Genome

All the genes and other DNA of an organism

- Each cell of the organism has its own copy of the organism's entire genome

Differential gene expression

The ability of cells to regulate gene expression (to express some genes in their genome but not others)

- Example: Cell type-specific gene expression
  - Expressed only in certain types of cells (for example, muscle cell-specific genes, liver cell specific genes, neuron cell specific genes, etc.)
  - Cells differentiate (become specialized as specific types of cells) early in the embryo stage
- Example: Developmentally regulated genes
  - Expressed only at certain times in the organism's lifetime
- Example: Environmentally regulated genes
  - Expressed only in the presence of certain molecules or certain changes in the organism's surroundings

Controlling the initiation of transcription is the major way cells regulate differential gene expression

- Transcription initiates when RNA Polymerase II becomes attached to the gene's promoter

  √ RNA Polymersase II cannot attach to the promoter without help from several transcription factor proteins that assemble at the promoter

- Some the transcription factors bind to specific DNA sequences (such as the TATA box) in the gene's promoter

  √ The transcription factors that bind sequences in the promoter are necessary but not sufficient to attach RNA Polymerase II to the gene

- Activators (transcription factors that bind to control elements (specific DNA sequences outside the promoter)) are the true regulators of differential gene expression

Control elements

DNA sequences outside the promoter region that activator proteins bind to

- Enhancer = A cluster of control elements

  √ Enhancers can be found far upstream or far downstream from the promoter

- There are many different activator proteins

  √ Each activator has a specific control element DNA sequence that it binds to

- When activators bind to a gene's enhancer, the DNA folds in a way that brings the enhancer (and the activators bound to it) close to the promoter region

  √ The activators are then able to help RNA Polymerase II attach to the gene

Cells have different types and amounts of activator proteins; This is what causes cells to have differential gene expression

- Some activators are only present in certain cell types

  √ This causes cell-specific expression of genes

- Some activators only function when activated by a signal transduction pathway

  √ This causes developmental and environmental specific gene expression

Repressors

Proteins that decrease transcription of a gene

- Some repressors bind to control elements
- Other repressors decrease transcription by changing the gene's region of the chromosome from euchromatin to heterochromatin by modifying the histone proteins

Euchromatin

The chromatin that is very loosely wrapped around the histone proteins

- Transcription proteins can easily bind to genes on euchromatin

Heterochromatin

The chromatin that is tightly bound to the histone proteins

- The tightly wound DNA blocks transcription proteins from binding, so genes in heterochromatin are not expressed
- By modifying the histones, enzymes can change euchomatin to heterochromatin and *visa versa*

Methylation of DNA

Adding a methyl (CH_3) functional group to the nucleotides of a gene inhibits expression of that gene

Although transcription is the major way cells regulate genes, there are other stages where expression can be regulated. Each of the following events is carried out by specific enzymes. By changing the level of those enzymes, the cell can regulate gene expression at that step.

- Alternative splicing of the pre-mRNA
  - Enzymes control which alternative exons are used

- Blocking translation initiation
  - Proteins that bind to the mRNA can block the ribosome from starting translation

- mRNA degradation
  - Several enzymes in the cytoplasm bind to and destroy mRNAs

- RNA interference
  - The cell can inhibit expression of any specific mRNA by producing small RNA molecules (called miRNA (microRNA) or siRNA (small interfering RNA)) that are complementary the mRNA

  - The interfering RNAs attach to blocking/degrading proteins, then bind to the target mRNA by complementary base pairing. This leads to target specific mRNA degradation or translation blocking

- Protein modification
  - Many proteins are initially made with inhibiting domains. These domains must be cleaved out by enzymes to make the protein functional

  - Many proteins are activated by being phosphorylated by enzymes

  - The protein ubiquitin is added to proteins to mark them for degradation. This system can be used to selectively degrade specific target protein

The genomes of higher eukaryotes (like human beings) contain genes and a larger amount of non-gene DNA

The human genome consists of

- 25% = Genes (including introns, regulatory sequences, and other noncoding parts)

  √ Only 1.5% of genome is protein-coding DNA (exons)

- 2% = Non-functioning genes (gene fragments and mutated genes)

- 5% = Large segment duplications

- 24% = Tandem repeat DNA

- 44% = Transposable elements

Large segment duplications

Large segments (five kilobases or larger) that are present in multiple copies in the genome

- Some large segment duplications are on the same chromosome, some are on different chromosomes

Tandem repeat DNA (also called satellite DNA)

Short simple DNA sequences that are repeated many times in a row

- Example: The sequence GTTAC is repeated thousands of times in a row at many locations in the genome

- Much of the tandem repeat DNA is found at the chromosome's centromere or its telomeres (the tips of the chromosome)

  √ The tandem repeats in the centromere act as binding sites for proteins that hold sister chromosomes together and for the kinetochore proteins (that attach the centromere to spindle fibers during mitosis)

  √ The tandem repeats at the telomeres act as binding sites for proteins that protect the ends of the chromosomes

  √ The tandem repeats at the telomeres also function as disposable DNA because the telomeres become shorter every time the chromosomes are duplicated

Transposable elements (TE's) (also called LINEs)

DNA sequences that move themselves from one location to another in the genome

- The two types of transposable elements are Transposons and Retrotransposons

Transposons

DNA sequences that move to a new location in the genome by means of a DNA intermediate

- The DNA sequence of the transposon is copied (or cut) into a short piece of DNA
- The DNA piece inserts at a random location in the genome

Retrotransposons

DNA sequences that move to a new location in the genome by means of an RNA intermediate

- The DNA sequence of the retrotransposon is transcribed into an mRNA that encodes an enzyme called Reverse Transcriptase
- The Reverse Transcriptase enzyme makes a DNA piece using the its own mRNA as a template
- The DNA piece inserts at a random location in the genome

SINEs

Partial retrotransposons that do not encode reverse transcriptase but that move by using the reverse transcriptase enzyme made by another retrotransposon

- The Alu sequence is the most abundant SINE in the human genome.
  - √ The genome contains over half a million copies of Alu (10% of the entire human genome)

Some facts about transposable elements in humans:

- There are over a million TE's in the human genome

  √ But only about 1 in 2000 TE's in the human genome are active. The inactive ones are inactivated by mutations in their sequence or by repressor proteins

- About 1 in 50 people's have a new TE copy in their genome

  √ The active transposable elements are the most powerful mutagens of our genome

TE's can change the genome

- TE's can alter a gene that they insert into

  √ TE's can make it non-functional if the insertion happens in a coding region of the gene

- They can bring exons or entire genes with them

  √ This duplicates genes or adds new exons to existing genes

- The repressor proteins that have evolved to repress transposable elements can be used to regulate other genes also

Multigene family

A group of identical (or very similar genes) at one location in the genome

- Example: There are two globin polypeptides that make hemoglobin protein. Each polypeptide comes from a multigene family containing several variations of the gene and several pseudogenes (non-transcribed versions of the gene)

- Multigene families originate when an error by the cell duplicates a gene

  √ Errors in crossing over or transposon moving can duplicate genes

- The copy of the gene will change in sequence by random mutations

  √ The altered gene may make a protein with a new function (a new gene) or become functionless (a pseudogene)

Figs 19.14, 19.17, 19.18, and 19.20