Cell Reports
Article

Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression

Bjoern Gaertner,[1,2,5] Jeff Johnston,[1,5] Kai Chen,[1] Nina Wallaschek,[1] Ariel Paulson,[1] Alexander S. Garruss,[1] Karin Gaudenz,[1] Bony De Kumar,[1] Robb Krumlauf,[1,3] and Julia Zeitlinger[1,4,*]
[1]Stowers Institute for Medical Research, Kansas City, MO 64110, USA
[2]PhD Program, University of Freiburg, Faculty of Biology, Schaenzlestrasse 1, D-79104 Freiburg, Germany
[3]Department of Anatomy and Cell Biology
[4]Department of Pathology
University of Kansas Medical Center, Kansas City, KS 66160, USA
[5]These authors contributed equally to this work
*Correspondence: jbz@stowers.org
http://dx.doi.org/10.1016/j.celrep.2012.11.024

SUMMARY

Poised RNA polymerase II (Pol II) is predominantly found at developmental control genes and is thought to allow their rapid and synchronous induction in response to extracellular signals. How the recruitment of poised RNA Pol II is regulated during development is not known. By isolating muscle tissue from *Drosophila* embryos at five stages of differentiation, we show that the recruitment of poised Pol II occurs at many genes de novo and this makes them permissive for future gene expression. A comparison with other tissues shows that these changes are stage specific and not tissue specific. In contrast, Polycomb group repression is tissue specific, and in combination with Pol II (the balanced state) marks genes with highly dynamic expression. This suggests that poised Pol II is temporally regulated and is held in check in a tissue-specific fashion. We compare our data with findings in mammalian embryonic stem cells and discuss a framework for predicting developmental programs on the basis of the chromatin state.

INTRODUCTION

The recruitment of RNA polymerase II (Pol II) has long been thought to be the rate-limiting step for transcription at most genes. However, in recent years it has become clear that at a large fraction of genes, Pol II initiates transcription but then pauses just downstream of the transcription start site (TSS), and that the regulation of Pol II elongation is also a critical step for transcription (Core et al., 2008; Guenther et al., 2007; Muse et al., 2007; Nechaev et al., 2010; Rahl et al., 2010; Zeitlinger et al., 2007). Strikingly, paused Pol II is preferentially found at developmental control genes, suggesting that these genes are frequently regulated at the level of elongation (Muse et al., 2007; Zeitlinger et al., 2007). However, exactly how the interplay of Pol II recruitment and elongation contributes to the regulation of developmental processes is not known.

Evidence so far suggests that paused Pol II helps the rapid and synchronous induction of genes in response to extracellular stimuli. For example, at *Drosophila* heat shock genes, where paused Pol II was originally discovered, gene induction in response to heat shock occurs very rapidly (Boehm et al., 2003; Gilmour and Lis, 1986; Rougvie and Lis, 1988). Furthermore, genes that are paused in the early *Drosophila* embryo tend to be activated in a more synchronous fashion (Boettiger and Levine, 2009). The exact mechanisms by which paused Pol II helps gene induction are not entirely understood. It has been proposed that paused Pol II keeps the promoter in an open state by displacing the promoter nucleosome just upstream of the TSS (Gilchrist et al., 2008, 2010). Furthermore, genes with paused Pol II are transcribed at low levels (Fuda et al., 2009; Zeitlinger et al., 2007), raising the possibility that occasional full-length transcription may also prime genes for activation. Thus, paused Pol II could mediate rapid gene activation directly, or indirectly by establishing a permissive state.

How is Pol II pausing regulated during development? The simplest model is that Pol II pausing occurs by default and thus may represent a transcriptional checkpoint for important, highly regulated genes. Indeed, Pol II pausing could be an intrinsic property of the promoter, because core promoter elements such as Inr, DPE, and PB are highly enriched among genes with Pol II pausing (Gilchrist et al., 2010; Hendrix et al., 2008; Lee et al., 2008; Rach et al., 2009; Rahl et al., 2010). However, there is also evidence that genes can lose paused Pol II and show a closed or inactive promoter state with high nucleosome occupancy (Gilchrist et al., 2010). This raises the possibility that recruitment of paused Pol II is developmentally regulated and that this may occur independently of gene induction. Such a mechanism could render genes either inaccessible or more permissive to activation in certain tissues or developmental stages. Thus, it may represent an additional

developmental checkpoint that ensures precise and robust gene regulation during development (Levine, 2011).

Paused Pol II has frequently been associated with Polycomb group (PcG) repression. Both paused Pol II and PcG proteins are preferentially found at developmental control genes (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006; Nègre et al., 2006; Oktaba et al., 2008; Schwartz et al., 2006; Tolhuis et al., 2006) and have been observed to co-occur (Bracken et al., 2006; Brookes et al., 2012; Enderle et al., 2011; Kharchenko et al., 2011; Lee et al., 2006; Marks et al., 2012; Schwartz et al., 2010), and there is mechanistic evidence that they antagonize each other (Brookes et al., 2012; Chopra et al., 2011; Dellino et al., 2004; Marks et al., 2012; Stock et al., 2007). In *Drosophila*, the co-occurrence of PcG repression and Pol II has been referred to as the balanced state (Schwartz et al., 2010), but its significance for development is unclear.

PcG repression is epigenetically inherited, making it an ideal mechanism for guiding and stabilizing cell fate. A classical example is the repression of Hox genes by PcG complexes, which maintains the segmental identity across the body axis throughout the life cycle of *Drosophila* (Ringrose and Paro, 2007; Schwartz and Pirrotta, 2007). PcG repression also restricts the expression of other important developmental control genes (Oktaba et al., 2008; Pelegri and Lehmann, 1994), but its relationship to paused Pol II is not known.

It is possible that the balanced state is related to the bivalent domain in mouse and human embryonic stem cells (ESCs), which is the co-occurrence of H3K27 trimethylation (H3K27me3) and H3K4me3 near the TSS (Bernstein et al., 2006; Mikkelsen et al., 2007). Bivalent domains are found at higher frequency in ESCs than in differentiated cells and are thought to poise genes for activation during differentiation (Bernstein et al., 2006). However, the universal role of bivalent domains in development has been questioned because they have not been found in either *Drosophila* (Gan et al., 2010; Schuettengruber et al., 2009) or *Xenopus* (Akkers et al., 2009), and even in mouse ESCs they may not be as prevalent as previously thought (Marks et al., 2012).

So far, the role of Pol II pausing and PcG repression in development has not been systematically examined. This is primarily because such studies require a large number of cells from various developmental stages and tissues, and techniques for isolating large quantities of specific cells from embryos have only recently been developed (Bonn et al., 2012; Deal and Henikoff, 2010). Second, measurements of paused Pol II are sensitive to the level of transcription (Lee et al., 2008; Nechaev et al., 2010), which makes it challenging to analyze the role of Pol II pausing during a developmental process in which gene expression is highly regulated.

In this study, we used fluorescence-activated cell sorting (FACS) to isolate muscle cells from *Drosophila* embryos at five time points during development, and analyzed the distribution of Pol II and H3K27me3 across the genome. We specifically focused our analysis on paused Pol II in the absence of significant transcription, a state we refer to as poised Pol II. We found that the set of genes occupied by poised Pol II changes dynamically during development, and that de novo recruitment of poised Pol II is indicative of future gene induction. Interestingly, though, this does not occur in a tissue-specific manner, suggesting that changes in poised Pol II occur globally as a function of developmental time. In contrast, the H3K27me3 mark is tissue specific, suggesting that PcG repression keeps Pol II in check in a tissue-specific fashion. Indeed, the combination of both marks, i.e., the balanced state, is associated with highly dynamic spatial and temporal expression during embryogenesis and is similar to the bivalent domain in mammals.

RESULTS

FACS-Based Isolation of Tissues from *Drosophila* Embryos

To analyze the chromatin state and transcription during the development of specific cell types, we developed a FACS-based method that can be coupled to immunoprecipitation (IP) experiments and messenger RNA (mRNA) isolation followed by deep sequencing (chromatin IP sequencing [ChIP-seq] and mRNA sequencing [mRNA-seq]; Figure 1A). We labeled muscle cells by expressing plasma membrane-targeted green fluorescent protein (GFP) under the control of *mef2-GAL4*, which drives expression in the developing mesoderm as well as in the somatic, visceral, and cardiac musculature starting from embryonic stage 9. This allowed us to sample various developmental stages, encompassing mesoderm subdivision (6–8 hr after egg laying [AEL], myoblast fusion (8–10 hr AEL), terminal differentiation (10–12 hr AEL), and the terminally differentiated musculature (14–17 hr AEL). To examine mesodermal tissue at the time of mesoderm specification and gastrulation (2–4 hr AEL), we used the $Toll^{10b}$ mutant, which produces embryos that consist of only mesodermal precursors (Furlong et al., 2001; Schneider et al., 1991).

For ChIP-seq experiments, embryos were dissociated into single cells, fixed, filtered, and then sorted (Figure 1A and Figure S1). Microscopic examination of sorted cells indicated a purity of >80%–90% (Figure S1B). Furthermore, PolI II binds to muscle-specific genes at the expected stages during our time course (Figure 1B), and GFP-positive versus GFP-negative cells sorted from the same cell suspension show large differences in Pol II binding (Figure S1E), indicating strong enrichment of muscle cells in our sample.

mRNA-seq was performed on live-sorted cells from the same tissues. We find that our mRNA-seq data are highly reproducible ($R^2 = 0.99$ for all samples) and show the expected dynamic regulation of known muscle genes (Figure 1C). Furthermore, using a false discovery rate (FDR) of <0.05 (corrected for multiple testing), we find that the function of upregulated and downregulated genes as determined by the Gene Ontology (GO) function are consistent with the known stages of muscle development (Figure S2A).

De Novo Recruitment of Poised Pol II over Developmental Time

To test whether Pol II pausing is regulated during development, we analyzed the occupancy of Pol II across the muscle time course using ChIP-seq. We used an antibody against the C-terminal domain (CTD) of Pol II (8WG16) in independent

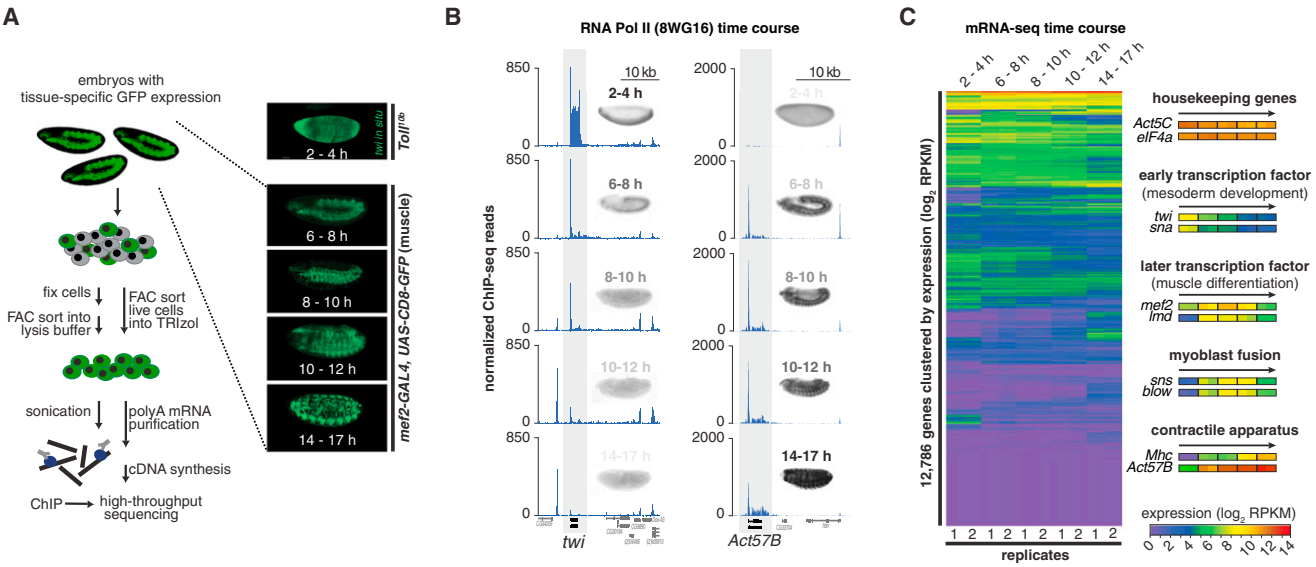

Figure 1. Experimental Strategy for Tissue-Specific Time-Course Analysis of Chromatin and Transcription

(A) Overview of the experimental procedure. Embryos expressing GFP in a tissue of interest are dissociated into single cells. GFP-positive cells are isolated by FACS and analyzed by ChIP-seq and mRNA-seq. Because *mef2-GAL4*-driven GFP expression in the embryonic musculature is only apparent from 6 hr AEL on, earlier events were studied using *Toll*^{10b} mutant embryos, which consist of mesodermal precursor cells (visualized by *twist* in situ hybridization). See also Figure S1.

(B) Genome browser snapshot of the dynamic changes in Pol II occupancy around the *twi* and *Act57B* genes during muscle development. Note that the changes correlate with changes in gene expression observed by in situ hybridization and mRNA-seq (C).

(C) Heatmap of the mRNA-seq data (left). The time-course data from replicate experiments for 12,786 individual genes were clustered by Euclidian distance. The color scale reflects their expression levels shown in RPKM. Based on spike-in mRNA, we estimate an RPKM value of one to correspond to 0.5–1 transcript/cell (see Extended Experimental Procedures). The timing of expression of well-characterized muscle genes (right) is consistent with the function of these genes.

See also Figure S1.

replicate experiments. Experiments with a different Pol II CTD antibody (4H8) gave similar results in our analyses (Figures S2E and S3A).

We previously defined paused Pol II by the pausing index, which is the ratio of Pol II enrichment around the TSS (Pol II$_{TSS}$) versus Pol II enrichment in the transcription unit (TU [Pol II$_{TU}$]; Muse et al., 2007; Zeitlinger et al., 2007; Figure 2A). Because Pol II$_{TU}$ depends on the transcription levels and is subject to noise at very low levels, we focused our analysis on poised Pol II, which we define as high levels of Pol II near the TSS (Pol II$_{TSS}$ in the top 20th percentile of all genes) with transcript levels below 10 reads per kilobase of exon model per million mapped reads (RPKM) as determined by mRNA-seq (because poised genes are transcribed above background; Fuda et al., 2009; Zeitlinger et al., 2007). This preferentially identifies developmental control genes similar to those published previously (Figures S2B and S2C).

To test whether the recruitment of poised Pol II changes during development, we selected all genes that have poised Pol II in at least one time point. Thus, these genes can have paused Pol II with active transcription or be in an inactive state without Pol II at other time points. Although 60% remain bound by Pol II with or without transcription throughout the time course (constant set), 40% are found to be in an inactive state with no Pol II at some point (Figure 2B). Strikingly, most of these genes lack Pol II at the first time point and gradually acquire Pol II promoter occupancy during our time course (opening set, n = 502; Figures 2B and S2E). Only a small fraction of genes lose Pol II occupancy over time (closing set, n = 65; Figures 2B and S2E). The de novo recruitment of poised Pol II also correlates with changes in chromatin accessibility as measured by increased DNase I hypersensitivity (DHS) in a whole-embryo time course (Figure 2C; note that the DHS time course ends at ~11 hr, and thus our last time point with maximum Pol II binding cannot be compared). Thus, for a large fraction of poised genes, the promoter becomes accessible and occupied by Pol II during the course of development.

When poised Pol II is established de novo, does it indicate that these genes are now more likely to be activated? Although this might be expected, it has not been formerly tested. To do so, we used each time point of the RNA-seq data sequentially as a reference time point (gray squares in Figure 2D) and identified all genes that are induced at future time points or were expressed in past time points (with different thresholds giving similar results, see Figure S3). We then asked what fractions of genes are induced among different Pol II groups (Figures 2D and 2E). We found that ~10%–45% of poised genes in the constant set are typically induced in the future. Interestingly, a very similar fraction (but containing mostly different genes) had been expressed in the past (Figure 2D,

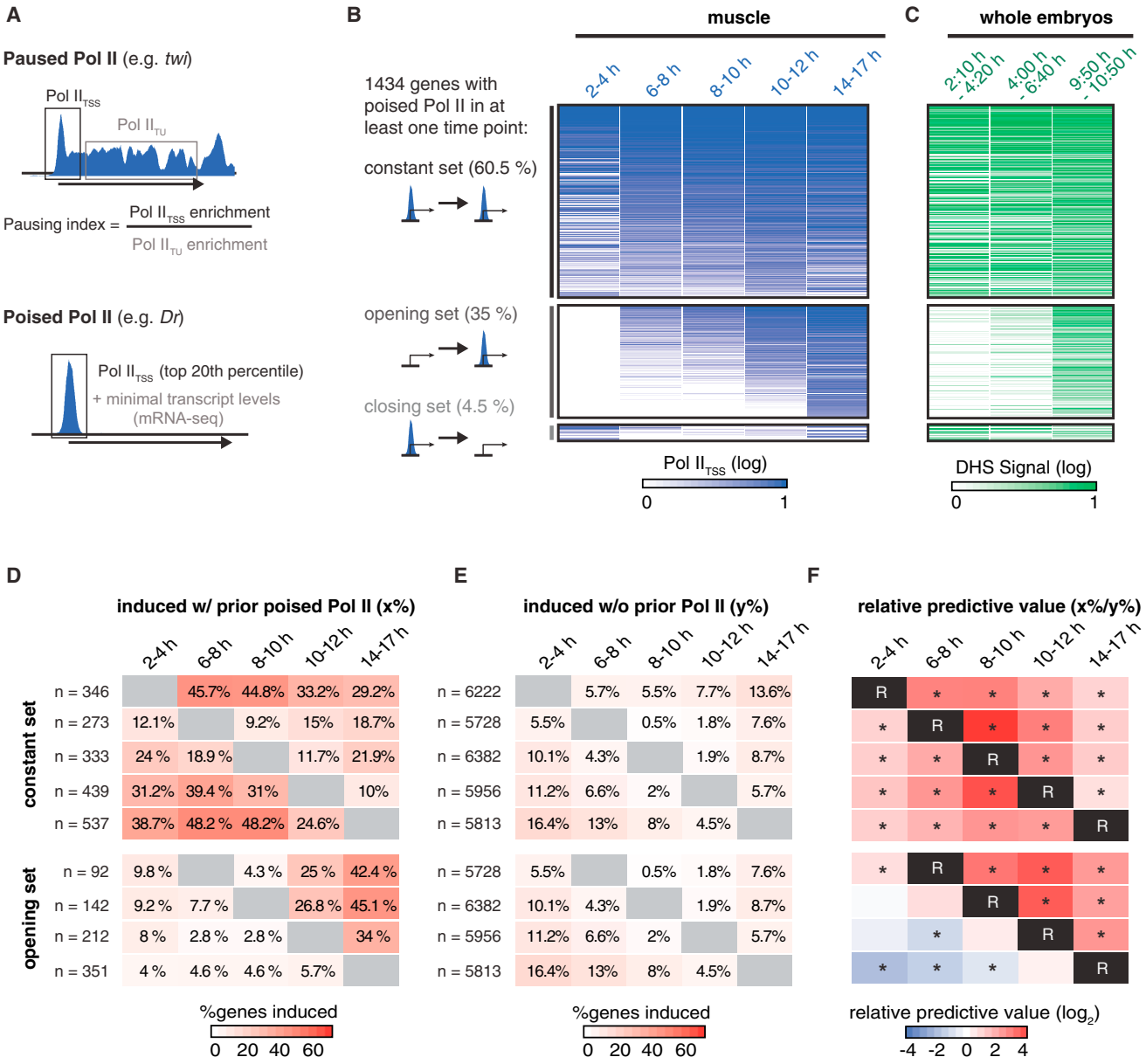

Figure 2. Recruitment of Poised Pol II Is Dynamically Regulated during Development

(A) Definitions of Pol II states. Stalled Pol II is defined by a high ratio of Pol II occupancy at the TSS over the Pol II occupancy in the TU of a gene. Poised Pol II is defined by high Pol II_{TSS} enrichment (top 20th percentile) and low expression as measured by mRNA-seq (RPKM < 10).

(B) Recruitment of poised Pol II is dynamically regulated during the time course. The Pol II occupancy at the TSS is shown across time for all genes (n = 1,434) that have poised Pol II in at least one of the time points of the time course; 60% of these genes (n = 867) remain occupied by Pol II at all times, and 40% switch between being occupied or not occupied by Pol II. Most of the genes that switch states lack Pol II occupancy at the first time point and gradually gain Pol II occupancy during the time course (opening set, n = 502), whereas only few genes (closing set, n = 65) are initially occupied by Pol II and subsequently lose it (0 = no enrichment, 1 = highest enrichment).

(C) DNase hypersensitivity (DHS) data from whole embryos at three time windows (data from Thomas et al., 2011) show increased DNase accessibility over time, consistent with promoter opening and Pol II recruitment (0 = no signal, 1 = highest signal).

(D) Fraction of poised genes induced at another time point (precision rate) for each reference time point (gray box; n is the total number of poised genes). Note that poised genes from the constant set are expressed in both the past and the future, whereas those from the opening set tend to be induced in the future only.

(E) The same calculation as in (D) was done for genes lacking Pol II at the reference time point.

(F) The ratio between the two percentages in (D) and (E) is the relative predictive value, which indicates how much more likely poised genes are to be activated than control genes without Pol II. Asterisks indicate $p < 0.05$.

See also Figure S2.

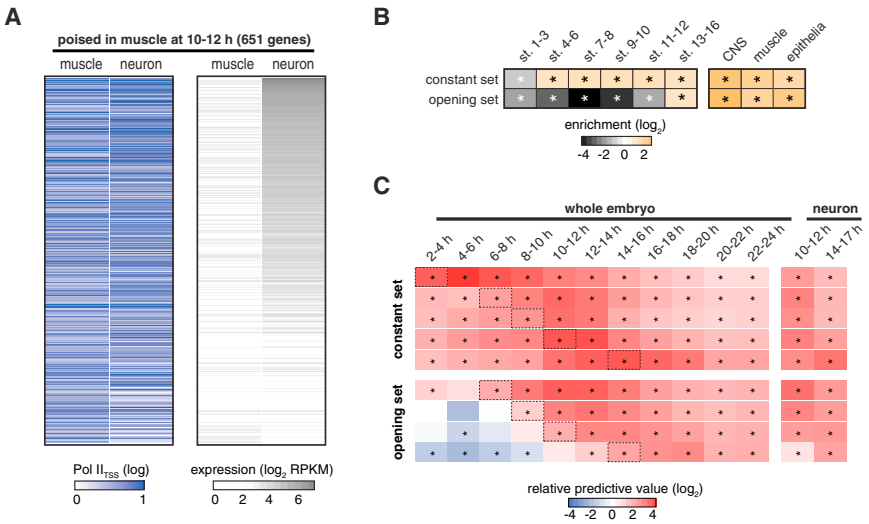

Figure 3. Poised Pol II Predicts Stage-Specific but Not Tissue-Specific Gene Expression

(A) Comparison of normalized Pol II occupancy between 10-12 hr muscle and neurons for genes poised in muscle. All genes poised in the 10–12 hr muscle sample also have Pol II bound in 10–12 hr neurons (left; 0 = no enrichment, 1 = highest enrichment). Many (49%) of the genes that are poised in 10–12 hr muscle are not only bound by Pol II in 10–12 hr neurons but are also expressed (right).

(B) Analysis of in situ expression of the genes of both the constant set and the opening set. As expected, genes in the constant set are enriched for all developmental stages, whereas the opening set genes are expressed late (stages 13–16) during development. Neither gene set is muscle specific, and both sets are also enriched for expression in the central nervous system and epithelial tissues.

(C) Relative predictive values for poised Pol II in the constant set and opening set for tissue-specific samples and whole embryos. Genes with poised Pol II at muscle reference time points are expressed not only in muscle but also in neurons or whole embryo, arguing that the recruitment of Pol II is not tissue specific. Note that the values for the constant set are strongly stage specific during the whole-embryo time course and that the opening set is not expressed in the entire early embryo. For the calculation, see Figure 2. Asterisks indicate significance ($p < 0.05$) and the dashed box emphasizes the reference time point.

See also Figure S3.

top), indicating that poised Pol II can also be a mark for past activation. In contrast, newly poised genes are much more likely to be expressed in the future (~45% versus ~10%, with past expression likely due to maternal transcripts; Figure 2D, bottom), supporting the idea that de novo recruitment of poised Pol II is a mechanism that prepares genes for future activation.

To obtain a more general measurement of the activity of gene groups, we refined our method. So far, the fraction of induced genes in each group varies and depends on the total number of genes induced, which in turn increases over developmental time (Figures 2D and 2E). To normalize, we defined the large number of genes with Pol II levels at or below background as control genes, and calculated the ratio between induced genes in the test set (poised Pol II) over control genes (no Pol II; Figures 2E and 2F). We call this normalized measurement the relative predictive value. At most time points, the fraction of induced genes among those with prior poised Pol II is significantly higher than the fraction of those without prior Pol II (shown in red in Figure 2E), with the highest values typically found near the reference sample. Only poised genes in the opening set are less likely to have been expressed in the past as compared with control genes (shown in blue in Figure 2E), and this overall pattern is robust for a variety of thresholds for identifying a poised gene and its activation (Figure S3B). This suggests that when genes switch from no Pol II to poised Pol II, their likelihood of activation becomes significantly higher.

Because on average poised genes tend to be expressed at higher levels than genes with no Pol II, we also used control genes with transcription levels similar to those of poised genes (Figure S3C). We found that the overall pattern of the predictive values for poised Pol II was still similar. Although this does not rule out the possibility that the permissive state associated with poised Pol II is in part mediated by low levels of transcription, it argues that low levels of transcripts per se do not have the same relative predictive value for future gene expression as poised Pol II itself.

Poised Pol II Marks Stage-Specific but Not Tissue-Specific Gene Expression

We next analyzed whether the recruitment of poised Pol II is tissue specific, but, surprisingly, found no evidence of this. First, we determined Pol II occupancy in differentiated neuronal tissue by sorting GFP-positive cells (GFP driven by *elav-GAL4*; Figure S1D). This showed that all genes that are poised in muscle cells have detectable levels of Pol II in neurons, and a large fraction (49%) of these genes are active in neurons (Figure 3A). Second, based on the large-scale in situ hybridization database ImaGO (Tomancak et al., 2007), the opening set of genes identified in our muscle time course are indeed expressed late in embryogenesis, but they are expressed in various tissue types, suggesting that Pol II is also recruited to these genes in many other tissues (Figure 3B).

Finally, when we analyzed the relative predictive value of poised genes using the method described in Figure 2F, we also found that poised genes in muscle, whether in the constant set or opening set, are frequently expressed in neuronal cells or whole embryos (data from Graveley et al., 2011; Figure 3C). Furthermore, the expression of the opening set is also restricted to later expression in whole embryos, consistent with the hypothesis that Pol II is recruited de novo throughout the embryo.

This suggests a model in which poised Pol II is dynamically recruited to genes over time, and these genes are then induced in a tissue-specific fashion. This explains why not all poised genes are induced in a particular tissue. For example, only ~50% of all poised genes are expressed during the entire muscle time course, whereas this cumulative percentage increases to ~70% when the expression data from neuronal cells and whole embryos are included (Figure S3).

Three Classes of Promoters Are Used during Embryogenesis

To test how promoter elements determine the dynamics of Pol II occupancy during development, we analyzed the core promoter elements in all of our gene groups. Studies so far have analyzed highly paused versus less-paused genes (Gilchrist et al., 2010; Hendrix et al., 2008; Lee et al., 2008), but whether this difference corresponds to focused and dispersed transcription (Rach et al., 2009) is not clear.

Here, we identified three promoter classes (Figure 4). First, so-called housekeeping genes, which are broadly expressed in the embryo (Tomancak et al., 2007), have dispersed promoter elements as previously shown (Rach et al., 2009). Second, we find that genes that are poised at any time point (constant set or opening set) are all highly enriched in promoter elements previously associated with Pol II stalling (GAGA, Inr, DPE, PB, and MTE). This suggests that these elements predispose genes for the recruitment of poised Pol II, but do not do so by default. Third, we find that genes that are induced without prior poised Pol II fall into a third class of promoters that are enriched for Inr and the TATA box. TATA-enriched promoters were previously identified as a separate class of promoters that are associated with cell-type-specific gene expression in adult somatic tissues (Engström et al., 2007). Thus, our results corroborate the notion that TATA-enriched promoters are a separate class, and suggest that these promoters do not require recruitment of poised Pol II prior to induction.

Because paused Pol II has been associated with a strong promoter nucleosome in the absence of transcription (Gilchrist et al., 2010), we analyzed the nucleosome organization in the three classes of promoters by performing micrococcal nuclease (MNase) treatment and paired-end sequencing at the first and last time points of the muscle time course (Figure 4B). We found that poised genes indeed show a strong promoter nucleosome when Pol II is not present at the first time point, whereas promoters occupied by poised Pol II are depleted for the promoter nucleosome. This difference is not intrinsic to the DNA sequence, because both sets of genes show similar predicted promoter nucleosome occupancy. In contrast, housekeeping genes or TATA-enriched genes do not have a strong promoter nucleosome, and the profile looks similar regardless of whether the genes are active or inactive. However, housekeeping genes were distinct from TATA-enriched genes in that the nucleosome occupancy at the first nucleosome was significantly higher. These results show that there are three distinct promoter classes at the level of nucleosome organization.

PcG Repression Is Tissue Specific

To analyze the role PcG repression, we mapped the genome-wide profile of H3K27me3 at all time points of muscle development, as well as in differentiated neuronal cells. We did not map PcG proteins directly, because at the well-characterized *Ubx* gene in *Drosophila*, PcG proteins bind independently of whether the gene is repressed or active (Papp and Müller, 2006), suggesting that PcG protein occupancy alone may not be a good indicator of PcG repression. On the other hand, the presence of H3K27me3 on the TU of genes has been found to correlate well with PcG repression (Papp and Müller, 2006; Schwartz et al., 2006).

We found that genes that are differentially marked by H3K27me3 were preferentially expressed in either muscle or nervous system based on mRNA-seq expression levels ($p < 0.02$, Scheirer-Ray-Hare test; Figure 5A) or whole-embryo in situ hybridizations ($p < 0.027$, Fisher's exact test; Figure 5B). An example is the *twist* gene, which shows high H3K27me3 levels across the TU in neuronal cells but lower levels in muscle cells (Figure 5C). Conversely, the *shaven* gene has high H3K27me3 levels in muscle cells but lower levels in neuronal cells (Figure 5C). Note that H3K27me3 is not completely absent in the other cell type, which is likely due to the segmentally modulated expression of *twist*, *shaven*, and many other developmental control genes. Thus, even if a PcG-regulated gene is active in muscle cells, it is rarely expressed in all cells of this tissue.

Next, we analyzed H3K27me3 across the muscle time course. We found that genes with H3K27me3 at the TU are less likely to be induced at future time points (blue in Figure 5D). However, unlike the predictions of poised Pol II, the predictions of H3K27me3 are tissue specific. The set of genes that are highly occupied by H3K27me3 in muscle cells does not negatively predict gene expression in neuronal cells of the same stage or the entire embryo (Figure 5D). This suggests that the gene set with PcG repression is tissue specific and tends to be maintained during *Drosophila* embryogenesis.

The Balanced State Correlates with Highly Dynamic Regulation

We then analyzed the co-occurrence of Pol II binding and H3K27me3, which defines the balanced state. For this purpose, we performed sequential ChIP (reChIP) analysis with chromatin from early wild-type embryos (2–4 hr AEL), using antibodies against H3K27me3 and then Pol II. The enrichment over input for single ChIPs and reChIPs was calculated following quantitative PCR (qPCR; Figures 6A and S4A) or deep sequencing (Figures 6B and S4B), after normalization to an intergenic control region or total read counts, respectively. An increase in enrichment from the first ChIP to the reChIP indicates some degree of co-occupancy, while equal enrichment or less is expected if the two antigens are mutually exclusive (Geisberg and Struhl, 2004a; see Extended Discussion). Indeed, we found that genes with Pol II and H3K27me3 enrichment in single ChIPs, but not genes with either H3K27me3 or Pol II enrichment only, showed higher enrichment after reChIP as compared with the first ChIP (Figure 6A). This effect increased with higher Pol II enrichment in single ChIPs and was statistically significant

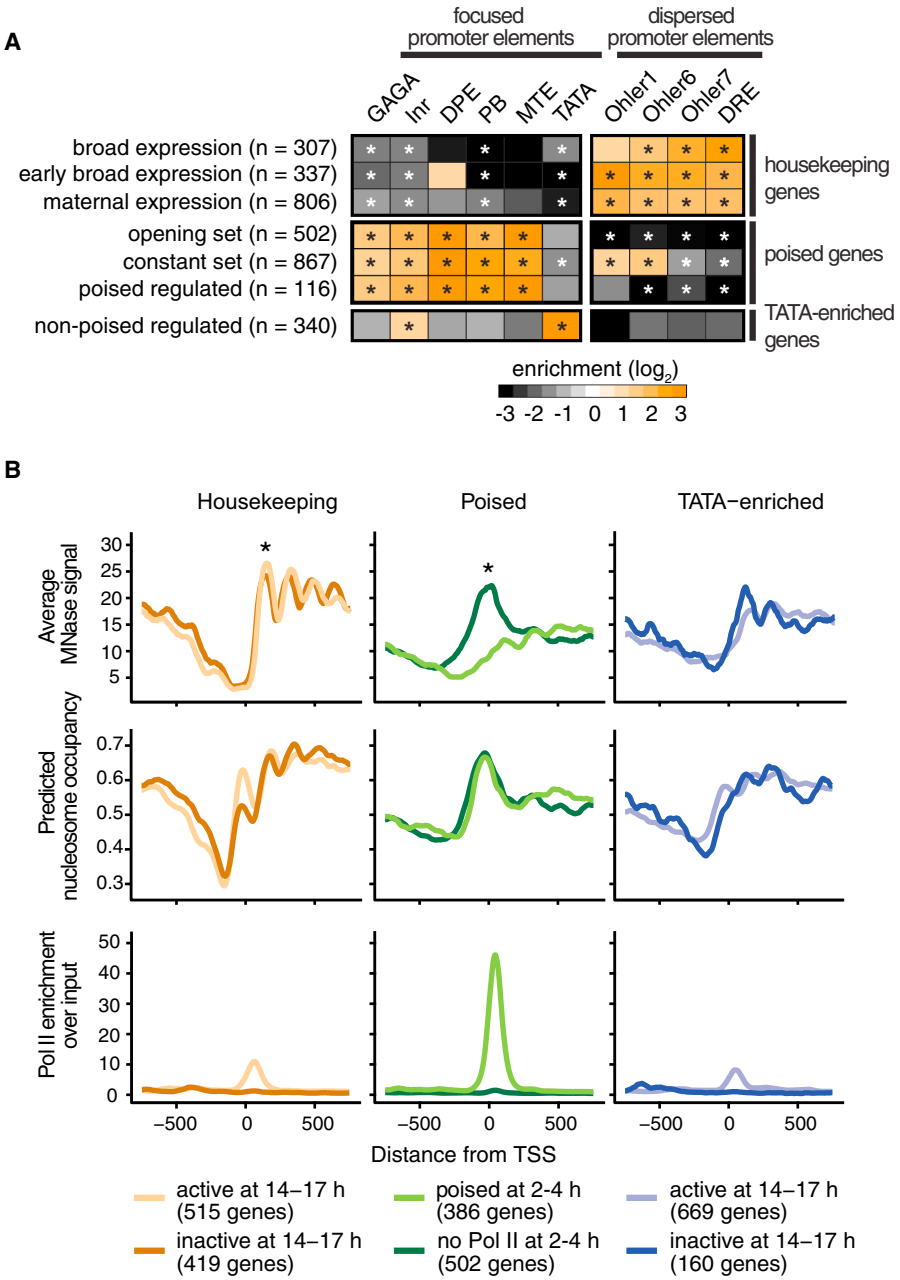

Figure 4. Different Core Promoters Are Associated with Distinct Pol II Occupancy Behavior

(A) Enrichment of core promoter elements in different gene groups. The asterisk indicates that the enrichment (yellow) or depletion (black) is significant ($p < 0.05$). First, housekeeping genes, as defined by broad expression throughout the embryo based on in situ hybridizations (Tomancak et al., 2007), are enriched for Ohler1, Ohler6, Ohler7, and DRE, which are found at dispersed promoters. Note that maternally expressed genes count here as housekeeping genes although they may not be expressed in the embryo. Second, genes that have the disposition for poised Pol II (opening set and constant set) are enriched for GAGA, Inr, DPE, PB, and MTE, which are found at focused promoters. The constant set is also enriched for some dispersed promoter elements, perhaps because of its higher average expression (not shown). The poised regulated and nonpoised regulated sets comprise genes that were induced at the last time point with or without prior poised Pol II. Third, the nonpoised regulated genes are enriched for Inr and TATA and thus have a different core promoter configuration that supports focused transcription. The TATA-enriched genes are depleted for housekeeping and developmental functions (not shown).

(B) The average nucleosome profile (top) and predicted nucleosome occupancy (middle) differ among the three promoter classes. For each class, the nucleosome profile as measured by MNase-seq was analyzed in the presence and absence of Pol II (shown at bottom). Only genes with poised Pol II tend to have a prominent promoter nucleosome in the absence of Pol II but not when Pol II is present (*$p < 10^{-48}$ with Wilcoxon rank sum test at +16 bp). Furthermore, housekeeping genes tend to have higher occupancy at the first nucleosome (asterisk, Wilcoxon rank sum test at +151 bp) than poised genes ($p < 10^{-40}$) or TATA-enriched genes ($p < 10^{-15}$). The average predicted nucleosome occupancy for each gene group was calculated based on Kaplan et al. (2009).

See also Table S2.

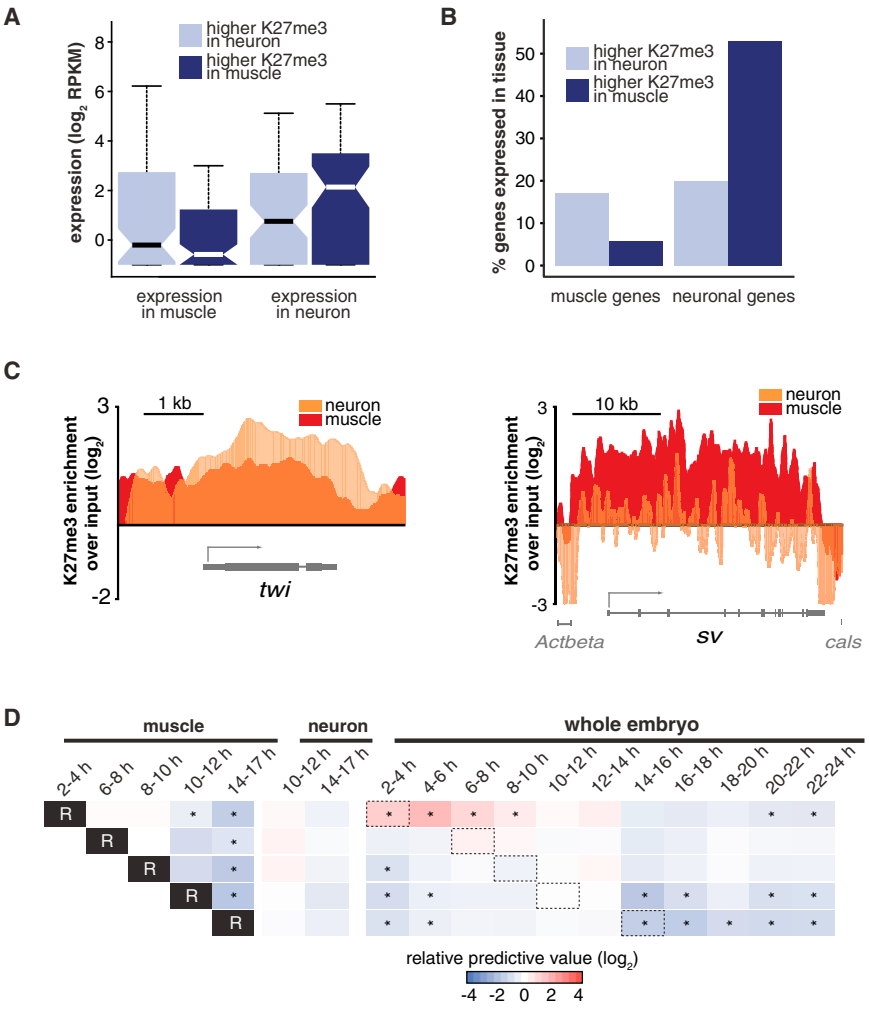

Figure 5. PcG Repression Predicts Tissue-Specific Repression

(A) Genes with differential H3K27me3 levels between muscle and neurons are differentially expressed between these tissues. mRNA-seq data from muscle cells or neuronal cells at 14–17 hr are shown as box plots (log2 RPKM) with whiskers as interquartile ranges. Genes that have either higher H3K27me3 in muscle (dark blue) or neurons (light blue) are differentially expressed (Scheirer-Ray-Hare test, $p < 0.018$).

(B) Based on in situ hybridizations, the differential H3K27me3 sets overlap significantly (*y* axis shows percent overlap) with genes that are expressed either in muscle only or neurons only at any time point during embryogenesis (Fisher's exact test, $p < 0.02$).

(C) Example of genes with differential H3K27me3 levels between muscle and neurons: *twist* (*twi*) has higher H3K27me3 enrichment in neurons (left), and *shaven* (*sv*) has higher H3K27me3 enrichment in muscle (right). Enrichment is shown as H3K27me3 reads over input reads, smoothened over 100 bp windows.

(D) In muscle cells, genes with high levels of H3K27me3 (top 2.5% of all genes) over the TU are less likely to be induced in the future as compared with genes without H3K27me3 (left). Such negative predictive values (blue) are not found when gene expression is predicted in neuronal tissue (middle) or the whole embryo (right), indicating that the repression is tissue specific. Asterisks indicate $p < 0.05$; dashed outlined boxes highlight the reference time points.

across all genes with H3K27me3 ($p < 10^{-32}$; Figure 6B). In contrast, increased reChIP enrichment was not observed with either control antibodies (FLAG) or H3K4me3 in the second ChIP, consistent with previous evidence arguing against the bivalent domain in *Drosophila* (Gan et al., 2010; Schuettengruber et al., 2009).

This suggests that H3K27me3 and Pol II co-occur to some degree at many genes. Although it is possible that Pol II occupancy levels are reduced upon PcG repression (see Extended Discussion, Figure 6D, and below), our data argue against the possibility that the balanced state is the result of mixed populations of cells. This is also consistent with reChIP experiments in human ESCs indicating the co-occurrence of a form of Pol II and PcG components (Brookes et al., 2012).

We next examined the relationship between the two marks over time. The overlap between genes with high Pol II and high H3K27me3 is highest at the first time point of our series (29.7% of all H3K27me3-marked genes) and decreases during later developmental stages (to 12.5%). This result is similar to observations on the bivalent domain in mammalian ESCs (Bernstein et al., 2006; Lee et al., 2006; Mikkelsen et al., 2007). Furthermore, although a large number of genes maintain both Pol II and H3K27me3 throughout the time course (cluster 1 in Figure 6D), many genes that are initially balanced lose Pol II, H3K27me3, or both over time (clusters 2–4 in Figure 6D). In fact, PcG-repressed genes significantly overlap with the closing set in Figure 2B ($p < 10^{-5}$; Fisher exact test), supporting the idea that PcG repression can reduce Pol II occupancy over time (Chopra et al., 2011; Dellino et al., 2004).

We also analyzed how Pol II and H3K27me3 occupancy at balanced genes correlates with gene expression (Figure 6D). The presence of Pol II correlates with higher expression levels, whereas the presence of H3K27me3 correlates with lower expression levels. Indeed, genes in the balanced state are expressed at low levels and are often poised. This supports the antagonistic relationship between Pol II and H3K27me3, hence the term "balanced state" is appropriate.

To test whether the balanced state confers specific dynamic expression properties, we analyzed the expression of balanced genes based on in situ hybridization data. We found that 68% of the balanced genes belong to a previously identified group of genes referred to as Blastoderm Patterning genes ($p < 10^{-23}$; Fisher's exact test), which are characterized by highly dynamic expression patterns from the blastoderm stage

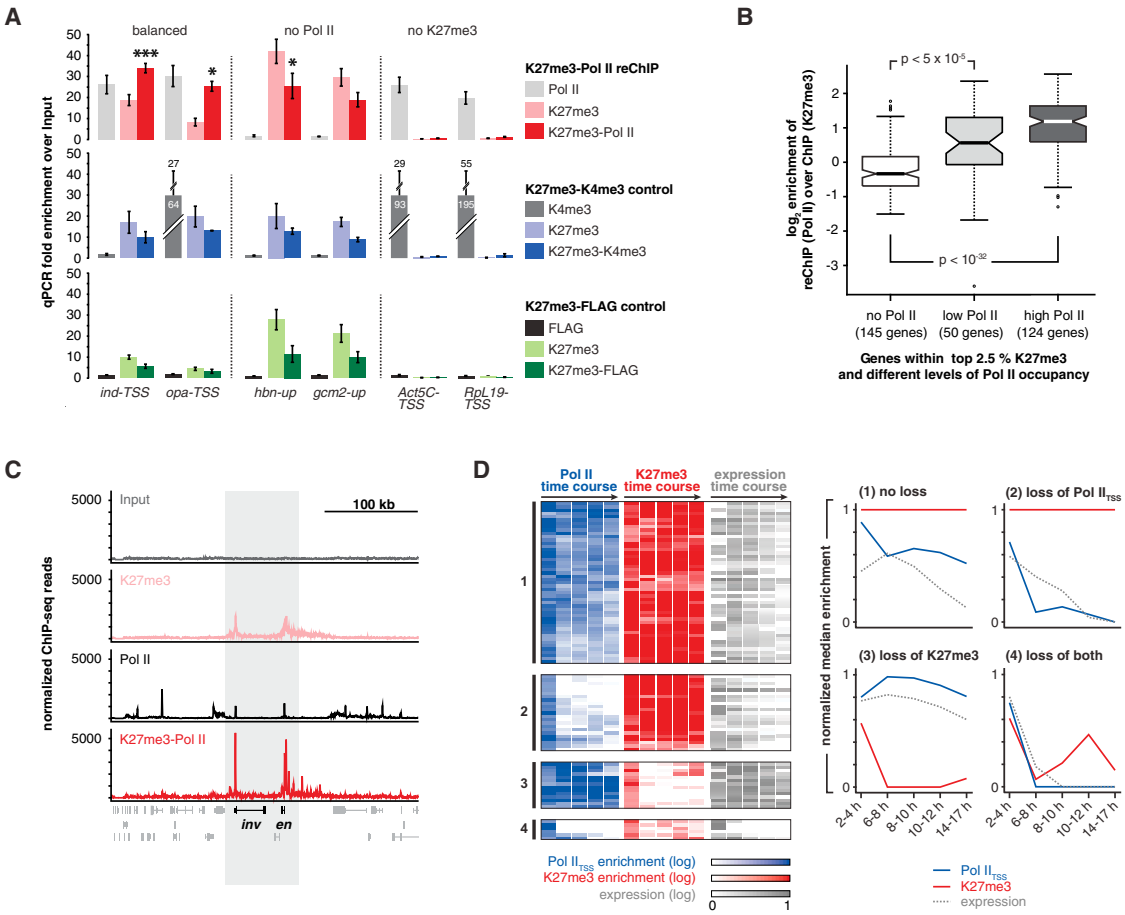

Figure 6. Behavior of the Balanced State during *Drosophila* Embryogenesis

(A) Sequential ChIPs against H3K27me3 and then Pol II confirm the co-occurrence of Pol II and H3K27me3 at genes (top panel). Co-occurrence results in higher reChIP enrichment compared to the first ChIP (measured over input and normalized to an intergenic control region). Control regions that are enriched for Pol II but lack H3K27me3 (*Act5C-TSS* and *RpL19-TSS*) show no significant enrichment in either ChIP or reChIP. Regions with H3K27me3 enrichment that lack Pol II (*hbn-up* and *gcm2-up*) show a decrease after the Pol II reChIP. At balanced genes (*opa-TSS* and *ind-TSS*), the reChIP enrichment is increased relative to the K27me3 ChIP enrichment. In contrast, an increase is not observed using either H3K4me3 (middle panel) or FLAG antibody (bottom panel) for the reChIP. Means for two to seven independent biological replicates are shown; error bars refer to the SEM. Sequential ChIPs were performed in wild-type embryos at 2–4 hr AEL because the assay requires large amounts of cells but not tissue homogeneity. Asterisks indicate $p < 0.05$; triple asterisk indicates $p < 0.001$ (t test).

(B) To analyze the global co-occupancy of Pol II and H3K27me3, we sequenced the single H3K27me3 ChIP and the H3K27me3-Pol II reChIP. The results confirm at a genome-wide level that the additional enrichment at the TSS of genes after the second ChIP (*y* axis = $\log_2$ Pol II ChIP – $\log_2$ H3K27me3 ChIP) strongly depends on whether Pol II is present at the gene (classified as no Pol II, low Pol II, and high Pol II based on an individual Pol II ChIP). Shown is a box plot with whiskers representing the interquartile ranges and circles showing outliers. The vast majority of genes with high Pol II show enrichment after the reChIP ($p < 10^{-32}$, t test).

(C) Genome browser snapshot of an ~320 kb genomic region encompassing the balanced genes *inv* and *en* (gray box).

(D) Time-course analysis of genes that have both Pol II and H3K27me3 enrichment at the first time point. The heatmap (left) represents the relative enrichments of Pol II and K27me3 as well as their respective relative expression levels (0 = no enrichment/expression, 1 = highest enrichment/expression). The line graphs (right) show the median levels for each of the four clusters in the heatmap. Note that the expression levels decrease over time and correlate positively with Pol II occupancy and negatively with H3K27me3 enrichment.

See also Figure S4 and Table S3.

onward (Tomancak et al., 2007). In comparison, genes selected by the presence of only H3K27me3 show less enrichment (35%; $p < 10^{-12}$; Fisher's exact test). Interestingly, many of these expression patterns are more dynamic than those of Hox genes and are not restricted to specific lineages. Thus, the balanced state marks genes with highly dynamic regulation. This suggests that PcG regulation keeps poised Pol II in check, and that this repression can be overcome in a tissue-specific fashion.

Balanced State in Mouse ESCs

Because the behavior of the balanced state during *Drosophila* embryogenesis is reminiscent of the bivalent state in mammalian ESCs, we investigated why the balanced state in

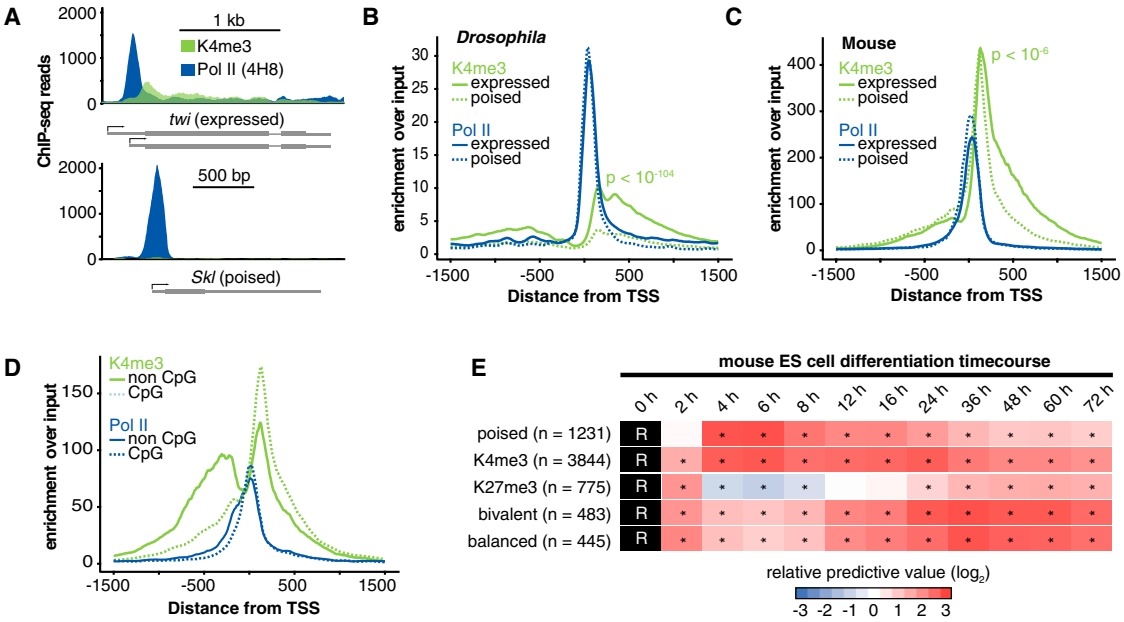

Figure 7. The Balanced State Is Similar to the Bivalent Domain in Mammals

(A–D) The lack of bivalent domains in *Drosophila* can be explained by the absence of detectable levels of H3K4me3 at poised genes.

(A) Significant H3K4me3 levels are detected at paused and expressed genes (*twi* example, top) but not at poised genes (*Skl* example, bottom).

(B) Average gene analysis of Pol II (blue) and H3K4me3 (green) at expressed (solid lines) or poised (dashed lines) genes in 10–12 hr *Drosophila* muscle cells confirms that poised genes in *Drosophila* lack significant levels of H3K4me3 ($p < 10^{-104}$ for H3K4me3 enrichment between expressed and poised genes at base position +163; Wilcoxon rank sum test).

(C) In mouse ESCs, an average gene analysis of Pol II (blue) and H3K4me3 (green) at expressed (solid lines) or poised (dashed lines) genes shows that poised mouse genes have high levels of H3K4me3 ($p < 10^{-6}$ for H3K4me3 enrichment between expressed and poised genes at base position +130; Wilcoxon rank sum test).

(D) The presence of CpG islands in mouse promoters does not explain the high levels of H3K4me3 at poised genes, as shown by the average gene analysis of Pol II (blue) and H3K4me3 (green) at poised mouse genes with (solid lines) or without (dashed lines) CpG island promoters.

(E) The balanced state behaves similarly to the bivalent state during mouse ESC differentiation induced by retinoic acid. The relative predictive values for future gene expression were calculated as described in Figure 2F. Both poised Pol II and H3K4me3 are positive predictors for future gene expression, but there are more genes with H3K4me3 and the predictions for poised Pol II are more stage specific. H3K27me3 is a negative predictor until ~8 hr and a positive predictor starting at ~24 hr. Balanced genes (Pol II above background and H3K27me3, n = 445) and bivalent genes (H3K4me3 and H3K27me3, n = 483) overlap highly (n = 340). See also Figure S5.

Drosophila is not associated with H3K4me3. We found that genes with poised Pol II do not have significant levels of H3K4me3 (Figures 7A and 7B). Only genes that are transcribed (but with similar Pol II_{TSS} occupancy) show H3K4me3 (Figures 7A and 7B).

This is in contrast to mouse ESCs, where the H3K4me3 signal is higher relative to Pol II (using the same antibodies in the two species), and genes with poised Pol II have high levels of H3K4me3 (Figure 7C). The H3K4me3 signal at poised genes is more narrowly distributed but is almost as high as it is at highly transcribed genes. The high levels of H3K4me3 at poised genes in mammals cannot be explained by the presence of CpG islands, because even promoters that are not within CpG islands show significant H3K4me3 levels (Figure 7D). Although the exact mechanisms that explain this species-specific difference remain to be shown, we found evidence that the lack of H3K4me3 at poised genes in *Drosophila* is due to their low nucleosome occupancy and higher nucleosome turnover (Figure S5).

Finally, to compare the dynamic behavior of balanced and bivalent genes during ESC differentiation, we analyzed published Pol II, H3K4me3, and H3K27me3 data in mouse ESCs and performed an extended time-course expression analysis in response to retinoic acid treatment (Lin et al., 2011). We found that the bivalent state is overall more frequent than the balanced state, but after adjustment of the analysis thresholds (Extended Experimental Procedures), the balanced genes and bivalent genes largely overlap and show a similar behavior in our analysis (Figure 7E).

For both poised Pol II and H3K4me3, the relative predictive values for future gene expression are high. Poised Pol II may be more stage specific, because the values are highest at time points just after the reference sample, whereas the values for H3K4me3 are high throughout the time course (Figure 7E). In combination with H3K27me3, though, poised Pol II and H3K4me3 behave very similarly, i.e., they tend to mark genes with late expression. This supports the hypothesis that the balanced state and the bivalent domain are in

principle related, and that differences in the relative levels of Pol II and H3K4me3 (and perhaps other regulatory differences) explain why the bivalent domain was discovered in mammals, while the balanced state was first described in *Drosophila*.

DISCUSSION

Regulation of Poised Pol II at the Level of Recruitment

We find that poised Pol II is frequently recruited to promoters de novo over developmental time, and that this recruitment helps establish a permissive state that can enable future activation. The mechanisms by which poised Pol II is recruited de novo are not known. Although it could be mediated by sequence-specific transcription factors, the transcription factors examined in vivo so far appear to affect both the recruitment and elongation of Pol II (Boehm et al., 2003), or may even preferentially regulate Pol II elongation (Rahl et al., 2010). It is also possible that the recruitment of poised Pol II is regulated at the level of the chromatin state, e.g., changes in the boundaries of heterochromatin could affect promoter accessibility.

It is clear, however, that not all genes that are poised will be expressed in these cells in the near future. Thus, the poised Pol II state is not simply an early sign of gene activation. This makes sense because the recruitment of poised Pol II is not tissue specific and thus the cell may not receive the appropriate developmental or environmental signal to activate a poised gene. Furthermore, poised Pol II can persist for some time after a gene is downregulated and marks past activation.

Developmental Implications

Regulation of a permissive state over developmental time has developmental implications. First, cells of a developing tissue sometimes have a time window in which they are competent to respond to certain signals (Pearson and Doe, 2004; Tran and Doe, 2008). Thus, changes in poised Pol II might alter the way a cell responds to extracellular signals over time. Second, it may be important during pattern formation that a wide range of cells are able to respond to activating signals such as morphogen gradients, although only a subset will receive sufficient signal to activate the appropriate genes. Because we find that the poised state is also present in mouse ESCs and predicts stage-specific gene expression, it is possible that the role of poised Pol II in development reflects a broadly conserved feature of animal development.

Induction of Genes without Prior Poised Pol II

Although much of our work focused on poised Pol II, we identified a significant number of genes that are induced without prior poised Pol II, consistent with previous studies (Gilchrist et al., 2012; Lin et al., 2011). Remarkably, these genes tend to have a distinct combination of core promoter elements. Their promoters are enriched for the TATA box and their nucleosome configuration is distinct from paused genes or housekeeping genes. It remains to be shown how different core promoter elements are differentially used in development and how they influence the dynamics of Pol II initiation and elongation in vivo.

Framework for Analyzing the Chromatin State during Development

We found that different aspects of the chromatin state, such as poised Pol II or H3K27me3, can be used to analyze transcription during development. For example, whereas the recruitment of poised Pol II is mostly stage specific, PcG repression is tissue specific and may keep poised Pol II in check. Thus, different properties of the transcription or chromatin state correlate with either spatial or temporal changes during development, suggesting that there are as yet undiscovered relationships between chromatin regulation and development. This is exciting because an important goal in biology is to predict cellular behavior and development based on genotype and epigenetic state. Thus, mapping the relationship between chromatin and development more systematically could serve as a roadmap for predicting the behavior of diseased cells in humans, e.g., by identifying the tissue of origin and the developmental potential of cells.

EXPERIMENTAL PROCEDURES

Isolation of GFP-Marked Muscle and Neurons for ChIP-Seq and mRNA-Seq Analysis

Briefly, 50 mg aliquots of tightly staged embryos expressing CD8-GFP in either muscle (*mef2-GAL4*) or neurons (*elav-GAL4*) were dissociated in 7 ml Dounce tissue grinders, filtered, prefixed for 5 min with 1% formaldehyde while being spun down, postfixed for 15 min, and passed through a 70 μm syringe filter (BD Medimachine). GFP-positive cells were isolated on a MoFlo high-speed sorter (Beckman Coulter). For a list of the fly lines used, refer to Table S1.

ChIP-Seq Experiments

ChIPs from whole embryos (*Toll*^{10b}, Oregon R) were performed as described in He et al. (2011). Chromatin from cells isolated by FACS was pelleted by high-speed centrifugation and sonicated to an average size of 200 bp; 2–7 μg soluble chromatin was used for each ChIP. Sequencing libraries were prepared from 5–20 ng immunoprecipitated DNA or 100 ng input DNA according to Illumina's instructions (see Extended Experimental Procedures).

Sequential K27me3-Pol II ChIP

Briefly, 60 μg chromatin was immunoprecipitated with 10 μg anti-H3K27me3 antibody (abcam ab6002 and Active Motif #39155), eluted, and subsequently diluted before precipitation with 10 μg anti-CTD4H8 antibody (Millipore). For an extended protocol, see Extended Experimental Procedures. Sequences of the qPCR primers used are listed in Table S3.

mRNA-Seq Library Preparation

Total RNA from sorted cells was isolated using TRIzol (Invitrogen). Polyadenylated *Bacillus subtilis* spike-in RNAs (in vitro transcribed from ATCC clones 87482–87486) were added to a defined amount of total RNA before mRNA-seq libraries were made according to Illumina's instructions.

MNase-Seq

For MNase-Seq, 50 mg cross-linked *Toll*^{10b} embryos were homogenized and washed, and aliquots were digested with increasing amounts of MNase and 20 μg RNaseA at 37°C for 1 hr. After purification by MinElute columns (QIAgen), samples were run on a 2% agarose gel, and DNA corresponding to mononucleosomes (in this case from the sample treated with

32 U MNase) was prepared for paired-end sequencing according to Illumina's instructions.

ChIP-Seq Data Analysis

Sequenced libraries (Illumina GAIIx) were aligned to the UCSC dm3 reference genome. Enrichment values were calculated for each protein-coding transcript in Flybase release 5.28, for Pol II TSS (200 bp wide region centered at +30 bp), Pol II TU (from +400 bp to the 3′ end), H3K27me3 (entire length of the transcript), and H3K4me3 (TSS to +500 bp). Enrichment values were the number of aligned reads overlapping each region in the IP sample divided by the corresponding input control after read-count normalization. To correct for artificially high ratios due to little signal in both the IP and control regions, high ratios with low IP signal were discarded. For genes with multiple annotated TSSs, the enrichment values for the transcript with the highest Pol II_{TSS} enrichment were used. The enrichment values for all genes are listed in Table S4.

mRNA-Seq Data Analysis

Libraries were sequenced on an Illumina GAIIx, and Tophat was used to align them to the reference genome (Flybase release 5.28 with the five spike-in mRNA sequences added as pseudo-chromosomes). Cufflinks was used for transcript abundance (in RPKM) and differential expression analysis (Cuffdiff).

Definition of Gene States

We define a gene as minimally expressed if its RPKM is < 10, and as poised if it is both minimally expressed and has a Pol II_{TSS} enrichment value in the top 20th percentile for both Pol II 8WG16 antibody replicates. Up- and downregulated genes are based on a default FDR of 0.05. A gene is considered induced if it crosses the minimally expressed threshold between two consecutive time points and qualifies as upregulated. A gene is considered PcG repressed if the H3K27me3 enrichment is in the top 2.5% of all genes.

Annotation of Core Promoter Elements

Sequences surrounding all annotated *Drosophila melanogaster* transcript start sites were scanned for the core promoter elements listed in Table S2. A core promoter element was scored as present if it was found with no mismatch within a specified basepair window relative to the TSS.

ACCESSION NUMBERS

All data have been submitted to the GEO database under the accession number GSE34304.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.11.024.

LICENSING INFORMATION

ACKNOWLEDGMENTS

We thank V. Weake for fly stocks, the staff of the Cytometry and Molecular Biology core facilities at Stowers for technical help, Y. Jiang for help in the analysis, and C. Seidel, J. Conaway, and J. Workman for discussions. This project was funded by an NIH New Innovator Award to J.Z. (1DP2 OD004561-01). The work on ESCs was, in part, performed to fulfill requirements for B.D.'s PhD thesis research as a student registered with the Open University. J.Z. is a Pew scholar.

Received: February 15, 2012
Revised: September 29, 2012
Accepted: November 27, 2012
Published: December 20, 2012

REFERENCES

Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., Françoijs, K.J., Stunnenberg, H.G., and Veenstra, G.J. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev. Cell *17*, 425–434.

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

Boehm, A.K., Saunders, A., Werner, J., and Lis, J.T. (2003). Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. Mol. Cell. Biol. *23*, 7628–7637.

Boettiger, A.N., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471–473.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. *44*, 148–156.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157–170.

Chopra, V.S., Hendrix, D.A., Core, L.J., Tsui, C., Lis, J.T., and Levine, M. (2011). The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. Mol. Cell *42*, 837–844.

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.

Deal, R.B., and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev. Cell *18*, 1030–1040.

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C., and Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. Mol. Cell *13*, 887–893.

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2011). Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res. *21*, 216–226.

Engström, P.G., Ho Sui, S.J., Drivenes, O., Becker, T.S., and Lenhard, B. (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Res. *17*, 1898–1908.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Furlong, E.E., Andersen, E.C., Null, B., White, K.P., and Scott, M.P. (2001). Patterns of gene expression during Drosophila mesoderm development. Science *293*, 1629–1633.

Gan, Q., Schones, D.E., Ho Eun, S., Wei, G., Cui, K., Zhao, K., and Chen, X. (2010). Monovalent and unpoised status of most genes in undifferentiated cell-enriched Drosophila testis. Genome Biol. *11*, R42.

Geisberg, J.V., and Struhl, K. (2004a). Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. Nucleic Acids Res. *32*, e151.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. *22*, 1921–1933.

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

Gilchrist, D.A., Fromm, G., dos Santos, G., Pham, L.N., McDaniel, I.E., Burkholder, A., Fargo, D.C., and Adelman, K. (2012). Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. Genes Dev. *26*, 933–944.

Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol. Cell. Biol. *6*, 3984–3989.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473–479.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nat. Genet. *43*, 414–420.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc. Natl. Acad. Sci. USA *105*, 7762–7767.

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

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480–485.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol. Cell. Biol. *28*, 3290–3300.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. Cell *145*, 502–511.

Lin, C., Garrett, A.S., De Kumar, B., Smith, E.R., Gogol, M., Seidel, C., Krumlauf, R., and Shilatifard, A. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). Genes Dev. *25*, 1486–1498.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507–1511.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335–338.

Nègre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. (2006). Chromosomal distribution of PcG proteins during Drosophila development. PLoS Biol. *4*, e170.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E., and Müller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Dev. Cell *15*, 877–889.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. *20*, 2041–2054.

Pearson, B.J., and Doe, C.Q. (2004). Specification of temporal identity in the developing nervous system. Annu. Rev. Cell. Dev. Biol. *20*, 619–647.

Pelegri, F., and Lehmann, R. (1994). A role of polycomb group genes in the regulation of gap gene expression in Drosophila. Genetics *136*, 1341–1353.

Rach, E.A., Yuan, H.Y., Majoros, W.H., Tomancak, P., and Ohler, U. (2009). Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. Genome Biol. *10*, R73.

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

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223–232.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795–804.

Schneider, D.S., Hudson, K.L., Lin, T.Y., and Anderson, K.V. (1991). Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal-ventral polarity in the Drosophila embryo. Genes Dev. *5*, 797–807.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol. *7*, e13.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.

Schwartz, Y.B., Kahn, T.G., Stenberg, P., Ohno, K., Bourgon, R., and Pirrotta, V. (2010). Alternative epigenetic chromatin states of polycomb target genes. PLoS Genet. *6*, e1000805.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

Thomas, S., Li, X.Y., Sabo, P.J., Sandstrom, R., Thurman, R.E., Canfield, T.K., Giste, E., Fisher, W., Hammonds, A., Celniker, S.E., et al. (2011). Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. Genome Biol. *12*, R43.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2

Polycomb chromatin binding in Drosophila melanogaster. Nat. Genet. *38*, 694–699.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *8*, R145.

Tran, K.D., and Doe, C.Q. (2008). Pdm and Castor close successive temporal identity windows in the NB3-1 lineage. Development *135*, 3491–3499.

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.