HudsonAlpha
INSTITUTE FOR BIOTECHNOLOGY

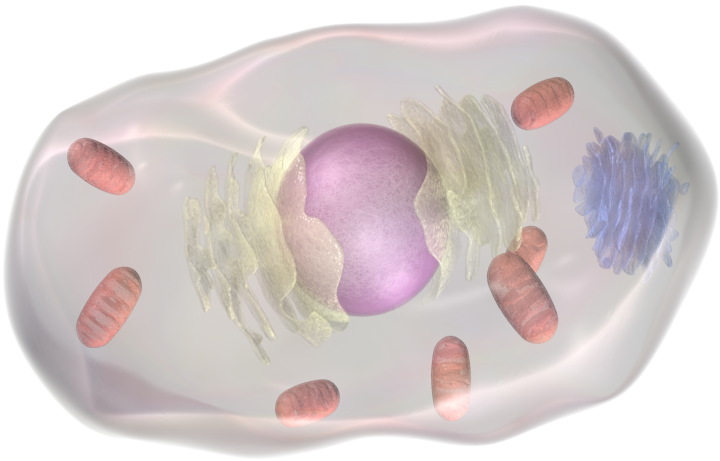
BIOTECH 101 – UNDERSTANDING THE BASICS

Genetics is at the forefront of investigations into human variation, disease and biotechnology. Newspapers, TV, magazines, radio and the internet have made genetics a high priority and almost every day there is an announcement trumpeting the discovery of a gene for disease "X" or a genetic test available to identify disorder "Y". How do we make sense of all the information and separate fact from hype? The answer is complex, in part because genetics is complex. The first step, however, begins with a solid understanding of the structure, function and malfunction of our genes. Education brings empowerment, and the ability to make informed decisions when faced with genetic questions.

A PRIMER ON THE BUILDING BLOCKS

Biotechnology

At its core, biotechnology is a field that uses biological processes, organisms or systems to develop products aimed to improve our life. The roots of biotechnology stretch back 7,000 years to the creation of bread, cheese wine and vinegar (which all depend on harnessing and modifying some biological process). The field has expanded dramatically over the thirty years, powered by our understanding of DNA.



Cell

A cell is the structural and functional unit of all known living organisms – the most basic unit or building block of life. Adults have an estimated 10 trillion cells in their bodies.

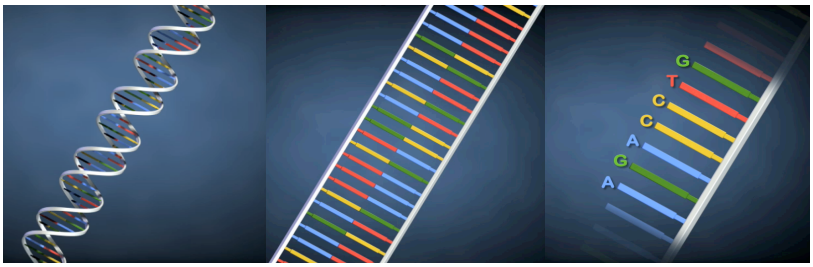
Nucleus

The nucleus is the command center of a cell, containing the cell's DNA. The nucleus also controls the cell's growth and division.

Chromosomes

Chromosomes are long strands of DNA, often wrapped tightly around proteins and condensed to fit into the cell nucleus. In humans, the nucleus contains two (2) sets of chromosomes, one set contributed by each parent. Each set contains 22 *autosomes* and one *sex chromosome* for a total of 23 chromosomes per set, or 46 chromosomes per cell. Males have both an "X" and "Y" sex chromosome, while females have two "X" sex chromosomes.

DNA

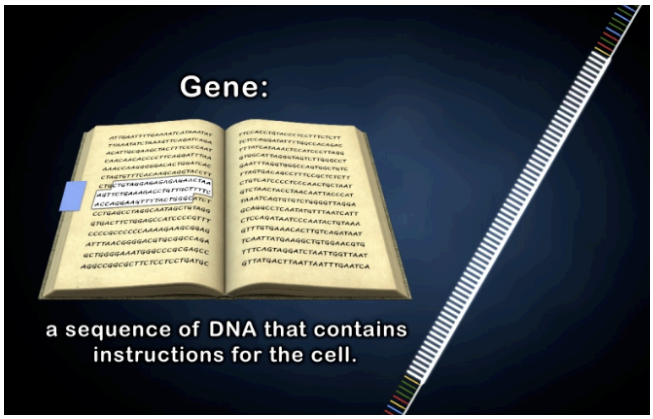


Deoxyribonucleic Acid (DNA) contains the genetic instructions that pass information about the organism from one generation to the next. Each chromosome consists of a long stretch of DNA, very tightly wrapped around a protein scaffold. The structure of DNA is similar to a twisted ladder (also known as a double helix). The "rungs" are composed of four building blocks called nucleotides. These nucleotides are Adenine, Thymine, Cytosine and Guanine, abbreviated as A, T, C and G. There are nearly 3 billion nucleotides in a complete set of human DNA. The *sequence* of DNA is the specific order of the nucleotides across a region of DNA.

More details: There are three components of a DNA molecule: the chemical base (A, T, G or C), deoxyribose (a sugar unit), and a phosphate group. The two strands of the DNA double helix are held together by hydrogen-bonding between the bases of each strand. The bases pair in a very specific manner, with A always pairing with T and G always matched with C. The DNA replicates by separating the two strands and creating a complementary copy of each, a process known as *replication.*

Genes

A gene is a specific stretch of DNA that provides instructions to the cell. Often, but not always, these instructions tell the cell how to assemble a certain protein. Genes help determine whether we will be male or female, brown or blue-eyed, tall or short. In altered forms (known as mutations) genes play major roles in the development of a wide variety of diseases and disorders. Genes are arranged along the chromosome in a linear fashion, each having a specific location. Only a minority of the human genome is composed of genes (about 2%). The majority of human DNA consists of regulatory and repetitive sequences, fragments of inactive genes and other DNA sequence with yet-unknown functions. Genes generally have a set of sequence-specific characteristics that allow them to be identified. Humans have approximately 20,000 genes.



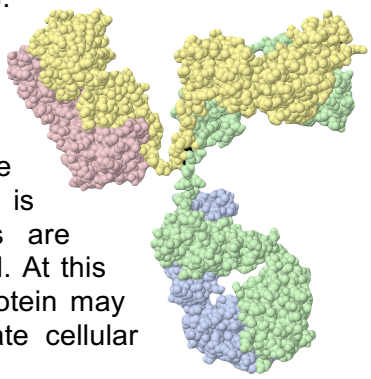
More details: The field of human genetics is dedicated to understanding the structure and function of our genes. Within this field are many areas of specialization: the study of chromosomes (cytogenetics); the study of the structure and function of individual genes (molecular and biochemical genetics); the study of the genome and its organization and function (genomics); the analysis and interpretation of various types of nucleotide and amino acid sequences, protein domains, and protein structures (bioinformatics and computational biology); the study of genetic variation in human populations (population genetics); the study of the genetic control of development (developmental genetics); and the application of genetics to diagnosis and patient care (clinical genetics).

RNA

Ribonucleic acid (RNA) is similar in structure to DNA, with some key differences, most notably that RNA is a single stranded molecule rather than a double stranded helix. There are several types of RNA, including messenger RNA (mRNA), which is copied from a DNA template: the RNA thus contains the DNA "message" used as the blueprint for a specific protein. The process of creating a mRNA strand from the DNA template is called *transcription*. This working copy leaves the nucleus and is transported to the cytoplasm of the cell. There, it directs the production of its specific protein.

Proteins

Proteins are composed of amino acid building blocks. There are twenty different amino acids used in the construction of the protein chain, also known as a *polypeptide*. Proteins are the structural and functional elements of our cells.



More details: The process of protein synthesis, called *translation*, takes place in the cytoplasm of the cell when a mRNA molecule interacts with a structure known as a ribosome. The mRNA sequence is read in three base segments, or *codons*. Each amino acid is represented by one of more possible triplet codons. Proteins are synthesized one codon at a time until a "stop" codon is encountered. At this point, the polypeptide chain is released from the ribosome. The protein may undergo additional modification and then is sent to the appropriate cellular compartment or secreted into the extracellular environment.

Gene Expression

Gene expression refers to the process where a gene is activated under certain conditions and "speaks out" producing a messenger RNA copy that directs the production of a protein. This is a highly specialized process and a number of different signals exist inside the cell to turn genes "on" or "off". Different cell types will activate subsets of genes depending upon their function and the signals they receive from other cells and the outside environment.

Genome

The genome represents all of the genetic information present in a cell. More precisely, a genome is the specific complement of DNA present on one complete set of chromosomes. Humans therefore have 2 sets of this information (one on the maternal chromosomes and another on the paternal chromosomes). This distinction is often overlooked when individuals speak about the genome of specific organisms.

The Human Genome Project

The effort to investigate the human genome in its entirety was called the Human Genome Project. This 13-year international effort included sequencing of the 3 billion nucleotide human genome, comparing it with the genomes of other organisms such as bacteria, yeast and roundworms, and addressing the ethical, legal and social issues arising from the study of the human genome.

Genomics

The study of the functions and interactions of all the genes in the genome, including their interactions with environmental factors. Thanks to current technology, the field of genomics is expanding rapidly allowing researchers to examine the connections between multiple genes.

GENETIC VARIATION & ITS IMPACT ON HEALTH & DISEASE

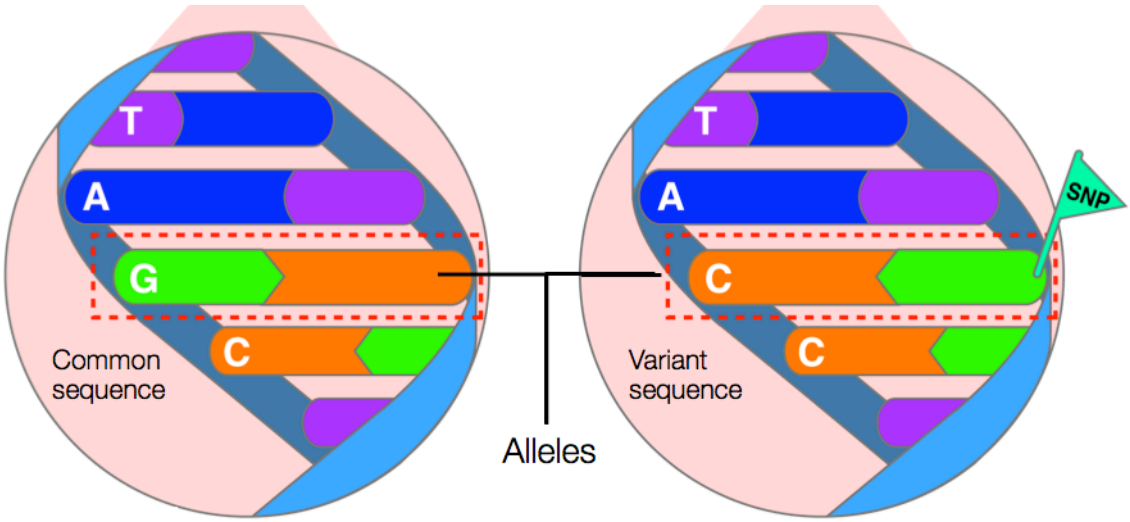
The sequence of DNA found on our chromosomes is more than 99% identical between any two humans. It is amazing to consider that the tiny amount of DNA variability plays such a large role in shaping our own uniqueness. Many DNA sequence differences have no effect on physical characteristics, whereas other differences are directly responsible for major, disease causing effects. In between these extremes, are the variations that result in small changes in anatomy, dietary intolerances, organ function, susceptibility to infection, and even some variability in personality, athletic aptitude, and artistic talent. Note that genetic information is not the sole determinant of these traits – the environment plays a major role as well.

Mutation

A ***mutation*** is defined as any change in the sequence or arrangement of DNA. In most cases the change is neutral in effect, but occasionally the change is harmful. In rare cases, the change is actually beneficial and confers some selective advantage to the organism. Mutations can alter the number of chromosomes in a cell, the structure of an individual chromosome, or the sequence of a stretch of DNA.

At any specific gene, variations due to mutation may exist, leading to the production of slightly different proteins. An ***allele*** is an alternative form of a given DNA sequence. You can think of this as a specific version or flavor of a gene, which may or may not have a physical or clinical impact.

These variations are not necessarily harmful (think of the different eye and hair colors). If the allele is present in at least 1% of the population, it is called a *polymorphism.* If the variation is due to a single nucleotide difference, the variant is called a *SNP – single nucleotide polymorphism.*



More details: Some alleles functions “dominantly”, in that they are sufficient to cause a specific physical trait regardless of the allele contributed by the other parent. Other alleles act “recessively” and require that both parents contribute similar alleles for the physical characteristic to be manifest.

Origin of Mutation

Mutations in genes occur primarily in one of two ways: errors introduced during the process of DNA replication, or failure to repair DNA that has been damaged in some way. DNA damage can be spontaneous, or in response to a physical or chemical agent (like ultraviolet light). The vast majority of all DNA mutations are recognized and repaired soon after they occur. If a mutation occurs in a subset of cells from a certain tissue - in the lung or breast, for example - this type of mutation is not passed on to the next generation. If, however, the mutation occurs in cells that will ultimately develop into eggs or sperm, the mutation becomes an inherited change that can be passed on to future generations.

A single nucleotide substitution can alter the mRNA code that directs the assembly of the protein. Such mutations may direct the addition of an incorrect amino acid to the growing protein or even lead to premature termination of protein construction. Mutations can also be caused by the insertion, inversion, fusion or deletion of DNA sequence. Some of these may involve only a few nucleotides while other insertions or deletions may be hundreds or even thousands of bases long.

Estimates of the rate of mutation in the human genome suggest each of us is born with approximately 60 genetic changes not present in either of our parents!

The Effect of Mutation on Protein Function

What are the basic ways that disease-causing mutations lead to physical manifestation of the disease? A genetic disease occurs when an alteration in the DNA of an essential gene changes the amount or function (or both) of the gene products (protein). This can occur in one of four possible ways:

1. partial or total loss of function of the protein – the mutation prevents the production of mRNA and subsequent protein
2. gain of function for the protein – the function may be enhanced (more is not necessarily better)
3. addition of a new function for the protein
4. alteration in the way the protein is regulated – leading to inappropriate activation at the wrong time or place

In addition, mutations can alter the way transcription or translation is regulated, the way the protein is modified, transported to the correct location, interacts with other molecules or is broken down. Depending upon the type of protein the DNA produces, the mutation may affect an enzyme, transport or storage machine, cell support molecule, regulatory factor, growth factor, hormone or cell signaling mechanism. Proteins carry out an incredible number of different functions and as a result, mutations in nearly every class of function can lead to a genetic disorder.

GENETIC DISORDERS

Nearly all disease is the result of the combined action of environment and genes. The relative genetic role, however, may be very small or very large. Among the disorders caused in whole or in part by genetic factors, three main types are recognized:

1. single-gene disorders
2. chromosome disorders
3. multifactorial disorders

Single Gene

Single-gene defects are caused by changes or *mutations* in single genes. Remember that for each gene, two copies are present (one on the chromosome inherited maternally, the other from the paternal contribution). The mutation may be present on only one of the copies of the gene, or on both copies. Single-gene disorders show very obvious physical symptoms and very often have recognizable inheritance patterns in the family. Most of these defects are rare (usually with a frequency much less than 1 in 1,000); however, as a group, single gene disorders are responsible for a significant proportion of disease and death, affecting 2% of the population over an entire life span. Some examples of single-gene disorders are Cystic Fibrosis, Huntington's syndrome, Tay-Sachs disease, Sickle Cell Anemia, Duchenne muscular dystrophy and Neurofibromatosis.

Chromosomal

In chromosomal disorders the defect is not due to a mistake in a single gene, but rather to an excess or deficiency of the genes contained in an entire chromosome or segment of a chromosome. For example, the presence of an entire extra copy of the 21st chromosome leads to a specific defect, Down syndrome. There are over 200 genes on chromosome 21 and individuals with Down syndrome have 3 normal copies of each gene. The increase in DNA results in more, mRNA being transcribed for many of these genes, leading to increases in protein levels. This is thought to disrupt the delicate balance maintained in cells, leading to the specific physical characteristics associated with Down syndrome.

As a group, chromosome disorders are quite common, affecting about 7 out of every 1,000 infants born. They are especially common among miscarriages, accounting for about half of all such cases. Clearly, large-scale disruptions in the genome are not usually compatible with life.

Multifactorial

Multifactorial disorders represent a number of both newborn and adult diseases. There appears to be no single error in the genetic information for many of these cases. Rather, the disease is the result of a combination of small variation in genes that together can produce or predispose to a serious defect. Often, these genetic factors work together with environmental risks. It is estimated that over 60% of the entire population is impacted by a multifactorial disorder. Examples include asthma, diabetes, hypertension, schizophrenia, coronary artery disease and cancer. Most of the disorders that are seen in adults are multifactorial. These are also called disorders with *complex inheritance*.