2022

The Central Dogma: Gene Expression

Ayisha Sookdeo
CUNY Guttman Community College

Lesson Overview:
In this lesson, students will learn the basic structure and function of DNA and RNA. They will also learn the process of gene expression. Finally, students will learn about the scientific contributor, Ernest Everest Just, and his contribution to biology.

Topics:

- The central dogma
- Gene expression
- Deoxyribonucleic acid
- Ribonucleic acid
- Proteins
- Transcription
- Translation

Learning Goals:

- Identify the structure and function of DNA.
- Identify the structure and function of RNA.
- Explain the process of transcription.
- Explain the process of translation.
- Discover Ernest Everest Just's contribution to biology

Written by: Ayisha R. Sookdeo PhD
Figures created by: Ayisha R. Sookdeo PhD

The **central dogma** is the process by which our genetic information (DNA) is transcribed into messenger RNA (mRNA) and then translated into protein by the cell's ribosomes. Proteins are made up of amino acids. Within a cell, proteins dictate cellular function. DNA contains the information needed to produce all of our proteins and RNA is a messenger that carries this information to the ribosomes to create the proteins. Essentially, the central dogma explains the process of gene expression. During the process of gene expression:

- The conversion of DNA into mRNA is called transcription.
- The conversion of mRNA into protein is called translation.

The Central Dogma
DNA $\xrightarrow[\text{transcription}]{}$ mRNA $\xrightarrow[\text{translation}]{}$ protein

DNA is short for deoxyribonucleic acid. DNA is the hereditary material that instructs an organism to develop and function. DNA is found in every living organism. In prokaryotes, it is found freely in the cytoplasm and in eukaryotes, it is found in the nucleus. The DNA molecule is a double helix that is made up of nucleotides. Each nucleotide is made up of three components: a phosphate group, a sugar molecule and a nitrogen base. There are four nitrogen bases: adenine, guanine, cytosine and tyrosine. Nucleotides attach together to form two long anti-parallel strands that spiral creating that double helix.

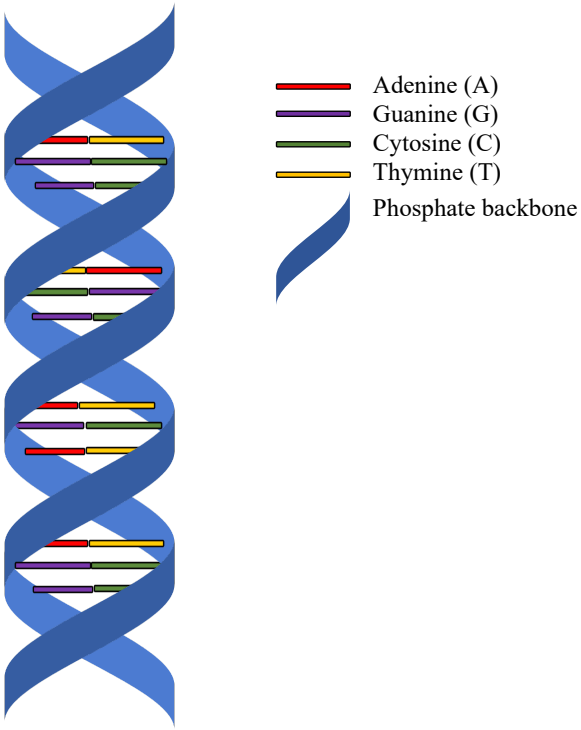


Figure 1: The structure of DNA. DNA is a double helix made up of a sugar phosphate backbone and nitrogen bases: adenine, guanine, cytosine and thymine.

The attachment that occurs between the two strands, occur at the nitrogen bases via hydrogen bonding: adenine binds with tyrosine and guanine binds with cytosine.

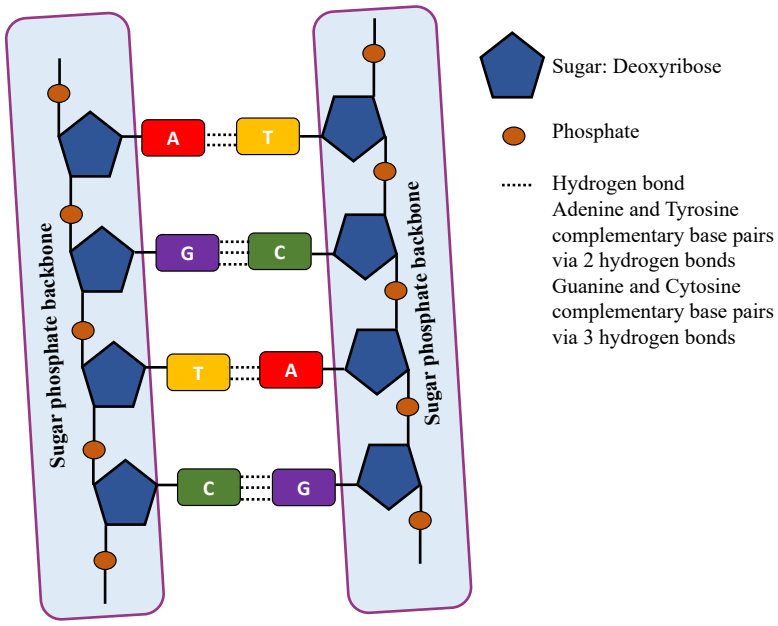


Figure 2: DNA is a double helix with the two strands of DNA being anti-parallel. The nitrogen bases are adenine, thymine, guanine and cytosine. Adenine complementary base pairs with thymine and guanine complementary base pairs with cytosine.

RNA is short for ribonucleic acid. Like DNA, RNA is also made up of three components, a phosphate group, a sugar molecule and a nitrogen base. However, the sugar molecule (ribose) contains one more hydroxyl group on the second carbon. The nitrogen bases are adenine, guanine, cytosine and uracil. Lastly, RNA is a single stranded molecule.

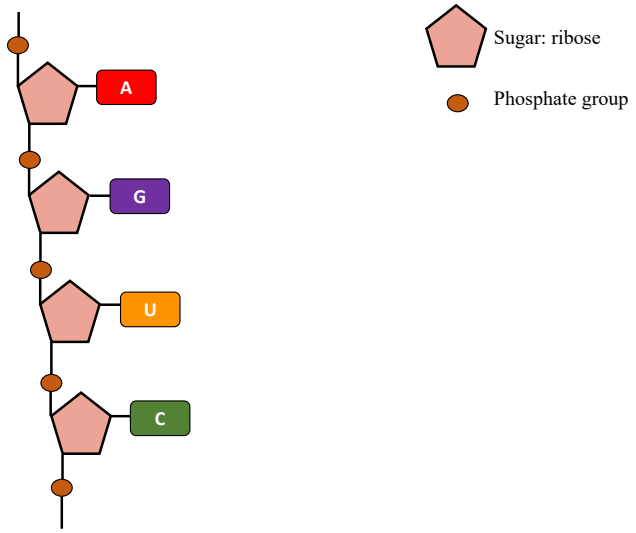


Figure 3: RNA is a single stranded molecule made up of a phosphate group, the sugar, ribose, molecule and a nitrogen base. Like DNA, the nitrogen bases are adenine, thymine, guanine and thymine.

There are three major forms of RNA:

mRNA (messenger RNA): When the cell needs to produce a particular protein, it activates the protein's gene and produces multiple copies of that piece of DNA in the form of mRNA (messenger RNA). mRNA is complementary to one of the DNA strands of a gene. In eukaryotic cells, mRNA contains both coding regions and untranslated regions. Once fully processed a 5' cap and a 3' poly-adenylation tail is added. The cap is a modified guanine (G) nucleotide and it protects the transcript from being broken down. The polyadenylation tail contains roughly 100-200 adenines. The multiple copies of mRNA are then used to translate the genetic code into protein through the action of the ribosomes.

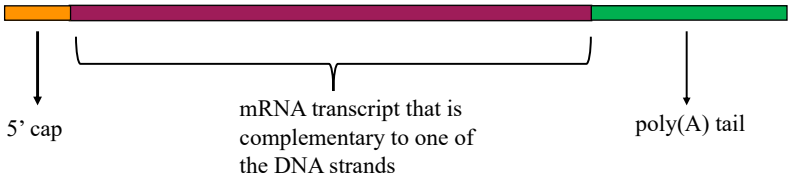


Figure 4: After the mRNA transcript is made, the molecule is further processed with the addition of a 5' modified guanine cap and a 3' poly-adenylation tail.

tRNA (transfer RNA): tRNA is a small RNA molecule that serves a crucial role in protein synthesis. Each tRNA molecule has two important regions: a trinucleotide region called the anticodon and a region for attaching a specific amino acid. Remember, a protein is made up of a chain of amino acids. During translation, when a new amino acid is added to the growing chain of a protein, a tRNA molecule forms a base pair with its complementary sequence on the messenger RNA (mRNA) molecule – this ensures that the appropriate amino acid is inserted into the protein.

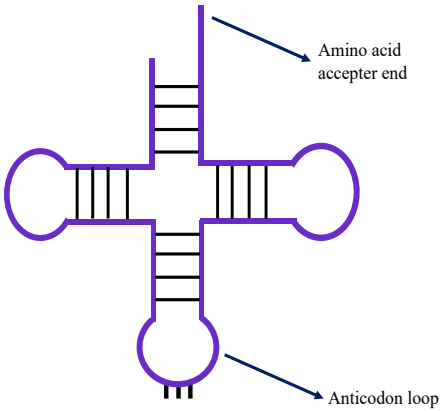


Figure 5: The tRNA molecule is visualized here in its cloverleaf structure highlighting two important components, the anticodon loop and the amino acid acceptor end. The anticodon loop contains the three-nucleotide anticodon which can recognize and interpret an mRNA codon. The amino acid acceptor end contains the corresponding amino acid. When a tRNA recognizes and binds to its codon of the mRNA transcript in the ribosome, the tRNA transfers the appropriate amino acid to the end of the growing amino acid chain during translation.

rRNA (ribosomal RNA): rRNA is a type of non-coding RNA that is the primary component of ribosomes. rRNA is a ribozyme which carries out protein synthesis in ribosomes. Molecules of rRNA are synthesized in a specialized region of the cell nucleus called the nucleolus. Ribosomes are largely made up of rRNA and associated with ribosomal proteins to form the small and large ribosome subunits.

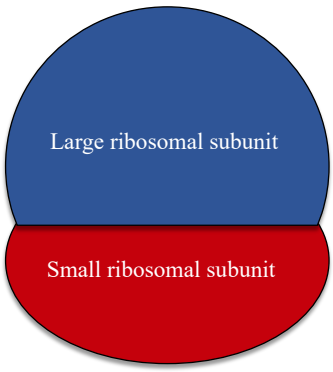


Figure 6: The ribosome is a macromolecule that is responsible for protein synthesis. The ribosome is made up of rRNA and ribosomal proteins. Ribosomes are made up of the large ribosomal subunit and small ribosomal subunit.

Proteins are large molecules that do most of the work within a cell. Proteins have multiple functions such as cellular metabolism, regulation, structural support, transport and defense. As previously mentioned, proteins are made up of smaller units called amino acids. There are 20 different types of amino acids.

The 20 amino acids			
alanine arginine asparagine aspartic acid cysteine	glutamine glutamic acid glycine histidine isoleucine	leucine lysine methionine phenylalanine proline	serine threonine tryptophan tyrosine valine

Each amino acid corresponds with a triplet nucleotide code called a codon. There are 64 different codons. With only 20 amino acids and 64 different codons, the genetic code is degenerate. In other words, one amino acid can have different codons. This is beneficial with respect to mutations - if there are any mutations within the genetic code, the same amino acid will still be read.

Let's take a deeper dive into the process of gene expression. Gene expression is the process of converting genetic information into functional proteins. The first step in gene expression is transcription.

During **transcription**, the genetic information on the strand of DNA is copied into mRNA. In eukaryotic cells, DNA is stored in the nucleus, therefore transcription occurs in the nucleus. In prokaryotic cells, there is no nucleus, therefore, transcription occurs in the cytoplasm. It is important to note that the mRNA molecule is not the exact copy of the DNA template, but rather complementary to it.

When a cell needs a particular protein, its mRNA is triggered to be synthesized. But how does the cell know which gene to transcribe? Simply put, accessory proteins called transcription factors bind to a specific DNA sequence called promoter sequences or enhancers. When the transcription factors bind to these regions, it sequesters RNA polymerase to the transcription site. The transcription factors and the RNA polymerase form the transcription initiation complex. The complex initiates transcription, RNA polymerase synthesizes mRNA using the DNA as the template strand by complementary base pairing. The mRNA molecule is elongated and transcription gets terminated once the strand is completely synthesized. The resultant mRNA copies of the gene act as a blueprint for protein synthesis during the process of translation.

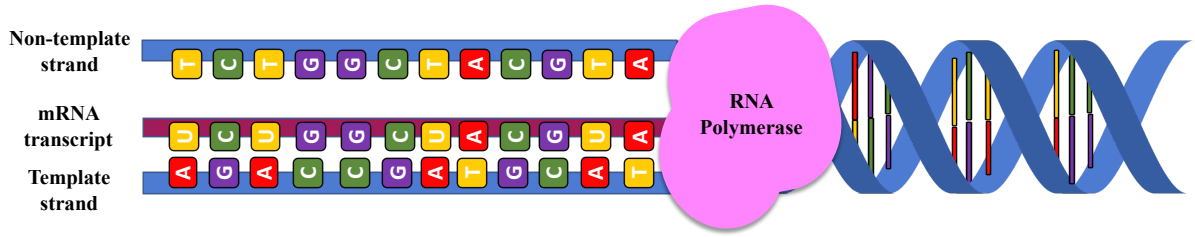


Figure 7: During transcription, one of the DNA strands of the gene of interest acts as a template. RNA polymerase is responsible for synthesizing the mRNA transcript by complementary base pairing to the template strand. Once the gene is completely synthesized, transcription is terminated and the mRNA transcript acts as a blueprint for protein synthesis.

Translation is the process by which proteins are synthesized and it occurs in the cytoplasm in both prokaryotic and eukaryotic cells. It is important to note that in eukaryotic cells, the mRNA transcript must be transported from the nucleus to the cytoplasm in order to be translated into its protein product. Translation occurs on the ribosomes. The ribosome has a small and large subunit and is also associated with many ribosomal RNA molecules and proteins. Translation of an mRNA molecule by the ribosome occurs in three stages: initiation, elongation, and termination.
During initiation:

1. The small ribosomal subunit binds to the start of the mRNA sequence. The start codon of all mRNA sequence is AUG and that encodes for the amino acid methionine. A transfer RNA (tRNA) molecule that carries the amino acid methionine binds to the start codon of the mRNA sequence and the A site of the ribosome.
2. The large ribosomal subunit binds to form the now complete initiation complex.

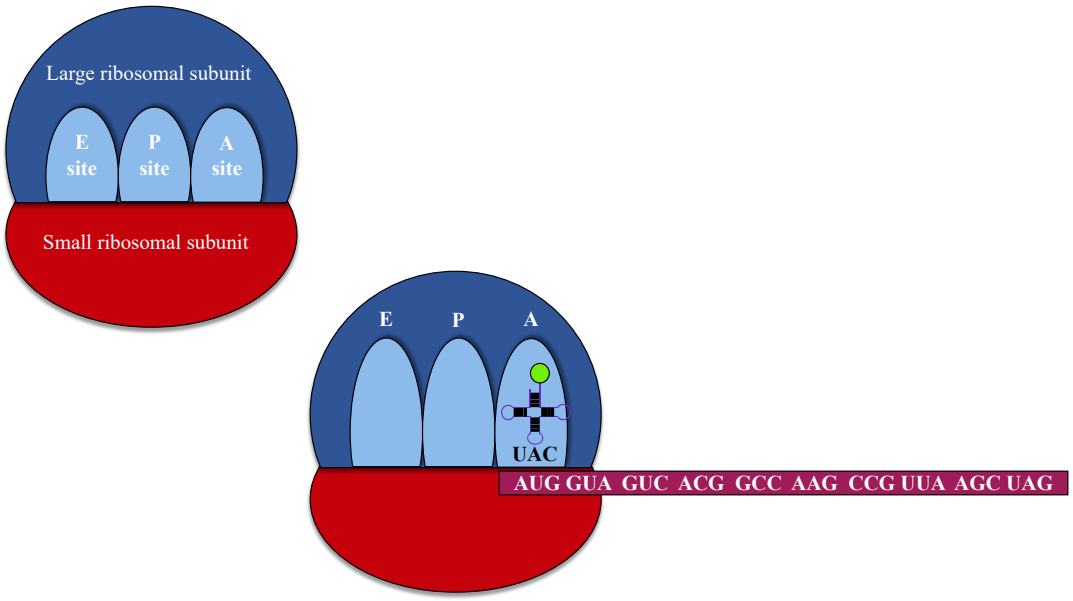
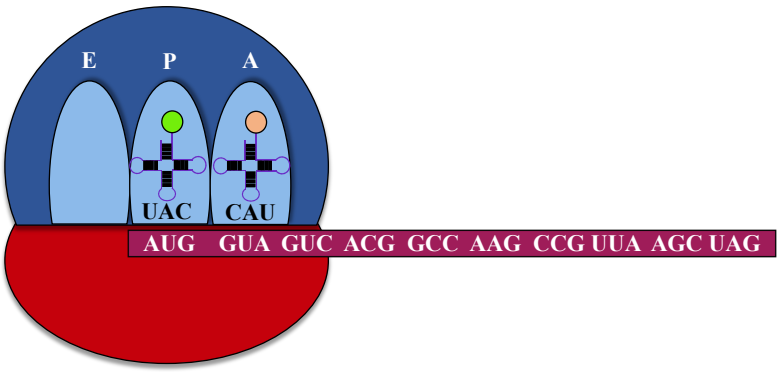


Figure 8: Translation begins at the start codon (AUG) which encodes for methionine. The Ribosomes have three sites: the A site, P site and E site. The A (aminoacyl) site is the first binding site in the ribosome. The P (peptidyl) site holds the tRNA which is linked to the growing polypeptide chain. The E (exit) site is where the tRNA molecules bind before they dissociate from the ribosome.

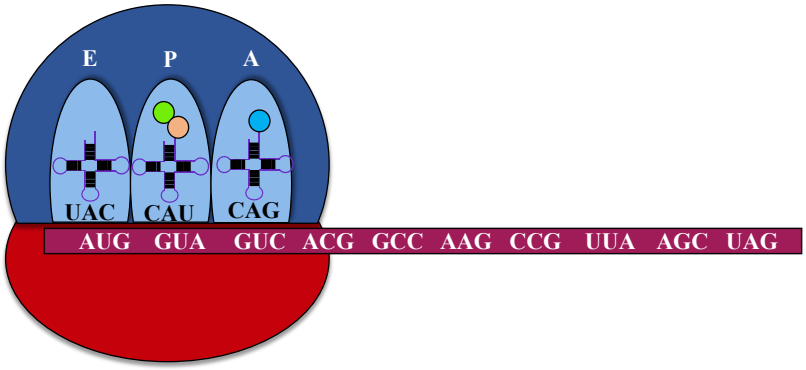
During the elongation:

1. The fully assembled ribosome moves along the transcript and translates each codon.
2. Each corresponding amino acid is added to the growing chain and linked via a bond called a peptide bond (hence why a protein is also referred to as a polypeptide). This occurs at the P site of the ribosome.
3. Elongation persists until all of the codons are read.

A.



B.



C.

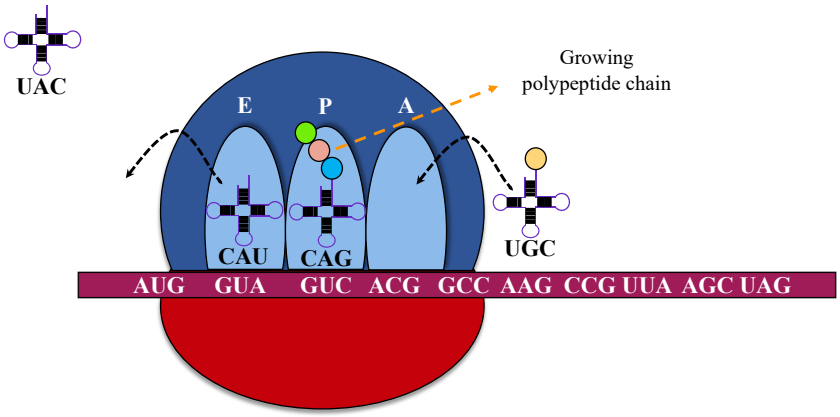


Figure 9: (**A**)During elongation, the ribosome translates each codon in order (from the 5' end to the 3' end). Each corresponding amino acid is added to the growing chain via the A-site (**B)** and linked via a bond called a peptide bond at the P-site. (**C**) Once the bond is formed, the tRNA moves to the E-site and exits the ribosome. Elongation continues until all of the codons are read.

During termination:

1. Once the ribosome reaches a stop codon (UAA, UAG, and UGA), translation ends. There are no tRNA molecules that recognize these codons; therefore, translation is complete. The new protein is released from the ribosome and the translation complex disassembles.

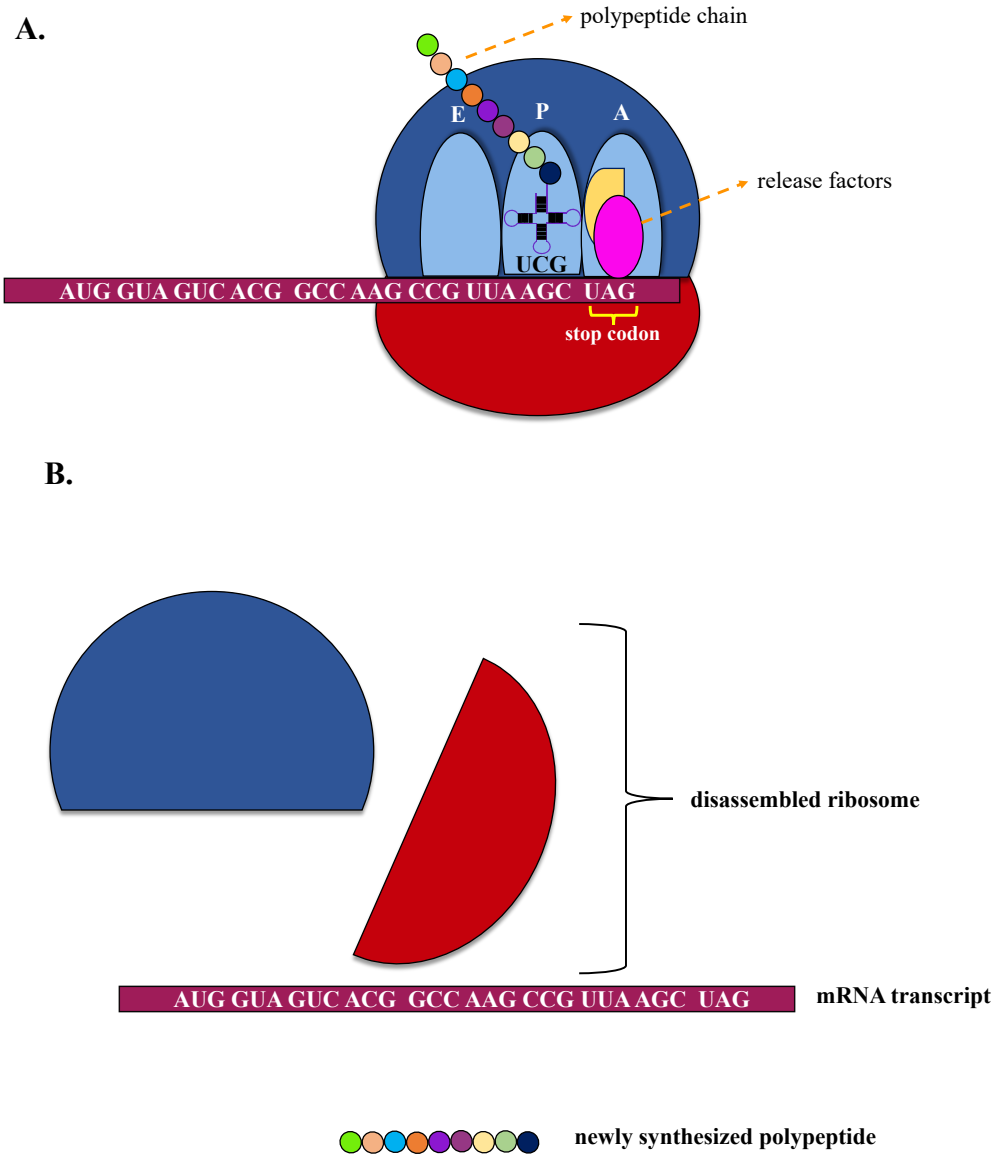


Figure 10: (**A**) When the ribosomes reach a stop codon, translation ends. Release factors bind at the A-site (**B**) This triggers the release of the polypeptide chain and the disassembling of the ribosome.

Scientist Spotlight
Ernest Everest Just

Ernest Everest Just (**E.E. Just**) was an African American biologist born on August 14, 1883 in Charleston, South Carolina. He lived through the Segregation era - navigating through a society that enforced diminished access to education, housing and opportunities for African Americans. Yet, E.E. Just was able to become a prominent contributor to the scientific world. His interest in biology began during his undergraduate studies in Dartmouth College. He conducted research with Dartmouth Professor William Patten, in the area of anatomy and evolution of vertebrates (frogs, mollusk and limpet). He graduated from Dartmouth College in 1907. Soon after graduation, Just accepted a faculty position at the HBCU Howard University in Washington DC where he initially taught English. He later joined the Department of Biology and Geology and excelled as a Professor of Zoology. In 1912, Just published his first scientific paper about the marine annelid *Nereis limbate*. Within that text, he explored how the sperm entrance point determines the first cleavage plane in the egg during fertilization.

Being a true scholar, Just earned a PhD in 1916 while teaching at Howard University from the University of Chicago. He was the first African American to do so. His doctoral studies focused on the breeding habits and the fertilization reaction of another marine annelid, *Platynereis megalops*. Just also conducted research at the Marine Biological Laboratory in Woods Hole, Massachusetts. While at Woods Hole, Just investigated how parthenogenesis in marine invertebrates are effected by UV irradiation, hypertonic and hypotonic sea water. He was infamous for his excellent research techniques in studying invertebrates and recreating their natural environment to that in the lab. Just published several books, one of which called The Biology of the Cell Surface, where he presented his theory of genetic restriction. This theory explored how genes and cytoplasmic factors interact during the developmental process. Just placed importance on the cytoplasmic factors being responsible for the differentiation of cells. He discusses his perspective on the inadequacy of Thomas Hunt Morgan's gene theory of heredity; if the genes are the same in every cell, then how can the genes be responsible for differentiation. Over a century later, we know more about the intricacies of cellular differentiation and understand that Just was not completely wrong in his theory. Cellular differentiation does indeed depend on the genes within a cell, however, there are many cytoplasmic factors (transcription factors) that turn on and turn off gene expression. The cytoplasm is indeed involved in differentiation and Thomas Hunt Morgan's gene-centric outlook on cell differentiation did not tell the complete story. If we were to explore this concept with respect to gene expression, though DNA is found in the nucleus, the process of converting DNA into protein occurs in the cytoplasm. There are many crucial machineries that contribute to gene expression that exist in the cytoplasm, like ribosomes, tRNA, etc. E.E. Just's theory was ahead of his time, and it is worth commemorating his intelligence and contribution to the world of science.

Test your knowledge

A. Answer questions listed below:

1. Describe three differences between DNA and RNA.

2. What are the four nitrogen bases in DNA?

3. What is the central dogma of molecular biology?

4. Which bases complimentary base pairs with each other?

5. Describe the three major types of RNA that is found in a cell.

6. What is transcription?

7. What is translation?

8. If the cell requires a particular protein to be synthesized, what will happen if RNA polymerase is non-functional?

9. What happens if termination occurs prematurely during protein synthesis?

10. What are the downstream effects of a non-functional nucleolus?

B. Create the mRNA transcript to the below sequences.

1. ACGATGCCATAGAAAA

2. TTTGCCCGCGTATATAA

3. CGCGATACAGATAGATC

C. Research Question: Write a three-paragraph essay answering the questions below.

Explain Thomas Hunt Morgan's gene theory of heredity. Explain E.E.Just's genetic restriction theory. Now that you are acquainted with the process of gene expression, in what ways were E.E. Just's theory accurate. Conversely, in what ways was his theory inaccurate?