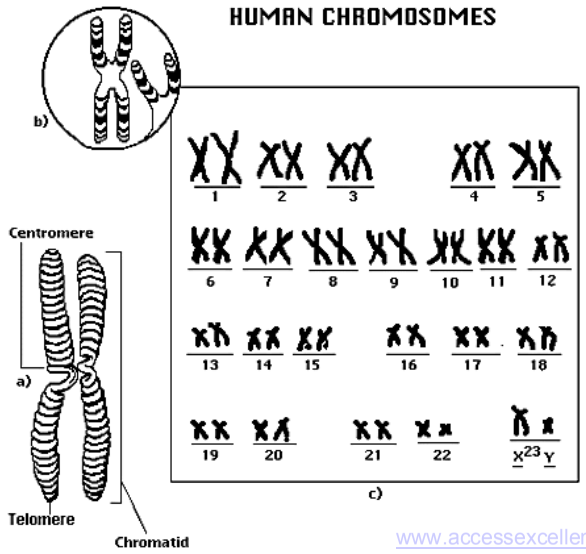
The genome

- The genome is distributed along chromosomes, which are made of compressed and entwined DNA.
- A (protein-coding) gene is a segment of chromosomal DNA that directs the synthesis of a protein.

The human genome

- The human genome is distributed along 23 pairs of chromosomes
  - 22 autosomal pairs;
  - the sex chromosome pair, **XX** for females and **XY** for males.
- In each pair, one chromosome is paternally inherited, the other maternally inherited (cf. meiosis).

Chromosomes

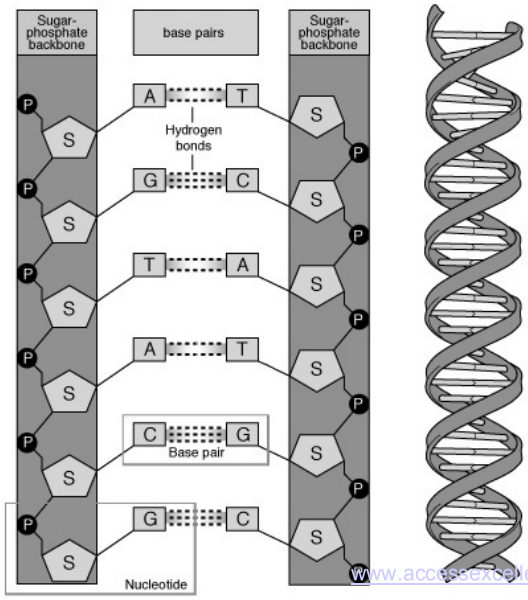


www.accessexcellence.com/AB/GG/

DNA structure

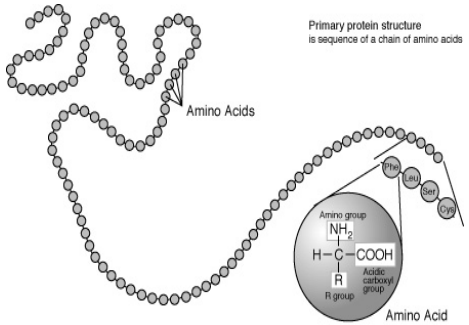
- A deoxyribonucleic acid or DNA molecule is a double-stranded polymer composed of four basic molecular units called nucleotides.
- Each nucleotide comprises
  - a phosphate group;
  - a deoxyribose sugar;
  - one of four nitrogen bases:
    - purines: adenine (A) and guanine (G),
    - pyrimidines: cytosine (C) and thymine (T).

DNA structure



www.accessexcellence.com/AB/GG/

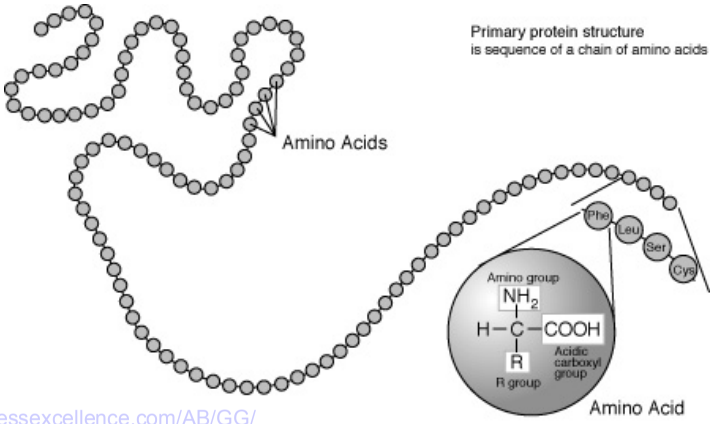
DNA replication



"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

J.D. Watson & F. H. C. Crick. (1953). Molecular structure of Nucleic Acids. *Nature.* **171**: 737-738.

Proteins

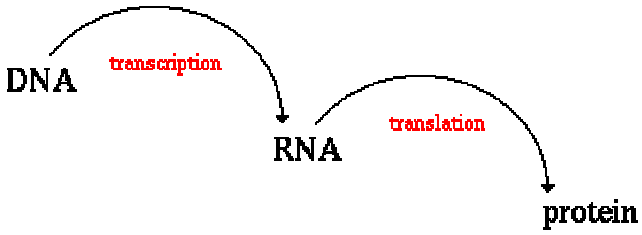


www.accessexcellence.com/AB/GG/

Proteins

- Proteins: large molecules composed of one or more chains of amino acids, polypeptides.
- Amino acids: class of 20 different organic compounds containing a basic amino group ($-NH_2$) and an acidic carboxyl group ($-COOH$).
- The order of the amino acids is determined by the base sequence of nucleotides in the gene coding for the protein.
- E.g. hormones, enzymes, antibodies.

Central dogma



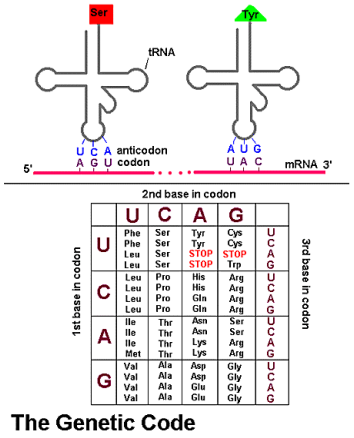
RNA

- A ribonucleic acid or RNA molecule is a nucleic acid similar to DNA, but
  - single-stranded;
  - ribose sugar rather than deoxyribose sugar;
  - uracil (U) replaces thymine (T) as one of the bases.
- RNA plays an important role in protein synthesis and other chemical activities of the cell.
- Several classes of RNA molecules, including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and other small RNAs.

The genetic code

- DNA: sequence of four different nucleotides.
- Proteins: sequence of twenty different amino acids.
- The correspondence between DNA's four-letter alphabet and a protein's twenty-letter alphabet is specified by the genetic code, which relates nucleotide triplets or codons to amino acids.

The genetic code



The Genetic Code

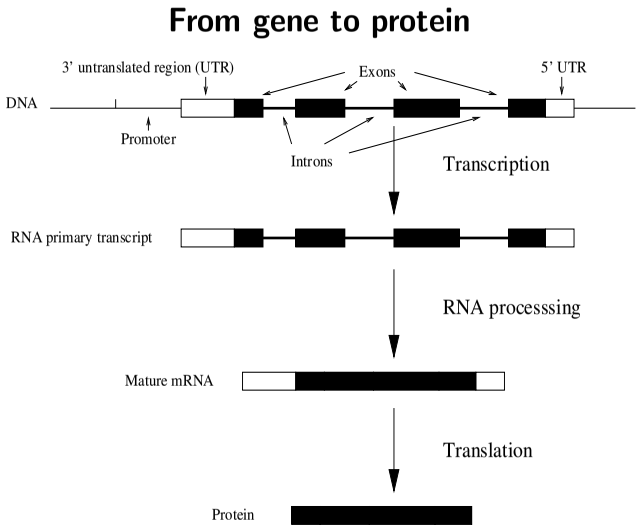
www.accessexcellence.com/AB/GG/

Start codon: initiation of translation (AUG, Met).
Stop codons: termination of translation.

Mapping between codons and amino acids is **many-to-one**: 64 codons but only 20 a.a..
Third base in codon is often redundant,
e.g., stop codons.

Exons and introns

- Genes comprise only about 2% of the human genome.
- The rest consists of non-coding regions
  - chromosomal structural integrity,
  - cell division (e.g. centromere)
  - regulatory regions: regulating when, where, and in what quantity proteins are made .
- The terms exon and intron refer to coding (translated into a protein) and non-coding DNA, respectively.



Differential expression

- Each cell contains a complete copy of the organism's genome.
- Cells are of many different types and states
  E.g. blood, nerve, and skin cells, dividing cells, cancerous cells, etc.
- What makes the cells different?
- Differential gene expression, i.e., when, where, and how much each gene is expressed.
- On average, 40% of our genes are expressed at any given time.

Alternative splicing

- There are more than 1,000,000 different human antibodies. How is this possible with only ~30,000 genes?
- Alternative splicing refers to the different ways of combining a gene's exons. This can produce different forms of a protein for the same gene.
- Alternative pre-mRNA splicing is an important mechanism for regulating gene expression in higher eukaryotes.
- E.g. in humans, it is estimated that approximately 30% of the genes are subject to alternative splicing.

Genotype and phenotype

- The term allele refer to a DNA sequence variant for a particular gene.
- The term genotype refers to DNA variants at a gene or set of genes for a given individual.
- The term phenotype refers to an observable trait or set of traits in an individual.
  E.g. blood pressure, disease status.

Alleles: Different versions of the genetic code at a particular locus

- **Types of Alleles (2 of many)**
  - **Single Nucleotide Polymorphism (SNP)**
    Allele 1: …TCGTGGCAGGTATTAGC…
    Allele 2: …TCGTGTCAGGTATTAGC…
  - **Short Tandem Repeat Polymorphism (STRP)**
    Allele 1: …AACGTCGAGAGAGACAGGTATTAGC…
    Allele 2: …AACGTCGAGAGAGAGACAGGTATTAGC…
    Allele 3: …AACGTCGAGAGAGAGAGACAGGTATTAGC…
    …

~~Disease Gene~~ vs. Disease Related Allele(s)

- **E.g., the PAH Gene codes for the enzyme Phenylalanine Hydroxylase**

SNP Markers

Single Nucleotide Polymorphism (SNP)

Allele 1: …TCGTGGCAGGTATTAGC…

Allele 2: …TCGTGTCAGGTATTAGC…

Alleles: Different versions of the genetic code at a particular locus

Microsatellite Markers

Short Tandem Repeat Polymorphisms (STRPs)

Repeating di-, tri-, tetra-, or penta- nucleotide units

E.g., a $(CA)_n$ di-nucleotide repeat

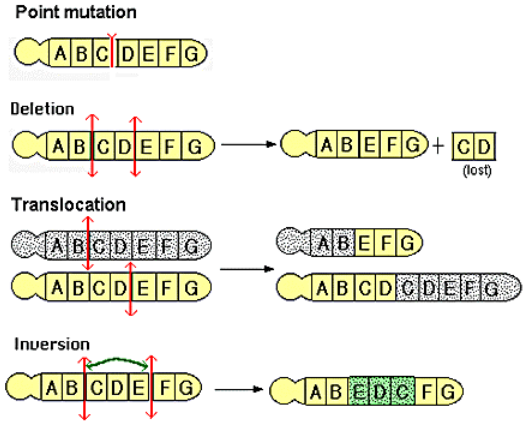
Allele 1: …AACGTCCACA…CAGAGGTATTAGC…
Allele 2: …AACGTCCACA…CACAGAGGTATTAGC…
Allele 3: …AACGTCCACA…CACACAGAGGTATTAGC…
…

Genetic diseases

- Genetic disease: the susceptibility or resistance to the disease is influenced by variants in the DNA sequence (alleles of the disease genes) that are transmitted from parent to offspring.

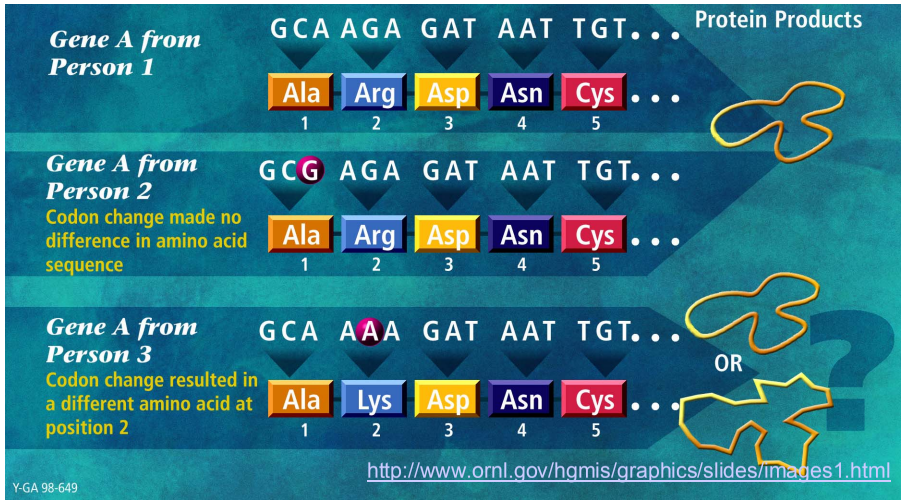
- E.g. Alzheimer's disease, asthma,cancer, color blindness, diabetes, hemophilia, sickle cell anemia, etc.

Chromosome mutations

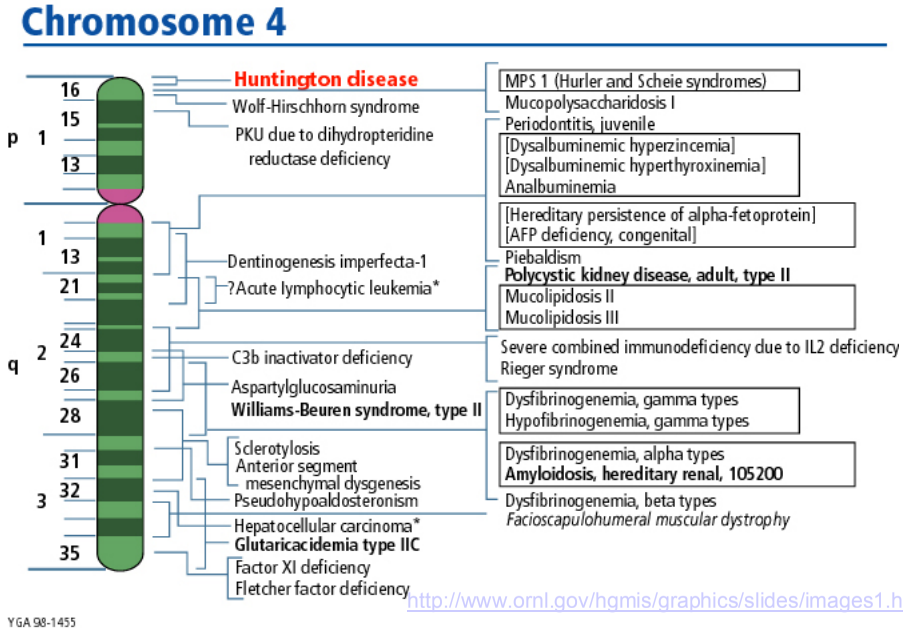


www.accessexcellence.com/AB/GG/

From DNA to proteins



http://www.ornl.gov/hgmis/graphics/slides/images1.html



http://www.ornl.gov/hgmis/graphics/slides/images1.html

Mendelian traits

- Usually influenced by a single gene.
- Usually a simple deterministic relationship between genotype and phenotype:
  0/1 penetrances.
- E.g.
  - Cystic fibrosis (autosomal recessive);
  - Hemophilia (X-linked recessive);
  - Huntington's disease (autosomal dominant);
  - Sickle cell anemia (autosomal recessive).

Sickle cell anemia

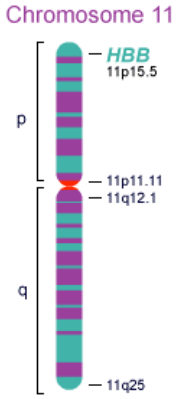
- **Phenotype:** Abnormal type of hemoglobin, HbS. Hemoglobin is a protein found in red blood cells and involved in oxygen transport. Individuals with the disease tire easily and often develop heart failure.

- **Molecular pathology:** Point mutation in the hemoglobin beta gene (HBB) found on the short arm of chromosome 11, 11p15.5.
  The amino acid Glu (Glutamic acid) is replaced by Val (Valine) in the in the HbA beta-chain
  → mutant HbS molecule;
  → sickle-shaped red blood cells.

Sickle cell anemia

The mutant hemoglobin HbS causes red blood cells to assume a sickle shape, become hard and sticky. This disrupts blood flow.

Sickle cell anemia



HBB Sequence in Normal Adult Hemoglobin (Hb A):

Nucleotide	CTG	ACT	CCT	GAG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Glu	Glu	Lys	Ser
	3			6			9

HBB Sequence in Mutant Adult Hemoglobin (Hb S):

Nucleotide	CTG	ACT	CCT	GTG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Val	Glu	Lys	Ser
	3			6			9

http://www.ornl.gov/hgmis/posters/chromosome/hbb.html

Sickle cell anemia

- **Mode of inheritance:** Autosomal recessive (gene on chr. 11).

  An individual is affected iff he/she has two copies of the mutant allele for HBB.

- **Prevalence**
  - Overall US population: 1/72,000.
  - Higher among African Americans: 1/500 are affected and1/12 are heterozygous carriers.

SCA and malaria

- The high incidence of SCA in certain populations is related to another disease: malaria.
- Heterozygotes for SCA are resistant to malaria.
- Malaria is caused by four protozoan parasites of the genus *Plasmodium*.

  P. falciparum is the most widespread and dangerous.
- The parasites are transmitted from person to person by female mosquito. They multiply in infected individuals and break down red blood cells.
- Worldwide incidence: each year 300-500 million cases, 1 million deaths (mainly sub-Saharan Africa).
- Symptoms: fever, headaches, etc.

Hemophilia A

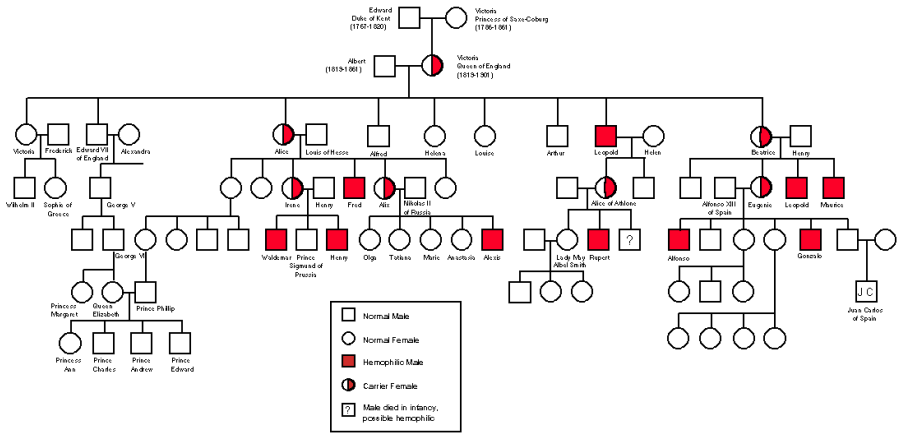
- **Phenotype:** Clotting factor VIII is reduced or missing → inability to properly form blood clots.

- **Molecular pathology:**
  - Different types of defects in the FVIII gene.
  - The FVIII gene is located in the most distal band of the long arm of the X chromosome, Xq28. The gene contains 26 exons (24 with length 69 to 262 bp; exons14 and 26 contain 3106 and 1958 bp, resp.).

Hemophilia A

- **Mode of inheritance:** X-linked recessive.
  - Males are affected iff they have one copy of the mutant allele;
  - Females are affected iff they have two copies of the mutant allele.
- **Prevalence:** 1/5,000 in males, much lower in females.

The royal disease



http://ublib.buffalo.edu/libraries/projects/cases/hemo.htm

Complex traits

- Multiple genes, multiple alleles of a gene.
- No clear deterministic relationship between genotype and phenotype:
  - unknown penetrances, usually not 0/1;
  - incomplete penetrances and phenocopies.
- Unknown population genetic model, non-random mating.
- Genetic heterogeneity.
- Diagnostic uncertainties (e.g., late onset).
- Gene-gene interactions.
- Gene-environment interactions.

  E.g. Alzheimer's disease, asthma, diabetes, hypertension, etc.

Type 1 diabetes

- Type 1 or Insulin Dependent Diabetes Mellitus (IDDM).
- **Phenotype:** the body does not produce any insulin, insulin is needed to convert sugar into energy.
- **Autoimmune disease:** T-cell destruction of pancreatic beta-cells which produce insulin.
- **Prevalence:** 300,000-500,000 in the US.

Type 1 diabetes

- **Mode of inheritance: complex.**
- 20 or so genes (most yet to be identified) are believed to influence susceptibility to IDDM -- the exact combination of genes differing with the individual.
- The major histocompatibility complex (MHC) of genes (called HLA in humans) may account for about half of the total genetic risk of developing IDDM.
- Environmental factors.