Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA $\rightarrow$ RNA)
- The Genetic Code
- Translation (RNA $\rightarrow$ protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA $\rightarrow$ RNA)
- The Genetic Code
- Translation (RNA $\rightarrow$ protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
• What do genes do?
• How do we define a gene?
• Discuss the derivation of the “one gene, one polypeptide” model, tracing the history through Garrod, Beadle and Tatum, and Pauling.
Genes generally are information for making specific proteins

- in connection with the rediscovery of Mendel’s work around the dawn of the 20th century, the idea that genes are responsible for making enzymes was advanced.

- this view was summarized in the classic work *Inborn Errors of Metabolism* (Garrod 1908)

Premise: certain diseases arise from metabolic disorders.
Genes generally are information for making specific proteins

- work by **Beadle and Tatum** in the 1940s refined this concept
  - found mutant genes in the fungus *Neurospora* that each affected a single step in a metabolic pathway
  - developed the “**one gene, one enzyme**” hypothesis

- Follow-up work by Srb and Horowitz illustrated this even more clearly (their work is actually what is presented in your textbook and in the figure here)
Genes generally are information for making specific proteins

- later work by Pauling and others showed that other proteins are also generated genetically

- also, some proteins have multiple subunits encoded by different genes

- this ultimately led to the “one gene, one polypeptide” hypothesis
• What do genes do?
• How do we define a gene?
• Discuss the derivation of the “one gene, one polypeptide” model, tracing the history through Garrod, Beadle and Tatum, and Pauling.
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA → RNA)
- The Genetic Code
- Translation (RNA → protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
• How does RNA differ from DNA structurally?

• What are the structural and functional differences between mRNA, tRNA and rRNA?
RNA (ribonucleic acid)

- RNA serves mainly as an intermediary between the information in DNA and the realization of that information in proteins.
RNA

- RNA has some structural distinctions from DNA
  - typically single-stranded (although often with folds and complex 3D structure)
  - sugar is ribose; thus, RNA polymers are built from ribonucleotides
    - OH at the #2 C on the ribose, vs. deoxyribose in DNA
    - uracil (U) functions in place of T
RNA (ribonucleic acid)

- three main forms of RNA are used: mRNA, tRNA, and rRNA

  - **mRNA** or messenger RNA: copies the actual instructions from the gene
  
  - **tRNA** or transfer RNA: links with amino acids and bring them to the appropriate sites for incorporation in proteins
  
  - **rRNA** or ribosomal RNA: main structural and catalytic components of ribosomes, where proteins are actually produced

- all are synthesized from DNA templates (thus, some genes code for tRNA and rRNA, not protein)
• How does RNA differ from DNA structurally?

• What are the structural and functional differences between mRNA, tRNA and rRNA?
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA → RNA)
- The Genetic Code
- Translation (RNA → protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
• Explain the “central dogma of gene expression”.

• What is the difference between transcription and translation?

• How will you keep these similar-sounding terms clear in your head?
Central Dogma of Gene Expression

\[ \text{DNA} \rightarrow \text{RNA} \rightarrow \text{protein} \]

- the gene is the DNA sequence with instructions for making a product
- the protein (or protein subunit) is the product
Central Dogma of Gene Expression

- **DNA \(\rightarrow\) RNA is transcription**
  - making RNA using directions from a DNA template
  - transcribe = copy in the same language (language used here is base sequence)
Central Dogma of Gene Expression

- **RNA → protein** is translation
  - making a polypeptide chain using directions in mRNA
  - translate = copy into a different language; here the translation is from base sequence to amino acid sequence
Central Dogma of Gene Expression

- there are exceptions to the central dogma
  - some genes are for an RNA final product, such as tRNA and rRNA (note: mRNA is NOT considered a final product)
  - for some viruses use RNA as their genetic material
    - some never use DNA
    - some use the enzyme reverse transcriptase to perform RNA \(\rightarrow\) DNA before then following the central dogma
Central Dogma of Gene Expression

DNA $\rightarrow$ RNA $\rightarrow$ protein
• Explain the “central dogma of gene expression”.

• What is the difference between transcription and translation?

• How will you keep these similar-sounding terms clear in your head?
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- **Transcription (DNA → RNA)**
- The Genetic Code
- Translation (RNA → protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
What three steps must most (perhaps all) biological processes have?
• Describe the events of initiation, elongation, and termination of transcription.

• Be sure to use key terms like upstream, downstream, promoter, etc.
Transcription (DNA → RNA)

- RNA is synthesized as a complementary strand using DNA-dependent RNA polymerases.
- The process is somewhat similar to DNA synthesis, but no primer is needed.
- Bacterial cells each only have one type of RNA polymerase.
- Eukaryotic cells have three major types of RNA polymerase.
  - RNA polymerase I is used in making rRNA.
  - RNA polymerase II is used in making mRNA and some small RNA molecules.
  - RNA polymerase III is used in making tRNA and some small RNA molecules.
Transcription (DNA → RNA)

- only one strand is transcribed, with RNA polymerase using ribonucleotide triphosphates (NTPs) to build a strand in the 5’→3’ direction
  - thus, the DNA is transcribed (copied or read) in the 3’→5’ direction
  - the DNA strand that is read is called the template strand
Transcription (DNA $\rightarrow$ RNA)

- **upstream** means toward the 5’ end of the RNA strand, or toward the 3’ end of the template strand (away from the direction of synthesis)

- **downstream** means toward the 3’ end of the RNA strand, or toward the 5’ end of the template strand
Transcription

- Nucleotide triphosphates are added to the growing strand at the 3’ end

- Phosphodiester bonds are made by DNA dependent RNA polymerases
  - Two phosphates are lost from each nucleotide triphosphate

- Note the antiparallel, complementary strands
Complementary Coding

- If the template DNA is:
  \[
  \]

- The transcribed mRNA is:
  \[
  \]
Transcription (DNA $\rightarrow$ RNA)

transcription has three stages:

- initiation
- elongation
- termination
Transcription (DNA $\rightarrow$ RNA)

- **initiation** requires a **promoter** – site where RNA polymerase initially binds to DNA
  - promoters are important because they are needed to allow RNA synthesis to begin
  - promoter sequence is upstream of where RNA strand production actually begins
  - promoters vary between genes; this is the main means for controlling which genes are transcribed at a given time

![Diagram of transcription showing promoter, transcribed DNA sense strand, and mRNA transcript](image-url)
Transcription (DNA $\rightarrow$ RNA)

- bacterial promoters
  - about 40 nucleotides long
  - positioned just before the point where transcription begins
  - recognized directly by RNA polymerase
Transcription (DNA $\rightarrow$ RNA)

- eukaryotic promoters (for genes that use RNA polymerase II)
  - initially, transcription factors bind to the promoter; these proteins facilitate binding of RNA polymerase to the site
  - transcription initiation complex
    - completed assembly of transcription factors and RNA polymerase at the promoter region
    - allows initiation of transcription (the actual production of an RNA strand complementary to the DNA template)
Transcription
(DNA $\rightarrow$ RNA)

- eukaryotic promoters (for genes that use RNA polymerase II)

- genes that use RNA polymerase II commonly have a "**TATA box**" about 25 nucleotides upstream of the point where transcription begins

- actual sequence is something similar to TATAAA on the non-template strand

- sequences are usually written in the 5’$\rightarrow$3’ direction of the strand with that sequence unless noted otherwise
Transcription (DNA \(\rightarrow\) RNA)

- regardless of promoter specifics, initiation begins when RNA polymerase is associated with the DNA
  - RNA polymerase opens and unwinds the DNA
  - RNA polymerase begins building an RNA strand in the 5’\(\rightarrow\)3’ direction, complementary to the template strand
  - only one RNA strand is produced
**Transcription (DNA → RNA)**

- **elongation**
  - transcription continues in a linear fashion, with DNA unwinding and opening along the way
  - the newly synthesized RNA strand easily separates from the DNA and the DNA molecule “zips up” behind RNA polymerase, reforming the double helix
Transcription (DNA $\rightarrow$ RNA)

- termination: the end of RNA transcription
  - in prokaryotes, transcription continues until a **terminator sequence** is transcribed – usually a GC hairpin or something similar
  - that terminator sequence (now in RNA) causes RNA polymerase to release the RNA strand and release from the DNA
Transcription (DNA $\rightarrow$ RNA)

- **termination**: the end of RNA transcription

  - termination in eukaryotes is more complicated and differs for different RNA polymerases
  
  - still always requires some specific sequence to be transcribed
  
  - for RNA pol II the specific sequence is usually hundreds of bases before the actual ending site
The Template Strand Codes mRNA

- First one, and then the other, DNA strand can be the template (coding, or sense) strand for different genes.
• Describe the events of initiation, elongation, and termination of transcription.

• Be sure to use key terms like upstream, downstream, promoter, etc.
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA → RNA)
- The Genetic Code
- Translation (RNA → protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
• What is a codon?

• What is the genetic code?

• Why are the “words” in the genetic code three bases long?
The genetic code

- the actual information for making proteins is called the **genetic code**

- the genetic code is based on **codons**: sequences of three bases that instruct for the addition of a particular amino acid (or a stop)
  
  - codons are thus read in sequences of 3 bases on mRNA, sometimes called the **triplet code**
  
  - codons are always written in 5’→3’ fashion
  
  - four base types allow $4^3 = 64$ combinations, plenty to code for the 20 amino acids typically used to build proteins
<table>
<thead>
<tr>
<th>First mRNA base (5' end)</th>
<th>Second mRNA base</th>
<th>Third mRNA base (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>U</td>
<td>UG</td>
</tr>
<tr>
<td>UC</td>
<td>UCC</td>
<td>GC</td>
</tr>
<tr>
<td>UCA</td>
<td>UCA</td>
<td>CG</td>
</tr>
<tr>
<td>UCG</td>
<td>UCA</td>
<td>CG</td>
</tr>
<tr>
<td>CU</td>
<td>C</td>
<td>CG</td>
</tr>
<tr>
<td>CCA</td>
<td>C</td>
<td>CG</td>
</tr>
<tr>
<td>CCG</td>
<td>C</td>
<td>CG</td>
</tr>
<tr>
<td>CA</td>
<td>C</td>
<td>CG</td>
</tr>
<tr>
<td>CAG</td>
<td>C</td>
<td>CG</td>
</tr>
<tr>
<td>AC</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>ACA</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>ACG</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>AU</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>AUC</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>ACAA</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>AAG</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>AUG</td>
<td>AUG</td>
<td>AG</td>
</tr>
<tr>
<td>AGU</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>AGC</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>AGA</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>AGG</td>
<td>A</td>
<td>AG</td>
</tr>
<tr>
<td>GU</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GUC</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GUA</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GUG</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GUU</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GCC</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GCA</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GCG</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GCU</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GGU</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GGC</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GGA</td>
<td>G</td>
<td>GC</td>
</tr>
<tr>
<td>GGG</td>
<td>G</td>
<td>GC</td>
</tr>
</tbody>
</table>

**Notes:**
- **Phe:** Phenylalanine
- **Ser:** Serine
- **Tyr:** Tyrosine
- **Cys:** Cysteine
- **UAA Stop:** Stop codon
- **UGA Stop:** Stop codon
- **UAG Stop:** Stop codon
- **Leu:** Leucine
- **Pro:** Proline
- **His:** Histidine
- **CGU:** Cysteine
- **Arg:** Arginine
- **Ile:** Isoleucine
- **Thr:** Threonine
- **Asn:** Asparagine
- **Ser:** Serine
- **Val:** Valine
- **Ala:** Alanine
- **Asp:** Aspartic acid
- **Lys:** Lysine
- **Met or start:** Start codon
- **Gly:** Glycine

**Start codons:** AUG, GUG
### The Genetic Code

<table>
<thead>
<tr>
<th>First Position (5' End)</th>
<th>Second Position</th>
<th>Third Position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>u</td>
<td>uuc, uuc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>uaa, uag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phenylalanine, Leucine</td>
</tr>
<tr>
<td>c</td>
<td>u</td>
<td>ucc, uca</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serine</td>
</tr>
<tr>
<td>g</td>
<td>u</td>
<td>ugc, uga</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tyrosine, (stop), (stop)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cysteine, (stop), Tryptophan</td>
</tr>
<tr>
<td>c</td>
<td>u</td>
<td>ccc, cca, cag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leucine, Proline</td>
</tr>
<tr>
<td>a</td>
<td>ca</td>
<td>cac, caa, cag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Histidine, Glutamine</td>
</tr>
<tr>
<td>g</td>
<td>cg</td>
<td>cgc, cga, cgg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arginine</td>
</tr>
<tr>
<td>a</td>
<td>au</td>
<td>auc, aua, aug</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isoleucine, Methionine</td>
</tr>
<tr>
<td>c</td>
<td>ac</td>
<td>acc, aca, aag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Threonine</td>
</tr>
<tr>
<td>a</td>
<td>aa</td>
<td>aac, aaa, aag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asparagine, Lysine</td>
</tr>
<tr>
<td>g</td>
<td>ag</td>
<td>agc, aga, agg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serine, Arginine</td>
</tr>
<tr>
<td>g</td>
<td>gu</td>
<td>guc, gua, gug</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Valine</td>
</tr>
<tr>
<td>c</td>
<td>gc</td>
<td>gcc, gca, ggg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alanine</td>
</tr>
<tr>
<td>a</td>
<td>ga</td>
<td>gac, gaa, gag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aspartic acid, Glutamic acid</td>
</tr>
<tr>
<td>g</td>
<td>gg</td>
<td>ggc, gga, ggg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glycine</td>
</tr>
</tbody>
</table>

- Don’t try to memorize the complete genetic code.
- Do know that the code is **degenerate** or **redundant**: some amino acids are coded for by more than one codon (some have only one, some as many as 6).
- Know that AUG is the “start” codon: all proteins will begin with methionine, coded by AUG.
- Know about the **stop codons** that do not code for an amino acid but instead will end the protein chain.
- Be able to use the table to “read” an mRNA sequence.
The genetic code

- The genetic code was worked out using artificial mRNAs of known sequence
  - The first “word” was determined by Nirenberg using poly-uracil RNA. Just a long string of U’s:
    
    5' - U - U - U - U - U - U - U - U - U - U - U - U - 3'
  
- When the polyU-RNA was added to a mixture of ribosomes, the resulting polypeptide was all phenylalanines: a long string of Phe’s
  
  Phe-Phe-Phe-Phe-Phe-Phe-Phe-Phe-Phe-Phe

- Thus UUU codes for Phe

- The complete genetic code was worked out by 1967
The genetic code

- the reading of the code 3 bases at a time establishes a reading frame; thus, AUG is very important as the first codon establishes the reading frame

- the genetic code is nearly universal – all organisms use essentially the same genetic code (strong evidence for a common ancestry among all living organisms; allows most of what is done in "genetic engineering")
• What is a codon?

• What is the genetic code?

• Why are the “words” in the genetic code three bases long?
Diagram a mature mRNA.
mRNA coding region

- each mRNA strand thus has a **coding region** within it that codes for protein synthesis
- the coding region starts with the AUG start, and continues with the established reading frame
- the coding region ends when a **stop codon** is reached
- the mRNA strand prior to the start codon is called the *5’ untranslated region* or **leader sequence**
- the mRNA strand after the stop codon is called the *3’ untranslated region* or **trailing sequence**
- collectively, the leader sequence and trailing sequence are referred to as noncoding regions of the mRNA
• Diagram a mature mRNA.
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA → RNA)
- The Genetic Code
- Translation (RNA → protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
• Describe the events of initiation, elongation, and termination of translation.

• Be sure to use key terms like ribosome, ribozyme, anticodon, activated tRNA, EPA sites, translocation, termination factor, etc. Also, be sure to note:
  – how the reading frame is established
  – the direction of reading mRNA (5’ and 3’ ends)
  – the direction of protein synthesis (N- and C-ends)
Prokaryotic and Eukaryotic Gene Expression

- Prokaryotes lack a nucleus; eukaryotes have nuclei. So:
  - Prokaryotes make RNA and protein in cytoplasm.
  - Eukaryotes make RNA in the nucleus, protein in cytoplasm.

<table>
<thead>
<tr>
<th>Prokaryotes</th>
<th>Eukaryotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA → RNA</td>
<td>DNA → RNA</td>
</tr>
<tr>
<td>RNA → Protein</td>
<td>RNA → Protein</td>
</tr>
<tr>
<td>transcription</td>
<td>translation</td>
</tr>
<tr>
<td>translation</td>
<td></td>
</tr>
</tbody>
</table>
Translation (RNA $\rightarrow$ protein)

- The site of translation is the ribosome.
- Ribosomes are complexes of RNA and protein, with two subunits.
- Ribosomes catalyze translation (more on this role later).
Translation (RNA → protein)

- Ultimately, peptide bonds must be created between amino acids to form a polypeptide chain.

- Recall that peptide bonds are between the amino group of one amino acid and the carboxyl group of another.

- The ribosome acts at the ribozyme that catalyzes peptide bond formation.

- Primary polypeptide structure is determined by the sequence of codons in mRNA.
Translation (RNA $\rightarrow$ protein)

- tRNAs bring amino acids to the site of translation

- tRNAs are synthesized at special tRNA genes

- tRNA molecules are strands about 70-80 bases long that form complicated, folded 3-dimensional structures

- tRNAs have attachment sites for amino acids

- each tRNA has an **anticodon** sequence region that will form a proper complementary basepairing with a codon on an mRNA molecule
Translation (RNA \(\rightarrow\) protein)

- tRNA is **linked** to the appropriate amino acid by enzymes called **aminoacyl-tRNA synthetases**

  - the carboxyl group of each specific amino acid is attached to either the 3' OH or 2' OH group of a specific tRNA
  - there is at least one specific aminoacyl-tRNA synthetase for each of the 20 amino acids used in proteins

- ATP is used as an energy source for the reaction

- the resulting complex is an **aminoacyl-tRNA**, also called a **charged tRNA** or **activated tRNA**

- the amino acid added must be the proper one for the anticodon on the tRNA
Translation (RNA \rightarrow protein)

- there are not actually 64 different tRNAs
- three stops have no tRNA
- some tRNAs are able to be used for more than one codon
- for these, the third base allows some “wobble” where basepairing rules aren’t strictly followed
- this accounts for some of the degeneracy in the genetic code
- for note how often the 3rd letter in the codon does not matter in the genetic code
- there are usually only about 45 tRNA types made by most organisms

<table>
<thead>
<tr>
<th>First Position (5' End)</th>
<th>Second Position</th>
<th>Third Position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>u</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>uuu</td>
<td>uuc</td>
</tr>
<tr>
<td></td>
<td>Leucine</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>c</td>
<td>ucc</td>
<td>uca</td>
</tr>
<tr>
<td></td>
<td>Serine</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>uau</td>
<td>uac</td>
</tr>
<tr>
<td></td>
<td>Tyrosine</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>ugu</td>
<td>ugc</td>
</tr>
<tr>
<td></td>
<td>Cysteine</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>cuu</td>
<td>cuu</td>
</tr>
<tr>
<td></td>
<td>Leucine</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>ccu</td>
<td>cca</td>
</tr>
<tr>
<td></td>
<td>Proline</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>cau</td>
<td>cac</td>
</tr>
<tr>
<td></td>
<td>Histidine</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>cgua</td>
<td>cgaa</td>
</tr>
<tr>
<td></td>
<td>Arginine</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>auu</td>
<td>auc</td>
</tr>
<tr>
<td></td>
<td>Isoleucine</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>acu</td>
<td>acc</td>
</tr>
<tr>
<td></td>
<td>Threonine</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>aau</td>
<td>aac</td>
</tr>
<tr>
<td></td>
<td>Asparagine</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>agu</td>
<td>agc</td>
</tr>
<tr>
<td></td>
<td>Serine</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>guu</td>
<td>guc</td>
</tr>
<tr>
<td></td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>gru</td>
<td>gcc</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>gau</td>
<td>gac</td>
</tr>
<tr>
<td></td>
<td>Aspartic acid</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>ggu</td>
<td>ggc</td>
</tr>
<tr>
<td></td>
<td>Glycine</td>
<td></td>
</tr>
</tbody>
</table>
the mRNA and aminoacyl-tRNAs bond at the ribosome for protein synthesis

- The large ribosome subunit has a groove where the small subunit fits.
- mRNA is threaded through the groove.
- The large subunit has depressions where tRNAs can fit.
  - The **E site** is where uncharged tRNA molecules are moved and then released.
  - The **P site** is where the completed part of the polypeptide chain will be attached to tRNA.
  - The **A site** is where the new amino acid will enter on an aminoacyl-tRNA as a polypeptide is made.
the mRNA and aminoacyl-tRNAs bond at the ribosome for protein synthesis

- the tRNAs that bond at these sites basepair with mRNA
  - pairing is anticodon to codon
  - must match to make proper basepairs, A-U or C-G, except for the allowed wobbles at the 3rd base
Translation has three stages:

- initiation,
- elongation, and
- termination

- all three stages have protein “factors” that aid the process
- many events within the first two stages require energy, which is often supplied by GTP (working effectively like ATP)
Translation (RNA → protein): initiation

- **initiation complex** is formed

- begins with the loading of a special **initiator tRNA** onto a small ribosomal subunit
  - the initiator tRNA recognizes the codon AUG, which is the initiation start codon
  - AUG codon codes for the amino acid methionine
  - the initiator tRNA thus is charged with methionine; written as tRNA\(^{\text{Met}}\)
Translation (RNA $\rightarrow$ protein): initiation

- next the small ribosomal subunit binds to an mRNA

- for prokaryotes, at the **ribosome recognition sequence** in the mRNA's leader sequence

- for eukaryotes, at the 5’ end of the mRNA (actually at the 5’ cap, more on that later)

- the initiator tRNA anticodon will then basepair with the start codon
Translation (RNA → protein): initiation

- the large ribosomal subunit then binds
- the initiator tRNA is at the P site
- proteins called **initiation factors** help the small subunit bind to the initiator tRNA and mRNA
- assembly of the initiation complex also requires energy from GTP (eubacteria) or ATP (eukaryotes)
Translation (RNA → protein): elongation

- the aminoacyl-tRNA coding for the next codon in the mRNA then binds to the A site of the ribosome
- has to have proper anticodon-codon basepairs form with the mRNA (again wobble occurs for some)
- the binding step requires energy, supplied by GTP
- proteins called **elongation factors** assist in getting the charged tRNA to bind
Translation (RNA $\rightarrow$ protein): elongation

- the amino group of the amino acid on the tRNA in the A site is then in alignment with the carboxyl group of the amino acid in the P site
  - peptide bond formation can spontaneously occur
  - the peptide bond formation is catalyzed by the ribosome itself, with energy that had been stored in the aminoacyl-tRNA molecule
  - in the process, the amino acid at the P site is released from its tRNA
  - this leaves an unacylated tRNA in the P site, and a tRNA in the A site which now contains the growing peptide chain of the protein
  - notice that protein synthesis proceeds from the amino end of the polypeptide to the carboxyl end (N$\rightarrow$C)
Translation (RNA $\rightarrow$ protein): elongation

- **Translocation** then takes place
  - the ribosome assembles essentially moves three nucleotides along the mRNA
  - the ribosome moves relative to the mRNA: a new codon now sits in the A site
  - the unacylated tRNA is moved from the P site to the E site, where it is released
  - the tRNA-peptide is moved from the A site to the P site
  - the translocation process also requires energy from GTP
  - elongation factor proteins assist with translocation
  - now everything is set up for another elongation step
Translation (RNA $\rightarrow$ protein): elongation

- note again that polypeptides are synthesized on ribosomes starting at the amino terminal end and proceeding to the carboxy terminal end (N$\rightarrow$C)

- note also that mRNA's are made from their 5' end to their 3' end, and they are also translated from their 5' end to their 3' end (5'$\rightarrow$3')
Translation (RNA $\rightarrow$ protein): termination

- A stop codon signals the end for translation (UAA, UGA, and UAG are universal stop codons).

- No tRNA matches the stop codon; instead, it's a termination factor (AKA release factor) binds there.

- The termination factor causes everything to dissociate, freeing the polypeptide, mRNA, last tRNA, and ribosomal subunits all from each other (think of the termination factor as a little molecular bomb).
• Describe the events of initiation, elongation, and termination of translation.

• Be sure to use key terms like ribosome, ribozyme, anticodon, activated tRNA, EPA sites, translocation, termination factor, etc. Also, be sure to note:
  – how the reading frame is established
  – the direction of reading mRNA (5’ and 3’ ends)
  – the direction of protein synthesis (N- and C-ends)
• Can mRNAs be used more than once? What are the consequences of this?
Translation (RNA $\rightarrow$ protein)

- for an average-sized polypeptide chain (~300-400 amino acids long) translation takes less than a minute

- **polyribosomes**
  - an mRNA is typically being translated by many ribosomes at the same time
  - typically as many as 20 ribosomes may be synthesizing protein from the same message
  - in prokaryotes, ribosomes initiate and begin elongation even before RNA polymerase ends transcription
    - thus, in prokaryotes transcription and translation are nearly simultaneous
    - that leads to polyribosomes of prokaryotes being closely associated with DNA
  - mRNAs do not stick around forever – they are quickly degraded (as fast as in about 2-5 minutes in most prokaryotes)
• Can mRNAs be used more than once? What are the consequences of this?
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA $\rightarrow$ RNA)
- The Genetic Code
- Translation (RNA $\rightarrow$ protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
What special things are different about eukaryotic mRNA production compare to prokaryotic mRNA production?

Be sure to address key terms such as:

- pre-mRNA
- 5’ cap
- poly-A tail
- RNA splicing
- introns
- exons
Differences between prokaryotes and eukaryotes in transcription and translation

- In eukaryotes, the mRNA is modified before leaving the nucleus.
- The initial transcript is called **precursor mRNA** (or pre-mRNA, or heterogeneous nuclear RNA, or hnRNA).
Differences between prokaryotes and eukaryotes in transcription and translation

- the first modification is 5’ mRNA capping
  - happens early, when eukaryotic mRNAs are just being formed and are 20 - 30 nucleotides long
  - a set of enzymes found in the nucleus adds a **5’ cap** to the message
  - the cap consists of a modified guanine residue, called 7-methylguanylate
  - this cap is required for binding to eukaryotic ribosomes (so an uncapped mRNA cannot be translated in eukaryotes)
  - also appears that the cap makes eukaryotic mRNAs less susceptible to degradation and to promote the transport of the mRNA out of the nucleus
Differences between prokaryotes and eukaryotes in transcription and translation

- the 3’ tail: **polyadenylation**
  - a **polyadenylation signal** in the mRNA trailing sequence signals for the addition of a “tail” on the 3’ end of the mRNA
  - the tail is a series of adenines, and is called a **poly-A tail**
  - polyadenylation is the process of putting the tail on
    - enzymes recognize the polyadenylation signal and cut the RNA strand at that site
    - the enzymes then add 100 - 250 adenine ribonucleotides to the mRNA chain
Differences between prokaryotes and eukaryotes in transcription and translation

- The roles of polyadenylation
  - Starting the process leads to termination of transcription
  - May make mRNAs less susceptible to degradation
  - May help get mRNA out of the nucleus
  - May help in initiation of translation
Differences between prokaryotes and eukaryotes in transcription and translation

- interrupted coding sequences: **introns** and **exons**

- the transcript made from the DNA in eukaryotes is often much larger than the final mRNA

- some stretches of bases called **introns** “interrupt” the sequence and must be removed
  - the number of introns varies, from none for some genes up to dozens or more for others
  - different alleles of the same gene may even vary in intron number
  - the regions that will not be removed are called **exons**
Differences between prokaryotes and eukaryotes in transcription and translation

- The process of removing introns is called RNA splicing.
- The signals for splicing are short sequences at the ends of introns.
- Particles called snRNPs associate with the mRNA in a complex called the spliceosome.
  - SnRNPs are made of small RNA molecules and proteins.
  - The spliceosome catalyzes cutting out and removing an intron and joining together the exons.
  - RNAs in some of the snRNPs act as ribozymes in the splicing process.
  - Note that the spliceosome is not always required, but it usually is needed.
• What special things are different about eukaryotic mRNA production compared to prokaryotic mRNA production?

• Be sure to address key terms such as:
  – pre-mRNA
  – 5’ cap
  – poly-A tail
  – RNA splicing
  – introns
  – exons
• How does alternative splicing work?
Why do exons exist?

- in some cases, **alternative RNA splicing** allows one DNA sequence to direct synthesis of two or more different polypeptides (this may be very common in humans)
• How does alternative splicing work?
• How does exon shuffling work?

• Be sure to include the term “domain” in your explanation.
Why do exons exist?

- Exons tend to code for specific **domains** within proteins
  - A domain is a region within the protein that has a specific function
  - Exons with “junk DNA” intron regions between them may be easy to move around and rearrange to make new proteins
  - This leads to the notion that many proteins consist of such functional domains which can be readily shuffled around during evolution to produce new proteins with novel functions

- Such **exon shuffling** does indeed appear to have played a prominent role in evolution in eukaryotes
• How does exon shuffling work?

• Be sure to include the term “domain” in your explanation.
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA → RNA)
- The Genetic Code
- Translation (RNA → protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
What is the modern definition of a gene?
Modern definition of genes

- complications in some scenarios make it necessary to modify the definition of a gene

- a more inclusive definition: a gene is a nucleotide sequence with information for making a final polypeptide or RNA product

- the usual flow of information is still

  DNA $\rightarrow$ RNA $\rightarrow$ polypeptide
What is the modern definition of a gene?
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA $\rightarrow$ RNA)
- The Genetic Code
- Translation (RNA $\rightarrow$ protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
• What are mutations, and how can they be good, bad, or neutral?
Mutations are changes in the DNA sequence

- mutations may occur as accidents during DNA replication, or may be induced by DNA-damaging radiation or chemicals
  - DNA-damage inducers are called mutagens
  - many mutagens increase the likelihood of cancer, and are thus carcinogens
  - some DNA regions are more prone to mutations; they are called mutational hot spots (trinucleotide repeats are one example)
  - organisms have mechanisms to repair damage to DNA and to proofread DNA during replication, but mutations still occur (usually at a very low rate)
  - the mutations that are most likely to lead to genetic changes (for good or bad) are those in the coding regions of genes
What are mutations, and how can they be good, bad, or neutral?
What is the difference between these three types of point mutation:
- silent mutation
- missense mutation
- nonsense mutation

What is a frameshift mutation, and why does it usually have a huge impact?

What are transposons?
Mutations are changes in the DNA sequence

- Mutations that result in the substitution of one base for another are referred to as **point mutations** or base substitution mutations.

  - If the point mutation does not actually cause a change in what amino acid is coded for (thus usually having no effect), it is called a **silent mutation**.

  - If the point mutation causes a change in what amino acid is coded for, it is called a **missense mutation**.

  - If the point mutation results in the formation of a stop codon where an amino acid previously was coded for, it is called a **nonsense mutation**.

- Nonsense mutations result in the premature termination of the protein sequence, and thus an active protein is usually not formed.
Missense Mutation Example: Sickle-Cell Anemia

- missense at 6th codon in hemoglobin β chain (counted after protein processing)
- in DNA a T is replaced with an A; this leads to valine instead of glutamic acid in the protein
- resulting hemoglobin is “sticky” with other hemoglobin chains, crystallizing easily

**Normal hemoglobin β chain**

**DNA:** CAC GTG GAC TGA GGA CTG CTC  
**RNA:** GUG CAC CUG ACU CCU **GAG** GAG-  
**Protein:** val-his-leu-thr-pro-glutamic acid-glutamic acid-

**Sickle cell anemia hemoglobin β chain**

**DNA:** CAC GTG GAC TGA GGA **CAC** CTC  
**RNA:** GUG CAC CUG ACU CCU **GUG** GAG-  
**Protein:** val-his-leu-thr-pro-val-glutamic acid-
Missense Mutation Example: Sickle-Cell Anemia

**Normal Hemoglobin**
- **Primary Structure**: Val 1, His 2, Leu 3, Thr 4, Pro 5, Glu 6, Glu 7, ...
- **Secondary and Tertiary Structures**: α subunit, β subunit
- **Quaternary Structure**: Normal hemoglobin (top view)
- **Function**: Molecules do not associate with one another; each carries oxygen.
- **Red Blood Cell Shape**: Normal cells are full of individual hemoglobin molecules, each carrying oxygen.

**Sickle-Cell Hemoglobin**
- **Primary Structure**: Val 1, His 2, Leu 3, Thr 4, Pro 5, Val 6, Glu 7, ...
- **Secondary and Tertiary Structures**: Exposed hydrophobic region
- **Quaternary Structure**: Sickle-cell hemoglobin
- **Function**: Molecules interact with one another to crystallize into a fiber; capacity to carry oxygen is greatly reduced.
- **Red Blood Cell Shape**: Fibers of abnormal hemoglobin deform cell into sickle shape.
Mutations are changes in the DNA sequence.

**Frameshift mutations** - mutations that shift the reading frame (occur when nucleotides are either added or deleted).
Frameshift Mutations

- Example using English as an analogous system – 2 types possible:
  - **ORIGINAL**: THEMANCANRUNNOW
    - Reads: (THE MAN CAN RUN NOW)
  - **INSERTION mutation**: THEMMTANCANRUNNOW
    - Reads: (THE MTA NCA NRU NNO W)
  - **DELETION mutation**: THEM|NCANRUNNOW
    - Reads: (THE MNC ANR UNN OW) – red bar indicates the removal of A
Mutations are changes in the DNA sequence

- some mutations are caused by pieces of DNA that can jump around the genome
  - such jumping DNA is called a transposon or transposable element
  - transposons exist in both prokaryotes and eukaryotes
  - for most their normal function (if any) is unknown, but some larger ones can provide benefits by moving copies of useful genes with them
• What is the difference between these three types of point mutation:
  – silent mutation
  – missense mutation
  – nonsense mutation

• What is a frameshift mutation, and why does it usually have a huge impact?

• What are transposons?
Chapter 17: Genes and How They Work

- Genes generally are information for making specific proteins
- RNA (ribonucleic acid)
- Overview of Gene Expression
- Transcription (DNA $\rightarrow$ RNA)
- The Genetic Code
- Translation (RNA $\rightarrow$ protein)
- Differences between prokaryotes and eukaryotes in transcription and translation
- Modern Definition of Genes
- Mutations
- Gene Regulation
• Why is regulation of gene expression important?

• How can, for example, a cell in the retina of your eye make different proteins from a cell in your liver when both cells have exactly the same DNA?

• What are constitutive genes, transcription factors, repressors, activators, and enhancers?
Gene Regulation

- Ch. 18
- gene expression is regulated
- regulation allows for different expression under different conditions
  - a given cell type will only express genes appropriate for that cell type
  - gene expression can be changed in response to the environment
  - constitutive genes (housekeeping genes) are constantly transcribed, with little or no regulation
Gene Regulation

- proteins that regulate transcription are called transcription factors
  - transcription factors often bind directly to DNA
  - transcription factors usually are activated or inactivated based on signals
  - signals are some sort of change in the internal environment of the cells
  - signals can be information from the environment (such as hormones), or as simple as running out of a food molecule or having a new food source
Gene Regulation

- most transcription factors associate with **promoters**
  - promoter sequence determines what transcription factions can bind to the promoter to help initiate transcription
  - different promoter sequences allow for differences in expression
- **repressors** – transcription factors that suppress or stop gene expression
- **activators** – transcription factors that either activate ("turn on") gene expression, or that enhance gene expression
Gene Regulation

- sometimes DNA sequences away from the promoter can also affect transcription
  - such sequences can be upstream or downstream of the coding region, or even within the coding region or introns
  - they are usually within a few kilobases of the coding region, and often within a few hundred bases
  - enhancers – DNA regions, often far from the promoter, where activators will bind either directly or indirectly
Gene Regulation

- A given cell type will only express genes appropriate for that cell type.
• Why is regulation of gene expression important?

• How can, for example, a cell in the retina of your eye make different proteins from a cell in your liver when both cells have exactly the same DNA?

• What are constitutive genes, transcription factors, repressors, activators, and enhancers?