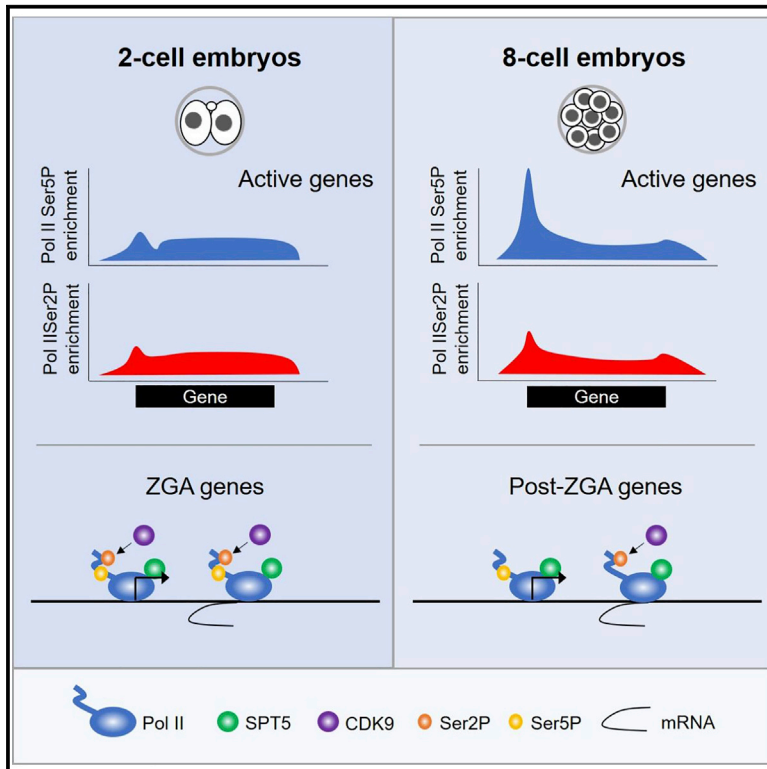


Distinct patterns of RNA polymerase II and transcriptional elongation characterize mammalian genome activation

Graphical abstract



Authors

Kenichiro Abe, Tamas Schauer,
Maria-Elena Torres-Padilla

Correspondence

torres-padilla@helmholtz-muenchen.de

In brief

Abe et al. performed ChIL-seq to map Ser2 and Ser5 phosphorylated RNA Pol II forms in mouse embryos before, during, and after major ZGA. Functional perturbations and RNA-seq demonstrate that CDK9 and SPT5 are major regulators of ZGA in mice.

Highlights

- Maps of Ser2 and Ser5 phosphorylated RNA Pol II during and after ZGA in mouse embryos
- Distribution of phosphorylated RNA Pol II forms is similar at ZGA and resolves after ZGA
- Serine 2 phosphorylation in 2-cell embryos precedes gene activation at later stages
- CDK9 and SPT5 are key regulators of global transcription during ZGA



Article

Distinct patterns of RNA polymerase II and transcriptional elongation characterize mammalian genome activation

Kenichiro Abe,^{1,4} Tamas Schauer,^{1,2,4} and Maria-Elena Torres-Padilla^{1,3,5,*}
¹Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, 81377 München, Germany

²Bioinformatics Unit, Biomedical Center, Ludwig-Maximilians-University, 82152 Planegg-Martinsried, Germany

³Faculty of Biology, Ludwig-Maximilians-Universität, München, Germany

⁴These authors contributed equally

⁵Lead contact

*Correspondence: torres-padilla@helmholtz-muenchen.de
<https://doi.org/10.1016/j.celrep.2022.111865>

SUMMARY

How transcription is regulated as development commences is fundamental to understand how the transcriptionally silent mature gametes are reprogrammed. The embryonic genome is activated for the first time during zygotic genome activation (ZGA). How RNA polymerase II (Pol II) and productive elongation are regulated during this process remains elusive. Here, we generate genome-wide maps of Serine 5 and Serine 2-phosphorylated Pol II during and after ZGA in mouse embryos. We find that both phosphorylated Pol II forms display similar distributions across genes during ZGA, with typical elongation enrichment of Pol II emerging after ZGA. Serine 2-phosphorylated Pol II occurs at genes prior to their activation, suggesting that Serine 2 phosphorylation may prime gene expression. Functional perturbations demonstrate that CDK9 and SPT5 are major ZGA regulators and that SPT5 prevents precocious activation of some genes. Overall, our work sheds molecular insights into transcriptional regulation at the beginning of mammalian development.

INTRODUCTION

The maternal-to-zygotic transition (MZT) is an essential process enabling the formation of a new organism following fertilization. The MZT involves the reprogramming of the gametes endowing embryonic cells with the capacity to generate all cells in our body. Molecularly, MZT is accomplished by the degradation of maternal transcripts inherited from the oocyte and the activation of the zygotic gene expression program. In mammals, the oocyte accumulates maternal transcripts synthesized during oocyte growth, but transcriptional activity ceases when the oocyte reaches its full size.^{1–3} After fertilization, the embryo activates its own genome, a process known as zygotic genome activation (ZGA), which in mice occurs in two waves. The first wave initiates in the zygote, is known as minor ZGA, and is indispensable for the second wave of transcription, which occurs at the late 2-cell stage and is referred to as major ZGA.^{4–7} Most maternal mRNAs are degraded by the end of major ZGA and replaced with embryonic transcripts. Whether and how these initial ZGA waves influence later transcriptional programs is unknown.

In mammals, ZGA has distinctive molecular properties, including, for example, a robust transcriptional activity of repetitive elements. Among them, MERVL is transcribed during minor ZGA^{8,9} and becomes fully activated during major ZGA.¹⁰ Likewise, LINE-1 elements are transcribed during both ZGA waves,

and their transcription regulates chromatin accessibility.^{11–13} In addition, intergenic regions are robustly transcribed during minor ZGA and splicing, and 3' processing was suggested to be inefficient in mouse zygotes.¹⁴ Interestingly, down-regulation or chemical inhibition of splicing factors in mouse embryonic stem (ES) cells induces 2-cell-like cells, which display similar expression pattern to 2-cell embryos.^{15,16} This suggests that specific features of RNA Polymerase II (Pol II) activity may be important for the developmental program that underlies cellular plasticity and totipotency.

Pol II's largest subunit RPB1 possesses a C-terminal repeat domain, CTD, which can be highly phosphorylated at different positions.¹⁷ Specific CTD residues are important for transcriptional regulation, for example, Y1 mutation leads to defective transcriptional termination¹⁸ and an S5A mutation of the yeast CTD leads to impaired promoter escape of Pol II due to Mediator retention at promoters.¹⁹ During a transcription cycle, non-phosphorylated Pol II is recruited to the promoter and together with the general transcription factors assembles into the pre-initiation complex.^{17,20,21} Subsequently, the CTD is phosphorylated on Serine 5 (Ser5), which induces promoter clearance and transcription initiation from the transcriptional start site (TSS).²² Productive transcription of mRNA is further regulated during elongation.²¹ Ser5 phosphorylated Pol II (Pol II Ser5P) undergoes promoter proximal pausing ~25–50 nt downstream of the TSS,



which is induced by the DRB Sensitivity-Inducing Factor (DSIF) and the Negative Elongation Factor (NELF).^{21,23} Release from pausing is driven by the recruitment and activation of positive transcription elongation factor b (P-TEFb)-containing complexes, for example, the Super Elongation Complex.²¹ Thus, the transition into productive elongation is also highly regulated. Namely, transition into productive elongation is triggered by the action of the P-TEFb complex containing CDK9, which phosphorylates the SPT5 subunit of DSIF, the NELF-E subunit of NELF and the Ser 2 residue of the CTD (Pol II Ser2P).²³ While NELF phosphorylation leads to its ejection from chromatin, phosphorylated DSIF travels with Pol II during elongation.²² Globally, Pol II Ser5P is enriched at promoter proximal regions and TSS, reflecting initiation and pausing, whereas Pol II Ser2P is enriched across gene bodies during transcriptional elongation.²⁴

Genome-wide analysis of nascent RNA in mouse, human, and *Drosophila* cells indicated that accumulation of Pol II around promoter regions is an important regulatory step of eukaryotic gene expression.^{25–29} In mouse ES cells, knockout of the NELF subunit NELF-B, leads to a change in Pol II localization from the promoter to the gene body, to altered transcription of FGF/ERK signaling pathway genes and to compromised differentiation and proliferation.³⁰ Interestingly, upon acute depletion of NELF-C in human colon cancer cells, Pol II travels downstream for further ~50 nt but does not reach the gene body.³¹ NELF-E is also lost from gene promoters, but SPT5 peaks remained largely unaffected and Pol II Ser2P and Ser5P phosphorylation levels at promoters are maintained.³¹ These experiments suggest that NELF plays a role in early elongation, which is distinct to pause-release regulation.

Likewise, a mutation of the DSIF component SPT5 leads to a severe reduction of promoter proximal Pol II and compromises ES cell differentiation.³² SPT5 stabilizes paused Pol II at promoters and promotes Pol II processivity across the entire gene body.^{33,34} This implies that the control of Pol II binding to promoters is not the only regulatory mechanism underlying transcription but also promoter clearance and elongation in the gene body are key regulatory steps.

How Pol II is regulated at the beginning of development is not fully understood. In particular, the regulatory proteins and potential function of transcriptional elongation remain to be investigated. Using an antibody raised against a Ser5 phosphorylated CTD peptide, a recent study documented the genome-wide localization of Pol II in mouse pre-implantation embryos.³⁵ Pol II Ser5P was found to be prevalent at promoters of both inactive and active genes in early embryos, suggesting that Pol II Ser5P may be subject to pausing at the beginning of development. However, how elongation is controlled and whether the engaged, elongating Ser2 phosphorylated Pol II is subject to additional regulation in early embryos is unclear. Treatment of mouse zygotes prior to minor ZGA with the specific CDK9 inhibitor flavopiridol results in arrested development primarily at the 2-cell stage,³⁶ suggesting that full developmental competency relies on the integrity of the elongation machinery, yet the molecular foundations behind are unknown.

Here, we optimized Chromatin integration labeling (ChIL-seq),³⁷ which we applied to as little as 100 mouse cells and

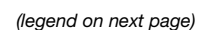
generated genome-wide maps of Pol II Ser5P and Pol II Ser2P during minor and major ZGA, as well as post-ZGA. In addition, we studied the function of elongation factors by performing acute depletion of NELF or DSIF proteins using Trim-Away to investigate their role in transcription elongation during ZGA. Our results indicate a need for specific NELF proteins and DSIF for transcription of ZGA genes. Accordingly, our results suggest that Ser2 phosphorylation is present during ZGA, but also that Pol II Ser2P at the 2-cell stage may act as a facilitator of embryonic transcription at later developmental stages.

RESULTS

Global profiling of phosphorylated Pol II

We established a modified version of ChIL-seq, which can be used to generate reliable genome-wide maps of phosphorylated Pol II with as little as 100 cells (Figure S1A and Table S1). Using this protocol, we recapitulated genome-wide profiles of Pol II Ser5P and Pol II Ser2P in ES cells obtained using a standard chromatin immunoprecipitation sequencing (ChIP-seq) protocol from 1 million ES cells (Figures S1B, S1C). ChIL-seq data showed overall high correlation with ChIP-seq profiles for Pol II Ser5P and Pol II Ser2P using either 100 or 1000 ES cells as starting material (Figure S1A). The distribution of Pol II Ser5P was strongly biased around the TSS, as expected, and similar to ChIP-seq profiles (Figures S1B, S1C). Likewise, Pol II Ser2P was enriched along the gene body of transcribed genes in ES cells and displayed the typical accumulation downstream of the transcription termination site (TTS), which was visible in both ChIL-seq and ChIP-seq datasets (Figure S1C). Pol II Ser2P also accumulated at the TSS with apparent high levels seen by ChIL (Figure S1C). However, this enrichment was lesser, relative to the gene body, compared with Pol II Ser5P (Figures S1B, S1C). The enrichment of Pol II Ser5P and Pol II Ser2P paralleled gene expression levels (Figure S1B, S1C). Thus, our ChIL-seq can be used to map Pol II with as little as 100 cells of bulk input material, a slight improvement compared with small-scale Tn5-assisted chromatin cleavage with sequencing (Stacc-seq) and Cut&Tag methods,³⁸ which have been used for 500 cells.³⁵

We generated genome-wide profiles of Pol II Ser5P and Pol II Ser2P in mouse embryos, as well as immunoglobulin (Ig)G controls, focusing on ZGA. We profiled zygotes at the time of minor ZGA, late 2-cell-stage embryos corresponding to major ZGA as well as 8-cell-stage embryos, a stage at which ZGA has concluded but the embryo still undergoes significant transcriptional activity.^{6,39,40} We performed biological triplicates for each stage, with the data showing similar overall Pol II profiles between replicates (Figure S1D). The IgG samples showed some enrichment at the TSS, albeit not reproducibly between replicates (Figure S1D), likely due to the Tn5 transposase in ChIL that has a bias toward accessible chromatin regions.³⁷ Globally, both Pol II Ser5P and Pol II Ser2P displayed lower enrichment around both the TSS and gene body in zygotes, compared with 2- and 8-cell embryos (Figures 1A–1C), presumably because transcriptional activity is lower during minor ZGA.⁵ Nonetheless, zygotes showed a visible enrichment of both Pol II Ser5P and Pol II Ser2P at the TSS, particularly on genes with



highest expression levels (Q5 and Q4) (Figure 1A). The enrichment of both phosphorylated forms of Pol II increased progressively from the 2- to the 8-cell stage (Figures 1A and 1B). As expected, highest expressed genes displayed the highest enrichment of Pol II Ser5P and Pol II Ser2P (Figure 1A). This is in line with recently published data for Pol II Ser5P using Stacc-seq (Figure S1E),³⁵ which also correlated well with our Pol II Ser5P datasets (Figure S1F).

Interestingly, we noted that Pol II Ser5P and Pol II Ser2P displayed very similar distribution patterns, particularly over the TSS, in the three stages analyzed. Namely, while Pol II Ser5P is known to display a strong enrichment at the TSS, a sharp enrichment at the TSS was visible for both phosphorylated Pol II forms and the shape of the normalized coverage was similar for both Pol II Ser5P and Pol II Ser2P (Figures 1A and 1B). This pattern was clearly visible in composite plots in which the coverage was scaled to enable qualitative comparisons between Pol II Ser5P and Pol II Ser2P (Figure 1D). While the absolute signal values of the two antibodies are not directly comparable, scaled plots using the maximum and minimal signal of each dataset enable comparison of their global distribution patterns across stages. The similarity between the Ser2P and the Ser5P forms of Pol II was most pronounced in zygotes (Figure 1D) and thus it may reflect a feature of Pol II in early embryonic chromatin. Interestingly, the distinctive distribution patterns between Pol II Ser5P and Pol II Ser2P emerged only from the 2-cell stage but were most evident at the 8-cell stage (Figure 1D), where Pol II Ser2P was more enriched downstream of the TTS compared to Pol II Ser5P (Figure 1D). These different patterns may not be related to levels of gene expression between developmental stages, since expression levels of genes at 2- and 8-cell stages are globally similar (Figure S2A).

Importantly, the Pol II Ser5P ChIL-seq signal at the 2-cell stage was sensitive to treatment with DRB, a potent inhibitor of transcription elongation that induces Pol II promoter proximal pausing.^{43,44} Upon DRB treatment, Pol II Ser5P sharply accumulated at the TSS and the Pol II Ser5P enrichment in the gene body was severely reduced (Figure 1E; compare to Figure 1A and see also below). This is unlikely to be due to changes in accessibility along the gene body that may result from transcriptional inhibition, since DRB treatment does not lead to changes in chromatin accessibility.^{45,46} Principal-component analysis (PCA) including

the DRB-treated samples revealed that the replicates were consistent between each other, and, most importantly that DRB-treated 2-cell-stage embryos clustered together with zygotes, rather than with untreated 2-cell embryos (Figure 1F), in agreement with previous reports^{4,35} and validating our ChIL-seq protocol.

To investigate potential differences between Pol II Ser5P and Pol II Ser2P, we classified the mapped reads into genomic features as intergenic, TSS, TTS, exons, and introns (Figure 1G). As expected, the fraction of mapped reads at TSS was greater in 2-cell-stage embryos treated with DRB than without DRB (Figure 1G). Globally, both Pol II Ser5P and Pol II Ser2P displayed similar distributions across the genomic features examined (Figure 1G). However, the proportion of Pol II reads at the TSS increased as development proceeds for Pol II Ser5P but not for Pol II Ser2P (Figures 1G and 1H), in agreement with our observations above. In addition, zygotes displayed a large fraction of reads from Pol II Ser5P in intergenic regions, which decreased at the 2-cell stage (Figures 1G and 1I). This is in line with previous RNA sequencing (RNA-seq) data showing that intergenic regions are pervasively transcribed in zygotes. Indeed, intergenic transcription is a well-known feature of mouse zygotes.¹⁴

Thus, our ChIL-seq data reveal distinctive features of Pol II Ser5P and Pol II Ser2P distribution in mouse embryos around ZGA, with the more typical elongation enrichment of Pol II Ser2P emerging only after ZGA.

The accumulation of Pol II around the TSS increases as development proceeds

To further understand how Pol II elongation is regulated, we next compared the Pol II enrichment at the TSS versus the gene body across the three embryonic stages, compared with ES cells. Metaplots using normalized coverage revealed that Pol II Ser5P is strongly enriched around the TSS and that this TSS enrichment increases progressively during development and is highest in ES cells (Figure 2A). This pattern may suggest that pausing increases as development proceeds. Our composite analysis also revealed that 2-cell-stage embryos show a distinctive Pol II Ser5P distribution, which displayed a sharp dip around the first ~2 nucleosomes downstream of the TSS (Figure 2A). This pattern was only observed in 2-cell-stage embryos and was particularly clear when we computed the scaled coverage on the same Pol II Ser5P metaplot (Figure 2B), but not that of

Figure 1. ChIL-seq maps of phosphorylated Pol II in mouse embryo

(A and B) Metaplots comparing Ser5P (left) and Ser2P (right) Pol II ChIL-seq mean coverage normalized by IgG centered at TSS (A) or at TTS (B) at indicated stages. Maternal RNA genes⁴¹ were removed and genes were grouped in quintiles based on their expression level at the corresponding stage.⁴² First quintile includes genes with zero expression. Average of three biological replicates is shown.

(C) Example genomic region of Pol II Ser5P ChIL-seq coverage in zygote, 2-cell, 2-cell treated with DRB, and 8-cell-stage embryos. Average of three biological replicates is shown.

(D) Scaled metaplots comparing Ser5P and Ser2P Pol II ChIL-seq mean coverage centered at TSS and TTS spanning –5 kb to +5 kb from each at indicated stages. Genes in the highest expression quintile are included. Average of three biological replicates is shown.

(E) Schematics of DRB treatment and metaplot showing Pol II Ser5P ChIL-seq mean coverage normalized by IgG centered at TSS in 2-cell-stage embryos treated with DRB. The same gene quintiles as in (A) were used. Average of three biological replicates is shown.

(F) Principal-component analysis on Pol II Ser5P ChIL-seq read counts at the gene body (TSS +500 bp – TTS –500 bp) in three biological replicates of embryos and ES cells. First two principal components are shown and percentage of the variance explained by each component is indicated.

(G) Fraction of read counts overlapping genomic annotation features comparing Pol II Ser5P (left) and Ser2P ChIL-seq (right) in embryos and ES cells. Counts were normalized by the total length of each feature. Bar plots with error bars show mean ± SEM of three biological replicates.

(H and I) Dot plots depicting the fraction of read counts at TSS (H) and intergenic regions (I) in three biological replicates (same values as in G).

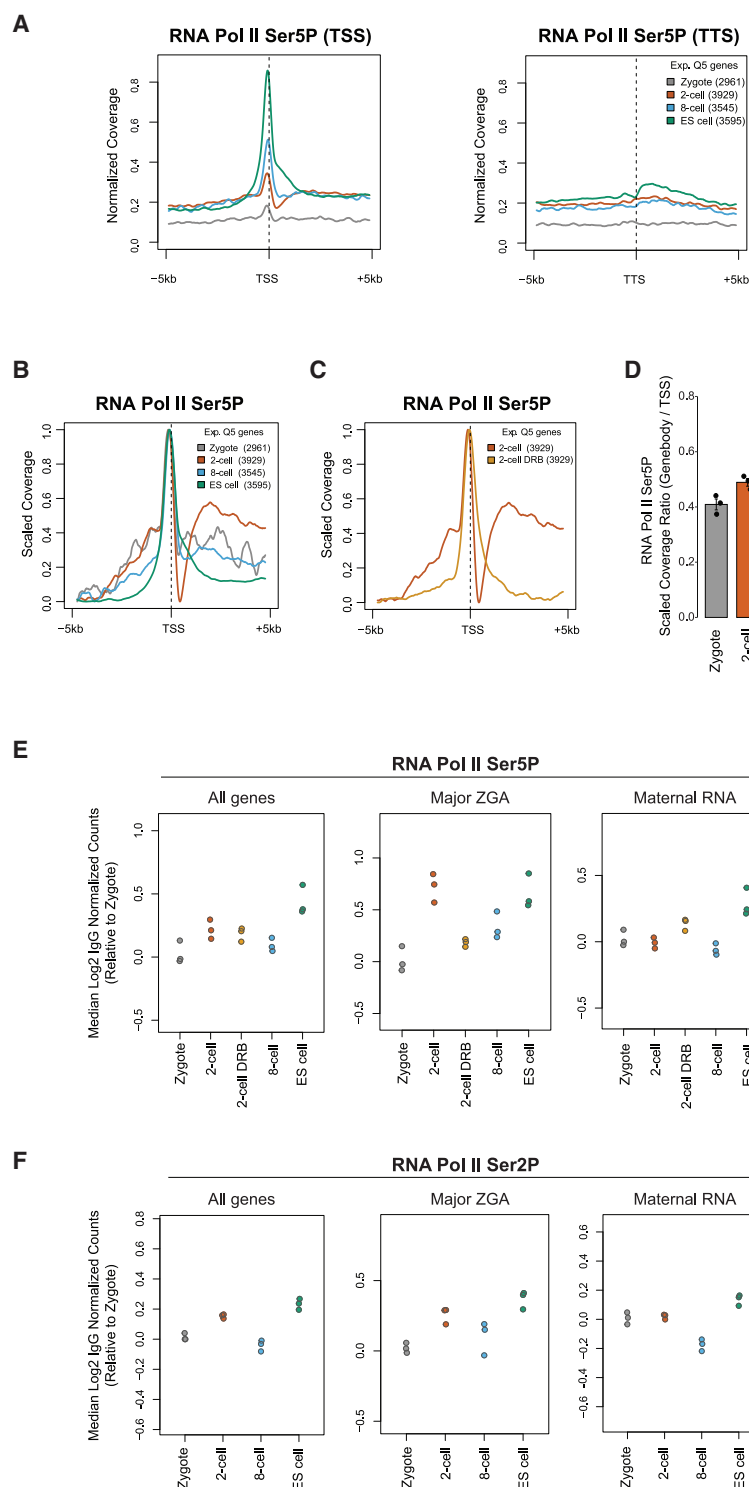


Figure 2. Dynamics of phosphorylated Pol II accumulation throughout development

(A) Metaplots of Pol II Ser5P ChIP-seq mean coverage centered at TSS (left) and TTS (right) and spanning -5 kb to $+5$ kb from each, for the indicated embryo stages and ES cells. From (A) to (D), genes in the highest expression quintile are included (maternal RNA genes were removed). Average of three biological replicates is shown.

(B and C) Scaled metaplots of Pol II Ser5P ChIP-seq mean coverage centered at TSS comparing embryos and ES cells (B) or 2-cell-stage embryos with or without DRB (C). Average of three biological replicates is shown.

(D) Ratio of Pol II Ser5P ChIP-seq scaled mean coverage at gene body (TSS $+1.5$ kb $- +5$ kb) over TSS (-200 bp $- +100$ bp) in embryos and ES cells. Bar plots indicate mean \pm SEM of three biological replicates.

(E and F) Median of Pol II Ser5P (E) and Pol II Ser2P (F) ChIP-seq read counts normalized by IgG at the gene body (TSS $+500$ bp $-$ TTS -500 bp) of all genes (left), major ZGA (middle), or maternal RNA genes (right). Gene classes are from DBTME database.⁴¹ Dot plots indicate values relative to the mean of zygote in three biological replicates.

We noted that the enrichment of Pol II Ser5P across the gene body, relative to the TSS, was highest at the 2-cell stage (Figures 2A and 2B). This was particularly visible from ~ 1 – 2 kb downstream of the TSS (Figures 2B and 2C) and suggests that Pol II is released more efficiently into active elongation across the gene body during ZGA. To address this possibility, we calculated the relative abundance of Pol II in the gene body versus the TSS in zygotes, 2- and 8-cell-stage embryos. Indeed, the ratio of Pol II Ser5P occupancy on the gene body over the TSS was highest at the 2-cell stage, followed by zygotes, and this value decreased at the 8-cell stage and was lowest in ES cells (Figure 2D). DRB treatment led to a strong decrease in the relative gene body/TSS ratio, as expected (Figure 2D). Overall, these observations are in line with our interpretation above, suggesting that pausing may increase as development proceeds, and is highest in ES cells.

To further understand the dynamics of Pol II Ser5P and Pol II Ser2P with respect to developmental gene programs, we

investigated Pol II abundance across different groups of genes. As a proxy for Pol II occupancy, we counted reads across the gene body, excluding the TSS and the TTS (± 500 base pairs [bp]), which we normalized by the sequencing depth and by the corresponding IgG control and computed values relative to

Pol II Ser2P (Figure S2). These observations were not an artifact of the ChIP, as the Pol II Ser5P dip downstream of the TSS was abolished by the addition of DRB, which inhibits transcriptional elongation (Figure 2C), and suggests that this feature may be linked to embryonic features of transcriptional elongation.

zygotes. We categorized genes based on their expression pattern as maternal and as genes expressed during major ZGA.⁴¹ We observed a strong increase in the median abundance of Pol II Ser5P across the gene body from the zygote to the 2-cell stage on major ZGA genes (Figure 2E). This shift in Pol II abundance was prevented in the DRB-treated samples and was not observed when we performed the same analysis in either all genes, or in maternally expressed genes (Figure 2E). Interestingly, Pol II Ser2P displayed a similar, albeit weaker, change from zygote to the 2-cell stage than Pol II Ser5P (Figure 2F). Pol II Ser2P levels along the body of major ZGA genes remained higher at the 8-cell stage and in ES cells, compared with zygotes (Figure 2F). These analyses are in agreement with our conclusions above (Figure 1D), indicating that both phosphorylated forms of Pol II display similar patterns of genic accumulation prior to the 8-cell stage.

Pol II Ser2P accumulates at post-ZGA genes at the 2-cell stage

The results above suggest that the genic distribution patterns of the phosphorylated forms of Pol II may differ as development proceeds. To investigate this, we examined Pol II occupancy in relationship with the global transcriptome of each developmental stage determined by RNA-seq.⁴² Globally, Pol II Ser5P showed the highest correlation between occupancy at the 2-cell stage and the corresponding stage transcriptome (Figure 3A). This correlation between Pol II Ser5P and RNA content remained high for the 4-cell stage (Figure 3A), presumably because many transcripts produced at the 2-cell stage persist at the 4-cell stage. Indeed, chromatin accessibility of gene bodies at the 2-cell stage showed a strong correlation with Pol II Ser5P levels at this stage (Figure S3A). Pol II Ser5P occupancy at the 8-cell stage correlated most strongly with its own 8-cell transcriptome; however, the zygote Pol II Ser5P distribution was poorly correlated with the transcriptome in non-fertilized MII oocytes and zygotes (Figure 3A). This could be in part because zygotes contain primarily transcripts that accumulated during oocyte growth. Interestingly, Pol II Ser2P at the 2-cell stage, while highly correlated to its corresponding transcriptome, was most strongly correlated with transcriptomes of 4- and 8-cell embryos (Figure 3B). Relatively, this pattern was more pronounced for Pol II Ser2P occupancy, than for Pol II Ser5P (Figures 3A and 3B) and suggested that Pol II may be phosphorylated as early as the 2-cell stage, on genes that will be expressed at later developmental stages. In support of this interpretation, H3K36me3 levels in 8-cell embryos correlated highly to Pol II Ser2P at the 2-cell stage (Figure S3B; and see below). Thus, we next investigated Pol II occupancy on genes belonging to MGA (Figure 3C). MGA corresponds to “Mid pre-implantation Gene Activation” and is considered as third wave of transcription during pre-implantation development.⁶ MGA includes genes that reach full transcriptional activation from the 4-cell stage and remain mostly active until the blastocyst stage (Hamatani et al.⁶ and Figure 3C). Indeed, MGA genes displayed higher enrichment of Pol II Ser5P at the 2-cell stage, relative to zygotes or maternal genes (Figure 3D). Similar results were obtained for Pol II Ser2P (Figure 3E). In line with their developmental dynamics (Figure 3C), major ZGA genes tended to display higher levels of Pol II Ser5P at the 2-cell

stage, compared with MGA genes (Figure 3D). However, Pol II Ser2P was slightly more enriched on MGA genes compared with major ZGA at the 2-cell stage (Figure 3E), despite MGA genes reaching highest transcript levels abundance only at the 4-cell stage (Figure 3C). This suggests that Pol II may be phosphorylated at Ser2 at the 2-cell stage on genes, which become activated only later.

To address whether Pol II phosphorylation occurs prior to transcriptional activation more specifically, we restricted our analysis to genes that are expressed exclusively at the 2- or 4-cell stage and are thus specific to these stages (Figure 3F). We refer to these genes as “transient,” and this analysis removes the confounding effect of steady-state transcripts produced in the previous developmental stage, yet present at subsequent stages. Pol II Ser5P enrichment was highest at the stages corresponding to their respective specific genes. For example, Pol II Ser5P enrichment at the 2-cell stage was highest at the 2-cell “transient” genes (Figure 3G). Pol II Ser2P enrichment across the gene body in 2-cell-stage embryos was higher in 2-cell transient genes compared with maternal genes (Figure 3H). An example of such a gene is shown in Figure 3I. At the 2-cell stage, Pol II Ser2P remained present at 1-cell transient genes and accumulated at 2-cell transient genes. Notably, Pol II Ser2P was enriched across the gene body of 4-cell “transient” genes at the 2-cell stage (Figure 3H). However, in contrast to 2-cell transient genes, Pol II Ser2P remained enriched in 4-cell transient genes at the 8-cell stage (Figure 3H). We note that in spite of their classification as 4-cell “transient” in the mouse DBTMEE database,⁴¹ this group of genes becomes activated at the 4-cell stage but reaches maximum transcript abundance at the 8-cell stage (Figure 3F). Indeed, examination of the genome tracks showed Pol II Ser2P at the 2-cell stage on individual genes, which are transcribed at the 4- and 8-cell stage, but whose transcripts are undetectable at the 2-cell stage (Figure 3J). Overall, these data indicate that Pol II Ser2P marks ongoing transcriptional activity in mouse embryos. Importantly, they also suggest that Pol II Ser2 is phosphorylated on the gene body of genes prior to their full activation at later developmental stages, thereby indicating a potential priming state at the 2-cell stage.

To further investigate potential elongation dynamics, we analyzed H3K36me3 levels according to gene expression in zygotes, 2- and 8-cell-stage embryos.⁴⁸ We observed a delay in the acquisition of H3K36me3 over transcribed genes during developmental progression (Figure S3C). For example, genes highly expressed at the 2-cell stage have significantly lower levels of H3K36me3 than genes equally highly transcribed at the 8-cell stage (Figures S3C–S3E). This suggests that H3K36me3 may not necessarily reflect elongation kinetics at these developmental stages and thus cautions as to whether the well-established role of H3K36me3 during transcriptional elongation also applies to embryos. These interpretations are consistent with earlier work indicating that H3K36me3 and other histone modifications associated with transcription are uncoupled from ZGA in mouse oocytes and embryos.^{49,50} Thus, H3K36me3 levels do not necessarily reflect levels of transcription during ZGA but rather H3K36me3 accumulation increases with developmental progression, which could suggest a gradual maturation of chromatin toward a more somatic type.

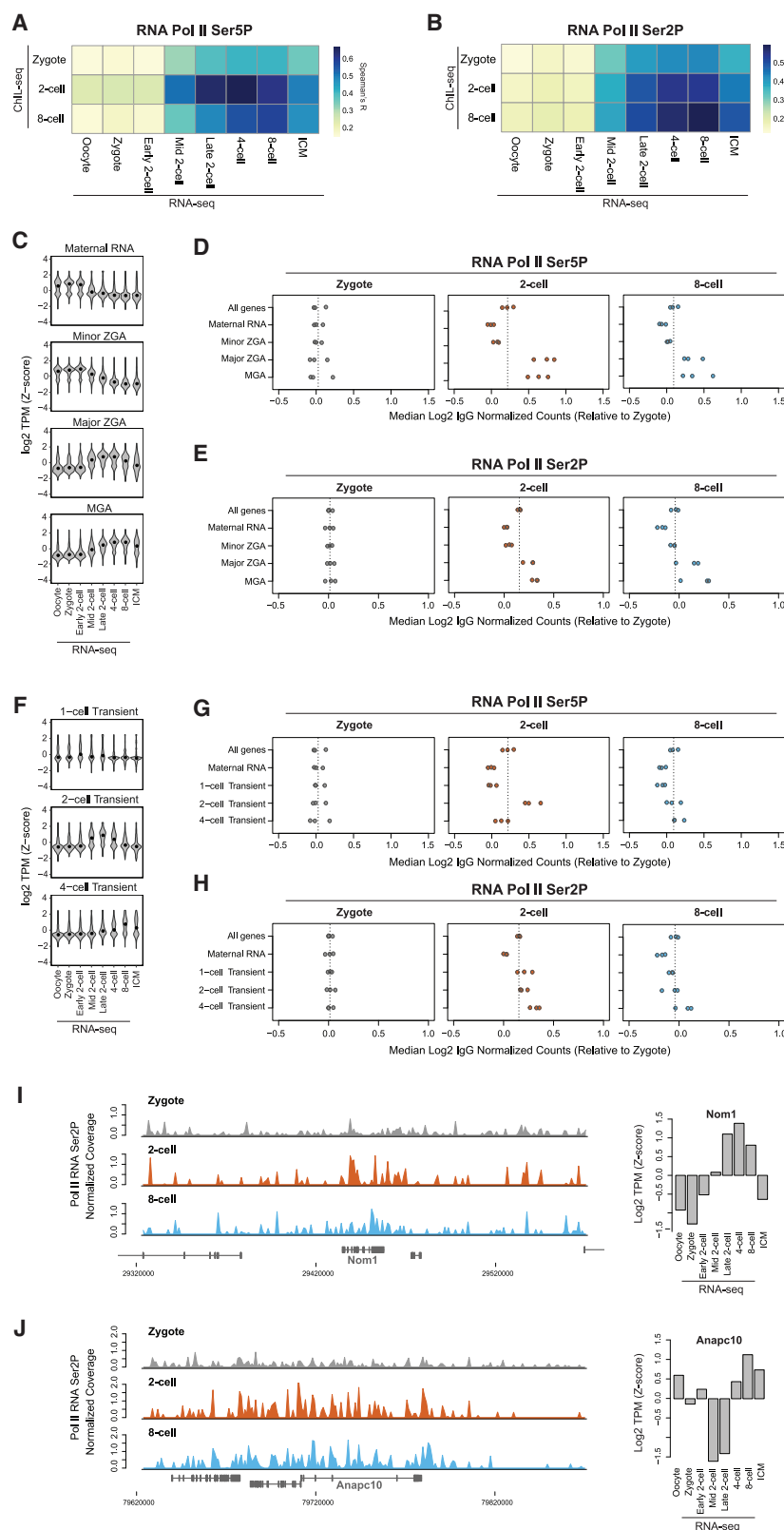


Figure 3. Pol II Ser2P accumulates at post-ZGA genes at the 2-cell stage

(A and B) Correlation heatmaps between Pol II Ser5P (A) or Ser2P (B) ChIL-seq read counts at the gene body and RNA-seq expression levels throughout development.⁴² Spearman's correlation coefficients were calculated on the mean of the replicates.

(C) Violin plots indicating mean expression levels (Z score of transcripts per million [TPM]) of the genes corresponding to DBTMEE classes analyzed in (D) and (E) throughout development.^{42,47}

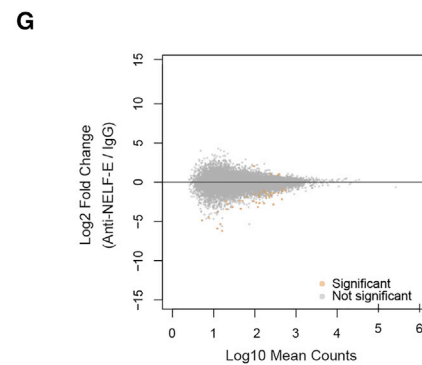
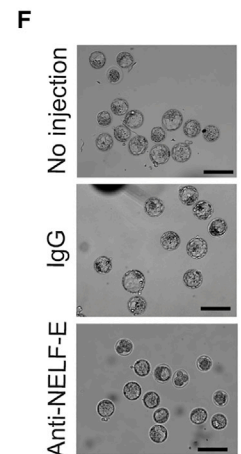
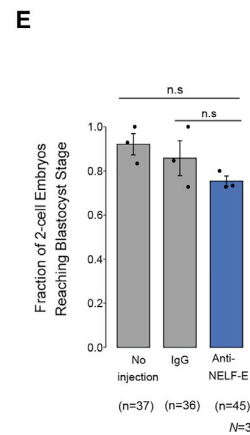
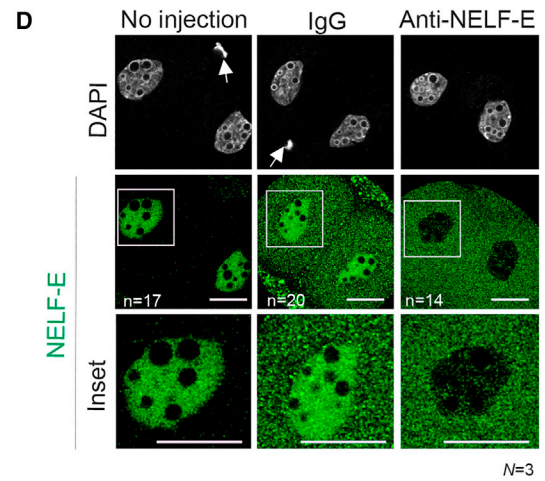
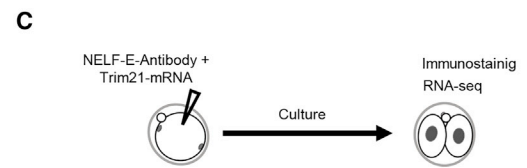
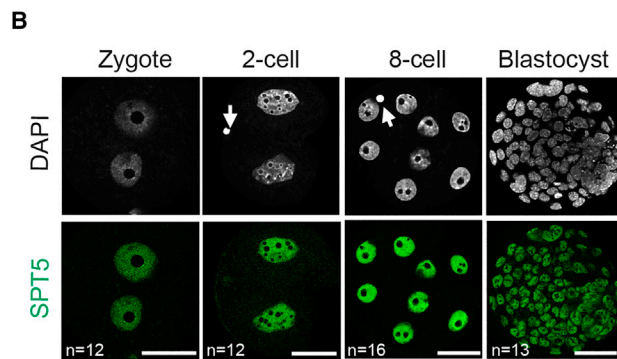
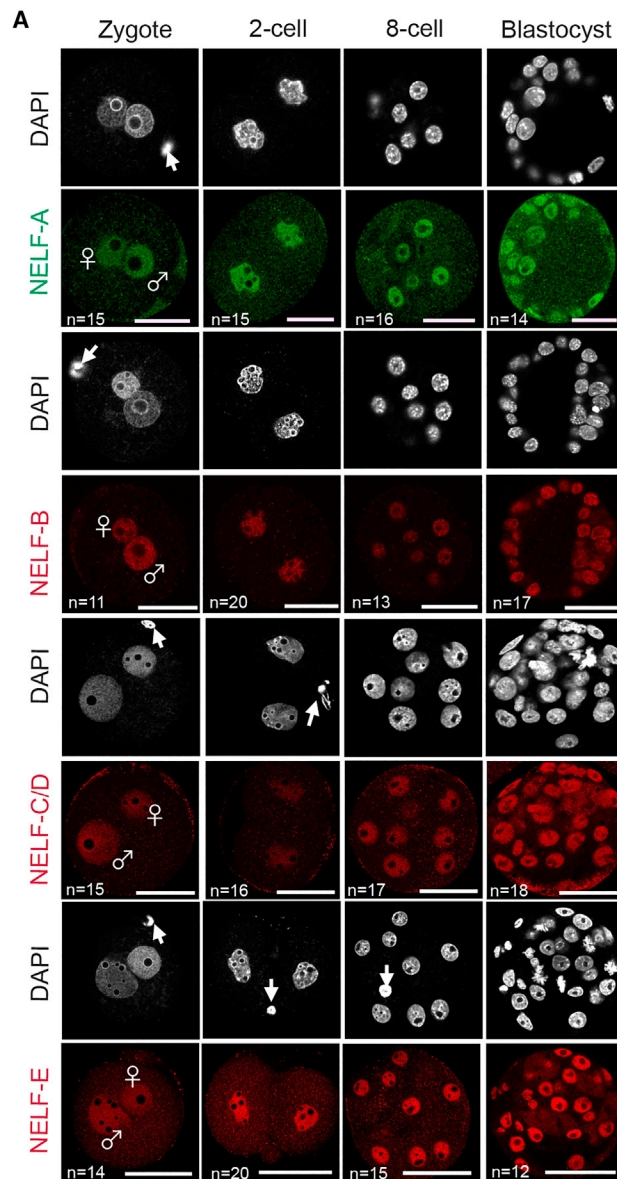
(D and E) Median of Pol II Ser5P (D) and Ser2P (E) ChIL-seq read counts normalized by IgG at gene body (TSS +500 bp – TTS –500 bp) grouped by DBTMEE gene classes indicated on the left.⁴¹ Dot plots indicate values relative to the mean of zygote in three biological replicates.

(F) Violin plots indicating mean expression levels (Z score of transcripts per million [TPM]) of the genes corresponding to the DBTMEE classes analyzed in (G) and (H) throughout development.

(G and H) Median of Pol II Ser5P (D) and Ser2P (E) ChIL-seq read counts normalized by IgG at the gene body (TSS +500 bp – TTS –500 bp) grouped by DBTMEE gene classes indicated on the left.⁴¹ Dot plots indicate values relative to the mean of zygote in three biological replicates.

(I) Example genomic regions showing Pol II Ser2P ChIL-seq mean coverage of three biological replicates at an example gene (*Nom1*) displaying highest Pol II Ser2P across the gene body in 2-cell-stage embryos. Corresponding gene expression levels during development are shown (right).

(J) Example genomic regions showing Pol II Ser2P ChIL-seq mean coverage of three biological replicates at an example of "4-cell transient" gene (*Anapc10*) in zygotes, 2- and 8-cell-stage embryos. Corresponding gene expression levels indicating undetectable transcripts at the 2-cell stage are shown (right).



(legend on next page)

NELF is dispensable for global gene expression in 2-cell-stage embryos

Our data above indicating highest Pol II gene body compared with TSS ratio suggest more productive Pol II elongation at the 2-cell stage. Thus, we next investigated the role of the pausing factors NELF and DSIF. We first examined NELF and DSIF expression in published RNA-seq datasets.⁴² *Nelfa* and *Nelfcd* transcripts are present from the zygote stage onward, presumably as maternal transcripts (Figure S4A), and as previously described for NELF-A.⁵¹ In contrast, *Nelfb* and *Nelfe* were barely detected in zygotes, their expression increased at the 2-cell stage, with *Nelfe* showing the highest levels of expression (Figure S4A). Among the DSIF components, *Spt4b* transcripts were detected at low levels and *Spt4a* was present as a maternal transcript, displaying highest abundance at the 2-cell stage. *Spt5* transcripts started to accumulate at the 2-cell stage, and remained at relatively low levels for the remainder of the pre-implantation period (Figure S4B). Because mRNA and protein expression are often uncoupled in early embryos, we performed immunostaining. We detected all NELF proteins, NELF-A, NELF-B, NELF-C/D, and NELF-E as early as the zygote stage in both pronuclei (Figure 4A). Likewise, SPT5 was robustly detected at all stages analyzed (Figure 4B), suggesting that NELF and DSIF components are maternally inherited.

To investigate the function of NELF in 2-cell embryos, we used Trim-Away to acutely deplete NELF proteins (Figure 4C).⁵² We first focused on NELF-E since NELF-E mediates contacts with RNA and DSIF.⁵³ Trim-Away using an antibody against NELF-E led to the efficient NELF-E depletion, in contrast to the IgG control (Figure 4D). Surprisingly, the majority of embryos depleted for NELF-E reached the blastocyst stage in a proportion similar to control embryos injected with IgG or non-injected (75.4% versus 85.8% and 92.1%, respectively) (Figures 4E and 4F). We obtained similar results upon Trim-Away for NELF-C/D (Figures S4C, S4D). We then performed RNA-seq in late 2-cell embryos upon Trim-Away for NELF-E (Figure S4E). Differential gene expression analysis revealed that depletion of NELF-E leads to only minor effects in gene expression (43 DE genes, $\text{padj} < 0.05$) (Figure 4G, Table S2).

We conclude that the NELF complex, in particular NELF-E and NELF-C/D, does not play a major role in the regulation of transcription during major ZGA. We note that this contrasts to NELF-A, which has been shown to regulate the “2-cell-like”

program in mouse ES cells.⁵¹ However, and in line with our observations, NELF-A regulation of the “2-cell-like” program is independent of its role as an elongation factor.⁵¹ Thus, the function of NELF as an elongation factor may be dispensable for pre-implantation development and ZGA.

CDK9 and SPT5 are key regulators of gene expression during major ZGA

CDK9, a component of the p-TEFb, phosphorylates NELF, which leads to NELF dissociation from Pol II resulting in the transition of Pol II from pausing to productive elongation.⁵⁴ In addition, CDK9 facilitates Pol II pause-release through the phosphorylation of DSIF and NELF.^{43,55} Nuclear CDK9 protein is most abundant at the late 2-cell stage.³⁶ Therefore, we reasoned that the abundant CDK9 may inhibit NELF-C/D and NELF-E function at this stage. Thus, we next addressed the role of CDK9 in transcriptional regulation using a specific inhibitor of CDK9 activity (BAY1251152, referred hereafter as BAY) (Figure 5A).⁵⁶ We verified the efficiency of BAY treatment using immunofluorescence, which revealed that the Pol II Ser2P signal was largely diminished in 2-cell embryos compared with the controls (Figure 5B). Of note, chemical inhibition of CDK9 in mouse zygotes using flavopiridol resulted in a dose-dependent developmental arrest at the 2-cell stage.³⁶ We next examined the effect of CDK9 inhibition on gene expression in late 2-cell-stage embryos using single-embryo RNA-seq. This analysis revealed significant changes in gene expression of a large number of genes upon BAY treatment ($n = 8,842$, $\text{padj} < 0.05$; Table S2). While we observed both up- and down-regulation, most genes were down-regulated ($n = 6,528$, $\text{padj} < 0.05$; Figure 5C). A large majority of major ZGA genes were among the down-regulated genes (Figure S5A and Table S3). Interestingly, minor ZGA genes were less affected and were more likely to be up-regulated (Figure 5B). Considering our data above and that of others showing that the CDK9 inhibitor, DRB, induced Pol II Ser5P accumulation at the TSS,³⁵ overall these results suggest that CDK9 plays a key role during elongation at the 2-cell stage.

To further investigate the role of elongation factors during ZGA, we next focused on DSIF. SPT5 works as an obligate dimer with SPT4 and is known to play a role as pausing factor together with the NELF complex,^{57,58} but is also involved in regulating Pol II processivity.³³ Phosphorylation of the C-terminus of SPT5 by CDK9 contributes to transcription elongation and 3'

Figure 4. NELF-C/D and NELF-E are dispensable for global gene expression in 2-cell-stage embryos

(A and B) Immunostaining with the indicated antibodies in zygotes, 2-cell and 8-cell embryos, and blastocyst. “n” indicates total number of embryos analyzed across two independent experiments. Arrows indicate polar bodies. Scale bars, 50 μm .

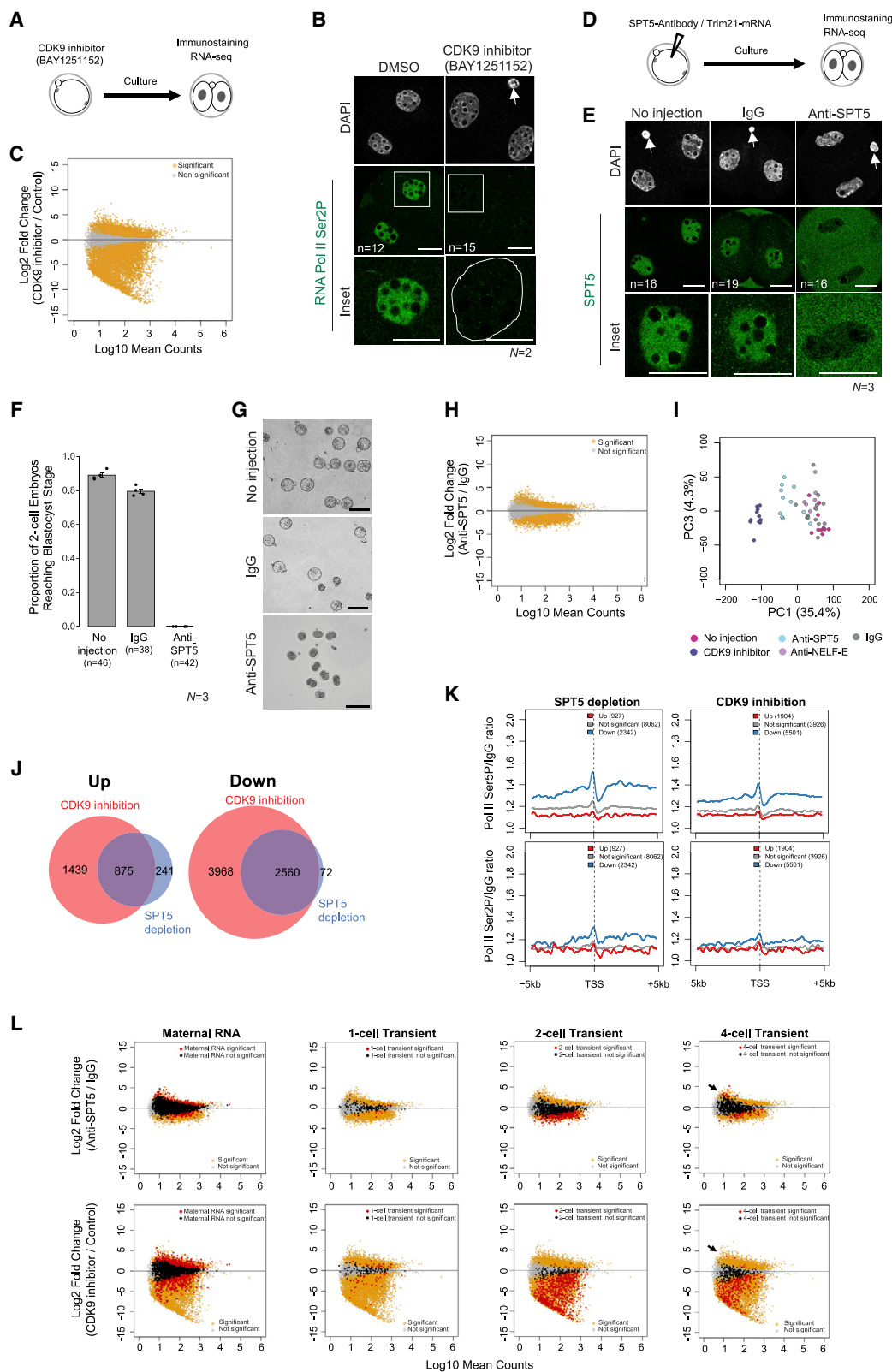
(C) Schematic of Trim-Away for NELF-E depletion.

(D) Immunostaining using anti-NELF-E antibody in 2-cell-stage embryos after Trim-Away for NELF-E or negative controls (Trim-Away for IgG and non-injected embryos); “n” indicates total number of embryos analyzed and “N” number of independent experiments. Note that Trim-Away involves the injection of an antibody into embryos, which is often picked-up during the immunofluorescence staining as cytoplasmic background by the secondary antibody. Scale bars, 50 μm .

(E) Developmental progression analysis. Proportion of control (non-injected or embryos under Trim-Away against IgG) or Trim-Away NELF-E embryos reaching the blastocyst stage; “n” indicates total number of embryos analyzed and “N” number of independent experiments. Graphs show mean \pm SEM obtained in individual experiments; p values were obtained by analysis of variance.

(F) Representative controls (non-injected or IgG) or embryos subjected to NELF-E Trim-Away 3 days after microinjection. Scale bar, 100 μm .

(G) MA plot comparing log2-fold change between 2-cell-stage embryos upon NELF-E depletion and control embryos injected with IgG against mean RNA-seq counts of 8 or 11 embryos, respectively. Three and five independent experiments were conducted for NELF-E depletion and IgG controls, respectively. Differentially expressed genes are labeled in orange ($\text{padj} < 0.05$), non-differential genes in gray.



(legend on next page)

processing.^{59,60} Thus, SPT5 may act as a central regulator for transcription elongation independently of NELF. To address this possibility, we investigated the impact of acute depletion of SPT5 on developmental progression and transcriptional regulation.

Trim-Away for SPT5 led to a strong depletion of SPT5 in 2-cell-stage embryos (Figures 5D and 5E). SPT5 depletion resulted in developmental arrest at the 2-cell stage with full penetrance (Figures 5F and 5G). In contrast, control IgG-injected or non-injected embryos formed blastocysts at normal developmental rates (79.5% and 89%, respectively, Figures 5F and 5G). This phenotype is reminiscent of that of embryos grown in the presence of global transcriptional inhibitors, such as α -amanitin,⁶¹ and suggests that SPT5 may be essential for ZGA. Thus, we next performed RNA-seq. Depletion of SPT5 led to major changes in gene expression at the 2-cell stage, although the number of mis-regulated genes was less than upon CDK9 inhibition ($n = 3,748$, $\text{padj} < 0.05$; Figure 5H and Table S2). PCA of all RNA-seq samples revealed that 2-cell-stage embryos in which CDK9 was inhibited cluster farthest away from controls (Figure 5I). Notably, embryos in which NELF-E was depleted clustered with controls, in line with our data above showing that NELF-E depletion caused only minor changes in gene expression (Figure 4G). Two-cell-stage embryos depleted of SPT5 localized in an intermediate position between controls and BAY-treated samples (Figure 5I), suggesting that SPT5 depletion leads to a similar, albeit milder transcriptional phenotype than CDK9 inhibition. Similarly to CDK9 inhibition, the absence of SPT5 led to up- and down-regulation of genes (Figure 5H). Indeed, the vast majority of down-regulated genes in SPT5-depleted embryos (2,560 of 2,632; 97.3%) were also down-regulated upon CDK9 inhibition (Figure 5J). In addition, 78.4% (875 of

1,116) of up-regulated genes upon SPT5 loss were up-regulated after CDK9 inhibition (Figure 5J).

Globally, our ChIL-seq of Pol II Ser5P and Pol II Ser2P revealed high Pol II enrichment at the 2-cell stage on the genes down-regulated upon SPT5 depletion and CDK9 inhibition (Figure 5K). These results indicate that highly transcribed genes in 2-cell-stage embryos depend upon SPT5 and/or CDK9. Conversely, up-regulated genes displayed mostly low to not-detectable Pol II enrichment (Figure 5K). Major ZGA genes were enriched among the down-regulated genes upon SPT5 depletion or CDK9 inhibition ($n = 628$ of 2,632 or $n = 1,064$ of 6,528, respectively; Figures S5A, S5C and Table S3). Maternal RNA genes were under-represented among the differential expressed genes, and more maternal RNA transcripts were up-regulated than down-regulated (Figure 5L and Table S3), which suggests that full transcriptional elongation and major ZGA are required for efficient degradation of a fraction of maternal transcripts. Indeed, SPT5 depletion resulted in up-regulation of 271 maternal RNA genes, although maternal genes were not over-represented among up-regulated genes (odds ratio 0.95) (Figure 5L). Up-regulated genes also included minor ZGA genes, and this group of genes was over-represented among up-regulated genes ($n = 510$) (Figure S5D and Table S3). Considering that most genes down-regulated upon the absence of SPT5 are also down-regulated upon CDK9 inhibition, we suggest that SPT5 regulates major ZGA downstream of CDK9 activity. Interestingly, we found that a group of 4-cell transient genes ($n = 20$) was up-regulated upon SPT5 depletion, but not upon CDK9 inhibition (Figure 5L, arrow). These data indicate that SPT5 can prevent precocious expression of some genes, which will become activated at later stages of development.

Figure 5. CDK9 and SPT5 regulate transcription in 2-cell-stage embryos

- (A) Experimental design to inhibit CDK9 activity.
- (B) Pol II Ser2P immunostaining in 2-cell embryos treated with BAY1251125 or vehicle (DMSO); "n" indicates total number of embryos analyzed across two independent experiments (N). Arrow indicates polar bodies. Scale bars, 50 μm .
- (C) MA plot comparing log₂-fold change between 2-cell-stage embryos treated with BAY1251125 and control embryos against mean RNA-seq counts of 11 and 10 embryos, respectively, derived from three independent experiments. Differentially expressed genes are labeled in orange ($\text{padj} < 0.05$), non-differential genes in gray.
- (D) Schematic of Trim-Away for SPT5 depletion.
- (E) Immunostaining using anti-SPT5 antibody in 2-cell embryos after Trim-Away for SPT5 or negative controls (Trim-Away for IgG and non-injected embryos); "n" indicates total number of embryos analyzed; "N" number of independent experiments. Scale bars, 50 μm .
- (F) Developmental progression analysis. Proportion of control (non-injected or embryos under Trim-Away against IgG) or Trim-Away SPT5 embryos reaching the blastocyst stage; "n" indicates total number of embryos analyzed and "N" number of independent experiments. Graphs show mean \pm SEM obtained in individual experiments.
- (G) Representative control (non-injected or IgG) embryos or embryos subjected to Trim-Away for SPT5 3 days after microinjection. Scale bar, 100 μm .
- (H) MA plot comparing log₂-fold change between 2-cell-stage embryos upon SPT5 depletion by Trim-Away and control embryos injected with IgG against mean RNA-seq counts of 11 embryos depleted of SPT5 and 11 control embryos from five independent experiments. Differentially expressed genes are in orange ($\text{padj} < 0.05$), non-differential genes in gray.
- (I) Principal-component analysis on RNA-seq read counts in 2-cell-stage embryos. Experiments were conducted three times for non-injected and NELF-E-depleted embryos and five times for IgG controls, SPT5-depleted embryos and CDK9 inhibition. First two principal components are shown. Percentage of the variance explained by each component is indicated.
- (J) Venn diagram comparing the overlap between up- (left) or down-regulated genes (right; $\text{padj} < 0.05$) upon BAY1251125 treatment relative to control embryos (for CDK9 inhibition) and embryos subjected to Trim-Away for SPT5 relative to their control embryos injected with IgG (for SPT5 depletion).
- (K) Metaplots comparing Pol II Ser5P (top) and Ser2P (bottom) ChIL-seq mean coverage normalized by IgG centered at TSS. Genes were grouped by their differential expression in 2-cell-stage embryos upon SPT5 depletion (left) or CDK9 inhibition (right). Average of three biological replicates is shown.
- (L) Same MA plots as (C) and (H) but with the indicated DBTMEE gene groups labeled.⁴¹ Differentially expressed genes within each gene group are marked in red ($\text{padj} < 0.05$), non-differential genes within the gene group in black, differentially expressed but not within the corresponding gene group in orange and non-differentially expressed and not in the gene group in gray.

DISCUSSION

The regulation of eukaryotic transcriptional elongation has been intensively investigated, but how this is regulated during early development is less well understood. By generating genome-wide maps of Pol II Ser2P in mouse embryos, our data provide insights into transcriptional elongation in the mammalian embryo at ZGA, a key process for development. We find that the distribution patterns of phosphorylated Pol II are practically undistinguishable during ZGA. This contrasts to ES and somatic cells, in which the initiating, Ser5 phosphorylated form of Pol II is typically found enriched at the TSS and the elongating, Ser2-phosphorylated form displays a distinctive pattern of accumulation across the gene body. It is noteworthy that the resolution of these two patterns between the Ser5P and Ser2P Pol II emerges after ZGA, suggesting a transition from an embryonic toward a more somatic configuration of Pol II after ZGA. Similarly, H3K4me3, which exists initially across ~25 kb broad domains in mouse oocytes and becomes resolved into typical peaks around the TSSs of active genes only upon ZGA.⁶² Thus, the embryo undergoes a global remodeling of chromatin and transcriptional features during MZT, which includes a potential change in the regulation of transcriptional elongation.

Our findings indicating that the ratio of Pol II Ser5P accumulation across the gene body compared with the TSS is highest at the 2-cell stage may indicate that Pol II does not stall during major ZGA. The accumulation of Pol II at the TSS and its regulation through promoter proximal pausing is not only important for productive elongation, but potentially also for promoter-enhancer communication.^{63,64} Hi-C analyses in mouse pre-implantation embryos revealed that topological associated domains, including enhancer-promoter interactions are unstructured in early stages.^{65,66} Consistently, our work suggests that Pol II is less accumulated around TSS in zygotes, 2- and 8-cell-stage embryos compared with ES cells.

We find that Pol II is already phosphorylated at Ser2 at the 2-cell stage on the gene body of genes, which will be activated at the 4- or 8-cell stage. Such Pol II Ser2P “primed” occupancy is also associated with some level of H3K36me3 enrichment at the 8-cell stage, which has been linked to poised transcription.^{67,68} Thus, Pol II Ser2 phosphorylation may function as an alternative mechanism to poise some genes for transcription to enable efficient gene expression, as the cell cycle length is progressively shortened at later stages. While our data do not have enough resolution to determine the position Pol II would stop at, it is possible that the Pol II Ser2P enrichment that we observe reflects an alternative pausing position. It is not unprecedented that Pol II stops at a “second” pause position downstream of the TSS. For example, the second “pause” position upon NELF-C and NELF-E depletion in human cells is enriched in Pol II Ser2P at this location.³¹ Alternatively, phosphorylated Pol II at Ser2 could transcribe along the gene body but transcription and/or processing may be inefficient. For example, poly-adenylation may be delayed and thus transcribed mRNAs would be unstable and/or non-detectable by RNA-seq. The time required for accumulating a significant amount of stable, fully processed transcripts may be delayed so that such transcripts are only robustly detected a few hours later, in this case at the 4-cell stage.

Last, it remains formally possible that Pol II Ser2P is uncoupled from productive elongation at these stages of development. For example, many chromatin marks are “atypical” in early embryos, including H3K9me3, which is non-repressive,⁶⁹ H3K4me3, which is distributed along broad ~25 kb domains,⁵⁰ or H3K36me3 which is uncoupled from ZGA.^{49,70} A similar scenario could apply to Pol II, perhaps due to absence of interactors or a different set of embryonic-specific CTD regulators.

The absence of any of the four NELF proteins in somatic cells in culture leads to dysregulation of promoter pausing.^{30,71,72} Our results suggest that NELF-C/D and NELF-E are not necessary for global transcription in 2-cell-stage embryos. Recent work indicates that depleting NELF-B from the zygote to the 2-cell stage results in progressive developmental arrest from the 4-cell stage and prevents blastocyst formation.⁷³ On the other hand, NELF-A overexpression in mouse ES cells induces 2-cell-like cells, with the corresponding transcriptional activation of many ZGA genes.⁵¹ However, NELF-A exerts this function as ZGA regulator independently of the NELF complex.⁵¹ Our work thus further raises the possibility that the composition of the NELF complex in embryos may differ from somatic cells. Alternatively, although NELF proteins are expressed during pre-implantation development, NELF complex assembly may not occur. While the roles of SPT5 and CDK9 in transcription have been extensively studied by many in other contexts, given the specificities of embryonic chromatin investigating the roles of these factors in their *in vivo* developmental context is essential. Whether proximal promoter pausing and the potential involvement of NELF as described in somatic cells and in *Drosophila* occurs in zygotes and 2-cell-stage embryos remains to be further investigated. Indeed, our results indicate that the ratio of Pol II Ser5P accumulation at the gene body compared with the TSS is highest in zygotes and 2-cell-stage embryos but decreases as development proceeds.

SPT5 regulates pausing in combination with NELF,^{57,58} but SPT5 also functions as an elongation factor when phosphorylated.^{59,60} The latter function is consistent with our results showing that acute SPT5 depletion leads to a substantial major ZGA defect. Interestingly, we found that some genes, which otherwise are expressed only at the 4-cell stage, become up-regulated upon SPT5 depletion. Although we observed this for a small number of genes, this result aligns with the recent study in which NELF-B depletion between the zygote and the 2-cell stage was found to induce up-regulation of some 4- and 8-cell stage “transient” genes at the 2-cell stage.⁷³ Our results showing Pol II Ser2 phosphorylation at the 2-cell stage provide a mechanistic basis for these observations and suggest that Ser2 phosphorylated Pol II may prime for gene expression at later developmental stages and point toward SPT5 as one of the regulators of this process.

Limitations of the study

We modified the ChIL-seq method as few as 100 cells. Compared with other methods, ChIL-seq offers flexibility, as it can be used in fixed samples and provides an alternative to, e.g., Cut&Run/Cut&Tag in cases in which antibodies prove inefficient in Cut&Run/Cut&Tag. However, ChIL-seq also has some disadvantages. We note that while the mapping rate is substantial, the percentage of filtered reads in embryos is

relatively low and thus the sequencing depth needed is high, which is often observed when assessing low-input chromatin mapping datasets. Like other low-input protocols, ChIL-seq may also generate data with low signal-to-noise ratio, potentially due to chromatin accessibility and thus it is important to perform side-by-side IgG controls and ideally more than two biological replicates, as we have done in our manuscript. In addition, ChIL-seq may also have some “transposase” effect due to Tn5. This effect would also be visible in IgG controls and may render some datasets difficult to interpret, depending on the transcription or chromatin factor studied. The choice of method for low-input approaches must be appropriately adapted to the sample and antibody in question. Likewise, the choice and testing of antibodies should be an obligate step since in our experience antibodies that work in ChIP will not necessarily work in ChIL.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Cell culture
 - Embryo collection and culture
- **METHOD DETAILS**
 - Trim-Away
 - DRB and BAY1251152 treatment
 - Immunofluorescence
 - ChIL-seq probe preparation
 - ChIL-seq sample preparation
 - ChIL-seq library preparation
 - ChIP-seq sample and library preparation
 - RNA-seq library preparation
 - ChIL-seq analysis
 - Analysis of published ATAC-seq, ChIP-seq and Stacc-seq datasets
 - RNA-seq analysis
 - Analysis of publicly available RNA-seq datasets
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2022.111865>.

ACKNOWLEDGMENTS

We thank Y. Ohkawa and H. Kimura for sharing the ChIL-seq protocol prior to publication and helpful discussions, A. Burton and S. Hamperl for critical reading of the manuscript, T. Straub for advice on computational analyses, and Inti de la Rosa and HMGU Genomics Facility for sequencing. K.A. was partially supported by the Uehara Foundation. Torres-Padilla lab is supported by the Helmholtz Association and by the DFG through the SFB CRC1604 “Chromatin Dynamics” (Project ID 213249687).

AUTHOR CONTRIBUTIONS

K.A. designed the study and performed all the experimental work; T.S. performed all the computational analysis; M.-E.T.-P. directed, conceived, and supervised the study. All authors contributed to and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 18, 2022

Revised: September 8, 2022

Accepted: September 30, 2022

Published: December 27, 2022

REFERENCES

1. Abe, K.-I., Inoue, A., Suzuki, M.G., and Aoki, F. (2010). Global gene silencing is caused by the dissociation of RNA polymerase II from DNA in mouse oocytes. *J. Reprod. Dev.* 56, 502–507.
2. De La Fuente, R., and Eppig, J.J. (2001). Transcriptional activity of the mouse oocyte genome: companion granulosa cells modulate transcription and chromatin remodeling. *Dev. Biol.* 229, 224–236.
3. Moore, G.P.M., and Lintern-Moore, S. (1974). A correlation between growth and RNA synthesis in the mouse oocyte. *J. Reprod. Fertil.* 39, 163–166.
4. Abe, K.-I., Funaya, S., Tsukioka, D., Kawamura, M., Suzuki, Y., Suzuki, M.G., Schultz, R.M., and Aoki, F. (2018). Minor zygotic gene activation is essential for mouse preimplantation development. *Proc. Natl. Acad. Sci. USA* 115, E6780–E6788.
5. Aoki, F., Worrad, D.M., and Schultz, R.M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev. Biol.* 181, 296–307.
6. Hamatani, T., Carter, M.G., Sharov, A.A., and Ko, M.S.H. (2004). Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* 6, 117–131.
7. Matsumoto, K., Anzai, M., Nakagata, N., Takahashi, A., Takahashi, Y., and Miyata, K. (1994). Onset of paternal gene activation in early mouse embryos fertilized with transgenic mouse sperm. *Mol. Reprod. Dev.* 39, 136–140.
8. Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell Embryos1. *Biol. Reprod.* 68, 651–654.
9. Peaston, A.E., Esvikov, A.V., Graber, J.H., De Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606.
10. Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., and Schultz, R.M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* 269, 276–285.
11. Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.-E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* 20, 332–338.
12. Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.-E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* 49, 1502–1510.
13. Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* 174, 391–405.e19.

14. Abe, K.I., Yamamoto, R., Franke, V., Cao, M., Suzuki, Y., Suzuki, M.G., Vlahovicek, K., Svoboda, P., Schultz, R.M., and Aoki, F. (2015). The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing. *EMBO J.* **34**, 1523–1537.
15. Rodriguez-Terrones, D., Gaume, X., Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.-E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119.
16. Shen, H., Yang, M., Li, S., Zhang, J., Peng, B., Wang, C., Chang, Z., Ong, J., and Du, P. (2021). Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843–2859.e20.
17. Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490.
18. Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forné, I., Decker, T.-M., Martin, D., Schüller, R., Krebs, S., et al. (2018). Tyrosine-1 of RNA polymerase II CTD controls global termination of gene transcription in mammals. *Mol. Cell* **69**, 48–61.e6.
19. Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. *Nat. Struct. Mol. Biol.* **21**, 449–455.
20. Osman, S., and Cramer, P. (2020). Structural biology of RNA polymerase II transcription: 20 Years on. *Annu. Rev. Cell Dev. Biol.* **36**, 1–34.
21. Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478.
22. Guo, J., and Price, D.H. (2013). RNA polymerase II transcription elongation control. *Chem. Rev.* **113**, 8583–8603.
23. Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483–508.
24. Odawara, J., Harada, A., Yoshimi, T., Maehara, K., Tachibana, T., Okada, S., Akashi, K., and Ohkawa, Y. (2011). The classification of mRNA expression levels by the phosphorylation state of RNAPII CTD based on a combined genome-wide approach. *BMC Genomics* **12**, 516.
25. Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848.
26. Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953.
27. Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540.
28. Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554.
29. Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516.
30. Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol. Cell* **58**, 311–322.
31. Aoi, Y., Smith, E.R., Shah, A.P., Rendleman, E.J., Marshall, S.A., Woodfin, A.R., Chen, F.X., Shiekhattar, R., and Shilatifard, A. (2020). NELF regulates a promoter-proximal step distinct from RNA Pol II pause-release. *Mol. Cell* **78**, 261–274.e5.
32. Tastemel, M., Gogate, A.A., Malladi, V.S., Nguyen, K., Mitchell, C., Banaszynski, L.A., and Bai, X. (2017). Transcription pausing regulates mouse embryonic stem cell differentiation. *Stem Cell Res.* **25**, 250–255.
33. Hu, S., Peng, L., Xu, C., Wang, Z., Song, A., and Chen, F.X. (2021). SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. *Mol. Cell* **81**, 4425–4439.e6.
34. Aoi, Y., Takahashi, Y.H., Shah, A.P., Iwanaszko, M., Rendleman, E.J., Khan, N.H., Cho, B.K., Goo, Y.A., Ganesan, S., Kelleher, N.L., and Shilatifard, A. (2021). SPT5 stabilization of promoter-proximal RNA polymerase II. *Mol. Cell* **81**, 4413–4424.e5.
35. Liu, B., Xu, Q., Wang, Q., Feng, S., Lai, F., Wang, P., Zheng, F., Xiang, Y., Wu, J., Nie, J., et al. (2020). The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature* **587**, 139–144.
36. Oqani, R.K., Kim, H.R., Diao, Y.F., Park, C.S., and Jin, D.I. (2011). The CDK9/Cyclin T1 subunits of P-TEFb in mouse oocytes and preimplantation embryos: a possible role in embryonic genome activation. *BMC Dev. Biol.* **11**, 33.
37. Harada, A., Maehara, K., Handa, T., Arimura, Y., Nogami, J., Hayashi-Takanaka, Y., Shirahige, K., Kurumizaka, H., Kimura, H., and Ohkawa, Y. (2019). A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* **21**, 287–296.
38. Kaya-Okur, H.S., Janssens, D.H., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2020). Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* **15**, 3264–3283.
39. Jukam, D., Shariati, S.A.M., and Skotheim, J.M. (2017). Zygotic genome activation in vertebrates. *Dev. Cell* **42**, 316–332.
40. Schultz, R.M. (1993). Regulation of zygotic gene activation in the mouse. *Bioessays* **15**, 531–538.
41. Park, S.J., Shirahige, K., Ohsugi, M., and Nakai, K. (2015). DBTMEE: a database of transcriptome in mouse early embryos. *Nucleic Acids Res.* **43**, D771–D776.
42. Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196.
43. Laitem, C., Zaborowska, J., Isa, N.F., Kufs, J., Dienstbier, M., and Murphy, S. (2015). CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nat. Struct. Mol. Biol.* **22**, 396–403.
44. Maderious, A., and Chen-Kiang, S. (1984). Pausing and premature termination of human RNA polymerase II during transcription of adenovirus in vivo and in vitro. *Proc. Natl. Acad. Sci. USA* **81**, 5931–5935.
45. Shoaib, M., Chen, Q., Shi, X., Nair, N., Prasanna, C., Yang, R., Walter, D., Frederiksen, K.S., Einarsson, H., Svensson, J.P., et al. (2021). Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes. *Nat. Commun.* **12**, 4800.
46. Stewart-Morgan, K.R., Reverón-Gómez, N., and Groth, A. (2019). Transcription restart establishes chromatin accessibility after DNA replication. *Mol. Cell* **75**, 408–414.
47. Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782.
48. Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A.B., Zhang, Y., Zhang, B., Yu, G., Xia, W., et al. (2019). SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat. Genet.* **51**, 844–856.
49. Bošković, A., Bender, A., Gall, L., Ziegler-Birling, C., Beaujean, N., and Torres-Padilla, M.-E. (2012). Analysis of active chromatin modifications in early mammalian embryos reveals uncoupling of H2A.Z acetylation and H3K36 trimethylation from embryonic genome activation. *Epigenetics* **7**, 747–757.
50. Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557.

51. Hu, Z., Tan, D.E.K., Chia, G., Tan, H., Leong, H.F., Chen, B.J., Lau, M.S., Tan, K.Y.S., Bi, X., Yang, D., et al. (2020). Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nat. Cell Biol.* 22, 175–186.
52. Clift, D., McEwan, W.A., Labzin, L.I., Konieczny, V., Mogessie, B., James, L.C., and Schuh, M. (2017). A method for the acute and rapid degradation of endogenous proteins. *Cell* 171, 1692–1706.e18.
53. Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* 560, 601–606.
54. Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell Biol.* 24, 787–795.
55. Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol. Cell Biol.* 12, 2078–2090.
56. Wells, C.I., Vasta, J.D., Corona, C.R., Wilkinson, J., Zimprich, C.A., Ingold, M.R., Pickett, J.E., Drewry, D.H., Pugh, K.M., Schwinn, M.K., et al. (2020). Quantifying CDK inhibitor selectivity in live cells. *Nat. Commun.* 11, 2743.
57. Rosen, G.A., Baek, I., Friedman, L.J., Joo, Y.J., Buratowski, S., and Gelles, J. (2020). Dynamics of RNA polymerase II and elongation factor Spt4/5 recruitment during activator-dependent transcription. *Proc. Natl. Acad. Sci. USA* 117, 32348–32357.
58. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* 12, 343–356.
59. Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., III, and Hartzog, G.A. (2003). Dual roles for SPT5 in pre-mRNA processing and transcription elongation revealed by identification of SPT5-associated proteins. *Mol. Cell Biol.* 23, 1368–1378.
60. Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-Mediated phosphorylation of hSPT5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* 21, 227–237.
61. Warner, C.M., and Versteegh, L.R. (1974). In vivo and in vitro effect of α -amanitin on preimplantation mouse embryo RNA polymerase. *Nature* 248, 678–680.
62. Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* 537, 548–552.
63. Abuhashem, A., Garg, V., and Hadjantonakis, A.-K. (2022). RNA polymerase II pausing in development: orchestrating transcription. *Open Biol.* 12, 210220.
64. Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512, 96–100.
65. Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., He, J., Xiang, Y., Wang, Q., Li, Y., et al. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 547, 232–235.
66. Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L., et al. (2017). 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* 170, 367–381.e20.
67. Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719.
68. Strobl-Mazzulla, P.H., Sauka-Spengler, T., and Bronner-Fraser, M. (2010). Histone demethylase Jmjd2A regulates neural crest specification. *Dev. Cell* 19, 460–468.
69. Burton, A., Brochard, V., Galan, C., Ruiz-Morales, E.R., Rovira, Q., Rodriguez-Terrones, D., Kruse, K., Le Gras, S., Udayakumar, V.S., Chin, H.G., et al. (2020). Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat. Cell Biol.* 22, 767–778.
70. Yano, S., Ishiuchi, T., Abe, S., Namekawa, S.H., Huang, G., Ogawa, Y., and Sasaki, H. (2022). Histone H3K36me2 and H3K36me3 form a chromatin platform essential for DNMT3A-dependent DNA methylation in mouse oocytes. *Nat. Commun.* 13, 4440.
71. Sun, J., and Li, R. (2010). Human negative elongation factor activates transcription and regulates alternative transcription initiation. *J. Biol. Chem.* 285, 6443–6452.
72. Sawarkar, R., Sievers, C., and Paro, R. (2012). Hsp90 globally targets paused RNA polymerase to regulate gene expression in response to environmental stimuli. *Cell* 149, 807–818.
73. Abuhashem, A., Lee, A.S., Joyner, A.L., and Hadjantonakis, A.-K. (2022). Rapid and efficient degradation of endogenous proteins in vivo identifies stage-specific roles of RNA Pol II pausing in mammalian development. *Dev. Cell* 57, 1068–1080.e6.
74. Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* 534, 652–657.
75. Miyazari, Y., and Torres-Padilla, M.-E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. *Nature* 483, 470–473.
76. Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. *F1000Res.* 10, 33.
77. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
78. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008.
79. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10.
80. R Core Team (2022). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). <https://www.r-project.org/>.
81. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121.
82. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
83. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
84. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
85. Freter, R., Osawa, M., and Nishikawa, S.-I. (2010). Adult stem cells exhibit global suppression of RNA polymerase II serine-2 phosphorylation. *Stem Cell* 28, 1571–1580.
86. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
87. Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol. Cell* 66, 568–576.e4.
88. Blakeley, P., Fogarty, N.M.E., Del Valle, I., Wamaitha, S.E., Hu, T.X., Elder, K., Snell, P., Christie, L., Robson, P., and Niakan, K.K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* 142, 3613–3165.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-SPT5	Santa Cruz	Cat#sc-136075; RRID:AB_2286723
Mouse anti-NELF-A	Abcam	Cat#ab167675; RRID:AB_2927462
Rabbit anti-NELF-B	Abcam	Cat#ab167401; RRID:AB_2721176
Rabbit anti-NELF-C/D	Abcam	Cat#ab170934; RRID:AB_2927463
Rabbit anti-NELF-E	Abcam	Cat#ab170104; RRID:AB_2827280
Goat IgG	Santa Cruz	Cat#sc-2028; RRID:AB_737167
Rabbit anti-Pol II Ser5P	Abcam	Cat#ab5131; RRID:AB_449369
Rabbit anti-Pol II Ser2P	Abcam	Cat#ab5095; RRID:AB_304749
Rabbit IgG	Cell Signaling	Cat#2729S; RRID:AB_1031062
Goat anti-Rabbit IgG Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat#A32731; RRID:AB_2633280
Goat anti-Mouse IgG Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat#A11029; RRID:AB_2534088
Goat anti-Rabbit IgG Secondary Antibody, Cy3	Invitrogen	Cat#A10520; RRID:AB_2534029
Donkey anti-rabbit IgG	Jackson ImmunoResearch Laboratories	Cat#711-005-152; RRID:AB_2340585
Biological samples		
Mouse pre-implantation embryos	This paper	N/A
Chemicals, peptides, and recombinant proteins		
5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside (DRB)	Santa Cruz	Cat#sc-200581
BAY1251152	Selleckchem	Cat#S8730
CHIR99021	Cayman Chemical Company	Cat#13122
PD0325901	Cayman Chemical Company	Cat#13034
Dimethyl sulfoxide	Sigma Aldrich	Cat#41639
RNasin	Promega	Cat#N2511
Pregnant Mare Serum Gonadotropin (PMSG)	Ceva	PREGMAGON [®]
human Chorionic Gonadotrophin (hCG)	MSD Animal Health	OVOGEST [®]
DMEM, high glucose, GlutaMAX Supplement, pyruvate	Gibco	Cat#31966021
FBS	PAN BIOTECH	Cat#P30-3302
2-mercaptoethanol	Gibco	Cat#21985023
Non-essential amino acids	Gibco	Cat#11140-050
Penicillin Streptomycin	Gibco	Cat#15070-063
LIF	Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)	N/A
Gelatine solution	PAN BIOTECH	Cat#P06-20410
KSOM	This paper	N/A
Vectashield	Vector Laboratories	Cat#H-2000
DBCO-PEG5-NHS ester	Santa Cruz	Cat#sc-499348
Tn5 transposase	Glow Biologics	Cat#GBRP-TN5
TAPS	MP Biomedicals	Cat#103007
N,N-dimethylformamide	Sigma Aldrich	Cat#D4551
T4 DNA ligase	New England Biolab	Cat#M0202
T4 DNA polymerase	New England Biolab	Cat#M0203

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
T7 RNA polymerase	New England Biolab	Cat#M0251
DNase I	Thermo Scientific	Cat#AM2238
SMARTScribe reverse transcriptase	Takara	Cat#639536
Prime star GXL	Takara	Cat#R050A
Proteinase K	Sigma Aldrich	Cat#P6556
ERCC RNA Spike in Mix	Invitrogen	Cat#4456740
SuperScript III Reverse Transcriptase	Invitrogen	Cat#18080093
KAPA HiFi ReadyMix	Roche	Cat#KK2602
Critical commercial assays		
mMESSAGE mMACHINE kit (Ambion)	Invitrogen	Cat#AM1344
Amicon Ultra–0.5 mL, 30 kDa or 100 kDa centrifugal filters	Millipore	Cat#UFC503008, Cat#UFC510024
PD Minitraps™ G-25	GE Healthcare	Cat#GE28-9180-07
Microcon Ultracel DNA Fast Flow Membrane	Millipore	Cat#MRCF0R100
RNeasy MinElute Cleanup kit	Qiagen	Cat#74204
Gene read Size selection kit	Qiagen	Cat#180514
AMPure XP beads	Beckman Coulter	Cat#A63882
RNAClean XP beads	Beckman Coulter	Cat#A66514
NEBNext Ultra II DNA Library Prep Kit	New England Biolab	Cat#E7645S
Nextera XT DNA Library Preparation Kit	Illumina	Cat#FC-131-1096
Deposited data		
Raw data	This paper	GEO: GSE195840
RNA-seq data	Deng et al., 2014 ⁴²	GEO: GSE45719
Stacc-seq data	Liu et al., 2020 ³⁵	GEO: GSE135457
H3K36me3 ChIP-seq data	Xu et al., 2019 ⁴⁸	GEO: GSE112835
ATAC-seq data	Wu et al., 2016 ⁷⁴	GEO: GSE66390
Experimental models: Cell lines		
Mouse: E14 embryonic stem cells	Miyazaki and Torres-Padilla, 2012 ⁷⁵	RRID:CVCL_C320
Experimental models: Organisms/strains		
Mouse: C57BL/6J × CBA/H F1 mouse	Janvier labs	RRID:MGI:5650652
Oligonucleotides		
Probe: ChIL DNA Fw	Sigma-Aldrich	N/A
Probe: ChIL DNA Rv	Sigma-Aldrich	N/A
Read2-2 primer: CGGAGATGTGTATAAGAGACAGNNNNNN	Sigma-Aldrich	N/A
Ad1 primer: AATGATACGGCGACCAACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG	Sigma-Aldrich	N/A
Ad2-2 primer: CAAGCAGAAGACGGCATACGAGAT[8mer_index]GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	Sigma-Aldrich	N/A
Oligo dT primer: AAGCAGTGGTATCAACGCAGAGTACT ₃₀ VN	Biomers.net	N/A
Template switching oligo: AAGCAGTGGTATCAACGCAGAGTACATrGrG+G	Exiqon	N/A
ISPCR oligo: AAGCAGTGGTATCAACGCAGAGT	Biomers.net	N/A
Recombinant DNA		
Plasmid: pGEMHE-mCherry-mTrim21	Clift et al., 2017 ⁵²	Addgene plasmid: 105522

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Snakemake	Mölder et al. ⁷⁶	https://snakemake.readthedocs.io/en/stable/index.html
bowtie2	Langmead and Salzberg, 2012 ⁷⁷	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Danecek et al. ⁷⁸	https://www.htslib.org/doc/samtools.html
Cutadapt	Martin, 2011 ⁷⁹	https://cutadapt.readthedocs.io/en/stable/
R	R Core Team ⁸⁰	https://www.r-project.org
Bioconductor	Huber et al., 2015 ⁸¹	https://bioconductor.org
STAR	Dobin et al., 2013 ⁸²	https://github.com/alexdobin/STAR
Bedtools	Quinlan and Hall, 2010 ⁸³	https://bedtools.readthedocs.io/en/latest/
Rsem	Li and Dewey, 2011 ⁸⁴	https://deweylab.github.io/RSEM/
Data analysis pipelines		https://zenodo.org/record/7313885#.Y25fAS8w2WY

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Maria-Elena Torres-Padilla (torres-padilla@helmholtz-muenchen.de).

Materials availability

Oligonucleotides generated in this study are available from the [lead contact](#).

Data and code availability

- ChIL-seq, ChIP-seq and RNA-seq datasets are available on GEO under the accession GSE195840.
- Reproducible pipelines were created by Snakemake (version snakemake-minimal 5.2.4) and are available on Github at https://github.com/tschauer/Abe_et_al_2022.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

Mouse embryonic stem (ES) cells were cultured in Dulbecco's modified Eagle's medium (DMEM)-GlutaMAX-I containing 15% FBS, 0.1 mM 2-beta-mercaptoethanol, non-essential amino acids, penicillin and streptomycin and 2× LIF over gelatin-coated plates. The medium was supplemented with 2i (3 μM CHIR99021 and 1 μM PD0324901, Cayman Chemical Company) for maintenance and expansion.

Embryo collection and culture

All experiments were approved by the Government of Upper Bavaria. Mice housed in Helmholtz Zentrum München were maintained and bred under institutional guidelines in a 12h light cycle at 20–24°C and 45–65% humidity. To obtain embryos, 5–7-week-old F1 (C57BL/6J × CBA/H) female mice were injected with 10 IU pregnant mare serum gonadotropin (PMSG) (Ceva) and human chorionic gonadotrophin (hCG) (MSD Animal Health) 48 h later. Embryos were collected from F1 females mated with F1 males (3–6 months old) at the following post hCG times: zygotes 26h, 2-cell stage 46h, 8-cell stage 68h and blastocysts 90h.

METHOD DETAILS

Trim-Away

Trim-Away was performed as described.⁵² mRNAs were transcribed *in vitro* from pGEMHE-mCherry-mTrim21 (Addgene: 105522) using the mMESSAGE mMACHINE kit (Ambion). Antibodies and RNasin (Promega) were purified and concentrated in PBS using Amicon Ultra-0.5 mL, 30 kDa or 100 kDa centrifugal filters (Millipore). Zygotes were collected from F1 (C57BL/6J × CBA/H) female mice crossed with F1 males at 17–19 h post hCG injection and microinjected with 1–2 pL of the mixture containing 500 ng/μL antibody

(anti-SPT5 (Santa cruz, sc-136075), anti-NELF-C/D (Abcam, ab170934), anti-NELF-E (Abcam, ab170104) or goat IgG (Santa Cruz, sc-2028)), 200 ng/ μ L Trim21-mCherry mRNA, 2 U/ μ L RNasin and 0.05% NP-40. Embryos were cultured in KSOM under paraffin oil at 37°C, 5% CO₂.

DRB and BAY1251152 treatment

Zygotes were collected at 18 h post hCG and cultured in KSOM containing 0.1% DMSO, 100 μ M DRB (Santa cruz) or 10 μ M BAY1251152 (Selleckchem) for 28 h until embryos reached at late 2-cell stage.

Immunofluorescence

Embryos or ES cells were fixed in 4% paraformaldehyde for 15 min at room temperature and permeabilized in 0.5% Triton X-100 in PBS for 15 min. After three times washing in PBS containing 0.5% BSA, the samples were incubated overnight at 4°C in primary antibodies diluted in PBS containing 2% BSA (1:100 dilution). Antibodies were used as follows: anti-SPT5 (Santa Cruz, sc-136075), anti-NELF-A (Abcam, ab167675), anti-NELF-B (Abcam, ab167401), anti-NELF-C/D (Abcam, ab170934), anti-NELF-E (Abcam, ab170104), anti-Pol II Ser2P (Abcam, ab5095). After overnight incubation, samples were washed at three times in PBS containing 0.5% BSA and incubated for one hour at room temperature in PBS containing 2% BSA and secondary antibodies coupled with Alexa 488 or Cy3 fluorophores (Invitrogen) (catalog numbers A-32731, A-11029 and A-10520, 1:500 dilution). After washing, samples were mounted in Vectashield (Vector Laboratories) containing DAPI. Confocal microscopy was performed using a 63 \times oil objective in a Leica SP8 confocal microscope.

ChIL-seq probe preparation

ChIL-seq probe was prepared following the protocol of Harada et al.³⁷ Briefly, 100 μ g of anti-rabbit IgG (Jackson ImmunoResearch Laboratories) in 100 μ L 100 mM NaHCO₃ (pH 8.3) was mixed with 0.2 μ L of DBCO-PEG5-NHS ester and incubated for 1 h at room temperature using a rotator. The antibody was filtered by a PD MiniTrap G-25 desalting column (GE Healthcare) to remove the free DBCO-PEG5-NHS. DBCO-PEG5-conjugated antibodies were concentrated using an Amicon Ultra-0.5 NMWL 100 kDa centrifugal filter (Merck Millipore) and then diluted to 1 μ g/ μ L with PBS. The concentration of the antibody was calculated from the absorbances at 280 and 309 nm measured using a spectrophotometer (Nanodrop). Double-stranded ChIL DNA was prepared by annealing two oligonucleotides (95°C for 5 min and then decreasing the temperature by 0.1°C/s to 20°C, in 10 mM Tris-Cl, pH 7.4, 1 mM EDTA, and 100 mM NaCl). 100 μ L of double-stranded ChIL DNA (annealed at 10 μ M) was mixed with 75 μ L of DBCO-PEG5-conjugated antibody (1 μ g/ μ L) and incubated for 1 week at 4°C using a rotator. After the Click reaction with azide-DNA, the antibody was filtered by Microcon Ultracel DNA Fast Flow Membrane (Merck Millipore) to remove the unreacted DNA, and the antibody-DNA conjugates were concentrated using an Amicon Ultra-0.5 NMWL 100 kDa centrifugal filter (Merck Millipore) and then diluted to 1 μ g/ μ L with PBS.

ChIL-seq sample preparation

ChIL-seq was performed following a modified protocol of Harada et al.³⁷ using an equivalent number of cells per condition. Briefly, for ES cells ChIL-seq, 100 or 1000 cells were plated on gelatin-coated 96-well plates, left to attach for 16 h and fixed. For embryos ChIL-seq, we used: 100x zygotes, 50x 2-cell stage embryos or 12x 8-cell embryos per replicate. Cells or embryos were fixed with 1% formaldehyde in PBS for 5 min and permeabilized with 1% Triton X-100 in PBS for 20 min. Cells or embryos were incubated in 2% BSA in PBS with the primary antibodies (anti-Pol II Ser5P (Abcam, ab5131), anti-Pol II Ser2P (Abcam, ab5095)) or Rabbit IgG (Cell Signaling) at 2 μ g/mL for 6 h at room temperature, washed three times in 0.5% BSA in PBS and then incubated in 2% BSA in PBS containing 500 mM NaCl with ChIL probe at 2 μ g/mL at 4°C overnight. After washing three times in PBS for ES cells or PBS containing 0.1% BSA for embryos, samples were incubated in 50 μ L Tn5 solution (50 mM HEPES-KOH, pH 7.2, 0.1 M NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, 0.1% Triton X-100 and 10% glycerol) containing 0.5 μ L HActive™ Tn5 transposase (Glow Biologics, GBRP-TN5) for 10 min at room temperature. Tn5-MEDS-B oligo (2 μ M) was then added at 0.5 μ L per well in 1 \times dialysis buffer (50mM HEPES-KOH, pH 7.2, 0.1M NaCl, 0.1mM EDTA, 1mM dithiothreitol, 0.1% Triton X-100 and 10% glycerol) and incubated for 60 min at room temperature. After washing once in 1 \times dialysis buffer and at three times in PBS for ES cells or PBS containing 0.1% BSA for embryos, the samples were incubated in 100 μ L 1 \times TAPS-DMF buffer (10 mM TAPS-NaOH, pH 8.5, 5 mM MgCl₂ and 10% N,N-dimethylformamide) for 1 h at 37°C and then washed at three times in PBS for ES cells or PBS containing 0.1% BSA for embryos. To stop the Tn5 reaction, cells were incubated in 100 μ L 0.2% SDS for 10 min at room temperature or embryos were incubated in 0.1% BSA/PBS containing 50 mM EDTA at 50°C for 30 min. After washing three times in PBS for ES cells or PBS containing 0.1% BSA for embryos, samples were incubated in 50 μ L reaction mixture (200 U T4 DNA ligase, NEB, and 1.5 U T4 DNA polymerase, NEB, in 1 \times T4 DNA ligase reaction buffer containing 0.1 mM dNTP) for 30 min at room temperature. After discarding the solution, for ES cells, 100 μ L 0.2% SDS were added to remove DNA-bound enzymes, followed by incubation for 10 min at room temperature and three washes in PBS. For embryos, 0.1% BSA/PBS containing 50 mM EDTA were added and incubated at 50°C for 30 min, and then washed at three times in 0.1% BSA/PBS. *In situ* transcription was then carried out using T7 RNA polymerase (800 U per well; NEB) for 20 h at 37°C. The samples were treated with DNase I (2.5 U per well; Thermo) for 30 min at 37°C and the transcribed RNA was then collected using a Qiagen MinElute Cleanup kit. The Ser5P and Ser2P Pol II antibodies

(ab5131 and ab5095, respectively) have been tested with peptide competition assays for specificity (see for example Freter et al.⁸⁵). The ab5131 Ser5P Pol II antibody has been tested using peptide competition by dot blot, which indicated that it is effectively competed by a Ser5P repeat peptide but does not cross-react with a non-phosphorylated CTD peptide (www.abcam.com/ab5131). Likewise, the ab5095 Ser2P Pol II antibody has also been tested for cross-reactivity and is effectively competed by a Ser2P peptide, but not by a Ser5P peptide (www.abcam.com/ab5095).

ChIL-seq library preparation

Purified RNA was reverse transcribed by SMARTScribe reverse transcriptase (Takara) following the manufacturer's instructions, with Read2-2 primer. Synthesized cDNA was amplified by Prime star GXL (Takara) following the manufacturer's instructions, with Ad1 and Ad2-2 primers. The PCR amplification conditions were followed by 14 cycles for 1000 cells with anti-Pol II Ser5P antibody, 15 cycles for 1000 cells with anti-Pol II Ser2P antibody, 16 cycles for 100 cells or embryos with anti-Ser5P antibody, or 17 cycles for 100 cells or embryos with anti-Ser2P antibody at 95°C for 10 s, 60°C for 15 s and 68°C for 100 min. We used corresponding IgGs as negative controls in our manuscript, considering that a similar result to the IgG negative control was obtained when no primary control was added in Harada et al.³⁷ After amplification, small fragments (<150 bp) were removed using Gene read Size selection kit (Qiagen) and large size of fragments (>600 bp) were removed using AMPure XP beads (Beckman Coulter) according to the manufacturer's instruction. The amplified DNA was finally purified using AMPure XP beads, excluding large and small fragments. The quality of the libraries was assessed using a Bioanalyzer 2100 (Agilent Technologies). Samples were paired-end sequenced at PE150 on an Illumina NovaSeq 6000 platform.

ChIP-seq sample and library preparation

ES cells were fixed with 1% formaldehyde in DMEM at 37°C for 10 min, neutralized in DMEM containing 125 mM glycine at room temperature for 5 min and washed with PBS at three times. Fixed cells were collected with 2 mL soft lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 0.1% NP-40 10% Glycerol) and centrifuged at 3000 rpm for 15 min. After removal of supernatant, the pellet was resuspended in 100 μ L SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0) and sonicated using E220 evolution (Covaris) at 140 peak incident power, 10% duty factor, 200 cycles per burst, 80 s treatment time and 7°C water temperature. Sheared chromatin samples were centrifuged at 13000 rpm for 10 min and then the supernatant was diluted with 900 μ L dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.0, 167 mM NaCl) and incubated with 2 μ g anti-Pol II Ser5P, anti-Pol II Ser2P antibody or IgG at 4°C overnight with rotation. After incubation with 20 μ L Protein A/G magnetic beads (Thermo Scientific) for 3 h at 4°C with rotation, the samples were washed with low salt buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100), high salt buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100) at three times and then washed with LiCl buffer (250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl, pH 8.0) at three times. DNA was eluted with 200 μ L elution buffer (1% SDS, 100 mM NaHCO₃) at 65°C for 10 min with shaking and then added with 200 mM NaCl as final concentration. Samples were decrosslinked after overnight 65°C incubation, followed by proteinase K treatment for 1 h and DNA was purified by phenol-chloroform extraction. ChIP-seq libraries were prepared using a NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's instructions. The quality of the libraries was assessed using a 2100 Bioanalyzer (Agilent). Samples were paired-end sequenced at PE150 on an Illumina Hiseq 4000.

RNA-seq library preparation

Zygotes were collected at 17 h post hCG injection and cultured in KSOM with or without BAY1251152 until 46 h post hCG or were subjected to Trim-Away for SPT5, NELF-E or IgG control. Embryos were washed in PBS and flash-frozen in lysis buffer containing ERCC RNA spike in (Invitrogen) and processed for single embryo RNA-seq according to the Smart-Seq2 protocol.⁸⁶ Briefly, the RNA is purified by RNAClean XP beads (Beckman Coulter) and then the purified RNA containing the magnet beads is reverse transcribed by SuperScript III Reverse Transcriptase (Invitrogen) with 20% Betaine, 1 μ M oligo dT primer and 1 μ M template switching oligo according to the manufacture's instruction. The synthesized cDNA is amplified at 16 cycles by KAPA HiFi ReadyMix (Roche) with 0.1 μ M ISPCR oligo and is purified by AMPure XP beads. RNA-seq libraries are prepared by Nextera XT DNA sample preparation kit (Illumina). The quality of the libraries was assessed using a 2100 Bioanalyzer (Agilent). Samples were paired-end sequenced at PE150 on an Illumina NovaSeq 6000 platform.

ChIL-seq analysis

The mouse reference genome and annotations were downloaded from Ensembl (version GRCm38.dna.primary_assembly.fa and GRCm38.101.gtf).

Paired-end sequencing reads were trimmed using cutadapt (version 3.4) with parameters -a CTGTCTCTTATA -A CTGTCTCTTATA -pair-filter = any -minimum-length = 20. Trimmed reads were mapped to the reference genome using bowtie2 (version 2.3.5) with parameters -end-to-end -very-sensitive -no-unal -no-mixed -no-discordant -I 10 -X 500. Aligned reads were filtered by quality using samtools (version 1.3) with parameter -q 12. Read pairs were read in to R (version 3.6.3) using readGAlignmentPairs from the GenomicAlignment package (version 1.22.0) and fragments were filtered for fragments that have both unique start and end positions. Coverage vectors were created using the coverage function from GenomicRanges package (version 1.38.0) and

were normalized by the number of fragments and multiplied by a million. Normalized coverages were centered in a 10 kb window surrounding transcription start sites (TSS) or termination sites (TTS) of protein coding genes using coverageWindowsCentered-Stranded from tsTools package (version 0.1.2). Biological replicates of centered coverage matrices were averaged and genes were grouped by their expression levels in the given stage i.e. Q1 included genes with zero counts and Q2-Q5 were equal sized groups of genes stratified by increasing TPM values (transcript per million, see published RNA-seq analysis). Maternal RNA genes (DBTMEE)⁴¹ were removed prior Q1-Q5 grouping. Composite plots were generated using plotComposite from HelpersforChIPSeq package (version 0.1.0). When indicated, IgG normalization was applied on the TSS or TTS centered coverage matrices (replicate averaged) by calculating the element-wise ratio of the Ser5P or the Ser2P matrices and the corresponding IgG matrix. Scaled composite plots were further scaled by the minimum and maximum values. The ratio of gene body over TSS was calculated on the mean scaled coverages in windows of TSS -200 bp – +100 bp for TSS, and TSS +1.5 kb – +5 kb for gene body. For genomic region tracks, the mean coverage in 1 kb non-overlapping bins was plotted. Genome-wide distribution of reads were analyzed at the following genomic features: TSS \pm 2.5 kb, exons, introns, TTS \pm 2.5 kb of protein coding genes that are longer than 1 kb and intergenic regions that are the gaps in between. The total sum of the reads per feature were normalized by the total length to obtain a rate. Finally, feature rates were divided by the sum of the rates for each sample. For count-based analysis, fragments were counted over protein coding gene bodies, that is 500 bp downstream of TSS and 500 bp upstream of TTS, as well as over 100 kb large consecutive bins. Size factors were calculated on the 100 kb bins using estimateSizeFactorsForMatrix from DESeq2 package (version 1.26.0) and applied on read counts over gene body. Normalized counts were log2 transformed, while zero counts were treated as NA (missing value). For principal-component analysis (PCA) and correlation heatmaps, genes with NAs in any of the considered samples were omitted. For PCA, log2 normalized counts were centered by the mean and scaled by the standard deviation. For dot plots, showing log2 normalized counts, the mean of zygote values was subtracted for each gene and the median of each gene classification group (e.g. DBTMEE class) was calculated while NAs in the given sample were omitted. When indicated, IgG normalization was applied on the log2 size factor normalized counts by gene-wise subtracting the average values of the corresponding IgG replicates. When ChIL-seq data were compared to published datasets, log2 normalized counts were scaled by the gene body length and Spearman's correlation coefficients across conditions were visualized as heatmaps.

Analysis of published ATAC-seq, ChIP-seq and Stacc-seq datasets

Published ATAC-seq, H3K36me3 ChIP-seq and Stacc-seq datasets were obtained from GEO with accession numbers GSE66581,⁷⁴ GSE112835⁴⁸ and GSE135457,³⁵ respectively. Paired-end sequencing reads were trimmed using cutadapt (version 3.4) with parameters -a CTGTCTCTTATA -A CTGTCTCTTATA -a AGATCGGAAGAGC -A AGATCGGAAGAGC --minimum-length = 20. Trimmed reads were mapped to the reference genome using bowtie2 (version 2.3.5) with parameters --end-to-end --very-sensitive --no-unal --no-mixed --no-discordant -I 10 -X 500. Aligned reads were filtered by quality using samtools (version 1.3) with parameter -q 12. Read pairs were read in to R (version 3.6.3) using readGAlignmentPairs from the GenomicAlignment package (version 1.22.0) and fragments were filtered for fragments that have either unique start or end positions. Downstream analyses were performed as for ChIL-seq.

RNA-seq analysis

The reference genome and annotation files were modified by concatenating the mouse genome with ERCC (source: ThermoFisher) and the construct 'pGEMHE-mCherry-mTrim21' fasta and gtf files, respectively. The endogenous Trim21 locus was masked using bedtools (version 2.29.0). The modified reference fasta file was used to prepare STAR index files using rsem-prepare-reference (rsem version 1.3.3). Paired-end reads were aligned using STAR (version 2.7.6a) with parameters --quantMode TranscriptomeSAM GeneCounts --outSAMtype BAM SortedByCoordinate. Gene expression levels were estimated as transcript per million (TPM) using rsem-calculate-expression (rsem version 1.3.3) with parameters --bam --paired-end --strandedness none. Samples were filtered by quality criteria i.e. the minimum number of reads and detected genes were set to 5×10^5 reads and 1.000 genes, respectively. The percentage of reads assigned to mitochondrial genes or ERCC spike-ins were thresholded by a maximum of 10 percent. The minimum read counts per million (CPM) of the mCherry-mTrim21 construct was set to 500 for injected samples. Differential expression analysis was carried out by DESeq2 (version 1.26.0) on read counts per gene calculated by STAR. Results were visualized as MA plots, i.e. log2 fold change in relation to mean counts, using HelpersforDESeq2 package (version 0.1). Adjusted p value (padj) cutoff was set to 0.05. Venn diagrams were created by using the Vennrable package (version 3.1.0.9). For exploratory analysis, normalized counts were log2 transformed after adding a pseudocount of 1 and the transformed values were batch corrected using ComBat from the sva package (version 3.34.0). For PCA, log2 normalized counts were centered by the mean and scaled by the standard deviation.

Analysis of publicly available RNA-seq datasets

Published RNA-seq datasets were obtained from GEO with accession numbers GSE38495,⁴⁷ GSE45719⁴² and GSE98063.⁸⁷ Data processing, read counting and TPM calculations were carried out as described for RNA-seq, except, without Trim21-related modifications of the genome. Early blastocyst cells were further divided to ICM and TE cells by hierarchical clustering on the TPM values of selected marker genes (gene list from Blakeley et al.⁸⁸). Downstream analyses were performed as for RNA-seq.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed by fitting an analysis of variance model on the fraction of embryos reaching blastocyst stage calculated for each experiment using R (version 4.0.2). Multiple comparisons were performed by using the `glht` function (multcomp package, version 1.4–18). Residuals were investigated using the DHARMA package (version 0.4.4). Error bars on [Figures 1G](#), [2D](#), [4E](#), [5F](#), [S4A](#), [S4B](#) and [S4D](#) indicate the standard error of the mean of the biological replicates (s.e.m.). Statistical test on [Figure 4E](#) was performed using analysis of variance followed by a single-step multiple testing procedure. All of the statistical details of experiments can be found in the figure legends.

Cell Reports, Volume 41

Supplemental information

**Distinct patterns of RNA polymerase II and
transcriptional elongation characterize
mammalian genome activation**

Kenichiro Abe, Tamas Schauer, and Maria-Elena Torres-Padilla

Figure S1

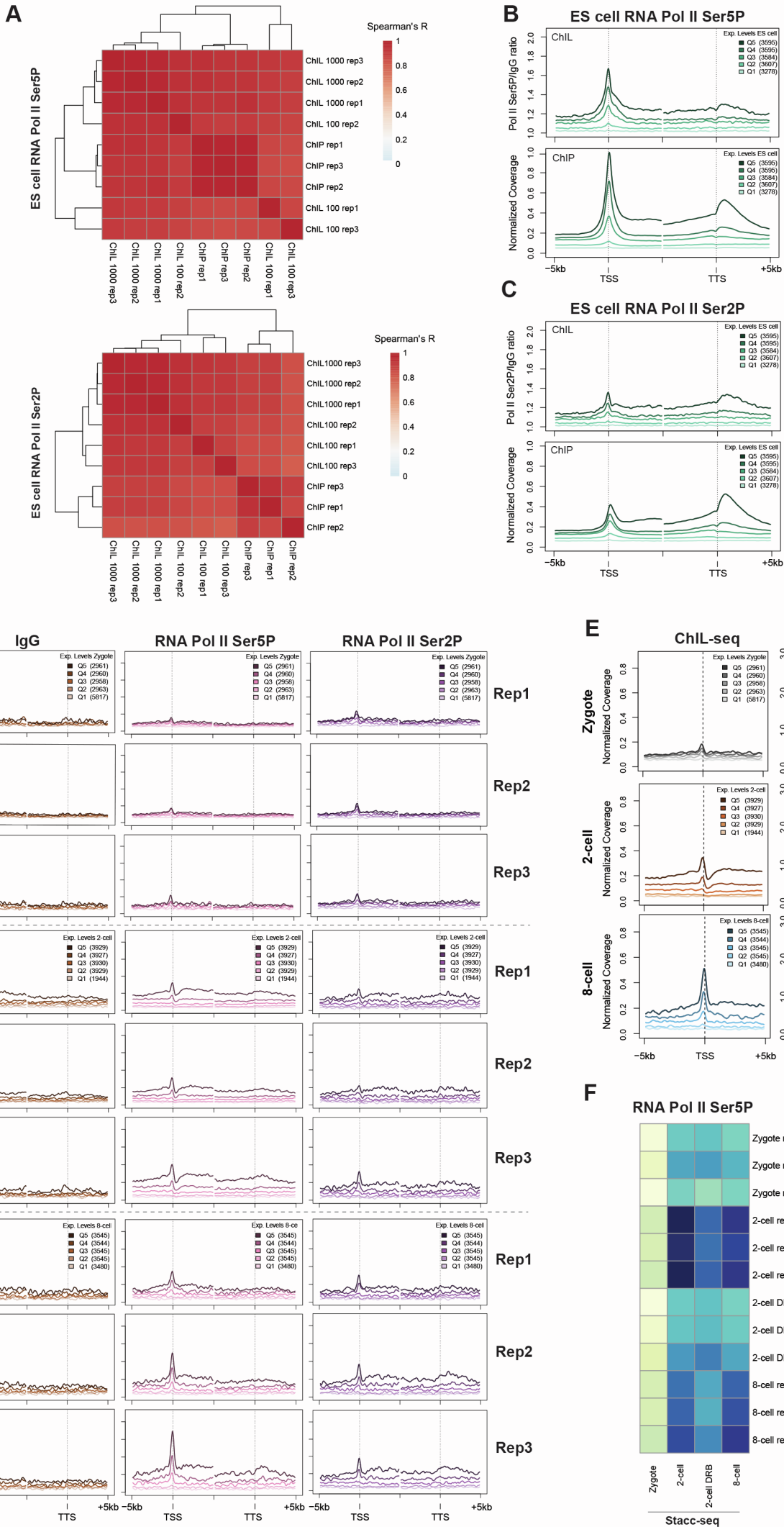


Figure S1. ChIL-seq maps phosphorylated Pol II, related to Figure 1

(A) Correlation heatmaps between Pol II Ser5P (top) or Pol II Ser2P (bottom) ChIL-seq data from 100 or 1000 ES cells and ChIP-seq data of bulk ES cells. Correlations are calculated based on read counts at the gene body from 3 biological replicates.

(B and C) Metaplots comparing Pol II Ser5P (B) and Pol II Ser2P (C) ChIL-seq data of 100 ES cells (top) and ChIP-seq data of bulk ES cells (bottom). Metaplots depict ChIL-seq mean coverage normalized by IgG and ChIP-seq mean coverage centered at TSS and TTS. Maternal genes⁴⁵ were removed and genes were grouped in quintiles based on their expression level⁴⁶. The first quintile includes genes with zero expression. The average of 3 biological replicates is shown.

(D) Metaplots comparing IgG (left), Pol II Ser5P (middle) and Pol II Ser2P (right) ChIL-seq mean coverage centered at TSS and TTS in zygote, 2-cell and 8-cell stage embryos. Three biological replicates are plotted independently. Maternal RNA genes⁴⁵ were removed and genes were grouped in quintiles based on their expression levels⁴⁶. The first quintile includes genes with zero expression.

(E) Metaplots comparing Pol II Ser5P ChIL-seq (left) and published Pol II Ser5P Stacc-seq³⁵ (right) mean coverage centered at TSS in embryos. Maternal RNA genes⁴⁵ were removed and genes were grouped in quintiles based on their expression levels in the given stage⁴⁶. The first quintile includes genes with zero expression. The average of 3 biological replicates of ChIL-seq and 2 biological replicates of Stacc-seq are shown. We note that some differences between the Stacc-seq and ChIL-seq could arise due to slightly different time points at which embryos were collected and the different antibodies for Pol II Ser5P used.

(F) Correlation heatmap in embryo datasets between Pol II Ser5P ChIL-seq and published Stacc-seq generated using an antibody against phosphorylated Ser5P³⁵. Spearman's correlation coefficients were calculated based on gene body counts from on 3 biological replicates of ChIL-seq and 2 biological replicates of Stacc-seq.

Figure S2

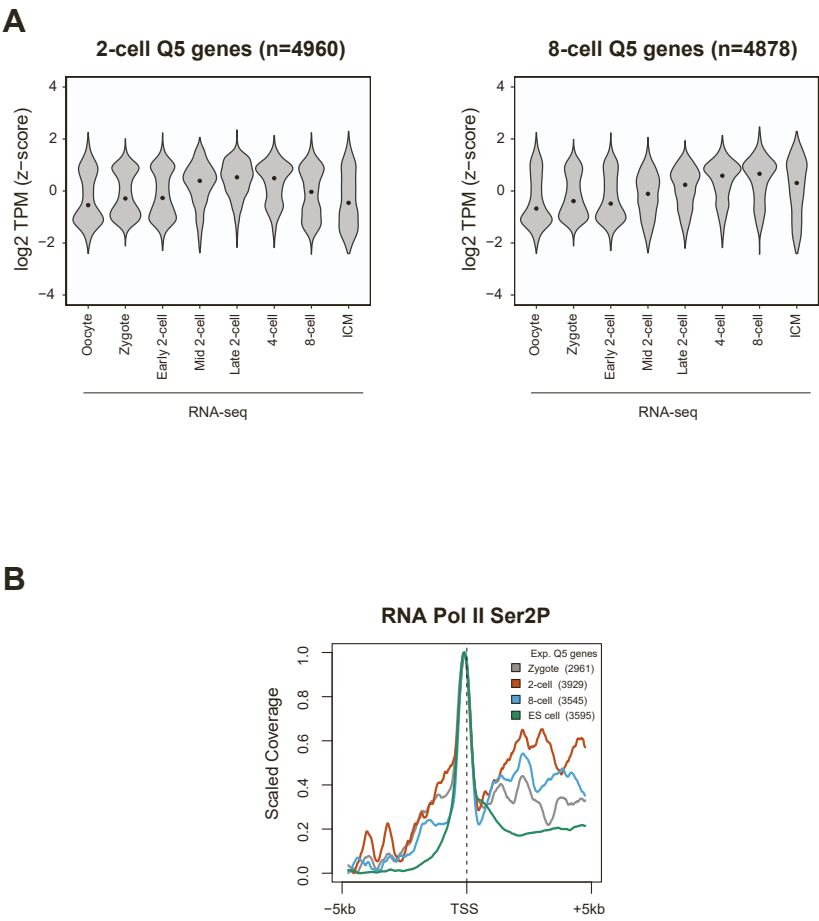


Figure S2. Dynamics of phosphorylated Pol II accumulation during development, related to Figure 2

(A) Violin plots indicating mean expression levels (z-score of transcripts per million [TPM]) of genes in the highest expression quintile of 2-cell (left) or 8-cell (right) stage embryos throughout development.^{46,73} ‘n’ indicates the number of genes.

(B) Scaled metaplots of Pol II Ser2P ChIL-seq mean coverage centered at TSS comparing embryos and ES cells. Maternal RNA genes⁴⁵ were removed and genes in the highest expression quintile are included. The average of 3 biological replicates is shown.

Figure S3

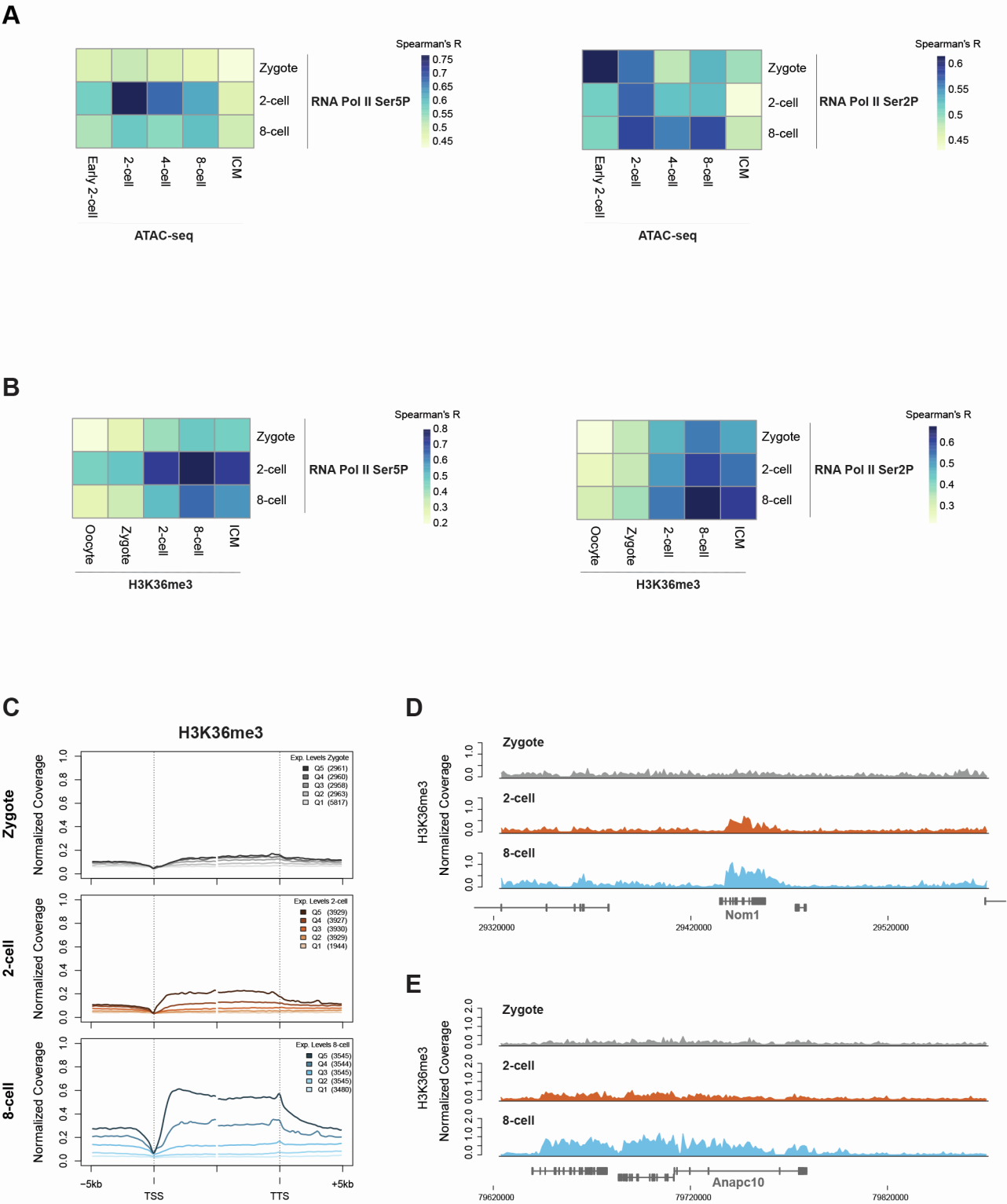


Figure S3. Pol II Ser5 and Ser2 phosphorylation patterns at stage-specific genes, related to Figure 3

(A and B) Correlation heatmaps between Pol II Ser5P (left) or Pol II Ser2P (right) ChIL-seq and ATAC-seq (A) or H3K36me3 ChIP-seq (B) embryonic datasets^{47,76} read counts at the gene body at the indicated developmental stages. Spearman's correlation coefficients were calculated on the mean of 3 biological replicates of ChIL-seq, 2 biological replicates of ATAC-seq or 2 biological replicates of H3K36me3 ChIP-seq.

(C) Metaplots of H3K36me3 ChIP-seq datasets⁴⁷ mean coverage centered at TSS and TTS. Genes were grouped in quintiles based on their expression levels in the given stage⁴⁶. The first quintile includes genes with zero expression. Maternal RNA genes⁴⁵ were removed from the analysis. The average of 2 biological replicates is shown.

(D) Example genomic regions showing H3K36me3 ChIP-seq⁴⁷ coverage in zygote, 2-cell and 8cell stage embryos at an example gene (Nom1) displaying highest Pol II Ser2P across the gene body in 2-cell stage embryos, corresponding to its transcriptional activation. The average of 2 biological replicates is shown.

(E) Example genomic regions showing H3K36me3 ChIP-seq⁴⁷ coverage at an example of '4-cell transient' gene (Anapc10) in zygotes, 2- and 8-cell stage embryos. The average of 2 biological replicates is shown.

Figure S4

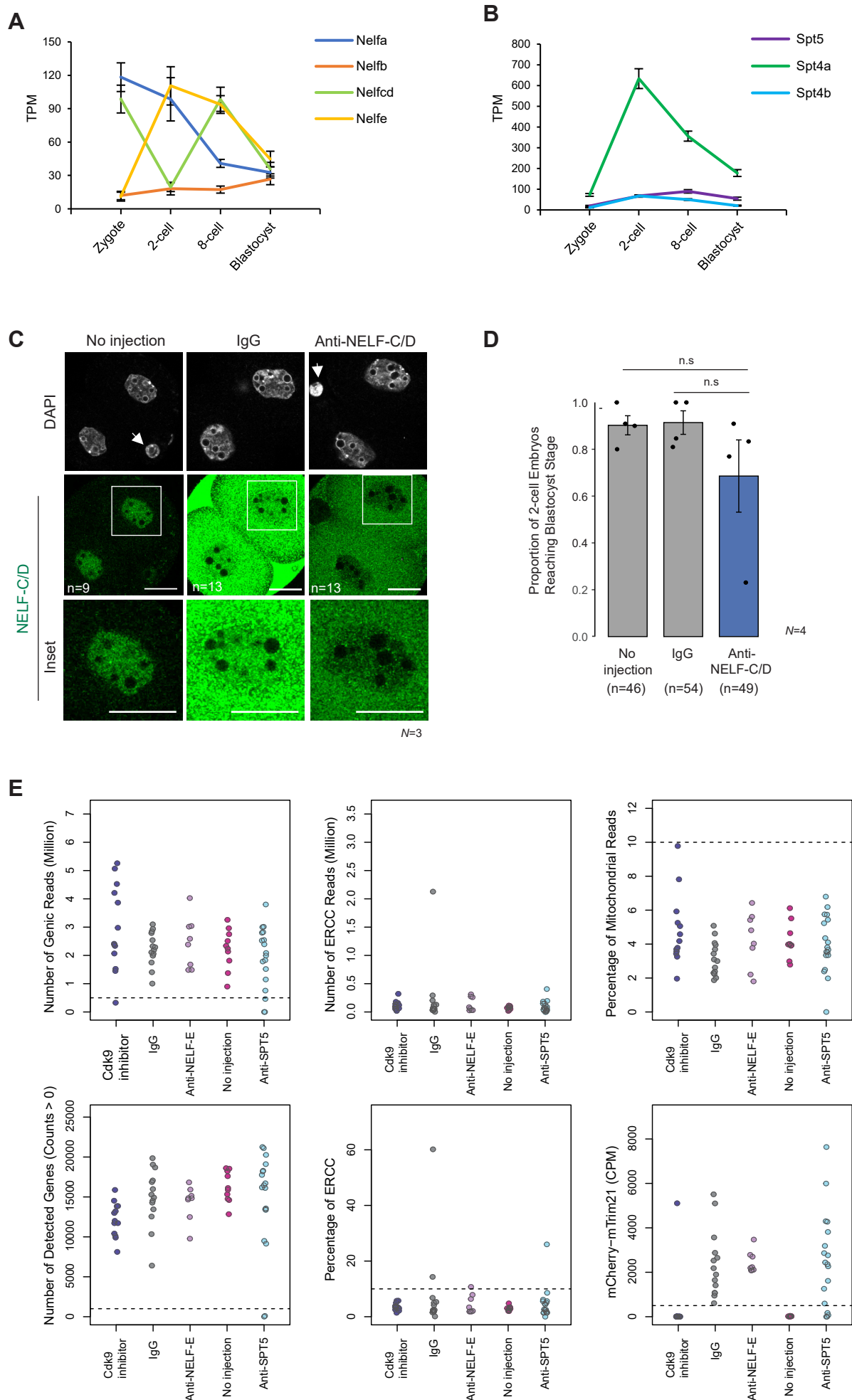


Figure S4. Expression of NELF and DSIF components, Trim-Away for NELF-C/D and quality control analysis of RNA-seq datasets, related to Figure 4

(A and B) Expression levels (transcripts per million [TPM]) of NELF (A) and DSIF subunits (B) from published embryonic RNA-seq datasets at the indicated stages⁴⁶. Error bars indicate mean \pm standard error of the mean (SEM) from single cells.

(C) Immunostaining using anti-NELF-C/D antibody in 2-cell stage embryos after Trim-Away for NELF-C/D or negative controls (Trim-Away for IgG and non-injected embryos). 'n' indicates the total number of embryos analyzed and 'N' the number of independent experiments. Scale bars indicate 50 μ m.

(D) Developmental progression plotted as the proportion of control (non-injected or embryos under Trim-Away against IgG) or Trim-Away NELF-C/D embryos reaching the blastocyst stage. 'n' indicates the total number of embryos analyzed and 'N' the number of independent. Graphs show the mean \pm standard error of the mean (SEM) obtained in individual experiments. P-values were obtained by analysis of variance.

(E) Quality filter of single embryo RNA-seq. Top-left, total sum of read counts per gene (minimum threshold 5×10^5 reads); top-middle, total sum of read counts per ERCC spike-in; top-right, percentage of read counts at mitochondrial genes (maximum threshold 10%); bottom-left, number of detected genes with a read count greater than zero (minimum threshold 1.000 genes); bottom-middle, percentage of read counts at ERCC spike-in (maximum threshold 10%); read counts per million (CPM) aligned to the mCherry-mTrim21 construct (minimum threshold: 500 counts in injected samples). Thresholds are indicated as dashed line. The experiments were conducted 3 times for the collection of non-injected embryos and NELF-E depleted embryos and 5 times for the collection of IgG controls, SPT5 depleted embryos and CDK9 inhibition.

Figure S5

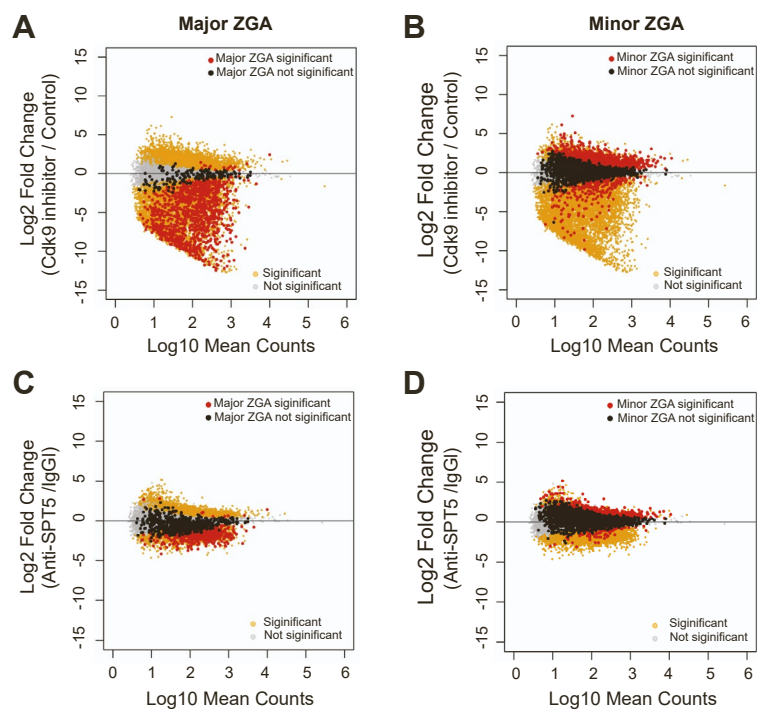


Figure S5. Differentially expressed major or minor ZGA genes upon CDK9 inhibition and SPT5 depletion, related to Figure 5

(A and B) MA plots comparing the log₂ fold change between 2-cell stage embryos treated with BAY1251125 and control, untreated embryos in relation to mean RNA-seq. Differentially expressed major ZGA (A) or minor ZGA (B) gene groups are marked in red ($\text{padj} < 0.05$), non-differential genes within the gene group in black, differentially expressed but not within the corresponding gene group in orange and non-differentially expressed and not in the gene group in grey.

(C and D) MA plots comparing the log₂ fold change between 2-cell stage embryos upon SPT5 depletion by Trim-Away and control embryos injected with IgG in relation to mean RNA-seq. Differentially expressed major ZGA (C) or minor ZGA (D) gene groups are marked in red ($\text{padj} < 0.05$), non-differential genes within the gene group in black, differentially expressed but not within the corresponding gene group in orange and non-differentially expressed and not in the gene group in grey.