

Module 12: Basic Genetics

People are made of organs, organs are made of cells, cells are composed with genetic information. Genetics are important because we can earlier detect genetic diseases when a newborn has a genetic predisposition to disease; Ex. Cystic Fibrosis is a recessive gene. Complex traits are made up of more than one gene.

We can also use genetic information to create "custom drugs" in a practice called **pharmacogenetics**. Genetic factors account for 20% to 40% of inter-individual differences in metabolism and response. Genetic variants can alter the pharmacodynamics of a drug, potentially increasing efficacy or toxicity.

Somatic Cell Structure of Eukaryotes

All living things are made of cells. There are two types of cells:

- Somatic - differentiate to create organs required by an organism
- Germ (sex)

From a genetics POV the most important part of the cell is the nucleus, where the chromosomes are stored. Humans have 23 pairs of chromosomes, each set has 1 from the mother and 1 from the father. They are ordered from largest to smallest, (1 is largest, 22 is smallest). The 23rd is the sex chromosome, which is XX in females and XY in males.

DNA

Chromosomes are made up of Deoxyribonucleic acid (DNA) molecules. DNA molecule is packaged into chromosomes in a double helix ladder structure. The "rungs" of this ladder are bases or nucleotides. The two strands of the helix are complementary.

Nucleotides:

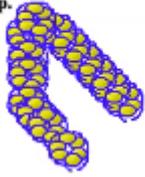
- Purines: Adenine and Guanine (A, G)
- Pyrimidines: Thymine and Cytosine (T, C)

DNA Packaging

DNA molecules are packaged in a complex manner into chromosomes.



(c): solenoids form a loop.



(d): idealized cross section of a chromosome: several solenoid loops.

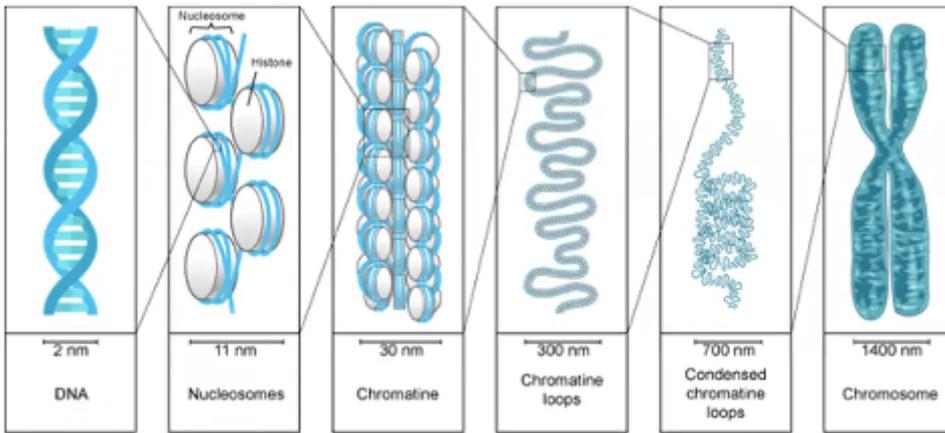


(e): portion of a chromosome without the protein scaffold.



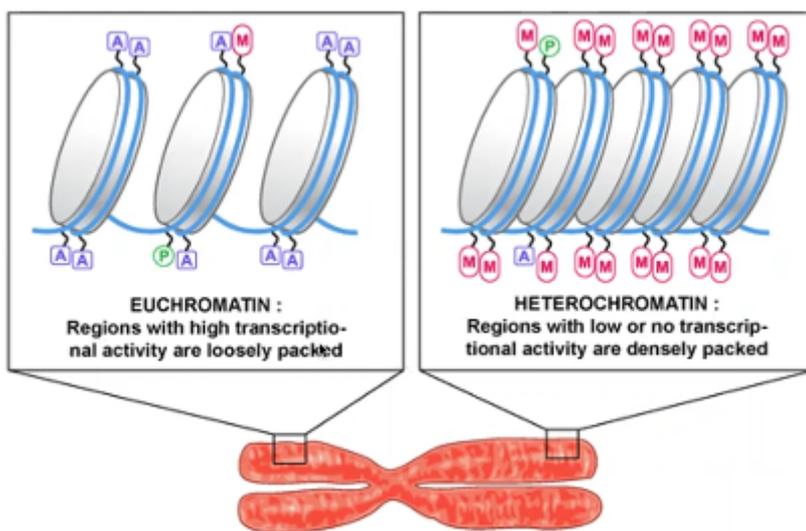
C

In each cell there's nearly seven ft of DNA.



<https://www.shmoop.com/dna/dna-packaging.html>

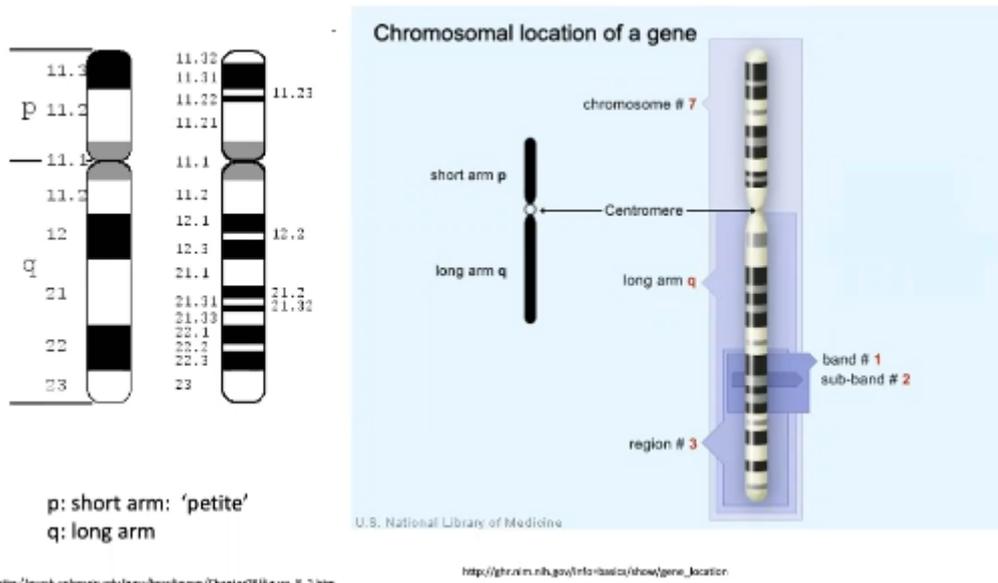
DNA is not tightly packed into chromosomes all the time, it only condenses when the cell is getting ready to divide.



○ : Histone ℓ : DNA A : Acetyl group P : Phosphoryl group M : Methyl group

Cytogenetic Location

Each human chromosome has a short arm ("p" for petite) and a long arm ("q" for queue), separated by a centromere. The ends of the chromosome are called telomere. Each chromosome arm is divided into cytogenetic bands that can be seen using a microscope and special stains. At higher resolutions sub-bands are seen within bands.



These bands are numbered p1, p2, p3... q1, q2... etc. Counting from the centromere out toward the telomeres. This is the process geneticists use to address the location of a band or a range of bands of a gene.

Ex. 7q31.2 indicates it is on chromosome 7, q arm, region 3, band 1, and sub-band 2. The ends of chromosomes are labeled p_{tel} and q_{tel}; 7q_{tel} refers to the end of the long arm on chromosome 7.

RNA

DNA codes for RNA which codes for proteins



- RNA is Ribonucleic acid
- Single strand
- Copies (transcribes) DNA to bring message to ribosomes for translation into proteins
- Uracil instead of thymine (U instead of T)

DNA "unzips" and allows RNA to transcribe its molecules, then is transported to ribosomes in the cytoplasm. RNA is synthesized from 5' to 3'. This strand is called the template strand of DNA, and is complementary to the newly synthesized RNA while the coding strand of DNA has the same

sequence as the new RNA. Either strand of the chromosome can take any role.

Proteins

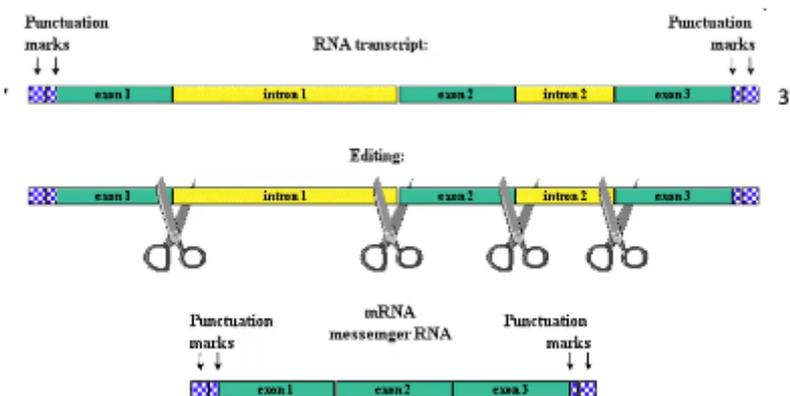
Proteins are often referred to as the machinery of the cell as they are responsible for nearly every task in a cell. There are 20 "letters" in the protein alphabet, called amino acids.

The DNA alphabet (A,C,G, T) is transcribed into mRNA and then translated to proteins using **codons**. Codons consist of 3 DNA bases or letters; $4^3 = 64$ combinations.

		Second Base				Third Base
		U	C	A	G	
First Base	U	UUU-Phe	UCU-Ser	UAU-Tyr	UGU-Cys	U
		UUC-Phe	UCC-Ser	UAC-Tyr	UGC-Cys	C
		UUA-Leu	UCA-Ser	UAA-Stop	UGA-Stop	A
		UUG-Leu	UCG-Ser	UAG-Stop	UGG-Trp	G
	C	CUU-Leu	CCU-Pro	CAU-His	CGU-Arg	U
		CUC-Leu	CCC-Pro	CAC-His	CGC-Arg	C
		CUA-Leu	CCA-Pro	CAA-Gln	CGA-Arg	A
		CUG-Leu	CCG-Pro	CAG-Gln	CGG-Arg	G
	A	AUU-Ile	ACU-Thr	AAU-Asn	AGU-Ser	U
		AUC-Ile	ACC-Thr	AAC-Asn	AGC-Ser	C
		AUA-Ile	ACA-Thr	AAA-Lys	AGA-Arg	A
		AUG-Met	ACG-Thr	AAG-Lys	AGG-Arg	G
	G	GUU-Val	GCU-Ala	GAU-Asp	GGU-Gly	U
		GUC-Val	GCC-Ala	GAC-Asp	GGC-Gly	C
		GUA-Val	GCA-Ala	GAA-Glu	GGA-Gly	A
		GUG-Val	GCG-Ala	GAG-Glu	GGG-Gly	G

You'll notice some of these result in the same amino acids.

Gene Structure



A gene, or a code for a specific protein sequence, does not lie in contiguous DNA - the RNA transcript is edited to create mRNA. The part that is removed is called the introns, and the parts kept are exons.

A gene is part of the DNA in the genome that is turned to RNA. They have several components:

- Promoter Region - Signals start of the gene. Often includes many A & T bases

- Exons - Translated from DNA to RNA then transcribed into proteins, expressed
- Introns - Cut out before protein transcription, unexpressed
- Splice site - junction of intron and exon, additional cutting may occur at these sites to produce different versions of proteins. Also called intervening sequence (IVS)

Intron and Exon sizes vary between genes and thus can contain different number of base pairs.

Non-coding DNA (ncDNA)

Defined as all of the DNA sequences within a genome that are not found within protein-coding exons; Both introns and IVS. Not represented within the amino acid sequence of expressed proteins. >98% of the human genomes is composed of ncDNA, and there are many different subtypes of ncDNA such as:

- Pseudogenes - regulatory DNA sequences, repetitive DNA sequences and sequences related to mobile genetic elements
- Sequences within genes:
 - Genes for non-coding RNA (e.g. tRNA, rRNA)
 - Untranslated components of protein-coding genes (introns, and 5' and 3' untranslated regions of mRNA)

Genetic Variability

There are genetic differences between humans. The different variations of a particular genetic location are called alleles. Since every chromosome is part of a pair each person has 2 alleles for every gene. The observed physical outcome of a gene is called the phenotype.

There are 3 billion base pairs in the human genetic sequence, taking up about 3GB of space on a computer. The DNA sequence across humans is 99.9% identical. That .1% results in differences across the genome.

Types of Genetic Variability

- Single base pair substitution - change in a single base pair.
 - Sometimes called point mutation, single nucleotide polymorphism or "SNP"
 - Polymorphism - DNA sequence variation that is common in the population (1% or higher). Alleles that occur with <1% frequency are called mutations
 - Also called single nucleotide variant or "SNV", which also applies to mutations.
- Single Nucleotide Variant (SNV)
 - Exonic
 - Silent/synonymous - it does not change amino acid due to degenerate code
 - e.g. AAA -> AAG still codes for Lysine
 - Nonsynonymous - changes amino acid sequence
 - Missense -> changes a single amino acid

- Non-Exonic - Intronic or intervening sequence
 - Can be a single base change, insertion, or deletion
 - Not part of the mRNA, so it will not alter the amino acid sequence

Sources of Genetic Variability

- Nucleotide repeats
 - Can be caused by disease such as Huntington's Disease
- Copy Number Variants
 - Gain and losses of large chunks of DNA sequence (10k - 5 million bases)
 - Structural change in genome
 - Found in exonic, intronic, and IVS
 - ~.4% of genomes of unrelated people differ with respect to copy number

Consequences of Genetic Polymorphism

- Change level of protein expression]
- Alter protein or make it non-functional
- Eliminate protein expression
- Create new protein
- Nothing

Representation of Genotypes

A person's genotype at a specific location in the genome consists of two alleles. There is not a standard way to record genotypes but some common conventions are concatenation, slashes and spaces:

11, 12, 22, 11 1/1,1/2,2/2,1/1 1 1, 1 2, 2 2, 1 1

For mutiallelic markers, markers with more than two alleles:

1. CCCACCCACCCA -- 3 repeats, 12 nucleotides
2. CCCACCCACCCACCCA -- 4 repeats, 16 nucleoties
3. CCCACCCACCCACCCACCCACCCA -- 6 repeats, 24 nucleotides

• Microsatellites/STRs/VNTRs may be recorded as

- Allele lengths (number of base pairs):
12/16 16/16 16/24
 - Number of repeats:
3/4 4/4 4/6
 - Consecutive numbers which may or may not correspond to allele size:
1/2 2/2 2/3
-

Revision #4

Created 30 August 2022 14:08:05 by Elkip

Updated 30 August 2022 22:53:39 by Elkip