold change reported by read counts (34% underestimate)

Figure S3 – Related to Fig. 2: Pol II ChIP-seq signal in genes is internally normalized to reads mapping to non-specific background, enabling interpretations of absolute binding. **A)** Model of the physical origins of Pol II ChIP-seq reads that map to genic (regions defined by Refseq annotations), versus intergenic regions (outside of Refseq annotations). Reads mapping to intergenic background regions likely arise from DNA bound non-specifically to surfaces of reagents used to pull-down the antibody. We assume that the amount of non-specifically bound DNA is constant across samples processed identically. **B)** The fraction of total mapped reads that constitute intergenic non-specific background empirically determined for each Pol II ChIP-seq library is >94%. Note that the slight dip at 60min and 90min time points is consistent with the overall spike in Pol II binding within genes described in main text. **C)** We illustrate the effect of the level of non-specific background on the ability of a sequencing-based assay to estimate absolute changes in signal. We consider a biological phenomenon that causes a global 5-fold change in Pol II binding from State A to State B, and assess the performance of our Pol II ChIP assay in reporting this change, compared to a hypothetical assay in which absolute level of non-specific background is 19 fold less. This analysis shows that our Pol II ChIP-seq would underestimate the true fold change by only 4%, whereas the hypothetical low-background assay would underestimate the fold change by 34%. **D)** For the same scenario as in C, we illustrate the relationship between error in estimating absolute changes by library size-normalized read counts with constant absolute background inherent to the assay.

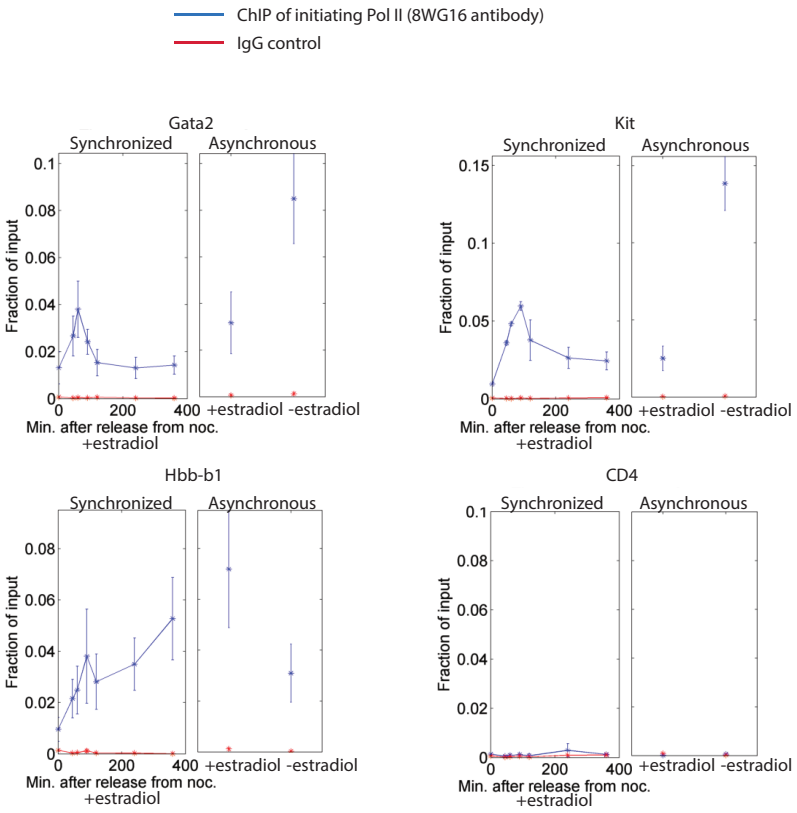


Figure S4 – Related to Fig. 2B: ChIP-qPCR of initiating Pol II confirms ChIP-seq patterns at individual loci. We performed ChIP using an antibody (8WG16) specific for the initiating form of Pol II in estradiol-induced G1E GATA1-ER cells, followed by qPCR of amplicons proximal to the transcriptional start site in a nocodazole arrest-release time course without additional FACS purification. Also shown are ChIP performed in asynchronous controls with and without estradiol treatment. Gata2 and Kit are known to be down-regulated by estradiol induction, Hbb-b1 is known to be up-regulated by estradiol induction, and CD4 is a silent gene that serves as a negative control. Error bars denote SEM (n=3-4)

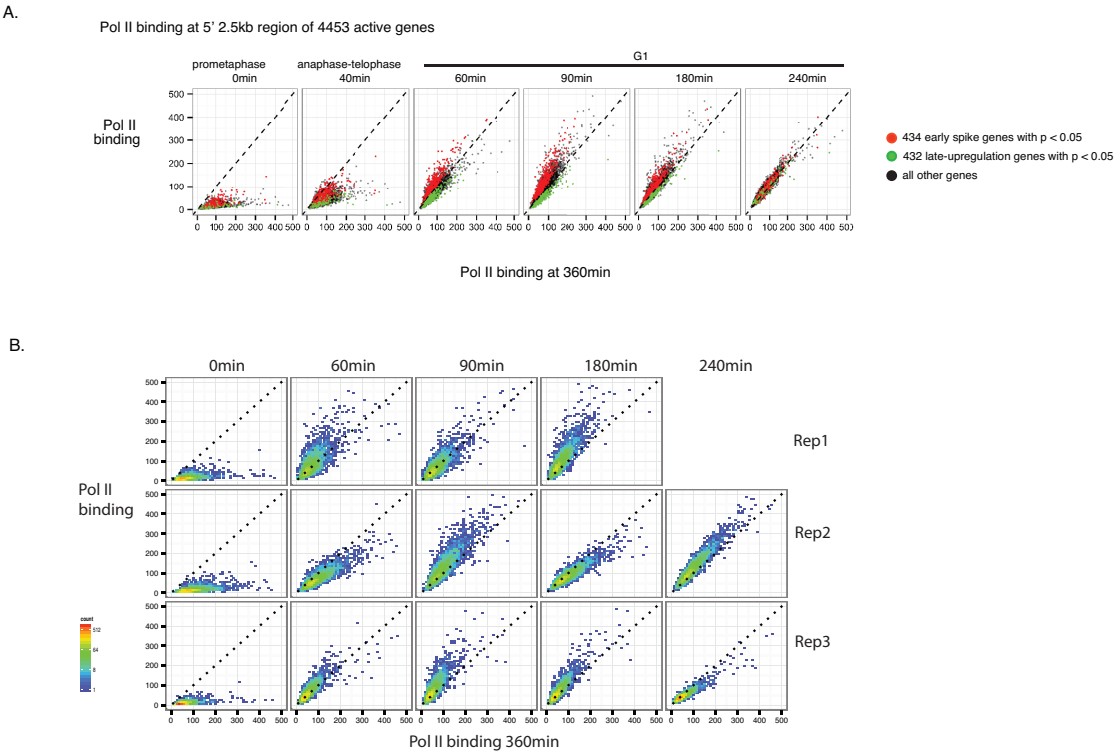
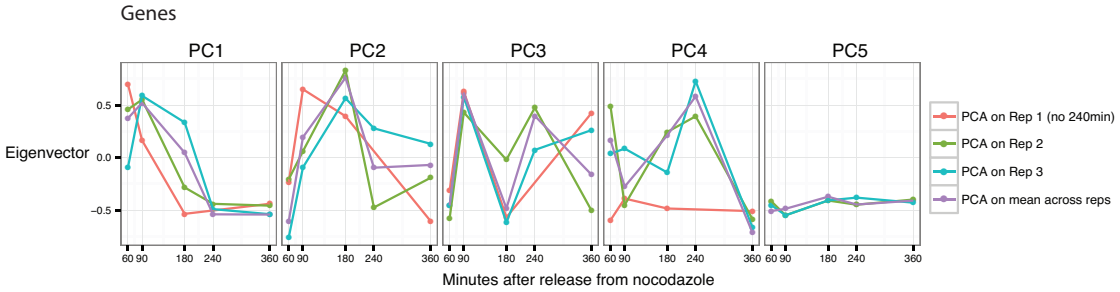


Figure S5 – Related to Fig. 2: Statistical significance of degree of match to the first principal component for individual genes and genome-wide Pol II binding patterns for each biological replicate. A) A bootstrapping approach was used to assess statistical significance of the degree of match to the first principal component for each gene. An empirical null distribution was obtained by scrambling the time points for each gene 10000 times without replacement and projecting the scrambled data onto the the first principal component eigenvector. Genes with $p < 0.05$ based on a one-sided test are highlighted in the graphs showing Pol II binding at each time point plotted against that at the 360min time point. **B)** Pol II binding at the 5' 2.5kb region of 4309 genes active in at least one time point were plotted for each time point against the 360min time point for each of the biological replicates.

A.



B.

Percentage of total variance accounted for

	PC1	PC2	PC3	PC4	PC5
Mean	47.2%	27.9%	13.9%	0.11%	0.001%
Rep 1	42.3%	34.4%	23.3%	<0. 1%	N/A
Rep 2	43.3%	22.4%	17.9%	16.4%	<0. 1%
Rep 3	41.5%	28.4%	15.5%	14.5%	0.1%

C.

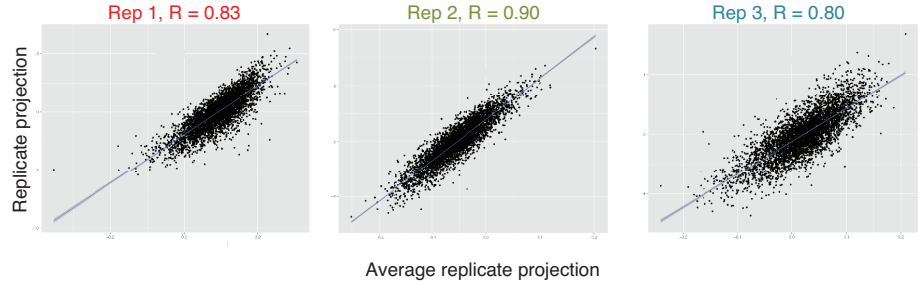


Figure S6 – Related to Fig. 2B: Principal component analysis for Pol II binding at genes. **A)** We performed principal component analysis on each of the three biological replicates using the 5' 2.5kb region of 4309 genes deemed active in at least one of the time points. Only G1 time points (60min, 90min, 180min, 240min, 360min) were used for principal component analysis. Shown are the principal components found for each replicate. Principal components found by using the mean Pol II binding across all replicates are also shown (same first principal component as shown in Fig. 2B). Replicate 1 does not have the 240min time point, so has only 4 principal components. **B)** The percentages of total variance accounted for by each of the replicate-derived and mean-derived principal components are shown. **C)** We evaluated replicate concordance by plotting the projection of the RPKM for each replicate onto the first principal component derived from the mean RPKM across all replicates on the y-axis, versus the projection of the mean RPKM across all replicates onto the first principal component derived from the mean RPKM across all replicates on the x-axis. Pearson correlation coefficients are shown.

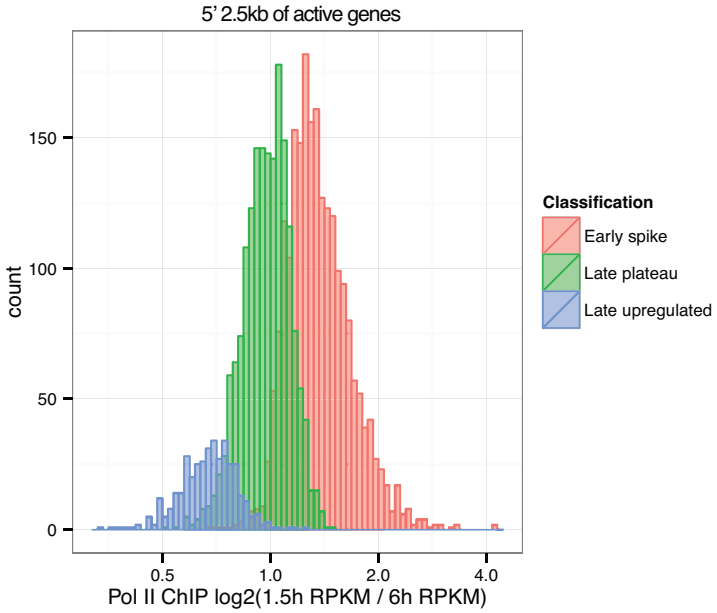


Figure S7 – Related to Fig. 2B: Magnitudes of transcriptional changes in G1. Pol II binding is quantified as ratio of RPKM from 1.5h time point to the 6h time point, and shown as histograms separately for the early spike, late plateau, and late upregulated classes defined in Fig. 2B.

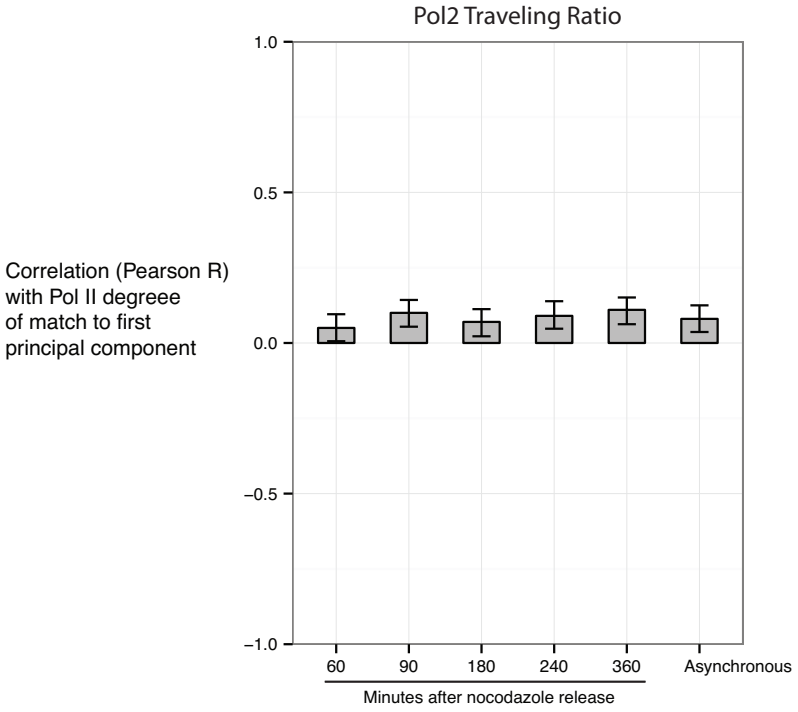


Figure S8 – Related to Fig. 2: Mitosis-G1 transcriptional spike is not associated with changes in Pol II promoter escape. Pol II degree of match with (projection onto) the first principal component is plotted against a traveling ratio of Pol II (a measure of the rate of Pol II promoter escape) at each time point. We modified the definition of the traveling ratio (Rahl et al., 2010) to be the Pol II ChIP-seq read density in the -30bp to +300bp region relative to the TSS, divided by the read density in the +1.3kb to +5kb region. This definition avoids confounding from incomplete elongation in the early time points after mitosis. The analysis was performed only on genes $>$5kb.

Top 200 early spike genes

Category	Subcategory	expected	observed	p-value (bon)
KEGG	p53 signaling pathway	1.53268	7	0.0127075
KEGG	Neuroactive ligand-receptor interaction	0.821078	5	0.0207315

Top 200 late up-regulation genes

Category	Subcategory	expected	observed	p-value (bon)
Gene Ontology	plasma membrane	17.6023	43	1.10481e-05
Gene Ontology	contractile fiber	0.586742	6	0.0088973
Gene Ontology	membrane part	41.433	66	0.0171488
Gene Ontology	plasma membrane part	7.31171	20	0.026343
Pfam domains	PDZ domain (Also known as DHR or GLGF)	1.14013	7	0.00216449
Pfam domains	LIM domain	0.68408	5	0.0096066
Pfam domains	Protein tyrosine kinase	0.820896	5	0.0245133
Pfam domains	Calcium-binding EGF domain	0.273632	3	0.042072

Figure S9 – Related to Fig. 2B: Analysis of gene sets enriched among early spike vs. late up-regulation genes. Using GeneTrail (Backes et al., 2007; Keller et al., 2008), we tested for gene sets that are overrepresented among the top 200 early spike genes or the top 200 late up-regulation genes, using the 4309 active genes as background. GeneTrail includes tests for the enrichments of KEGG pathways, Gene Ontology and Pfam domains. Shown are the overrepresented gene sets that are statistically significant (Bonferroni-corrected p = 0.05).

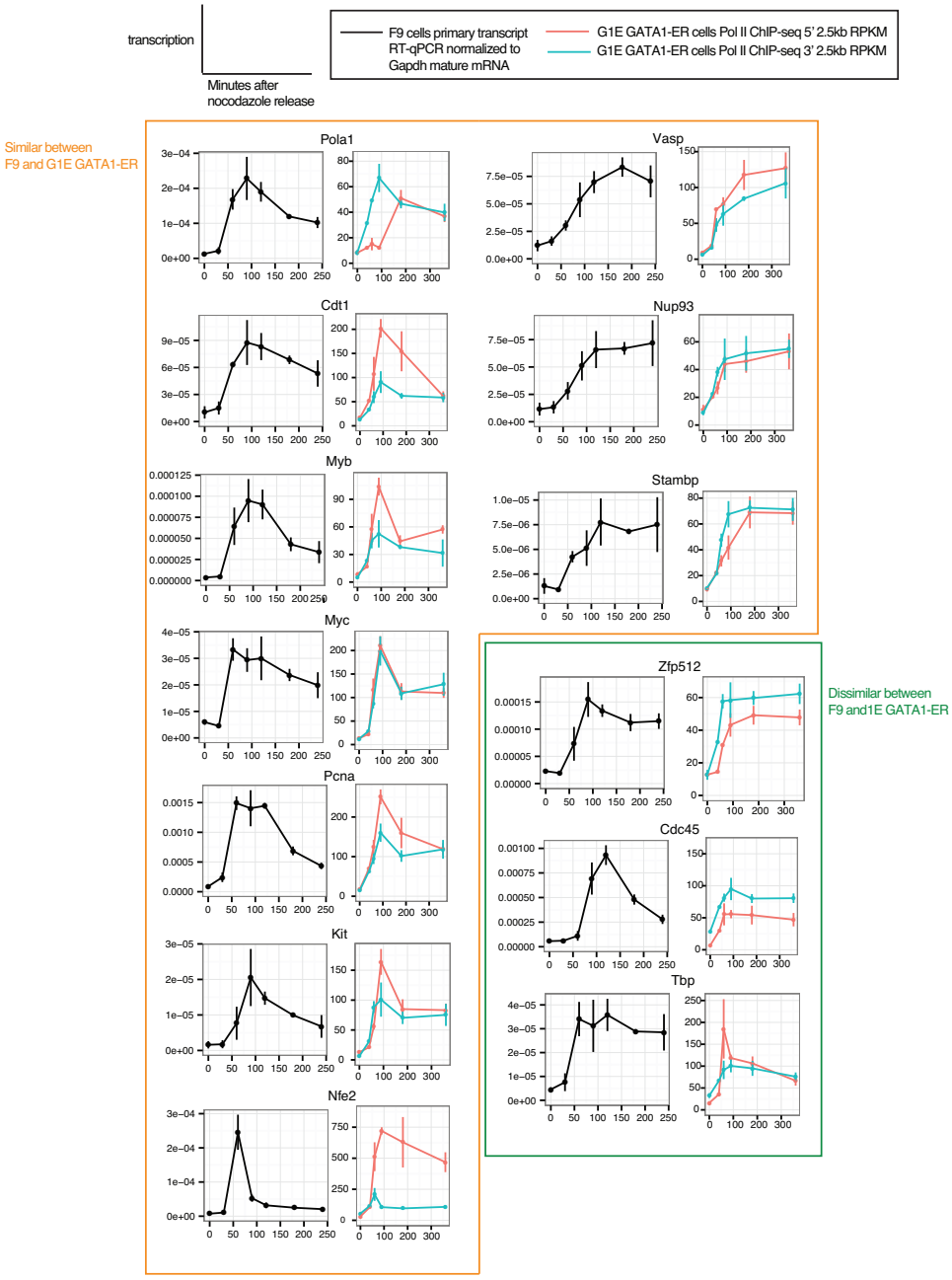
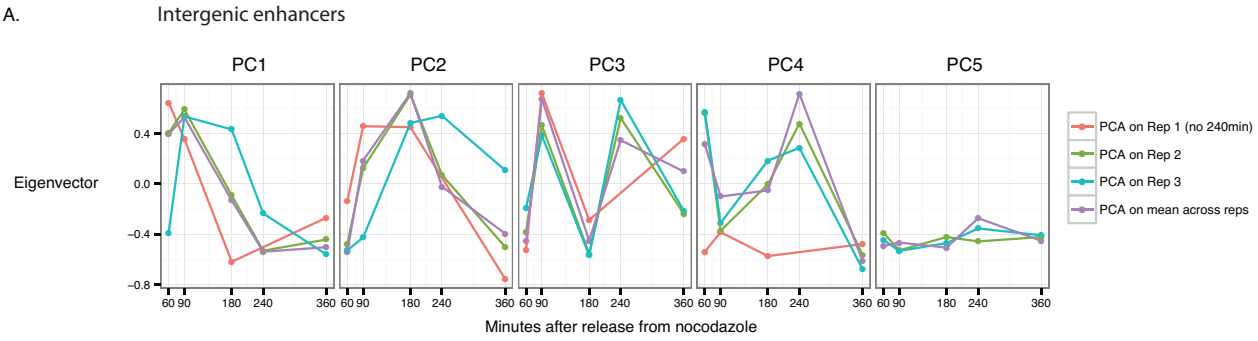


Figure S10 – Comparison of G1 transcriptional patterns between developmentally distinct murine cell types. Shown are nocodazole arrest-release transcriptional profiles for an embryonic carcinoma cell line (F9) and erythroid cell line (G1E GATA1-ER). Genes expressed in both cell lines were examined by primary transcript RT-qPCR using primers flanking intron-exon junctions for F9 cells, and Pol II ChIP-seq for G1E GATA1-ER as described in the main text. Several of the G1E GATA1-ER Pol II ChIP-seq plots are duplicated from Fig. 1 and Fig. 3 for ease of comparison. Genes showing similar or dissimilar G1 transcriptional profiles are highlighted. Error bars denote SEM (n = 4 for F9; n = 3 for G1E GATA1-ER).

A.



B.

Percentage of total variance accounted for

	PC1	PC2	PC3	PC4	PC5
Mean	42.4%	31.1%	15.2%	11.3%	<0.1%
Rep 1	42.4%	32.9%	24.7%	<0.1%	N/A
Rep 2	36.3%	25.9%	21.5%	16.3%	<0.1%
Rep 3	35.9%	30%	17.8	16.4%	<0.1%

C.

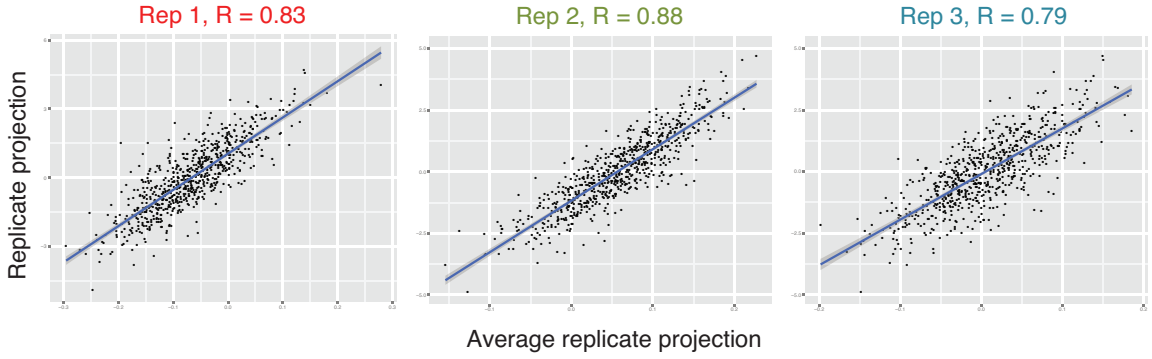


Figure S11 – Fig. 3A: Principal component analysis for Pol II binding at intergenic enhancers. Principal components analysis was performed in for 809 intergenic enhancers in the same manner as described for genes in Fig. S6, and the information shown in A-C are exactly analogous to that in Fig. S6.

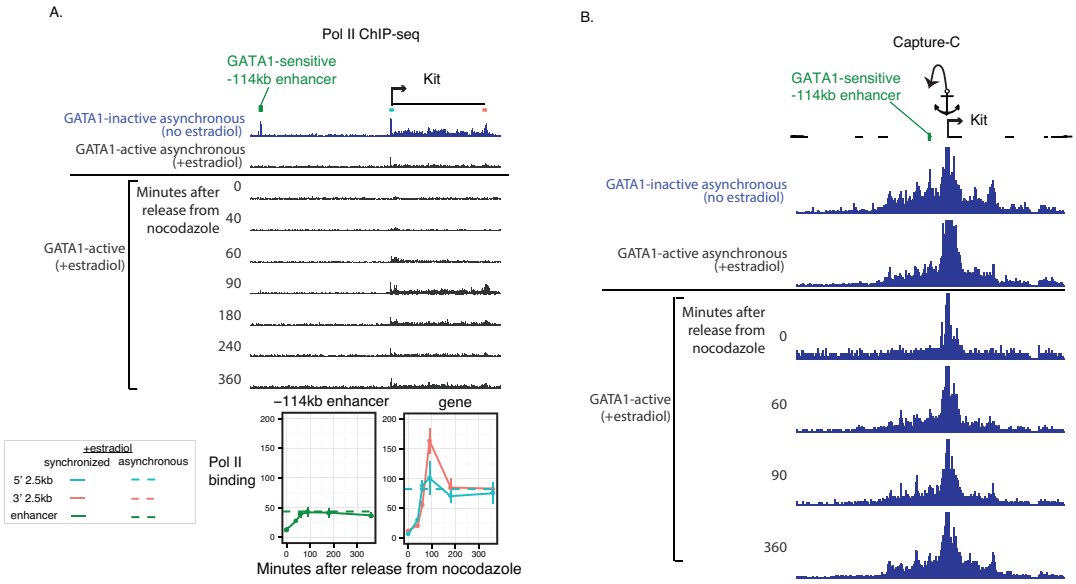
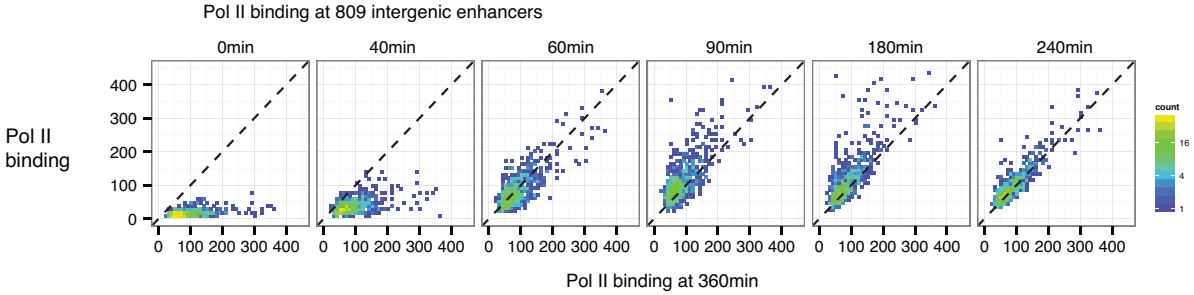
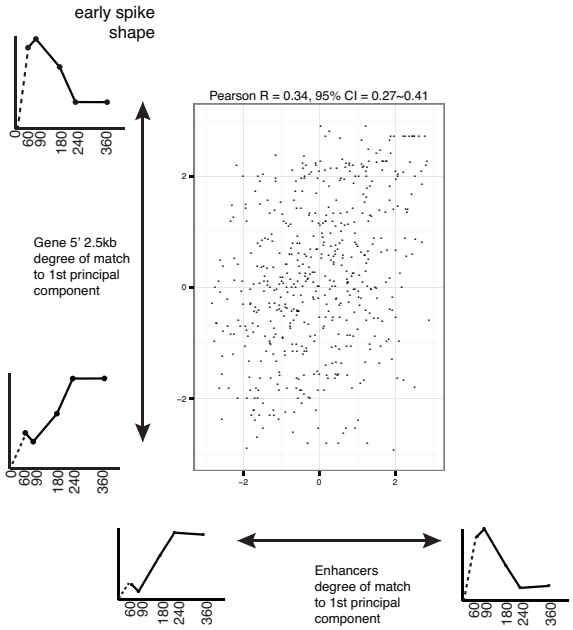


Figure S12 – Related to Fig. 3: Mitosis-G1 spike in Kit gene transcription occurs in the absence of spike in Pol II binding or enhancer-promoter looping. A) Browser track views of Pol II binding at the Kit locus, including the enhancer located at -114kb upstream of the gene. Tracks are shown for GATA1-inactive condition (no estradiol treatment), in which there is high Pol II binding to the gene and enhancer, and for GATA1-active conditions (13h treatment with estradiol), in which there is basal level of Pol II binding at the gene and very little at the enhancer. A time course of release from nocodazole arrest is shown in the GATA1-active condition. Quantification of Pol II binding at the -114kb enhancer and the 5' 2.5kb and 3' 2.5kb regions of the gene are shown below. **B)** Browser track views of Capture-C signal performed with anchor at the Kit promoter, under conditions as described in A). Quantification of Capture-C enhancer-promoter contacts are shown in Fig. 3D.

A.



B.



C.

Gene ontology enrichment of genes neighboring top 200 early spike enhancers

ID	Description	Bonferroni-corrected P-value	Fold Enrichment
GO:0006955	immune response	7.13E-08	3.62E
GO:0045087	innate immune response	1.13E-07	5.01
GO:0002252	immune effector process	7.54E-06	3.64
GO:0006952	defense response	2.94E-05	3.24

Figure S13 – Related to Fig. 3: Global patterns of Pol II binding at intergenic enhancers and their nearest gene. A) Pol II binding at 809 intergenic enhancers at each time point is plotted against binding at the 360min time point. **B)** Each enhancer is assigned to the nearest gene. The degree of match to (projection onto) the 1st principal component for enhancers (as described in Fig. 3A) is plotted against the degree of match to (projection onto) the 1st principal component of the nearest gene (as described in Fig 2B). The Pearson correlation coefficient is shown. **C)** We used GREAT 3.0.0 McLean et al. (2010) to assess Gene Ontology enrichments among neighboring genes of the top 200 early spike intergenic enhancers, using the set of 809 intergenic enhancers as background. Shown are enriched ontologies with at least 6 genes associated with the target regions and have $p < 0.05$ after Bonferroni correction.

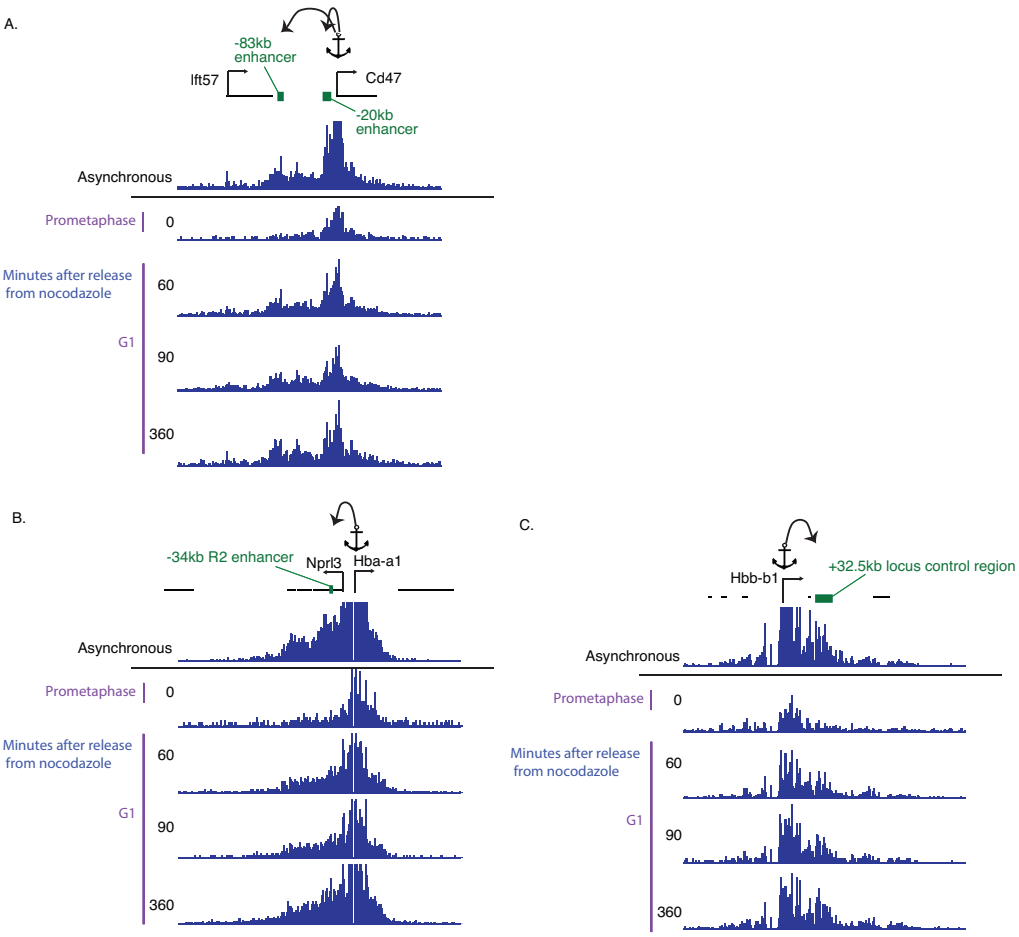


Figure S14 – Related to Fig. 3: Browser tracks for enhancer-promoter contacts at the mitosis-G1 transition measured by Capture-C. Capture-C was performed using anchors targeting the **A)** Cd47 promoter, **B)** Hba-a1 promoter, and **C)** Hbb-b1 promoter, as described in Fig. 3C. The corresponding quantification of read densities at individual enhancers are shown in Fig. 3C.

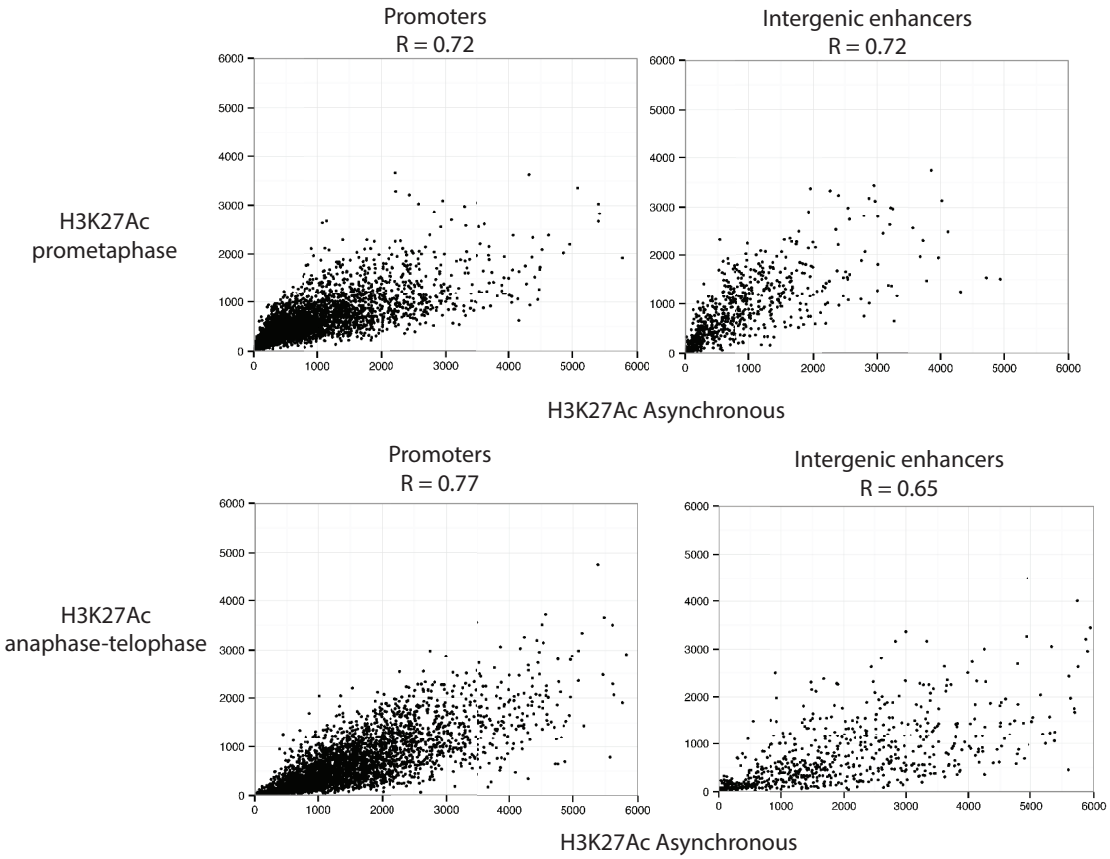


Figure S15 – Related to Fig. 4: Locus-specific differences in H3K27Ac in mitosis and interphase. Top row: Scatter plots of H3K27Ac ChIP-seq for cells arrested by nocodazole in prometaphase (FACS purified for MPM2-positivity) vs. asynchronous (FACS purified for MPM2-negativitiy). Bottom row: Scatter plots of H3K27Ac ChIP-seq for cells in the anaphase-telophase compartment (40min after release from nocodazole arrest, sorted for YFP-MD low and 4N DNA content by DAPI) vs. matched asynchronous sample. Each data point corresponds to a DNase hotspot annotated as promoters or intergenic enhancers. Values shown are library size-normalized read densities (RPKM). Pearson correlation coefficients are shown at the top of each graph.

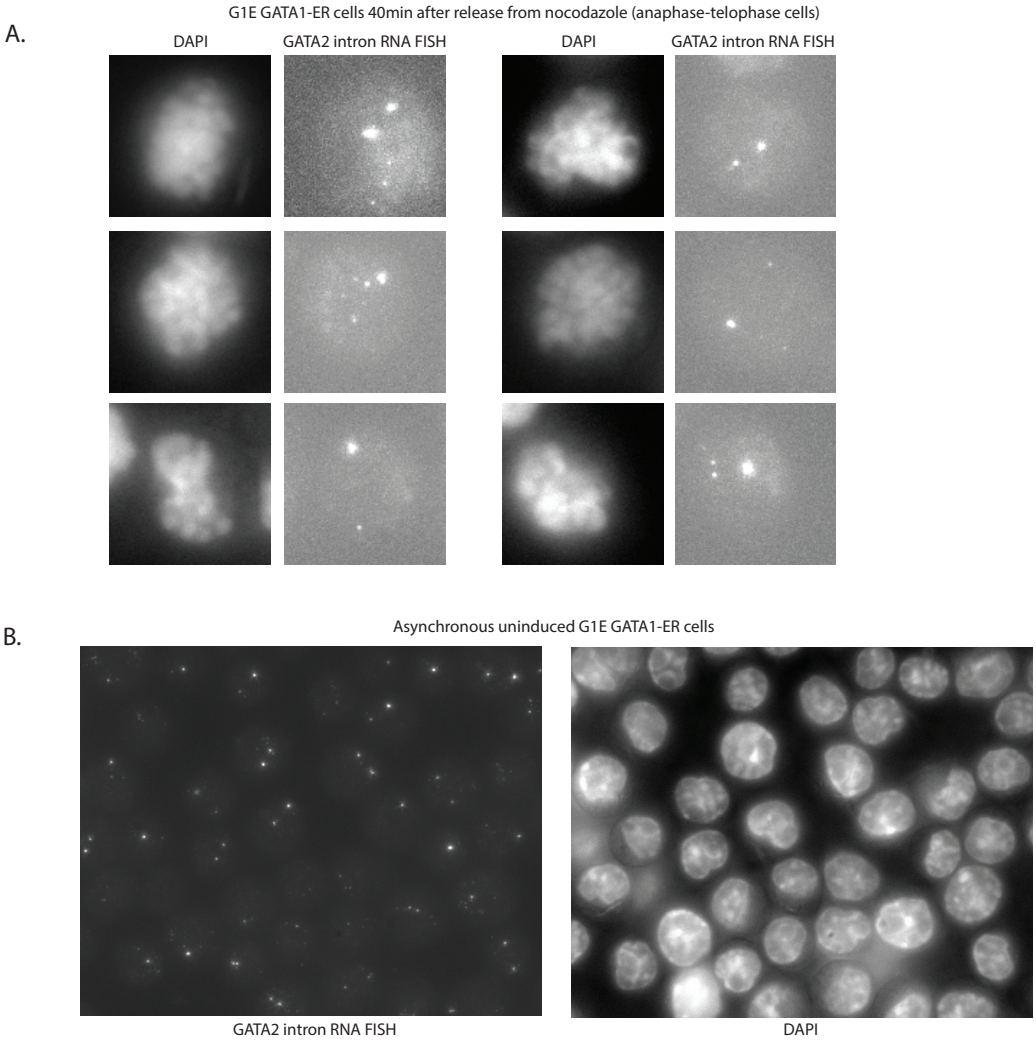


Figure S16 – **A)** Additional images of anaphase-telophase G1E GATA1-ER cells (estradiol induced, 40min after release from nocodazole) with grossly condensed chromosomes and active transcription detected by single-molecule RNA FISH for GATA2 introns. B) Images of an asynchronous population of uninduced G1E GATA1-ER cells obtained from RNA FISH for GATA2 introns and DAPI stain. GATA2 is highly transcribed n this condition.

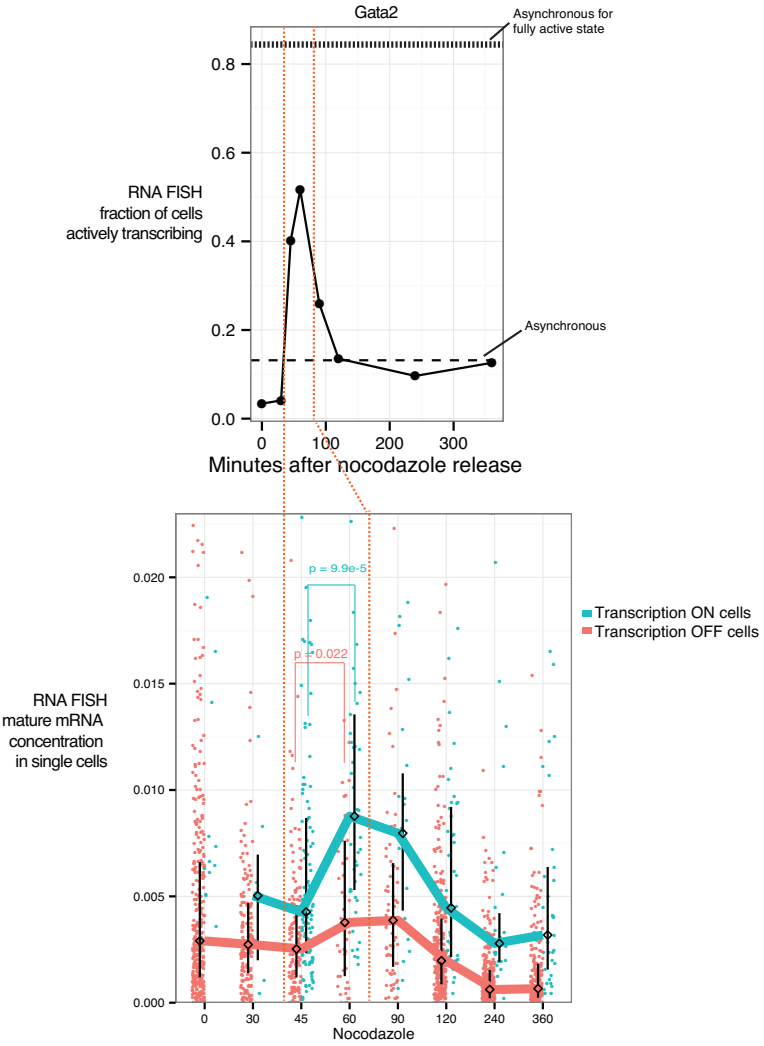


Figure S17 – Related to Fig. 5: Additional biological replicates for primary and mature mRNA FISH. We simultaneously imaged Gata2 primary and mature mRNAs by probing for exons and introns in cells synchronized with nocodazole. Shown are data pooled from two biological replicates (combined 89-355 cells for each time point) performed similarly as in Fig. 5. For these replicates, cells were synchronized by nocodazole in the absence of subsequent cell sorting purification, so the 45min-90min time points in particular represent a mixed population of 4N and 2N cells (anaphase to G1). P-values from one-sided Wilcoxon test are shown.

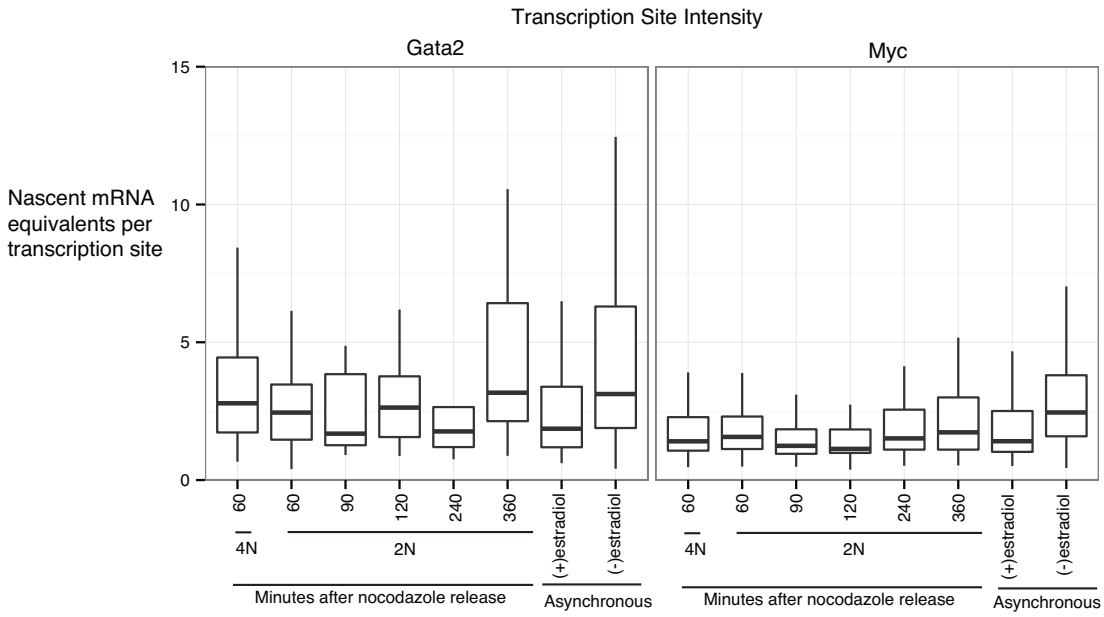


Figure S18 – Related to Fig. 5: Transcription site intensities are relatively unchanged with G1 progression. Boxplots of transcription site intensities are shown. Transcription sites are defined as spots where intron and exon spots colocalize, and the intensities are quantified from the exon channel and expressed as equivalents of the average intensity of single mature mRNA molecules.

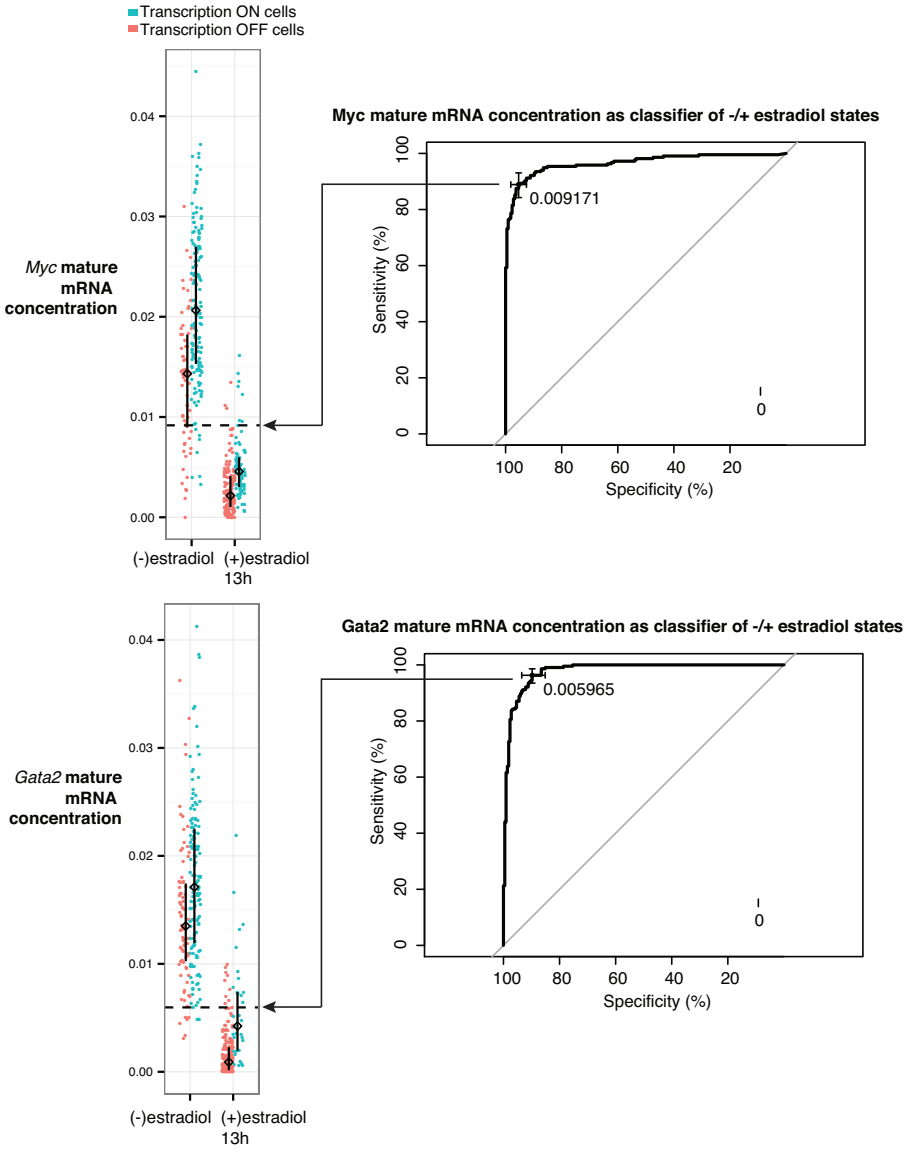


Figure S19 – Related to Fig. 5D: Receiver operating characteristics curves for Gata2 and Myc mature mRNA concentrations as classifier for cells in the absence vs. presence of estradiol. Gata2 and Myc mature mRNA concentrations were quantified by RNA FISH for transcriptionally ”on” and ”off” cells in an asynchronously dividing population of G1E GATA1-ER cells in the absence (-) and presence (+) of estradiol for 13h. The optimal threshold of Gata2 and Myc mature mRNA concentration for discriminating the -/+ estradiol populations was determined by the receiver operating characteristics curves shown above. These the indicated are labeled as “fully active state threshold” in Fig. 5D.

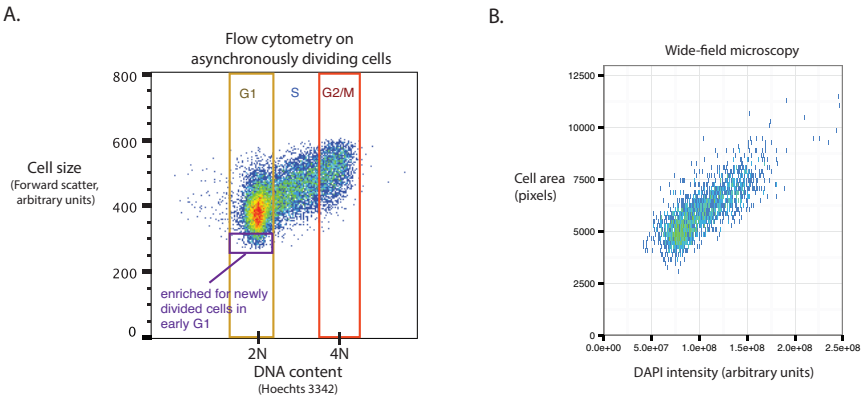


Figure S20 – Related to Fig. 5: Cell size is proportional to DNA content in G1E GATA1-ER cells. A) Asynchronously dividing G1E GATA1-ER cells were stained with Hoechts 3342 (DNA content) and subjected to flow cytometry. For each cell, forward scatter (reflection of cell size) is plotted against Hoechts 3342, and the boundaries of cell cycle phases are indicate based on DNA content. Within G1, we reason that the smallest cells, highlighted in purple box, are enriched for the newly divided cells in early G1. **B)** We measured cell area (based on manual segmentation of cell boundaries) versus DAPI intensity (sum over single optical plane) by wide-field microscopy, and use their proportionality to estimate thresholds for cell cycle phases in Fig. 6. We further confirmed these cell area thresholds work as approximations of DNA content by arrest-release with nocodazole (analysis not shown).

Supplemental Experimental Procedures

As described in the Data Access section, we provide scripts that reproduce the majority of figures starting from processed data at an online repository, in addition to raw and processed sequencing data available at GSE83293. We summarize our experimental protocols and computational analyses here.

1 Cell culture, cell cycle synchronization and cell sorting

G1E cells were previously derived through deletion of GATA1 in mouse embryonic stem cells, followed by in vitro differentiation (Weiss et al., 1997). We cultured a subline of G1E cells, G1E-ER4, in which GATA1-ER was retrovirally transduced (referred to in main text as "G1E GATA1-ER"), as previously described (Weiss et al., 1997). We retrovirally transduced G1E-ER4 cells with the YFP-MD construct and sorted for a pool of stably YFP-positive cells. Except where indicated in the text as uninduced, we induced cell to mature with 100nM estradiol to activate GATA1-ER. During estradiol induction, we simultaneously treated cells with nocodazole for 7h-13h, washed once, and replated into fresh medium lacking nocodazole for varying times (40min-360min), ensuring all samples are exposed to estradiol for the same duration of 13h.

For ChIP, we harvested cells by resuspension in PBS 2mM EDTA and fixed with 1% formaldehyde at room temperature for 10 minutes with constant mixing, then quenched with 1M glycine, stained with DAPI, and sorted on a BD FACSAria based on YFP-MD and DAPI signal. Sorted fixed cell pellets were snap-frozen with liquid nitrogen and stored at -80 until ready for further processing.

2 Chromatin immunoprecipitation (ChIP)

2.1 Reagent preparation

ChIP-seq of total Pol II was performed in 3 biological replicates using N-20 (Santa Cruz, cat# sc899). For ChIP-qPCR of initiating form of Pol II, we used 8WG16 (Covance, cat# MMS-126R). For H3K27Ac ChIP-seq, we used anti-H3K27Ac from Active-Motif, cat# 39685. For ChIP was performed as follows on approximately 7-20 million cells for each sample. Protease inhibitor (P8340, Sigma) was added to the following buffers right before use: Cell Lysis Buffer (10mM Tris pH 8, 10mM NaCl, 0.2% NP-40/Igpal), Nuclear Lysis Buffer (50mM Tris pH 8, 10mM EDTA, 1% SDS) , and IP Dilution Buffer (20mM Tris pH 8, 2mM EDTA, 150mM NaCl, 1% Triton X-100, 0.01% SDS).

Agarose beads slurry was prepared by mixing Protein A (Invitrogen 15918014) and Protein G (Invitrogen 15920010) agarose beads at 1:1 ratio, washed with PBS 3 times, and the mixed beads resuspended in 1:1 volume in PBS (volumes indicated below are for the slurry in PBS, but PBS is removed by centrifugation and aspiration prior to bead use). For use in "pre-clearing" step, 50 µgrabbit IgG was mixed with 50 µL Protein A and G mixed agarose beads slurry. For use for immunoprecipitation step, 70 µL bead slurry was mixed with antibody (20 µg for N20, 10 µg for 8WG1), and 10.2 µg for

anti-H3K27Ac for each immunoprecipitation and incubated for >8h at 4°C to allow binding ("pre-bound").

Additional buffers were prepared as follows: IP Wash Buffer 1 (20mM Tris pH 8, 2mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS), High Salt Buffer (20 mM Tris pH 8, 2mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.01% SDS), IP Wash Buffer 2 (10 mM Tris pH 8, 1mM EDTA, 0.25 M LiCl, 1% NP-40/Igepal, 1% Na-deoxycholate), Elution Buffer (100 mM NaHCO3, 1% SDS).

2.2 Protocol

All steps performed on ice or at 4°C unless otherwise noted. Formaldehyde-fixed cell pellets stored in -80C were thawed on ice, resuspended in 1ml Cell Lysis Buffer and incubated for 20min on ice, then resuspended in 1ml Nuclear Lysis Buffer and incubated for 20min on ice, then diluted with 0.6 ml of IP Dilution Buffer. Samples sonicated at 4°C for 45 minutes using either the Bioruptor (Diagenode) or Epishear (Active Motif) (same machine was used within each biological replicate), then centrifuged at 21130gfor 10min to remove cellular debris. Supernatant was transferred to new tube and mixed with "pre-clear" beads and rotated at 4°C for >5h, then centrifuged at 821g to pellet beads. 200 µL of supernatant was saved as "input." The remaining supernatant was mixed with beads prebound with antibodies and rotated for >8h at 4°CḂeads were washed sequentially once with IP Wash Buffer 1, twice with High Salt Buffer 1, once with IP Wash Buffer 2, then twice with Tris-EDTA pH 8.0 (BP2473-1, Fisher Scientific) (with centrifugation at 5283g for 2min and aspiration of supernatant before resuspension with each subsequent buffer). Then, beads were pelleted by centrifugation and supernatant aspirated.

All steps from this point on performed at room temperature unless otherwise noted. Beads were resuspended in 100 µL Elution Buffer twice sequentially and the supernatant from both steps combined for a final eluate volume of 200 µL. 12 microliter of 5M NaCl and 2 µL of 10mg/ml RNase A (10109169001, BMB) were added to the 200 µL eluted samples and 200 µL input samples and incubated overnight at 65°CṪhen, 3 µL of 20mg/ml Proteinase K (3115879, BMB) was added and incubation at 65°C continued for an additional 2h. 10 µL of 3M sodium acetate pH 5 was added to each sample and DNA purified per the instructions of the QIAquick PCR Purification Kit (cat# 28106, Qiagen). Inputs were eluted in 133.4 µL Buffer EB (Qiagen), and immunoprecipitated samples in 60 µL Buffer EB. DNA was stored at -20°C until further processing.

3 ChIP-seq library preparation and Illumina sequencing

All samples, including input, were processed for library construction for Illumina sequencing using Illumina's TruSeq ChIP Sample Preparation Kit (Illumina cat# IP-202-1012). In brief, DNA fragments were repaired to generate blunt ends, purified using Agencourt AMPure XP Beads (Beckman Coulter cat# A63881), and a single "A" nucleotide was added to each end. Double-stranded Illumina adaptors were ligated to the fragments. Ligation products were purified using Agencourt AMPure XP Beads, and subject to size selection using SPRIselect Beads (Beckman Coulter cat# B23318)

in which both a left side size selection was performed at 0.9x volume, and a right side size selection was performed at 0.6x volume according to manufacturer's specifications. Library fragments were then amplified for 16 cycles of PCR and products were purified using Agencourt AMPure XP Beads. Constructed libraries were run on the Agilent Bioananlyzer 2100 (Agilent Technologies) using either the DNA 7500 kit (cat# 5067-1504) or the High Sensitivity DNA kit (cat# 5067-4626) as appropriate to determine the average size and confirm the absence of unligated adaptors. The mean library size is approximately 330 bp.

The ChIP-seq libraries were quantitated by qPCR using the Kapa SYBR FAST Universal kit (Kapa Biosystems) according to the Illumina's Sequencing Library qPCR Quantification Guide. Libraries were multiplexed and sequenced on the Illumina HiSeq 2000 using Illumina's kits and reagents as appropriate.

We sequenced 3 biological replicates for the 0min, 60min, 90min, 180min, and 360min time points; two biological replicates for 240min time point; and one replicate for 40min time point.

4 Bioinformatic analysis of ChIP-seq data

4.1 Read processing

Reads were mapped to mouse mm9 genome using Bowtie. Mapped reads were passed to MACS with a matched control (input) dataset for peak calling for producing bigwig files with reads shifted to account for fragment size. We filtered peaks which had an overlap of at least one base pair with blacklisted regions described in either Pimkin et al. or mm9 blacklisted regions identified by ENCODE. The ENCODE blacklisted regions were obtained from `https://sites.google.com/site/anshulkundaje/projects/blacklists`. Parameters for the above steps are further detailed in Supplemental File.

4.2 Generation of browser tracks

Bigwig files output by MACS were loaded into Integrated Genomics Viewer. Y-axes were adjusted to normalize for total number of reads mapped in each library.

4.3 Identification of active genes

Because Pol II binding tends to show the greatest enrichment near the 5' and 3' ends of genes, we focused on quantification within the 5' 2.5kb (500bp upstream to 2kb downstream relative to Refseq transcriptional start site) and 3' 2.5kb (500bp upstream and 2kb downstream relative to Refseq transcriptional end site) regions. If the 5' or 3' 2.5kb region of a gene overlapped at least one Pol II peak called by MACS in at least one sample (arrest-release and asynchronous samples with estradiol induction), then we deemed the gene active. From this set of active genes, we removed those whose boundaries (after 500bp extension upstream of transcriptional start site, and 2kb extension downstream of transcriptional end site) overlapped with another active gene. This step is meant to avoid mis-assignment of Pol II signal to the wrong gene. After

these steps we arrive at 4309 active and non-overlapping genes that are the subject of Fig. 2 and used for subsequent steps outlined here.

4.4 Identification of intergenic enhancers

We previously predicted distal cis-regulatory modules, or likely enhancers, in G1E ER4 cells based largely on presence of DNase sensitivity coincide with relative absence of H3K4me3 and that do not overlap known transcriptinoal start sites based on Refseq annotation (Hsiung et al., 2014). Among these predicted enhancers, we additionally filtered for intergenic enhancers, defined as those that are $>$3kb away from 5' end, and $>$20kb away from the 3' end of Refseq gene boundaries and overlap at least one Pol II peak called by MACS in at least one sample (arrest-release and asynchronous samples with estradiol induction). This step is intended to avoid mis-assignment of Pol II signal from binding at genes to enhancers. After these steps we arrive at 809 intergenic enhancers that are the subject of Fig. 3 and used for subsequent steps outlined here.

4.5 Quantitation of Pol II binding

We used bigWigAverageOverBed to count mapped reads and calculated read count per kilobase per million mapped reads in library (RPKM). Changes in RPKM reflect absolute changes in Pol II binding very well even in the context of global changes in Pol II binding Fig. S3, so we use RPKM for all subsequent analyses. For individual replicates and for the mean across replicates, we calculated RPKM at the 5' and 3' 2.5kb regions of genes and intergenic enhancers and used them for all analyses that refer to "Pol II binding" in the main text and figures.

4.6 Principal component analysis

We performed principal component analysis on Pol II binding at the 5' 2.5kb regions of genes, and separately intergenic enhancers, as follows. We confined the principal component analysis to G1 time points (60min, 90min, 180min, 240min, and 360min) in order to focus the analysis on the part of the time course with the patterns of interest. We normalized RPKM of 5' 2.5kb regions of genes by the sum of RPKM across G1 time points. On these normalized values, we performed principal component analysis using the R package prcomp, with variables scaled to have unit variance prior to analysis (scale. = TRUE). Principal components ("rotation" of the output from prcomp) are plotted against the time points. Projection onto principal components ("x" of the output from prcomp) are referred to as the "degree of match to principal component" in text and figures. The above was performed for individual replicates and for the mean RPKM across replicates. For assessment of replicate concordance in Fig. S6 and Fig. S11, RPKM from individual replicates were projected onto the principal components derived from mean RPKM.

Heatmaps in Fig. 2B and Fig. 3A were generated using the heatmap.2 function from the R package gplots. Rows are median-normalized. The thresholds for separating "early spike" from "late plateau" in Fig. 2B and Fig. 3A were chosen based on

the inflection of projection onto the first principal component from positive to negative. The thresholds for separating "late plateau" and "late up-regulated" were chosen manually based on the appearance of the heatmaps.

4.7 Analysis of enriched gene sets

We used GeneTrail (Backes et al., 2007; Keller et al., 2008) to query gene sets from the Gene Ontology, KEGG, and Pfam databases for overrepresentation among the extreme 200 genes based on their projection onto the first principal component for 5' 2.5kb regions of genes (top and bottom of heatmaps in Fig. 2B). The background is the set of all 4309 active, non-overlapping genes.

4.8 Analysis of chromatin features for association with early G1 transcriptional spike

Below describes how we obtained the Pearson correlation coefficients shown in Fig. 4.

For ChIP-seq signals of H3K27Ac, H3K4me3, H4K4me1, H3K27me3, and H3K9me3, we quantified the read densities of those features within known DNase "hotspots", which are $>$250bp regions of enriched DNase sensitivity previously categorized as a promoter or distal enhancer hotspot (Hsiung et al., 2014). For 4309 such gene-promoter hotspot pairs, we calculated the Pearson correlation coefficient between the log2 read densities of each histone modification within the promoter DNase hotspot with that of Pol II ChIP-seq projection onto the first principal component for the associated 5' 2.5kb gene region. For intergenic enhancers, we calculated the Pearson correlation coefficient between the read densities of each histone modification within the intergenic enhancer DNase hotspot with that of Pol II ChIP-seq projection onto the first principal component for those intergenic enhancers.

For DNase-seq signal in mitotic and asynchronous populations, we quantified the read densities of those features within known DNase "peaks," which are 150bp regions of DNase hypersensitivity (contained within DNase hotspots), previously categorized as residing at promoters or distal enhancers (Hsiung et al., 2014). We calculated the Pearson correlation coefficient between the read densities of each of these features within the promoter DNase peak with that of Pol II ChIP-seq projection onto the first principal component for the associated 5' 2.5kb gene region. We chose to base these analyses on DNase peaks, rather than hotspots, because they tend to be more dynamic between mitosis and interphase for promoters (**?**). For intergenic enhancers, we calculated the Pearson correlation coefficient between the read densities of each feature within the intergenic enhancer DNase hotspot with that of Pol II ChIP-seq projection onto the first principal component for these intergenic enhancers.

For GATA1 ChIP-seq signal in mitotic and asynchronous populations, we obtained the union of previously defined GATA1 binding peaks in mitosis and asynchronous cells (Kadauke et al., 2012). We quantified the mitotic or asynchronous GATA1 ChIP-seq read densities within these GATA1 peaks. We assigned each of the 5' 2.5kb regions of the 4309 active genes to the nearest GATA1 peak (must be $<$150kb away). Where there are multiple GATA1 peaks overlapping or equidistant from a given 5' 2.5kb region, we took the average of read densities across those multiple GATA1 peaks. We

calculated the Pearson correlation coefficient between the GATA1 ChIP-seq read densities and the projection onto the Pol II first principal component for each 5' 2.5kb gene region. The same procedure was applied to intergenic enhancers, except we required that the GATA1 peaks must overlap the intergenic enhancer.

We also performed a similar analysis centered on GATA1 ChIP-seq signal within all DNase peaks (rather than GATA1 binding peaks) that overlap with either the 5' 2.5kb gene regions or the intergenic enhancer regions, thereby allowing for the correlation coefficient to also include DNase-sensitive regions that do not meet the threshold for calling GATA1 binding peaks. We obtained essentially the same results as in Fig. 4 (not shown).

5 Image analysis

We manually segmented boundaries of cells from brightfield images and localized RNA spots using custom software written in MATLAB (Raj and Tyagi, 2010), with subsequent analyses performed in R. The area within segmentation borders is used for cell area. For Fig. 6, we adjusted for minor systematic variations in the distributions of cell area found across imaging sessions by adding a constant to the cell area, such that the median across all biological replicates are equal. Mature mRNA concentrations per cell are quantified by the spot counts in the exon channel, divided by the cell area.

We took two approaches to quantify co-localization of intron exon spots: 1) manual inspection of each spot by eye to identify relatively bright spots that show co-localization and are thus likely transcription sites, and 2) automated co-localization defined as Gaussian-fitted spots less than 2 pixels apart to identify all likely primary transcripts. The manual inspection approach is more selective for the spots marking nascent transcripts emanating from transcriptionally active gene loci, excluding some co-localized spots of dimmer intensity in the single molecule range that correspond to primary transcripts that have diffused away from the transcription site. On the other hand, the automated approach is easier to apply to a large number of spots and is more precisely defined. Since the presence of primary transcripts that have diffused away from transcription sites indicates recent transcription, the two approaches indeed yield very similar results for Fig. 5 and Fig. S17, so we show the results using the manual inspection approach. For Fig. 6, we used the automated approach because it is easier to implement on a large number of cells. In this case we sum the intensities of primary transcript spot in the exon channel for each cell, expressed as equivalents of the average intensity of a single-molecule mature mRNA spot in the exon channel in a field of cells. Primary transcript equivalents are shown both directly and as values normalized to estimated DNA copy number in Fig. 6.

6 Live-cell imaging

Uninduced G1E GATA1-ER cells expressing mCherry-MD were grown in media and adhered to imaging surface with poly-L-lysine. Microscopy was performed as described for RNA FISH except as follows. Cells were maintained at 37°Cand 5% CO2 during imaging. We used a 60x oil objective and imaged at 500ms exposure every 5

minutes at a single optical plane for approximately 13 hours using the 41004 Texas Red filter (Chroma) and in the bright-field channel.

7 Plotting and graphics

We used R (R Core Team, 2013) packages reshape2 (Wickham, 2007), dplyr (Wickham and Francois, 2015), and ggplot2 (Wickham, 2009) to produce nearly all figures, followed by cosmetic adjustments in Adobe Illustrator.

References

Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y. A., Müller, R., Meese, E. and Lenhof, H.-P. (2007). GeneTrail–advanced gene set enrichment analysis. Nucleic Acids Res. *35*, W186–92.

Hsiung, C. C.-S., Morrissey, C. S., Udugama, M., Frank, C. L., Keller, C. A., Baek, S., Giardine, B., Crawford, G. E., Sung, M.-H., Hardison, R. C. and Blobel, G. A. (2014). Genome accessibility is widely preserved and locally modulated during mitosis. Genome Res. *25*, gr.180646.114–225.

Kadauke, S., Udugama, M. I., Pawlicki, J. M., Achtman, J. C., Jain, D. P., Cheng, Y., Hardison, R. C. and Blobel, G. A. (2012). Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. Cell *150*, 725–737.

Keller, A., Backes, C., Al-Awadhi, M., Gerasch, A., Küntzer, J., Kohlbacher, O., Kaufmann, M. and Lenhof, H.-P. (2008). GeneTrailExpress: a web-based pipeline for the statistical evaluation of microarray experiments. BMC Bioinformatics *9*, 552.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, nbt.1630–9.

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria.

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.

Raj, A. and Tyagi, S. (2010). Single Molecule Tools: Fluorescence Based Approaches, Part A, vol. 472,. Elsevier Inc.

Weiss, M. J., Yu, C. and Orkin, S. H. (1997). Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. Mol. Cell. Biol. *17*, 1642–1651.

Wickham, H. (2007). Reshaping Data with the reshape Package. Journal of Statistical Software *21*, 1–20.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer New York.

Wickham, H. and Francois, R. (2015). dplyr: A Grammar of Data Manipulation. R package version 0.4.1.