Absence of canonical marks of active chromatin in developmentally regulated genes

Silvia Pérez-Lluch1–4,7, Enrique Blanco1,3,4,7, Hagen Tilgner1,2,6,7, Joao Curado1,2,5,7, Marina Ruiz-Romero3,4, Montserrat Corominas3,4 & Roderic Guigó1,2

The interplay of active and repressive histone modifications is assumed to have a key role in the regulation of gene expression. In contrast to this generally accepted view, we show that the transcription of genes temporally regulated during fly and worm development occurs in the absence of canonically active histone modifications. Conversely, strong chromatin marking is related to transcriptional and post-transcriptional stability, an association that we also observe in mammals. Our results support a model in which chromatin marking is associated with the stable production of RNA, whereas unmarked chromatin would permit rapid gene activation and deactivation during development. In the latter case, regulation by transcription factors would have a comparatively more important regulatory role than chromatin marks.

Post-translational modifications of histones define an evolutionarily conserved ‘code’ that governs differential gene expression1. Trimethylation of histone H3 at lysine 4 (H3K4me3) and lysine 36 (H3K36me3), for instance, correlates with active transcription, whereas trimethylation of histone H3 at lysine 9 (H3K9me3) and lysine 27 (H3K27me3) is usually linked to transcriptional repression2,3. The combinatorial behavior of histone modifications along regulatory regions—reflecting and/or influencing the specific arrangement of transcription factors—modulates the expression levels of genes, conferring a unique temporal and spatial transcriptional program. Computational models have been developed that can predict gene expression from histone modifications with great accuracy4,5.

A number of recent reports, however, indicate that expression of certain genes may occur in the absence of histone modifications canonically associated with active genes. The modENCODE project reported that some expressed genes lacked H3K4me3 marks6. Hödl and Basler found that cells that lack H3K4 methylation respond to developmental signaling pathways by activating target gene expression in Drosophila melanogaster wing imaginal discs (WIDs)7. Chen et al. observed that pre-midblastula transition (pre-MBT) genes have particularly low levels of H3K4me3 (ref. 8). More recently, Zhang et al. reported that genes within yeast heterochromatic regions can be transcribed in the absence of active histone marks9. Here we show that active transcription in the absence of chromatin marking is actually a general feature of genes that are strongly regulated during development. We analyzed data produced by modENCODE in whole animals and tissues for fly and worm, characterized the fly transcriptome by RNA sequencing (RNA-seq) and the fly epigenome by chromatin immunoprecipitation and sequencing (ChIP-seq) in two spatially well-defined and relatively homogeneous developmental tissues in fly, and carried out targeted experimental validations in isolated cells. All these analyses strongly suggest that expression of genes regulated during fly development can occur in the absence of marks typically associated with active genes, and, indeed, this expression does not seem to be affected by perturbations of the histone methyltransferase system. Conversely, we found that chromatin marking is associated not only with transcriptional levels but also with transcriptional and post-transcriptional stability—an association that appears to have been conserved throughout metazoan evolution.

RESULTS

Expression without histone modification during development

To investigate the dynamics of chromatin marking in genes regulated during development, we analyzed data generated by the Drosophila modENCODE project6,10. We specifically analyzed RNA-seq data and ChIP-seq data for H3K4me3, acetylation of histone H3 at lysine 9 (H3K9ac), monomethylation of histone H3 at lysine 4 (H3K4me1), acetylation of histone H3 at lysine 27 (H3K27ac), H3K27me3 and H3K9me3 in whole flies (Supplementary Fig. 1a). To measure transcriptional stability, we computed the coefficient of variation for gene expression over 12 developmental time points (Online Methods and Supplementary Fig. 1b), with lower values corresponding to higher transcriptional stability. The distribution for the coefficients of variation uncovered a large class of genes that showed constant expression during development and two other minor classes containing genes whose expression was highly variable and often restricted to a limited number of developmental stages.

1Centre for Genomic Regulation, Barcelona, Spain. 2Departament de Ciències Experimental i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain. 3Departament de Genètica (Facultat de Biologia), Universitat de Barcelona, Barcelona, Spain. 4Institut de Biomedicina de la Universitat de Barcelona, Barcelona, Spain. 5Graduate Program in Areas of Basic and Applied Biology, Abel Salazar Biomedical Sciences Institute, University of Porto, Porto, Portugal. 6Present address: Department of Genetics, Stanford University, Stanford, California, USA. 7These authors contributed equally to this work. Correspondence should be addressed to M.C. (mcorominas@ub.edu) or R.G. (roderic.guigo@crg.cat).

Received 2 June; accepted 22 July; published online 17 August 2015; doi:10.1038/ng.3381
We arbitrarily selected the 1,000 genes with the highest coefficients of variation and defined them as developmentally ‘regulated’ genes because of their variable pattern of expression throughout developmental time. Conversely, we selected the 1,000 genes with the lowest coefficients of variation and defined them as developmentally ‘stable’ genes. For each gene, we determined the time point at which its expression was the highest. At this time point, we did not observe strong differences in the expression levels for stable and regulated genes (Fig. 1a). At the same time point, we measured the levels of histone modifications for each gene (Online Methods). We found that, at the time point with highest expression, stable genes were strongly marked by histone modifications typically associated with active transcription, H3K4me3 and H3K9ac, and also with enhancers, H3K4me1 and H3K27ac. Unexpectedly, however, regulated genes showed very low levels of these modifications, with levels comparable to those of silent genes (Fig. 1b and Supplementary Fig. 2).

In Figure 1c, we compare the pattern of H3K4me3 along the timeline of fly development in CG8636, a gene stably expressed during development, and St2, a gene specifically expressed in pupa. CG8636 showed a strong H3K4me3 peak downstream of the transcription start site (TSS), whereas St2 lacked any marking, even in the pupa stage where it is expressed at higher levels than CG8636 (Supplementary Fig. 3). This contrasting pattern of histone marking...
was not only apparent when comparing genes with extreme behavior but was also a distinct feature of the partitioning of the entire set of fly genes according to transcriptional stability (Supplementary Fig. 4). We observed that regulated genes showed higher levels of histone modifications typically associated with inactive genes, H3K27me3 and H3K9me3, than did stable genes, with the levels of these marks similar to those found for silent genes (Fig. 1d and Online Methods). The levels of inactive marks, however, were generally low in comparison to the levels of active marks, even for genes silent during development—a large proportion of which lacked any evidence of these histone modifications (Supplementary Fig. 5a,b), as has been previously reported11. We found only a weak relationship between the level of repressive marks and gene expression (Supplementary Fig. 5c,d).

Given that the developmental chromatin maps generated by the modENCODE project are for whole organisms, it could be argued that an apparent lack of chromatin marking is the consequence of the expression of regulated genes being spatially confined to specific organs, tissues or subcellular domains. Although, indeed, regulated genes in general showed a spatially restricted pattern of expression, chromatin marking could actually be detected in stable genes that exhibited a restricted expression pattern comparable to that of the regulated genes (Supplementary Fig. 3). To further investigate the potential effect of restricted expression on the ability to detect chromatin marking, we used tissue-specific RNA-seq data from modENCODE12. Third instar larva (L3) was the developmental stage with the largest number of tissues available, including carcass, central nervous system, digestive system, fat body, imaginal discs and salivary glands. Using L3 tissue-specific RNA-seq data, we identified seven regulated genes expressed in all six available tissues at L3 (‘regulated, broadly expressed’ genes; Fig. 2a, left). Conversely, we identified 130 stable genes specifically expressed in only one of the aforementioned tissues at L3 (‘stable, tissue-specific’ genes; Fig. 2a, right). Regulated, broadly expressed genes had much higher expression levels than stable, tissue-specific genes when expression was measured for the whole body (almost fourfold higher; Fig. 2b), as well as in general when expression was measured for individual tissues (Supplementary Fig. 6). Regulated genes also had higher expression levels than stable genes overall. However, the levels of H3K4me3, H3K9ac, H3K4me1 and H3K27ac in regulated, broadly expressed genes were significantly (P < 0.05) lower than those in stable genes, even than those in stable, tissue-specific genes, and were comparable to those in silent genes (Fig. 2c). We confirmed gene expression and H3K4me3 and H3K9ac levels by quantitative PCR (qPCR) (Fig. 2b) and ChIP-qPCR (Fig. 2d), respectively.

### Figure 2 Gene expression and histone modifications in regulated, broadly expressed and stable, tissue-specific genes at L3.

(a) Diagrams of developmentally regulated genes broadly expressed across multiple tissues at L3 (left) and stable genes expressed in only one tissue at L3 (right). E, embryo; AM, adult male; AF, adult female. (b) Left, gene expression levels at L3 measured by whole-organism RNA-seq. The number of genes in each category is given under the box plots. Right, validation by qPCR of the expression at L3 of regulated, broadly expressed genes in comparison to a stable gene (Bmcp) and a silent gene (CG5367). Error bars, s.e.m. from three independent replicates. (c) Levels of H3K4me3, H3K9ac, H3K4me1, H3K27ac and H3K27me3 for whole L3 individuals. The seven regulated genes broadly expressed at L3 are depicted as red dots in the box plots. P values were computed using the Wilcoxon test (two-sided). (d) Validation by individual ChIP assays and qPCR of H3K4me3, H3K9ac and H3K27me3 levels in regulated genes broadly expressed at L3. H3K4me3, H3K9ac and H3K27me3 ChIP assay results are presented as the enrichment of the marks in regulated genes relative to a silent gene (CG5367). Error bars, s.e.m. from three independent replicates. Negative control experiments were performed without antibody.
Figure 3 Association of histone modifications with transcriptional stability in metazoans. (a) Scatterplot of H3K4me3 levels at the time point with highest expression during fly development and transcriptional stability measured as the coefficient of variation for gene expression across time points. The correlation is computed as the partial correlation (coefficient of correlation, cc) given gene expression. (b) Partial correlations between active marks and transcriptional stability (coefficient of variation). Correlations were computed controlling for gene expression. All correlations are statistically significant (P < 2.2 × 10^{-16}). P values were computed using Student’s t test (two-sided). NA, not available. (c) Expression of stable, regulated and silent genes during worm development at the time point with maximum expression. RPKM, reads per kilobase per million mapped reads. (d) Levels of H3K4me3 and H3K36me3 at the time point with maximum expression during worm development. (e) Expression of genes with constant and variable expression in the tissue or cell line with the highest expression across multiple samples from the Roadmap Epigenomics Mapping Consortium. (f) Levels of H3K4me3 in the human tissue with maximum expression. (g) Expression of genes with constant and variable expression in the tissue with highest expression across ten mouse tissues from the mouse ENCODE project. (h) Levels of H3K4me3 in the mouse tissue with maximum expression. These levels correspond to the maximum height of the ChIP-seq peak within the gene body. P values were computed using the Wilcoxon test (two-sided).

All these results strongly suggest that activation of genes regulated during development occurs mostly in the absence of histone modifications canonically linked to active genes. Our results also point to the association of strong chromatin marking not only with transcriptional levels but also with transcriptional stability. We calculated the coefficient of correlation across all genes between the coefficient of variation for gene expression across developmental time points, as computed above, and the level of histone modifications at the developmental time point with the highest expression. We used partial correlations to control for a potential confounding effect of gene expression levels (Online Methods). For all active histone modifications, the partial correlations were negative and significant (as low as coefficient of correlation = −0.68 for H3K4me3; Fig. 3a,b and Supplementary Fig. 7), strongly supporting association between transcriptional stability and active chromatin marks.

To investigate whether the lack of chromatin marking in regulated genes and the association between chromatin marking and transcriptional stability are conserved in other metazoans, we first analyzed RNA-seq–based gene expression at seven time points throughout Caenorhabditis elegans development and ChIP-chip data on two histone modifications (H3K4me3 and H3K36me3) available for these time points in modENCODE. Although both the resolution and reliability of the chromatin data obtained through ChIP-chip analysis in worm were lower than in the fly ChIP-seq data, we observed the same trend: expression levels at the time point of maximum expression were very similar for regulated and stable genes (Fig. 3c), whereas regulated genes showed lower levels of H3K4me3 and H3K36me3, with levels more similar to those of silent genes (Fig. 3d).

Unfortunately, genome-wide transcriptomic and epigenetic developmental maps of the resolution of those from modENCODE are not yet available for mammalian (or vertebrate) systems. Nevertheless, using transcriptomic and epigenomic data across multiple tissues and cell lines in human and mouse, we did find that active chromatin marking was also associated with transcription stability in mammalian systems. We used RNA-seq data and ChIP-seq data for H3K4me3, H3K36me3 and H3K4me1 for 56 human adult and fetal tissues, primary cells and cultured cell lines from the Roadmap Epigenomics Mapping Consortium. We found strong negative correlation between the coefficient of variation for gene expression across these samples and histone levels (Fig. 3b). The gene set with the highest variation in expression across human tissues is likely to show some enrichment for regulated genes. Thus, we selected the 1,000 genes with the highest coefficients of variation as variably expressed genes and the 1,000 genes with the lowest coefficients of variation as constantly expressed genes. In the cell type in which the expression of each gene was highest, variable genes showed higher expression than constant genes (Fig. 3e). Yet, the levels of active histone modifications in these cell types were much lower in variable genes than in constant genes (Fig. 3f and Supplementary Fig. 8a). Very similar results
were obtained in mouse when using Encyclopedia of DNA Elements (ENCODE) data15 (Fig. 3b,g,h and Supplementary Fig. 8b).

Expression without histone modifications in imaginal discs
Data generated by the modENCODE project monitor complex systems encapsulating great cellular heterogeneity. To investigate the dynamics of chromatin marking during development in a more homogeneous cellular environment, we characterized the transcriptome by RNA-seq (Supplementary Fig. 9a,b and Supplementary Table 1) and the epigenome by ChIP-seq in two Drosophila L3 tissues: WID and eye-antenna imaginal disc (EID). We specifically monitored the genic nucleosome (H3) and the active marks H3K4me3, H3K9ac, H3K4me1 and H3K27ac, plus the transcription elongation mark H3K36me3 (Supplementary Fig. 9c). Both, WIDs and EIDs are epithelial tissues in early differentiation stages, and differentially expressed genes are likely to be under temporal developmental control. Although the WID and EID epigenomes and transcriptomes were very similar (Supplementary Fig. 9d,e), differentially expressed genes did exhibit functions strongly consistent with the known biology of these tissues (Supplementary Fig. 9f and Supplementary Tables 2 and 3).

We then investigated the marking of regulated and stable genes in WIDs and EIDs. To focus on genes under stronger regulation, we identified 55 developmentally regulated genes expressed in EIDs but not in WIDs and 10 developmentally regulated genes expressed in WIDs but not in EIDs. We also identified a set of 284 stable genes highly expressed in both EIDs and WIDs, as well as a set of 30 genes silent in both tissues (Online Methods and Supplementary Tables 4–7).

We next compared the histone marks at stable, silent and regulated WID- and EID-specific genes. Consistent with previous observations16,17, the WID and EID profiles for stable genes were very similar, as were those for silent genes (Fig. 4a). Stable and silent genes were both characterized by higher stable nucleosome occupancy than nearby intergenic regions, but H3 enrichment was greater for stably expressed genes than for silent genes. Stable genes were also strongly marked by H3K4me3, H3K9ac and H3K36me3 and, as observed in modENCODE, by H3K4me1 and H3K27ac. Silent genes mostly lacked these histone modifications. Regulated tissue-specific genes, however, exhibited contrasting behavior. As expected, WID-specific genes lacked active modifications in EIDs (Fig. 4b), and, conversely, EID-specific genes were not marked in WIDs (Fig. 4c).

Unexpectedly but consistent with the behavior that we observed in modENCODE data, WID-specific genes were also not marked in WIDs nor were EID-specific genes marked in EIDs. Absence of active histone marking cannot be attributed to the lack of nucleosomes because H3 was observed in these genes (Fig. 4b,c). It is also unlikely that it originates from higher nucleosome turnover in regulated genes as, at least in Drosophila S2 cells18, nuclear turnover is similar for stable and regulated genes (Supplementary Fig. 10). Lack of histone marking is also due to the relatively low expression levels of WID- and EID-specific genes, as, even when these genes had high levels of expression comparable to those of constitutively expressed genes, there was no marking by active histone modifications (Fig. 5; see Supplementary Fig. 11 for more examples). The WID-specific gene CG4382 and the EID-specific gene CG14516 had expression levels similar to those of the stable gene nub. This gene, however, was strongly marked by histone modifications in both WIDs and EIDs, whereas CG4382 and CG14516 were marked in neither tissue. Lack of chromatin marking cannot be attributed to the restricted expression of tissue-specific genes, as the expression of nub is also restricted to specific regions in both WIDs and EIDs19,20. H3 levels for tissue-specific and stable genes were comparable and only depended weakly according to the expression status of the genes (Fig. 5).

Active transcription without histone modifications
Although WIDs and EIDs are relatively homogeneous tissues, they already show cellular subspecialization at the L3 stage. For instance, the WID-specific gene pdm2 (POU domain protein 2), like nub (nubbin)21, which has strong temporal and spatial regulation during development, is only expressed in the wing primordium (wing pouch) at the L3 stage (Fig. 6a). To unequivocally demonstrate lack of chromatin marking in developmentally regulated genes, we took advantage of the nub-GAL4 construct to drive expression of GFP only in the wing pouch, where pdm2 is expressed. Thus, we collected all cells expressing pdm2 and investigated chromatin marking for this gene only in the cells in which it was expressed. More specifically, dissection and dissociation of wing discs followed by cell sorting analyses allowed the isolation of two populations of cells: the wing pouch (nub domain, GFP positive) and the rest of the wing (GFP negative) (Fig. 6a).
and Online Methods). By using qPCR, we found that the expression of pdm2, restricted to sorted GFP-positive cells, was even higher than the expression of crm, a gene expressed at the same level throughout the WID (Fig. 6b). ChIP-qPCR on sorted cells showed that the levels of H3K4me3 and H3K36me3 in pdm2 were significantly lower than those in crm and comparable to those in CG10013, a gene silent in the whole WID (Fig. 6c). High RNA levels of pdm2 in the wing pouch (Fig. 6b) did not necessarily demonstrate active transcription, as transcription could have occurred at an earlier time point. To assess active gene expression, we directly measured newly transcribed RNA (nascent RNA) in sorted cells. Active transcription of pdm2 in GFP-positive cells was as high as transcription of the control gene crm (Fig. 6d).

To investigate marking by repressive histone modifications in expressed genes (Fig. 1d), we monitored the levels of H3K27me3 in pdm2, a gene exhibiting this modification at L3 when levels were measured in the whole organism. We performed individual ChIP-qPCR assays in sorted cells and found that pdm2 was indeed marked by H3K27me3 in the WID but only outside the wing pouch. No marking was observed in the wing pouch, where pdm2 is expressed (Fig. 6e). This finding suggests that the repressive

---

**Figure 5** Profiles of RNA expression, H3 and histone modifications in WIDs and EIDs. noc is a gene stably expressed in WIDs and EIDs, CG4382 is a WID-specific gene and, CG14516 is an EID-specific gene. Levels of gene expression (in RPKM) are depicted at the bottom. Screenshots were obtained from the UCSC Genome Browser56.

**Figure 6** Active transcription of pdm2 without chromatin modifications. (a) Expression of pdm2 in a WID (left) and EID (middle) as demonstrated by labeling with a pdm2-specific probe. The gene is only expressed in the wing pouch of the WID (green). Scale bars, 100 μm. (b) Expression of pdm2 in sorted cells analyzed by qPCR. Gene expression is normalized by that of the control gene crm. Error bars, s.e.m. from three biological replicates. (c) ChIP analysis of H3K4me3 and H3K36me3 (negative controls without antibody) for sorted cells. ChIP results are presented as enrichment relative to a silent gene not marked with H3K4me3 or H3K36me3 (CG10013). crm is used as a positive control for these modifications. Error bars, s.e.m. from at least three biological replicates. P values were computed using the Student's t test (two-sided). (d) Newly transcribed RNA from GFP-sorted cells. Nascent RNA levels are normalized by the levels for the control gene crm. Error bars, s.e.m. from at least three biological replicates. (e) ChIP analysis of H3K27me3 (negative controls without antibody) for sorted cells. H3K27me3 ChIP results are presented as enrichment relative to a constitutively expressed gene not marked with H3K27me3 (Rpl32). Abd-B is used as positive control for this modification. Error bars, s.e.m. from at least three biological replicates.
modifications detected in whole organisms in regulated genes (Fig. 1d) could originate from organs or tissues in which these genes are not expressed.

Lack of active marking suggests that genes regulated throughout development may not respond to histone modification systems. Therefore, we specifically investigated the response of regulated genes to the lack of ASH2 (absent, small or homeotic disc 2), a key cofactor for H3K4 methylation22. First, we characterized ASH2 occupancy along fly genes using ChIP-seq data obtained in WIDs17 and found a very strong depletion of ASH2 binding to the promoters of regulated genes in comparison to those of stable genes (Fig. 7a). Second, we used the ash2I1 mutant allele (which corresponds to an imprecise excision of a P-element inserted in the ash2 gene) to interfere with H3K4me3. Because this allele is lethal at the late L3–early pupa stage23, we performed clonal analyses in WIDs and EIDs. We specifically analyzed two stable genes—en (engrailed), expressed in the posterior compartment of the WID, and CycA (Cyclin A), ubiquitously expressed in the WID—as well as two regulated genes, pdm2, expressed in the wing pouch, and boss (bride of sevenless), expressed in the differentiated photoreceptor R8 cell of the EID. We confirmed lack of H3K4me3 in ash2I1 mutant clones (Fig. 7b) and observed a clear reduction in the levels of En and CycA, whereas the expression levels of Boss and pdm2 were not affected (Fig. 7c–p).

Genome organization of regulated genes

It has been suggested that developmental control genes are under a characteristic regulatory program24. They tend to harbor an increased number of transcription factor binding sites25 and are characterized by ‘peaked’ (or narrow) promoters, in comparison to housekeeping genes, which are associated with more ‘dispersed’ (or broad) promoters26–28. Using the promoter classification of Ni et al.30, we found that stable genes were strongly enriched in broad (and weak) promoters in comparison to regulated genes (444 versus 12). In contrast, the
Figure 8  Promoter architecture and genome organization for stable and developmentally regulated genes. (a) PCA of genes expressed in the *D. melanogaster* embryo between 0 and 12 h, based on the ChIP-chip binding profiles of 20 transcription factors. (b) Fraction of stable and regulated genes in different states from chromatin segmentations at LE and L3 (ref. 43). Right, proportion of regulated genes when considering only genes expressed at LE or L3. (c) The same as in (b), for segmentations in the BG3 and S2 cell lines (ref. 42). (d) The same as in (b), for segmentation in Kc167 (ref. 41). Black chromatin corresponds to repressive chromatin; yellow and red chromatin is typical of transcriptionally active regions; green and blue chromatin corresponds to repressive chromatin. (e) Proportion of stable and regulated genes mapping to spatial chromatin domains, considering the 1,169 domains inferred by HiC in fly embryos (ref. 44). (f) Distribution of Shannon entropy of splicing for stable and regulated genes. Shannon entropy is computed at the developmental time point at which gene expression is the maximum. The number of genes for each category appears below the x axis. (g) Distribution of the relative usage of the major isoform. The y axis shows the fraction of the total transcriptional output of the gene that is captured by the most abundant isoform.

The proportion of peaked promoters was similar in stable and regulated genes (42 versus 38). Overall, however, our set of regulated genes exhibited most of the characteristics that have been reported for developmental and/or peaked-promoter genes in *Drosophila* and other species (see Lenhard et al.31 for a review). Thus, the promoters of regulated genes show stronger conservation29–31, particularly in predicted transcription factor binding motifs (Supplementary Fig. 12). They are depleted in DNA replication–related element (DRE) sequences, which are associated with the disperse initiation of transcription8 (15% of regulated compared to 39% of stable genes) and enriched in TATA-binding protein (TBP) boxes, characteristic of tighter gene regulation32,33 (49% versus 15%). In contrast, the promoters of stable genes overlapped modENCODE high-occupancy target (HOT) regions, associated with open chromatin and ubiquitous expression34,35, more often than the promoters of regulated genes (67% versus 8%). We also found that the overall pattern of transcription factor binding clearly separated regulated from stable genes, as demonstrated by principal-component analysis (PCA) based on ChIP-chip data for 20 transcription factors in fly embryos (Fig. 8a). Finally, analyses of published data36–40 of knockdown or overexpression of several transcription factors showed a frequently larger impact of these perturbations on the expression of regulated genes in comparison to stable genes (Supplementary Table 8).

Regulated genes also exhibited a characteristic genome organization. We mapped our sets of stable and regulated genes to a number of genome segmentations, representing epigenomic domains, recently obtained in *Drosophila* cell lines (Kc167 (ref. 41), BG3 and S2 (ref. 42)) and developmental time points (late embryo (LE) and L3 (ref. 43)). We systematically found that regulated genes tended to occur in chromatin states that are depleted in histone modifications (ref. 43). Right, proportion of regulated genes when considering only genes expressed in the developmental and/or peaked-promoter genes in *Drosophila* and other species (see Lenhard et al.31 for a review). Thus, the promoters of regulated genes show stronger conservation29–31, particularly in predicted transcription factor binding motifs (Supplementary Fig. 12). They are depleted in DNA replication–related element (DRE) sequences, which are associated with the disperse initiation of transcription8 (15% of regulated compared to 39% of stable genes) and enriched in TATA-binding protein (TBP) boxes, characteristic of tighter gene regulation32,33 (49% versus 15%). In contrast, the promoters of stable genes overlapped modENCODE high-occupancy target (HOT) regions, associated with open chromatin and ubiquitous expression34,35, more often than the promoters of regulated genes (67% versus 8%). We also found that the overall pattern of transcription factor binding clearly separated regulated from stable genes, as demonstrated by principal-component analysis (PCA) based on ChIP-chip data for 20 transcription factors in fly embryos (Fig. 8a). Finally, analyses of published data36–40 of knockdown or overexpression of several transcription factors showed a frequently larger impact of these perturbations on the expression of regulated genes in comparison to stable genes (Supplementary Table 8).

Regulated genes also exhibited a characteristic genome organization. We mapped our sets of stable and regulated genes to a number of genome segmentations, representing epigenomic domains, recently obtained in *Drosophila* cell lines (Kc167 (ref. 41), BG3 and S2 (ref. 42)) and developmental time points (late embryo (LE) and L3 (ref. 43)). We systematically found that regulated genes tended to occur in chromatin states that are depleted in histone modifications (Fig. 8b–d), even when considering only regulated genes expressed in the developmental time point at which the segmentation was obtained (Fig. 8b).
Epigenomic domains in turn are spatially organized into well-defined physical domains within the nucleus\(^4\). Silent chromatin regions in particular fold into modular chromosomal entities, which we found enriched in regulated genes (Fig. 8e). The nuclear lamina has a key role in this physical organization, through interaction with large continuous chromosomal domains. These lamina-associated domains (LADs) are generally depleted of chromatin marks\(^45\), and we consistently found that regulated genes were strongly enriched in LADs (52% compared to 5% of stable genes in 412 LADs from Kc167 cells\(^46\)).

**Histone modifications and alternative splicing**

Beyond its role in primary RNA production, chromatin structure has also been implicated in subsequent steps of RNA processing. In particular, a number of studies have uncovered a relationship between nucleosome occupancy and exon-intron structure\(^47,48\) and between specific histone modifications and alternative splicing\(^49–51\). We found in fly WID and EID that highly included exons were characterized by higher H3 occupancy than less frequently included ones, as previously reported in mammals\(^48\) (Online Methods, Supplementary Fig. 13a,b and Supplementary Tables 9 and 10) and the correlation between H3 occupancy and exon inclusion peaks very close to the acceptor site (Supplementary Fig. 13c,d).

We thus speculated that strong chromatin marking might not only be associated with more stable RNA production but also with tighter regulation of alternative splicing. To measure the complexity of alternative splicing, we computed Shannon entropy on the relative abundance of a gene's alternative splicing isoforms (Online Methods). The splicing entropy grew with the number of isoforms and with the evenness of their relative abundances. Higher entropic values can be interpreted as tight regulation of alternative splicing, whereas lower values would correspond with more stochastic production of alternative isoforms. As hypothesized, splicing entropy, measured at the time point of maximum expression for a gene, was lower for strongly marked, stable genes than for unmarked, developmentally regulated genes (Fig. 8f). Further supporting tighter regulation of splicing, we also found that the major isoform captured a larger fraction of the total transcriptional output in stable genes than in regulated ones (Fig. 8g).

**DISCUSSION**

Cell type–specific transcriptional regulation is crucial to maintain cell identity throughout the lifetime of organisms, yet it must be flexible enough to allow for responses to endogenous and exogenous stimuli. This regulation is mediated by specific molecular factors (for example, cell type–specific transcription factors and chromatin modifications), as well as by the topological organization of the genome. In particular, modifications occurring on DNA and on histones regulate gene expression by establishing and maintaining specific chromatin states\(^22,53\). The association of certain modifications with transcriptional activation or repression has become widely accepted. Nevertheless, expression of genes in the absence of chromatin marks has also been reported\(^6–9\). Here we found that transcription in the absence of most canonically active chromatin marks is actually a characteristic feature of genes that are regulated during fly and worm development. These genes are not necessarily equivalent to developmental control genes, many of which are known to be marked\(^41,52\).

Analyses of tissue-specific gene expression data, as well as our targeted validation experiments, support the notion that our observations do not arise from the expression of developmentally regulated genes being low or confined to small cell populations, from limited detection sensitivity and/or from persistence in the cell of RNA molecules transcribed at some earlier time point. Thus, although factors not accounted for cannot be completely ruled out, our observations appear to reflect a true biological property of genes regulated throughout development, which may be a consequence of these genes being partially unresponsive to histone modification systems.

We also found that the strongly marked chromatin state is associated with more tightly controlled transcriptional and post-transcriptional regulation, in particular of splicing. This is consistent with earlier observations\(^54\) of simultaneous enrichment in the expression of chromatin-modifying enzymes and splicing factors in cell-enriched testis and with the higher levels of H3K36me3 found by de Almeida et al.\(^49\) in mammalian constitutive exons as compared to alternative exons.

Overall, our results lead us to hypothesize that the relative contributions of transcription factors and histone modifications to the regulation of gene expression differentiates the transcriptional programs of stable and regulated genes. In stable genes that are constitutively expressed, strong chromatin marking leads to transcriptional stability and tightly controlled RNA production. In these genes, regulation by transcription factors would have a comparatively smaller role. In contrast, genes regulated during development that need to be rapidly activated and deactivated are characterized by an unmarked chromatin state. In these other genes, transcription factors binding to chromatin would have the predominant regulatory role. These distinct regulatory programs would be reflected in the topological organization of the chromatin fiber within the nucleus, with regulated genes located in silent chromosomal modular domains that physically interact with the nuclear lamina\(^45\).

Although we found evidence for this model of transcriptional regulation specifically in the fly, preliminary results suggest that it may be generalizable to other metazoans. Although detailed transcriptional, epigenetic and topological maps of genomes are being produced in an increasing number of cell lines and tissues, developmental maps are still sparse in mammalian species. Exhaustive monitoring through a much larger variety of conditions, differentiation states and developmental stages, as well as characterization of enhancers, is required to fully understand the layer of epigenetic regulation that mediates between genome sequence and RNA production.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** ChIP-seq and RNA-seq raw data and profiles of read counts have been deposited in the Gene Expression Omnibus (GEO) under accession GSE56551.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank D. Gonzalez-Knowles, A. Breschi and M. Melé for help with data analysis, F. Serras, M. Morey and A. Kornblith for insightful suggestions and G. Cavalli for discussing data before publication, as well as the anonymous reviewers for their critical input. We thank R. Garrido for administrative assistance. We thank the modEncode project, the ENCODE Project (human and mouse data) and the Roadmap Epigenomics Mapping Consortium for granting open access of these resources to the scientific community. We also thank the Ultrasound Sequencing Unit of the Centre for Genomic Regulation (CRG, Barcelona, Spain) for sample processing and the Confocal Unit of CCCiTUB (Centres Científics i Tecnològics de la Universitat de Barcelona) (Universitat de Barcelona, Barcelona, Spain). This work was performed under the financial support of the Spanish Ministry of Economy and Competitiveness with grants BIO2011-26205 to R.G., CSD2007-00008 and BFU2012-36888 to M.C., and ´Centro de Excelencia Severo Ochoa 2013–2017; SEV-2012-0208 and the European Research Council/European
1. Li, B., Carey, M. & Workman, J.L. The role of chromatin during transcription. Cell Development 131, 767–774 (2004).
2. Black, J.C., Van Rechem, C. & Whetstine, J.R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol. Cell 48, 491–507 (2012).
3. Wagner, E.J. & Carpenter, P.B. Understanding the language of Lys36 methylation at histone H3. Nat. Rev. Mol. Cell Biol. 13, 115–126 (2012).
4. Dong, X. et al. Modeling gene expression using chromatin features in various cellular contexts. Genome Biol. 13, R53 (2012).
5. Karlic, R., Chung, H.R., Lasserre, J., Vlahovicek, K. & Vingron, M. Histone modification levels are predictive for gene expression. Proc. Natl. Acad. Sci. USA 107, 2926–2931 (2010).
6. Nègre, N. et al. A cis-regulatory map of the Drosophila genome. Nature 471, 527–531 (2011).
7. Hödl, M. & Basler, K. Transcription in the absence of histone H3.2 and H3K4 methylation. Curr. Biol. 22, 2253–2257 (2012).
8. Chen, K. et al. A global change in RNA polymerase II pausing during the Drosophila midblastula transition. elife 2, e00861 (2013).
9. Zhang, H., Gao, L., Anandhakumar, J. & Gross, D.S. Uncoupling transcription from covalent histone modification. PLoS Genet. 10, e1004202 (2014).
10. Grønveld, B.R. & van Bemmel, J.G. Integrative analysis of 111 reference human epigenomes. Nature 471, 129–143 (2011).
11. Sullivan, P.J., Johnson, J.E., Bejerano, G. & Guigo, R. Integrative analysis of active transcriptional programs. PLoS Genet. 9, e1003627 (2013).
12. Tokuyasu, K.T. et al. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143, 212–224 (2010).
13. Greil, A. et al. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471, 480–485 (2011).
14. Tukey, J.W. et al. The BTB–zinc finger transcription factor Aplabut acts as an epithelial oncogene in Drosophila melanogaster through maintaining a progenitor-like cell state. PLoS Genet. 9, e1002981 (2013).
15. Grønveld, B.R. et al. A comprehensive encyclopedia of DNA elements in the mouse genome. Nature 518, 317–330 (2015).
16. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature 518, 355–364 (2014).
17. Barski, A. et al. High-resolution profiling of histone modifications in the human genome. Cell 129, 823–837 (2007).
18. Pérez-Lluch, S. et al. Genome-wide chromatin occupancy analysis reveals a role for TFIID in transcriptional pausing. Nucleic Acids Res. 39, 4626–4639 (2011).
19. Teves, S.S. & Henikoff, S. Transcription-generated torsional stress destabilizes nucleosomes. Nat. Struct. Mol. Biol. 21, 88–94 (2014).
20. Luque, C.M. & Milan, M. Growth control in the proliferative region of the Drosophila eye-antenna primordium: the elbow-roc gene complex. Dev. Biol. 301, 527–339 (2007).
21. Weihe, U., Dorfman, R., Wernet, M.F., Cohen, S.M. & Milan, M. Proximodistal subdivision of Drosophila legs and wings: the elbow-roc gene complex. Dev. Biol. 28, 665–676 (2007).
22. Ng, M., Diaz-Benjumea, F.J. & Cohen, S.M. Numb encodes a POU-domain protein required for proximal-distal patterning in the Drosophila wing. Development 121, 589–599 (1995).
23. Beltrán, S., Angulo, M., Pignatelli, M., Serras, F. & Corominas, M. Functional dissection of the ash2 and ash1 transcriptions provides insights into the transcriptional basis of wing phenotypes and reveals conserved protein interactions. Genome Biol. 8, R67 (2007).
24. Beltrán, S. et al. Transcriptional network controlled by the trithorax-group gene ash2 in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 100, 3293–3298 (2003).
25. Zellinger, J. & Stark, A. Developmental gene regulation in the era of genomics. Dev. Biol. 339, 230–239 (2010).
26. Stark, A. et al. Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature 450, 219–232 (2007).
27. Carinci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. 38, 626–635 (2006).
28. Gaertner, B. et al. Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep. 2, 1670–1683 (2012).
29. Hoskins, R.A. et al. Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Res. 21, 182–192 (2011).
30. Rach, E.A., Yuan, H.Y., Majors, W.H., Tomancak, P. & Ohler, U. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. Genome Biol. 10, R73 (2009).
31. Luque, C.M. & Milan, M. Growth control in the proliferative region of the Drosophila eye-antenna primordium: the elbow-roc gene complex. Dev. Biol. 301, 527–339 (2007).
32. Beltrán, S., Angulo, M., Pignatelli, M., Serras, F. & Corominas, M. Functional dissection of the ash2 and ash1 transcriptions provides insights into the transcriptional basis of wing phenotypes and reveals conserved protein interactions. Genome Biol. 8, R67 (2007).
33. Beltrán, S. et al. Transcriptional network controlled by the trithorax-group gene ash2 in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 100, 3293–3298 (2003).
ONLINE METHODS

*Drosophila* strains. The strains used included Canton *S* as a wild-type strain and *nub-GAL4/+; UAS-GFP/+*. Flies were kept on standard medium at 25 °C.

Tissue disaggregation and cell sorting. WIDs from *nub-GAL4/+; UAS-GFP/+* flies were dissected in PBS and incubated for 1 h in 10× trypsin solution (Sigma, T4174) at room temperature on a rotating wheel. Cells were vigorously pipetted and kept on ice in Schneider’s insect medium. To discard dead cells, DAPI was added to the samples at 1 µg/ml final concentration. Cells were sorted on a FACSAria (BD) with the 85-µm nozzle. We were able to recover around 2.5 × 10^6 GFP-negative cells and 2 × 10^6 GFP-positive cells from 400 WIDs. An independent sorting experiment was carried out for each replicate, both for ChIP assays and gene expression analyses.

RNA extraction, reverse transcription and RT-PCR. We used 120 WIDs and 250 EIDs as starting material for RNA-seq. For *pdm2* gene expression analysis, the WIDs from *nub-GAL4/+; UAS-GFP/+* flies were disaggregated. RNA from sorted cells was extracted with the ZR-RNA MicroPrep kit (Zymo Research). For L3-specific gene expression, five L3 larvae were crossbred, fixed and sonicated as described above. Immunocomplexes from sorted cells were analyzed on a 1% agarose gel.

Genetic mosaics. Clones with the *asa*21 mutant allele were obtained by mitotic recombination using the FLP/FRT technique57, yw; FRT82Bash211/TM3C flies were crossed with *ywhsp75A FRT82B GFP/*TM6B flies, and WIDs and EIDs from L3 *Tb*-negative larvae were dissected. Heat shock was carried out for 45 min at 37 °C (52 ± 4 h after egg laying) to induce clone generation.

In *situ* hybridizations and immunohistochemistry. In *situ* hybridizations and immunostaining were carried out according to standard protocols. The cDNA for *pdm2* was PCR amplified using the primers listed in Supplementary Table 11 and cloned into a pBSK+/– vector at the EcoRI restriction site. Riboprobe was synthesized using T7 polymerase and digoxigenin-labeled ribonucleotides (Roche). Alkaline phosphatase conjugated antibody to digoxigenin (Roche, 1163716001) and NBT and BCIP (Roche) were used to develop the in *situ* hybridization. Peroxidase-conjugated antibody to digoxigenin and Tyramide signal amplification ( TSA, Life Technologies) were used for FISH analysis. WIDs and EIDs were analyzed with a DMLB microscope and an SPE confocal microscope (Leica). The primary antibodies used were rabbit antibody to H3K4me3 (1:1,000 dilution; Abcam, ab8580) and Millipore-Upstate, 07-473), H3K9ac (Abcam, ab4441), H3K36me3 (Abcam, ab9050), H3K4me1 (Diagenode, CS-037-100), H3K27ac (Abcam, ab4729) and H3K27me3 (Upstate-Millipore, 07-449).

Chromatin immunoprecipitation. L3 WIDs or EIDs isolated from Canton *S* flies were fixed, pooled in 700 µl of sonication buffer (10 mM Tris-HCl, pH 8.0, 2 mM EDTA and 1 mM EGTA) and processed as described17. Around 300-400 nuclei were used in these experiments. Trypsin-treated cells from GFP-transgenic flies were fixed after sorting for 10 min at room temperature and sonicated with a Diagenode Bioruptor for 15 min at high power in lysis buffer (1% SDS, 10 mM Tris-HCl, pH 8.0 and 2 mM EDTA). Immunoprecipitations were performed in RIP buffer. For L3 ChIP assays and immunoprecipitation experiments, we used 1 µg of the corresponding antibody. For ChIP assays in sorted cells, we used 0.45 µg of antibody to H3K4me3, 0.3 µg of antibody to H3K36me3, 0.33 µg of antibody to H3K27ac and 1 µg of antibody to H3K27me3. For L3-specific ChIP assays, five wall-wandering L3 Canton *S* larvae were disrupted, fixed and sonicated as described above. Immunocomplexes were recovered by incubation with Invitrogen Protein A magnetic beads for 2 h. The beads were washed three times in RIPA or immunoprecipitation buffer, once in lithium chloride buffer and twice in TE buffer17. The primers used for RT-PCR are listed in Supplementary Table 11. The antibodies used for the ChIP assays were to H3 (Abcam, ab1791), H3K4me3 (Abcam, ab8580 and Millipore-Upstate, 07-473), H3K9ac (Abcam, ab4441), H3K36me3 (Abcam, ab9050), H3K4me1 (Diagenode, CS-037-100), H3K27ac (Abcam, ab4729) and H3K27me3 (Upstate-Millipore, 07-449).

Nascent RNA. For nascent RNA assays, 400 WIDs from *nub-GAL4/+; UAS-GFP/+* flies were dissected and disaggregated as described above. The Click-IT Nascent RNA Capture kit (Molecular Probes, C10635) was used according to the manufacturer’s instructions. Briefly, disaggregated cells were incubated with 0.5 µM 5-ethyl thymidine in Schneider’s insect medium for 1 h at room temperature. Total RNA was extracted and biotinylated with 0.25 mM biotin azide for 30 min at room temperature. Biotinylated RNA was precipitated overnight at −80 °C and purified with streptavidin-conjugated beads for 30 min at room temperature. Nascent RNA was eluted in 0.1% SDS for 5 min at 99 °C, and reverse transcription was carried out as described above. Four biological replicates were analyzed. The primers used for RT-PCR are listed in Supplementary Table 11.

Solexa/Illumina sequencing. Solexa/Illumina sequencing was carried out at the Ultrassequencing Unit of the Centre for Genomic Regulation (CRG, Barcelona, Spain). All protocols for Solexa/Illumina ChIP-seq and RNA-seq analyses were carried out according to the manufacturer’s instructions. For ChIP-seq, 10 ng of each sample was used and fragments between 300 and 350 bp were size selected before sequencing. For RNA-seq, 5 µg of total RNA was used for sequencing.

*Drosophila* genome and annotation. We used FlyBase12 annotation release 5.12 for genome version dm3.

RNA-seq and ChIP-seq read mapping. Reads of 36 and 40 bp obtained from single-end sequencing in RNA-seq and ChIP-seq for WID and EID cells were aligned using GEM59, allowing up to two mismatches to the *Drosophila* genome (version dm3), and reads for RNA were aligned to all possible junctions of 5’ to 3’-ordered exon pairs occurring within the same annotated gene. ChIP-seq and RNA-seq raw data and profiles of read counts have been deposited in GEO under accession GSE56551.

Gene and transcript expression analysis. Reads mapping uniquely to the genome were used to quantify genes and transcripts separately in each tissue using the FluxCapacitor60. Expression levels are given in RPKM. Linear regression analysis between log-transformed WID and EID RPKM values gave a highly significant slope and intercept. Thus, we identified 628 genes at least one unit above the linear regression line (differentially expressed genes in EIDs) and 184 genes at least one unit below the linear regression line (differentially expressed genes in WIDs). To build our collection of regulated, tissue-specific genes from each differentially expressed gene set, we required coefficient of variation of at least 1.2 and an RPKM of at least 1.5 in one tissue and less than 0.1 RPKM in the other tissue (55 EID-specific genes and 10 WID-specific genes resulted from these criteria). Finally, genes with coefficient of variation <1.2 that were expressed in both tissues (>2.3 RPKM) with a difference in expression of less than 20% were selected as stable expressed genes in the 2 tissues (284 genes), and the genes whose expression in both tissues was 0 RPKM were considered to be silent (30 genes).

ChIP-seq analyses. ChIP-seq reads for H3, H3K4me3, H3K9ac, H3K36me3, H3K4me1 and H3K27ac were extended to the full average fragment length in the corresponding experiment. For each position in the genome, the number of extended ChIP-seq reads overlapping this position was recorded. Each sample was normalized by the total number of sequenced reads and the average fragment length. The genome-wide correlation between WID and EID samples was computed using the UCSC Table browser on windows of 1,000 bp56. To compute the correlation between ChIP-seq data and RNA-seq expression data were aligned using GEM 59, allowing up to two mismatches to the *Drosophila* genome (version dm3), and reads for RNA were aligned to all possible junctions of 5’ to 3’-ordered exon pairs occurring within the same annotated gene. ChIP-seq and RNA-seq raw data and profiles of read counts have been deposited in GEO under accession GSE56551.

Gene and transcript expression analysis. Reads mapping uniquely to the genome were used to quantify genes and transcripts separately in each tissue using the FluxCapacitor60. Expression levels are given in RPKM. Linear regression analysis between log-transformed WID and EID RPKM values gave a highly significant slope and intercept. Thus, we identified 628 genes at least one unit above the linear regression line (differentially expressed genes in EIDs) and 184 genes at least one unit below the linear regression line (differentially expressed genes in WIDs). To build our collection of regulated, tissue-specific genes from each differentially expressed gene set, we required coefficient of variation of at least 1.2 and an RPKM of at least 1.5 in one tissue and less than 0.1 RPKM in the other tissue (55 EID-specific genes and 10 WID-specific genes resulted from these criteria). Finally, genes with coefficient of variation <1.2 that were expressed in both tissues (>2.3 RPKM) with a difference in expression of less than 20% were selected as stable expressed genes in the 2 tissues (284 genes), and the genes whose expression in both tissues was 0 RPKM were considered to be silent (30 genes).

ChIP-seq analyses. ChIP-seq reads for H3, H3K4me3, H3K9ac, H3K36me3, H3K4me1 and H3K27ac were extended to the full average fragment length in the corresponding experiment. For each position in the genome, the number of extended ChIP-seq reads overlapping this position was recorded. Each sample was normalized by the total number of sequenced reads and the average fragment length. The genome-wide correlation between WID and EID samples was computed using the UCSC Table browser on windows of 1,000 bp56. To compute the correlation between ChIP-seq data and RNA-seq expression data.
data, we assigned to each gene the highest peak of the corresponding ChIP signal within the gene body and correlated this value to the expression of the gene. To produce the graphical distribution of reads for each sample around a particular site, we calculated the weighted number of reads at each position from −500 bp to +500 bp for each TSS, polyadenylation site and splice acceptor site, according to FlyBase. To graphically represent an idealized gene, we normalized the location of the reads within the gene using a window of 100 units and calculated the mean at each position. We extended this representation 500 bp upstream and downstream of the gene. To compare WID and EID samples, we calculated the weighted number of reads at each position in the normalized ChIP-seq profiles.

**ENCODE and Roadmap Epigenomic analyses.** Stable and developmentally regulated genes in *Drosophila.* To define the transcriptional stability of genes, we calculated the coefficient of variation of gene expression, as reported by the modENCODE project 50. For each protein-coding gene that had detectable expression at 12 selected developmental time points (Supplementary Fig. 1a). From the full ranking of 13,635 genes, we defined the bottom 1,000 genes with the lowest variation in expression during development as stable genes and the top 1,000 genes with the highest variation in expression as developmentally regulated genes. In addition, at each time point, we selected the same number of silent genes as regulated genes expressed at that time point, for a total of 1,000 silent genes. For these genes, we measured the strength of the highest peak (measured as the log of the number of reads reported by modENCODE) within the gene body at the time point when each gene’s expression was the highest for H3K4me3, H3K9ac, H3K4me1 and H3K27ac and the average signal within the gene body for H3K27me3 and H3K9me3 modENCODE ChIP-seq profiles (GEO GSE16013). Because of issues with the ChIP-seq data for three samples, H3K9ac (adult male) and H3K9me3 (L3 and adult male), we used ChIP-chip data in these cases instead. The Wilcoxon test (two-sided) was used to evaluate the statistical significance of the difference between ChIP values for stable, regulated and silent genes in each sample. To build the subsets of genes with low, medium and high regulation, we ranked the top 1,000 regulated genes by their expression (at the time point with maximum expression), and we classified them into 3 groups with the same number of genes. Partial correlations between the coefficient of variation and the histone marking of genes, with the effect of the expression of such genes removed, were calculated with the ggm R package.

**L3-specific genes analysis.** To compare the expression and histone modification levels for regulated, broadly expressed and stable, tissue-specific genes, we used anatomy RNA-seq data from the modENCODE project available in FlyBase 52. We used the gene sets previously defined for modENCODE analysis to create new subgroups of genes:

- **Stable:** the 1,000 genes with the lowest coefficients of variation for gene expression across modENCODE time points.
- **Silent:** genes identified as silent at L3 (RPKM = 0).
- **Regulated:** broadly expressed at L3: developmentally regulated genes that were detected in L3 whole-body data and that were furthermore expressed with at least 1 RPKM in each of the six tissues with L3 tissue-specific RNA-seq data available.
- **Stable, tissue-specific at L3:** from the set of extended stably expressed genes (P1 in Supplementary Fig. 4), we selected the genes that, using L3 tissue-specific RNA-seq data, were detected as expressed with at least 10 RPKM in one of the tissues and had expression of no higher than 1 RPKM in all the other tissues. We identified 26 carcass-specific genes, 8 central nervous system–specific genes, 36 digestive-specific genes, 21 fat body–specific genes, 36 imaginal disc–specific genes and 4 salivary gland–specific genes.

Expression and histone modification levels were calculated using L3 data from modENCODE following the methodology of the previous analysis.

**Stable and developmentally regulated genes in *C. elegans.** We estimated H3K3me3 and H3K36me3 levels in seven developmental stages (early embryo, late embryo, larval L1, L2, L3 and L4, and young adult) from the array signal files in Gerstein et al. 35. To define developmentally stable and regulated genes, we used the same procedure as in fly. To obtain gene and transcript quantifications, we mapped the RNA-seq reads from modENCODE *C. elegans* 35 to the Wbcde215.68 version of the genome using GEM 50 and used the FluxCapacitor 60 to produce the quantifications. Partial correlations between the coefficient of variation and the histone marking of genes, with the effect of the expression of such genes removed, were calculated with the ggm R package.

**Human and mouse analyses.** To define the transcriptional stability of human genes, we calculated the coefficient of variation for gene expression, as reported by the Roadmap Epigenomics Consortium, for each protein-coding gene that had detectable expression in the set of 56 consolidated epigenomes 14. From the full ranking of 18,064 genes, we defined as constant genes the bottom 1,000 genes with the lowest variation in expression across the 56 tissues and cell lines and as variable genes the top 1,000 genes with the highest variation in expression. In addition, for each epigenomes, we selected the same number of silent genes as variable genes expressed in that tissue, for a total of 1,000 silent genes. For these genes, we measured the strength of the highest peak (measured as the log of the number of reads reported by the Roadmap Epigenomics Consortium) within the gene body in the tissue in which expression was the highest for H3K4me3, H3K36me3 and H3K4me1. The Wilcoxon test (two-sided) was used to evaluate the statistical significance of the difference between ChIP values for constant, variable and silent genes in each sample. Partial correlations between the coefficient of variation and the histone marking of genes, with the effect of the expression of such genes removed, were calculated with the ggm R package.

**Nucleosome turnover.** Using the provided nascent RNA signal tracks in S2 cells 48 with no treatment, we calculated the average signal in the gene body of genes in the previously defined stable, regulated and silent gene sets. Stable and regulated genes with average normalized counts greater than 1 were kept (986 stable and 56 regulated genes), and silent genes with a signal of 0 were also kept (258 genes). For these selected genes, we defined the nucleosome turnover rate as the average CATCH-IT signal within the gene body in S2 cells with no treatment 19. The Wilcoxon test (two-sided) was used to evaluate the statistical significance of the signal among the gene sets.

**Promoter analyses.** To measure the conservation of the promoters of regulated and stable genes across 12 drosophilds, we computed the average of the UCSC phastCons multiz15way track 36 along the promoter sequences of each gene set (promoter length, 200 bp). To characterize the promoters of regulated and stable genes, we used the MatScan program 61 with the full collection of 827 predictive matrices available in Jaspar and Transfac 62,63. From each initial pool of predictions, we removed binding sites in genome regions in the UCSC Genome Browser that presented on average a probability lower than 0.95 of being conserved across the 12-drosophild phastCons multiz15way alignments 64. The Wilcoxon test (one-sided) was used to evaluate the statistical significance of the difference for the stable and developmentally regulated gene sets in each comparison (phastCons scores and number of conserved sites). To identify focused and dispersed initiation sites 65,66, we searched for putative binding sites for TBP and DRE in the promoter sequences of the top 1,000 stable genes and the top 1,000 regulated genes (promoter length, 100 bp). We selected TBP as a marker of focused initiation and DRE as a representative of dispersed initiation. The weight matrix for TBP is from Jaspar 6, and the weight matrix for DRE is from Fly Factor Survey 65. PCA was performed using the ChIP-chip levels of 20 transcription factors in embryos at 0–12 h in the promoter regions of stable and regulated genes with microarray signal intensity greater than 10 at this time point 39.

**Genome segmentation.** To match the states of a particular map of genome segmentation and our sets of stable and regulated genes, we counted how many genes from these two groups overlapped with the segments of each state. To annotate our collection of genes, we used the modENCODE ChromHMM 46 maps for BG3 and S2 cell lines 42, the hiHMM maps for late embryo and L3 larvae 43, and the chromatin types identified by Filion et al. 41. To annotate
the topological information for stable and regulated genes, we conducted a similar analysis on the HiC genome domains previously identified in late embryo and the LADs reported in Kc cells.

**Transcription factor perturbation analysis.** To study the effects of transcription factors at stable and regulated genes, we analyzed publicly available data on knockdown or overexpression of various *Drosophila* transcription factors. First, we checked how many stable and regulated genes were expressed in the tissue or cell type used in each study before the perturbation of the transcription factor, using published expression data on L3 brain, Kc cells and S2 cells and our data on L3 EIDs. Genes with RPKM > 1 were considered expressed. Then, we intersected the stable and regulated expressed genes with the genes identified as differentially expressed in each study.

**Splicing entropy.** For each gene, we computed the Shannon's entropy (or diversity index) on the basis of the relative frequencies of the gene's annotated isoforms in a given cell line. Letting \( g \) be a gene with \( n \) annotated isoforms with relative frequencies \( p_1, p_2, \ldots, p_n \) under a given condition, the entropy of \( g \), \( H(g) \), is computed as

\[
H(g) = - \sum_{i=1}^{n} p_i \log p_i
\]

\( H(g) \) increases with the number of annotated isoforms and with the evenness of their frequencies. \( H(g) \) is zero when there is only one expressed isoform (which would correspond to tight regulation of isoform expression), and it is at its maximum when all isoforms are equally expressed (which would correspond to a lack of splicing regulation and stochastic production of alternative splicing isoforms). On the basis of transcript quantifications generated by the modENCODE project for fly and computed by us for worm, we calculated the splicing entropy of each gene at the developmental time point when its expression was at its maximum. The box plots in Figure 8f display the distributions of \( H(g) \) separately for genes with different numbers of isoforms.

57. Xu, T. & Rubin, G.M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* 117, 1223–1237 (1993).
58. Cagan, R.L., Kramer, H., Hart, A.C. & Zipursky, S.L. The bride of sevenless and sevenless interaction: internalization of a transmembrane ligand. *Cell* 69, 393–399 (1992).
59. Derrien, T. et al. Fast computation and applications of genome mappability. *PLoS ONE* 7, e30377 (2012).
60. Montgomery, S.B. et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464, 773–777 (2010).
61. Blanco, E., Messeguer, X., Smith, T.F. & Guigo, R. Transcription factor map alignment of promoter regions. *PLoS Comput. Biol.* 2, e49 (2006).
62. Portales-Casamar, E. et al. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 38, D105–D110 (2010).
63. Wingender, E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.* 9, 326–332 (2008).
64. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050 (2005).
65. Zhu, L.J. et al. FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* 39, D111–D117 (2011).
66. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9, 215–216 (2012).
67. Cherbas, L. et al. The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res.* 21, 301–314 (2011).