Chapter 18

Regulation of Gene Expression

Overview: Conducting the Genetic Orchestra

- Both prokaryotes and eukaryotes alter their patterns of gene expression in response to changes in environmental conditions.
- Multicellular eukaryotes also develop and maintain multiple cell types.
    - Each cell type contains the same genome but expresses a different subset of genes.
    - During development, gene expression must be carefully regulated to ensure that the right genes are expressed only at the correct time and in the correct place.
- Gene expression in eukaryotes and bacteria is often regulated at the transcription stage.
    - Control of other levels of gene expression is also important.
- RNA molecules play many roles in regulating eukaryotic gene expressions.
- Disruptions in gene regulation may lead to cancer.

Concept 18.1 Bacteria often respond to environmental change by regulating transcription

- Natural selection favors bacteria that express only those genes whose products are needed by the cell.
    - A bacterium in a tryptophan-rich environment that stops producing tryptophan conserves its resources.
- Metabolic control occurs on two levels.
- First, cells can adjust the activity of enzymes already present.
    - This may happen by *feedback inhibition*, in which the activity of the first enzyme in a pathway is inhibited by the pathway's end product.
    - Feedback inhibition, typical of anabolic (biosynthetic) pathways, allows a cell to adapt to short-term fluctuations in the supply of a needed substance.
- Second, cells can vary the number of specific enzyme molecules they make by regulating gene expression.
    - The control of enzyme production occurs at the level of transcription, the synthesis of messenger RNA coding for these enzymes.
    - Genes of the bacterial genome may be switched on or off by changes in the metabolic status of the cell.

- The basic mechanism for the control of gene expression in bacteria, known as the *operon model*, was described by Francois Jacob and Jacques Monod in 1961.

The operon model controls tryptophan synthesis.

- *Escherichia coli* synthesizes tryptophan from a precursor molecule in a series of steps, with each reaction catalyzed by a specific enzyme.
- The five genes coding for the subunits of these enzymes are clustered together on the bacterial chromosome as a transcription unit, served by a single promoter.
- Transcription gives rise to one long mRNA molecule that codes for all five polypeptides in the tryptophan pathway.
- The mRNA is punctuated with start and stop codons that signal where the coding sequence for each polypeptide begins and ends.
- A key advantage of grouping genes with related functions into one transcription unit is that a single on-off switch can control a cluster of functionally related genes.
    - In other words, these genes are *coordinately controlled*.
- When an *E. coli* cell must make tryptophan for itself, all the enzymes are synthesized at one time.
- The switch is a segment of DNA called an **operator**.
- The operator, located within the promoter or between the promoter and the enzyme-coding genes, controls the access of RNA polymerase to the genes.
- The operator, the promoter, and the genes they control constitute an **operon**.
    - The *trp* operon (*trp* for tryptophan) is one of many operons in the *E. coli* genome.
- By itself, an operon is turned on: RNA polymerase can bind to the promoter and transcribe the genes of the operon.
- The operon can be switched off by a protein called the *trp* **repressor**.
    - The repressor binds to the operator, blocks attachment of RNA polymerase to the promoter, and prevents transcription of the operon's genes.
- Each repressor protein recognizes and binds only to the operator of a particular operon.
- The *trp* repressor is the protein product of a **regulatory gene** called *trpR*, which is located at some distance from the operon it controls and has its own promoter.
- Regulatory genes are transcribed continuously at slow rates, and a few *trp* repressor molecules are always present in an *E. coli* cell.
- Why is the *trp* operon not switched off permanently?
- First, binding by the repressor to the operator is reversible.
    - An operator vacillates between two states, with and without a repressor bound to it.
    - The relative duration of each state depends on the number of active repressor molecules around.
- Second, repressors contain allosteric sites that change shape depending on the binding of other molecules.
    - The *trp* repressor has two shapes: active and inactive.
    - The *trp* repressor is synthesized in an inactive form with little affinity for the *trp* operator.

- Only if tryptophan binds to the *trp* repressor at an allosteric site does the repressor protein change to the active form that can attach to the operator, turning the operon off.

- Tryptophan functions in the *trp* operon as a **corepressor**, a small molecule that cooperates with a repressor protein to switch an operon off.
- When concentrations of tryptophan in the cell are high, more tryptophan molecules bind with *trp* repressor molecules, activating them.
  - The active repressors bind to the *trp* operator and turn the operon off.
- At low levels of tryptophan, most of the repressors are inactive, and transcription of the operon's genes resumes.

There are two types of operons: repressible and inducible.

- The *trp* operon is an example of a *repressible operon*, one that is inhibited when a specific small molecule (tryptophan) binds allosterically to a regulatory protein.
- In contrast, an *inducible operon* is stimulated (induced) when a specific small molecule interacts with a regulatory protein.
- The classic example of an inducible operon is the *lac* operon (*lac* for lactose).
- Lactose (milk sugar) is available to *E. coli* in the human colon if the host drinks milk.
  - Lactose metabolism begins with hydrolysis of lactose into its component monosaccharides, glucose and galactose.
  - This reaction is catalyzed by the enzyme ß-galactosidase.
- Only a few molecules of β-galactosidase are present in an *E. coli* cell grown in the absence of lactose.
  - If lactose is added to the bacterium's environment, the number of ß-galactosidase molecules increases by a thousandfold within 15 minutes.
- The gene for ß-galactosidase is part of the *lac* operon, which includes two other genes coding for enzymes that function in lactose metabolism.
- The regulatory gene, *lacI*, located outside the operon, codes for an allosteric repressor protein that can switch off the *lac* operon by binding to the operator.
- Unlike the *trp* operon, the *lac* repressor is active all by itself, binding to the operator and switching the *lac* operon off.
  - An **inducer** *inactivates* the repressor.
  - When lactose is present in the cell, allolactose, an isomer of lactose, binds to the repressor.
  - This inactivates the repressor, and the *lac* operon can be transcribed.
- *Repressible enzymes* generally function in anabolic pathways, synthesizing end products from raw materials.
  - When the end product is present in sufficient quantities, the cell can allocate its resources to other uses.
- *Inducible enzymes* usually function in catabolic pathways, digesting nutrients to simpler molecules.
  - By producing the appropriate enzymes only when the nutrient is available, the cell avoids making proteins that are not needed.

- Both repressible and inducible operons demonstrate *negative* control of genes because active repressors switch off the active form of the repressor protein.
    - It may be easier to see this for the *trp* operon, but it is also true for the *lac* operon.
    - Allolactose induces enzyme synthesis not by acting directly on the genome, but by freeing the *lac* operon from the negative effect of the repressor.

Some gene regulation is positive.

- *Positive* gene control occurs when a protein molecule interacts directly with the genome to switch transcription on.
- The *lac* operon is an example of positive gene regulation.
- When glucose and lactose are both present, *E. coli* preferentially uses glucose.
    - The enzymes for glucose breakdown in glycolysis are always present in the cell.
- Only when lactose is present *and* glucose is in short supply does *E. coli* use lactose as an energy source and synthesize the enzymes for lactose breakdown.
- When glucose levels are low, **cyclic AMP (cAMP)** accumulates in the cell.
- The regulatory protein *catabolite activator protein (CAP)* is an **activator** of transcription.
- When cAMP is abundant, it binds to CAP, and the regulatory protein assumes its active shape and can bind to a specific site at the upstream end of the *lac* promoter.
    - The attachment of CAP to the promoter increases the affinity of RNA polymerase for the promoter, directly increasing the rate of transcription.
    - Thus, this mechanism qualifies as positive regulation.
- If glucose levels in the cell rise, cAMP levels fall.
    - Without cAMP, CAP detaches from the operon and *lac* operon is transcribed only at a low level.
- The *lac* operon is under dual control: negative control by the *lac* repressor and positive control by CAP.
    - The state of the *lac* repressor (with or without bound allolactose) determines whether or not the *lac* operon's genes are transcribed.
    - The state of CAP (with or without bound cAMP) controls the *rate* of transcription if the operon is repressor-free.
    - The operon has both an on-off switch and a volume control.
- CAP works on several operons that encode enzymes used in catabolic pathways. It affects the expression of more than 100 *E. coli* genes.
    - If glucose is present and CAP is inactive, then the synthesis of enzymes that catabolize other compounds is slowed.
    - If glucose levels are low and CAP is active, then the genes that produce enzymes that catabolize whichever other fuel is present are transcribed at high levels.

Concept 18.2 Eukaryotic gene expression is regulated at many stages

- Like unicellular organisms, the tens of thousands of genes in the cells of multicellular eukaryotes turn on and off in response to signals from their internal and external environments.
- Gene expression must be controlled on a long-term basis during cellular differentiation.

Differential gene expression is the expression of different genes by cells with the same genome.

- A typical human cell probably expresses about 20% of its genes at any given time.
    - Highly specialized cells, such as nerves or muscles, express a tiny fraction of their genes.
    - Although all the cells in an organism contain an identical genome, the subset of genes expressed in the cells of each type is unique.
- The differences between cell types are due to **differential gene expression**, the expression of different genes by cells with the same genome.
- The function of any cell, whether a single-celled eukaryote or a particular cell type in a multicellular organism, depends on the appropriate set of genes being expressed.
    - Problems with gene expression and control can lead to imbalance and disease, including cancer.
- Our understanding of the mechanisms that control gene expression in eukaryotes has been enhanced by new research methods, including advances in DNA technology.
- In all organisms, a common control point for gene expression is at transcription, often in response to signals coming from outside the cell.
    - For this reason, the term *gene expression* is often equated with transcription.
- With their greater complexity, eukaryotes have opportunities for controlling gene expression at additional stages.

Chromatin modifications affect the availability of genes for transcription.

- The DNA of eukaryotic cells is packaged with proteins in a complex called chromatin.
    - The basic unit of chromatin is the nucleosome.
- The location of a gene's promoter relative to nucleosomes and to the sites where the DNA attaches to the chromosome scaffold or nuclear lamina affect whether the gene is transcribed.
- Genes of densely condensed heterochromatin are usually not expressed.
- Chemical modifications of the histone proteins and DNA of chromatin play a key role in chromatin structure and gene expression.
- The N-terminus of each histone molecule in a nucleosome protrudes outward from the nucleosome.
    - These histone tails are accessible to various modifying enzymes, which catalyze the addition or removal of specific chemical groups.
- **Histone acetylation** (addition of an acetyl group, —$COCH_3$) and deacetylation of lysines in histone tails appear to play a direct role in the regulation of gene transcription.
- Acetylation of lysines neutralizes their positive charges and reduces the binding of histone tails to neighboring nucleosomes, easing access for transcription proteins.
    - Some of the enzymes responsible for acetylation or deacetylation are associated with or are components of transcription factors that bind to promoters.

- Thus, histone acetylation enzymes may promote the initiation of transcription not only by modifying chromatin structure but also by binding to and recruiting components of the transcription machinery.
- Other chemical groups, such as methyl and phosphate groups, can be reversibly attached to amino acids in histone tails.
    - The attachment of methyl groups (—CH_3) to histone tails leads to condensation of chromatin.
    - The addition of a phosphate group (phosphorylation) to an amino acid next to a methylated amino acid has the opposite effect.
- The recent discovery that modifications to histone tails can affect chromatin structure and gene expression has led to the *histone code hypothesis*.
    - This hypothesis proposes that specific combinations of modifications, as well as the order in which they have occurred, determine chromatin configuration.
    - Chromatin configuration in turn influences transcription.

DNA methylation reduces gene expression.

- While some enzymes methylate the tails of histone proteins, other enzymes methylate certain bases in DNA itself, usually cytosine.
    - DNA methylation occurs in most plants, animals, and fungi.
- Inactive DNA is generally more highly methylated than actively transcribed regions.
    - For example, the inactivated mammalian X chromosome is heavily methylated.
    - Individual genes are usually more heavily methylated in cells where they are not expressed. Removal of extra methyl groups can turn on some of these genes.
- In some species, DNA methylation is responsible for the long-term inactivation of genes during cellular differentiation.
    - Deficient DNA methylation leads to abnormal embryonic development in organisms as different as mice and the plant *Arabidopsis*.
- Once methylated, genes usually stay that way through successive cell divisions in a given individual.
- Methylation enzymes recognize sites on one strand that are already methylated and correctly methylate the daughter strand after each round of DNA replication.
- This methylation pattern accounts for *genomic imprinting*, in which methylation turns off either the maternal or paternal alleles of certain mammalian genes at the start of development.
- The chromatin modifications just discussed do not alter the DNA sequence, and yet they may be passed along to future generations of cells.
- Inheritance of traits by mechanisms not directly involving the nucleotide sequence is called **epigenetic inheritance**.
- The molecular systems for chromatin modification may well interact with each other in a regulated way.
    - In *Drosophila*, experiments suggest that a particular histone-modifying enzyme recruits a DNA methylation enzyme to one region and that the two enzymes collaborate to silence a particular set of genes.

- Working in the opposite order, proteins have also been found that bind to methylated DNA and then recruit histone deacetylation enzymes.
- Thus, a dual mechanism, involving both DNA methylation and histone deacetylation, can repress transcription.

- Researchers are amassing more and more evidence for the importance of epigenetic information in the regulation of gene expression.
  - Epigenetic variations may explain why one identical twin acquires a genetically based disease, such as schizophrenia, while another does not, despite their identical genomes.
  - Alterations in normal patterns of DNA methylation are seen in some cancers, where they are associated with inappropriate gene expression.
- Enzymes that modify chromatin structure are integral parts of the cell's machinery for regulating transcription.

Transcription initiation is controlled by proteins that interact with DNA and with each other.

- Chromatin-modifying enzymes provide initial control of gene expression by making a region of DNA more available or less available for transcription.
- A cluster of proteins called a *transcription initiation complex* assembles on the promoter sequence at the upstream end of the gene.
  - One component, RNA polymerase II, transcribes the gene, synthesizing a primary RNA transcript or pre-mRNA.
  - RNA processing includes enzymatic addition of a 5′ cap and a poly-A tail, as well as splicing out of introns to yield a mature mRNA.
- Multiple **control elements** are associated with most eukaryotic genes.
  - Control elements are noncoding DNA segments that serve as binding sites for protein transcription factors.
  - Control elements and the transcription factors they bind are critical to the precise regulation of gene expression in different cell types.
- To initiate transcription, eukaryotic RNA polymerase requires the assistance of proteins called transcription factors.
- *General transcription factors* are essential for the transcription of *all* protein-coding genes.
  - Only a few general transcription factors independently bind a DNA sequence such as the TATA box within the promoter.
  - Others are involved in protein-protein interactions, binding each other and RNA polymerase II.
- Only when the complete initiation complex has been assembled can the polymerase begin to move along the DNA template strand to produce a complementary strand of RNA.
- The interaction of general transcription factors and RNA polymerase II with a promoter usually leads to only a slow rate of initiation and the production of few RNA transcripts.
- In eukaryotes, high levels of transcription of particular genes depend on the interaction of control elements with *specific transcription factors*.
- Some control elements, named *proximal control elements*, are located close to the promoter.

- *Distal control elements*, grouped as **enhancers**, may be thousands of nucleotides away from the promoter or even downstream of the gene or within an intron.
- A given gene may have multiple enhancers, each active at a different time or in a different cell type or location in the organism.
    - Eukaryotic gene expression can be altered by the binding of specific transcription factors, either activators or repressors, to the control elements of enhancers.
- Two structural elements are common to many activator proteins: a DNA-binding domain and one or more activation domains.
    - Activation domains bind other regulatory proteins or components of the transcription machinery to facilitate transcription.
- Protein-mediated bending of DNA brings bound activators in contact with a group of *mediator proteins* that interact with proteins at the promoter.
    - These interactions help assemble and position the initiation complex on the promoter.
- Eukaryotic repressors can inhibit gene expression by blocking the binding of activators to their control elements or to components of the transcription machinery.
    - Other repressors bind directly to control-element DNA, turning off transcription even in the presence of activators.
- Some activators and repressors act indirectly to influence chromatin structure.
    - Some activators recruit proteins that acetylate histones near the promoters of specific genes, promoting transcription.
    - Some repressors recruit proteins that deacetylate histones, reducing transcription or *silencing* the gene.
- Recruitment of chromatin-modifying proteins seems to be the most common mechanism of repression in eukaryotes.

The control of transcription in eukaryotes depends on the binding of activators to DNA control elements.

- The number of different nucleotide sequences found in control elements is surprisingly small: about a dozen.
- On average, each enhancer is composed of about ten control elements, each of which can bind to only one or two specific transcription factors.
    - The particular *combination* of control elements in an enhancer may be more important than the presence of a unique control element in regulating transcription of the gene.
- Even with only a dozen control element sequences, a large number of combinations are possible.
- A particular combination of control elements is able to activate transcription only when the appropriate activator proteins are present, at a precise time during development or in a particular cell type.
- The use of different combinations of control elements allows fine regulation of transcription with a small set of control elements.
- In prokaryotes, coordinately controlled genes are often clustered into an operon with a single promoter and other control elements upstream.
    - The genes of the operon are transcribed into a single mRNA and translated together.

- In contrast, very few eukaryotic genes are organized this way.
- More commonly, co-expressed genes coding for the enzymes of a metabolic pathway are scattered over different chromosomes.
    - Coordinate gene expression depends on the association of a specific control element or combination of control elements with every gene of a dispersed group.
    - A common group of transcription factors binds to all the genes in the group, promoting simultaneous gene transcription.
- For example, a steroid hormone enters a cell and binds to a specific receptor protein in the cytoplasm or nucleus, forming a hormone–receptor complex that serves as a transcription activator.
    - Every gene whose transcription is stimulated by that steroid hormone has a control element recognized by that hormone–receptor complex.
- Other signal molecules control gene expression indirectly by triggering signal-transduction pathways that lead to activation of transcription.
    - The principle of coordinate regulation is the same: Genes with the same control elements are activated by the same chemical signals.
- Systems for coordinating gene regulation probably arose early in evolutionary history.
- The nucleus has a defined architecture and regulated movements of chromatin.
- Recent techniques allow researchers to cross-link and identify regions of chromosomes that associate with each other during interphase.
- Loops of chromatin extend from individual chromosomal territories into specific sites in the nucleus.
    - Different loops from the same chromosome and loops from other chromosomes congregate in such sites, some of which are rich in RNA polymerases and other transcription-associated proteins.
    - These sites are likely areas specialized for a common function or *transcription factories*.

Post-transcriptional mechanisms play supporting roles in the control of gene expression.

- Regulatory mechanisms that operate after transcription allow a cell to rapidly fine-tune gene expression in response to environmental changes, without altering its transcriptional patterns.
    - RNA processing in the nucleus and the export of mRNA to the cytoplasm provide opportunities for gene regulation that are not available in prokaryotes.
- In **alternative RNA splicing**, different mRNA molecules are produced from the same primary transcript, depending on which RNA segments are treated as exons and which as introns.
    - Regulatory proteins specific to a cell type control intron-exon choices by binding to regulatory sequences within the primary transcript.
- Alternative RNA splicing significantly expands the repertoire of a set of genes.
    - It may explain the surprisingly low number of human genes: similar to those of a soil worm, a mustard plant, or a sea anemone.
    - Between 75% and 100% of human genes that have multiple exons probably undergo alternative splicing.

- The extent of alternative splicing increases the number of possible human proteins, likely correlated with complexity of form.

- The life span of an mRNA molecule is an important factor in determining the pattern of protein synthesis.
  - Prokaryotic mRNA molecules are typically degraded after only a few minutes, while eukaryotic mRNAs typically last for hours, days, or weeks.
  - In red blood cells, mRNAs for hemoglobin polypeptides are unusually stable and are translated repeatedly.
- Nucleotide sequences in the untranslated trailer region (UTR) at the 3′ end affect mRNA stability.
  - Transferring such a sequence from a short-lived mRNA to a normally stable mRNA results in quick mRNA degradation.

Translation presents an opportunity for the regulation of gene expression.

- The initiation of translation of an mRNA can be blocked by regulatory proteins that bind to specific sequences within the 5′ or 3′ UTR of the mRNA, preventing ribosome attachment.
- The mRNAs present in the eggs of many organisms lack poly-A tails of sufficient length to allow initiation of translation.
  - During embryonic development, a cytoplasmic enzyme adds more adenine nucleotides so that translation can begin at the appropriate time.
- Translation of *all* the mRNAs in a eukaryotic cell may be regulated simultaneously by the activation or inactivation of the protein factors required to initiate translation.
  - This mechanism starts the translation of mRNAs that are stored in eggs.
  - Just after fertilization, translation is triggered by the sudden activation of translation initiation factors, resulting in a burst of protein synthesis.
- Some plants and algae store mRNAs during periods of darkness. Light triggers the reactivation of the translational apparatus.

The final opportunities for controlling gene expression occur after translation.

- Often, eukaryotic polypeptides are processed to yield functional proteins.
  - For example, cleavage of pro-insulin forms the active hormone.
- Many proteins must undergo chemical modifications before they are functional.
  - Regulatory proteins may be activated or inactivated by the reversible addition of phosphate groups.
  - Proteins destined for the surface of animal cells acquire sugars.
- Regulation may occur at any of the steps involved in modifying or transporting a protein.
- The length of time a protein functions before it is degraded is strictly regulated.
  - Proteins such as the cyclins that regulate the cell cycle must be relatively short-lived.
- To mark a protein for destruction, the cell attaches a small protein called ubiquitin to it.
  - Giant protein complexes called **proteasomes** recognize and degrade the tagged proteins.

- Mutations making specific cell cycle proteins impervious to proteasome degradation can lead to cancer.
- The scientists worked out the regulated process of protein degradation won the 2004 Nobel Prize in Chemistry.

Concept 18.3 Noncoding RNAs play multiple roles in controlling gene expression

- Only 1.5% of the human genome codes for proteins. Of the remainder, only a very small fraction consists of genes for ribosomal RNA and transfer RNA.
- Until recently, it was assumed that most of the rest of the DNA was untranscribed. Recent data have challenged that assumption, however.
    - Study of a region comprising 1% of the human genome found that over 90% of the region was transcribed.
    - Introns accounted for only a fraction of this transcribed, nontranslated RNA.
- A significant amount of the genome may be transcribed into non–protein-coding RNAs (or *noncoding RNAs* or *ncRNAs)*, including a variety of small RNAs.
- A large, diverse population of RNA molecules may play crucial roles in regulating gene expression in the cell.

MicroRNAs can bind to complementary sequences in mRNA molecules.

- In the past few years, researchers have found small, single-stranded RNA molecules called **microRNAs** (**miRNAs**) that bind to complementary sequences in mRNA molecules.
- miRNAs are formed from longer RNA precursors that fold back on themselves to form one or more short, double-stranded hairpin structures stabilized by hydrogen bonding.
- An enzyme called Dicer cuts each hairpin into a short, double-stranded fragment of about 22 nucleotide pairs.
- One of the two strands is degraded. The other strand (miRNA) associates with a protein complex and directs the complex to any mRNA molecules that have a complementary sequence of 7-8 nucleotides.
- The miRNA–protein complex either degrades the target mRNA or blocks its translation.
- Expression of up to one-half of all human genes may be regulated by miRNAs.
- The phenomenon of inhibition of gene expression by RNA molecules is called **RNA interference** (**RNAi**).
- Injecting double-stranded RNA molecules into a cell somehow turns off expression of a gene with the same sequence as the RNA.
    - This RNA interference is due to **small interfering RNAs** (**siRNAs**), similar in size and function to miRNAs and are generated by similar mechanisms in eukaryotic cells.
- Both miRNAs and siRNAs can associate with the same proteins, with similar results.
    - The distinction between these molecules is the nature of the precursor molecules from which they are formed.
    - Each miRNA forms from a single hairpin in the precursor RNA, while multiple siRNAs form from a longer, double-stranded RNA molecule.

- Cellular RNAi pathways lead to the destruction of RNAs and may have originated as a natural defense against infection by double-stranded RNA viruses.
    - The fact that the RNAi pathway can also affect the expression of nonviral cellular genes may reflect a different evolutionary origin for the RNAi pathway.
- Many species, including mammals, possess long, double-stranded precursors to small RNAs that interfere with various steps in gene expression.

Small RNAs can remodel chromatin and silence transcription.

- Small RNAs can cause remodeling of chromatin structure.
    - In yeast, siRNAs are necessary for the formation of heterochromatin at the centromeres of chromosomes.
- An RNA transcript produced from DNA in the centromeric region of the chromosome is copied into double-stranded RNA by a yeast enzyme and then processed into siRNAs.
    - The siRNAs associate with a protein complex, targeting the complex back to the RNA sequences made from the centromeric sequences of DNA.
    - The proteins in the complex recruit enzymes to modify the chromatin, turning it into the highly condensed centromeric heterochromatin.
- A newly discovered class of small ncRNAs, called piwi-associated RNAs (piRNAs) also induce formation of heterochromatin, blocking expression of parasitic DNA elements in the genome known as transposons.
    - piRNAs, 24–31 nucleotides in length, are processed from single-stranded RNA precursors.
    - In germ cells of many animal species, piRNAs help re-establish appropriate methylation patterns in the genome during gamete formation.
- Chromatin remodeling not only blocks expression of large regions of the chromosome; RNA-based mechanisms may also block the transcription of specific genes.
    - Some plant miRNAs have sequences that bind to gene promoters and can repress transcription; piRNAs can also block expression of specific genes.
    - In some cases, miRNAs and piRNAs activate gene expression.
- Small ncRNAs regulate gene expression at multiple steps and in many ways.
    - Extra levels of gene regulation may allow evolution of a higher degree of complexity of form.
    - An increase in the number of miRNAs encoded in the genomes of species may have allowed morphological complexity to increase over evolutionary time.
- A survey of species suggests that siRNAs evolved first, followed by miRNAs and later piRNAs, which are found only in animals.
    - While there are hundreds of types of miRNA, there appear to be many thousands of types of piRNAs, allowing the potential for very sophisticated gene regulation by piRNAs.
- Many ncRNAs play important roles in embryonic development, the ultimate example of an elaborate program of regulated gene expression.

Concept 18.4 A program of differential gene expression leads to the different cell types in a multicellular organism

- In the development of most multicellular organisms, a single-celled zygote gives rise to cells of many different types.
    - Each type has a different structure and corresponding function.
    - Cells of different types are organized into tissues, tissues into organs, organs into organ systems, and organ systems into the whole organism.
- Thus, the process of embryonic development must give rise not only to cells of different types but also to higher-level structures arranged in a particular way in three dimensions.

A genetic program is expressed during embryonic development.

- As a zygote develops into an adult organism, its transformation results from three interrelated processes: cell division, cell differentiation, and morphogenesis.
- Through a succession of mitotic cell divisions, the zygote gives rise to many cells.
    - Cell division alone would produce only a great ball of identical cells.
- During development, cells become specialized in structure and function, undergoing **cell differentiation**.
- Different kinds of cells are organized into tissues and organs.
- The physical processes that give an organism its shape constitute **morphogenesis**, the "creation of form."
- Cell division, cell differentiation, and morphogenesis have their basis in cellular behavior.
    - Morphogenesis can be traced back to changes in the shape and motility of cells in the various embryonic regions.
    - The activities of a cell depend on the genes it expresses and the proteins it produces.
    - Because almost all cells in an organism have the same genome, differential gene expression results from differential gene regulation in different cell types.
- Why are different sets of activators present in different cell types?
- One important source of information early in development is the egg's cytoplasm, which contains both RNA and proteins encoded by the mother's DNA, distributed unevenly in the unfertilized egg.
- Maternal substances that influence the course of early development are called **cytoplasmic determinants**.
    - These substances regulate the expression of genes that affect the developmental fate of the cell.
    - After fertilization, the cell nuclei resulting from mitotic division of the zygote are exposed to different cytoplasmic environments.
    - The set of cytoplasmic determinants a particular cell receives helps determine its developmental fate by regulating expression of the cell's genes during cell differentiation.
- The other important source of developmental information is the environment around the cell, especially signals impinging on an embryonic cell from nearby cells.
    - In animals, these signals include contact with cell-surface molecules on neighboring cells and the binding of growth factors secreted by neighboring cells.

- These signals cause changes in the target cells, a process called **induction**.
    - The molecules conveying these signals within the target cells are cell-surface receptors and other proteins expressed by the embryo's own genes.
    - The signal molecules send a cell down a specific developmental path by causing a change in its gene expression that eventually results in observable cellular changes.

Cell differentiation is due to the sequential regulation of gene expression.

- During embryonic development, cells become visibly different in structure and function as they differentiate.
- The earliest changes that set a cell on a path to specialization show up only at the molecular level.
    - Molecular changes in the embryo drive the process, called **determination**, which leads to the observable differentiation of a cell.
- Once it has undergone determination, an embryonic cell is irreversibly committed to its final fate.
    - If a determined cell is experimentally placed in another location in the embryo, it will differentiate as if it were in its original position.
- The outcome of determination—observable cell differentiation—is caused by the expression of genes that encode *tissue-specific proteins*.
    - These proteins give a cell its characteristic structure and function.
- Differentiation begins with the appearance of cell-specific mRNAs and is eventually observable in the microscope as changes in cellular structure.
- In most cases, the pattern of gene expression in a differentiated cell is controlled at the level of transcription.
- Cells produce the proteins that allow them to carry out their specialized roles in the organism.
    - For example, liver cells specialize in making albumin, while lens cells specialize in making crystalline.
    - Skeletal muscle cells have high concentrations of proteins specific to muscle tissues, such as a muscle-specific version of the contractile proteins myosin and actin, as well as membrane receptor proteins that detect signals from nerve cells.
- Muscle cells develop from embryonic precursors that have the potential to develop into a number of alternative cell types.
    - Although the committed cells are unchanged, they are now *myoblasts*.
    - Eventually, myoblasts begin to synthesize muscle-specific proteins and fuse to form mature, elongated, multinucleate skeletal muscle cells.
- Researchers have worked out the events at the molecular level that lead to muscle cell determination by growing myoblasts in culture and analyzing them with molecular biology techniques.
    - Researchers isolated different genes, caused each to be expressed in a separate embryonic precursor cell, and looked for differentiation into myoblasts and muscle cells.
    - They identified several "master regulatory genes" that, when transcribed and translated, commit the cells to become skeletal muscle.

- One of these master regulatory genes is called *myoD*.
    - *myoD* encodes MyoD protein, a transcription factor that binds to specific control elements in the enhancers of various target genes and stimulates their expression.
    - Some target genes for MyoD encode for other muscle-specific transcription factors.
    - MyoD also stimulates expression of the *myoD* gene itself, helping to maintain the cell's differentiated state.
- All the genes activated by MyoD have enhancer control elements recognized by MyoD and are thus coordinately controlled.
- The secondary transcription factors activate the genes for proteins such as myosin and actin to confer the unique properties of skeletal muscle cells.
- The MyoD protein is capable of changing fully differentiated fat and liver cells into muscle cells.
- Not *all* cells can be transformed by MyoD, however.
    - Nontransforming cells may lack a *combination* of regulatory proteins in addition to MyoD.

Pattern formation sets up the embryo's body plan.

- Cytoplasmic determinants and inductive signals contribute to **pattern formation**, the development of spatial organization in which the tissues and organs of an organism are all in their characteristic places.
- Pattern formation begins in the early embryo, when the major axes of an animal are established.
- Before specialized tissues and organs form, the relative positions of a bilaterally symmetrical animal's three major body axes (anterior-posterior, dorsal-ventral, right-left) are established.
- The molecular cues that control pattern formation, **positional information**, are provided by cytoplasmic determinants and inductive signals.
    - These signals tell a cell its location relative to the body axes and to neighboring cells and determine how the cell and its progeny will respond to future molecular signals.
- Studies of pattern formation in *Drosophila melanogaster* have established that genes control development and have identified the key roles of specific molecules in defining position and directing differentiation.
- Combining anatomical, genetic, and biochemical approaches in the study of *Drosophila* development, researchers have discovered developmental principles common to many other species, including humans.
- Fruit flies and other arthropods have a modular construction.
    - An ordered series of segments make up the three major body parts: the head, thorax (with wings and legs), and abdomen.
- Cytoplasmic determinants in the unfertilized egg provide positional information for two developmental axes (anterior-posterior and dorsal-ventral axis) before fertilization.
- The *Drosophila* egg develops in the female's ovary, surrounded by ovarian cells called nurse cells and follicle cells that supply the egg cell with nutrients, mRNAs, and other substances.
- During fruit fly development, the egg forms a segmented larva, which goes through three larval stages.

- The fly larva forms a pupal cocoon within which it metamorphoses into an adult fly.

- In the 1940s, Edward B. Lewis used mutants to investigate *Drosophila* development.
  - Bizarre developmental mutations were on the fly's genetic map, providing the first concrete evidence that genes somehow direct the developmental process.
  - These **homeotic genes** control pattern formation in the late embryo, larva, and adult.
- In the late 1970s, Christiane Nüsslein-Volhard and Eric Weischaus set out to identify *all* the genes that affect segmentation in *Drosophila*. They faced three problems.
- First, because *Drosophila* has about 13,700 genes, there could be either only a few genes affecting segmentation or so many that the pattern would be impossible to discern.
- Second, mutations that affect segmentation are likely to be **embryonic lethals**, leading to death at the embryonic or larval stage.
  - Flies with embryonic lethal mutations never reproduce, and cannot be bred for study.
  - Nüsslein-Volhard and Wieschaus focused on recessive mutations that could be propagated in heterozygous flies.
- Third, because of maternal effects on axis formation in the egg, the researchers also needed to study maternal genes.
- After exposing flies to mutagenic chemicals, Nüsslein-Volhard and Wieschaus looked for dead embryos and larvae with abnormal segmentation.
  - Through appropriate crosses, they found heterozygotes carrying embryonic lethal mutations.
- Nüsslein-Volhard and Wieschaus identified 1,200 genes essential for embryonic development.
  - About 120 of these were essential for normal segmentation.
- The researchers grouped the genes by general function, mapped them, and cloned many of them.
- In 1995, Nüsslein-Volhard, Wieschaus, and Lewis were awarded a Nobel Prize.

Gradients of maternal molecules in the early Drosophila embryo control axis formation.

- Cytoplasmic determinants produced under the direction of maternal effect genes are deposited in the unfertilized egg.
- A **maternal effect gene** is a gene that, when mutant in the mother, results in a mutant phenotype in the offspring, regardless of the offspring's own genotype.
  - In fruit fly development, maternal effect genes encode proteins or mRNA that are placed in the egg while it is still in the ovary.
  - When the mother has a mutation in a maternal effect gene, she makes a defective gene product (or none at all) and her eggs will not develop properly when fertilized.
- Maternal effect genes are also called **egg-polarity genes** because they control the orientation of the egg and consequently the fly.
  - One group of genes sets up the anterior-posterior axis, while a second group establishes the dorsal-ventral axis.
- One gene called ***bicoid*** affects the front half of the body.

- An embryo whose mother has a mutant *bicoid* gene lacks the front half of its body and has duplicate posterior structures at both ends.
    - This suggests that the product of the mother's *bicoid* gene is essential for setting up the anterior end of the fly and might be concentrated at the future anterior end.
- This is a specific version of the *morphogen gradient hypothesis*, in which gradients of **morphogens** establish an embryo's axes and other features.
- Using DNA technology and biochemical methods, researchers were able to clone the *bicoid* gene and use it as a probe for *bicoid* mRNA in the egg.
    - As predicted, the *bicoid* mRNA is concentrated at the extreme anterior end of the egg cell.
- After the egg is fertilized, *bicoid* mRNA is transcribed into protein, which diffuses from the anterior end toward the posterior, resulting in a gradient of proteins in the early embryo.
    - Injections of pure *bicoid* mRNA into various regions of early embryos resulted in the formation of anterior structures at the injection sites.
- The *bicoid* research is important for three reasons.
    1. It identified a specific protein required for some of the earliest steps in pattern formation.
    2. It increased our understanding of the mother's role in the development of an embryo.
    3. It demonstrated a key developmental principle: a gradient of molecules can determine polarity and position in the embryo.
- Maternal mRNAs are crucial during development of many species.
    - In *Drosophila*, gradients of specific proteins encoded by maternal mRNAs determine the posterior and anterior ends and establish the dorsal-ventral axis.
- Later, positional information encoded by the embryo's genes establishes a specific number of correctly oriented segments and triggers the formation of each segment's characteristic structures.

Concept 18.5 Cancer results from genetic changes that affect cell cycle control

- Cancer is a set of diseases in which cells escape the control mechanisms that normally regulate cell growth and division.
    - The gene regulation systems that go wrong during cancer are the systems that play important roles in embryonic development and immune response.
- The genes that normally regulate cell growth and division during the cell cycle include genes for growth factors, their receptors, and the intracellular molecules of signaling pathways.
    - Mutations altering any of these genes in somatic cells can lead to cancer.
    - The agent of such changes can be random spontaneous mutations or environmental influences such as chemical carcinogens, X-rays, and some viruses.

Proto-oncogenes can become oncogenes, contributing to the development of cancer.

- Cancer-causing genes, **oncogenes**, were initially discovered in viruses.
    - Close counterparts have been found in the genomes of humans and other animals.å

- Normal versions of cellular genes, called **proto-oncogenes**, code for proteins that stimulate normal cell growth and division.
- A proto-oncogene becomes an oncogene following genetic changes that lead to an increase in the proto-oncogene's protein production or in the intrinsic activity of each protein molecule.
    - These genetic changes include movement of DNA within the genome, amplification of the proto-oncogene, and point mutations in a control element or the proto-oncogene itself.
- Cancer cells frequently have chromosomes that have been broken and rejoined incorrectly.
    - A fragment may be moved to a location near an active promoter or other control element.
- Amplification increases the number of copies of the proto-oncogene in the cell.
- A point mutation in the promoter or enhancer of a proto-oncogene may increase its expression.
- A point mutation in the coding sequence may lead to translation of a protein that is more active or longer-lived.
- All of these mechanisms can lead to abnormal stimulation of the cell cycle, putting the cell on the path to malignancy.

Mutations to tumor-suppressor genes may contribute to cancer.

- The normal products of **tumor-suppressor genes** *inhibit* cell division.
- Some tumor-suppressor proteins normally repair damaged DNA, preventing the accumulation of cancer-causing mutations.
- Other tumor-suppressor proteins control the adhesion of cells to each other or to an extracellular matrix, which is crucial for normal tissues and often absent in cancers.
- Still others are components of cell-signaling pathways that inhibit the cell cycle.
    - Decreases in the normal activity of a tumor-suppressor protein may contribute to cancer.
- The proteins encoded by many proto-oncogenes and tumor-suppressor genes are components of cell-signaling pathways.
- Mutations in the products of two key genes, the *ras* proto-oncogene and the *p53* tumor-suppressor gene, occur in 30% and over 50% of human cancers, respectively.
- The Ras protein, the product of the ***ras* gene**, is a G protein that relays a growth signal from a growth factor receptor on the plasma membrane to a cascade of protein kinases.
    - At the end of the pathway is the synthesis of a protein that stimulates the cell cycle.
- Many *ras* oncogenes have a point mutation that leads to a hyperactive version of the Ras protein that trigger the kinase cascade in the absence of growth factor, resulting in excessive cell division.
- The ***p53* gene**, named for its 53,000-dalton protein product, is a tumor-suppressor gene.
    - The p53 protein is a specific transcription factor for the synthesis of several cell cycle-inhibiting proteins.
    - The *p53* gene has been called the "guardian angel of the genome."
- Once activated by DNA damage, the p53 protein functions as an activator for several genes.
    - The p53 protein can activate the *p21* gene, whose product halts the cell cycle by binding to cyclin-dependent kinases, allowing time for DNA repair.

- p53 also activates expression of a group of miRNAs, which inhibit the cell cycle.
- The p53 protein can also turn on genes directly involved in DNA repair.
- When DNA damage is irreparable, the p53 protein can activate "suicide genes" whose protein products cause cell death by apoptosis.

- A mutation that knocks out the *p53* gene can lead to excessive cell growth and cancer.

Multiple mutations underlie the development of cancer.

- More than one somatic mutation is generally needed to produce the changes characteristic of a full-fledged cancer cell.
- If cancer results from an accumulation of mutations, and if mutations occur throughout life, then the longer we live, the more likely we are to develop cancer.
- Colorectal cancer, with 135,000 new cases and 60,000 deaths in the United States each year, illustrates a multistep cancer path.
  - The first sign is often a polyp, a small benign growth in the colon lining.
  - The cells of the polyp look normal but divide unusually frequently.
  - Through gradual accumulation of mutations that activate oncogenes and knock out tumor-suppressor genes, the polyp can develop into a malignant tumor.
  - A *ras* oncogene and a mutated *p53* tumor-suppressor gene are usually involved.
- About a half dozen DNA changes must occur for a cell to become fully cancerous.
- These changes usually include the appearance of at least one active oncogene and the mutation or loss of several tumor-suppressor genes.
  - Because mutant tumor-suppressor alleles are usually recessive, mutations must knock out *both* alleles.
  - Most oncogenes behave like dominant alleles and require only one mutation.

Cancer can run in families.

- The fact that multiple genetic changes are required to produce a cancer cell helps explain the predispositions to cancer that run in families.
  - An individual inheriting an oncogene or a mutant allele of a tumor-suppressor gene is one step closer to accumulating the necessary mutations for cancer to develop.
- Geneticists are devoting much effort to finding inherited cancer alleles so that a predisposition to certain cancers can be detected early in life.
- About 15% of colorectal cancers involve inherited mutations.
- Many of these mutations affect the tumor-suppressor gene *adenomatous polyposis coli* or *APC*.
  - Normal functions of the *APC* gene include regulation of cell migration and adhesion.
  - Even in patients with no family history of the disease, *APC* is mutated in about 60% of colorectal cancers.
- Between 5% and 10% of breast cancer cases show an inherited predisposition.
  - Breast cancer is the second most common type of cancer in the United States, annually striking more than 180,000 women and leading to 40,000 deaths.

- Mutations in one gene, *BRCA1*, increase the risk of breast and ovarian cancer.
    - Mutations in *BRCA1* and the related gene *BRCA2* are found in at least half of inherited breast cancers.
- A woman who inherits one mutant *BRCA1* allele has a 60% probability of developing breast cancer before age 50 (versus a 2% probability in an individual with two normal alleles).
    - Both *BRCA1* and *BRCA2* are considered tumor-suppressor genes because their wild-type alleles protect against breast cancer and their mutant alleles are recessive.
- BRCA1 and BRCA2 proteins function in the cell's DNA damage repair pathway.
    - BRCA2, in association with another protein, helps repair breaks that occur in both strands of DNA.
- Because DNA breakage can contribute to cancer, the risk of cancer can be lowered by minimizing exposure to DNA-damaging agents, such as ultraviolet radiation in sunlight and the chemicals found in cigarette smoke.
- In addition to mutations and other genetic alterations, a number of tumor viruses can cause cancer in various animals, including humans.
    - In 1911, Peyton Rous, an American pathologist, discovered a virus that causes cancer in chickens.
    - The Epstein-Barr virus, which causes infectious mononucleosis, has been linked to several types of cancer in humans, notably Burkitt's lymphoma.
    - Papillomaviruses are associated with cancer of the cervix, and a virus called HTLV-1 causes a type of adult leukemia.
- Worldwide, viruses seem to play a role in about 15% of the cases of human cancer.
- Viruses can interfere with gene regulation in several ways if they integrate their genetic material into a cell's DNA.
    - Viral integration may donate an oncogene to the cell, disrupt a tumor-suppressor gene, or convert a proto-oncogene to an oncogene.
    - Some viruses produce proteins that inactivate p53 and other tumor-suppressor proteins, making the cell more likely to become cancerous.