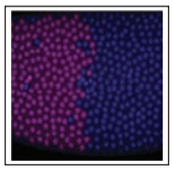
CHAPTER 21



Gene Regulation in Development and Evolution

Animal development depends on the differential expression of a constant genome to produce diverse cell types during embryogenesis. A typical animal genome contains approximately 20,000 genes. This is not only true for comparatively simple creatures such as nematode worms, but also pertains to the "crown and summit" of animal evolution, the human genome.

Differential gene expression can be defined as the synthesis of a protein (or RNA in the case of non-coding genes) in a subset of the cells comprising an embryo. Differential expression most commonly hinges on de novo transcription. Thus, the β-globin gene is selectively expressed in developing red blood cells, but not other tissues, because the gene is transcribed only in blood cells. However, there are examples of post-transcriptional mechanisms of differential gene expression. For example, mRNAs transcribed from the segmentation gene *hunchback* are distributed throughout the early *Drosophila* embryo, but are translated to produce functional proteins only in anterior (head and thorax), but not posterior (abdomen), regions.

How do we know that differential expression of an invariant genome is the key to animal development? A variety of classical and contemporary studies showed that different cell types contain the same genome. The first conclusive evidence came from the cloning of the frog *Xenopus laevis* in the 1960s and 1970s. These studies culminated with the replacement of the egg nucleus with the nucleus of a gut cell of an adult frog. The gut nucleus was able to sustain embryogenesis, the formation of a *Xenopus* tadpole, and its metamorphosis into an adult frog. The resulting frog is said to be a "clone" of the one that donated its gut cell because the two are genetically identical. Subsequent studies in the late 1990s and early 2000s extended cloning to sheep (Dolly), and it is now possible, at least in principle, to clone most animals.

The most spectacular demonstration of "genetic equivalence" among the different tissues of a developing animal is the transformation of virtually any cell type into an induced pluripotent stem (iPS) cell. Most mammalian embryos, including the human fetus, contain a small group of cells, the inner cell mass (ICM), which form all of the tissues and organs of the adult.

OUTLINE

The ICM cells are said to be "pluripotent" because they can produce many different cell types. The formation of ICM cells depends on the activities of three sequence-specific transcription factors—Oct4, Sox2, and Nanog. The forced expression of these three factors in a differentiated cell type, such as a fibroblast cell (connective tissue), is sufficient to transform them into iPS cells, which have the properties of ICM cells (see **Box 21-1, Formation of iPS Cells**). Indeed, iPS cells can be used to replace ICM cells within an

MEDICAL CONNECTIONS

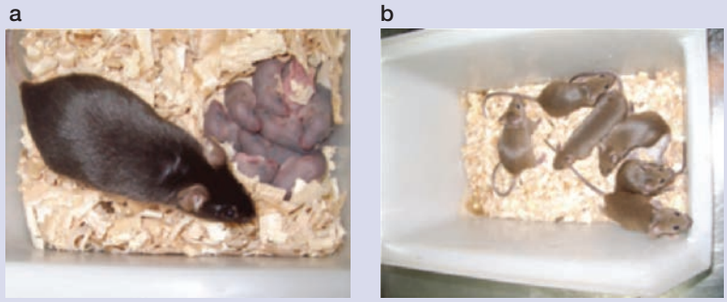
BOX 21-1 Formation of iPS Cells

The ICM cells of mammalian embryos undergo diverse pathways of differentiation and produce all of the tissues and organs of adults.

In the early 2000s, when stem cell fever gripped the biomedical research community, it was thought that the isolation of ICM cells would be the rate-limiting step for the use of stem cells in regenerative medicine. For example, insulin-dependent diabetics lack β-cells, secretory cells in the pancreas that produce insulin in response to increases in blood sugar levels after a meal. There is the hope that it will be one day possible to replace these β-cells with those produced in laboratory culture using stem cells. But the isolation of ICM cells from human fetuses presented a dizzying maze of technical and ethical challenges. This controversy, which became quite heated and political, has dissipated into obscurity because of a remarkable series of experiments conducted by Takahashi and Yamanaka in 2006. As a postdoctoral fellow, Yamanaka had identified a gene that is selectively expressed in ICM cells. He inserted *lacZ* into this gene and used it as a "marker" for identifying mouse fibroblasts that had been converted into stem cells (these converted cells are called induced pluripotent stem [iPS] cells). The marker gene is not normally expressed in fibroblasts but is activated when the cells are transformed into iPS cells. A variety of research groups had identified about 30 different transcription factors (TFs) that showed expression in cultured ICM stem cells. Takahashi and Yamanaka systematically forced the expression of these different TFs in fibroblasts, resulting in the induction of the *lacZ* marker gene. They then coexpressed different combinations of the TFs and found that three of these factors—Oct4, Sox2, and Nanog—were particularly potent in converting or reprogramming fibroblasts into iPS cells. These reprogrammed cells have most or all of the properties of bona fide ICM cells. The iPS cells can be induced to form just about any cell type, such as cardiomyocytes (heart muscle). In a further remarkable experiment, Yamanaka and coworkers showed that it was possible to obtain adult mice from iPS cells injected into embryos. The results revealed in **Box 21-1 Figure 1** show that the characteristics associated with the iPS cells are transmitted in the germline of the resulting offspring (Box 21-1 Fig. 1).

The competence of different adult tissues to be transformed into iPS cells, which, in turn, can be induced to produce any tissue, is a clear demonstration of genetic equivalence. These studies also raise the possibility of "replacement medicine," whereby skin fibroblasts from a sick individual can be used to produce iPS cells, which are subsequently programmed to generate the missing tissues causing illness, for example, the dopaminergic neurons for sufferers of Parkinson's or the β-cells for diabetics.

Gurdon and Yamanaka were awarded the Nobel Prize in Medicine in 2012 for their discovery that differentiated animal cells can be reprogrammed into any tissue.



BOX 21-1 FIGURE 1 The developmental potential of iPS cells. (a) Reprogrammed (iPS) cells, derived from a mouse of black coat color, were injected into the blastocyst (early-stage embryo) of a female mouse of white coat color, producing the black adult mouse (a male) shown here. Next to the adult are its progeny, newborn pups resulting from mating the iPS male with a white female. (b) The newborn pups in panel a have developed into young mice with a brown coat color, which is the typical result seen when a black male is crossed with a white female. (Reproduced from Zhao X.Y. et al. 2009. *Nature* **461:** 86; a and b are Fig. 2f and 2g, respectively.)

embryo and produce adult mice whose tissues are derived solely from the iPS cells.

In this chapter, we consider the different mechanisms for achieving differential gene expression in animal development. In the first half of this chapter, we describe how cells communicate with each other during development to ensure that each expresses a particular set of genes required for their proper development. Simple examples of each of these strategies are then described. In the next part, we describe how these strategies are used in combination with the transcriptional regulatory mechanisms described in **Chapter 19** to control the development of an entire organism—in this case, the fruit fly. In the final part of the chapter, we discuss how changes in gene regulation can cause diversity of animal morphology during evolution. A particularly important class of developmental control genes, the homeotic genes, is described.

THREE STRATEGIES BY WHICH CELLS ARE INSTRUCTED TO EXPRESS SPECIFIC SETS OF GENES DURING DEVELOPMENT

We have already seen how gene expression can be controlled by "signals" received by a cell from its environment. For example, the sugar lactose activates the transcription of the *lac* operon in *Escherichia coli*, whereas viral infection activates the expression of the β-interferon gene in mammals. In this chapter, we focus on the strategies that are used to instruct genetically identical cells to express distinct sets of genes and thereby differentiate into diverse cell types. The three major strategies are **mRNA localization**, **cell-to-cell contact**, and **signaling through the diffusion of a secreted signaling molecule** (**Fig. 21-1**). Each of these strategies is introduced briefly in the following sections.

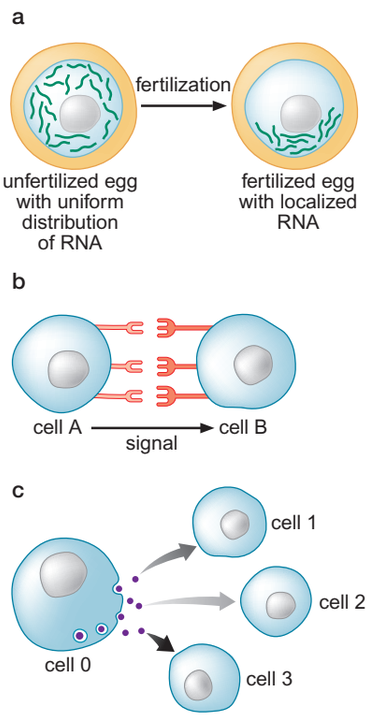


FIGURE 21-1 The three strategies for initiating differential gene activity during development. (a) In some animals, certain "maternal" RNAs present in the egg become localized either before or after fertilization. In this example, a specific mRNA (green squiggles) becomes localized to vegetal (bottom) regions after fertilization. (b) Cell A must physically interact with cell B to stimulate the receptor present on the surface of cell B. This is because the "ligand" produced by cell A is tethered to the plasma membrane. (c) In this example of long-range cell signaling, cell 0 secretes a signaling molecule that diffuses through the extracellular matrix. Different cells (1, 2, 3) receive the signal and ultimately undergo changes in gene activity.

Some mRNAs Become Localized within Eggs and Embryos Because of an Intrinsic Polarity in the Cytoskeleton

One strategy to establish differences between two genetically identical cells is to distribute a critical regulatory molecule asymmetrically during cell division, thereby ensuring that the daughter cells inherit different amounts of that regulator and thus follow different pathways of development. Typically, the asymmetrically distributed molecule is an mRNA. These mRNAs can encode RNA-binding proteins or cell-signaling molecules, but most often they encode transcriptional activators or repressors. Despite this diversity in the function of their protein products, a common mechanism exists for localizing mRNAs. Typically, they are transported along elements of the cytoskeleton, actin filaments, or microtubules. The asymmetry in this process is provided by the intrinsic asymmetry of these elements.

Actin filaments and microtubules undergo directed growth at the + ends (**Fig. 21-2**). An mRNA molecule can be transported from one end of a cell to the other end by means of an "adaptor" protein, which binds to a specific sequence within the non-coding **3′ untranslated trailer region** (**3′ UTR**) of an mRNA. Adaptor proteins contain two domains. One recognizes the 3′ UTR of the mRNA, whereas the other associates with a specific component of the cytoskeleton, such as myosin. Depending on the specific adaptor used, the mRNA–adaptor complex either "crawls" along an actin

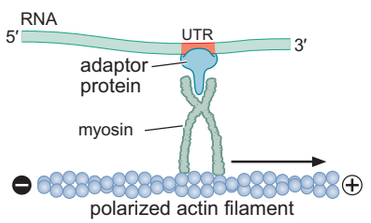


FIGURE 21-2 An adaptor protein binds to specific sequences within the 3′ UTR of the mRNA. The adaptor also binds to myosin, which "crawls" along the actin filament in a directed fashion, from the "–" end to the growing "+" end of the filament.

filament or directly moves with the + end of a growing microtubule. We will see how this basic process is used to localize mRNA determinants within the egg or to restrict a determinant to a single daughter cell after mitosis.

Cell-to-Cell Contact and Secreted Cell-Signaling Molecules Both Elicit Changes in Gene Expression in Neighboring Cells

A cell can influence which genes are expressed in neighboring cells by producing extracellular signaling proteins. These proteins are synthesized in the first cell and then either deposited in the plasma membrane of that cell or secreted into the extracellular matrix. Because these two approaches have features in common, we consider them together here. We then see how secreted signals can be used in other ways.

A given signal (of either sort) is generally recognized by a specific receptor on the surface of recipient cells. When that receptor binds to the signaling molecule, it triggers changes in gene expression in the recipient cell. This communication from the cell-surface receptor to the nucleus often involves **signal transduction pathways** of the sort we considered in **Chapter 19**. Here, we summarize a few basic features of these pathways.

Sometimes, ligand–receptor interactions induce an enzymatic cascade that ultimately modifies regulatory proteins already present in the nucleus (**Fig. 21-3**a). In other cases, activated receptors cause the release of DNA-binding proteins from the cell surface or cytoplasm into the nucleus (Fig. 21-3b). These regulatory proteins bind to specific DNA-recognition sequences and either activate or repress gene expression. Ligand binding can also cause proteolytic cleavage of the receptor. Upon cleavage, the intracytoplasmic domain of the receptor is released from the cell surface and enters the nucleus, where it associates with DNA-binding proteins and influences how those proteins regulate transcription of the associated genes (Fig. 21-3c). For example, the transported protein might convert what was a transcriptional repressor into an activator. In this case, target genes that were formerly repressed before signaling are now induced. We consider examples of each of these variations in cell signaling in this chapter.

Signaling molecules that remain on the surface control gene expression only in those cells that are in direct, physical contact with the signaling cell. We refer to this process as **cell-to-cell contact**. In contrast, signaling molecules that are secreted into the extracellular matrix can work over greater distances. Some travel over a distance of just one or two cell diameters, whereas others can act over a range of 50 cells or more. Long-range signaling molecules are sometimes responsible for positional information, which is discussed in the next section.

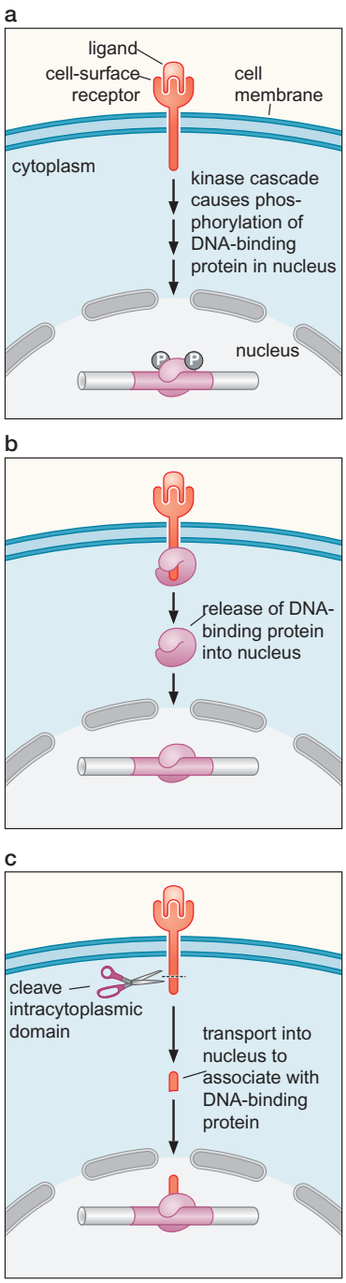


FIGURE 21-3 **Different mechanisms of signal transduction.** A ligand (or "signaling molecule") binds to a cell-surface receptor. (a) The activated receptor induces latent cellular kinases that ultimately cause the phosphorylation of DNA-binding proteins within the nucleus. This phosphorylation causes the regulatory protein to activate (or repress) the transcription of specific genes. (b) The activated receptor releases a dormant DNA-binding protein from the cytoplasm so that it can now enter the nucleus. Once in the nucleus, the regulatory protein activates (or represses) the transcription of specific genes. (c) The activated receptor is cleaved by cellular proteases that cause a carboxy-terminal portion of the receptor to enter the nucleus and interact with specific DNA-binding proteins. The resulting protein complex activates the transcription of specific genes.

Gradients of Secreted Signaling Molecules Can Instruct Cells to Follow Different Pathways of Development Based on Their Location

A recurring theme in development is the importance of a cell's position within a developing embryo or organ in determining what it will become. Cells located at the front of a fruit fly embryo (i.e., in **anterior** regions) will form portions of the adult head such as the antenna or brain but will not develop into **posterior** structures such as the abdomen or genitalia. Cells located on the top, or **dorsal**, surface of a frog embryo can develop into portions of the backbone in the tadpole or adult but do not form **ventral**, or "belly," tissues such as the gut. These examples illustrate the fact that the fate of a cell—what it will become in the adult—is constrained by its location in the developing embryo. The influence of location on development is called **positional information**.

The most common way of establishing positional information involves a simple extension of one of the strategies we have already encountered in **Chapter 19**—the use of secreted signaling molecules (**Fig. 21-4**). A small group of cells synthesizes and secretes a signaling molecule that becomes distributed in an **extracellular gradient** (Fig. 21-4a). Cells located near the "source" receive high concentrations of the secreted protein and develop into a particular cell type. Those cells located at progressively farther distances follow different pathways of development as a result of receiving lower

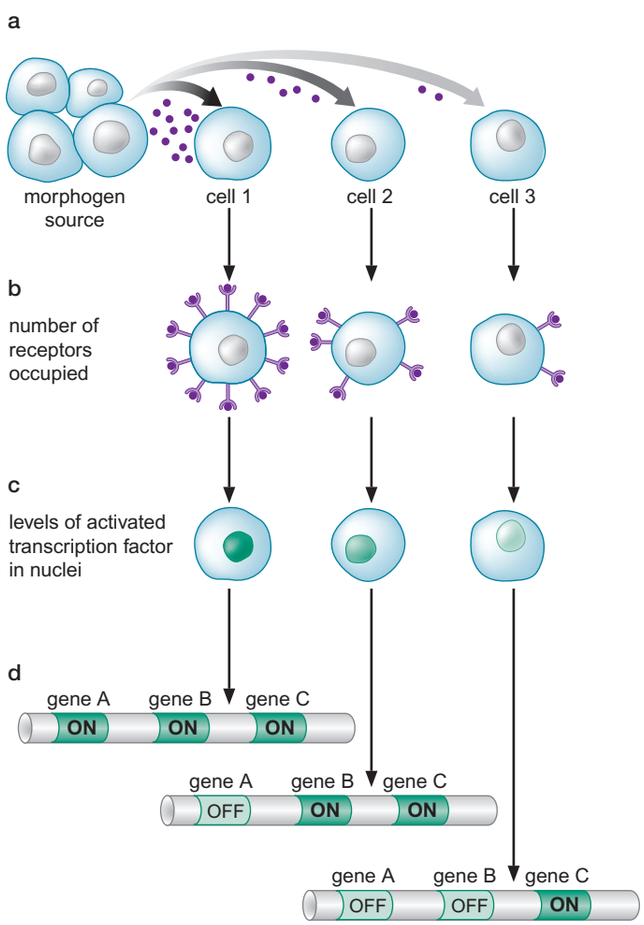


FIGURE 21-4 A cluster of cells produces a signaling molecule, or morphogen, that diffuses through the extracellular matrix. (a) Cells 1, 2, and 3 receive progressively lower amounts of the signaling molecule because they are located progressively farther from the source. (b) Cells 1, 2, and 3 contain progressively lower numbers of activated surface receptors. (c) The three cells contain different levels of one or more regulatory proteins. In the simplest scenario, there is a linear correlation between the number of activated cell-surface receptors and the numbers of activated regulatory proteins in the nucleus. (d) The different levels of the regulatory factor lead to the expression of different sets of genes. Cell 1 expresses genes A, B, and C because it contains the highest levels of the regulatory factor. Cell 2 expresses genes B and C, but not A, because it contains intermediate levels of the regulatory factor. These levels are not sufficient to activate gene A. Finally, cell 3 contains the lowest levels of the regulatory factor and expresses only gene C because expression of genes A and B requires higher levels.

concentrations of the signaling molecule. Signaling molecules that control position information are sometimes called **morphogens**.

Cells located near the source of the morphogen receive high concentrations of the signaling molecule and therefore experience peak activation of the specific cell-surface receptors that bind it. In contrast, cells located far from the source receive low levels of the signal, and, consequently, only a small fraction of their cell-surface receptors are activated. Consider a row of three cells adjacent to a source of a secreted morphogen. Something like 1000 receptors are activated in the first cell, whereas only 500 receptors are activated in the next cell, and just 200 in the next (Fig. 21-4b). These different levels of receptor occupancy are directly responsible for differential gene expression in the responding cells.

As we have seen, binding of signaling molecules to cell-surface receptors leads (in one way or another) to an increase in the concentration of specific transcriptional regulators, in an active form, in the nucleus of the cell. Each receptor controls a specific transcriptional regulator (or regulators), and this controls expression of particular genes. The number of cell-surface receptors that are activated by the binding of a morphogen determines how many molecules of the particular regulatory protein appear in the nucleus. The cell closest to the morphogen source—containing 1000 activated receptors—will possess high concentrations of the transcriptional activator in its nucleus (Fig. 21-4c). In contrast, the cells located farther from the source contain intermediate and low levels of the activator, respectively. Thus, there is a correlation between the number of activated receptors on the cell surface and the amount of transcriptional regulator present in the nucleus. How are these different levels of the same transcriptional regulator able to trigger different patterns of gene expression in these different cells?

In **Chapter 18**, we learned that a small change in the level of the λ repressor determines whether an infected bacterial cell is lysed or lysogenized. Similarly, small changes in the amount of morphogen, and hence small differences in the levels of a transcriptional regulator within the nucleus, determine cell identity. Cells that contain high concentrations of a given transcriptional regulator express a variety of target genes that are inactive in cells containing intermediate or low levels of the regulator (Fig. 21-4d). The differential regulation of gene expression by different concentrations of a regulatory protein is one of the most important and pervasive mechanisms encountered in developmental biology. We consider several examples in the course of this chapter.

EXAMPLES OF THE THREE STRATEGIES FOR ESTABLISHING DIFFERENTIAL GENE EXPRESSION

The Localized Ash1 Repressor Controls Mating Type in Yeast by Silencing the *HO* Gene

Before describing mRNA localization in animal embryos, we first consider a case from a relatively simple single-cell eukaryote, the yeast *Saccharomyces cerevisiae*. This yeast can grow as haploid cells that divide by budding (**Fig. 21-5**). Replicated chromosomes are distributed between two asymmetric cells—the larger progenitor cell, or mother cell, and a smaller bud, or daughter cell (Fig. 21-5a). These cells can exist as either of two mating types, called **a** and α, as discussed in **Chapters 11** and **19**.

A mother cell and its daughter cell can show different mating types. This difference arises by a process called **mating-type switching**. After budding

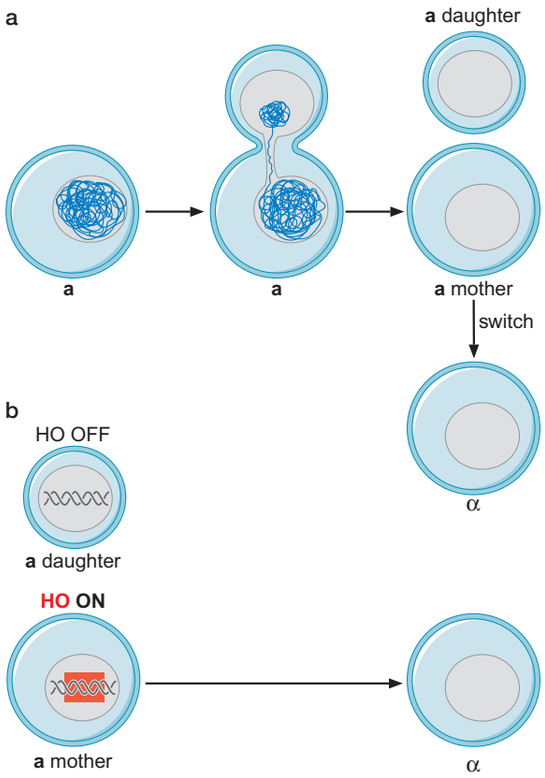


FIGURE 21-5 A haploid yeast cell of mating type a undergoes budding to produce a mother cell and a smaller daughter cell. (a) Initially, both cells are mating type **a**, but sometimes the mother cell can undergo switching to the α type. (b) Because of the localized Ash1 transcriptional repressor, the daughter cell is unable to express the *HO* gene and thus cannot undergo switching. In contrast, the mother cell can switch because it lacks Ash1 and is able to express *HO*.

to produce a daughter, a mother cell can "switch" mating type with, for example, an **a** cell giving rise to an **a** daughter, but subsequently switching to the α mating type (Fig. 21-5b).

Switching is controlled by the product of the *HO* gene. We saw in **Chapter 11** that the HO protein is a sequence-specific endonuclease. HO triggers gene conversion within the mating-type locus by creating a double-strand break at one of the two silent mating-type cassettes. We also saw in **Chapter 19** how HO is activated in the mother cell. It is kept silent in the daughter cell because of the selective expression of a repressor called Ash1 (**Fig. 21-6**), and this is why the daughter cell does not switch mating type. The *ash1* gene is transcribed in the mother cell before budding, but the encoded RNA becomes localized within the daughter cell through the following process. During budding, the *ash1* mRNA attaches to the growing ends of microtubules. Several proteins function as "adaptors" that bind the 3′ UTR of the *ash1* mRNA and also to the microtubules. The microtubules extend from the nucleus of the mother cell to the site of budding, and in this way, the *ash1* mRNA is transported to the daughter cell. Once localized within the daughter cell, the *ash1* mRNA is translated into a repressor protein that binds to, and inhibits the transcription of, the *HO* gene. This silencing of *HO* expression in the daughter cell prevents that cell from undergoing mating-type switching.

In the second half of this chapter, we will see the localization of mRNAs used in the development of the *Drosophila* embryo. Once again, this localization is mediated by adaptor proteins that bind to the mRNAs, specifically, to sequences found in their 3′ UTRs (see **Box 21-2, Review of Cytoskeleton: Asymmetry and Growth**).

A second general principle that emerges from studies on yeast mating-type switching is seen again when we consider *Drosophila* development:

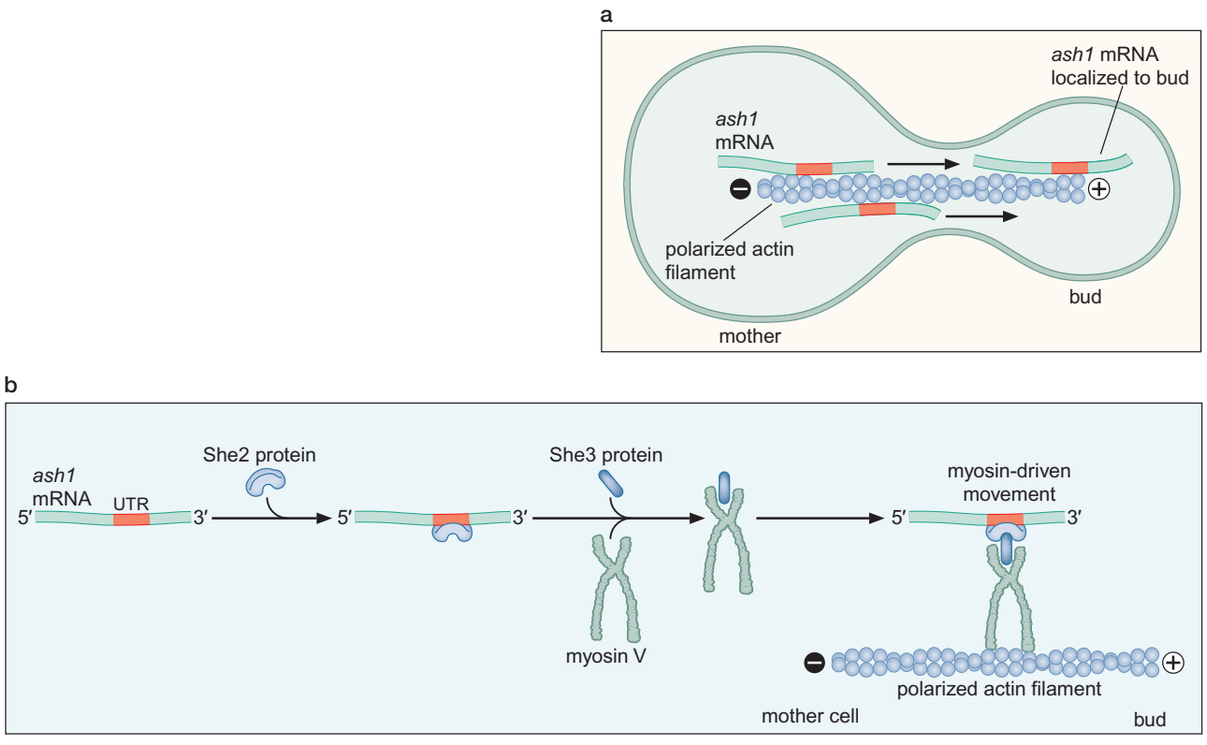


FIGURE 21-6 Localization of *ash1* mRNA during budding. (a) The *ash1* gene is transcribed in the mother cell during budding. The encoded mRNA moves from the mother cell into the bud by sliding along polarized actin filaments. Movement is directed and begins at the "−" ends of the filament and extends with the growing "+" ends. (b) The *ash1* mRNA transport depends on the binding of the She2 and She3 adaptor proteins to specific sequences contained within the 3′ UTR. These adaptor proteins bind myosin, which "crawls" along the actin filament and brings the *ash1* mRNA along for the ride. (Adapted, with permission, from Alberts B. et al. 2002. *Molecular biology of the cell*, 4th ed., p. 971, Fig. 16-84a. © Garland Science/Taylor & Francis LLC.)

the interplay between broadly distributed activators and localized repressors to establish precise patterns of gene expression within individual cells. In yeast, the SWI5 protein is responsible for activating expression of the *HO* gene (see **Chapter 19**). This activator is present in both the mother cell and the daughter cell during budding, but its ability to turn on *HO* is restricted to the mother cell because of the presence of the Ash1 repressor in the daughter cell. In other words, Ash1 keeps the *HO* gene off in the daughter cell despite the presence of SWI5.

A Localized mRNA Initiates Muscle Differentiation in the Sea Squirt Embryo

Localized mRNAs can establish differential gene expression among the genetically identical cells of a developing embryo. Just as the fate of the daughter cell is constrained by its inheritance of the *ash1* mRNA in yeast, the cells in a developing embryo can be instructed to follow specific pathways of development through the inheritance of localized mRNAs.

As an example, we consider muscle differentiation in sea squirts. Macho-1 is a major determinant for programming cells to form tail muscles in early sea squirt embryos.

Macho-1 mRNA is initially distributed throughout the cytoplasm of unfertilized eggs but becomes restricted to the vegetal (bottom) cytoplasm shortly after fertilization (**Fig. 21-7**). It is ultimately inherited by just two of the cells in eight-cell embryos, and as a result these two cells go on to form the tail muscles.

The Macho-1 mRNA encodes a zinc finger DNA-binding protein that is believed to activate the transcription of muscle-specific genes, such as tropomysin. Thus, these genes are expressed only in muscles because Macho-1 is made only in those cells. In the second part of this chapter, we see how regulatory proteins synthesized from localized mRNAs in the *Drosophila* embryo activate and repress gene expression and control the formation of different cell types.

ADVANCED CONCEPTS

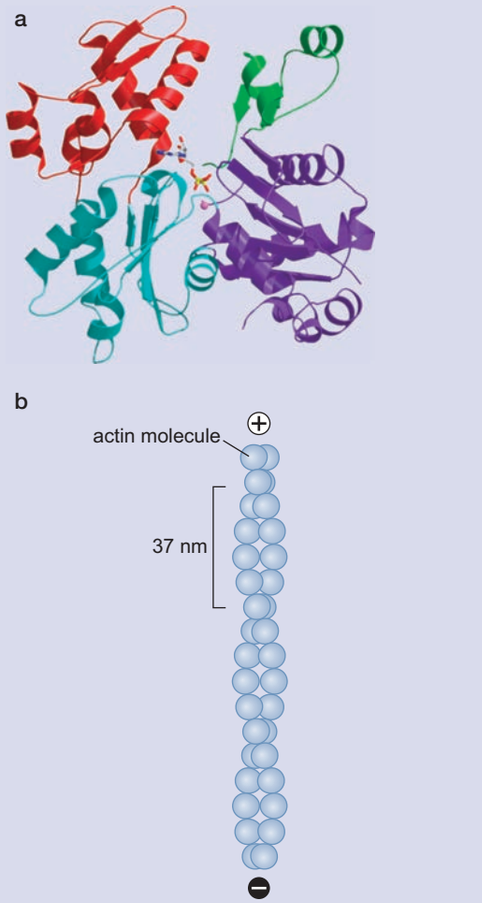
BOX 21-2 Review of Cytoskeleton: Asymmetry and Growth

The cytoskeleton is composed of three types of filaments: intermediate filaments, actin filaments, and microtubules. Actin filaments and microtubules are used to localize specific mRNAs in a variety of different cell types, including budding yeast and *Drosophila* oocytes. Actin filaments are composed of polymers of actin. The actin polymers are organized as two parallel helices that form a complete twist every 37 nm. Each actin monomer is located in the same orientation within the polymer, and as a result, actin filaments contain a clear polarity. The plus (+) end grows more rapidly than the minus (–) end, and consequently, mRNAs slated for localization move along with the growing + end (**Box 21-2 Fig. 1**).

Microtubules are composed of polymers of a protein called **tubulin**, which is a heterodimer composed of related α and β chains. Tubulin heterodimers form extended, asymmetric protofilaments. Each tubulin heterodimer is located in the same orientation within the protofilament. Thirteen different protofilaments associate to form a cylindrical microtubule, and all of the protofilaments are aligned in parallel. Thus, as seen for actin filaments, there is an intrinsic polarity in microtubules, with a rapidly growing "+" end and more stable "–" end (**Box 21-2 Fig. 2**).

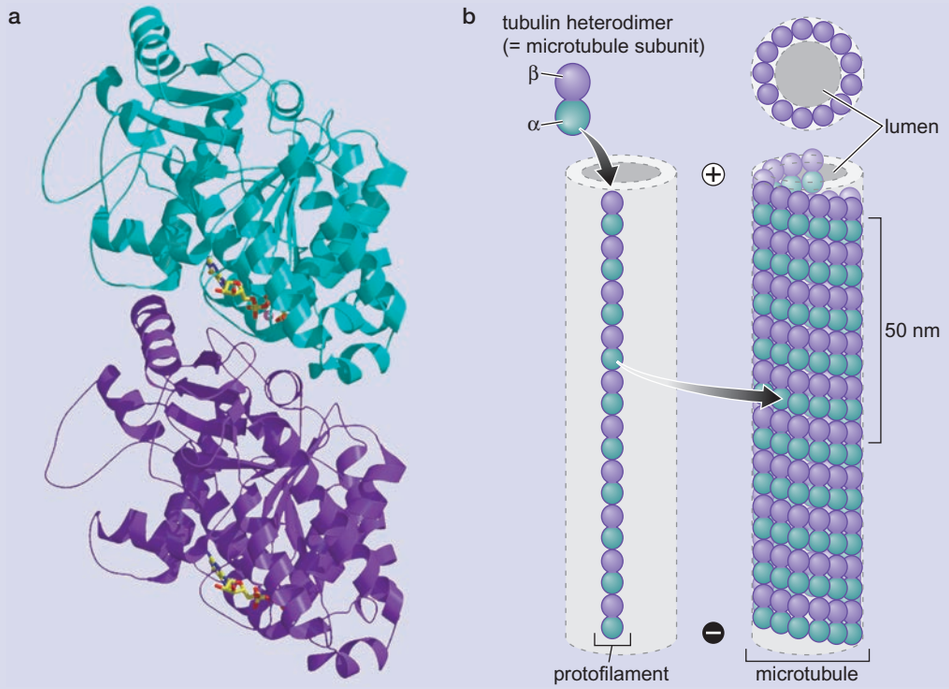
Both actin and tubulin function as enzymes. Actin catalyzes the hydrolysis of ATP to ADP, whereas tubulin hydrolyzes GTP to GDP. These enzymatic activities are responsible for the dynamic growth, or "treadmilling," seen for actin filaments and microtubules. Typically, it is the actin or tubulin subunits at the "–" end of the filament that mediate the hydrolysis of ATP or GTP, and as a result, these subunits are somewhat unstable and lost from the "–" end. In contrast, newly added subunits at the "+" end have not hydrolyzed ATP or GTP, and this causes them to be more stable components of the filament.

Directed growth of actin filaments or microtubules at the "+" ends depends on a variety of proteins that associate with the cytoskeleton. One such protein is called profilin, which interacts with actin monomers and augments their incorporation into the "+" ends of growing actin filaments. Other proteins have been shown to enhance the growth of tubulin protofilaments at the "+" ends of microtubules.



BOX 21-2 FIGURE 1 Structures of the actin monomer and filament. Crystal structure of the actin monomer. (a) The four domains of the monomer are shown, in different colors, with ATP (in red and yellow) in the center. The "–" end of the monomer is at the top, and the "+" end is at the bottom. (Otterbein L.R. et al. 2001. *Science* **293:** 708–711. Image prepared with MolScript, BobScript, and Raster 3D.) (b) The monomers are assembled, as a single helix, into a filament.

BOX 21-2 *(Continued)*



BOX 21-2 FIGURE 2 Structures of the tubulin monomer and filament. (a) The crystal structure of the tubulin monomer shows the α subunit (turquoise) and the β subunit (purple). The GTP molecules in each subunit are red and yellow. (From Lowe J. et al. 2001. *J. Mol. Biol.* **313:** 1045–1057. Image prepared with MolScript, BobScript, and Raster 3D.) (b) The protofilament of tubulin consists of adjacent monomers assembled in the same orientation.

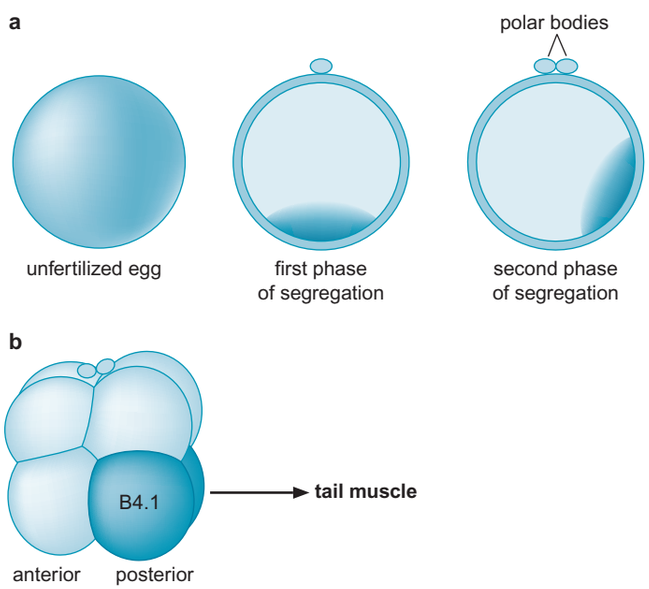


FIGURE 21-7 The Macho-1 mRNA becomes localized in the fertilized egg of a sea squirt. (a) The mRNA is initially distributed throughout the cytoplasm of unfertilized eggs. At fertilization, the egg is induced to undergo a highly asymmetric division to produce a small polar body (top). At this time, the Macho-1 mRNA becomes localized to bottom (vegetal) regions. Shortly thereafter, and well before the first division of the one-cell embryo, the Macho-1 mRNA undergoes a second wave of localization. This occurs during the second highly asymmetric meiotic division of the egg. (b) The Macho-1 mRNA becomes localized to a specific quadrant of the one-cell embryo that corresponds to the future B4.1 blastomeres. These are the cells that generate the tail muscles. (a, Adapted, with permission, from Nishida H. and Sawada K. 2001. *Nature* **409:** 725, Fig. 1c–e. © Macmillan.)

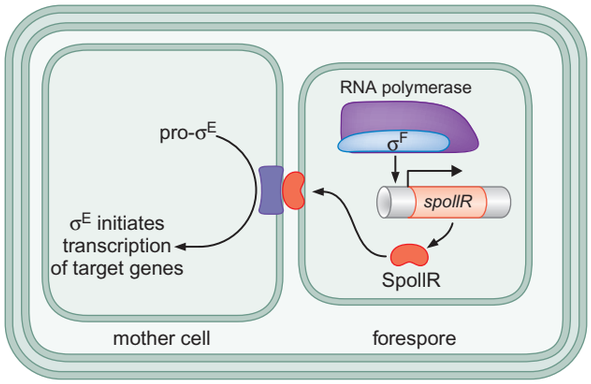


FIGURE 21-8 Asymmetric gene activity in the mother cell and forespore of *B. subtilis* depends on the activation of different classes of σ factors. The *spoIIR* gene is activated by σ^F in the forespore. The encoded SpoIIR protein becomes associated with the septum separating the mother cell (on the *left*) and forespore (on the *right*). It triggers the proteolytic processing of an inactive form of σ^E (pro-σ^E) in the mother cell. The activated σ^E protein leads to the recruitment of RNA polymerase and the activation of specific genes in the mother cell. (Redrawn, with permission, from Stragier P. and Losick R. 1996. *Annu. Rev. Genet.* **30:** 297–341, Fig. 3a. © *Annual Reviews*.)

Cell-to-Cell Contact Elicits Differential Gene Expression in the Sporulating Bacterium, *Bacillus subtilis*

The second major strategy for establishing differential gene expression is cell-to-cell contact. Again, we begin our discussion with a relatively simple case, this one from the bacterium *Bacillus subtilis*. Under adverse conditions, *B. subtilis* can form spores. The first step in this process is the formation of a septum at an asymmetric location within the sporangium, the progenitor of the spore. The septum produces two cells of differing sizes that remain attached through abutting membranes. The smaller cell is called the **forespore**; it ultimately forms the spore. The larger cell is called the mother cell; it aids the development of the spore (**Fig. 21-8**). The forespore influences the expression of genes in the neighboring mother cell, as described below.

The forespore contains an active form of a specific σ factor, σ^F, that is inactive in the mother cell. In **Chapter 18**, we saw how σ factors associate with RNA polymerase and select specific target promoters for expression. σ^F activates the *spoIIR* gene, which encodes a secreted signaling protein. SpoIIR is secreted into the space between the abutting membranes of the mother cell and the forespore, where it triggers the proteolytic processing of pro-σ^E in the mother cell. Pro-σ^E is an inactive precursor of the σ^E factor. The pro-σ^E protein contains an amino-terminal inhibitory domain that blocks σ^E activity and tethers the protein to the membrane of the mother cell (Fig. 21-8). SpoIIR induces the proteolytic cleavage of the amino-terminal peptide and the release of the mature and active form of σ^E from the membrane. σ^E activates a set of genes in the mother cell that is distinct from those expressed in the forespore. In this example, SpoIIR functions as a signaling molecule that acts at the interface between the forespore and the mother cell and elicits differential gene expression in the abutting mother cell through the processing of σ^E. Induction requires cell-to-cell contact because the forespore produces small quantities of SpoIIR that can interact with the abutting mother cell but are insufficient to elicit the processing of σ^E in the other cells of the population.

A Skin–Nerve Regulatory Switch Is Controlled by Notch Signaling in the Insect Central Nervous System

We now turn to an example of cell-to-cell contact in an animal embryo that is surprisingly similar to the one just described in *B. subtilis*. In that earlier example, SpoIIR causes the proteolytic activation of σ^E, which, in its active state, directs RNA polymerase to the promoter sequences of specific genes. In the following example, a cell-surface receptor is cleaved, and the intracytoplasmic

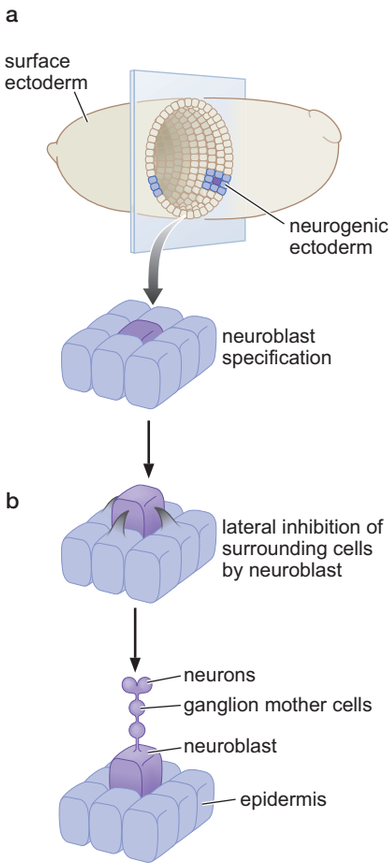


FIGURE 21-9 The neurogenic ectoderm forms two major cell types: neurons and skin cells (or epidermis). (a) Cells in the early neurogenic ectoderm can form either type of cell. However, once one of the cells begins to form a neuron or "neuroblast" (dark cell in the center of the grid of cells), it inhibits all of the neighboring cells that it directly touches. (b) This inhibition causes most of the cells to remain on the surface of the embryo and form skin cells. In contrast, the developing neuron moves into the embryo cavity and forms neurons.

domain moves to the nucleus, where it binds a sequence-specific DNA-binding protein that activates the transcription of selected genes.

For this example, we must first briefly describe the development of the ventral nerve cord in insect embryos (**Fig. 21-9**). This nerve cord functions in a manner that is roughly comparable to the spinal cord of humans. It arises from a sheet of cells called the **neurogenic ectoderm**. This tissue is subdivided into two cell populations: one group remains on the surface of the embryo and forms ventral skin (or epidermis), whereas the other population moves inside the embryo to form the neurons of the ventral nerve cord (Fig. 21-9a). This decision whether to become skin or neuron is reinforced by signaling between the two populations.

The developing neurons contain a signaling molecule on their surface called **Delta**, which binds to a receptor on the skin cells called **Notch** (Fig. 21-9b). The activation of the Notch receptor on skin cells by Delta renders them incapable of developing into neurons, as follows. Activation causes the intracytoplasmic domain of Notch ($Notch^{IC}$) to be released from the cell membrane and enter nuclei, where it associates with a DNA-binding protein called Su(H). The resulting Su(H)–$Notch^{IC}$ complex activates genes that encode transcriptional repressors that block the development of neurons.

Notch signaling does not cause a simple induction of the Su(H) activator protein but instead triggers an on/off regulatory switch. In the absence of signaling, Su(H) is associated with several corepressor proteins, including Hairless, CtBP, and Groucho (**Fig. 21-10**). Su(H) complexed with any of these proteins actively represses Notch target genes. When $Notch^{IC}$ enters the nucleus, it displaces the repressor proteins in complex with Su(H), turning that protein into an activator instead. Thus, Su(H) now activates the very same genes that it formerly repressed.

Delta–Notch signaling depends on cell-to-cell contact. The cells that present the Delta ligand (neuronal precursors) must be in direct physical contact with the cells that contain the Notch receptor (epidermis) in order to activate Notch signaling and inhibit neuronal differentiation. In the next section, we see an example of a secreted signaling molecule that influences gene expression in cells located far from those that send the signal.

A Gradient of the Sonic Hedgehog Morphogen Controls the Formation of Different Neurons in the Vertebrate Neural Tube

We now turn to an example of a long-range signaling molecule, a **morphogen**, that imposes positional information on a developing organ. For this example, we continue our discussion of neuronal differentiation, but this time, we consider the neural tube of vertebrates. In all vertebrate embryos, there is a stage when cells located along the future back—the dorsal ectoderm—move in a coordinated fashion toward internal regions of the embryo and form the neural tube, the forerunner of the adult spinal cord.

Cells located in the ventralmost region of the neural tube form a specialized structure called the **floorplate**. The floorplate is the site of expression of a secreted cell-signaling molecule called Sonic hedgehog (Shh), which is thought to function as a gradient morphogen.

Shh is secreted from the floorplate and forms an extracellular gradient in the ventral half of the neural tube. Neurons develop within the neural tube into different cell types based on the amount of Shh protein they receive. This is determined by their location relative to the floorplate; cells located near the floorplate receive the highest concentrations of Shh, and those located farther away receive lower levels. The extracellular Shh gradient leads to the specification of three neuronal cell types: V3, MN, and V2. These cells are located

progressively farther from the floorplate and differentially express three regulatory genes: Nkx2.2, Olig2, and Pax6, respectively (see **Fig. 21-11a**).

Initially, Pax6 is expressed throughout the presumptive neural tube (time t_0, Fig. 21-11b). Cells located near the floorplate—those receiving the highest concentrations of Shh—acquire the highest activity of Gli, the transcriptional effector of Shh signaling. At early stages, time t_1, the initial concentrations of the Gli activator are sufficient to induce Olig2 expression (Fig. 21-11b). At subsequent stages, sustained Shh induction raises the levels of Gli in the ventral neural tube, leading to the activation of Nkx2.2 (time t_2, Fig. 21-11b). Cross-repressive interactions maintain sequential expression of Nkx2.2, Olig2, and Pax6, leading to the specification of the V3, MN, and V2 neurons.

According to a simple "gradient affinity" model, the regulatory DNAs of the Olig2 and Nkx2.2 genes might be expected to contain Gli-binding sites with differing affinities. For example, Olig2 might be activated before Nkx2.2 because it contains high-affinity Gli-binding sites that are occupied by low levels of the Gli activator. In contrast, Nkx2.2 might be regulated by low-affinity Gli-binding sites, requiring higher, sustained levels of Shh and the Gli activator.

Recent studies suggest an alternative view, namely: Differential expression of Olig2 and Nkx2.2 is controlled by a network of gene interactions underlying the patterning of the neural tube (see Fig. 21-11b). Once Olig2 is activated in ventral regions, it represses Pax6, thereby creating a "window" for the induction of Nkx2.2. Pax6 is a potent repressor of Nkx2.2 but not Olig2. Differential repression by Pax6 might be a critical mechanism for sequential expression of Olig2 and Nkx2.2. Perhaps Nkx2.2 regulatory DNAs contain Pax6 repressor binding sites, whereas Olig2 regulatory sequences lack such sites.

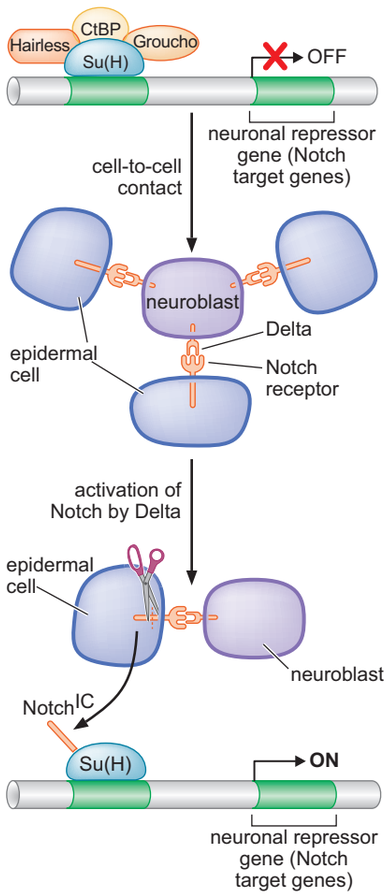


FIGURE 21-10 Notch–Su(H) regulatory switch. The developing neuron (neuroblast) does not express neuronal repressor genes (*top*). These genes are kept off by a DNA-binding protein called Su(H) and associated corepressor proteins (Hairless, CtBP, Groucho). The neuroblast expresses a signaling molecule, called Delta, that is tethered to the cell surface. Delta binds to the Notch receptor in neighboring cells that are in direct physical contact with the neuron. Delta–Notch interactions cause the Notch receptor to be activated in the neighboring cells, which differentiate into epidermis. The activated Notch receptor is cleaved by cellular proteases (scissors), and the intracytoplasmic region of the receptor is released into the nucleus. This piece of the Notch protein causes the Su(H) regulatory protein to function as an activator rather than a repressor. As a result, the neuronal repressor genes are activated in the epidermal cells so that they cannot develop into neurons.

THE MOLECULAR BIOLOGY OF *DROSOPHILA* EMBRYOGENESIS

WEB ANIMATION

In this section, we focus on the early embryonic development of the fruit fly, *Drosophila melanogaster* (see **Interactive Animation 21-1**). The molecular details of how development is regulated are better understood in this system than in any other animal embryo. The various mechanisms of cell communication discussed in the first half of this chapter and those of gene regulation discussed in the previous chapters are brought together in this example.

Localized mRNAs and cell-signaling pathways are both used to establish positional information that results in gradients of regulatory proteins that pattern the anteroposterior (head–tail) and dorsoventral (back–belly) body axes. These regulatory proteins—activators and repressors—control the expression of genes whose products define different regions of the embryo. A recurring theme is the use of complex regulatory DNAs—particularly complex enhancers—for the combinatorial control of sharp on/off patterns of gene expression.

An Overview of *Drosophila* Embryogenesis

Life begins for the fruit fly as it does for humans: Adult males inseminate females. A single sperm cell enters a mature egg, and the haploid sperm and egg nuclei fuse to form a diploid, "zygotic" nucleus. This nucleus undergoes a series of nearly synchronous divisions within the central regions of the egg. Because there are no plasma membranes separating the nuclei, the embryo now becomes what is called a **syncytium**—that is, a single cell with multiple nuclei. With the next series of divisions, the nuclei

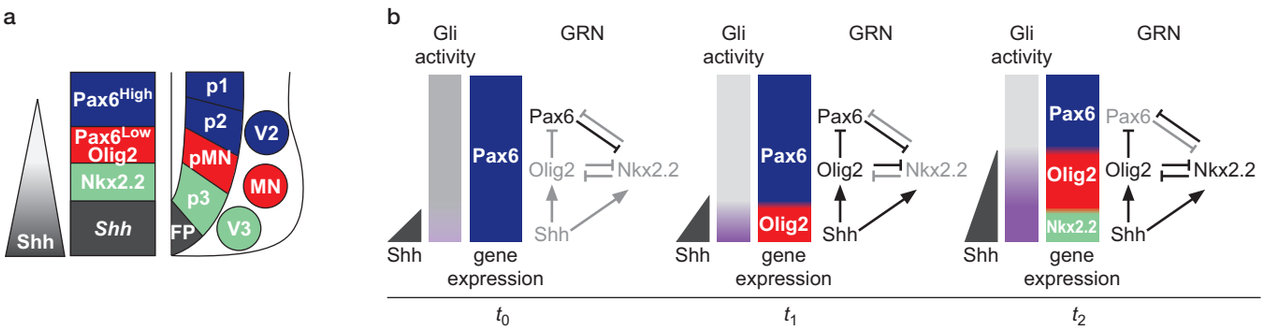


FIGURE 21-11 The extracellular Shh gradient leads to the specification of three neuronal cell types. (a) Shh forms a gradient in the neural tube. (b) Model for Shh signal-mediated patterning as development progresses. (Adapted from Balaskas N. et al. 2012. *Cell* **148:** 273–284; part a is Fig. 1A, p. 274; part b is Fig. 7A, p. 281.)

begin to migrate toward the cortex, or periphery, of the egg. Once located in the cortex, the nuclei undergo another three divisions leading to the formation of a monolayer of approximately 6000 nuclei surrounding the central yolk. During a 1-h period, from 2 to 3 h after fertilization, cell membranes form between adjacent nuclei.

The rapid nuclear divisions that occur during the first 2 h of *Drosophila* embryogenesis preclude precocious expression of critical patterning genes. Consider the short gastrulation (*sog*) gene as an example. The *sog* gene encodes an inhibitor of BMP signaling that is important for the patterning of the dorsal ectoderm during 2.5–3 h after fertilization. The *sog* transcription unit is 20 kb in length. RNA polymerase II (Pol II) has a remarkably slow rate of elongation—just 20 bp per second. As a result, it takes nearly 20 min for "tip-to-toe" transcription of *sog* and the synthesis of full-length, mature mRNAs. The first 11 rounds of nuclear divisions occur at a frequency of just 6–8 min, and consequently, there is no time for Pol II to complete transcription of the *sog* gene during the brief interphase periods of these division cycles. During mitosis, Pol II is released from the chromatin template and must reinitiate transcription at the onset of the subsequent division cycle. As a result, no meaningful *sog* mRNAs can be synthesized during the first 2 h of embryogenesis. In this example, the size of the *sog* transcription unit helps ensure that *sog* products are not synthesized until they are needed, at 2.5 h after fertilization. There are additional examples of gene size and large introns in the timing of gene expression during *Drosophila* embryogenesis. For example, the homeotic gene *Ubx* is >80 kb in length, and its expression is delayed for more than an hour relative to coexpressed segmentation genes containing small introns.

Before the formation of cell membranes, the nuclei are **totipotent** or uncommitted: They have not yet taken on an identity and can still give rise to any cell type. Just after cellularization, however, nuclei have become irreversibly "**determined**" to differentiate into specific tissues in the adult fly. This process is described in **Box 21-3, Overview of *Drosophila* Development**. The molecular mechanisms responsible for this dramatic process of determination are described in the following sections of this chapter.

A Regulatory Gradient Controls Dorsoventral Patterning of the *Drosophila* Embryo

The dorsoventral patterning of the early *Drosophila* embryo is controlled by a regulatory protein called Dorsal, which is initially distributed throughout

ADVANCED CONCEPTS

BOX 21-3 Overview of *Drosophila* Development

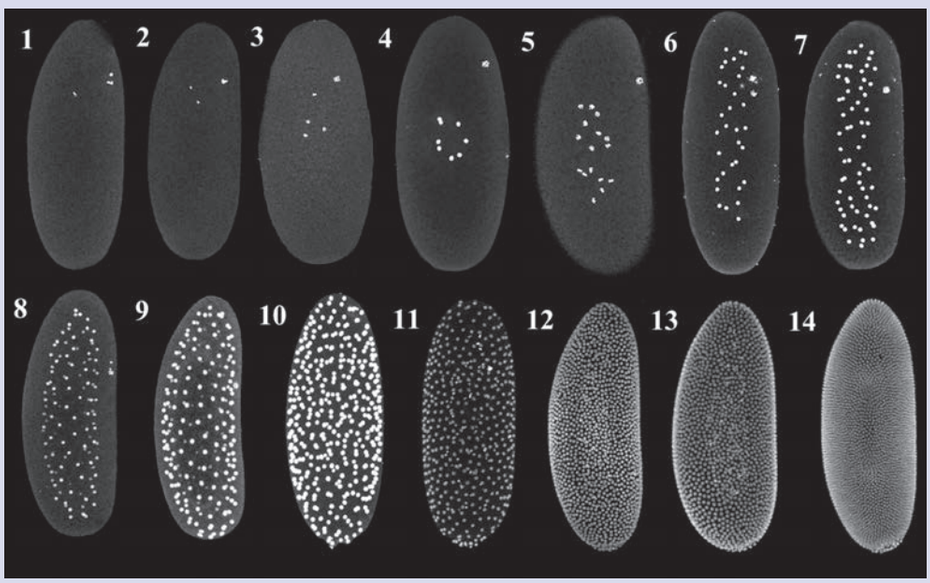
After the sperm and egg haploid nuclei fuse, the diploid, zygotic nucleus undergoes a series of 10 rapid and nearly synchronous cleavages within the central yolky regions of the egg. Large microtubule arrays emanating from the centrioles of the dividing nuclei help direct the nuclei from central regions toward the periphery of the egg (**Box 21-3 Fig. 1**). After eight cleavages, the 256 zygotic nuclei begin to migrate to the periphery. During this migration, they undergo two more cleavages (Box 21-3 Fig. 1, nuclear cleavage cycle 9). Most, but not all, of the resulting approximately 1000 nuclei enter the cortical regions of the egg (Box 21-3 Fig. 1, nuclear cleavage cycle 10). The others ("vitellophages") remain in central regions, where they have a somewhat obscure role in development.

Once the majority of the nuclei reach the cortex at ~90 min following fertilization, they first acquire competence to transcribe Pol II genes. Thus, as in many other organisms such as *Xenopus*, there seems to be a "midblastula transition," whereby early blastomeres (or nuclei) are transcriptionally silent during rapid periods of mitosis. Although causality is unclear, it does seem that DNA undergoing intense bursts of replication cannot simultaneously sustain transcription. These and other observations have led to the suggestion that there is competition between the large macromolecular complexes promoting replication and transcription.

After the nuclei reach the cortex, they undergo another three rounds of cleavage (for a total of 13 divisions after fertilization), leading to the dense packing of about 6000 columnar-shaped nuclei enclosing the central yolk (Box 21-3 Fig. 1, nuclear cleavage cycle 14). Technically, the embryo is still a syncytium, although histochemical staining of early embryos with antibodies against cytoskeletal proteins indicates a highly structured meshwork surrounding each nucleus. During a 1-h period, from 2 to 3 h after fertilization, the embryo undergoes a dramatic cellularization process, whereby cell membranes are formed between adjacent nuclei (Box 21-3 Fig. 1, nuclear cleavage cycle 14). By 3 h after fertilization, the embryo has been transformed into a cellular blastoderm, comparable to the "hollow ball of cells" that characterizes the blastulae of most other embryos.

The recently developed SPIM (single plane illumination microscopy) method has been used for the detailed imaging of *Drosophila* embryogenesis. High-resolution movies of early embryogenesis can be found on the following website: http://www.nature.com/nmeth/journal/vaop/ncurrent/extref/nmeth.2062-sv1.mov.

When the nuclei enter the cortex of the egg, they are totipotent and can form any adult cell type. The location of each nucleus, however, now determines its fate. The 30 or so nuclei



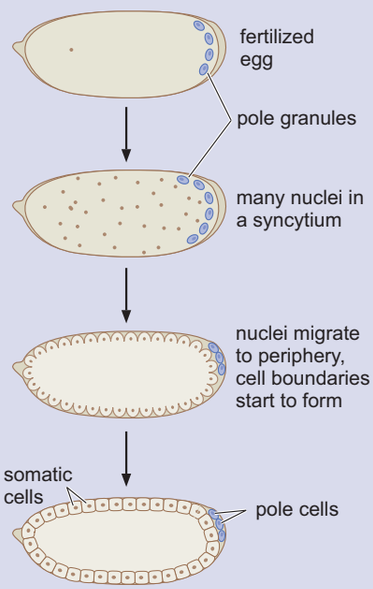
BOX 21-3 FIGURE 1 ***Drosophila* embryogenesis.** *Drosophila* embryos are oriented with the future head pointed up. The numbers refer to the number of nuclear cleavages. Nuclei are stained white within the embryos. For example, stage 1 contains the single zygotic nucleus resulting from the fusion of the sperm and egg pronuclei. The stained material in the *upper right* areas of stages 1–7 are polar bodies. The zygotic nucleus of stage 1 and the nuclei of stages 2, 3, . . ., are in central regions of the embryo. Stage 2 contains two nuclei arising from the first division of the zygotic nucleus. At stage 10, there are approximately 500 nuclei, and most are arranged in a single layer at the cortex (periphery of the embryo). At nuclear cleavage cycle 14, there are more than 6000 nuclei densely packed in a monolayer in the cortex. Cellularization occurs during this stage. (Courtesy of W. Baker and G. Shubiger.)

BOX 21-3 *(Continued)*

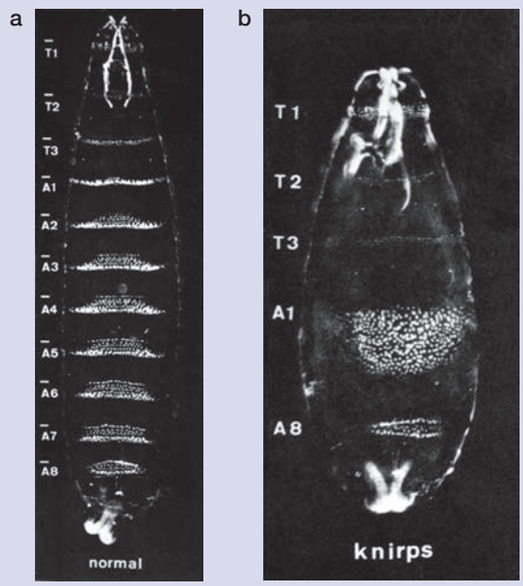
that migrate into posterior regions of the cortex encounter localized protein determinants, such as Oskar, which program these naive nuclei to form the germ cells (**Box 21-3 Fig. 2**). Among the putative determinants contained in the polar plasm are large nucleoprotein complexes, called polar granules. The posterior nuclei bud off from the main body of the embryo along with the polar granules, and the resulting pole cells differentiate into either sperm or eggs, depending on the sex of the embryo. The microinjection of polar plasm into abnormal locations, such as central and anterior regions, results in the differentiation of supernumerary pole cells.

Cortical nuclei that do not enter the polar plasm are destined to form the somatic tissues. Again, these nuclei are totipotent and can form any adult cell type. However, within a very brief period (no more than an hour), each nucleus is rapidly programmed (or specified) to follow a particular pathway of differentiation. This specification process occurs during the period of cellularization, although there is no reason to believe that the deposition of cell membranes between neighboring nuclei is critical for determining cell fate. Different nuclei show distinct patterns of gene transcription before the completion of cell formation. By 3 h after fertilization, each cell possesses a fixed positional identity, so that cells located in anterior regions of the embryo will form head structures in the adult fly, whereas cells located in posterior regions will form abdominal structures.

Systematic genetic screens by Eric Wieschaus and Christiane Nüsslein-Volhard identified approximately 30 "segmentation genes" that control the early patterning of the *Drosophila* embryo. This involved the examination of thousands of dead embryos. At the midpoint of embryogenesis, the ventral skin, or epidermis, secretes a cuticle that contains many fine hairs, or denticles. Each body segment of the embryo contains a characteristic pattern of denticles. Three different classes of segmentation genes were identified on the basis of causing specific disruptions in the denticle patterns of dead embryos. Mutations in the so-called "gap" genes cause the deletion of several adjacent segments (**Box 21-3 Fig. 3**). For example, mutations in the gap gene *knirps* cause the loss of the second through seventh abdominal segments (normal embryos possess eight such segments). Mutations in the "pair-rule" genes cause the loss of alternating segments.



BOX 21-3 FIGURE 2 Development of germ cells. Polar granules located in the posterior cytoplasm of the unfertilized egg contain germ cell determinants and the Nanos mRNA, which is important for the development of the abdominal segments. Nuclei (central dots) begin to migrate to the periphery. Those that enter posterior regions sequester the polar granules and form the pole cells, which form the germ cells. The remaining cells (somatic cells) form all of the other tissues in the adult fly. (Adapted, with permission, from Schneiderman H.A. 1976. *Symp. R. Entomol. Soc. Lond.* **8:** 3–34. © Royal Entomological Society.)



BOX 21-3 FIGURE 3 Dark-field images of normal and mutant cuticles. (a) The pattern of denticle hairs in this normal embryo is slightly different among the different body segments (labeled T1 through A8 in the image). (b) The Knirps mutant (having a mutation in the gap gene *knirps*), shown here, lacks the second through seventh abdominal segments. (Reprinted, with permission, from Nüsslein-Volhard C. and Wieschaus E. 1980. *Nature* **287:** 795–801. © Macmillan. Images courtesy of Eric Wieschaus, Princeton University.)

the cytoplasm of the unfertilized egg. After fertilization, and after the nuclei reach the cortex of the embryo, the Dorsal protein enters nuclei in the ventral and lateral regions but remains in the cytoplasm in dorsal regions (**Fig. 21-12**). The formation of this Dorsal gradient in nuclei across the embryo is very similar, in principle, to the formation of the Gli activator gradient within the vertebrate neural tube (see **Fig. 21-11**).

Regulated nuclear transport of the Dorsal protein is controlled by a cell-signaling molecule called **Spätzle**. This signal is distributed in a ventral-to-dorsal gradient within the extracellular matrix present between the plasma membrane of the unfertilized egg and the outer egg shell. After fertilization, Spätzle binds to the cell-surface Toll receptor. Depending on the concentration of Spätzle, and thus the degree of receptor occupancy in a given region of the syncytial embryo, Toll is activated to a greater or lesser extent. There is peak activation of Toll receptors in ventral regions—where the Spätzle concentration is highest—and progressively lower activation in more lateral regions. Toll signaling causes the degradation of a cytoplasmic inhibitor, Cactus, and the release of Dorsal from the cytoplasm into nuclei. This leads to the formation of a corresponding Dorsal nuclear gradient in the ventral half of the early embryo. Nuclei located in the ventral regions of the embryo contain peak levels of the Dorsal protein, whereas those nuclei located in lateral regions contain lower levels of the protein.

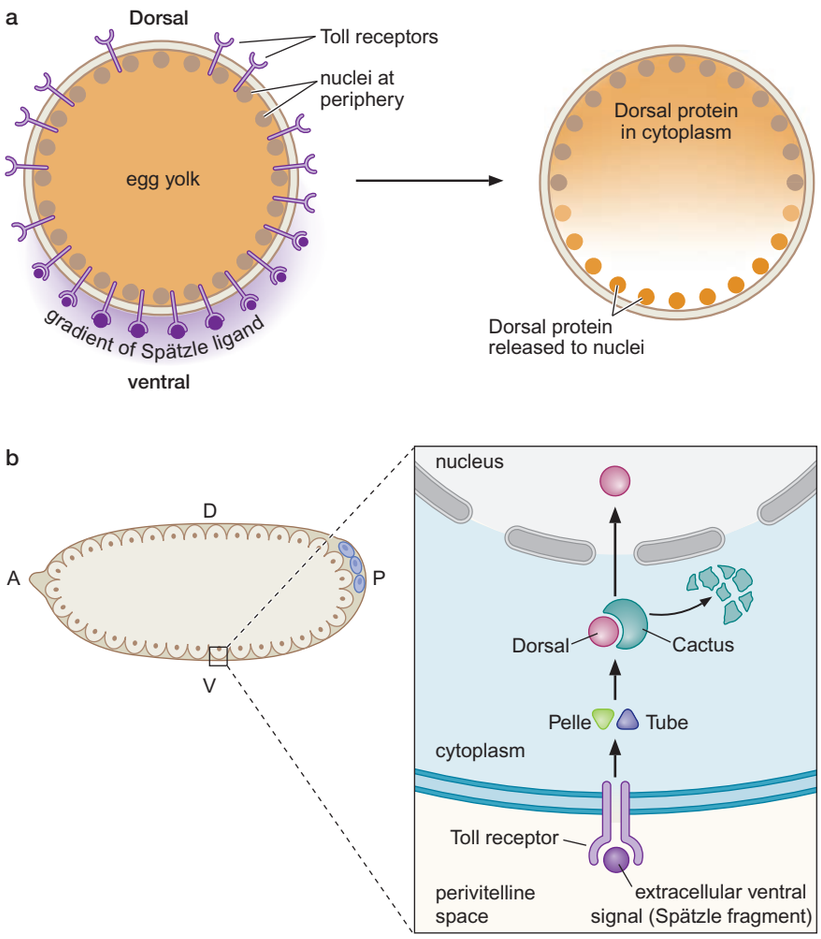


FIGURE 21-12 Spätzle-Toll and Dorsal gradient. (a) The circles represent cross sections through early *Drosophila* embryos. The Toll receptor is uniformly distributed throughout the plasma membrane of the precellular embryo. The Spätzle signaling molecule is distributed in a gradient with peak levels in the ventralmost regions. As a result, more Toll receptors are activated in ventral regions than in lateral and dorsal regions. This gradient in Toll signaling creates a broad Dorsal nuclear gradient. (b) Side view of the embryo with anterior to the *left* and dorsal surface *up*; details of the Toll signaling cascade to the *right*. Activation of the Toll receptor leads to the activation of the Pelle kinase in the cytoplasm. Pelle either directly or indirectly phosphorylates the Cactus protein, which binds and inhibits the Dorsal protein. Phosphorylation of Cactus causes its degradation, so that Dorsal is released from the cytoplasm into nuclei.

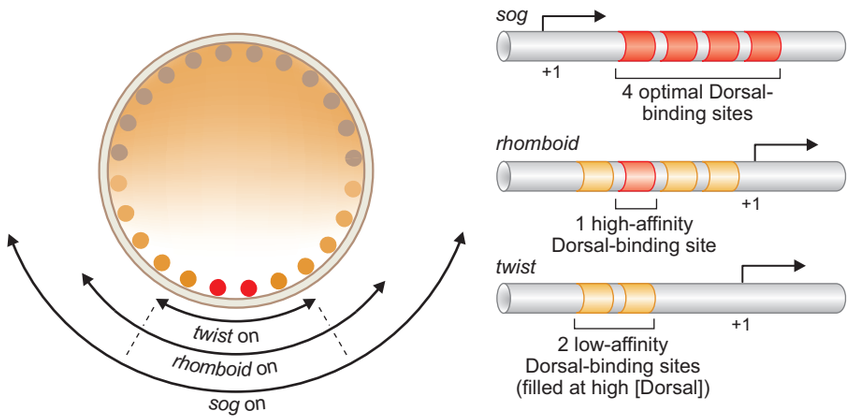


FIGURE 21-13 Three thresholds and three types of regulatory DNAs. The *twist* 5′ regulatory DNA contains two low-affinity Dorsal-binding sites that are occupied only by peak levels of the Dorsal gradient. As a result, *twist* expression is restricted to ventral nuclei. The *rhomboid* 5′ enhancer contains a cluster of Dorsal-binding sites. Only one of these sites represents an optimal, high-affinity Dorsal recognition sequence. This mixture of high- and low-affinity sites allows both high and intermediate levels of the Dorsal gradient to activate *rhomboid* expression in ventrolateral regions. Finally, the *sog* intronic enhancer contains four evenly spaced optimal Dorsal-binding sites. These allow high, intermediate, and low levels of the Dorsal gradient to activate *sog* expression throughout ventral and lateral regions.

The activation of some Dorsal target genes requires peak levels of the Dorsal protein, whereas others can be activated by intermediate and low levels, respectively. In this way, the Dorsal gradient specifies three major thresholds of gene expression across the dorsoventral axis of embryos undergoing cellularization ~2 h after fertilization. These thresholds initiate the differentiation of three distinct tissues: mesoderm, ventral neurogenic ectoderm, and dorsal neurogenic ectoderm (**Fig. 21-13**). Each of these tissues goes on to form distinctive cell types in the adult fly. The mesoderm forms flight muscles and internal organs, such as the fat body, which is analogous to our liver. The ventral and dorsal neurogenic ectoderm form distinct neurons in the ventral nerve cord.

We now consider the regulation of three different target genes that are activated by high, intermediate, and low levels of the Dorsal protein: *twist*, *rhomboid*, and *sog*. The highest levels of the Dorsal gradient—that is, in nuclei with the highest levels of Dorsal protein—activate the expression of the *twist* gene in the ventralmost 18 cells that form the mesoderm (Fig. 21-13). The *twist* gene is not activated in lateral regions, the neurogenic ectoderm, where there are intermediate and low levels of the Dorsal protein. The reason for this is that the *twist* 5′ regulatory DNA contains two low-affinity Dorsal-binding sites (Fig. 21-13). Therefore, peak levels of the Dorsal gradient are required for the efficient occupancy of these sites; the lower levels of Dorsal protein present in lateral regions are insufficient to bind and activate the transcription of the *twist* gene.

The *rhomboid* gene is activated by intermediate levels of the Dorsal protein in the ventral neurogenic ectoderm. The rhomboid 5′-flanking region contains a 300-bp enhancer located ~1.5 kb 5′ of the transcription start

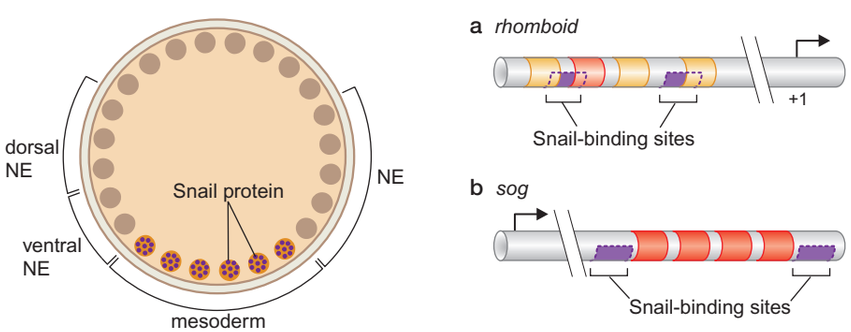


FIGURE 21-14 Regulatory DNAs. (a) The *rhomboid* enhancer contains binding sites for both Dorsal and the Snail repressor. Because the Snail protein is only present in ventral regions (the mesoderm), *rhomboid* is kept off in the mesoderm and restricted to ventral regions of the neurogenic ectoderm (ventral NE). (b) The intronic *sog* enhancer also contains Snail repressor sites. These keep *sog* expression off in the mesoderm and restricted to broad lateral stripes that encompass both ventral and dorsal regions of the neurogenic ectoderm (NE).

site (**Fig. 21-14**a). This enhancer contains a cluster of Dorsal-binding sites, mostly low-affinity sites as seen in the *twist* 5′ regulatory region. At least one of the sites, however, is an optimal, high-affinity site that permits the binding of intermediate levels of Dorsal protein—the amount present in lateral regions. In principle, the *rhomboid* enhancer can be activated by both the high levels of Dorsal protein present in the mesoderm and the intermediate levels present in the ventral neurogenic ectoderm, but it is kept off in the mesoderm by a transcriptional repressor called **Snail**. The Snail repressor is only expressed in the mesoderm; it is not present in the neurogenic ectoderm. The 300-bp *rhomboid* enhancer contains binding sites for the Snail repressor, in addition to the binding sites for the Dorsal activator. This interplay between the broadly distributed Dorsal gradient and the localized Snail repressor leads to the restricted expression of the *rhomboid* gene in the ventral neurogenic ectoderm. We have already seen how the localized Ash1 repressor blocks the action of the SWI5 activator in the daughter cell of budding yeast, and further along in this chapter, we see the extensive use of this principle in other aspects of *Drosophila* development.

The lowest levels of the Dorsal protein, present in lateral regions of the early embryo, are sufficient to activate the *sog* gene in broad lateral stripes that encompass both the ventral and dorsal neurogenic ectoderm. Expression of *sog* is regulated by a 400-bp enhancer located within the first intron of the gene (Fig. 21-14b). This enhancer contains a series of four evenly spaced high-affinity Dorsal-binding sites that can therefore be occupied even by the lowest levels of the Dorsal protein. As seen for *rhomboid*, the presence of the Snail repressor precludes activation of *sog* expression in the mesoderm despite the high levels of Dorsal protein found there. Thus, the differential regulation of gene expression by different thresholds of the Dorsal gradient depends on the combination of the Snail repressor and the affinities of the Dorsal-binding sites.

The occupancy of Dorsal-binding sites is determined by the intrinsic affinities of the sites, as well as protein–protein interactions between Dorsal and other regulatory proteins bound to the target enhancers. For example, we have seen that the 300-bp *rhomboid* enhancer is activated by intermediate levels of the Dorsal gradient in the ventral neurogenic ectoderm. This enhancer contains mostly low-affinity Dorsal-binding sites. However, intermediate levels of Dorsal are sufficient to bind these sites because of protein–protein interactions with another activator protein called **Twist**. However, intermediate levels of Dorsal are sufficient to bind these sites because of protein–protein interactions with additional activators that bind to the rhomboid enhancer. Different mechanisms of cooperative interactions are discussed in **Chapter 19** and in **Box 21-4 (Activator Synergy)**.

Segmentation Is Initiated by Localized RNAs at the Anterior and Posterior Poles of the Unfertilized Egg

At the time of fertilization, the *Drosophila* egg contains two localized mRNAs. One, the *bicoid* mRNA, is located at the anterior pole, and the other, the *oskar* mRNA, is located at the posterior pole (**Fig. 21-15**a). The *oskar* mRNA encodes an RNA-binding protein that is responsible for the assembly of **polar granules**. These are large macromolecular complexes composed of a variety of different proteins and RNAs. The polar granules control the development of tissues that arise from posterior regions of the early embryo, including the abdomen and the pole cells, which are the precursors of the germ cells (Fig. 21-15b).

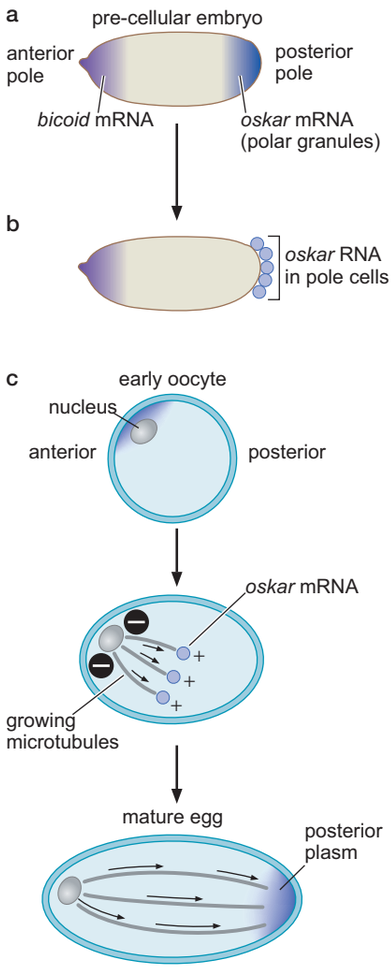


FIGURE 21-15 Localization of maternal mRNAs in the *Drosophila* egg and embryo. (a) The unfertilized *Drosophila* egg contains two localized mRNAs: *bicoid* in anterior regions and *oskar* in posterior regions. (b) The Oskar protein helps coordinate the assembly of the polar granules in the posterior cytoplasm. Nuclei that enter this region bud off the posterior end of the embryo and form the pole cells. (c) During the formation of the *Drosophila* egg, polarized microtubules are formed that extend from the oocyte nucleus and grow toward the posterior plasm. The *oskar* mRNA binds adaptor proteins that interact with the microtubules and thereby transport the RNA to the posterior plasm. The "–" and "+" symbols indicate the direction of the growing strands of the microtubules.

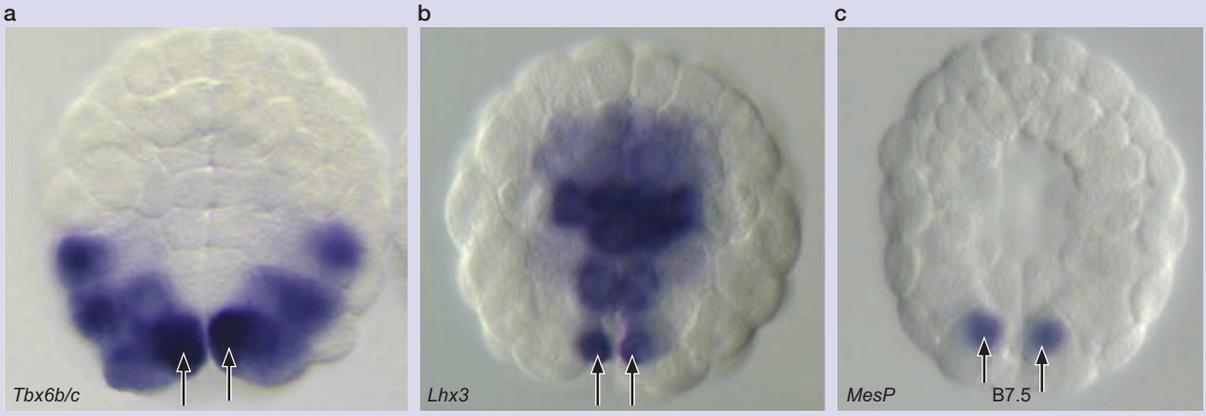
KEY EXPERIMENTS

BOX 21-4 Activator Synergy

Bacterial regulatory proteins such as the *lac* and λ repressors bind as dimers with high affinity. In yeast, the Gal4 activator binds as a dimer with high affinity to induce the expression of Gal1 and other genes required for galactose metabolism (see **Chapter 19**). In contrast, animal cells tend to lack such "dedicated" transcription factors. Many or most such factors bind to DNA as monomers with low affinities. Consequently, gene regulation is inherently more combinatorial in animal cells than in bacteria or yeast. Multiple proteins binding to multiple sites are required to achieve the activation or repression of gene expression.

This principle of combinatorial gene control is a pervasive feature of animal development. Quite often, activators A and B function in a synergistic manner to delineate a restricted pattern of gene expression. Neither A nor B alone is sufficient to do the job. There are many examples of activator synergy in animal development, but we illustrate the principle by considering the specification of the cardiac mesoderm (heart precursor cells) in the sea squirt embryo.

A regulatory gene called *MesP* is a critical determinant of cardiac mesoderm in both sea squirts and vertebrates. It is selectively activated in the B7.5 blastomeres of 110-cell embryos (arrows, **Box 21-4 Fig. 1**). These cells give rise to the beating heart of adult sea squirts. MesP is activated by two transcription factors, Tbx6b/c and Lhx3. Tbx6b/c is expressed throughout the developing tail muscles, as well as the B7.5 blastomeres (arrows, Box 21-4 Fig. 1a). Lhx3 is expressed throughout the presumptive gut, along with the B7.5 blastomeres (arrows, Box 21-4 Fig. 1b). Only the B7.5 blastomeres contain both Tbx6b/c and Lhx3, and in these cells they work synergistically to activate MesP (Box 21-4 Fig. 1c). Because neither transcription factor alone is sufficient for activation, MesP expression is restricted to B7.5 and is inactive in the gut and tail muscles.



BOX 21-4 FIGURE 1 MesP is synergistically activated by two transcription factors. Cells expressing each protein are stained blue. (a) Expression of *Tbx6b/c*. (b) Expression of *Lhx3*. (c) Expression of *MesP*. (Courtesy of Lionel Christiaen. Reproduced from Christiaen L. et al. 2009. *Dev. Biol.* **328:** 552. Parts a, b, and c are from Fig. 3A, 3B, p. 556 and Fig. 6C, p. 558.)

The *oskar* mRNA is synthesized within the ovary of the mother fly. It is first deposited at the anterior end of the immature egg, or **oocyte**, by "helper" cells called **nurse cells**. Both the oocyte and associated nurse cells arise from specialized stem cells within the ovary (see **Box 21-5, Stem Cell Niche**). As the oocyte enlarges to form the mature egg, the *oskar* mRNA is transported from anterior to posterior regions. This localization process depends on specific sequences within the 3′ UTR of the *oskar* mRNA (**Fig. 21-16**). We have already seen how the 3′ UTR of the *ash1* mRNA mediates its localization to the daughter cell of budding yeast by interacting with the growing ends of microtubules. A remarkably similar process controls the localization of the *oskar* mRNA in the *Drosophila* oocyte.

The *Drosophila* oocyte is highly polarized. The nucleus is located in anterior regions; growing microtubules extend from the nucleus into the posterior cytoplasm. The *oskar* mRNA interacts with adaptor proteins that are

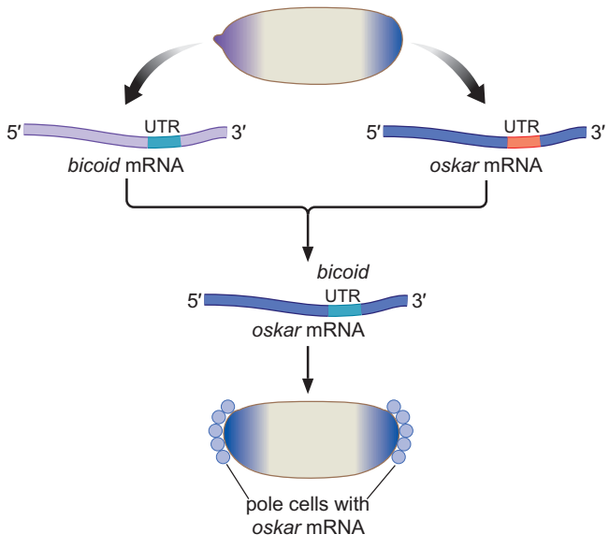


FIGURE 21-16 The *bicoid* and *oskar* mRNAs contain different UTR sequences. The *bicoid* UTR causes it to be localized to the anterior pole, and the distinct *oskar* UTR sequence causes localization in the posterior plasm. An engineered *oskar* mRNA that contains the *bicoid* UTR is localized to the anterior pole, just like the normal *bicoid* mRNA. This mislocalization of *oskar* causes the formation of pole cells in anterior regions. Pole cells also form from the posterior pole because of localization of the normal *oskar* mRNA in the posterior plasm.

associated with the growing "+" ends of the microtubules and are thereby transported away from anterior regions of the egg, where the nucleus resides, into the posterior plasm. After fertilization, the cells that inherit the localized *oskar* mRNA (and polar granules) form the pole cells.

The localization of the *bicoid* mRNA in anterior regions of the unfertilized egg also depends on sequences contained within its 3′ UTR. The nucleotide sequences of the *oskar* and *bicoid* mRNAs are distinct. As a result, they interact with different adaptor proteins and become localized to different regions of the egg. The importance of the 3′ UTRs in determining where each mRNA becomes localized is revealed by the following experiment. If the 3′ UTR from the *oskar* mRNA is replaced with that from *bicoid*, the hybrid *oskar* mRNA is located to anterior regions (just as *bicoid* normally is). This mislocalization is sufficient to induce the formation of pole cells at abnormal locations in the early embryo (see Fig. 21-16). In addition, the mislocalized polar granules suppress the expression of genes required for the differentiation of head tissues. As a result, embryonic cells that normally form head tissues are transformed into germ cells.

Bicoid and Nanos Regulate *hunchback*

The Bicoid regulatory protein is synthesized before the completion of cellularization. As a result, it diffuses away from its source of synthesis at the anterior pole and becomes distributed in a broad concentration gradient along the length of the early embryo. Both high and intermediate concentrations of Bicoid are sufficient to activate *hunchback*, which is essential for the subdivision of the embryo into a series of segments (**Fig. 21-17**). The *hunchback* gene is actually transcribed from two promoters: one is activated by the Bicoid gradient, and the other controls expression in the developing oocyte. The latter, "maternal" promoter leads to the synthesis of a *hunchback* mRNA that is evenly distributed throughout the cytoplasm of unfertilized eggs. The translation of this maternal transcript is blocked in posterior regions by an RNA-binding protein called **Nanos** (Fig. 21-17). Nanos is found only in posterior regions because its mRNA is, in turn, selectively localized there through interactions between its 3′ UTR and the polar granules we encountered earlier.

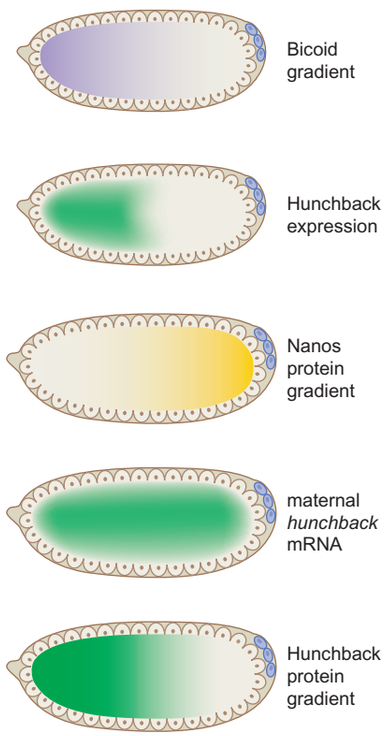


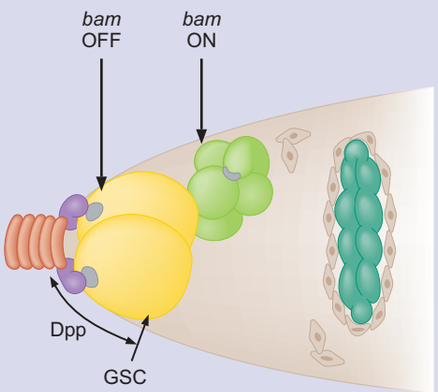
FIGURE 21-17 Hunchback protein gradient and translation inhibition by Nanos. The broad anteroposterior Bicoid protein gradient produces a sharp threshold of *hunchback* gene expression, as *hunchback* is activated by both high and intermediate levels of the Bicoid gradient. The *Nanos* mRNA is associated with polar granules; after its translation, the protein diffuses from posterior regions to form a gradient. The maternal *hunchback* mRNA is distributed throughout the early embryo, but its translation is arrested by the Nanos protein, which binds to specific sequences in the *hunchback* 3′ UTR. The Nanos gradient thereby leads to the formation of a reciprocal Hunchback gradient in anterior regions.

MEDICAL CONNECTIONS

BOX 21-5 Stem Cell Niche

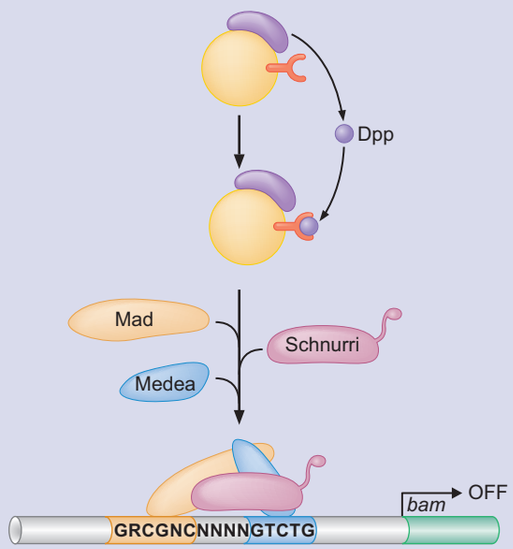
In *Drosophila*, the egg or oocyte arises from a stem cell precursor called the **germline stem cell** (GSC). Quite a lot is known regarding the transition of GSCs into oocytes within the *Drosophila* ovary, and it is likely that many aspects of this mechanism will apply to the development of other classes of stem cells in both flies and humans. Stem cells proliferate only when in direct physical contact with specialized cells, collectively known as the "niche," which produce a signal that triggers proliferation. When stem cells become detached from the niche, proliferation stops and the cells undergo differentiation into specialized cell types. In the *Drosophila* example, detachment of GSCs from the ovary niche causes them to develop into nondividing oocytes, in a process mediated by signal-induced repression. This process is now well understood at the molecular level and works as follows:

Niche cells within the *Drosophila* ovary, called Cap cells, secrete a diffusible signaling molecule called Dpp. Activation of the Dpp receptor within the associated GSCs results in silencing of a critical regulatory gene called *bam*: When transcription of *bam* is blocked, GSCs proliferate. This silencing of *bam* expression depends on direct physical contact between Cap cells and GSCs, similar to the process that results in activation of Notch signaling during formation of the insect nervous system. As GSCs proliferate, some of the daughter cells become detached from the Cap cells and thus are no longer targets of Dpp signaling. In the absence of signaling, *bam* transcription is activated, and the cell stops proliferating; instead, it differentiates into an oocyte (Box 21-5 Fig. 1).



BOX 21-5 FIGURE 1 *bam* expression in developing oocytes. The scheme represents the patterns of expression and distribution of *bam* mRNA and protein. Cap cells (purple) secrete Dpp, which activates its receptor on germline stem cells (GSCs, yellow), resulting in a signaling process that ultimately represses *bam* expression. As GSCs detach from the Cap cells, Dpp signaling is lost, and *bam* mRNA is expressed, leading to production of high levels of its protein in the cytoplasm. In the presence of Bam protein, the detached daughter cells develop into oocyte progenitor cells (green) and further into eight-cell cysts (dark green). (Adapted, with permission, from Chen D. and McKearin D.M. 2003. *Development* **130:** 1159–1170, Fig. 1. © Company of Biologists.)

The basic choice between stem cell proliferation and oocyte differentiation therefore depends on the on/off regulation of *bam* expression. And this regulation is now known to be mediated by a silencer element in the 5′ regulatory region of *bam*, having the sequence GRCGNC(N)$_5$GTCTG (Box 21-5 Fig. 2). Dpp signaling triggers nuclear transport of two Smad regulatory proteins, called Mad and Medea. These proteins bind the two half-sites in the silencer element and, in turn, recruit a transcriptional repressor, called ZF6-6 or Schnurri, that prevents transcription of *bam*. This recruitment of Schnurri and consequent repression of *bam* occurs only in GSCs that remain in contact with the Cap cells. As a result, these cells divide to produce more stem cells. In contrast, in GSC daughter cells that detach from the Cap cells, *bam* is actively transcribed because the signaling pathway leading to gene silencing is disrupted. In these cells, the Dpp receptor is not activated (because signaling is disrupted), and Mad and Medea are not transported to the nucleus and thus do not bind the 5′ silencer element or recruit the Schnurri repressor. Under these conditions, *bam* is expressed, and the daughter cells no longer proliferate but rather differentiate into oocytes. This requirement for direct physical contact between the niche and stem cell and resulting signal-induced repression may be a general mechanism for continuing stem cell proliferation.



BOX 21-5 FIGURE 2 The Dpp pathway actively represses key developmental genes. The binding of Dpp, secreted by Cap cells (purple), to the Dpp receptor on germline stem cells (GSC, yellow) initiates a signal that prompts the transport of Mad (orange) and Medea (blue) into the nucleus. Repressed target genes (here, *bam* is shown as an example) contain a *cis*-acting silencing element that binds Mad and Medea, which together recruit Schnurri (pink) to effectively block transcription. (Adapted, with permission, from Pyrowolakis G. et al. 2004. *Dev. Cell* **7:** 229–240, Fig. 7. © Elsevier.)

Nanos protein binds specific RNA sequences, NREs (Nanos response elements), located in the 3′ UTR of the maternal *hunchback* mRNAs, and this binding causes a reduction in the *hunchback* poly-A tail, which, in turn, destabilizes the RNA and inhibits its translation (see **Chapter 15**). Thus, we see that the Bicoid gradient activates the zygotic *hunchback* promoter in the anterior half of the embryo, whereas Nanos inhibits the translation of the maternal *hunchback* mRNA in posterior regions (see Fig. 21-17). This dual regulation of *hunchback* expression produces a steep Hunchback protein gradient, with the highest concentrations located in the anterior half of the embryo and sharply diminishing levels in the posterior half. Further considerations of gradients and their implications in development are discussed in **Box 21-6, Gradient Thresholds**.

Multiple Enhancers Ensure Precision of *hunchback* Regulation

Many patterning genes are regulated by "redundant" or multiple enhancers. As an example, consider the early activation of the *hunchback* gene by the Bicoid gradient. High and intermediate levels of the gradient activate *hunchback* expression in the anterior half of the embryo (see Fig. 21-17).

ADVANCED CONCEPTS

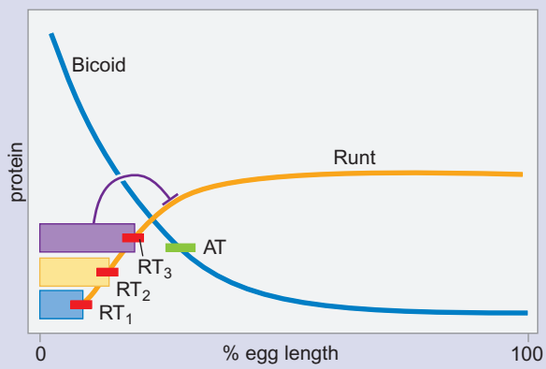
Box 21-6 Gradient Thresholds

We have encountered several examples of regulatory gradients producing different patterns of gene expression. Sonic hedgehog and its transcriptional effector Gli establish differential patterns of Nkx2.2, Olig2, and Pax6 expression in the developing neural tube of vertebrate embryos (Fig. 21-11). The Dorsal gradient generates different patterns of gene expression in the ventral mesoderm, lateral (neurogenic) ectoderm, and dorsal ectoderm of precellular *Drosophila* embryos (Fig. 21-12). The famous Bicoid gradient establishes sequential patterns of "gap" gene expression across the anterior–posterior axis of the precellular embryo (**Box 21-3 Fig. 3**).

Until recently, it was generally assumed that the affinity of Gli-, Dorsal-, and Bicoid-binding sites determined the spatial limits of gene expression. Indeed, we have discussed the evidence that such a mechanism is used for the Dorsal gradient. But there is emerging evidence that different binding affinities might not be sufficient to account for the diverse patterns of gene expression produced by the Gli and Bicoid gradients. For example, it was recently shown that Bicoid target genes activated by high levels of the Bicoid gradient contain similar binding affinities as those regulated by low levels of the gradient. In contrast, a simple binding affinity model would predict that genes activated by high levels of the gradient contain low-affinity sites, whereas genes activated by low levels contain optimal, high-affinity sites.

In fact, it appears that different threshold readouts of the Bicoid gradient depend on opposing repressor gradients, including the Runt repressor (**Box 21-6 Fig. 1**). Target genes RT_1, RT_2, and RT_3 are activated by progressively lower levels of the Bicoid gradient. But RT_1 and RT_2 contain Bicoid-binding sites with similar affinities. Their distinctive limits of expression appear to depend on differential repression by Runt. RT_3 is repressed by high levels of the Runt gradient, whereas RT_2 and RT_1 are repressed by progressively lower levels. It is currently uncertain whether these target genes contain similar Runt-binding sites. Perhaps their differential responses depend on different numbers of sites, with RT_1 containing more Runt repressor sites than RT_2 or RT_3.

As discussed above, the Gli activator gradient in the vertebrate neural tube might also rely on the use of transcriptional repressors to produce different readouts of the Sonic hedgehog gradient (see Fig. 21-11).



BOX 21-6 FIGURE 1 Cooperation of activator and repressor gradients. See text for details. (Adapted from Roth S. and Lynch J. 2012. *Cell* **149:** 511, Fig. 1, p. 512.)

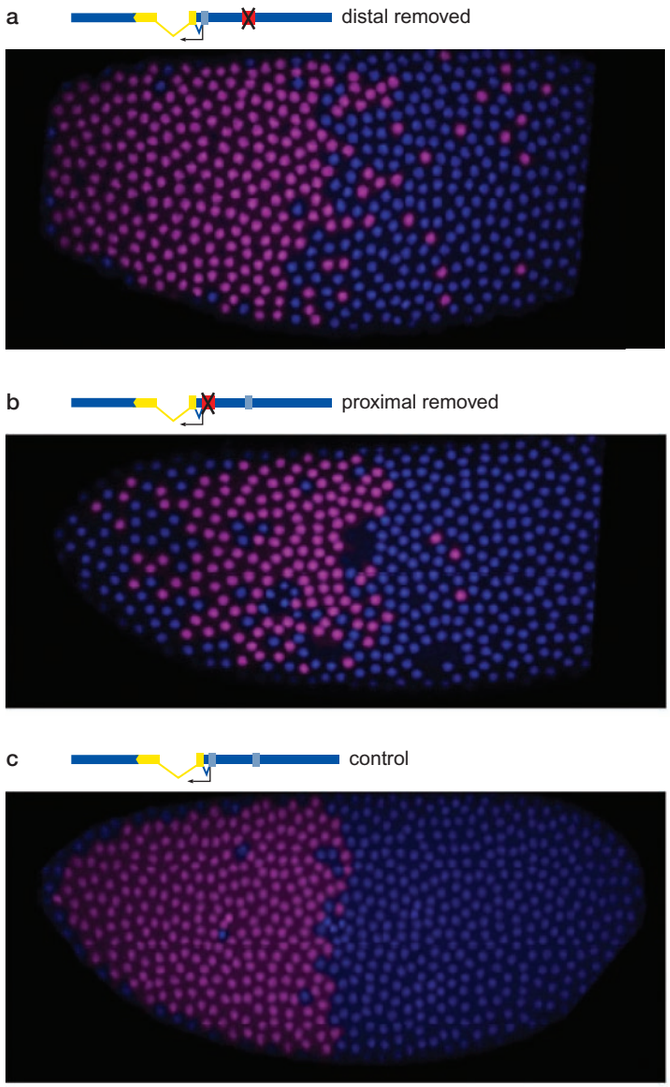


FIGURE 21-18 Hunchback is regulated by two enhancers with similar activities. (a) Early activation of Hunchback transcription occurs from a transgene containing only the proximal enhancer intact. The distal shadow enhancer was inactivated by mutation (indicated by X in diagram). Note that expression is not restricted to the anterior half (*left* half) of the embryo. (b) This panel shows activation obtained when the proximal enhancer is inactivated, leaving only the distal shadow enhancer intact. Expression is sporadic in the anterior regions. (c) Uniform activation and a sharp border are observed when both enhancers are intact. (Courtesy of Mike Levine; described in Perry M.W. et al. 2011. *Proc. Natl. Acad. Sci.* **108:** 13570–13575, Fig. 2A–C, p. 13572.)

Activation is mediated by two separate enhancers, which possess similar arrangements of Bicoid-binding sites and similar regulatory activities (**Fig. 21-18**). Why two enhancers rather than one? Two enhancers produce a sharper, more precise pattern of gene activation than either enhancer alone. Moreover, two enhancers help ensure reliable activation of the gene in large populations of embryos subjected to environmental variations such as changes in temperature. In some cases, multiple enhancers possessing overlapping activities prevent such variations to alter normal development. For example, a regulatory gene called *shavenbaby* is important for the development of tiny sensory hairs along the dorsal surface of advanced-stage embryos. *shavenbaby* is regulated by five separate enhancers distributed over an interval of 40 kb upstream of the transcription start site. Deletions of individual enhancers do not cause significant defects in the morphology of the hairs at optimal temperatures. But under adverse conditions, such as high (30°C) or low (15°C) temperatures, the removal of an enhancer results in fewer or misshapen sensory hairs.

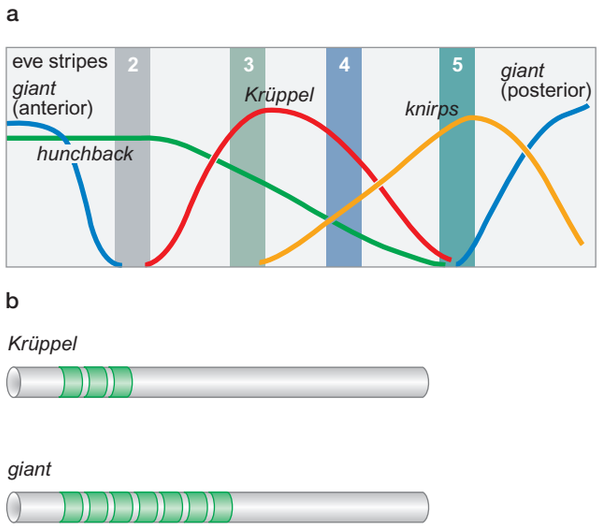


FIGURE 21-19 Expression of *hunchback* forms sequential gap expression patterns. (a) The anteroposterior Hunchback repressor gradient establishes different limits of *Krüppel*, *knirps*, and *giant* expression. High levels of Hunchback are required for the repression of *Krüppel*, but low levels are sufficient to repress *giant*. (b) The *Krüppel* and *giant* 5′ regulatory DNAs contain different numbers of Hunchback repressor sites. There are three sites in *Krüppel*, but seven sites in *giant*. The increased number of Hunchback sites in the *giant* enhancer may be responsible for its repression by low levels of the Hunchback gradient. (a, Redrawn, with permission, from Gilbert S.E. 1997. *Developmental biology*, 5th ed., p. 565, Fig. 14-23. © Sinauer.)

The Gradient of Hunchback Repressor Establishes Different Limits of Gap Gene Expression

Hunchback functions as a transcriptional repressor to establish different limits of expression of the so-called gap genes: *Krüppel*, *knirps*, and *giant* (discussed in Box 21-3). We will see that Hunchback also works in concert with the proteins encoded by these gap genes to produce segmentation stripes of gene expression, the first step in subdividing the embryo into a repeating series of body segments.

The Hunchback protein is distributed in a steep gradient that extends through the presumptive thorax and into the abdomen. High levels of the Hunchback protein repress the transcription of *Krüppel*, whereas intermediate and low levels of the protein repress the expression of *knirps* and *giant*, respectively (**Fig. 21-19**a). We have seen that the binding affinities of the Dorsal activator are responsible for producing different thresholds of gene expression. The Hunchback repressor gradient might not work in the same way. Instead, the *number* of Hunchback repressor sites may be a more critical determinant for distinct patterns of *Krüppel*, *knirps*, and *giant* expression (Fig. 21-19b). The *Krüppel* enhancer contains only three Hunchback-binding sites and is repressed by high levels of the Hunchback gradient. In contrast, the *giant* enhancer contains seven Hunchback sites and is repressed by low levels of the Hunchback gradient. The underlying mechanism here is unknown. Perhaps different thresholds of repression are produced by the additive effects of the individual Hunchback repression domains.

Hunchback and Gap Proteins Produce Segmentation Stripes of Gene Expression

A culminating event in the regulatory cascade that begins with the localized *bicoid* and *oskar* mRNAs is the expression of a "**pair-rule**" gene called *even-skipped*, or simply *eve*. The *eve* gene is expressed in a series of seven alternating, or pair-rule, stripes that extend along the length of the embryo (**Fig. 21-20**). Each *eve* stripe encompasses four cells, and neighboring stripes are separated by interstripe regions—also four cells wide—that express little or no *eve*. These stripes foreshadow the subdivision of the embryo into a repeating series of body segments.

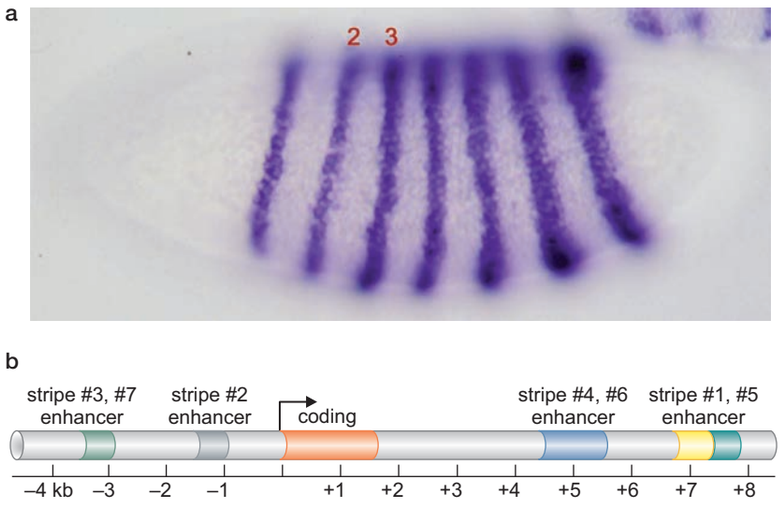


FIGURE 21-20 Expression of the *eve* gene in the developing embryo. (a) *eve* expression pattern in the early embryo. (b) The *eve* locus contains more than 12 kb of regulatory DNA. The 5′ regulatory region contains two enhancers, which control the expression of stripes 2, 3, and 7. Each enhancer is 500 bp in length. The 3′ regulatory region contains three enhancers that control the expression of stripes 4 and 6, stripe 1, and stripe 5, respectively. The five enhancers produce seven stripes of *eve* expression in the early embryo. (a, Image courtesy of Michael Levine.)

The *eve* protein-coding sequence is rather small, <2 kb in length. In contrast, the flanking regulatory DNAs that control *eve* expression encompass more than 12 kb of genomic DNA: ~4 kb located 5′ of the *eve* transcription start site, and ~8 kb in the 3′-flanking region (see Fig. 21-20). The 5′ regulatory region is responsible for initiating stripes 2, 3, and 7, and the 3′ region regulates stripes 1, 4, 5, and 6. The 12 kb of regulatory DNA contains five separate enhancers that together produce the seven different stripes of *eve* expression seen in the early embryo. Each enhancer initiates the expression of just one or two stripes. (In **Box 21-7**, ***cis*-Regulatory Sequences in Animal Development and Evolution**, we discuss further aspects and examples of the modular organization of regulatory elements within animal genomes.) We now consider the regulation of the enhancer that controls the expression of *eve* stripe 2.

The stripe 2 enhancer is 500 bp in length and located 1 kb upstream of the *eve* transcription start site. It contains binding sites for four different regulatory proteins: Bicoid, Hunchback, Giant, and Krüppel (**Fig. 21-21**). We have

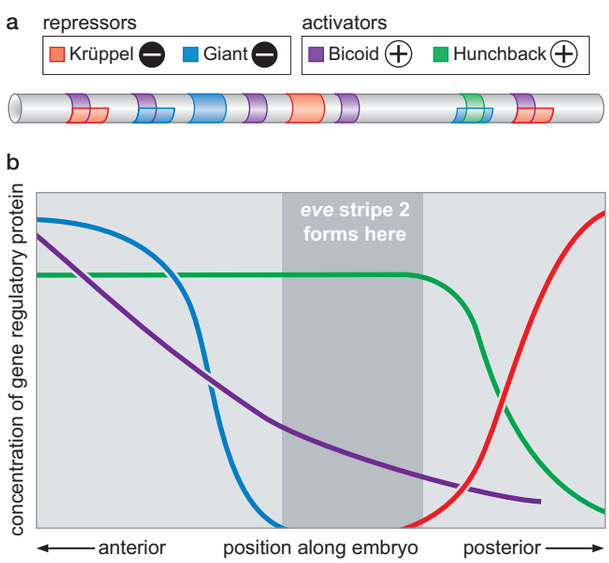


FIGURE 21-21 Regulation of *eve* stripe 2. (a) The 500-bp enhancer contains a total of 12 binding sites for the Bicoid, Hunchback, Krüppel, and Giant proteins. The distribution of these regulatory proteins in the early *Drosophila* embryo is summarized in the diagram shown in b. There are high levels of the Bicoid and Hunchback proteins in the cells that express *eve* stripe 2. The borders of the stripes are formed by the Giant and Krüppel repressors. (Giant is expressed in anterior and posterior regions. Only the anterior pattern is shown; the posterior pattern, which is regulated by Hunchback, is not shown.) (Adapted, with permission, from Alberts B. et al. 2002. *Molecular biology of the cell*, 4th ed.: a, p. 409, Fig. 7-55; b, p. 410, Fig. 7-56. © Garland Science/Taylor & Francis LLC.)

KEY EXPERIMENTS

BOX 21-7 *cis*-Regulatory Sequences in Animal Development and Evolution

cis-regulatory sequences are organized in a modular fashion within animal genomes. In general, there are separate enhancers for the individual components of a complex expression pattern. Consider a gene that is expressed in multiple tissues and organs within a developing mouse embryo, such as the liver, pancreas, and pituitary gland. Odds are that the gene contains separate enhancers for each of these sites of expression. We have seen that the *eve* locus contains five separate enhancers located in the 5′- and 3′-flanking regions (see Fig. 21-20). Each enhancer directs the expression of just one or two of the seven eve stripes in the early *Drosophila* embryo. This type of modular organization facilitates morphological diversity via evolution of *cis*-regulatory sequences, as we discuss below.

Modular Organization Circumvents Pleiotropy

How do patterns of gene expression change during evolution? There is emerging evidence that nucleotide changes within critical activator binding sites eliminate gene expression within a specific tissue or cell type during evolution. Consider the example of pelvic fins in stickleback fish. There are natural variants of sticklebacks that lack pelvic fins. When mated with individuals containing fins, it was possible to identify a major genetic locus responsible for reduced fins. It maps within the 5′-flanking region of the *Pitx1* gene. *Pitx1* is a developmental control gene that is essential for the development of several different tissues in mice, including the thymus, olfactory pit, and hindlimbs. In sticklebacks, it would appear that reduced fins result from point mutations in critical activator sites within the pelvic fin ("hindlimb") enhancer (**Box 21-7 Fig. 1**). These mutations disrupt expression in the developing pelvic fins, but they do not interfere with the activities of the other enhancers required for regulating *Pitx1* in the thymus, olfactory pit, and other tissues where the *Pitx1* gene is active.

Specific alterations within a modular, *cis*-regulatory region are also responsible for the evolution of distinct pigmentation patterns in different species of *Drosophila*. The classical *yellow* (*y*) locus is critical for pigmentation, and simple mutations in the gene result in flies with a yellow body color that lack localized foci of melanin. The *y* gene is regulated by separate enhancers for expression in the bristles, wings, and abdomen, as we now describe.

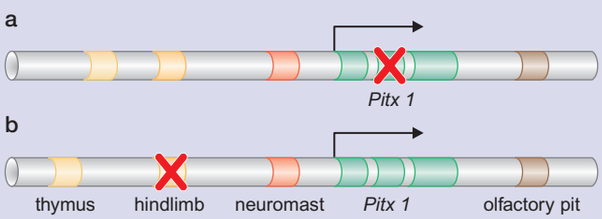
D. melanogaster adults (particularly males) contain intense pigmentation in the posterior abdominal segments. This pigmentation is due to the direct activation of the *y* abdominal enhancer by the Hox protein Abd-B. Drosophilids lacking abdominal segmentation, such as *Drosophila kikkawai*, contain point mutations in a critical Abd-B activator site. This causes a loss of *y* expression in the abdomen and the observed loss of pigmentation.

A separate enhancer controls *y* expression in the wings. In some *Drosophila* species, this enhancer directs a spot of pigmentation in a specific quadrant of the adult male wing (**Box 21-7 Fig. 2**). This spot is a critical component of the courtship ritual. Species lacking the mating spot contain point mutations in the wing enhancer, causing the restricted loss of *y* gene activity without compromising its function in other tissues such as the bristles and abdominal cuticle.

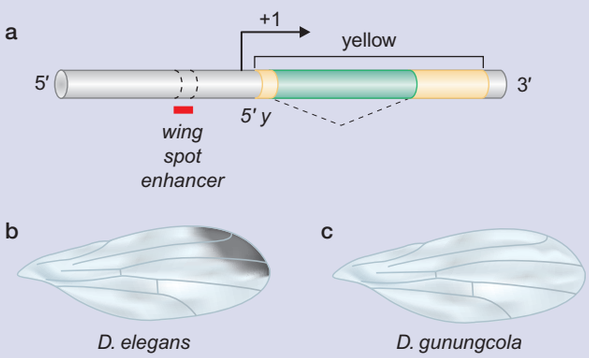
Changes in Repressor Sites Can Produce Big Changes in Gene Expression

The simple loss of critical activator sites within discrete enhancer modules can explain the localized loss of *Pitx1* and *y* gene activities. New patterns of gene expression might arise through the loss of repressor elements.

Most or all of the enhancers active in the early *Drosophila* embryo have repressor binding sites that are responsible for creating sharp boundaries of gene expression. For example, the *eve* stripe 2 enhancer contains binding sites for the Giant and Krüppel repressors, which produce sharp anterior and posterior



BOX 21-7 FIGURE 1 The developmental control gene *Pitx1*. (a) The structure of the *Pitx1* gene with 5′ upstream sequences. Shown here is a lethal null mutation (of a laboratory mouse) within the coding region (second exon) of the gene. (b) In the wild stickleback, a viable regulatory mutation within the 5′ upstream sequence results in reduced pelvic fin size.



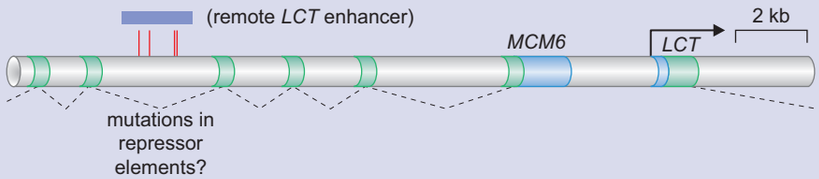
BOX 21-7 FIGURE 2 The *yellow* (*y*) locus of *Drosophila*. (a) The panel shows the structure and upstream regulatory sequences (enhancer sequences) of the *yellow* gene. (b) The normal pigmentation (the "mating spot") of the adult male wing in one species of *Drosophila*. (c) The wing of a species lacking pigmentation; this species carries a mutation in the 5′ spot enhancer.

BOX 21-7 *(Continued)*

borders of gene expression (see Fig. 21-21). Mutations in these sites cause a dramatic expansion in the normal expression pattern: a broad band of expression rather than a tight stripe.

A possible example of evolution via repressor elements is seen for the lactase (*LCT*) gene in human populations. In most primates, the *LCT* gene is expressed at high levels in the small intestines of infants, during the time they obtain milk from their mothers. However, the *LCT* gene is shut off after adolescence. Certain populations of humans are unusual in retaining *LCT* gene expression as adults. This persistence correlates with pastoral societies that use dietary milk long after weaning. Individual populations with persistent *LCT* expression contain nucleotide substitutions in an intronic sequence within the *MCM6* gene, located immediately 5′ of *LCT* (**Box 21-7 Fig. 3**).

These nucleotide changes might damage repressor elements that normally bind a silencer protein responsible for repressing *LCT* expression in the small intestines of adolescents and adults. Such a loss of critical *cis*-regulatory elements would be comparable to the inactivation of the hindlimb/pelvic fin enhancer in the *Pitx1* gene in sticklebacks or the inactivation of the abdominal and wing enhancers in the *y* gene of *Drosophila*. But in the case of the lactase gene, a novel pattern of gene expression is evolved, temporal persistence of LCT activity, because of the loss of repression elements.



BOX 21-7 FIGURE 3 **Structure of the *LCT* gene and its 5′ upstream regulatory region.**

seen how Hunchback functions as a repressor when controlling the expression of the gap genes; in the context of the *eve* stripe 2 enhancer, it works as an activator. In principle, Bicoid and Hunchback can activate the stripe 2 enhancer in the entire anterior half of the embryo because both proteins are present there, but Giant and Krüppel function as repressors that establish the edges of the stripe 2 pattern—the anterior and posterior borders, respectively (see Fig. 21-21).

Gap Repressor Gradients Produce Many Stripes of Gene Expression

eve stripe 2 is formed by the interplay of broadly distributed activators (Bicoid and Hunchback) and localized repressors (Giant and Krüppel). The same basic mechanism applies to the regulation of the other *eve* enhancers as well. For example, the enhancer that directs the expression of *eve* stripe 3 can be activated throughout the early embryo by ubiquitous transcriptional activators. The stripe borders are defined by localized gap repressors: Hunchback establishes the anterior border, whereas Knirps specifies the posterior border (**Fig. 21-22**).

The enhancer that controls the expression of *eve* stripe 4 is also repressed by Hunchback and Knirps. However, different concentrations of these repressors are required in each case. Low levels of the Hunchback gradient that are insufficient to repress the *eve* stripe 3 enhancer are sufficient to repress the *eve* stripe 4 enhancer (Fig. 21-22). This differential regulation of the two enhancers by the Hunchback repressor gradient produces distinct anterior borders for the stripe 3 and stripe 4 expression patterns. The Knirps protein is also distributed in a gradient in the precellular embryo. Higher levels of this gradient are required to repress the stripe 4 enhancer than are needed to repress the stripe 3 enhancer. This distinction

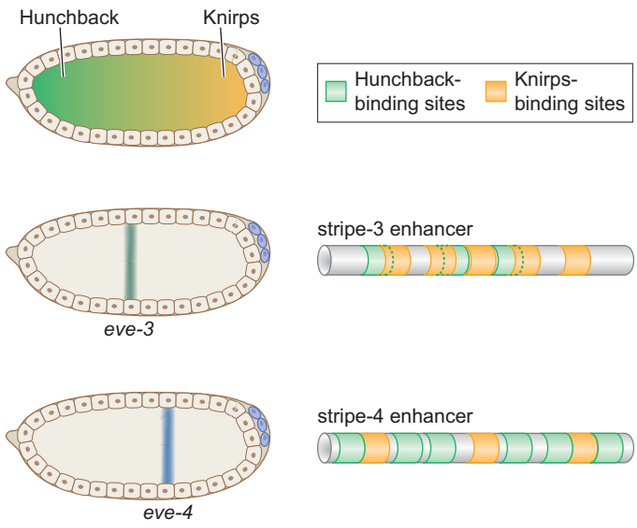


FIGURE 21-22 Differential regulation of the stripe 3 and stripe 4 enhancers by opposing gradients of the Hunchback and Knirps repressors. The two stripes are positioned in different regions of the embryo. The *eve* stripe 3 enhancer is repressed by high levels of the Hunchback gradient but low levels of the Knirps gradient. Conversely, the stripe 4 enhancer is repressed by low levels of the Hunchback gradient but high levels of Knirps. The stripe 3 enhancer contains just a few Hunchback-binding sites, and as a result, high levels of the Hunchback gradient are required for its repression. The stripe 3 enhancer contains many Knirps-binding sites, and consequently, low levels of Knirps are sufficient for repression. The stripe 4 enhancer has the opposite organization of repressor-binding sites. There are many Hunchback sites, and these allow low levels of the Hunchback gradient to repress stripe 4 expression. The stripe 4 enhancer contains just a few Knirps sites, so that high levels of the Knirps gradient are required for repression. Note that the stripe 3 enhancer actually directs the expression of two stripes, 3 and 7. The stripe 4 enhancer directs the expression of stripes 4 and 6. For simplicity, we consider only one of the stripes from each enhancer.

produces discrete posterior borders of the stripe 3 and stripe 4 expression patterns.

We have seen that the Hunchback repressor gradient produces different patterns of Krüppel, Knirps, and Giant expression. This differential regulation might be due to the increasing number of Hunchback-binding sites in the Krüppel, Knirps, and Giant enhancers. A similar principle applies to the differential regulation of the stripe 3 and stripe 4 enhancers by the Hunchback and Knirps gradients. The *eve* stripe 3 enhancer contains relatively few Hunchback binding sites but many Knirps sites, whereas the *eve* stripe 4 enhancer contains many Hunchback sites but relatively few Knirps sites (see Fig. 21-22). Similar principles are likely to govern the regulation of the remaining stripe enhancers that control the *eve* expression pattern (as well as the expression of other pair-rule genes).

Short-Range Transcriptional Repressors Permit Different Enhancers to Work Independently of One Another within the Complex *eve* Regulatory Region

We have seen that *eve* expression is regulated in the early embryo by five separate enhancers. In fact, there are additional enhancers that control *eve* expression in the heart and central nervous system (CNS) of older embryos. This type of complex regulation is not a peculiarity of *eve*. There are genetic loci that contain even more enhancers distributed over even larger distances. For example, several genes are known to be regulated by as many as 10 different enhancers, perhaps more, that are scattered over distances approaching 100 kb (as we discuss below). Thus, genes engaged in important developmental processes are often regulated by multiple enhancers. How do these enhancers work independently of one another to produce additive patterns of gene expression? In the case of *eve*, five separate enhancers produce seven different stripes.

Short-range transcriptional repression is one mechanism for ensuring enhancer autonomy—the independent action of multiple enhancers to generate additive patterns of gene expression. This means that repressors bound to one enhancer do not interfere with the activators bound to another enhancer within the regulatory region of the same gene. For example, we have seen that the Krüppel repressor binds to the *eve* stripe 2 enhancer

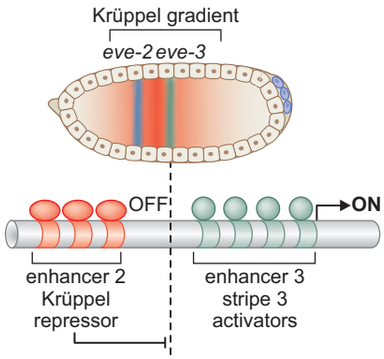


FIGURE 21-23 Short-range repression and enhancer autonomy. Different enhancers work independently of one another in the *eve* regulatory region because of short-range transcriptional repression. Repressors bound to one enhancer do not interfere with activators in the neighboring enhancers. For example, the Krüppel repressor binds to the stripe 2 enhancer and keeps stripe 2 expression off in central regions of the embryo. The *eve* stripe 3 enhancer is expressed in these regions. It is not repressed by Krüppel because it lacks the specific DNA sequences that are recognized by the Krüppel protein. In addition, Krüppel repressors bound to the stripe 2 enhancer do not interfere with the stripe 3 activators because they map too far away. Krüppel must bind no more than 100 bp from upstream activators to block their ability to stimulate transcription. The stripe 2 and stripe 3 enhancers are separated by a 1.5-kb spacer sequence.

FIGURE 21-24 A dominant mutation in the *Antp* gene results in the homeotic transformation of antennae into legs. The fly on the *right* is normal. Note the rudimentary set of antennae at the front end of the head. The fly on the *left* is heterozygous for a dominant *Antp* mutation (*AntpD*/+). It is fully viable and mainly normal in appearance except for the remarkable set of legs emanating from the head in place of antennae. (Courtesy of Matthew Scott.)

and establishes the posterior border of the stripe 2 pattern. The Krüppel repressor works only within the limits of the 500-bp stripe 2 enhancer. It does not repress the core promoter or the activators contained within the stripe 3 enhancer, both of which map more than 1 kb away from the Krüppel repressor sites within the stripe 2 enhancer (**Fig. 21-23**). If Krüppel could function over long distances, or if it mapped near the promoter (like bacterial repressors), then it would interfere with the expression of *eve* stripe 3, because high levels of the Krüppel repressor are present in that region of the embryo where the *eve* stripe 3 enhancer is active.

HOMEOTIC GENES: AN IMPORTANT CLASS OF DEVELOPMENTAL REGULATORS

The genetic analysis of *Drosophila* development led to the discovery of an important class of regulatory genes, the homeotic genes, which cause the morphological diversification of the different body segments. Some homeotic genes control the development of mouth parts and antennae from head segments, whereas others control the formation of wings and halteres from thoracic segments. The two best-studied homeotic genes are *Antp* and *Ubx*, responsible for suppressing the development of antennae and wings, respectively.

Antp (*Antennapedia*) controls the development of the middle segment of the thorax, the mesothorax. The mesothorax produces a pair of legs that are morphologically distinct from the forelegs and hindlegs. *Antp* encodes a homeodomain regulatory protein that is normally expressed in the mesothorax of the developing embryo. The gene is not expressed, for example, in the developing head tissues. But a dominant *Antp* mutation, caused by a chromosome inversion, brings the Antp protein-coding sequence under the control of a "foreign" regulatory DNA that mediates gene expression in head tissues, including the antennae (see **Fig. 21-24**). When misexpressed in the head, *Antp* causes a striking change in morphology: Legs develop instead of antennae.

Ubx (*Ultrabithorax*) encodes a homeodomain regulatory protein that controls the development of the third thoracic segment, the metathorax. *Ubx* specifically represses the expression of genes that are required for the development of the second thoracic segment, or mesothorax. Indeed, *Antp* is one of the genes that it regulates: Ubx represses *Antp* expression in the metathorax and restricts its expression to the mesothorax of developing embryos. Mutants that lack the Ubx repressor show an abnormal pattern of *Antp* expression. The gene is not only expressed within its normal site of action in the developing mesothorax but also misexpressed in the developing metathorax. This misexpression of *Antp* causes a transformation of the metathorax into a duplicated mesothorax.

In adult flies, the mesothorax contains a pair of legs and wings, whereas the metathorax contains a pair of legs and halteres (see **Fig. 21-25**). The halteres are considerably smaller than the wings and function as balancing structures during flight. *Ubx* mutants show a spectacular phenotype: They have four fully developed wings, because of the transformation of the halteres into wings.

The expression of *Ubx* in the different tissues of the metathorax depends on the regulatory sequences that encompass more than 80 kb of genomic DNA. A mutation called *Cbx* (*Contrabithorax*) disrupts this Ubx regulatory DNA without changing the *Ubx* protein-coding region. The *Cbx* mutation causes *Ubx* to be misexpressed in the mesothorax, in addition to its normal site of expression

in the metathorax (Fig. 21-26). Ubx now represses the expression of *Antp*, as well as the other genes needed for the normal development of the mesothorax. As a result, the mesothorax is transformed into a duplicated copy of the normal metathorax. This is a striking phenotype: The wings are transformed into halteres, and the resulting *Cbx* mutant flies look like wingless ants.

Changes in Homeotic Gene Expression Are Responsible for Arthropod Diversity

The interdisciplinary field known as "evo–devo" lies at the cusp of two traditionally isolated areas of research: evolutionary biology and developmental biology. The impetus for evo–devo research is that genetic analysis of development in flies, nematode worms, and other model organisms has identified the key genes responsible for evolutionary diversity. The homeotic genes represent premiere examples of such genes.

The *Drosophila* genome contains a total of eight homeotic genes organized in two gene clusters or complexes: the Antennapedia complex and the Bithorax complex (see **Box 21-8, The Homeotic Genes of *Drosophila* Are Organized in Special Chromosome Clusters**). A typical invertebrate genome contains eight to 10 homeotic genes, usually located within just one complex. Vertebrates have duplicated the ancestral Hox complex and contain four clusters. Changes in the expression and function of individual homeotic genes are responsible for altering limb morphology in arthropods and the axial skeletons of vertebrates. We describe below how changes in Ubx activity have produced evolutionary modifications in insects and other arthropods.

Changes in *Ubx* Expression Explain Modifications in Limbs among the Crustaceans

Crustaceans include most, but not all, of the arthropods that swim. Some live in the ocean, whereas others prefer fresh water. They include some of our favorite culinary dishes, such as shrimp, crab, and lobster. One of the most popular groups of crustaceans for study is *Artemia*, also known as "sea monkeys." Their embryos arrest as tough spores that can be purchased at toy stores. The spores quickly resume development upon addition of salt water.

The heads of these shrimp contain feeding appendages. The thoracic segment nearest the head, T1, contains swimming appendages that look like those further back on the thorax (the second through 11th thoracic segments, T2–T11). *Artemia* belongs to an order of crustaceans known as **branchiopods**. Consider a different order of crustaceans, called **isopods**. Isopods

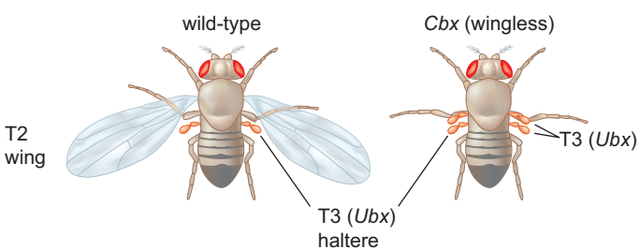


FIGURE 21-26 Misexpression of *Ubx* in the mesothorax results in the loss of wings. The *Cbx* mutation disrupts the regulatory region of *Ubx*, causing its misexpression in the mesothorax, and results in its transformation into the metathorax.

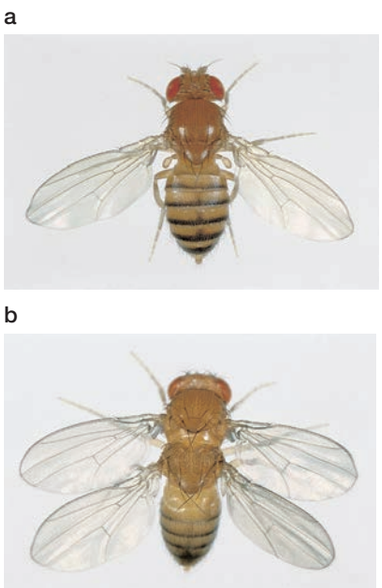


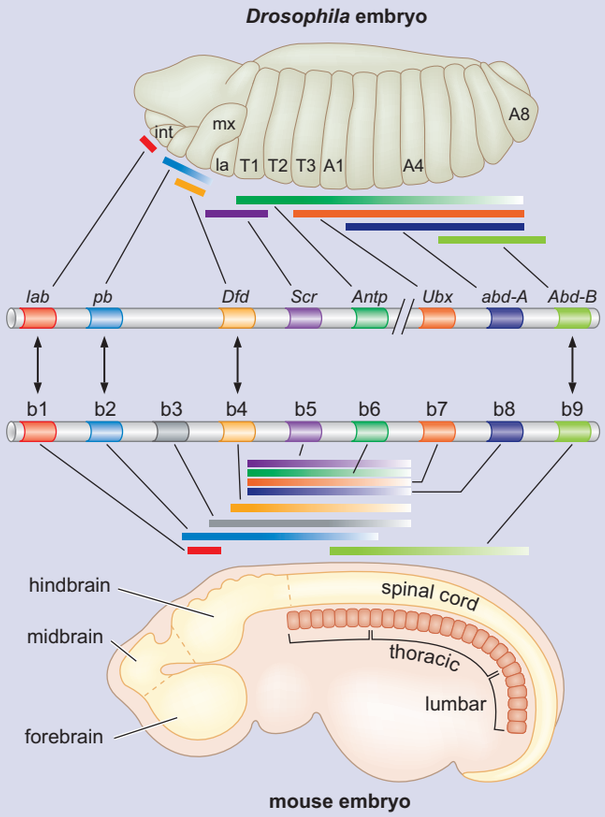
FIGURE 21-25 *Ubx* mutants cause the transformation of the metathorax into a duplicated mesothorax. (a) A normal fly is shown that contains a pair of prominent wings and a smaller set of halteres just behind the wings. (b) A mutant that is homozygous for a weak mutation in the *Ubx* gene is shown. The metathorax is transformed into a duplicated mesothorax. As a result, the fly has two pairs of wings rather than one set of wings and one set of halteres. (Courtesy of E.B. Lewis.)

ADVANCED CONCEPTS

BOX 21-8 Homeotic Genes of *Drosophila* Are Organized in Special Chromosome Clusters

Antp and *Ubx* represent only two of the eight homeotic genes in the *Drosophila* genome. The eight homeotic genes of *Drosophila* are located in two clusters, or gene complexes. Five of the eight genes are located within the Antennapedia complex, and the remaining three genes are located within the Bithorax complex (see **Box 21-8 Fig. 1**). Do not confuse the names of the complex with the individual genes within the complex. For example, the Antennapedia complex is named in honor of the *Antennapedia* (*Antp*) gene, which was the first homeotic gene identified within the complex. There are four other homeotic genes in the Antennapedia complex: *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), and *Sex combs reduced* (*Scr*). Similarly, the Bithorax complex is named in honor of the *Ultrabithorax* (*Ubx*) gene, but there are two others in this complex: *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*). Another insect, the flour beetle, contains a single complex of homeotic genes that includes homologs of all eight homeotic genes contained in the *Drosophila* Antennapedia and Bithorax complexes. The two complexes probably arose from a chromosomal rearrangement within a single ancestral complex.

There is a collinear correspondence between the order of the homeotic genes along the chromosome and their patterns of expression across the anteroposterior axis in developing embryos (see Box 21-8 Fig. 1). For example, the *lab* gene, located in the 3′-most position of the Antennapedia complex, is expressed in the anteriormost head regions of the developing *Drosophila* embryo. In contrast, the *Abd-B* gene, which is located in the 5′-most position of the Bithorax complex, is expressed in the posteriormost regions (see Box 21-8 Fig. 1). The significance of this colinearity has not been established, but it must be important because it is preserved in each of the major groups of arthropods (including flour beetles), as well as all vertebrates that have been studied, including mice and humans.



BOX 21-8 FIGURE 1 Organization and expression of *Hox* genes in *Drosophila* and in the mouse. The figure compares the collinear sequences and transcription patterns of the *Hox* genes in *Drosophila* and in the mouse. (Adapted, with permission, from McGinnis W. and Krumlauf R. 1992. *Cell* **68:** 283–302, Fig. 2. © Elsevier.)

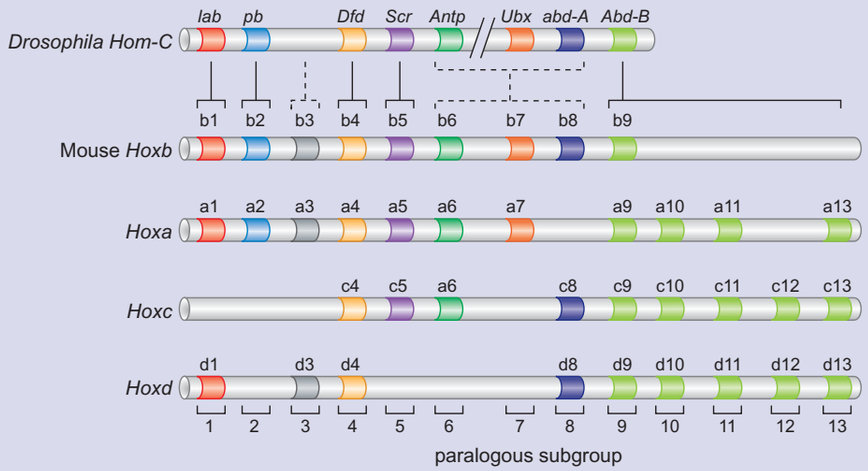
Mammalian Hox *Gene Complexes Control Anteroposterior Patterning*

Mice contain 38 *Hox* genes arranged within four clusters (Hoxa, Hoxb, Hoxc, Hoxd). Each cluster or complex contains nine or 10 *Hox* genes and corresponds to the single homeotic gene cluster in insects that formed the Antennapedia and Bithorax complexes in *Drosophila* (**Box 21-8 Fig. 2**). For example, the *Hoxa-1* and *Hoxb-1* genes are most closely related to the *lab* gene in *Drosophila*, whereas *Hoxa-9* and *Hoxb-9*—located at the other end of their respective complexes—are similar to the *Abd-B* gene.

In addition to this "serial" homology between mouse and fly *Hox* genes, each mouse Hox complex shows the same type of colinearity as that seen in *Drosophila*. For example, *Hox* genes located at the 3′ end of each complex, such as the *Hoxa-1* and *Hoxb-1*, are expressed in the anteriormost regions of developing mouse embryos (future hindbrain). In contrast, *Hox* genes located near the 5′ end of each complex, such as *Hoxa-9* and *Hoxb-9*, are expressed in posterior regions of the embryo (thoracic and lumbar regions of the developing spinal cord). The *Hoxd* complex shows sequential expression across the anteroposterior axis of the developing limbs. A comparable pattern is not observed in insect limbs, suggesting that the *Hoxd* genes have acquired "novel" regulatory DNAs during vertebrate evolution. Indeed, we have already seen in **Chapter 19** that a specialized global control region (GCR) coordinates the expression of the individual *Hoxd* genes in developing limbs.

Altered Patterns of Hox *Expression Create Morphological Diversity in Vertebrates*

Mutations in mammalian *Hox* genes cause disruptions in the axial skeleton, which consists of the spinal cord and the different vertebrae of the backbone. These alterations are evocative of some of the changes in morphology we have seen for the *Antp* and *Ubx* mutants in *Drosophila*.



BOX 21-8 FIGURE 2 Conservation of organization and expression of the homeotic gene complexes in *Drosophila* and in the mouse. (Adapted, with permission, from Gilbert S.E. 2000. *Developmental biology*, 6th ed., Fig. 11.36a. © Sinauer.)

Consider the *Hoxc-8* gene in mice, which is most closely related to the *abd-A* gene of the *Drosophila* Bithorax complex. It is normally expressed near the boundary between the developing rib cage and lumbar region of the backbone, the anterior "tail." (The *abd-A* gene is expressed in the anterior abdomen of the *Drosophila* embryo.) The first lumbar vertebra normally lacks ribs. However, mutant embryos that are homozygous for a knockout mutation in the *Hoxc-8* gene show a dramatic mutant phenotype. The first lumbar vertebra develops an extra pair of vestigial ribs. This type of developmental abnormality is sometimes called a "homeotic" transformation, one in which the proper structure develops in the wrong place. In this case, a vertebra that is typical of the posterior thoracic region develops within the anterior lumbar region.

Maintenance of Hox *Gene Expression Patterns*

Localized patterns of *Hox* gene expression are established in early fly and mouse embryos by combinations of sequence-specific transcriptional activators and repressors. Some of these regulatory proteins are modulated by cell signaling pathways, such as the FGF and Wnt pathways. In *Drosophila*, many of the same gap repressors that establish localized stripes of *eve* expression also control the initial patterns of *Hox* gene expression. These patterns are maintained throughout the life cycle long after the gap repressors are lost.

Consider, as an example, the *Abd-B Hox* gene in *Drosophila*. It is specifically expressed in the posterior abdomen, including the primordia of the fifth through eighth abdominal segments. *Abd-B* expression is initially repressed by the Hb, Kr, and Kni gap repressors in the head, thorax, and anterior abdomen of the early *Drosophila* embryo. These are the same repressors that establish localized stripes of *eve* expression (see Figs. 21-19 and 21-20). These repressors restrict *Abd-B* expression to the posterior abdominal segments.

The maintenance of *Abd-B* expression, as well as the expression of most other *Hox* genes in flies and mammals, depends on a large protein complex, called the Polycomb repression complex (PRC). The PRC binds to *Abd-B* regulatory sequences in cells that fail to activate the gene in the early embryo: the progenitors of the head, thorax, and anterior abdomen. In all of these cells, the PRC causes methylation of lysine 27 on histone H3, and this methylation correlates with the repression of the associated *Abd-B* transcription unit. Conversely, a ubiquitous activator complex, the Trithorax complex (TRC), binds to *Abd-B* regulatory sequences in cells that express the gene in the early embryo (i.e., the posterior abdominal segments). The binding of the TRC leads to the methylation of lysine 4 on histone H3, and this correlates with active transcription of *Abd-B*.

Thus, the PRC and TRC maintain on/off states of *Hox* gene expression depending on the initial expression patterns of these genes in the early embryo. If a given *Hox* gene is repressed in a particular cell, PRC binds and keeps the gene off in all the descendants of that cell. Conversely, if a given *Hox* gene is activated in a particular cell, then TRC will bind and ensure stable expression of the gene in all of its descendants. TRC and PRC serve to maintain a regulatory "memory" of *Hox* gene expression patterns.

MicroRNAs Modulate Hox *Activity*

Many *Hox* gene complexes contain microRNA (miRNA or miR) genes. For example, the fly *ANT-C* contains *miR-10*, and *BX-C* contains *miR-iab4*. The encoded miRNAs are thought to inhibit or attenuate the synthesis of different Hox proteins. The *iab4* miRNA inhibits Ubx protein synthesis in abdominal tissues. Vertebrate *Hox* complexes also contain miR genes, including *miR-10*. The *miR-196* gene is located in 5′ regions of several vertebrate *Hox* complexes. The encoded miRNA is thought to inhibit the synthesis of the Hoxb8 protein in posterior regions of mouse embryos. See **Chapter 20** for more details regarding how miRNAs block or attenuate protein synthesis.

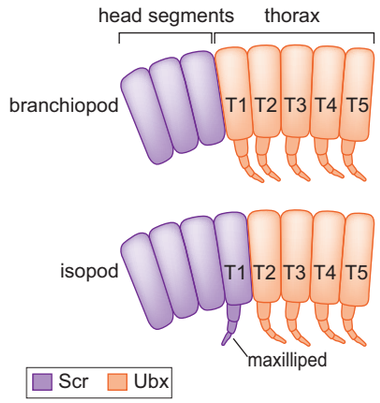


FIGURE 21-27 Changing morphologies in two different groups of crustaceans. In branchiopods, *Scr* expression is restricted to head regions, where it helps promote the development of feeding appendages, whereas *Ubx* is expressed in the thorax, where it controls the development of swimming limbs. In isopods, *Scr* expression is detected in both the head and the first thoracic segment (T1), and as a result, the swimming limb in T1 is transformed into a feeding appendage (the maxilliped). This posterior expansion of Scr was made possible by the loss of *Ubx* expression in T1 because Ubx normally represses *Scr* expression. (Adapted from Levine M. 2002. *Nature* **415:** 848–849, Fig. 2. © Macmillan.)

contain swimming limbs on the second through eighth thoracic segments, just like the branchiopods. But the limbs on the first thoracic segment of isopods have been modified. They are smaller than the others and function as feeding limbs (**Fig. 21-27**). These modified limbs are called maxillipeds (otherwise known as jaw feet), and look like appendages found on the head (although these are not shown in the figure).

Slightly different patterns of *Ubx* expression are observed in branchiopods and isopods. These different expression patterns are correlated with the modification of the swimming limbs on the first thoracic segment of isopods. Perhaps the last shared ancestor of the present branchiopods and isopods contained the arrangement of thoracic limbs seen in *Artemia* (which is itself a branchiopod): All thoracic segments contain swimming limbs. During the divergence of branchiopods and isopods, the *Ubx* regulatory sequences changed in isopods. As a result of this change, *Ubx* expression was eliminated in the first thoracic segment and restricted to segments T2–T8. This shift in *Ubx* expression permitted the formation of a maxilliped in place of the T1 swimming limb. There is a tight correlation between the absence of *Ubx* expression in the thorax and the development of feeding appendages in different crustaceans. For example, lobster embryos lack *Ubx* expression in the first two thoracic segments and contain two pairs of maxillipeds. Cleaner shrimp lack *Ubx* expression in the first three thoracic segments and contain three pairs of maxillipeds.

How Insects Lost Their Abdominal Limbs

All insects have six legs, two on each of the three thoracic segments; this applies to every one of the more than 1 million species of insects. In contrast, other arthropods, such as crustaceans, have a variable number of limbs. Some crustaceans have limbs on every segment in both the thorax and abdomen. This evolutionary change in morphology, the loss of limbs on the abdomen of insects, is not due to altered expression of pattern-determining genes, as seen in the case of maxilliped formation in isopods. Rather, the loss of abdominal limbs in insects is due to functional changes in the Ubx regulatory protein.

In insects, *Ubx* and *abd-A* repress the expression of a critical gene that is required for the development of limbs, called *Distal-less* (*Dll*). In developing *Drosophila* embryos, *Ubx* is expressed at high levels in the metathorax and anterior abdominal segments; *abd-A* expression extends into more posterior abdominal segments. Together, *Ubx* and *abd-A* keep *Dll* off in the first seven abdominal segments. Although *Ubx* is expressed in the metathorax, it does not interfere with the expression of *Dll* in that segment, because *Ubx* is not expressed in the developing T3 legs until after the time when *Dll* is activated. As a result, *Ubx* does not interfere with limb development in T3.

In crustaceans, such as the branchiopod *Artemia* already mentioned, there are high levels of both *Ubx* and *Dll* in all 11 thoracic segments. The expression of *Dll* promotes the development of swimming limbs. Why does Ubx repress *Dll* expression in the abdominal segments of insects but not crustaceans? The answer is that the Ubx protein has diverged between insects and crustaceans. This was shown in the following experiment.

The misexpression of *Ubx* throughout all of the tissues of the presumptive thorax in transgenic *Drosophila* embryos suppresses limb development because of the repression of *Dll* (**Fig. 21-28**). In contrast, the misexpression of the crustacean Ubx protein in transgenic flies does not interfere with *Dll* gene expression and the formation of thoracic limbs. These observations indicate that the *Drosophila* Ubx protein is functionally distinct from Ubx in

FIGURE 21-28 Evolutionary changes in Ubx protein function. (a) The *Dll* enhancer (*Dll*304) is normally activated in three pairs of "spots" in *Drosophila* embryos. These spots go on to form the three pairs of legs in the adult fly. (b) The misexpression of the *Drosophila* Ubx protein (DmUbxHA) strongly suppresses expression from the *Dll* enhancer. (c) In contrast, the misexpression of the Ubx protein from the brine shrimp *Artemia* (AfUbxHA) causes only a slight suppression of the *Dll* enhancer. (Adapted, with permission, from Ronshaugen M. et al. 2002. *Nature* **415:** 914–917, Fig. 2c. © Macmillan. Images courtesy of William McGinnis and Matt Ronshaugen.)

crustaceans. The fly protein represses *Dll* gene expression, whereas the crustacean Ubx protein does not.

What is the basis for this functional difference between the two Ubx proteins? (They share only 32% overall amino acid identity, but their homeodomains are virtually identical—59/60 matches.) It turns out that the crustacean protein has a short motif containing 29 amino acid residues that blocks repression activity. When this sequence is deleted, the crustacean Ubx protein is just as effective as the fly protein at repressing *Dll* gene expression (**Fig. 21-29**).

Both the crustacean and fly Ubx proteins contain multiple repression domains. As discussed in **Chapter 19**, it is likely that these domains interact with one or more transcriptional repression complexes. The "antirepression" peptide present in the crustacean Ubx protein might interfere with the ability of the repression domains to recruit these complexes. When this peptide is attached to the fly protein, the hybrid protein behaves like the crustacean Ubx protein and no longer represses *Dll*.

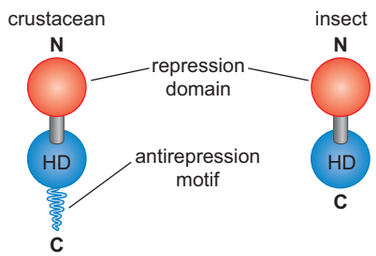


FIGURE 21-29 Comparison of Ubx in crustaceans and in insects. (*Left*) Ubx in crustaceans. The carboxy-terminal antirepression peptide blocks the activity of the amino-terminal repression domain. (*Right*) Ubx in insects. The carboxy-terminal antirepression peptide was lost through mutation. (Adapted, with permission, from Ronshaugen M. et al. 2002. *Nature* **415:** 914–917, Fig. 4b. © Macmillan.)

Modification of Flight Limbs Might Arise from the Evolution of Regulatory DNA Sequences

Ubx has dominated our discussion of morphological change in arthropods. Changes in the Ubx expression pattern appear to be responsible for the transformation of swimming limbs into maxillipeds in crustaceans. Moreover, the loss of the antirepression motif in the Ubx protein likely accounts for the suppression of abdominal limbs in insects. In this final section on that theme, we review evidence that changes in the regulatory sequences in *Ubx* target genes might explain the different wing morphologies found in fruit flies and butterflies.

In *Drosophila*, *Ubx* is expressed in the developing halteres, where it functions as a repressor of wing development. Approximately five to 10 target genes are repressed by Ubx. These genes encode proteins that are crucial for the growth and patterning of the wings (**Fig. 21-30**), and all are expressed in the developing wing. In *Ubx* mutants, these genes are no longer repressed in the halteres, and as a result, the halteres develop into a second set of wings.

Fruit flies are dipterans, and all of the members of this order contain a single pair of wings and a set of halteres. It is likely that Ubx functions as a repressor of wing development in all dipterans. Butterflies belong to a different order of insects, the lepidopterans. All of the members of this order (which also includes moths) contain two pairs of wings rather than a single

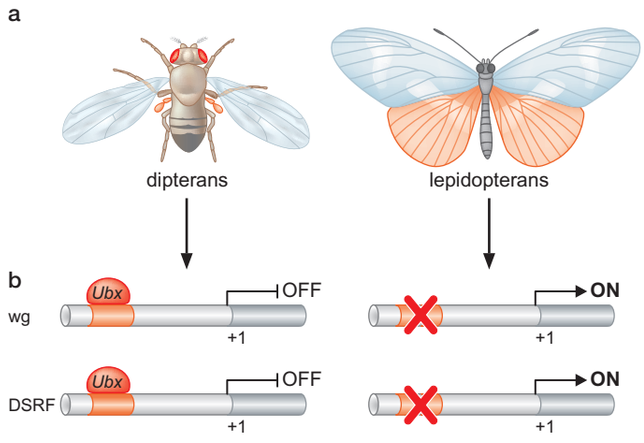


FIGURE 21-30 Changes in the regulatory DNA of *Ubx* target genes. (a) The Ubx repressor is expressed in the halteres of dipterans and hindwings of lepidopterans (orange). (b) Different target genes contain Ubx repressor sites in dipterans. These have been lost in lepidopterans.

pair of wings and a set of halteres. What is the basis for these different wing morphologies in dipterans and lepidopterans?

The two orders diverged from a common ancestor more than 250 million years ago. This is about the time of divergence that separates humans and nonmamalian vertebrates such as frogs. It would seem to be a sufficient period of time to alter *Ubx* gene function through any or all of the three strategies that we have discussed. The simplest mechanism would be to change the *Ubx* expression pattern so that it is lost in the progenitors of the hindwings in Lepidoptera. Such a loss would permit the developing hindwings to express all of the genes that are normally repressed by Ubx. The transformation of swimming limbs into maxillipeds in isopods provides a clear precedent for such a mechanism. However, there is no obvious change in the *Ubx* expression pattern in flies and butterflies; *Ubx* is expressed at high levels throughout the developing hindwings of butterflies.

That leaves us with two possibilities. First, the Ubx protein is functionally distinct in flies and butterflies. The second is that each of the approximately five to 10 target genes that are repressed by Ubx in *Drosophila* has evolved changes in its regulatory DNAs so that they are no longer repressed by Ubx in butterflies (see Fig. 21-30). It seems easier to modify repression activity than to change the regulatory sequences of five to 10 different *Ubx* target genes.

Surprisingly, it appears that the less likely explanation—changes in the regulatory sequences of several *Ubx* target genes—accounts for the different wing morphologies. The Ubx protein appears to function in the same way in fruit flies and butterflies. For example, in butterflies, the loss of Ubx in patches of cells in the hindwing causes them to be transformed into forewing structures (see Fig. 21-30a for the difference between forewings and hindwings). This observation suggests that the butterfly Ubx protein functions as a repressor that suppresses the development of forewings. Although not proven, it is possible that the regulatory DNAs of the wing-patterning genes have lost the Ubx-binding sites (Fig. 21-30b) and they are no longer repressed by Ubx in the developing hindwing.

GENOME EVOLUTION AND HUMAN ORIGINS

We now consider specific examples of comparative genome analysis, with a particular focus on the comparison of animal genomes. Our final discussion

of the comparison of the Neanderthal genome with those of chimpanzee and human provides a few startling insights into human origins.

Diverse Animals Contain Remarkably Similar Sets of Genes

About 100 different animal genomes have been fully sequenced and assembled, but the majority of these sequences correspond to just a few animal groups, centered around the human genome, as well as those of key model organisms such as the fruit fly, *Drosophila melanogaster*, and the nematode worm, *Caenorhabditis elegans*. Thus, several primate genomes (chimpanzees, rhesus monkey, etc.) have been determined to help identify the distinctive features of the human genome (see below). Twelve different species of *Drosophila* have been sequenced to help understand the diversification of distinct species of fruit flies. Currently, just one-third of all animal phyla are represented by a member species with a complete genome sequence assembly.

By far, the most startling discovery arising from comparative genome sequence analysis is the fact that wildly divergent animals, from the sea anemone to humans, possess a highly conserved set of genes. A typical invertebrate genome (e.g., sea anemone, worm, insect) contains approximately 15,000 protein-coding genes. Vertebrates contain a larger number, with an average of about 25,000 genes. However, this larger gene number is not generally due to the invention of "new" genes unique to vertebrates; rather, it is due to the duplication of "old" genes already present in invertebrate genomes. For example, invertebrates contain just a few copies of genes encoding a growth factor called fibroblast growth factor (FGF), whereas a typical vertebrate genome contains more than 20 different FGF genes.

A glimpse into the set of genes required for the distinctive attributes of all animals is provided by the genome sequence assembly of a single-cell eukaryote, a protozoan, called *Monosiga*. This organism is the closest living relative of modern animals. Yet, it lacks many of the genes required for animal development, including those encoding signaling molecules, such as Wingless, the transforming growth factor-β (TGF-β), Hedgehog, and Notch. It also lacks critical regulatory genes responsible for differential gene activity in developing animal embryos, including *Hox* genes and *Hox* clusters. Thus, the evolutionary transition of simple eukaryotes into modern animals required the creation of a large number of novel genes not seen among the simple organisms that lived in the ancient oceans more than 1 billion years ago.

Many Animals Contain Anomalous Genes

Despite a constant set, or "toolkit," of basic genes required for the development of all animals, every genome contains its own distinctive—and sometimes surprising—attribute. Consider the case of the sea squirt. It contains a gene encoding cellulose synthase (**Fig. 21-31**). This enzyme is used by plants to produce cellulose, the major biopolymer of wood. It is absent in virtually all animals, so what is it doing in the sea squirt? The adult is immobile and sits in tide pools where it filters seawater. It contains a rubbery protective sheath composed of tunicin, a biopolymer related to plant cellulose. However, prior to the genome assembly, it was unclear whether the sea squirt contained its own endogenous cellulose synthase gene or employed a symbiotic organism for producing the tunicin sheath. Indeed, there are numerous examples of animals using simple symbionts for unusual genetic functions. For example, termites and wood-eating cockroaches contain symbiotic bacteria in their hindguts that contain the necessary genes required for digesting wood.

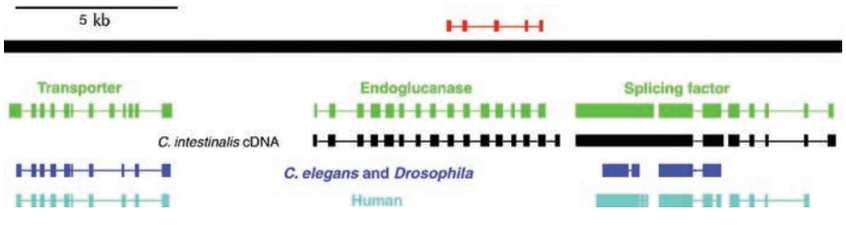


FIGURE 21-31 A plant gene in the *Ciona* genome compared with sequences from other animals. A 20-kb region of one of the *Ciona* scaffolds is shown. This sequence contains an endoglucanase gene, which encodes an enzyme that is required for the degradation and synthesis of cellulose, a major component of plant cell walls. The red rectangles on top represent the *Kerrigan-1* gene of *Arabidopsis*. The gene finder program identified 15 putative exons in the *Ciona* gene, indicated as green rectangles. In reality, there is a 5′ exon present in the cDNA (black rectangles below) that was missed by the computer program. Similarly, a flanking gene, which encodes an RNA splicing factor, is predicted to contain a small intron in a large coding region, whereas the cDNA sequence suggests that there is no intron. There is also a discrepancy in the size of the 5′-most exon. The flanking genes are conserved in worms, flies, and humans, whereas the endoglucanase gene is unique to *Ciona*, which contains a cellulose sheath. Note differences in the detailed intron–exon structures of the flanking genes among the different animal genomes. (Reprinted, with permission, from Dehal et al. 2002. *Science* **298:** 2157–2167, Fig. 8. © AAAS.)

Another surprise came from the analysis of the sea urchin genome; it contains two genes, *RAG1* and *RAG2*, required for the rearrangement of immunoglobin genes in humans and other vertebrates (see **Chapter 12**). One of the distinctive attributes of vertebrates is the ability to mount an adaptive immune response upon infection or injury. This includes the production of specific antibodies that recognize foreign antigens with great specificity and precision. Invertebrates possess a general innate immunity, but they lack the capacity to produce an adaptive immune response. Prior to the sea urchin genome assembly, it was thought that an ancestor of the modern vertebrates acquired a virus or transposon containing the *RAG1* and *RAG2* genes. However, the identification of these genes in sea urchins suggests that this is not true. Instead, the *RAG* genes were acquired by a much more distant ancestor, a progenitor of the so-called Deuterostomes, which diverged into modern echinoderms (e.g., sea urchins) and chordates (e.g., vertebrates) (see **Fig. 21-32**). It would appear that several descendants of this hypothetical ancestor, such as sea squirts, lost the *RAG* genes.

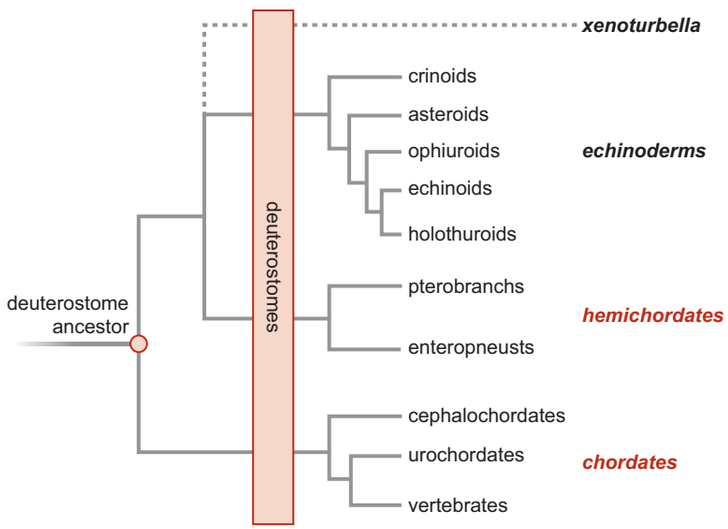


FIGURE 21-32 Deuterostome phylogeny. The deuterostomes include four animal phyla: Xenoturbellida, Echinodermata, Hemichordata, and Chordata. There are five classes of organisms within the echinoderms, two classes of hemichordates, and three classes of chordates. Note that the closest living relatives of the vertebrates are the urochordates, which include the sea squirts (see **Box 19-3**). (Adapted, with permission, from Gerhart J. 2006. *J. Cell Physiol.* **209:** 677–685. © Wiley-Liss, Inc.)

Synteny Is Evolutionarily Ancient

One of the striking findings of comparative genome analysis is the high degree of **synteny**, conservation in genetic linkage, between distantly related animals. There is extensive synteny between mice and humans. In many cases, this linkage even extends to the pufferfish, which last shared a common ancestor with mammals more than 400 million years ago. What is even more remarkable is that some of the linkage relationships are conserved between humans and simple invertebrates, such as sea anemones, which last shared a common ancestor more than 700 million years ago, well before the Cambrian radiation that produced most of the modern animal phyla (**Fig. 21-33**).

Genetic linkage is essential in prokaryotes, where linked genes are coregulated within a common operon (Chapter 18). Such linkage is generally absent in metazoan genomes, although the nematode worm *C. elegans* has been shown to contain a few operons. In other words, neighboring genes are no more likely to be coexpressed (e.g., in blood cells) than unlinked genes. Early comparative genome analyses appeared to confirm that genetic linkage bore no impact on gene regulation. For example, there is no obvious synteny in the arrangement of related genes in mammalian genomes (e.g., mouse and human) and invertebrate genomes such as *C. elegans* and *Drosophila*. However, there is emerging evidence that the genomes of nematode worms and fruit flies are highly "derived." That is, they have undergone distinctive rearrangements and changes not seen in other genomes. Evidence for this view stems from the analysis of the genome of *Nematostella*, a simple sea anemone.

Sea anemones are ancient creatures. They appear in pre-Cambrian fossils, before the first appearance of Arthropods (e.g., trilobites) and annelids. Despite their simplicity and ancient history, they contain several genes that have been lost in flies and worms. What is even more remarkable is that about half of the genetic linkages seen in the human genome are retained, albeit in a somewhat scrambled order, in the *Nematostella* genome (**Fig. 21-33**). Consider the q24 region of human chromosome 10. This region contains 11 genes within a 4-Mb interval, including the gene for actin and *SLK*, which encodes a kinase required for cell division. In the smaller *Nematostella* genome, these 11 genes are not only present but also linked within a 1-Mb interval. The conservation of this local synteny raises the possibility that linkage might influence gene function in some subtle manner, which we are currently unable to explain. By sequencing additional animal genomes, particularly those representing ancient creatures such as sponges and flatworms, it might be possible to reconstruct the ancestral karyotype—the exact chromosome complement and genetic linkages of the metazoan ancestor that generated all the modern animal phyla seen today.

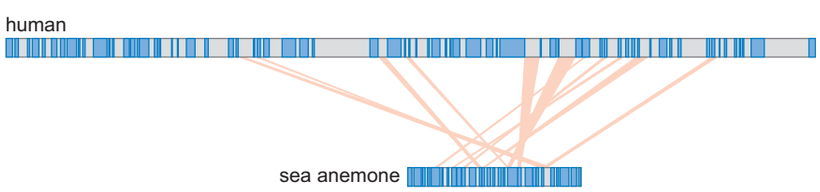


FIGURE 21-33 Conservation of genetic linkage between sea anemones and humans. The top diagram shows a 4-Mb region of human chromosome 10 (the q24 region). The lines show alignments between 11 different genes in this interval and corresponding sequences within a 1-Mb region of a sea anemone chromosome. All 11 genes are located together in both chromosomes, but the exact order of the genes has changed during the course of the ~700 million years since humans and sea anemones last shared a common ancestor.

Deep Sequencing Is Being Used to Explore Human Origins

The ability to sequence large quantities of DNA quickly and inexpensively has created an opportunity to perform experiments that were impossible to imagine even a year ago. One recent example concerns the analysis of the Neanderthal genome.

Modern humans appeared approximately 100,000 years ago and last shared a common ancestor with Neanderthals about 500,000 years ago. There is evidence that modern humans and Neanderthals coexisted in certain locations prior to the disappearance of the Neanderthals about 30,000 years ago. It has been suggested that the two groups mated, resulting in the occurrence of at least some "Neanderthal genes" in the modern human genome. To test this possibility, scientists have recently determined the complete sequence of the Neanderthal genome.

Neanderthal DNA samples have been obtained from well-preserved fossils. However, the DNA is heavily contaminated with bacteria and fungi. Nonetheless, the ability to generate hundreds of thousands of short DNA sequence "reads" (see Chapter 7) permits the identification of authentic Neanderthal DNA among the mixture of contaminating DNAs. In fact, just 2%–3% of the total DNA obtained from a well-preserved Neanderthal fossil corresponds to authentic Neanderthal DNA that matches chimpanzee and human reference genome sequences. The detailed comparison of these Neanderthal sequences with the chimpanzee and human genomes suggests that there was indeed comingling of Neanderthals and modern humans. It is amazing to think that the genomes of extinct organisms can be "resurrected."

SUMMARY

The cells of a developing embryo follow divergent pathways of development by expressing different sets of genes. Most differential gene expression is regulated at the level of transcription initiation. There are three major strategies: mRNA localization, cell-to-cell contact, and the diffusion of secreted signaling molecules.

mRNA localization is achieved by the attachment of specific 3′-UTR sequences to the growing ends of microtubules. This mechanism is used to localize the *ash1* mRNA to the daughter cells of budding yeast. It is also used to localize the *oskar* mRNA to the posterior plasm of the unfertilized egg in *Drosophila*.

In cell-to-cell contact, a membrane-bound signaling molecule alters gene expression in neighboring cells by activating a cell signaling pathway. In some cases, a dormant transcriptional activator, or coactivator protein, is released from the cell surface into the nucleus. In other cases, a quiescent transcription factor (or transcriptional repressor) already present in the nucleus is modified so that it can activate gene expression. Cell-to-cell contact is used by *B. subtilis* to establish different programs of gene expression in the mother cell and forespore. A remarkably similar mechanism is used to prevent skin cells from becoming neurons during the development of the insect central nervous system.

Extracellular gradients of secreted cell-signaling molecules can establish multiple cell types during the development of a complex tissue or organ. These gradients produce intracellular gradients of activated transcription factors, which, in turn, control gene expression in a concentration-dependent fashion. An extracellular Sonic hedgehog gradient leads to a Gli activator gradient in the ventral half of the vertebrate neural tube. Different levels of Gli regulate distinct sets of target genes and thereby produce different neuronal cell types. Similarly, the Dorsal gradient in the early *Drosophila* embryo elicits different patterns of gene expression across the dorsoventral axis. This differential regulation depends on the binding affinities of Dorsal-binding sites in the target enhancers.

The segmentation of the *Drosophila* embryo depends on a combination of localized mRNAs and gradients of regulatory factors. Localized *bicoid* and *oskar* mRNAs, at the anterior and posterior poles, respectively, lead to the formation of a steep Hunchback repressor gradient across the anteroposterior axis. This gradient establishes sequential patterns of Krüppel, Knirps, and Giant in the presumptive thorax and abdomen. These four proteins are collectively called gap proteins; they function as transcriptional repressors that establish localized stripes of pair-rule gene expression. Individual stripes are regulated by separate enhancers located in the regulatory regions of pair-rule genes such as *eve*. Each enhancer contains multiple binding sites for both activators and gap repressors. It is the interplay of broadly distributed activators, such as Bicoid, and localized gap repressors that establish the anterior and posterior borders of individual pair-rule stripes. Separate stripe enhancers work independently of one another to produce composite, seven-stripe patterns of pair-rule expression. This enhancer autonomy is due, in

part, to short-range transcriptional repression. A gap repressor bound to one enhancer does not interfere with the activities of a neighboring stripe enhancer located in the same gene.

Homeotic genes encode regulatory proteins responsible for making the individual body segments distinct from one another. The two best-studied homeotic genes, *Antp* and *Ubx*, control the development of the second and third thoracic segments, respectively, of the fruit fly. The misexpression of *Ubx* in the developing wings causes the development of wingless flies, whereas the misexpression of *Antp* in the head causes a transformation of antennae into legs.

In terms of sheer numbers and diversity, the arthropods can be considered the most successful of all animal phyla. More is known regarding the molecular basis of arthropod diversity than any other group of animals. For example, changes in the expression profile of the *Ubx* gene are cor related with the conversion of swimming limbs into maxillipeds in different groups of crustaceans. Functional changes in the Ubx protein might account for the repression of abdominal limbs in insects. Finally, changes in Ubx target enhancers might explain the different morphologies of the halteres in dipterans and the hindwings of butterflies.

Whole-genome assemblies of diverse animal groups reveal remarkable conservation of the core "genetic toolkit." Most animal genomes contain a similar set of genes, and most differences result from duplication and divergence of "old" genes rather than the invention of new genes. Not only are most genes conserved in different animal groups, but there is also conservation of genetic linkage, or synteny. As many as one-half of all the genes in the human genome are located next to the same neighbors in highly divergent animal groups such as sea anemones. Whole-genome assemblies are being used to obtain insights into our own human origins. A comparison of the chimpanzee and Neanderthal genomes suggest that modern humans contain significant contributions from "extinct" Neanderthals.

BIBLIOGRAPHY

Books

Carroll S.B., Grenier J.K., and Weatherbee S.D. 2005. *From DNA to diversity: Molecular genetics and the evolution of animal design*, 2nd ed. Blackwell, Malden, Massachusetts.

Gilbert S.F. 2010. *Developmental biology*, 9th ed. Sinauer Associates, Sunderland, Massachusetts.

Wolpert L. and Tickle C. 2010. *Principles of development*, 4th ed. Oxford University Press, New York.

mRNA Localization

Macdonald P.M. 2011. mRNA localization: Assembly of transport complexes and their incorporation into particles. *Curr. Opin. Genet. Dev.* **21:** 407–413.

Martin K.C. and Ephrussi A. 2009. mRNA localization: Gene expression in the spatial dimension. *Cell* **136:** 719–730.

Medioni C., Mowry K., and Besse F. 2012. Principes and roles of mRNA localization in animal development. *Development* **139:** 3263–3276.

Cell-to-Cell Contact

Barad O., Hornstein E., and Barkai N. 2011. Robust selection of sensory organ precursors by the Notch-Delta pathway. *Curr. Opin. Cell Biol.* **23:** 663–667.

Schweisguth F. 2004. Notch signaling activity. *Curr. Biol.* **14:** R129–R138.

Shapiro L., McAdams H.H., and Losick R. 2002. Generating and exploiting polarity in bacteria. *Science* **298:** 1942–1946.

Morphogen

Ashe H.L. and Briscoe J. 2006. The interpretation of morphogen gradients. *Development* **133:** 385–394.

Rogers K.W. and Schier A.F. 2011. Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* **27:** 377–407.

Roth S. and Lynch J. 2012. Does the Bicoid gradient matter? *Cell* **149:** 511–512.

Developmental Enhancers

Arnosti D.N. and Kulkarni M.M. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell Biochem.* **94:** 890–898.

Jaeger J. and Reinitz J. 2006. On the dynamic nature of positional information. *BioEssays* **28:** 1102–1111.

Levine M. 2010. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20:** R754–R763.

Segmentation

Lemons D. and McGinnis W. 2006. Genomic evolution of Hox gene clusters. *Science* **313:** 1918–1922.

Lewis E.B. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276:** 565–570.

Tschopp P. and Duboule D. 2011. A genetic approach to the transcriptional regulation of Hox gene clusters. *Ann. Rev. Genet.* **45:** 145–166.