Developmental Biology 339 (2010) 230–239

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Review

Developmental gene regulation in the era of genomics

Julia Zeitlinger [a,b,*], Alexander Stark [c]

[a] Stowers Institute for Medical Research, Kansas City, MO 64110, USA
[b] University of Kansas Medical Center, Kansas City, KS 66103, USA
[c] Research Institute of Molecular Pathology (IMP), A-1030 Vienna, Austria

ARTICLE INFO

Article history:
Received for publication 24 July 2009
Revised 4 December 2009
Accepted 23 December 2009
Available online 4 January 2010

Keywords:
Genetics
Genomics
Development
Drosophila
RNA Pol II stalling
Polycomb
Chromatin

ABSTRACT

Genetic experiments over the last few decades have identified many developmental control genes critical for pattern formation and cell fate specification during the development of multicellular organisms. A large fraction of these genes encode transcription factors and signaling molecules, show highly dynamic expression patterns during development, and are deeply evolutionarily conserved and deregulated in various human diseases such as cancer. Because of their importance in development, evolution, and disease, a fundamental question in biology is how these developmental control genes are regulated in such an extensive and precise fashion. Using genomics methods, it has become clear that developmental control genes are a distinct group of genes with special regulatory characteristics. However, a systematic analysis of these characteristics has not been presented. Here we review how developmental control genes were discovered, evaluate their genome-wide regulation and gene structure, discuss emerging evidence for their mode of regulation, and estimate their overall abundance in the genome. Understanding the global regulation of developmental control genes may provide a new perspective on development in the era genomics.

Introduction

Around the turn of this century, biology witnessed a tremendous increase in the development of new technology. Systematic RNAi screens, microarrays, genome sequencing, and next-generation sequencing technologies now allow the genome-wide assessment of sequence conservation, transcript abundance, protein–DNA interactions, and gene function. These advances have shifted many fields, such as developmental biology, to adopt genome-wide perspectives.

A marriage with new technology is not new to the field of developmental biology, which examines how a single cell develops into a complex multicellular organism. Initially restricted to detailed descriptions and transplantation experiments (e.g., Spemann and Mangold, 2001), the systematic use of genetics and molecular biology in the 20th century transformed developmental biology and gave rise to the field of developmental genetics. Through systematic genetic screens, key developmental control genes were identified and placed into molecular pathways. These molecular pathways now form the basis of our current understanding of cell–cell communication and pattern formation during development.

A surprising finding was that pattern formation across the animal kingdom occurs by similar principles and even similar cellular and molecular mechanisms. There are relatively few developmental pathways within an organism, and they are evolutionarily highly conserved among diverse organisms such as hydra, flies, worms, and vertebrates (Wolpert, 1994). For this reason, much of our knowledge about human development and disease today comes from studies of model organisms, which are experimentally more accessible.

Will genomics technologies, which allow the systematic investigation of biological questions on a genome-wide scale, have an impact on developmental biology similar to genetics in the 20th century? At first glance, it seems obvious that genomics has information to offer that is welcomed by developmental biologists. First, genomics experiments can often quickly generate testable hypotheses on how to fill in gaps of knowledge in a specific developmental system. Second, genomics methods provide an alternative approach to consolidate or reevaluate some of the principles uncovered by developmental biology. For example, developmental control genes and pathways are surprisingly limited in number, and developmental biologists have often wondered whether the majority of genes are not relevant for development (Nusslein-Volhard, 1994). With the recent progress in technology, we can now ask the following questions in a more systematic fashion: how many developmental genes are there in the genome? What distinguishes them from other genes? What is the role of other genes in development?

At a second glance though, it appears that there is some tension that arises from the different conceptual approaches utilized in genomics and developmental genetics. Traditional developmental genetics focuses on careful molecular analysis on a few relevant genes, while genomics examines all genes and often sacrifices depth for breadth. Thus, the genome-wide perspective is broader than traditional developmental genetics, but it also tends to be noisier, more

* Corresponding author.
E-mail address: jbz@stowers.org (J. Zeitlinger).

0012-1606/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2009.12.039

descriptive, and less mechanistic. In order to maximize the complementarity of these approaches, it is therefore important to appreciate the strengths and weaknesses of each. On one hand, genomics would not be possible without the strong foundation of detailed knowledge from developmental genetics. On the other hand, the new technology in genomics facilitates and accelerates biological insights in an unprecedented fashion. New insights can then be tested and followed up by more traditional analyses of specific examples.

In this review, we discuss how some of the recent progress in genomics has converged with more traditional approaches to advance our understanding of developmental biology. We will revisit the discovery of developmental control genes and discuss their properties from a genomics perspective with an emphasis on gene structure and regulation. We then review recent genomics studies that revealed some surprising properties of developmental control genes identified by genetic approaches. Developmental control genes are not only extensively regulated but also appear to be transcriptionally regulated by distinct mechanisms since they preferentially show stalled RNA polymerase II and occupancy by polycomb-group proteins. These findings strengthen the idea that developmental control genes represent a distinct class of genes with defined properties.

What are developmental control genes?

Developmental control genes are those genes that are critical for pattern formation and cell fate specification during the development of multicellular organisms and are defined by their specific mutant phenotypes in genetic screens (Dickmeis and Muller, 2005). Since the detection of a genetic phenotype depends on the experimental assay and the developmental context studied, it has not been possible to generate a complete list of all developmental control genes in the genome. It is generally thought, however, that not all genes behave as developmental control genes. For example, some genes are required in all cells ('housekeeping genes'), while other genes have highly specialized functions and are only required in certain differentiated cells or only under certain environmental conditions ('effector genes'). The loss-of-function phenotype of these two groups can be cell-lethal, pleiotropic, nonspecific, or very subtle and therefore has been difficult to study. Developmental control genes not only have a defined mutant phenotype, but also, their structure, function, and identity have been remarkably consistent across various genetic screens and across various model organisms.

The first systematic genetic screen for embryonic phenotypes was performed in Heidelberg, Germany, in the late 1970s by C. Nüsslein-Volhard and E. Wieschaus using *Drosophila* as model system. One hundred thirty-nine mutations were identified that caused an abnormal zygotic cuticle phenotype in the embryo. This result was revolutionary since genetics had previously been mostly applied to study the transmission of adult phenotypes and it was not clear whether mutations would frequently lead to distinct developmental defects. The impact of this screen on the field of developmental biology began to unfold during the 1980s, and the screen was awarded the Nobel Prize in 1995.

When the first genes were cloned from such genetic screens, it was surprising how many of them encoded transcription factors and components of signal transduction pathways. Based on recent gene annotations, three-quarters of the mutated genes from the Heidelberg screen have been identified. More than 42% of them encode site-specific transcription factors (FlyBase.org), although this class of transcription factors only represents 6% of all *Drosophila* genes (Adryan and Teichmann, 2006). When only patterning defects are analyzed, the fraction of transcription factor mutations is even higher: for example, among the 24 genes involved in segmentation, 75% encode site-specific transcription factors, and the remaining ones encode highly conserved signal transduction molecules. These numbers are intriguing, as the screen merely selected for developmental phenotypes but was unbiased with respect to genes.

The prevalence of transcription factors and signaling molecules in pattern formation is also observed in vertebrates. For example, among 70 known genes identified in genetics screens for axis formation and patterning in the zebrafish embryo, 34% encode transcription factors and 57% signaling molecules (Schier and Talbot, 2005).

A more systematic way to identify the relationship between genes and phenotype has become possible with the development of genome-wide RNAi screens. The first screens in *Caenorhabditis elegans* have confirmed that the knockdown of different classes of genes tend to cause different developmental defects (Kamath et al., 2003). While genes that cause lethality when knocked down are mostly involved in protein, RNA, and DNA metabolism, genes that cause a specific phenotype are enriched for transcription factor and signaling genes. The distinction is not clear cut, however, since some functional classes of genes, such as those involved in chromatin regulation and cellular architecture, are enriched among both lethal and viable phenotypes.

Not all developmental control genes encode transcription factors and signaling molecules. Specific developmental defects can be observed by loss-of-function phenotypes in some cell adhesion molecules, cell cycle regulators, chromatin regulators, or even structural molecules and ion channels (FlyBase.org, Schier and Talbot, 2005). Likewise, not all transcription factors are regulated and have tissue-specific functions. For example, general transcription factor genes are thought to be 'housekeeping' genes, although some of them may be expressed in a tissue-specific fashion and promote cell-type-specific terminal differentiation (Deato and Tjian, 2007, see also the review by Kadonaga in this issue). Furthermore, many developmental control genes are also required for the function of differentiated cells, making it hard to distinguish them from 'effector' genes. In summary, it is not straightforward to identify developmental control genes based on their molecular function, but the prevalence of transcription factors and signaling molecules among developmental control genes is significant.

The significance of developmental control genes

The strong enrichment of transcription factors and signaling molecules among developmental control genes has led to our current understanding of the mechanisms by which pattern formation during development is controlled. Briefly, pattern formation occurs through the interaction and communication of cells through signals and their corresponding signal transduction pathways. These signaling pathways specify cell fate by inducing a particular combination of transcription factors. These transcription factors in turn regulate the expression of further developmental control genes, and eventually tissue-specific 'effector' genes, which control morphogenesis, differentiation, and other tissue-specific functions (Busser et al., 2008; Davidson, 2006).

The importance of transcription factors in development and differentiation is well illustrated by what are often called master regulators. These transcription factors are both necessary and sufficient to initiate specific developmental programs. Examples include *Hox* genes, which can lead to the transformation of specific body segments when mutated (Hueber and Lohmann, 2008; Krumlauf, 1994; Morata, 1993), Pax6, which controls eye development (Baker, 2001), and MyoD, which is crucial for muscle formation (Olson, 1990).

More recent evidence for the dominant role of transcription factors in cell fate specification is the ability of ectopically expressed transcription factors to reprogram differentiated cells. Most notably, a variety of cells, including adult human somatic cells, can be reprogrammed to become pluripotent embryonic stem cells by a specific combination of transcription factors (Jaenisch and Young, 2008; Lewitzky and Yamanaka, 2007; Takahashi and Yamanaka,

2006). Thus, it appears that transcription factors can be sufficient for cell fate determination even when expressed in an artificial context.

Since signal transduction pathways and transcription factors work in a combinatorial manner, they can act in many different biological contexts during development and can regulate many different gene expression programs in different organisms. It is because of this versatility that the systematic genetic screens in *Drosophila* and *C. elegans* have been so relevant to human disease. Virtually all of the major signaling transduction pathways discovered in model organisms have been found to have homologous pathways in humans, and many have been shown to play a role in the etiology and metastasis of cancer (Dreesen and Brivanlou, 2007; Vidal and Cagan, 2006; Yang and Weinberg, 2008). Targeting these pathways by designing specific drugs is a major ongoing effort in the pharmaceutical industry (e.g., Arslan et al., 2006).

Extensive regulation of the regulators

A major hallmark of developmental control genes is that they are highly spatially and temporally regulated. Such dynamic regulation is expected since the genes have to act locally in order to affect specific developmental processes. Regulation occurs at all levels, from the transcriptional to the post-translational level. For example, many protein kinases, such as mitogen-activated protein kinases (MAPKs), are expressed at relative constant levels, but their activity is highly regulated as indicated by immunostainings with antibodies against the activated form (Gabay et al., 1997). Other examples include transcription factors whose activity is regulated by proteolytic cleavage or localization (e.g., Brivanlou and Darnell, 2002). Since global studies of gene regulation have so far been largely restricted to the DNA or RNA level, however, we will focus on those types of regulatory mechanisms.

Most developmental control genes are highly regulated at the level of transcription and have specific and dynamic expression patterns during development (Fig. 1). Many developmental control genes are expressed as highly dynamic and characteristic stripes in the early *Drosophila* embryo. With the recent availability of genome-wide datasets on gene expression in *Drosophila* (Arbeitman et al., 2002; Chintapalli et al., 2007; Manak et al., 2006; Tomancak et al., 2007), a more systematic assessment of the relationship between developmental control genes and expression patterns is possible.

A genome-scale *in situ* hybridization study of spatial expression patterns for 6003 (44%) genes in the *Drosophila* embryo was performed by Tomancak et al. (2007). The results suggest that genes can be classified into distinct classes. While 34% of all genes showed a spatially restricted expression pattern, 46% were expressed ubiquitously or broadly, and 19% appeared not to be expressed (Fig. 1A–C). The transcription factors from this study were specifically enriched among the spatially restricted genes throughout most of embryonic development and were depleted among the ubiquitously expressed genes. Likewise, signaling molecules were over-represented in the restricted gene category. At later stages of development, however, a restricted expression pattern was also observed for some specialized tissue-specific 'effector' genes. More broadly expressed genes, on the other hand, were likely to have general 'housekeeping' roles in cellular metabolism.

Gene expression during *Drosophila* embryogenesis has also been systematically examined by a time course of microarray expression analysis (Arbeitman et al., 2002; Hooper et al., 2007). Again, the relationship between gene expression dynamics and developmental function was striking. Developmental control genes such as those encoding transcription factors were particularly enriched among genes that are sharply up- and down-regulated during early embryogenesis (Fig. 1D), while genes that increased in expression during later stages, without subsequent decrease, encode more tissue-specific 'effector' genes such as structural constituents of the cuticle.

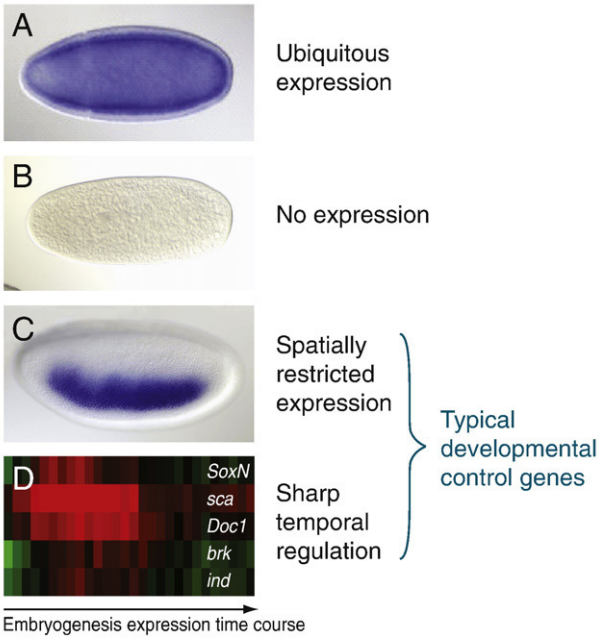

Fig. 1. Characteristic regulation of developmental control genes in *Drosophila* embryos. Genome-scale *in situ* hybridization of gene transcripts (blue) in *Drosophila* embryos show three classes of genes: (A) those that are broadly expressed, (B) those that are not expressed during development, and (C) those that are expressed in a spatially restricted fashion (Tomancak et al., 2007). The latter group is highly enriched for developmental control genes. (D) Genome-scale expression analysis during *Drosophila* embryogenesis (red = increased transcript abundance, green = decreased transcript abundance) shows that developmental control genes are often sharply up- and down-regulated in a temporal fashion (Hooper et al., 2007).

Transcripts that are present early in the embryo through maternal contribution mostly encode proteins involved in RNA and DNA metabolism as well as cell cycle ('housekeeping genes'), although transcription factors are also overrepresented in this category (Hooper et al., 2007).

Taken together, these results support the idea that the expression of developmental control genes is highly regulated in a spatial and temporal fashion during embryogenesis and that developmental control genes represent a distinct group of genes.

Size matters—evidence for regulation at the genome level

Gene expression during development is regulated by sequences around the gene called *cis*-regulatory modules or enhancers. Each enhancer is typically around 500 bp long and drives expression of the target gene in a certain spatial and temporal pattern during development. The overall expression pattern of a gene is determined by the sum of its enhancers. Consequently, genes with highly complex expression patterns are expected to have more enhancers and longer intergenic regions.

Consistent with this hypothesis, developmental control genes with highly regulated expression patterns typically have large regulatory regions with numerous *cis*-regulatory elements, each directing the expression of the gene in a different pattern (Fig. 2; Carroll, 2008; Nelson et al., 2004). Some regulatory regions are so large that classic genetic screens identified several distinct regulatory mutations for individual genes that cause defects in different parts of the body. Examples include the *Hox* genes (Hueber and Lohmann, 2008; Morata, 1993), *even-skipped* (*eve*) (Fig. 2A; Lifanov et al., 2003; Small et al., 1992), and *decapentaplegic* (*dpp*) (Fig. 2B; Blackman et al., 1991; Hepker et al., 1999; Jackson and Hoffmann, 1994; Johnson et al., 2003; Manak et al., 1994; Schwyter et al., 1995; St Johnston et al., 1990). In fact, the gene name *decapentaplegic* refers to mutant alleles

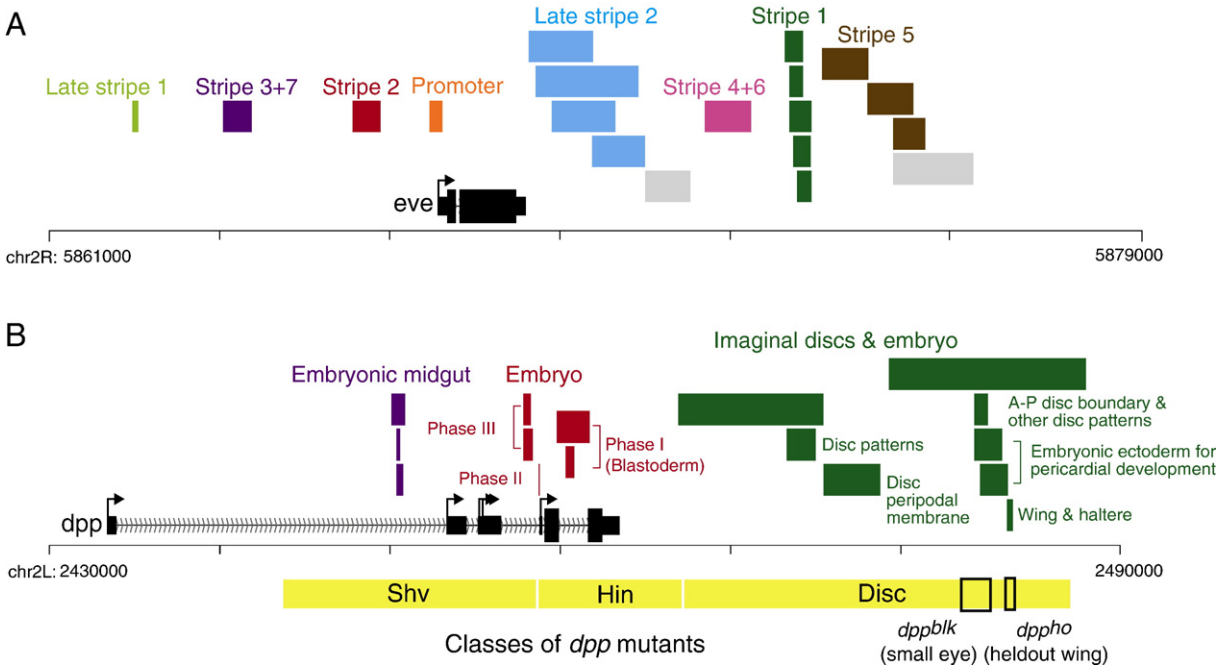

Fig. 2. Example of developmental control genes with large regulatory regions. (A) The even-skipped gene (*eve*, black) is expressed as seven stripes in the *Drosophila* embryo. Most stripes are driven by independent enhancers (shown in different colors), and some stripes even require different enhancers for earlier and later stages. These enhancers cover a large region around the gene. (B) The *decapentaplegic* gene (*dpp*, black) has numerous enhancers that drive reporter gene expression in various patterns in the embryo and in the imaginal discs (purple, red and green). Mutations in these regulatory regions of *dpp* have been identified and classified into three groups (yellow): those that produce a shortvein phenotype (Shv), those that are haploinsufficient (Hin), and those that affect imaginal discs (Disc) (St Johnston et al., 1990). Two well-known *dpp* regulatory mutants are dpp^{blk}, which produces a small eye, and dpp^{ho}, which produces a heldout wing phenotype. Both the reporter gene analysis and the genetic analysis of regulatory mutants shown here represent only a fraction of the regulatory information required to express *dpp* in the appropriate patterns during development.

that affect all 15 imaginal discs and the corresponding adult body structures (*decapenta-*, fifteen, *-plegic*, paralysis; Spencer et al., 1982).

An analysis using the most recent gene annotations shows that developmental control genes cover much larger intergenic regions, that is, the distance to the two neighboring genes is longer (Fig. 3; see also Nelson et al., 2004). The majority of genes identified from the Heidelberg screen by Nüsslein-Volhard and Wieschaus, as well as *bona fide* transcription factors and segmentation genes, have substantially larger intergenic regions than 'housekeeping' genes such as those involved in protein, DNA, or RNA metabolism (Fig. 3A). Long intergenic regions are also found for genes from Gene Ontology categories that contain a high number of developmental control genes such as "transcription factor activity," "signal transduction", or even "cell adhesion." In contrast, budding yeast has much smaller intergenic regions and does not show the same variation in intergenic length between the same Gene Ontology categories (Fig. 3B, see also Kristiansson et al., 2009). This is consistent with the fact that budding yeast is not a multicellular organism and does not require an extensive list of developmental control genes for patterning and cell fate specification.

Developmental control genes tend to have longer introns in addition to intergenic regions. This is consistent with the fact that introns also frequently harbor enhancers. Thus, developmental control genes do not only have longer intergenic regions, but also, the transcript itself covers much larger regions. The correlation between gene region sizes and gene function (e.g., as determined by Gene Ontology) is so striking that most analyses that are influenced by gene region size would inevitably enrich for Gene Ontology categories related to developmental regulation (Stanley et al., 2006).

Recent computational studies using comparative genomics across the twelve entirely sequenced *Drosophila* species suggest that the longer intergenic and intronic regions of developmental control genes also harbor an increased number of transcription factor binding sites. Genes with tissue-specific expression patterns and those involved in morphogenesis, organogenesis, and neurogenesis have the highest number of predicted regulatory motif instances, while ubiquitously expressed genes and those with housekeeping functions have the fewest, even when corrected for the different lengths (Stark et al., 2007). Transcription factors appear to be particularly heavily targeted by other transcription factors; they have twice as many predicted regulatory sites as the average gene. In addition, both insect and mammalian developmental regulators are significantly associated with ultraconserved elements, i.e., long, almost perfectly conserved non-coding regions (Bejerano et al., 2004; Glazov et al., 2005).

The same trend has been observed for 3′UTR lengths, suggesting extensive regulation of developmental control genes at the post-transcriptional level. While housekeeping genes tend to have short 3′ UTRs, developmental control genes tend to have long 3′UTRs with higher densities of microRNA target sites (Stark et al., 2005). In fact, transcription factors are one of the gene categories most frequently targeted by microRNAs across animals (Enright et al., 2003; Lewis et al., 2005; Stark et al., 2005, 2007). This argues that tight post-transcriptional regulation by microRNAs is an important feature of developmental control genes. Organisms like yeast that lack microRNAs do not show a significant difference of 3′UTR lengths between different functional gene categories (Stark et al., 2005).

Developmental control genes are 'poised for transcription'

Recent evidence suggests that the tight transcriptional regulation of developmental control genes may also be achieved by additional regulatory mechanisms. Several independent studies probing for chromatin or transcriptional states across the genome have concluded that developmental control genes tend to have an open chromatin structure in their promoter regions and are poised for activation (Bernstein et al., 2006; Boyer et al., 2006; Guenther et al., 2007; Heintzman et al., 2009; Lee et al., 2006; Muse et al., 2007; Zeitlinger et al., 2007).

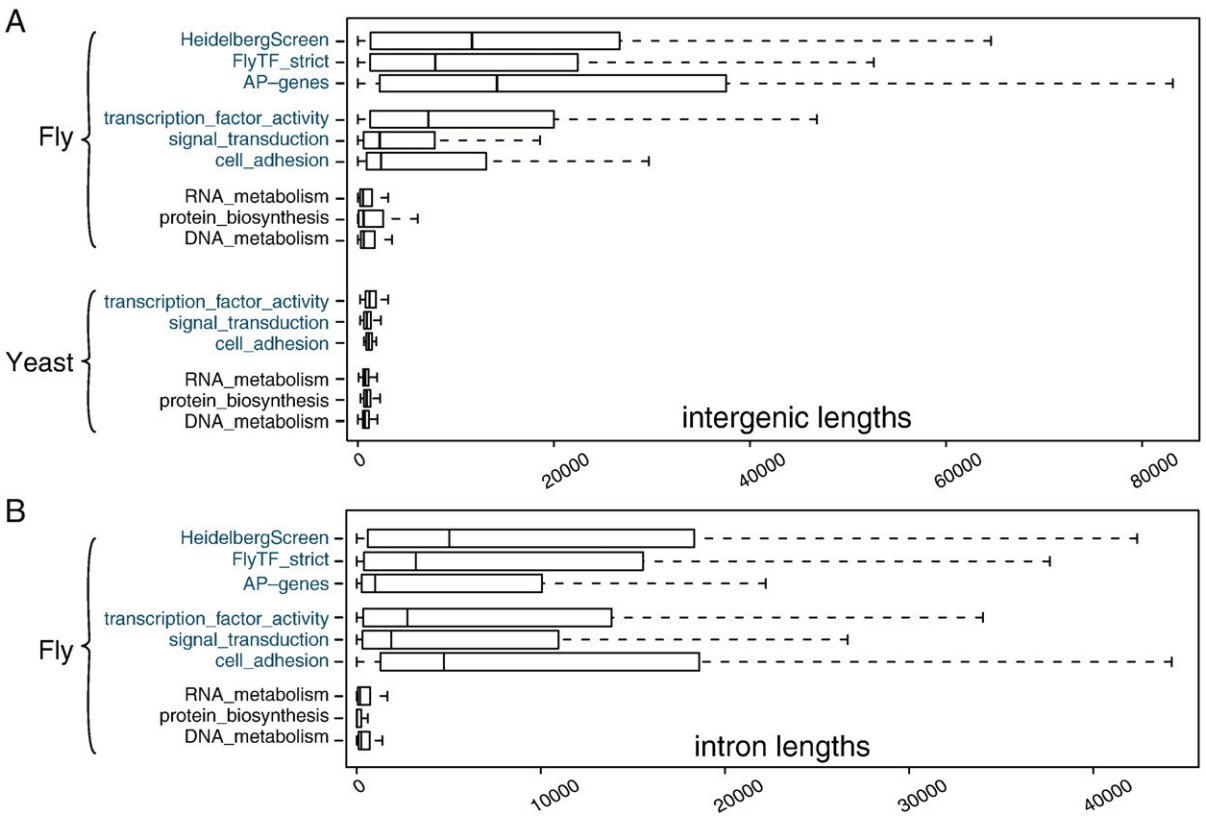

Fig. 3. Larger intergenic and intron regions at developmental control genes. Box and whisker plot of (A) the intergenic length of each gene, as measured by the distance in bp to the two neighboring genes, and (B) the total length of all introns in bp of each gene among various functional gene groups. Groups with a high fraction of developmental control genes (turquoise) have significantly longer regions in *Drosophila*. Shown are the genes identified by C. Nüsslein-Volhard and E. Wieschaus in their screen for embryonic mutants (HeidelbergScreen), sequence-specific transcription factors as defined by Adryan and Teichmann (2006) (FlyTF_strict) and well-known segmentation genes (AP genes) as defined by text books. Gene ontology categories that are also enriched for developmental control genes include "transcription_factor_activity," "signal_transduction," as well as "cell_adhesion." Gene ontology categories that represent 'housekeeping' genes include "RNA_metabolism," "protein_metabolism", and "DNA_metabolism". In budding yeast, the difference in intergenic region lengths between different functional groups is much smaller. This is because the yeast genome is more compact and does not include the regulatory regions required for the development of complex multicellular organisms.

This unexpected finding is well illustrated by genome-wide studies on RNA polymerase II. The traditional view on gene regulation has been that transcription factors bind to *cis*-regulatory elements and activate transcription by recruiting RNA polymerase II (Pol II) to the core promoter. This model predicts that Pol II is only found at promoters that are actively transcribed. However, genome-wide ChIP-chip and ChIP-seq studies of Pol II show that Pol II is present at many genes—especially those encoding developmental regulators—in the absence of transcription (Guenther et al., 2007; Muse et al., 2007; Schones et al., 2008; Wang et al., 2007; Zeitlinger et al., 2007). Thus, Pol II is recruited to these genes and stalls either as pre-initiation complex (PIC) (Soutoglou and Talianidis, 2002) or after promoter melting during the transition to productive elongation. The latter has been referred to as Pol II pausing or promoter–proximal pausing and has been well studied for *Drosophila hsp70* genes by John Lis and colleagues (reviewed by Lis, 2007). Since all stalled forms of Pol II are thought to prepare or facilitate rapid transcription, they are also referred to as poised Pol II.

Analysis of Pol II enrichment across the genome suggests that genes generally fall into three categories (Fig. 4; Muse et al., 2007; Zeitlinger et al., 2007). First, genes may have actively transcribing Pol II, which can be found uniformly across the entire open reading frame. Second, they may have the stalled form of Pol II, which accumulates at the 5′ end near the promoter. Finally, they may show no detectable Pol II. Interestingly, genes with uniform Pol II enrichment indicative of active transcription are often expressed broadly and correspond mostly to genes involved in protein, DNA, and RNA metabolism. Genes with stalled Pol II, on the other hand, are highly enriched for developmental control genes whose expression is regulated in a spatial and temporal fashion (Zeitlinger et al., 2007).

Functional anatomy of open promoters

The genome-wide studies raise the intriguing possibility that developmental control genes are regulated by mechanisms distinct from the traditional paradigm of transcription. A central theme proposed by many investigators has been that Pol II is present at developmental control genes in a poised form, thereby facilitating their rapid and tight regulation. The exact mechanisms, however, are not well understood. For example, it is not clear whether poised Pol II predominantly corresponds to paused Pol II, PIC, or other forms of Pol II and whether different regulatory mechanisms regulate the different forms.

Several lines of evidence indicate that the most common form of poised Pol II is paused Pol II rather than Pol II found in a PIC. First, the average peak of Pol II across the *Drosophila* genome lies approximately +50 bp downstream of the transcription start site (Fig. 5), at the expected location of paused Pol II. Second, close to 80% of genes with stalled Pol II in the genome are bound by Negative Elongation Factor (NELF) (Lee et al., 2008), a factor known to be associated with Pol II pausing (Yamaguchi et al., 1999). NELF is found at thousands of sites across the *Drosophila* genome, typically immediately downstream of the transcription start site, similar to paused Pol II (Fig. 5; Lee et al., 2008; see Peterlin and Price, 2006; Saunders et al., 2006 for further reading). Finally, permanganate footprints that probe for single-stranded DNA and therefore indicate open transcription bubbles are

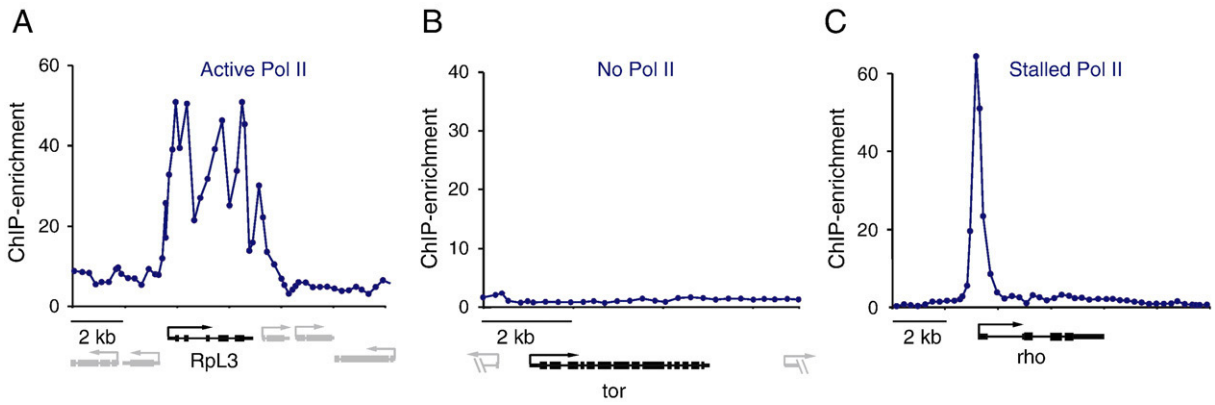

Fig. 4. Three classes of RNA polymerase II profiles in the early *Drosophila* embryo. Similar to the classification of genes based on their expression pattern during development, genes also fall into distinct classes based on their RNA polymerase II profile from ChIP-chip studies (Muse et al., 2007; Zeitlinger et al., 2007). (A) 'Housekeeping genes' such as ribosomal genes mostly show RNA polymerase II across the entire transcript region. (B) Genes that are not expressed during embryogenesis often have no RNA polymerase II enrichment at the gene. (C) Developmental control genes often have high levels of RNA polymerase II near the transcription start site but not in the transcribed region, indicating stalled RNA polymerase II (Muse et al., 2007; Zeitlinger et al., 2007).

found around +35 bp downstream of the transcription start site at many genes in the genome (Fig. 5; Lee et al., 2008).

Another important question is how Pol II pausing is established. In principle, there are two models, which are not necessarily mutually exclusive. Pol II could initiate and pause by default to be released upon transcriptional activation, or each of the steps could be independently regulated. Several lines of evidence support the default model of Pol II

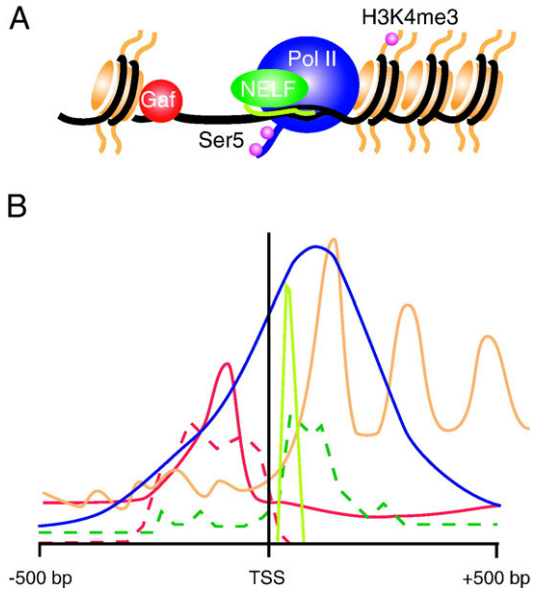

Fig. 5. Model of the open promoter structure at developmental control genes. (A) RNA polymerase II (blue) initiates transcription, leading to phosphorylation at serine 5 (Ser5) of its C-terminal domain (CTD) and resulting in trimethylation of histone H3 on lysine 4 (H3K4me3). Transcriptional initiation is facilitated by GAGA factor (Gaf, red) and is associated with a relative depletion of nucleosomes (orange) at the promoter. When the gene is not transcribed, RNA polymerase II fails to transition to productive elongation and stalls. Stalled RNA polymerase II is associated with a short transcript and an open transcription bubble (light green). An important regulator of stalling is Negative Elongation Factor (NELF, dark green). (B) Position of various components based on genome-wide data (solid lines) and or single gene data (dotted lines). The average profile of RNA polymerase II (blue) is from Zeitlinger et al. (2007). The average position of Gaf-binding sequences (red line) is from Hendrix et al. (2008). The position of Gaf binding (dotted red line) is from Lee et al. (2008). The average position of nucleosomes (orange) is from Mavrich et al. (2008). The average position of the open transcription bubble, as determined by permanganate footprinting (light green line), is from Mavrich et al. (2008) and Lee et al. (2008). The position of NELF binding (dotted dark green line) is from Lee et al. (2008).

pausing, but they do not exclude more extensive regulation. For example, NELF is not only found at paused genes but also at genes that are highly transcribed, suggesting that NELF-mediated pausing could be an important checkpoint, or an "integral but transient part of the transcription cycle" (Lee et al., 2008).

The model of default Pol II pausing is also consistent with an association of Pol II pausing and certain DNA sequence elements (Hendrix et al., 2008; Lee et al., 2008). Genes with Pol II pausing in *Drosophila* are enriched for core promoter elements that are bound by TFIID, including initiator (Inr) at the transcription start site and the downstream promoter element (DPE), which is located between +28 and +33 bp downstream of the Inr (see Juven-Gershon et al., 2008, for a review on core promoter elements). The "pause button" (PB), a novel CG-rich sequence motif, is also highly enriched at paused genes around the position of the DPE (Hendrix et al., 2008). Whether or not the PB and the core promoter elements directly regulate Pol II pausing remains to be shown. It is interesting, however, that developmental control genes are generally enriched for Inr and DPE elements (Engstrom et al., 2007).

In addition to core promoter elements, the GAGA motif is also enriched at paused genes in *Drosophila,* found most frequently around −100 bp relative to the transcription start site (Fig. 5B; Hendrix et al., 2008; Lee et al., 2008). At *Drosophila* heat shock genes, the GAGA motif has been shown to be required for the establishment of paused Pol II (Lee et al., 1992; Leibovitch et al., 2002; Shopland et al., 1995; Wang et al., 2005, reviewed in Adkins et al., 2006). It is recognized by the GAGA factor (Gaf), which is known to bind to many genes, including many developmental genes (van Steensel et al., 2003), and is indeed highly enriched at stalled and NELF-associated genes (Lee et al., 2008). Gaf is thought to promote nucleosome displacement by recruiting chromatin-remodeling complexes (Adkins et al., 2006; Tsukiyama et al., 1994), and there is evidence that Gaf together with TFIID is involved in generating an open chromatin structure at *Drosophila* heat shock genes (Leibovitch et al., 2002; Shopland et al., 1995). Taken together, these results suggest a global role for Gaf in the maintenance of an open chromatin structure at promoters and the establishment of Pol II pausing.

A main characteristic of an open promoter structure is a relative loss of nucleosome occupancy. Studies in yeast have shown that nucleosomes are disassembled during gene activation (Boeger et al., 2008; Boeger et al., 2003; Boeger et al., 2004; Henikoff, 2008; Reinke and Horz, 2003) and that the genome-wide nucleosome occupancy at promoters inversely correlates with gene activity (Bernstein et al., 2004; Lee et al., 2004; Pokholok et al., 2005; Yuan et al., 2005). In human CD4+ T cells, the correlation between nucleosome depletion

at promoters (referred to as −1 nucleosome) and gene activity is less pronounced. Instead, nucleosome depletion at promoters better correlates with Pol II occupancy (Schones et al., 2008). These results suggest that developmental control genes that have high levels of stalled Pol II also show high levels of nucleosome depletion upstream of the transcriptional start site.

The idea that developmental control genes have an open promoter structure is also consistent with other genome-wide studies. For example, the chromatin state at promoters of genes that are regulated in a tissue-specific fashion is largely invariant across diverse cell types in the mouse embryo (Heintzman et al., 2009). While the chromatin state dynamically changes at tissue-specific enhancers, the promoter state remains relatively constant (Heintzman et al., 2009). Promoters are also the most open sites based on DNAse hypersensitivity (DHS) mapping in human CD4+ T cells, more so than enhancers (Boyle et al., 2008). Thus, it appears likely that the dynamic regulation of developmental control genes during development is in part achieved by an open promoter structure that is amenable to rapid and precise regulatory inputs.

Actively transcribed genes in *Drosophila* and human cells also display a characteristic positioning and phasing of nucleosomes downstream of the transcription start sites. The first nucleosome appears to dictate the positioning of further downstream nucleosomes since its position is the most consistent across cells (Mavrich et al., 2008; Schones et al., 2008, Fig. 5). Promoters of genes with stalled Pol II also exhibit strong nucleosome phasing but the position of the first nucleosome downstream of the transcription start tends to be shifted relative to that of actively transcribed genes (Mavrich et al., 2008; Schones et al., 2008). At genes that are stalled in the *Drosophila* embryo, the first nucleosome is on average +145 bp downstream of the transcription start site, which is ~10 bp further downstream than at genes that are actively transcribed (Mavrich et al., 2008). In human CD4+ T cells, stalled promoters also have distinct first nucleosome positions. In this case, however, the first nucleosome in stalled promoters is found 30 bp further upstream compared to active promoters, i.e., +10 bp vs. +40 bp downstream of the transcription start site (Schones et al., 2008). It remains to be seen whether the position of the first nucleosome influences pausing or whether Pol II pausing causes a shift in nucleosomes.

Regulation of developmental control genes by polycomb

Polycomb-group (PcG) proteins have been shown to play an important role in the maintenance of gene repression during development. Numerous genome-wide studies have mapped components of PcG complexes by ChIP-chip in *Drosophila* (Bracken et al., 2006; Kwong et al., 2008; Negre et al., 2006; Oktaba et al., 2008; Schuettengruber et al., 2009; Schwartz et al., 2006; Tolhuis et al., 2006), mouse (Boyer et al., 2006; Squazzo et al., 2006), and human cells (Lee et al., 2006; Squazzo et al., 2006). In all studies, PcG proteins were found to be predominantly associated with developmental control genes.

The classic model suggests that PcG proteins, together with the counteracting trithorax group (trxG) proteins, are part of a cellular memory system that stably maintains the appropriate expression of developmental control genes through cell divisions after the inductive or repressive signals have disappeared (Ringrose and Paro, 2004; Schwartz and Pirrotta, 2007). This model is based on the regulation of *Hox* genes, whose compartmentalized expression is set up during segmentation in the *Drosophila* embryo and is maintained by PcG and trxG proteins throughout development into adult.

While the classic model of PcG function emphasizes stable maintenance of gene regulatory states, other studies including those using genome-wide approaches suggest a much broader role for PcG proteins in developmental gene regulation (Ringrose, 2007; Schwartz and Pirrotta, 2007). In mouse and human embryonic stem cells, PcG proteins are preferentially found at genes that are likely to be activated during differentiation (Boyer et al., 2006; Lee et al., 2006). In *Drosophila*, PcG proteins often bind to genes that are dynamically regulated throughout development (Ringrose, 2007; Schwartz and Pirrotta, 2007; see also Pelegri and Lehmann, 1994 for an example). Furthermore, PcG proteins bind to some extent in a cell-type-specific fashion, suggesting that their function changes during development (Mohn et al., 2008; Negre et al., 2006; Oktaba et al., 2008; Rinn et al., 2007; Squazzo et al., 2006).

A more refined model suggests that *PcG* and *trxG* genes maintain genes not only in a repressed and active state but also in a balanced/poised or mute state (Maurange et al., 2006; Schwartz and Pirrotta, 2008). The poised state is supported by genome-wide data on histone modifications in human embryonic stem cells. In these cells, many developmental control genes have 'bivalent domains,' that is, they have both H3K4 trimethylation and H3K27 trimethylation, which are active and repressive marks, respectively (Bernstein et al., 2004; Mikkelsen et al., 2007). H3K4 trimethylation is associated with trxG-mediated activation and is set right after RNA polymerase II (Pol II) has initiated transcription. H3K27 trimethylation is associated with PcG-mediated repression. The authors propose that bivalent domains silence developmental control genes while keeping them poised for activation.

While the poised or bivalent state of this model is reminiscent of the role proposed for stalled Pol II, the relationship between the two is still under investigation. Both PcG complex recruitment and stalled Pol II are facilitated by Gaf but the genome-wide Gaf binding sites only partially overlap with those of PcG proteins (Negre et al., 2006; Schuettengruber et al., 2009) and with stalled Pol II (Lee et al., 2008). This is in agreement with the finding that Gaf regulates genes that are not regulated by PcG proteins (e.g., *hsp70*, Tsukiyama et al., 1994). The number of PcG-bound genes is usually estimated to be smaller than the number of genes with stalled Pol II (Guenther et al., 2007; Lee et al., 2006; Muse et al., 2007; Ringrose, 2007; Zeitlinger et al., 2007). While one study finds that PcG proteins inhibit transcriptional initiation rather than elongation (Dellino et al., 2004), another study suggests that PcG proteins promote Pol II stalling at bivalent genes in mouse embryonic stem cells (Stock et al., 2007). Thus, the precise relationship between PcG function and Pol II stalling remains to be established.

While the exact mechanisms are unclear, genome-wide studies strongly support the conclusion that developmental control genes are regulated by special mechanisms. This is consistent with the fact that their expression is extensively regulated in a precise spatial and temporal fashion and is critical for development and disease. It will therefore be important to identify the molecular mechanisms by which developmental control genes are regulated and to determine the exact role of chromatin and epigenetic cellular memory during development.

Developmental control genes in the genomics era

In this review, we revisited the concept of developmental control genes as defined by developmental genetics and presented findings from recent genomic studies that have complemented and advanced our knowledge of developmental control genes. Our motivation stems from the observation that developmental control genes demonstrate unique features in such diverse aspects of gene regulation as Pol II stalling, nucleosome positioning, chromatin modifications, and binding by PcG proteins. Clearly, these characteristics, as well as the genomic organization of developmental control genes, are tightly connected to their dynamic and complex regulation during development. A combined effort of traditional biochemical and molecular genetics methods, as well as genomics approaches, will be required to further elucidate the mechanisms by which this occurs.

An additional benefit of extensive genomics studies is that they allow us to reevaluate, expand, or better quantify previously established

concepts in developmental biology. The evidence presented here has strengthened the idea that developmental control genes stand out as a group of highly regulated genes and can be distinguished from groups that are not subject to the same degree of regulation. However, are these characteristics unique enough to reliably identify or even redefine developmental control genes? While a highly desirable goal, we found that none of the characteristics are sufficient to define a comprehensive set of developmental regulators. We can only make use of the strong trends and enrichments to get estimates of their overall abundance.

In *Drosophila*, there are ~6% site-specific transcription factors (Adryan and Teichmann, 2006) and ~7% signal transduction genes (Ashburner et al., 2000). Based on *in situ* hybridization patterns in the *Drosophila* embryo, 34% of genes show spatially restricted expression (which includes many 'effector' genes) and 7% of genes are expressed in a restricted fashion during the blastoderm stages where the majority of pattern formation takes place (Tomancak et al., 2007). Twelve to twenty percent of all genes have intronic, intergenic or 3′UTR lengths comparable to the known developmental control genes (Fig. 3 and unpublished data), 10–12% of genes are estimated to show Pol II stalling (Muse et al., 2007; Zeitlinger et al., 2007), and 1–5% are bound by PcG proteins (Ringrose, 2007), although some PcG components may be found at a much larger fraction of genes (Schuettengruber et al., 2009). A rough estimate based on these criteria would thus predict that between 10% and 20% of all genes in the *Drosophila* genome correspond to developmental control genes. Since these estimates exceed the number of developmental control genes previously identified in genetic screens, it should be possible to use this information to help predict the function of uncharacterized genes during development.

Genomics studies have also begun to significantly contribute to other areas of developmental biology. For example, genomics methods now allow the systematic identification and analysis of *cis*-regulatory elements and how they control gene expression (Bonn and Furlong, 2008; Levine, 2008). Together with a systems-level understanding of how genes form developmental regulatory networks and control development (Alon, 2007; Busser et al., 2008; Jaenisch and Young, 2008), we could enter an era where development can be intimately linked to genomic information. We might be able to use genomic information to predict development, trace evolutionary changes, identify functionally relevant allelic variants, and understand the cellular and organismic response to environmental factors, diseases, and drug treatments.

Lewis Wolpert has argued more than a decade ago that the principles of development are understood, although many crucial details at the molecular level are missing (Wolpert, 1994). We anticipate that genomics approaches will not only help filling in crucial molecular details but will also provide new perspectives and surprises to developmental biology. Thus, genomics could become an integral part of traditional developmental biology and play a role similar to that of genetics in the last century.

Acknowledgments

We would like to thank Rick Young for initial discussions; Brianne Patton for researching and editing Fig. 2; Anaïs Bardet for help with Fig. 3; and Steven Chao, Brianne Patton, Carla Anderson, and Leonie Ringrose for their comments on the article.

References

Adkins, N.L., Hagerman, T.A., Georgel, P., 2006. GAGA protein: a multi-faceted transcription factor. Biochem. Cell. Biol. 84, 559–567.
Adryan, B., Teichmann, S.A., 2006. FlyTF: a systematic review of site-specific transcription factors in the fruit fly *Drosophila melanogaster*. Bioinformatics 22, 1532–1533.
Alon, U., 2007. Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8, 450–461.
Arbeitman, M.N., Furlong, E.E., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., White, K.P., 2002. Gene expression during the life cycle of *Drosophila melanogaster*. Science 297, 2270–2275.
Arslan, M.A., Kutuk, O., Basaga, H., 2006. Protein kinases as drug targets in cancer. Curr. Cancer Drug Targets. 6, 623–634.
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.
Baker, N.E., 2001. Master regulatory genes; telling them what to do. Bioessays 23, 763–766.
Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. Science 304, 1321–1325.
Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., Schreiber, S.L., 2004. Global nucleosome occupancy in yeast. Genome Biol. 5, R62.
Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., Lander, E.S., 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326.
Blackman, R.K., Sanicola, M., Raftery, L.A., Gillevet, T., Gelbart, W.M., 1991. An extensive 3′ cis-regulatory region directs the imaginal disk expression of decapentaplegic, a member of the TGF-beta family in *Drosophila*. Development 111, 657–666.
Boeger, H., Griesenbeck, J., Strattan, J.S., Kornberg, R.D., 2003. Nucleosomes unfold completely at a transcriptionally active promoter. Mol. Cell 11, 1587–1598.
Boeger, H., Griesenbeck, J., Strattan, J.S., Kornberg, R.D., 2004. Removal of promoter nucleosomes by disassembly rather than sliding in vivo. Mol. Cell 14, 667–673.
Boeger, H., Griesenbeck, J., Kornberg, R.D., 2008. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. Cell 133, 716–726.
Bonn, S., Furlong, E.E., 2008. *cis*-Regulatory networks during development: a view of Drosophila. Curr. Opin. Genet. Dev. 18, 513–520.
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., Bell, G.W., Otte, A.P., Vidal, M., Gifford, D.K., Young, R.A., Jaenisch, R., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349–353.
Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., Crawford, G.E., 2008. High-resolution mapping and characterization of open chromatin across the genome. Cell 132, 311–322.
Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., Helin, K., 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. 20, 1123–1136.
Brivanlou, A.H., Darnell Jr, J.E., 2002. Signal transduction and the control of gene expression. Science 295, 813–818.
Busser, B.W., Bulyk, M.L., Michelson, A.M., 2008. Toward a systems-level understanding of developmental regulatory networks. Curr. Opin. Genet. Dev. 18, 521–529.
Carroll, S.B., 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134, 25–36.
Chintapalli, V.R., Wang, J., Dow, J.A., 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. 39, 715–720.
Davidson, E.H., 2006. The Regulatory Genome: Gene Regulatory Networks In Development And Evolution, 1st ed. Academic Press.
Deato, M.D., Tjian, R., 2007. Switching of the core transcription machinery during myogenesis. Genes Dev. 21, 2137–2149.
Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C., Pirrotta, V., 2004. Polycomb silencing blocks transcription initiation. Mol. Cell 13, 887–893.
Dickmeis, T., Muller, F., 2005. The identification and functional characterisation of conserved regulatory elements in developmental genes. Brief Funct. Genomic Proteomic 3, 332–350.
Dreesen, O., Brivanlou, A.H., 2007. Signaling pathways in cancer and embryonic stem cells. Stem Cell Rev. 3, 7–17.
Engstrom, P.G., Ho Sui, S.J., Drivenes, O., Becker, T.S., Lenhard, B., 2007. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Res. 17, 1898–1908.
Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S., 2003. MicroRNA targets in Drosophila. Genome Biol. 5, R1.
Gabay, L., Seger, R., Shilo, B.Z., 1997. In situ activation pattern of *Drosophila* EGF receptor pathway during development. Science 277, 1103–1106.
Glazov, E.A., Pheasant, M., McGraw, E.A., Bejerano, G., Mattick, J.S., 2005. Ultraconserved elements in insect genomes: a highly conserved intronic sequence implicated in the control of homothorax mRNA splicing. Genome Res. 15, 800–808.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., Young, R.A., 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77–88.
Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X., Green, R.D., Lobanenkov, V.V., Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M., Ren, B., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459, 108–112.
Hendrix, D., Hong1, J.-W., Zeitlinger, J., Rokhsar, D.S., Levine, M., 2008. Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. Proc. Natl. Acad. Sci. U. S. A. 105, 7762–7767.
Henikoff, S., 2008. Nucleosome destabilization in the epigenetic regulation of gene expression. Nat. Rev. Genet. 9, 15–26.
Hepker, J., Blackman, R.K., Holmgren, R., 1999. *Cubitus interruptus* is necessary but not sufficient for direct activation of a wing-specific decapentaplegic enhancer. Development 126, 3669–3677.

Hooper, S.D., Boue, S., Krause, R., Jensen, L.J., Mason, C.E., Ghanim, M., White, K.P., Furlong, E.E., Bork, P., 2007. Identification of tightly regulated groups of genes during *Drosophila melanogaster* embryogenesis. Mol. Syst. Biol. 3, 72.

Hueber, S.D., Lohmann, I., 2008. Shaping segments: *Hox* gene function in the genomic age. Bioessays 30, 965–979.

Jackson, P.D., Hoffmann, F.M., 1994. Embryonic expression patterns of the *Drosophila* decapentaplegic gene: separate regulatory elements control blastoderm expression and lateral ectodermal expression. Dev. Dyn. 199, 28–44.

Jaenisch, R., Young, R., 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132, 567–582.

Johnson, A.N., Bergman, C.M., Kreitman, M., Newfeld, S.J., 2003. Embryonic enhancers in the dpp disk region regulate a second round of Dpp signaling from the dorsal ectoderm to the mesoderm that represses Zfh-1 expression in a subset of pericardial cells. Dev. Biol. 262, 137–151.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., Kadonaga, J.T., 2008. The RNA polymerase II core promoter—the gateway to transcription. Curr. Opin. Cell Biol. 20, 253–259.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J., 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature 421, 231–237.

Kristiansson, E., Thorsen, M., Tamas, M.J., Nerman, O., 2009. Evolutionary forces act on promoter length: identification of enriched *cis*-regulatory elements. Mol. Biol. Evol. 26, 1299–1307.

Krumlauf, R., 1994. Hox genes in vertebrate development. Cell 78, 191–201.

Kwong, C., Adryan, B., Bell, I., Meadows, L., Russell, S., Manak, J.R., White, R., 2008. Stability and dynamics of polycomb target sites in *Drosophila* development. PLoS Genet. e1000178, 4.

Lee, H., Kraus, K.W., Wolfner, M.F., Lis, J.T., 1992. DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes Dev. 6, 284–295.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., Lieb, J.D., 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat. Genet. 36, 900–905.

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., Koseki, H., Fuchikami, T., Abe, K., Murray, H.L., Zucker, J.P., Yuan, B., Bell, G.W., Herbolsheimer, E., Hannett, N.M., Sun, K., Odom, D.T., Otte, A.P., Volkert, T.L., Bartel, D.P., Melton, D.A., Gifford, D.K., Jaenisch, R., Young, R.A., 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125, 301–313.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., 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.

Leibovitch, B.A., Lu, Q., Benjamin, L.R., Liu, Y., Gilmour, D.S., Elgin, S.C., 2002. GAGA factor and the TFIID complex collaborate in generating an open chromatin structure at the *Drosophila melanogaster* hsp26 promoter. Mol. Cell. Biol. 22, 6148–6157.

Levine, M., 2008. A systems view of *Drosophila* segmentation. Genome Biol. 9, 207.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20.

Lewitzky, M., Yamanaka, S., 2007. Reprogramming somatic cells towards pluripotency by defined factors. Curr. Opin. Biotechnol. 18, 467–473.

Lifanov, A.P., Makeev, V.J., Nazina, A.G., Papatsenko, D.A., 2003. Homotypic regulatory clusters in *Drosophila*. Genome Res. 13, 579–588.

Lis, J.T., 2007. Imaging *Drosophila* gene activation and polymerase pausing in vivo. Nature 450, 198–202.

Manak, J.R., Mathies, L.D., Scott, M.P., 1994. Regulation of a decapentaplegic midgut enhancer by homeotic proteins. Development 120, 3605–3619.

Manak, J.R., Dike, S., Sementchenko, V., Kapranov, P., Biemar, F., Long, J., Cheng, J., Bell, I., Ghosh, S., Piccolboni, A., Gingeras, T.R., 2006. Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. Nat. Genet. 38, 1151–1158.

Maurange, C., Lee, N., Paro, R., 2006. Signaling meets chromatin during tissue regeneration in *Drosophila*. Curr. Opin. Genet. Dev. 16, 485–489.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., Gilmour, D.S., Albert, I., Pugh, B.F., 2008. Nucleosome organization in the *Drosophila* genome. Nature 453, 358–362.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., Bernstein, B.E., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., Schubeler, D., 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol. Cell 30, 755–766.

Morata, G., 1993. Homeotic genes of *Drosophila*. Curr. Opin. Genet. Dev. 3, 606–614.

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

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

Nelson, C.E., Hersh, B.M., Carroll, S.B., 2004. The regulatory content of intergenic DNA shapes genome architecture. Genome Biol. 5, R25.

Nusslein-Volhard, C., 1994. Of flies and fishes. Science 266, 572–574.

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

Olson, E.N., 1990. MyoD family: a paradigm for development? Genes Dev. 4, 1454–1461.

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

Peterlin, B.M., Price, D.H., 2006. Controlling the elongation phase of transcription with P-TEFb. Mol. Cell 23, 297–305.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K., Young, R.A., 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517–527.

Reinke, H., Horz, W., 2003. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Mol. Cell 11, 1599–1607.

Ringrose, L., 2007. Polycomb comes of age: genome-wide profiling of target sites. Curr. Opin. Cell Biol. 19, 290–297.

Ringrose, L., Paro, R., 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. 38, 413–443.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., Chang, H.Y., 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311–1323.

Saunders, A., Core, L.J., Lis, J.T., 2006. Breaking barriers to transcription elongation. Nat. Rev. Mol. Cell. Biol. 7, 557–567.

Schier, A.F., Talbot, W.S., 2005. Molecular genetics of axis formation in zebrafish. Annu. Rev. Genet. 39, 561–613.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., Zhao, K., 2008. Dynamic regulation of nucleosome positioning in the human genome. Cell 132, 887–898.

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

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

Schwartz, Y.B., Pirrotta, V., 2008. Polycomb complexes and epigenetic states. Curr. Opin. Cell Biol. 20, 266–273.

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

Schwyter, D.H., Huang, J.D., Dubnicoff, T., Courey, A.J., 1995. The decapentaplegic core promoter region plays an integral role in the spatial control of transcription. Mol. Cell. Biol. 15, 3960–3968.

Shopland, L.S., Hirayoshi, K., Fernandes, M., Lis, J.T., 1995. HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. Genes Dev. 9, 2756–2769.

Small, S., Blair, A., Levine, M., 1992. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. EMBO J. 11, 4047–4057.

Soutoglou, E., Talianidis, I., 2002. Coordination of PIC assembly and chromatin remodeling during differentiation-induced gene activation. Science 295, 1901–1904.

Spemann, H., Mangold, H., 2001. Induction of embryonic primordia by implantation of organizers from a different species. 1923. Int. J. Dev. Biol. 45, 13–38.

Spencer, F.A., Hoffmann, F.M., Gelbart, W.M., 1982. Decapentaplegic: a gene complex affecting morphogenesis in *Drosophila melanogaster*. Cell 28, 451–461.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.W., Margueron, R., Reinberg, D., Green, R., Farnham, P.J., 2006. Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res. 16, 890–900.

St Johnston, R.D., Hoffmann, F.M., Blackman, R.K., Segal, D., Grimaila, R., Padgett, R.W., Irick, H.A., Gelbart, W.M., 1990. Molecular organization of the decapentaplegic gene in *Drosophila melanogaster*. Genes Dev. 4, 1114–1127.

Stanley, S.M., Bailey, T.L., Mattick, J.S., 2006. GONOME: measuring correlations between GO terms and genomic positions. BMC Bioinformatics 7, 94.

Stark, A., Brennecke, J., Bushati, N., Russell, R.B., Cohen, S.M., 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. Cell 123, 1133–1146.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N., Ruby, J.G., Brennecke, J., Hodges, E., Hinrichs, A.S., Caspi, A., Paten, B., Park, S.W., Han, M.V., Maeder, M.L., Polansky, B.J., Robson, B.E., Aerts, S., van Helden, J., Hassan, B., Gilbert, D.G., Eastman, D.A., Rice, M., Weir, M., Hahn, M.W., Park, Y., Dewey, C.N., Pachter, L., Kent, W.J., Haussler, D., Lai, E.C., Bartel, D.P., Hannon, G.J., Kaufman, T.C., Eisen, M.B., Clark, A.G., Smith, D., Celniker, S.E., Gelbart, W.M., Kellis, M., 2007. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. Nature 450, 219–232.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., 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.

Takahashi, K., Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., 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., Rubin, G.M., 2007. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. Genome Biol. 8, R145.

Tsukiyama, T., Becker, P.B., Wu, C., 1994. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 367, 525–532.

van Steensel, B., Delrow, J., Bussemaker, H.J., 2003. Genomewide analysis of *Drosophila* GAGA factor target genes reveals context-dependent DNA binding. Proc. Natl. Acad. Sci. U. S. A. 100, 2580–2585.
Vidal, M., Cagan, R.L., 2006. *Drosophila* models for cancer research. Curr. Opin. Genet. Dev. 16, 10–16.
Wang, Y.V., Tang, H., Gilmour, D.S., 2005. Identification in vivo of different rate-limiting steps associated with transcriptional activators in the presence and absence of a GAGA element. Mol. Cell. Biol. 25, 3543–3552.
Wang, X., Lee, C., Gilmour, D.S., Gergen, J.P., 2007. Transcription elongation controls cell fate specification in the *Drosophila* embryo. Genes Dev. 21, 1031–1036.
Wolpert, L., 1994. Do we understand development? Science 266, 571–572.
Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H., 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell 97, 41–51.
Yang, J., Weinberg, R.A., 2008. Epithelial–mesenchymal transition: at the crossroads of development and tumor metastasis. Dev. Cell 14, 818–829.
Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., Rando, O.J., 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science 309, 626–630.
Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., Young, R.A., 2007. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. Nat. Genet. 39, 1512–1516.