

# 7

## Information Processing

Man cannot discover new oceans unless he has the courage to lose sight of the shore.

-Andre Gide



*"The blueprints for the construction of one human being requires only a meter of DNA and one tiny cell. ... even Mozart started out this way."* — L.L. Larison Cudmore

As creatures used to regarding ourselves as exceptional, humans must surely be humbled to realize that the instructions, for making one of our own, reside in a molecule so simple that scientists, for a very long time, did not believe

could possibly contain enough information to build even a simple cell. But a large body of evidence, built up over the past century, supports Larison Cudmore's assertion that the information for making you and me (and all the other kinds of living things in the world) is encoded in DNA. Tying in with Mendel's observations about how characteristics are passed on from one generation to the next, the discovery that there was a molecule that carried this information, altered forever how people thought about heredity.

The elucidation of the structure of DNA provided greater insights into how traits might be encoded in a molecule, and the ways in which the information is used by cells. As we learn more about this topic, scientists have remarked on how the information in our DNA resembles the programs that drive computers. While this analogy is a simplification, there is definitely a sense in which, as Richard Dawkins put it, "the machine code of the genes is uncannily computer-like", with information in our DNA directly determining the properties of the proteins that run our cells. We know, as Ada Yonath described it, that, "DNA is a code of four letters; proteins are made up of amino acids which come in 20 forms. So the ribosome is a very clever machine that reads one language and operates in another. "

If this sounds strange, it is even more intriguing to realize DNA is copied and passed on

from cell to cell, from one generation to the next. There is an unbroken line of inheritance from the first cell to every organism alive today. In the words of Lewis Thomas, "All of today's DNA, strung through all the cells of the earth, is simply an extension and elaboration of [the] first molecule."

The nature of this information, how it is copied and passed on, how it is read and interpreted, and how it gives rise to the cellular activities that we can observe, is the subject of this chapter. Another kind of information is also considered, towards the end of the chapter- the molecular information that cells receive from, and send to, each other. Overlaid on the instructions in the genes, this information provides cells with ongoing clues about both their own inner state and the environment around them. The interplay of these two kinds of information is responsible for the form and behavior of all living organisms.

# Information Processing: Genes and Genomes



## Introduction

For many years, scientists wondered about the nature of the information that directed the activities of cells. What kind of molecules carried the information, and how was the information passed on from one generation to the next? Key experiments, done between the 1920s and the 1950s, established convincingly that this genetic information was carried by DNA. In 1953, with the elucidation of the structure of DNA, it was pos-

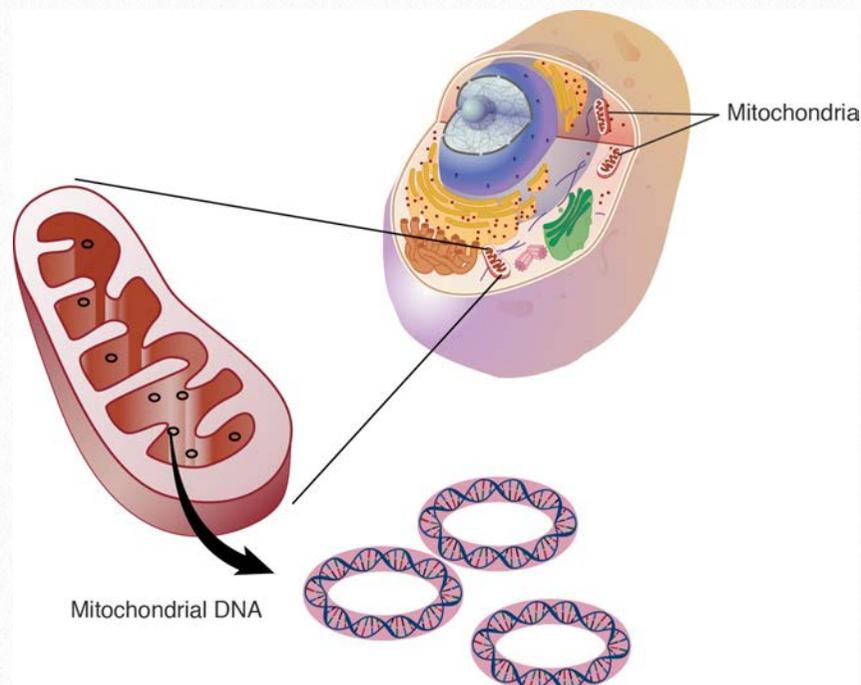
sible to begin investigating how this information is passed on, and how it is used.

## Genomes

We use the word “genome” to describe all of the genetic material of the cell. That is, a genome is the entire sequence of nucleotides in the DNA that is in all of the chromosomes of a cell. When we use the term genome without further qualification, we are generally referring to the chromosomes in the nucleus of a eukaryotic cell. As

**YouTube Lectures  
by Kevin  
[HERE](#) & [HERE](#)**

you know, eukaryotic cells have organelles like mitochondria and chloroplasts that have their own DNA (Figure 7.1 & 7.2). These are referred to as the mitochondrial or chloroplast genomes to distinguish them from the nuclear genome.



**Figure 7.1 - Mitochondria carry their own genome**

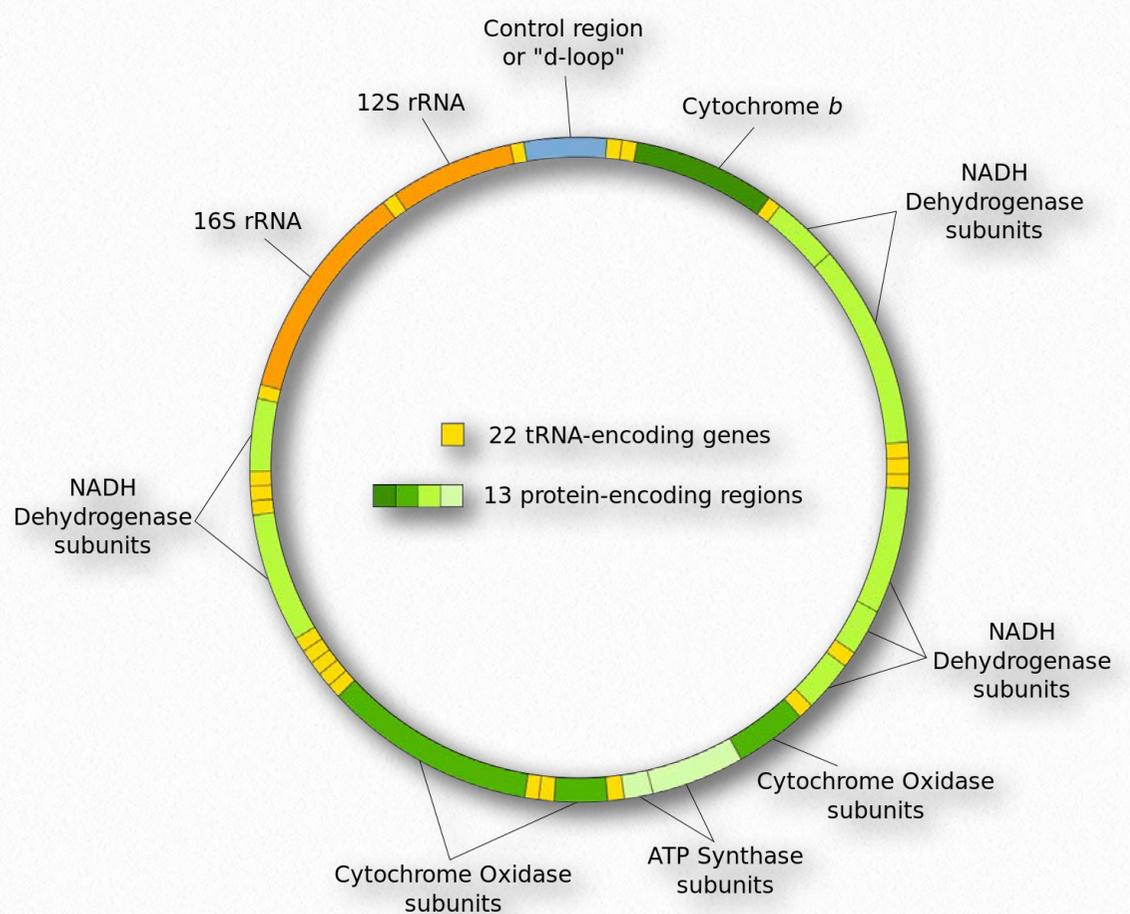
Starting in the 1980s, scientists began to determine the complete sequence of the genomes of many organisms, in the hope of better understanding how the DNA sequence specifies cellular functions. Today, the complete genome sequences have been determined for thousands of species from all domains of life, and many more are in the process of being worked out by groups of scientists across the world.

### Global genome initiative

The Global Genome Initiative, a collaborative effort to sequence at least one species from each of the 9,500 described invertebrate, vertebrate, and plant families is one of many such

ventures. The information from these various efforts is collected in enormous online repositories, so that it is freely available to scientists. As the sequence databases compile ever more information, the fields of computational biology and bioinformatics have arisen, to analyze and organize the data in a

way that helps biologists understand what the information in DNA means in the cellular context.



**Figure 7.2 - The human mitochondrial genome**

Wikipedia

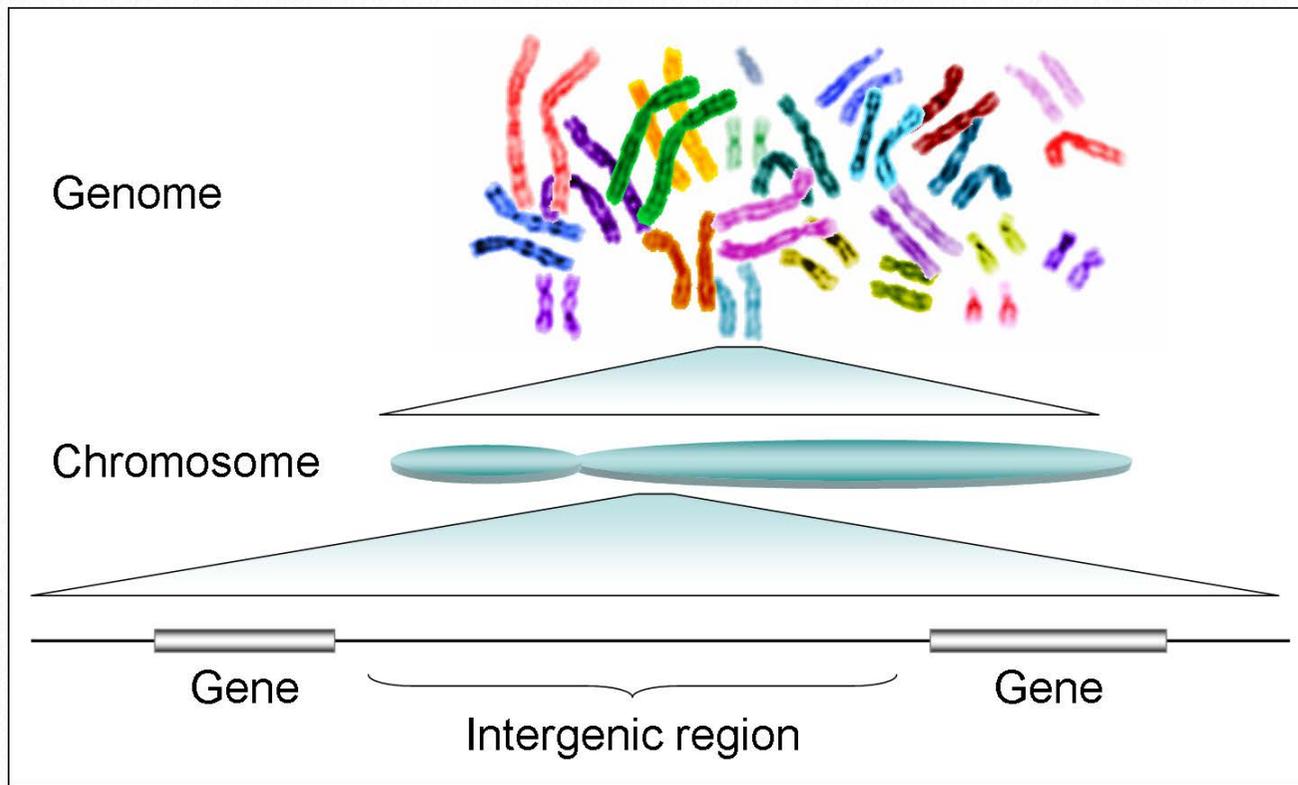


Figure 7.3 - From genomes to genes

Wikipedia

DNA was envisioned as a long string of nucleotides, in which certain regions, the genes, were separated by non-coding regions that were simply referred to as intergenic sequences (*inter*=between; *genic*=of genes). Early experiments in molecular biology suggested a simple relationship between the DNA sequence of a gene and

## Genes

It has been known for many years that phenotypic traits are controlled by specific regions of the DNA that were termed "genes". Thus,

its product, and led scientists to believe that each gene carried the information for a single protein. Changes, or mutations in the base sequence of a gene would be reflected in

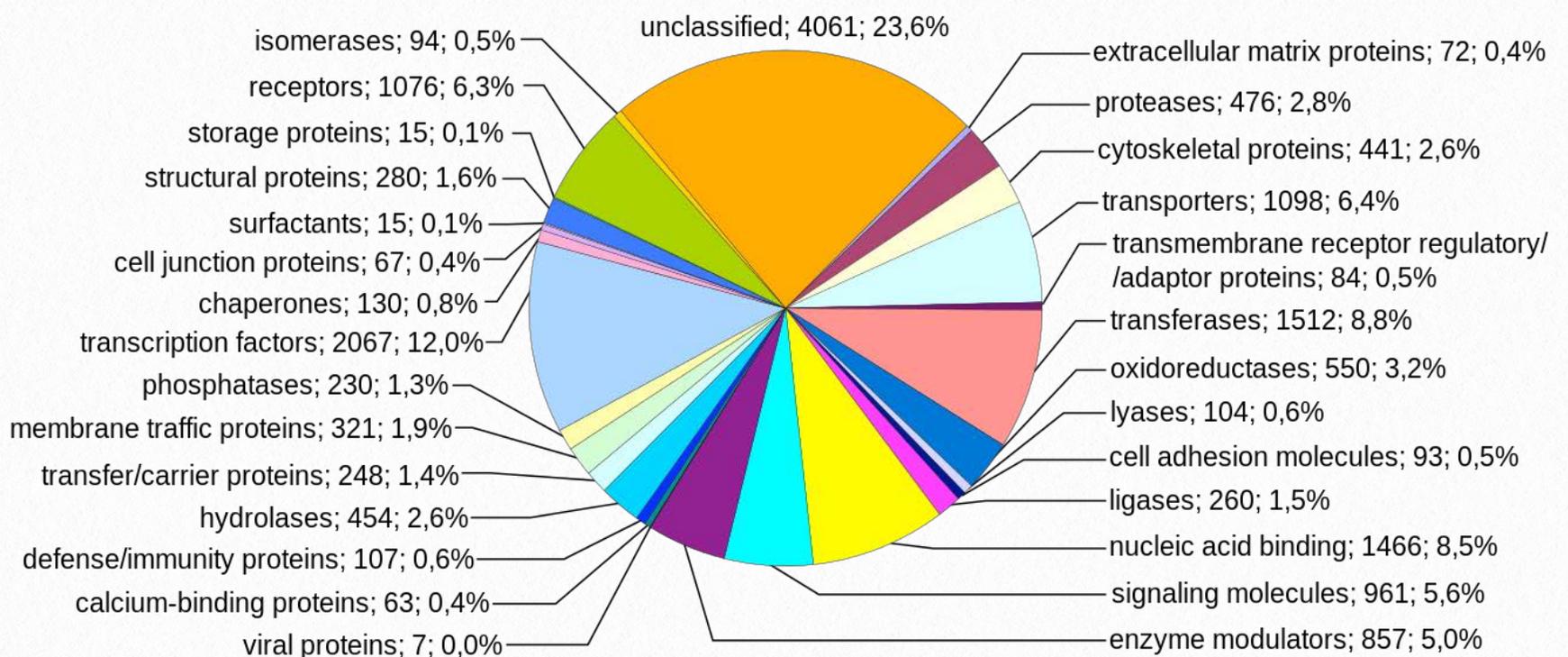


Figure 7.4 - Human genes sorted by class

changes in the gene product, which in turn, would manifest itself in the phenotype or observable trait. This simple picture, while still useful, has been modified by subsequent discoveries that demonstrated that the use of genetic information by cells is somewhat more complicated. Our definition of a gene is also evolving to take new knowledge into consideration.

## Matters of size

A common-sense assumption about genomes would be that if genes specify proteins, then the more proteins an organism made, the more genes it would need to have, and thus, the larger its genome would be. Comparison of various genomes shows, surprisingly, that there is not necessarily a direct relationship between the complexity of an organism and the size of its genome (Figure 7.5). To understand how this could be true, it is necessary to recognize that while genes are made up of DNA, all DNA does not consist of genes (for purposes of our discussion, we define a gene as a section of DNA that encodes an RNA or protein product). In the human genome, less than 2% of the total

DNA seems to be the sort of coding sequence that directs the synthesis of proteins. For many years, non-coding DNA in genomes was believed to be useless, and was described as “junk DNA” although it was perplexing that there seemed to be so much “useless” sequence. Recent discoveries have, however, demonstrated that much of this so-called junk

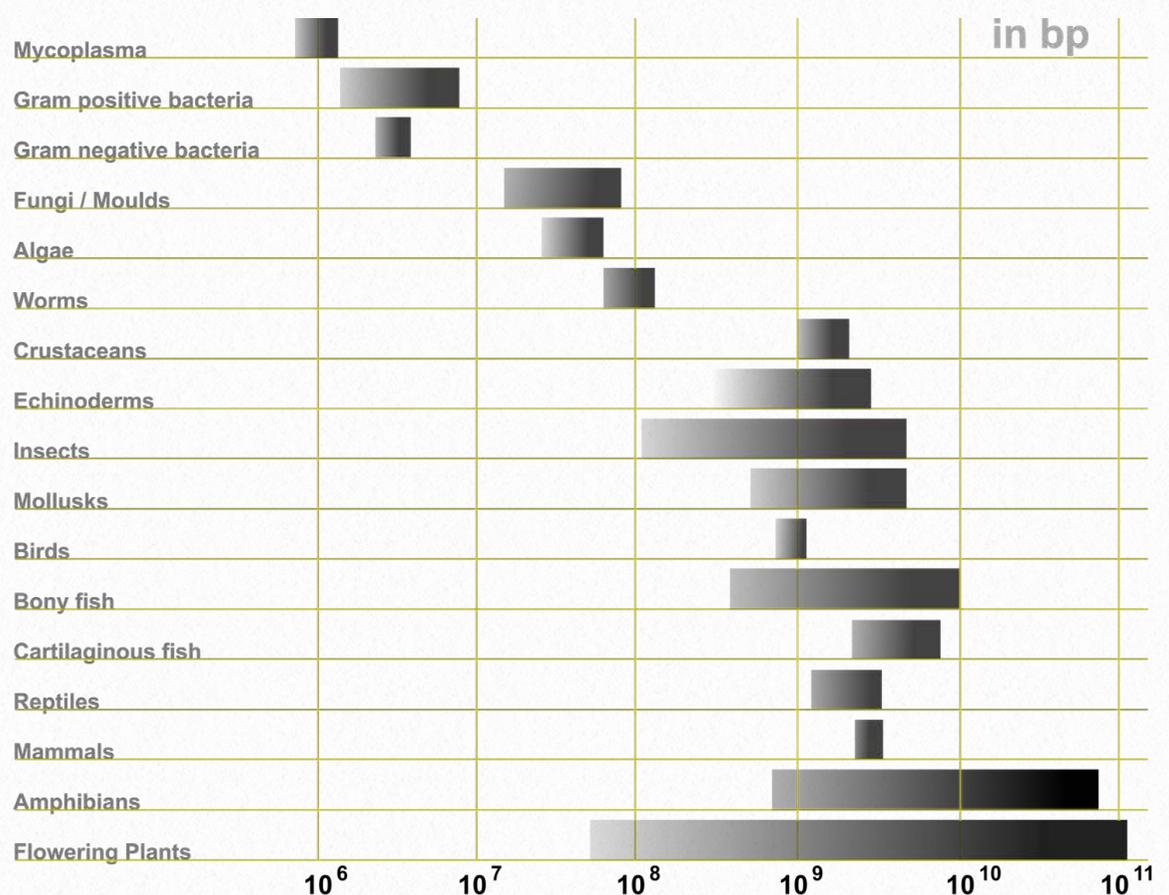


Figure 7.5 - Sizes of various genomes

Wikipedia

DNA may play important roles in evolution, as well as in regulation of gene expression.

## Introns

So, what is all the non-coding DNA doing there? We know that even coding regions in our DNA are interrupted by non-coding sequences called introns. This is true of most eukaryotic genomes. An examination of

genes in eukaryotes shows that non-coding intron sequences can be much longer than the coding sections of the gene, or exons. Most exons are relatively small, and code for fewer than a hundred amino acids, while introns can vary in size from several hundred base-pairs to many kilobase-pairs (thousands of base-pairs) in length. For many genes in humans, there is much more of intron sequence than coding (a.k.a. exon) sequence. Intron sequences account for roughly a quarter of the genome in humans.

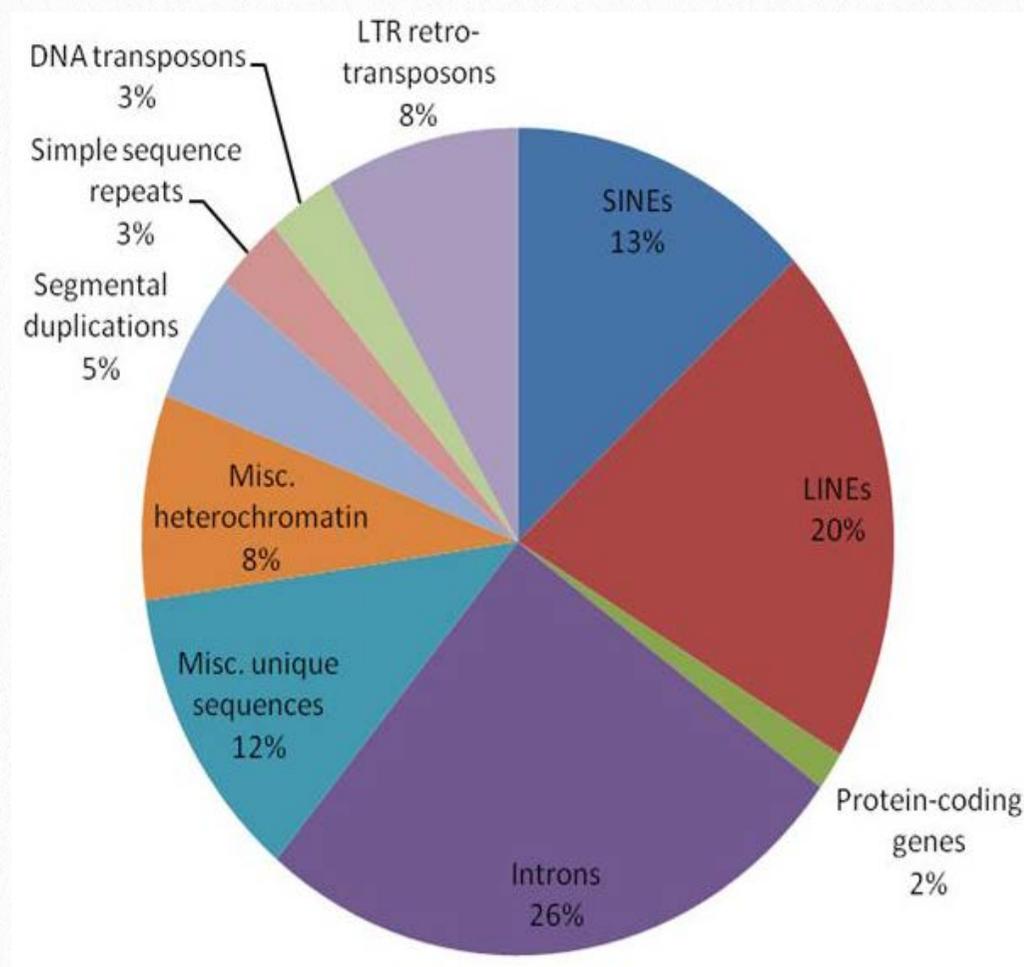


Figure 7.6 - Components of the human genome

Wikipedia

## Other non-coding sequences

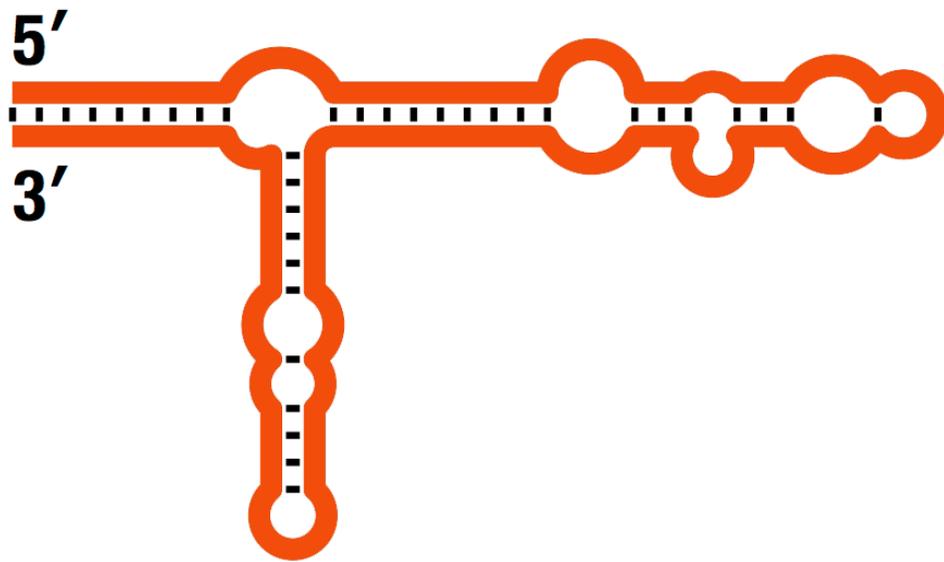
What other kinds of non-coding sequences are there? One function for some DNA sequences that do not encode RNA or proteins is in specifying when and to what extent a gene is used, or expressed. Such regions of DNA are called regulatory regions and each gene has one or more regulatory sequences that control its expression. However, regulatory sequences do not account for all the rest of the DNA in our genomes, either.

## Transposable sequences

Surprisingly, almost half of the human genome appears to consist of several kinds of re-

petitive sequences. Many of the repetitive sequences are known to be transposable elements (transposons), sections of DNA that can move around within the genome. Sometimes referred to as “jumping genes” these transposable elements can move from one chromosomal location to another, either through a simple “cut and paste” mechanism that cuts the sequence out of one region of the DNA and inserts it into another location, or through a process called retrotransposition involving an RNA intermediate.

**YouTube Lectures  
by Kevin  
[HERE](#) & [HERE](#)**



**Figure 7.7 - 5S rRNA structure**

Wikipedia

## LINES & SINES

There are millions of copies of each of two major classes of such transposable elements, the LINEs (Long Interspersed Elements) and SINEs (Short Interspersed Elements) in our genomes.

LINEs and SINEs are both a kind of transposable element called retrotransposons, sequences that are copied into RNA, then reverse transcribed back into DNA before being inserted into new locations. This movement is typically not sequence specific, meaning that the transposons can be inserted randomly in the genome, in many cases within coding regions. As might be expected, this can disrupt the function of the gene. Transposons may also insert within regulatory regions, and change the expression of the genes they control. As a major cause of mutation in genomes, transposons play an important role in evolution.

Finally, recent findings have shown that much of the genome is transcribed into RNAs, even though only about 2% encodes proteins. What are the RNAs that do not encode proteins? Ribosomal RNAs (Figure 7.7) and transfer RNAs, together with the small nuclear RNAs that function in splicing, account for some of these non-translated transcripts, but not all. The remaining RNAs are regulatory RNAs, small molecules that play an important role in regulating gene expression. As we understand more about genomes, it is becoming evident that the so-called "junk" DNA is anything but.

Graphic images in this book were products of the work of several talented students.  
Links to their Web pages are below

Click [HERE](#) for  
Aleia Kim's  
Web Page

Click [HERE](#) for  
Pehr Jacobson's  
Web Page

Click [HERE](#) for  
Penelope Irving's  
Web Page

Click [HERE](#) for  
Martha Baker's  
Web Page

Problem set related to this section [HERE](#)

Point by Point summary of this section [HERE](#)

To get a certificate for mastering this section of the book, click [HERE](#)

Kevin Ahern's free iTunes U Courses - [Basic](#) / [Med School](#) / [Advanced](#)

Biochemistry Free & Easy (our other book) [HERE](#) / [Facebook Page](#)

Kevin and Indira's Guide to Getting into Medical School - [iTunes U Course](#) / [Book](#)

To see Kevin Ahern's OSU ecampus courses - [BB 350](#) / [BB 450](#) / [BB 451](#)

To register for Kevin Ahern's OSU ecampus courses - [BB 350](#) / [BB 450](#) / [BB 451](#)

Biochemistry Free For All [Facebook Page](#) (please like us)

Kevin Ahern's [Web Page](#) / [Facebook Page](#) / Taralyn Tan's [Web Page](#)

Kevin Ahern's free downloads [HERE](#)

OSU's Biochemistry/Biophysics program [HERE](#)

OSU's College of Science [HERE](#)

Oregon State University [HERE](#)

Email [Kevin Ahern](#) / [Indira Rajagopal](#) / [Taralyn Tan](#)

# Transposon

To the tune of "Delta Dawn"  
**Metabolic Melodies** Website [HERE](#)

*(sung in a twangy, country style)*

Trans-po-son

All that movin' hither yon  
You're insertin' yourself in the chromosomes  
I guess it's why they say  
You're the jumpin' DNA  
Movin' and affectin' proteomes

You're carryin' a gene that gives assistance  
Helpin' tiny bugs to pass resistance  
To a drug that otherwise might kill 'em dead  
But with it they go on and live instead

McClintock studied you in *Zea maize*  
Marveling at your mysterious ways  
But the things you did went counter to what's known  
Somethin' like a genetic Twilight Zone

Transposon

All the battle lines are gone  
Over mechanisms Barbara did surmise  
It's good the truth came out  
There will never be a doubt  
She truly did deserve the Nobel prize

Transposon

All the battle lines are gone  
Over mechanisms Barbara did surmise  
It's good the truth came out  
There will never be a doubt  
She truly did deserve the Nobel prize

Lyrics by Kevin Ahern

No Recording Yet For This Song