Engineering Genetic Circuits

Chris J. Myers

Lecture 1: An Engineers Guide to Genetic Circuits

Francis Crick

DNA makes RNA, RNA makes protein, and proteins make us.

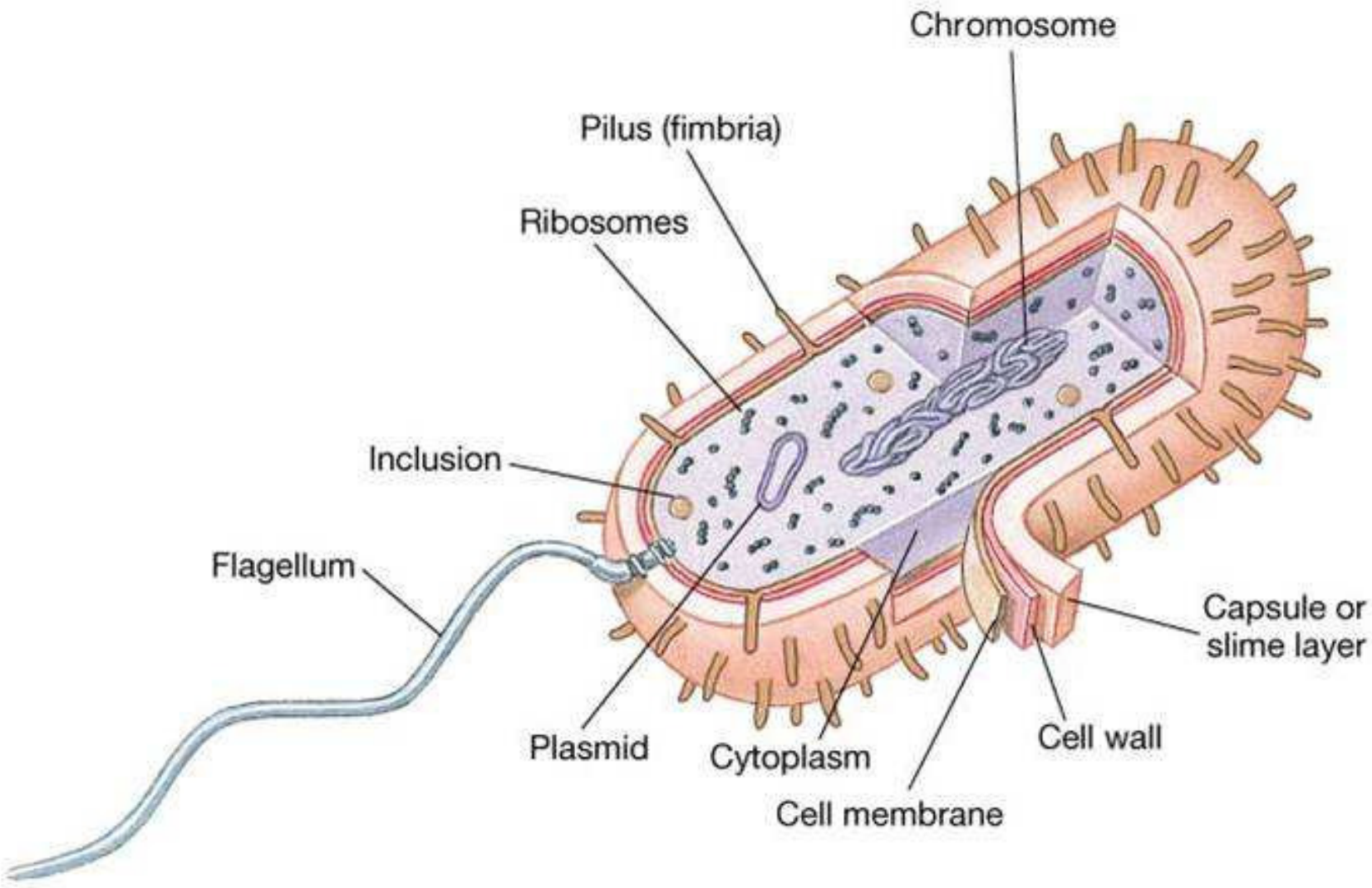
What is a Cell?

- The structural and functional unit of all living organisms.
- Some organisms, such as bacteria, are *unicellular*.
- Other organisms, such as humans, are *multicellular*.
- Humans have an estimated 30 trillion human cells and additional 39 trillion bacteria cells!
- Each cell can take in nutrients, convert these into energy, carry out specialized functions, and reproduce as necessary.
- Each cell stores its own set of instructions in its *genome* for carrying out each of these activities.

Prokaryotic Organisms

- Life arose on earth about 3.5 billion years ago.
- The first types of cells were *prokaryotic* cells.
- They are unicellular organisms that lack a nuclear membrane.
- They do not develop or differentiate into multicellular forms.
- Bacteria are the best known and most studied form.
- Some bacteria grow in masses, but each cell is independent.
- They are capable of inhabiting almost every place on the earth.
- They lack intracellular *organelles* and structures.
- Most functions of organelles are taken over by the *plasma membrane*.

Prokaryotic Features

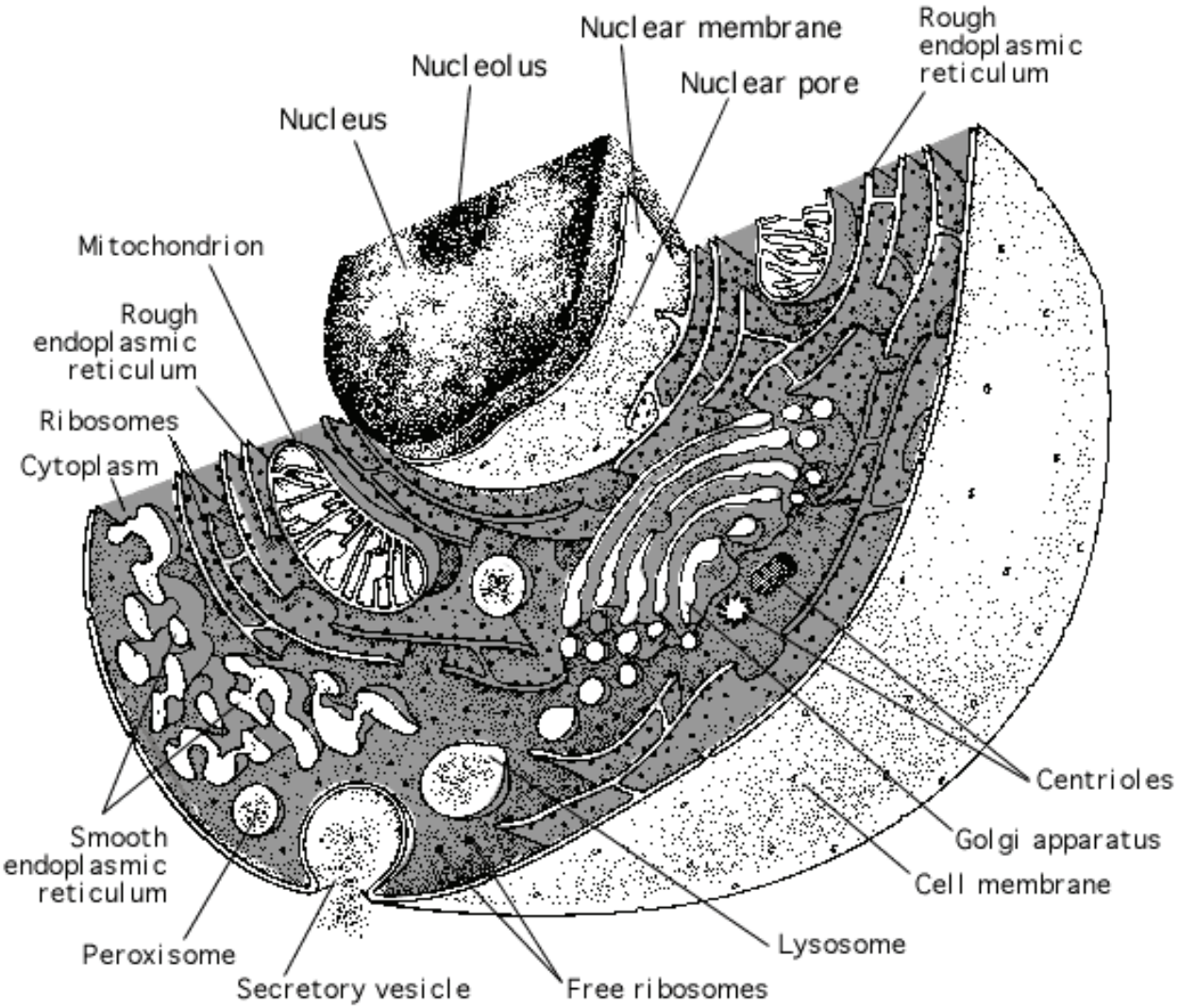


(Courtesy: National Human Genome Research Institute)

Eukaryotic Organisms

- *Eukaryotes* appear in the fossil record about 1.5 billion years ago.
- They include fungi, mammals, birds, fish, invertebrates, mushrooms, plants, and complex single-celled organisms, such as yeast.
- Eukaryotic cells are about 10 times the size of a prokaryote and can be as much as 1000 times greater in volume.
- Use same genetic code and metabolic processes as prokaryotes, but higher level organizational complexity permits multicellular organisms.
- Have membrane-bounded compartments called *organelles*.
- Most important is the *nucleus* that houses the cell's DNA.

Eukaryotic Features



(Courtesy: National Human Genome Research Institute)

The Nucleus: A Cell's Center

- A spheroid membrane-bound region that contains genetic information in long strands of DNA called *chromosomes*.
- Most conspicuous organelle found in a eukaryotic cell.
- Separated from the cytoplasm by *nuclear envelope* which isolates and protects DNA from molecules that could damage its structure or interfere with its processing.
- Where almost all DNA replication and RNA synthesis occurs.
- During processing, DNA is transcribed into mRNA.
- mRNA is transported out of the nucleus, where it is translated into a specific protein molecule.
- In prokaryotes, DNA processing takes place in the cytoplasm.

The Ribosome: The Protein Production Machine

- Found in both prokaryotes and eukaryotes.
- It is a large complex composed of RNAs and proteins.
- They process genetic instructions carried by mRNA.
- *Translation* is the process of converting a mRNA's genetic code into the exact sequence of amino acids that make up a protein.
- Protein synthesis is extremely important, so there are a large number of ribosomes (100s or 1000s) in a cell.
- They float freely in the cytoplasm or sometimes bind to another organelle called the *endoplasmic reticulum*.

Where Do Viruses Fit?

- They are not classified as cells and therefore are neither unicellular nor multicellular organisms.
- They are not “living” because they lack a metabolic system and are dependent on host cells they infect to reproduce.
- They have genomes of either DNA or RNA which are either double-stranded or single-stranded.
- Their genomes code for both proteins to package its genetic material and those needed to reproduce.

Viral Reproduction

- Because viruses are acellular, they must utilize the machinery and metabolism of a host cell to reproduce.
- For this reason, known as *obligate intracellular parasites*.
- Before entering host, it is a *virion*–package of genetic material.
- They can be passed from host to host either through direct contact or through a vector, or carrier.
- *Bacteriophages* attach to the cell wall surface, make a small hole, and inject their DNA into the cell.
- Others (such as HIV) enter the host via *endocytosis*, the process in which a cell takes in material from its environment.
- After entering the cell, its genetic material takes over cell and forces it to produce new viruses.

What are Macromolecules?

- Nearly 70 percent of all living organisms are made up of water.
- Remainder largely *macromolecules* of 1000s of atoms.
- There are four types:
  - *Carbohydrates*
  - *Lipids*
  - *Nucleic acids*
  - *Proteins*

Carbohydrates

- Made up of carbon and water ($C_n(H_2O)_m$ where $m \approx n$).
- Often called sugars.
- An example is *glucose*.
- Important source of chemical energy.
- Powers nearly all processes of a cell.
- Also part of the backbone for DNA and RNA.

Lipids

- Made up mostly of carbon and hydrogen atoms.
- Often have a *hydrophilic* (water-loving) part and a *hydrophobic* (water-fearing) part.
- Primary use is to form *membranes*.
- Membranes separate cells from one another and create compartments within cells as well as having other functions.
- Make good membranes because their hydrophobic parts attract to form *lipid bilayers* where exterior allows water, but interior repels water.
- This allows the lipid bilayers to form between areas containing water, but they do not allow water to easily pass through.
- Examples include fats, oils, and waxes.

Nucleic Acids

- Store information within living organisms.
- Composed of base bound to sugar and *phosphate* molecule.
- Two forms: DNA (*deoxyribonucleic acid*) and RNA (*ribonucleic acid*).
- Sequence of nucleotides encode the instructions to construct proteins.
- Most organisms use DNA, but a few viruses use RNA.

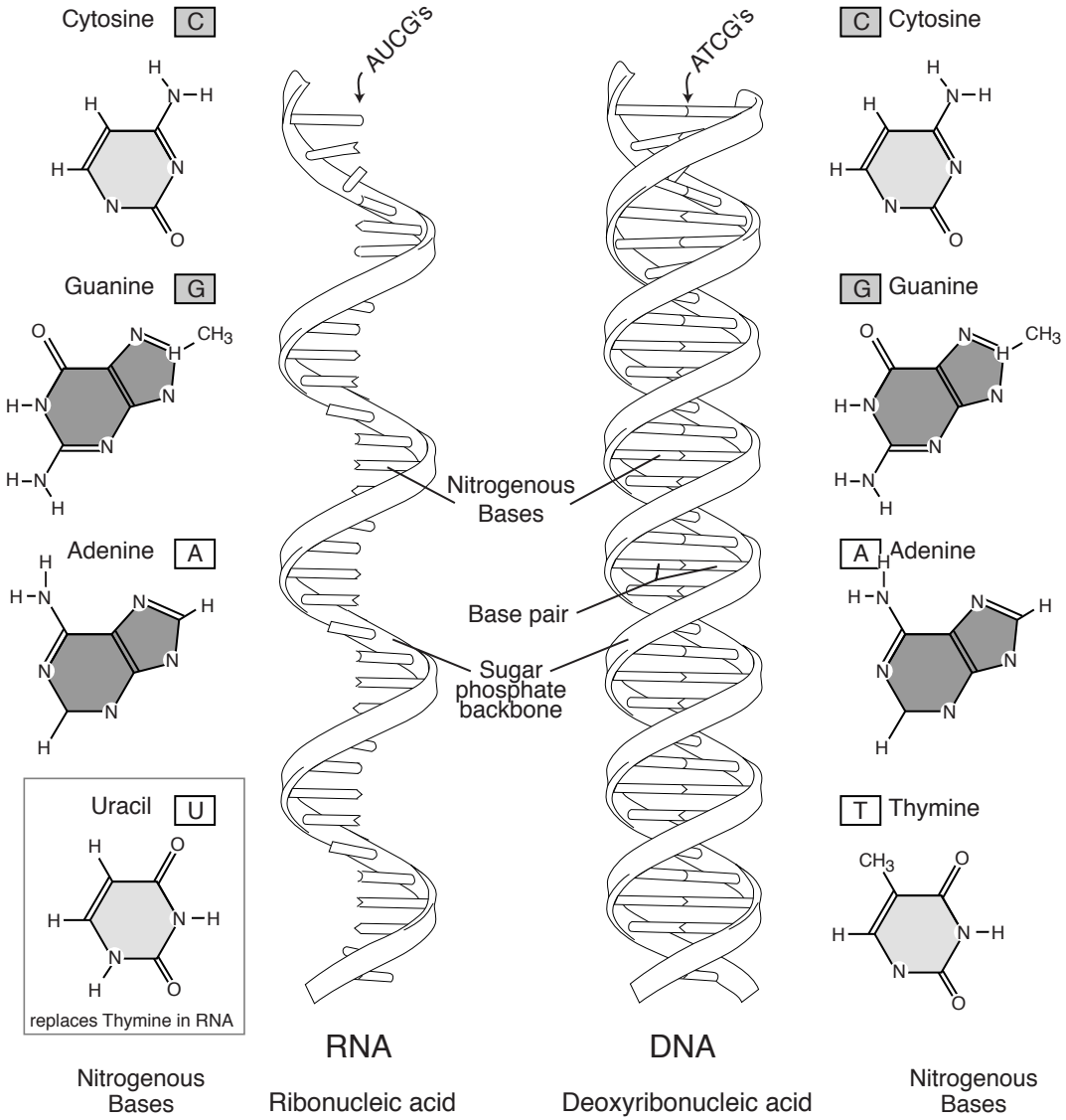
Deoxyribonucleic acid (DNA)

- A *DNA strand* (chain) is made up of four *chemical bases*: *adenine* (A) and *guanine* (G), which are called *purines*, and *cytosine* (C) and *thymine* (T), referred to as *pyrimidines*.
- Each base has slightly different composition of *O*, *C*, *N*, *H*.
- A strand of DNA is always synthesized in the 5' to 3' direction.
- The so-called *5' end* terminates in a 5' phosphate group (-PO4); the *3' end* terminates in a 3' hydroxyl group (-OH).
- DNA is a double-stranded with each strand running in opposite directions.
- *A*-*T* and *G*-*C* base pairs are *complementary*.
- Chemical makeup of this base pairing creates a force that twists the DNA into its coiled double helix structure.
- DNA is readily copied since one strand of DNA can act as a *template* to direct the synthesis of a complementary strand.

Ribonucleic acid (RNA)

- A single-stranded chain of nucleotides with the same 5' to 3' direction.
- Uses a different suger and *uracil* replaces the thymine nucleotide.
- All genes that code for proteins are first made into an RNA strand called a *messenger RNA* (mRNA).
- mRNA carries the information encoded in DNA to the protein assembly machinery, or *ribosome*.
- The ribosome complex uses mRNA as a template to synthesize the exact protein coded for by the gene.
- DNA also codes for *ribosomal RNAs* (rRNAs), *transfer RNAs* (tRNAs), and *small nuclear RNAs* (snRNAs).

Nucleic Acids Diagram



(Courtesy: National Human Genome Research Institute)

Proteins

- Basic building blocks of nearly all the machinery of a cell.
- Each cell contains thousands of different proteins.
- Long chains with as many as 20 kinds of amino acids.
- Genetic code carried by DNA specifies order and number of amino acids and, therefore, shape and function of the protein.
- Code from DNA is transferred to RNA through *transcription*.
- mRNA is *translated* by a *ribosome* into protein.
- mRNA decoded in blocks of three bases, or *codons*.
- Protein built one amino acid at a time, with order determined by the order of the codons in the mRNA.

The Genetic Code

	U	C	A	G
U	UUU Phenylalanine UUC Phenylalanine UUA Leucine UUG Leucine	UCU Serine UCC Serine UCA Serine UCG Serine	UAU Tyrosine UAC Tyrosine UAA Stop UAG Stop	UGU Cysteine UGC Cysteine UGA Stop UGG Tryptophan
C	CUU Leucine CUC Leucine CUA Leucine CUG Leucine	CCU Proline CCC Proline CCA Proline CCG Proline	CAU Histidine CAC Histidine CAA Glutamine CAG Glutamine	CGU Arginine CGC Arginine CGA Arginine CGG Arginine
A	AUU Isoleucine AUC Isoleucine AUA Isoleucine AUG Methionine	ACU Threonine ACC Threonine ACA Threonine ACG Threonine	AAU Asparagine AAC Asparagine AAA Lysine AAG Lysine	AGU Serineine AGC Serineine AGA Arginine AGG Arginine
G	GUU Valine GUC Valine GUA Valine GUG Valine	GCU Alanine GCC Alanine GCA Alanine GCG Alanine	GAU Aspartate GAC Aspartate GAA Glutamate GAG Glutamate	GGU Glycine GGC Glycine GGA Glycine GGG Glycine

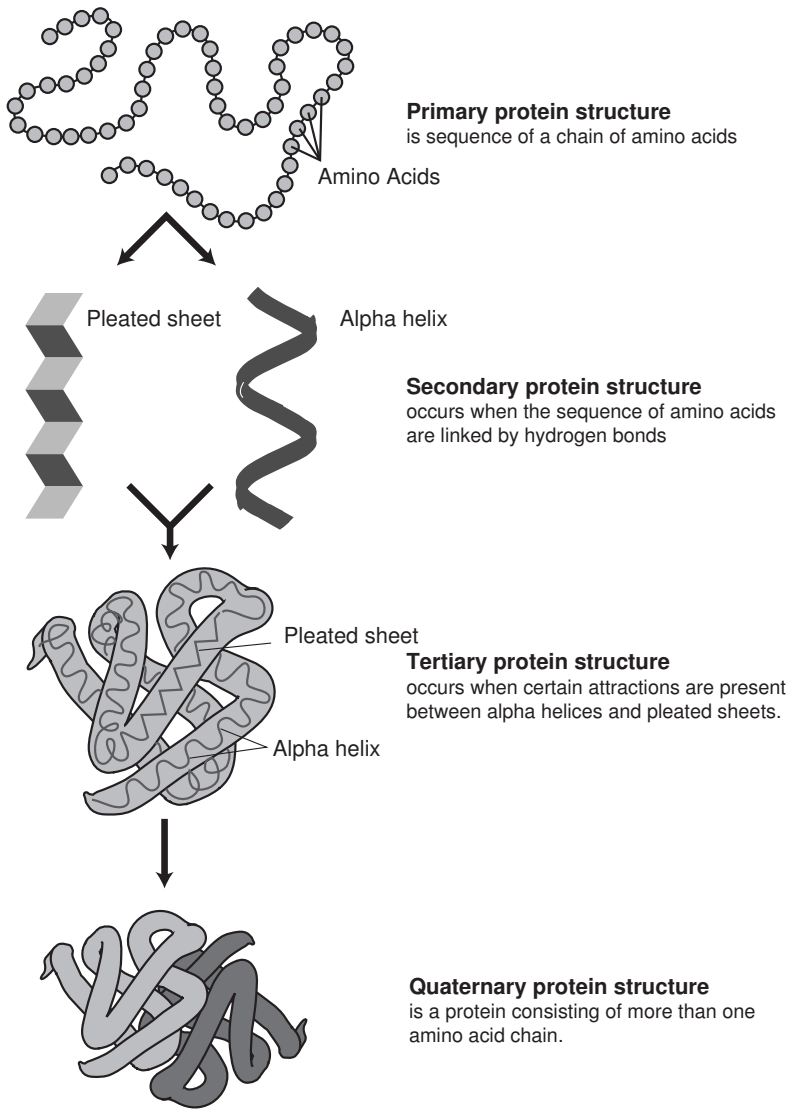
Proteins (cont)

- In 1961, Nirenberg and Matthaei correlated the first codon (UUU) with the amino acid phenylalanine.
- A given amino acid can have more than one codon.
- These redundant codons usually differ at the third position.
- Serine is encoded by UCU, UCC, UCA, and/or UCG.
- Redundancy is key to accommodating mutations that occur naturally as DNA is replicated and new cells are produced.
- Some codons do not code for an amino acid at all but instruct the ribosome when to stop adding new amino acids.

Protein Structure

- Protein folds into a specific 3-dimensional configuration.
- Shape and position of the amino acids in this folded state determines the function of the protein.
- Understanding and predicting *protein folding* is an important area of research.
- The structure of a protein is described in four levels.
  - *Primary structure* - sequence of amino acids.
  - *Secondary structure* - patterns formed by amino acids that are close (ex. α*-helicies* and β*-pleated sheets*).
  - *Ternary structure* - arrangement of far apart amino acids.
  - *Quaternary structure* - arrangement of proteins that are composed of multiple amino acid chains.

Protein Structure

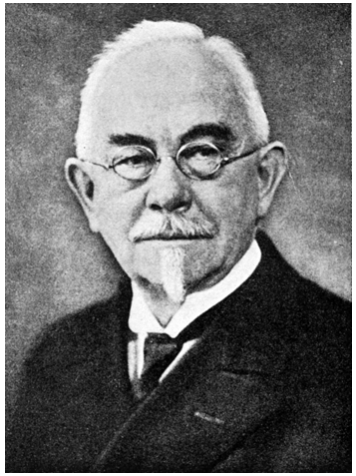


(Courtesy: National Human Genome Research Institute)

What is a Genome?

- All of the 30 million types of organisms use the same basic materials and mechanisms to produce building blocks necessary for life.
- Information encoded in the DNA within its *genome* is used to produce RNA which produces proteins.
- A genome is divided into genes where each gene encodes the information necessary for constructing a protein.
- Some also control the production of proteins by other genes.

What are Genes?



Wilhelm Johanssen

Gregor Mendel

- In 1909, Danish botanist Wilhelm Johanssen coined the word *gene* for the hereditary unit found on a chromosome.
- 50 years earlier, Gregor Mendel, an Austrian monk, characterized hereditary units as *factors*–differences passed from parent to offspring.
- Mendel experimented with his pea plants in the monastery gardens using a pair of clippers to manipulate their parentage.
- Discovery went largely ignored for nearly 50 years until three researchers essentially duplicated his results.

Where are Genes?


James Watson


Francis Crick


Maurice Wilkins


Rosalind Franklin

- Until 1953, it was not known for certain that genes are made of DNA.
- In 1953, Watson and Crick, with support from x-ray data from Franklin and Wilkens, discovered the double helix structure of DNA.
- This discovery showed that DNA is composed of two strands composed of complementary bases.
- This base pairing idea shed light on how DNA could encode genetic information and be readily duplicated during cell division.
- Between 1953 and 1965, work by Crick and others showed how the DNA codes for amino acids and thus proteins.

How Many Genes Do Humans Have?

- In February 2001, two largely independent draft versions of the human genome were published in Nature.
- Both estimated between 30,000 to 40,000 genes in the human genome (today's estimate is between 19,000 and 20,000).
- How do scientists estimate the number of genes in a genome?
  - *Open reading frames*, a 100 bases without a stop codon;
  - *Start codons* such as ATG;
  - Specific sequences found at *splice junctions*; and
  - *Gene regulatory sequences*.
- When complete mRNA sequences known, software can align start and end sequences with the DNA sequence.

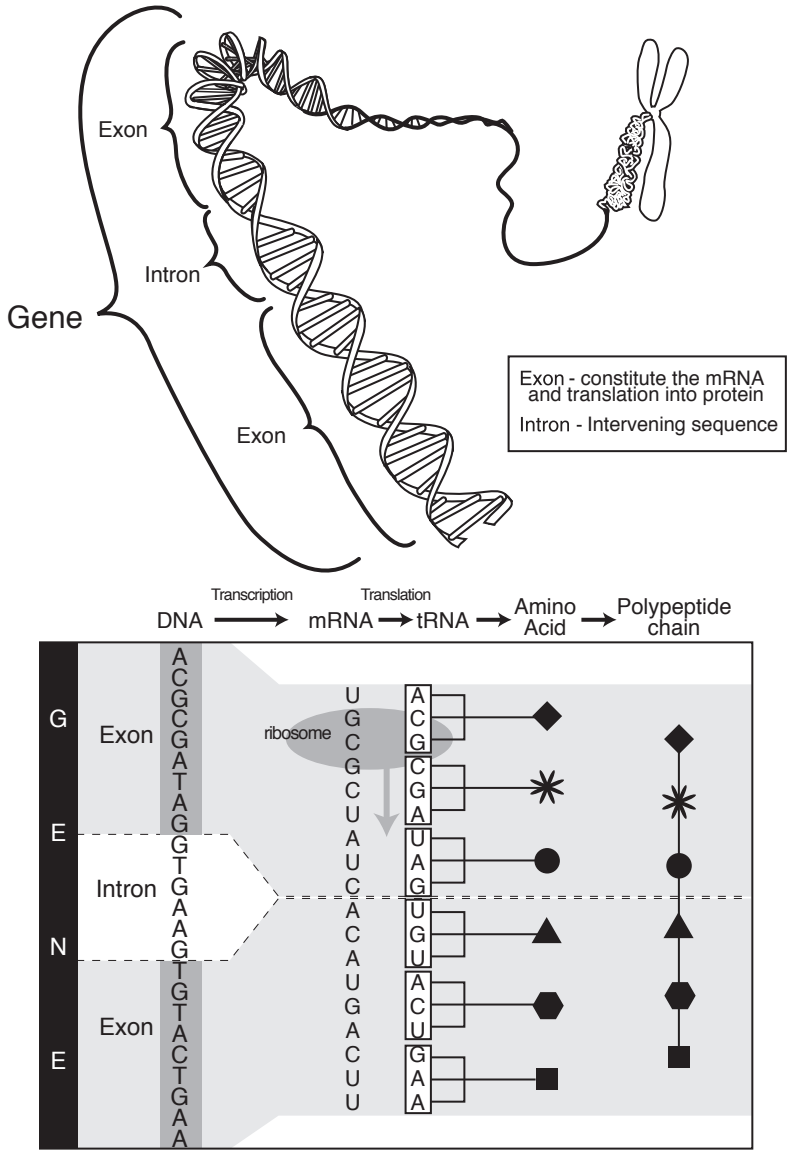
What is Contained in Our Genome?

- *Coding sequences* are sequences that code for proteins.
- Over 98% of our genome do not code for proteins.
- Includes some genes for *non-coding RNA* (ncRNA).
- *Regulatory sequences* are the start/end of genes, sites for initiating replication/recombination, or sites to turn genes on/off.
- *Pseudogenes* are believed to be remnants of real genes that have suffered mutations and are no longer functional.
- *Repetitive DNA*, short sequences repeated 100s of times, make up about 50 percent of our genome.
- *Transposons*, mobile genetic elements that can replicate and insert themselves at other locations, may account for as much as half the DNA.

Introns and Exons

- A eukaryotic gene is not found in a continuous stretch.
- The coding portions of a gene, called *exons*, are interrupted by intervening sequences, called *introns*.
- Both exons and introns are transcribed into mRNA, but before being transported to the ribosome, the mRNA transcript is edited.
- Removes introns, joins exons together, and adds unique features to end of transcript to make a "*mature*" mRNA.
- It is still unclear what all the functions of introns are, but may serve as the site for recombination.

Introns and Exons



(courtesy: National Human Genome Research Institute)

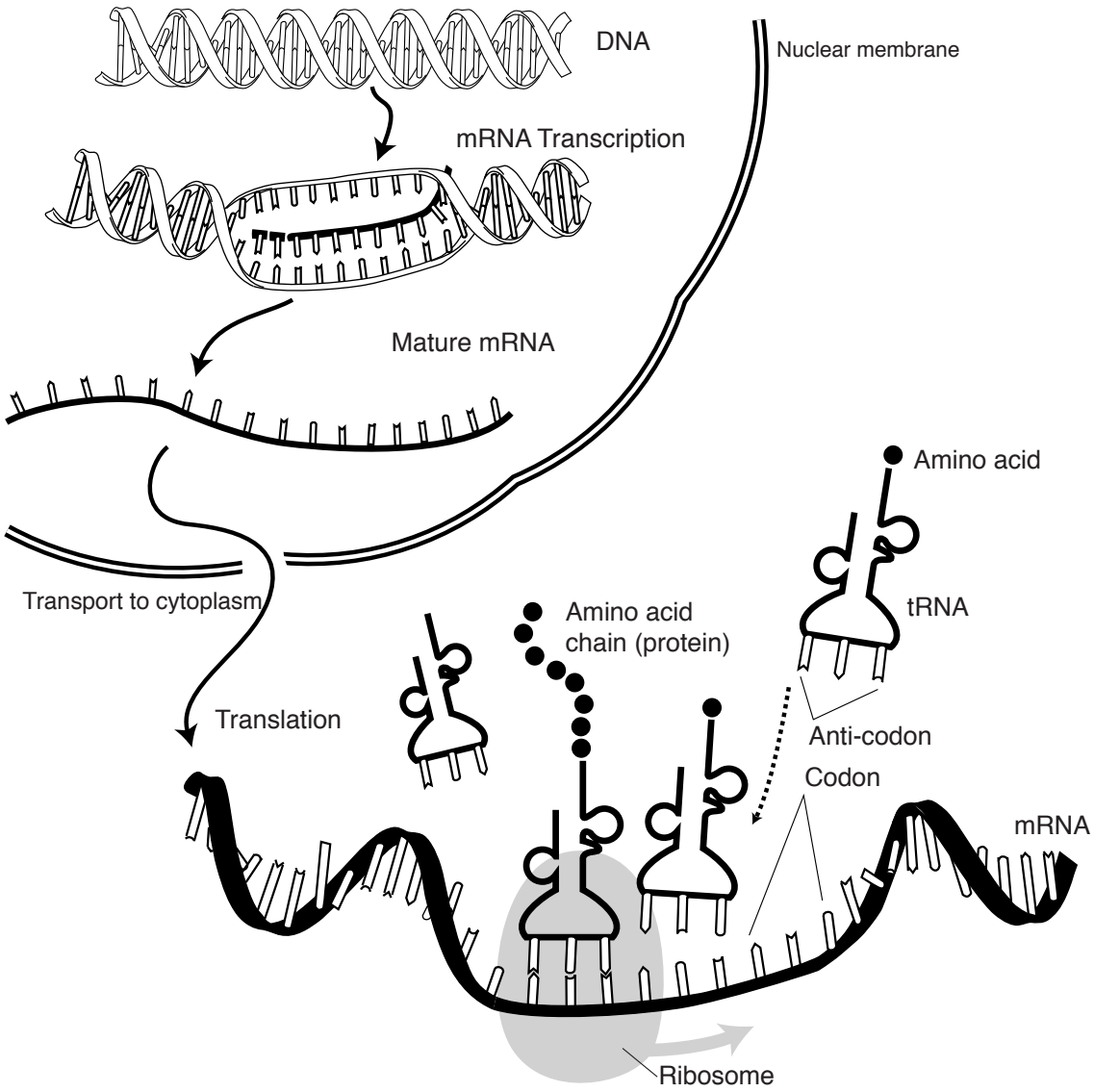
One Gene–One Protein?

- About 40 percent of the expressed genome is *alternatively spliced* to produce multiple proteins from a single gene.
- This process may have evolved to limit effects of mutations.
- Genetic mutations occur randomly, and the effect of a small number of mutations on a single gene may be minimal.
- However, an individual having many genes each with small changes could weaken the individual, and thus the species.
- If single mutation affects several alternate transcripts, it is likely that the individual will not survive.

Genetic Circuits

- Genes encoded in DNA used as templates to synthesize mRNA through the process of transcription.
- Genes include *coding sequences* and *regulatory sequences*.
- Regulatory sequences can bind to other proteins which in turn either activate or repress transcription.
- Transcription is also regulated through *post-transcriptional modifications*, *DNA folding*, and other feedback mechanisms.
- This regulatory network increases an organism's complexity.
- Behavior analogous to electrical circuits in which multiple inputs are processed to determine multiple outputs.
- Therefore, these regulatory networks known as *genetic circuits*.

Overview of Transcription and Translation

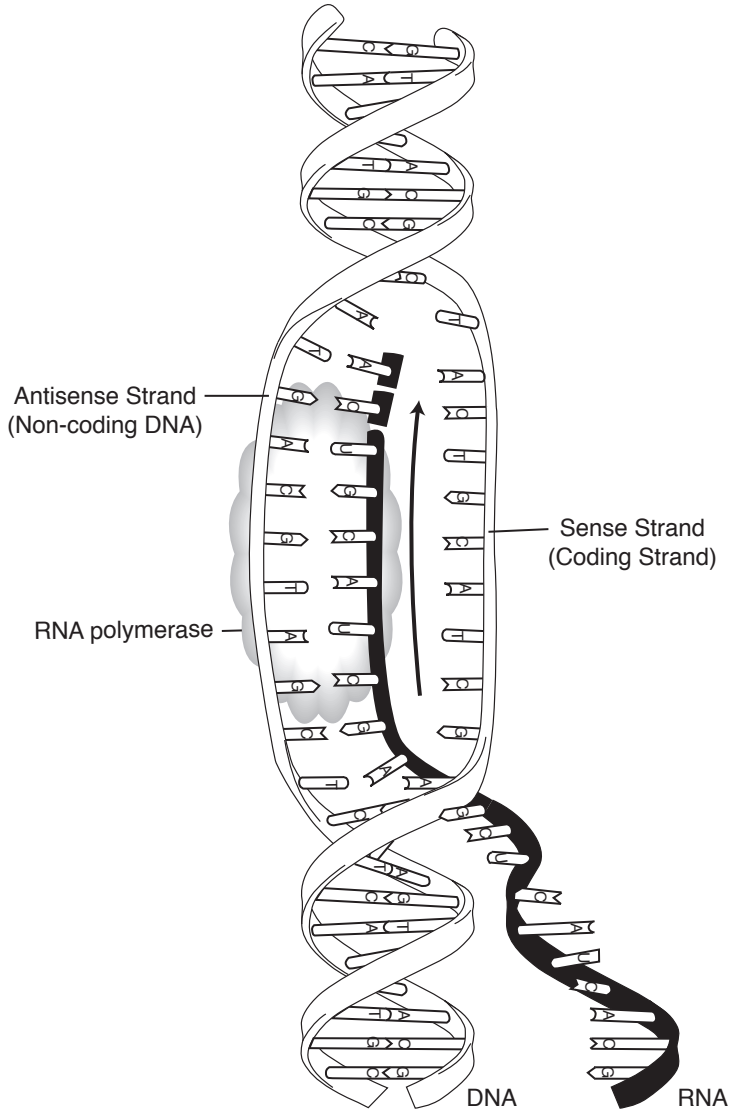


(Courtesy: National Human Genome Research Institute)

Transcription

- Initiated at *promoter site* by *RNA polymerase* (RNAP).
- Promoter is a unidirectional sequence found on one strand which instructs RNAP where to start and in which direction.
- RNAP unwinds double helix at that point and begins synthesis of an mRNA complementary to one of the strands of DNA.
- This strand is called the *antisense* or *template* strand, whereas the other strand is referred to as the *sense* or *coding* strand.
- Synthesis proceeds in a unidirectional manner.
- Terminates when polymerase stumbles upon a stop signal.
- In eukaryotes, not fully understood, but prokaryotes have short region of G's and C's that folds in on itself causing polymerase to trip and release the *nascent*, or newly formed, mRNA.

Transcription



(Courtesy: National Human Genome Research Institute)

Regulation of Transcription

- Ability of RNAP to bind to promoter site can be either enhanced or precluded by *transcription factors*.
- They recognize portions of the DNA sequence near the promoter region known as *operator sites*.
- Those that help RNAP bind are *activators* and those that block RNAP from binding are *repressors*.
- These sequences can be *cis-acting* (affecting adjacent genes), or *trans-acting* (affecting distant genes).

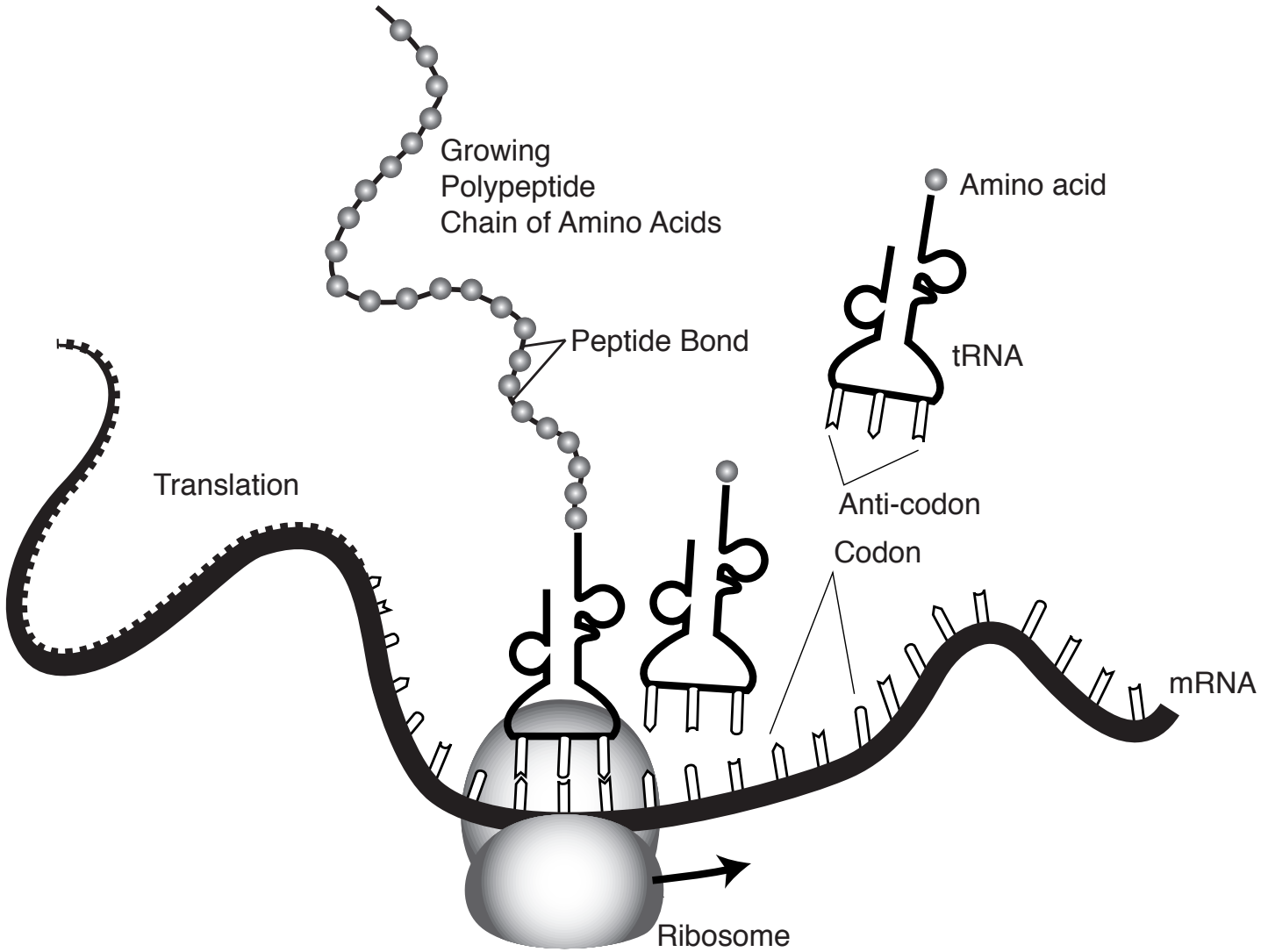
Translation

- *Ribosome* has two subunits.
- *Small subunit* finds mRNA to begin *translating*.
- *Large subunit* has two sites for amino acids to bind.
- *A site* accepts *transfer RNA* (tRNA) bearing an amino acid.
- *P site* binds the tRNA to the growing chain.
- Each tRNA has a specific *acceptor site* that binds a particular triplet of nucleotides, called a *codon*,
- Also has an *anti-codon site* that binds a sequence of three unpaired nucleotides, the *anti-codon*, which binds to the codon.
- Also has specific *charger protein* that only binds to a specific tRNA and attaches correct amino acid to the acceptor site.

Translation (cont)

- *Start signal* is the codon ATG that codes for methionine.
- A tRNA charged with methionine binds to the start signal.
- Large subunit binds to the mRNA and the small subunit, and so begins *elongation*, the formation of the polypeptide chain.
- After the first charged tRNA appears in the A site, the ribosome shifts so that the tRNA is now in the P site.
- New tRNAs, corresponding to codons of the mRNA, enter the A site, and a bond is formed between the two amino acids.
- The first tRNA is now released, and the ribosome shifts again so that a tRNA carrying two amino acids is now in the P site.
- This continues until the ribosome reaches a *stop codon*.
- Ribosome breaks apart releasing the mRNA and new protein.

Translation

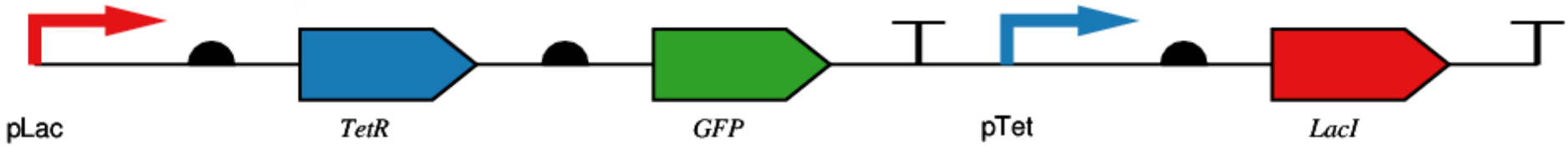


(Courtesy: National Human Genome Research Institute)

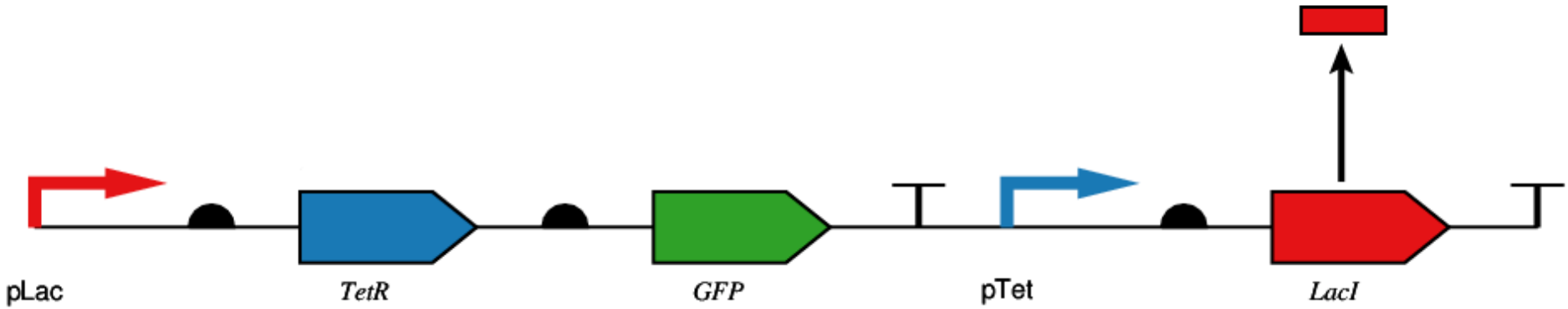
Translational Control

- Note that a protein will often undergo further modification, called *post-translational modification*.
- *Translational regulation* occurs through the binding of *repressor proteins* to a sequence found on an RNA molecule.
- Translational control plays a significant role in the process of embryonic development and cell differentiation.

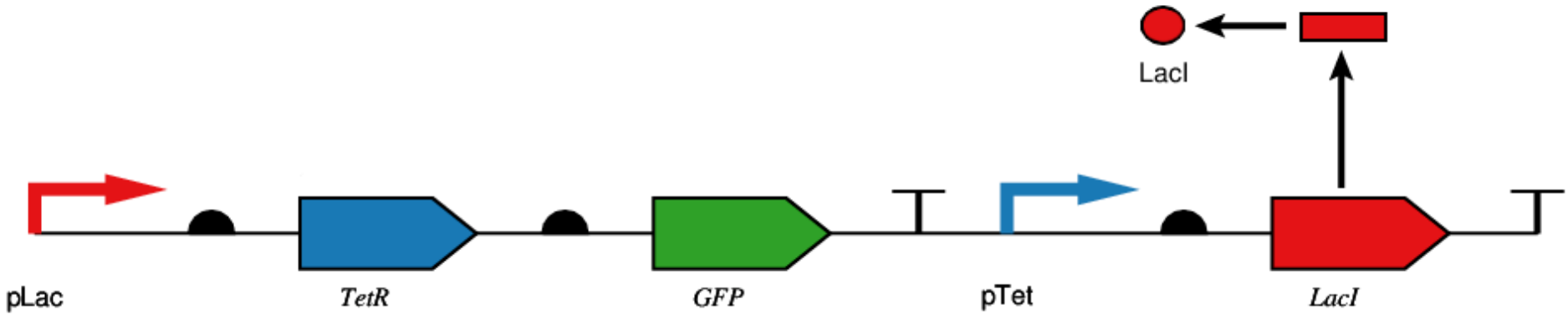
Genetic Toggle Switch (Gardner et al. 2000)



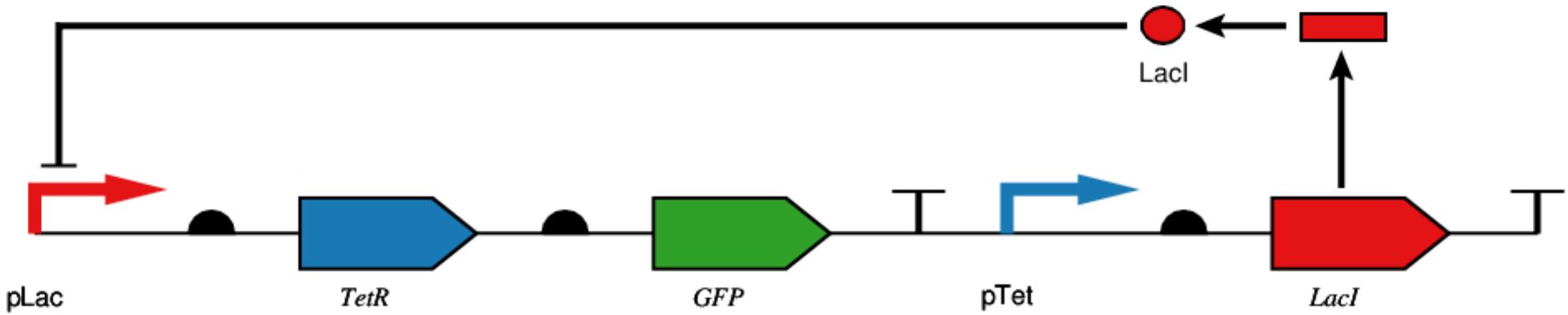
Genetic Toggle Switch (Gardner et al. 2000)



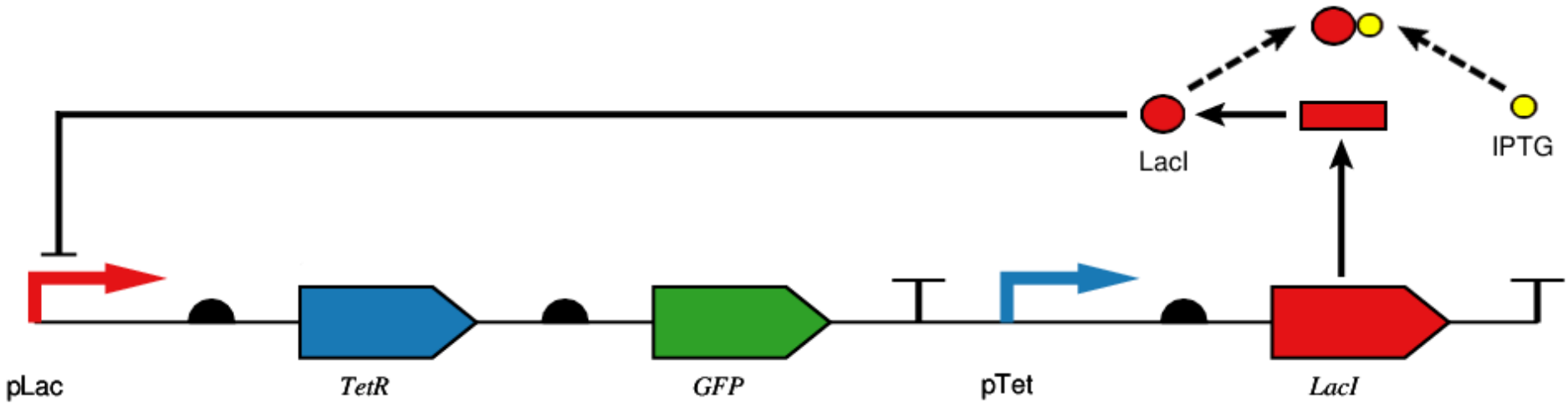
Genetic Toggle Switch (Gardner et al. 2000)



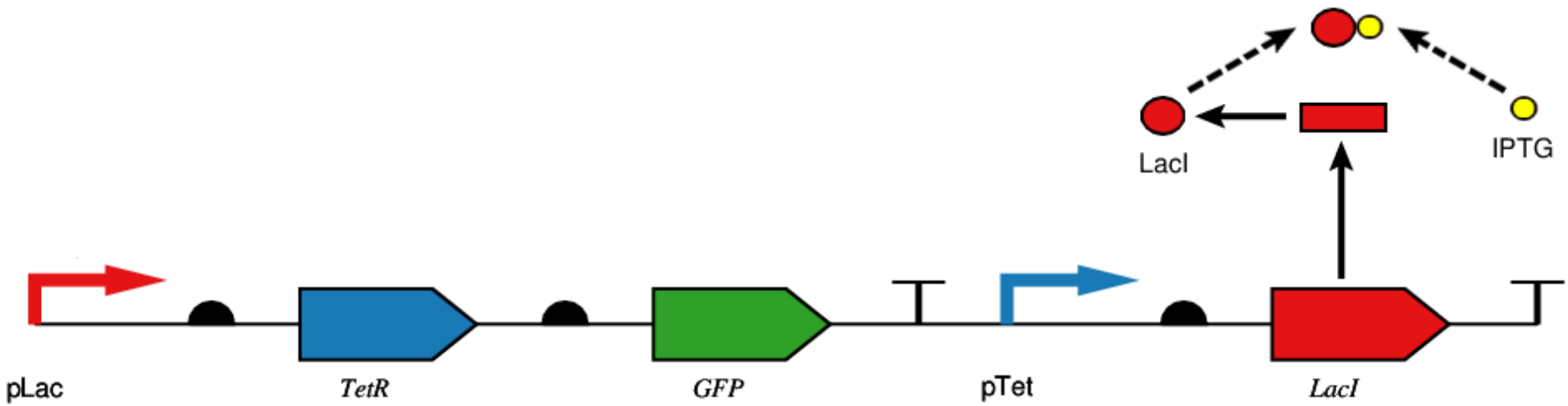
Genetic Toggle Switch (Gardner et al. 2000)



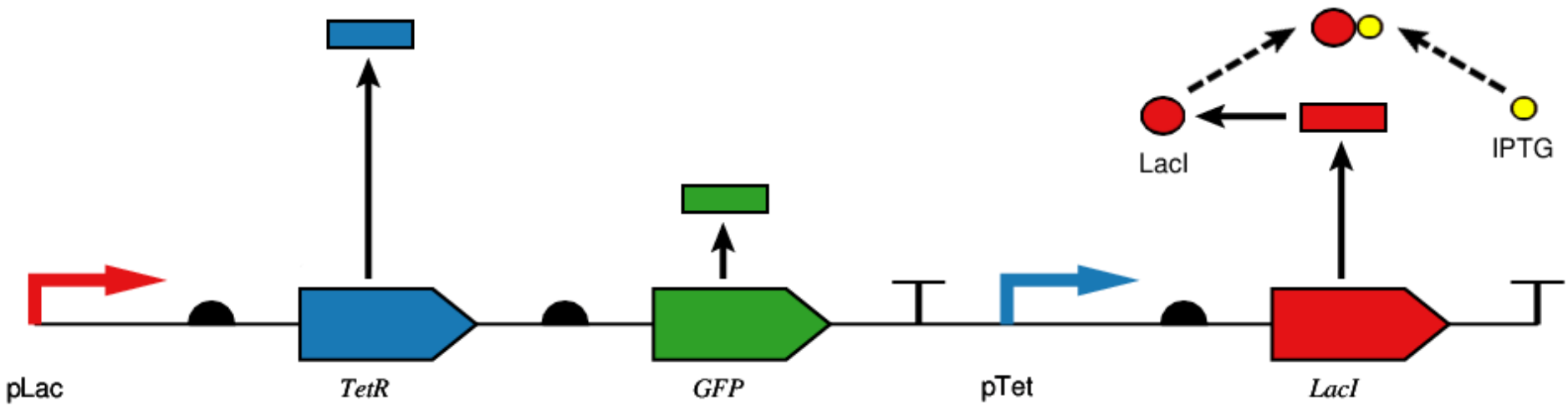
Genetic Toggle Switch (Gardner et al. 2000)



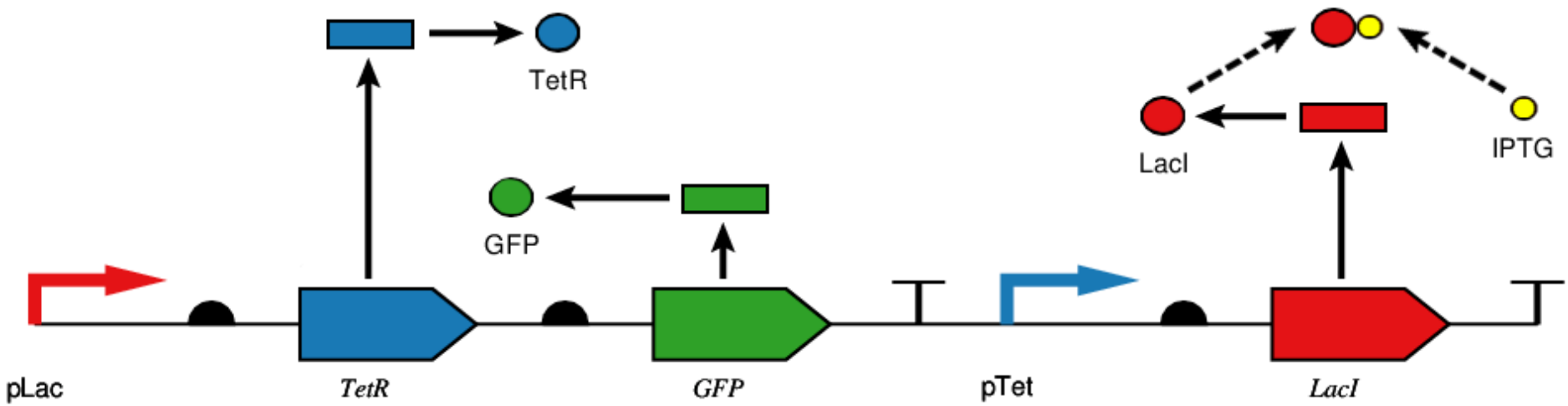
Genetic Toggle Switch (Gardner et al. 2000)



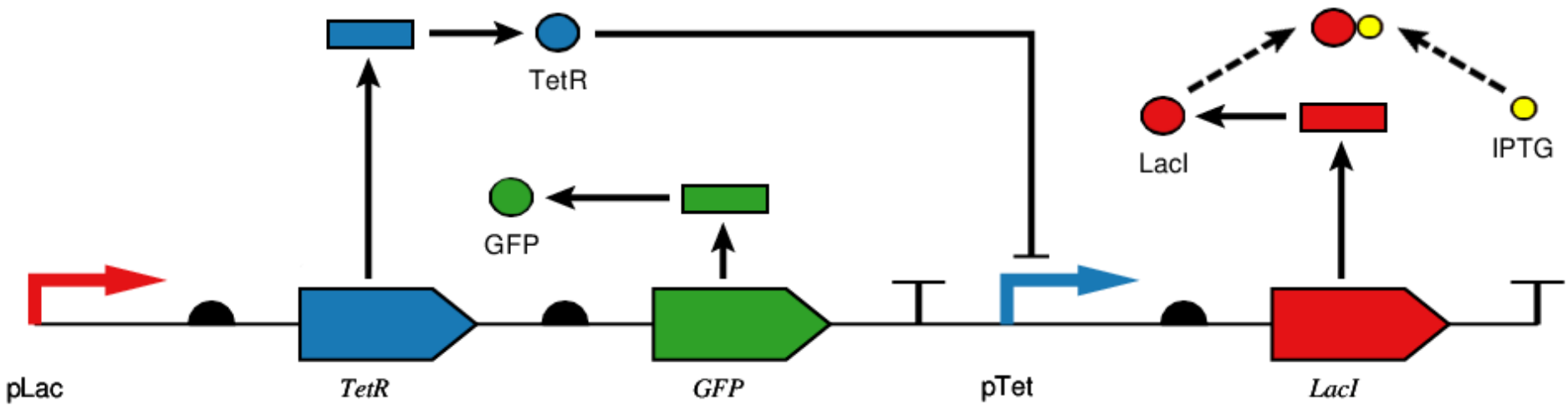
Genetic Toggle Switch (Gardner et al. 2000)



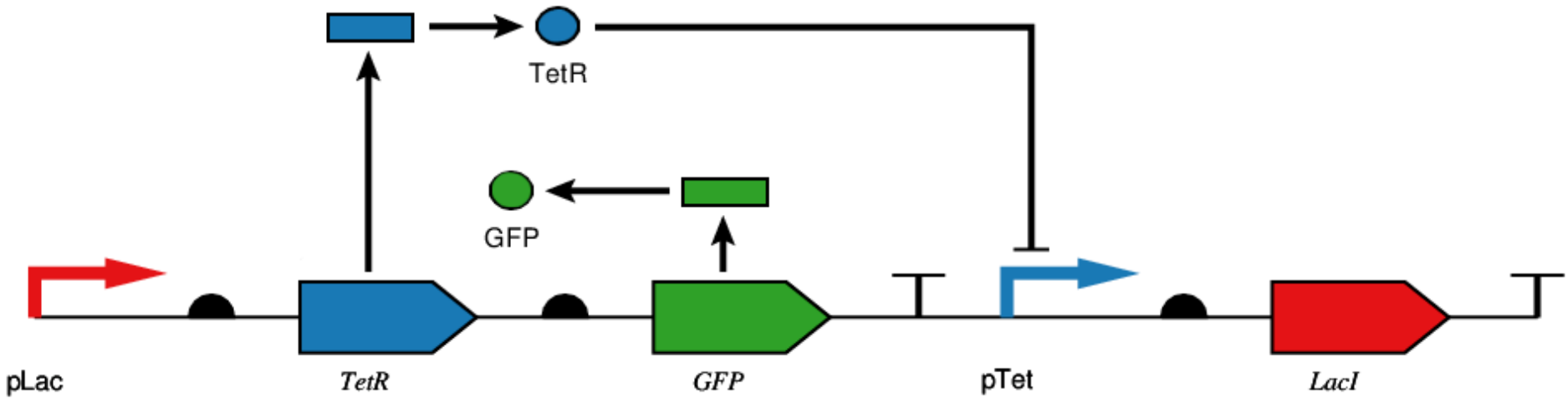
Genetic Toggle Switch (Gardner et al. 2000)



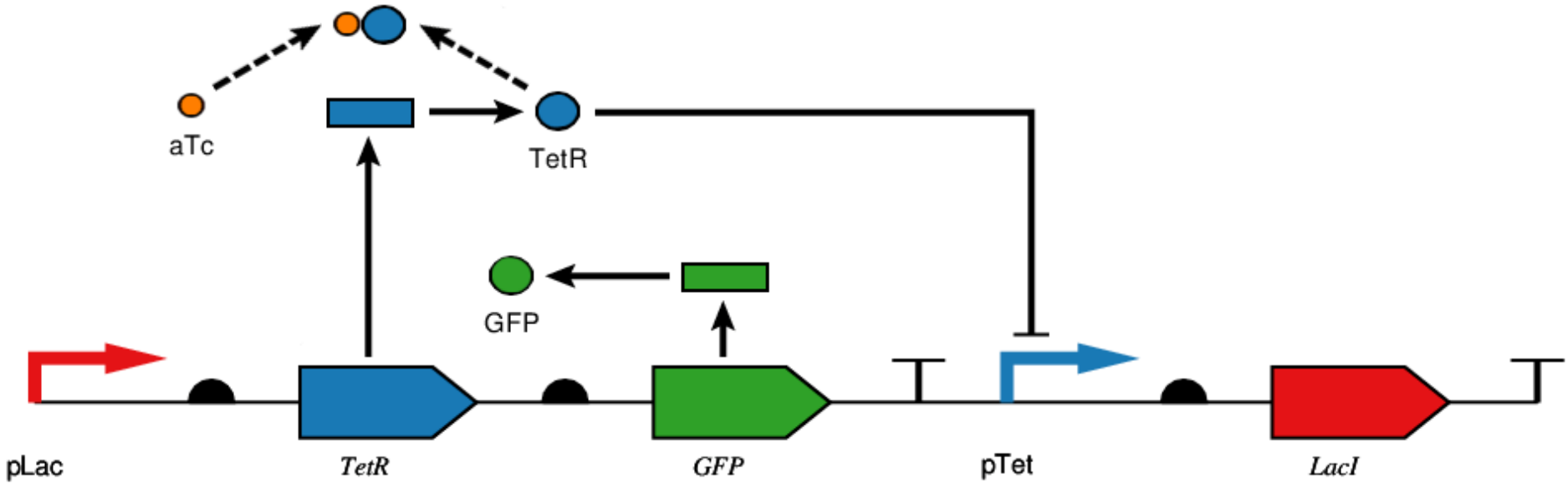
Genetic Toggle Switch (Gardner et al. 2000)



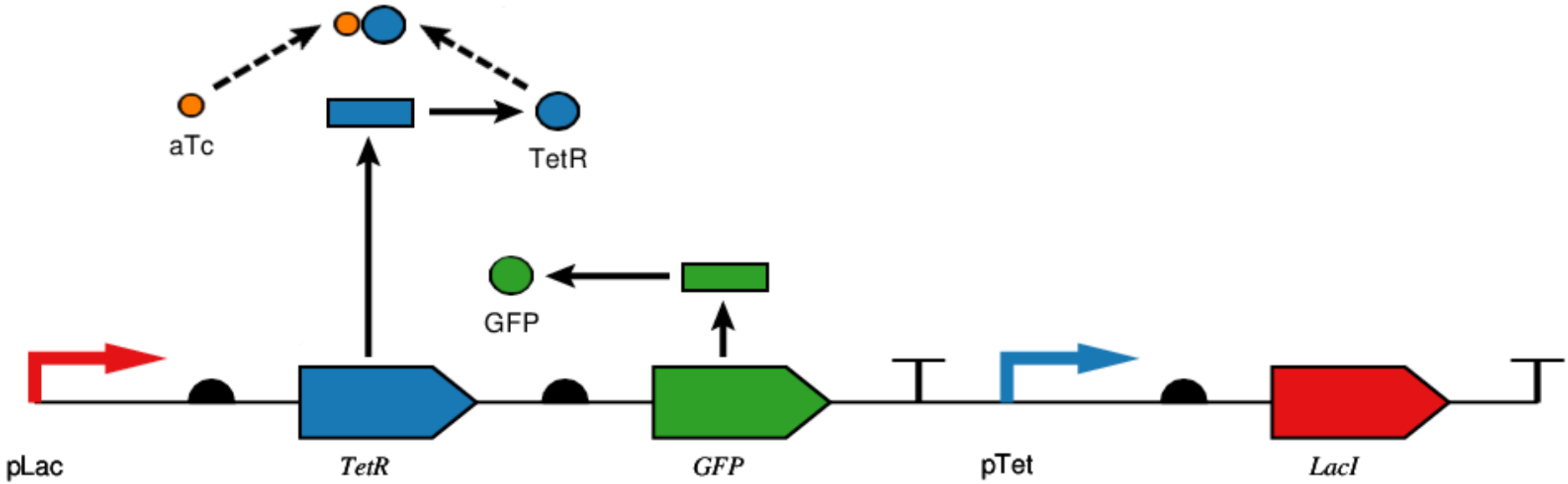
Genetic Toggle Switch (Gardner et al. 2000)



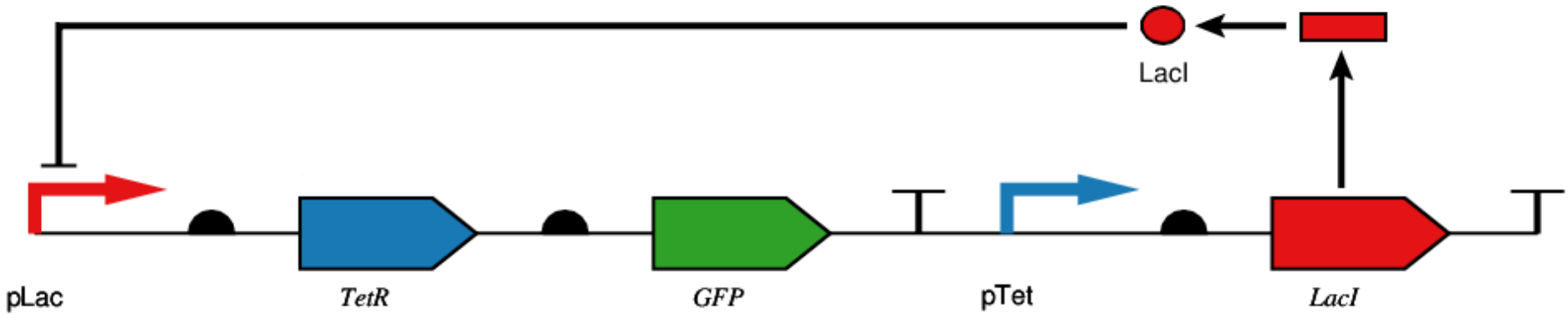
Genetic Toggle Switch (Gardner et al. 2000)



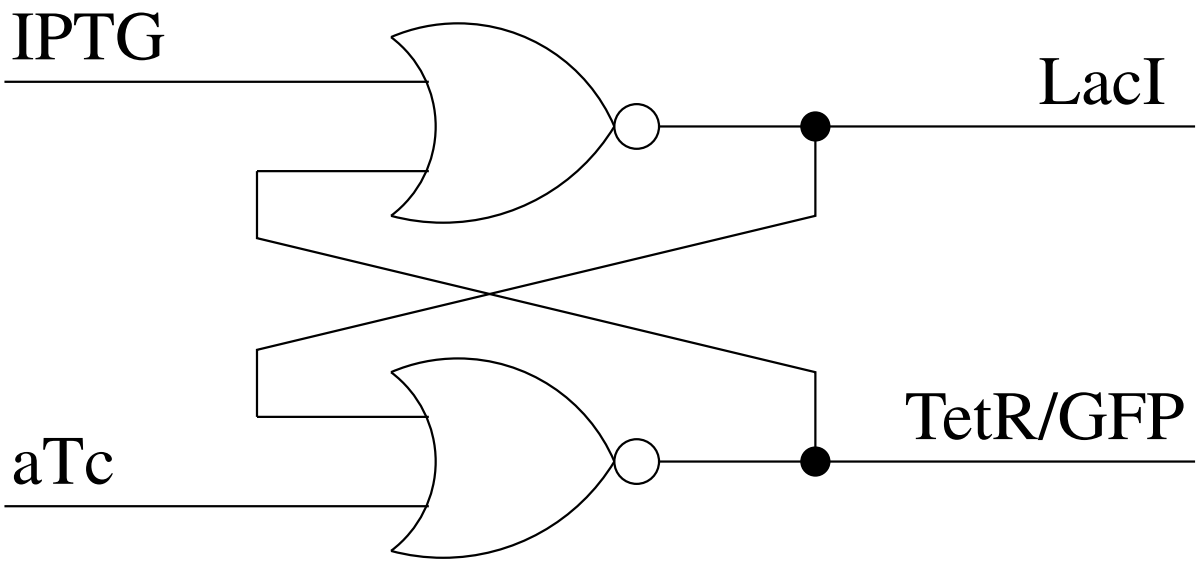
Genetic Toggle Switch (Gardner et al. 2000)



Genetic Toggle Switch (Gardner et al. 2000)



Genetic Toggle Switch (SR Latch) Logic Diagram



Sources

- Engineering Genetic Circuits - Chapter 1
- NCBI's Science Primer - http://www.ncbi.nlm.nih.gov.
- Tozeren/Byers - "New Biology for Eng. and Comp. Scientists".
- Gonick/Wheelis - "The Cartoon Guide to Genetics".
- King/Stansfield - "A Dictionary of Genetics".
- Wikipedia - http://en.wikipedia.org.
- Berg/Tymoczko/Stryer - "Biochemistry".
- Watson et al. - "Molecular Biology of the Gene".
- Alberts et al. - "Molecular Biology of the Cell".

Assignment #1

- Ensure that you understand genetic circuits to the level of detail presented in this lecture.
- Read the genetic toggle switch paper: `http://www.nature.com/nature/journal/v403/n6767/full/403339a0.html`
- Read the paper that you selected for this course.
- Download and examine the supplemental material for your paper.
- Locate and submit information about parts for the genetic toggle switch and your genetic circuit (details in the next lecture).
- **This assignment is due by Friday August 31st.**