03-621 Week 2

Advanced Quantative Genetics

Aidan Jan

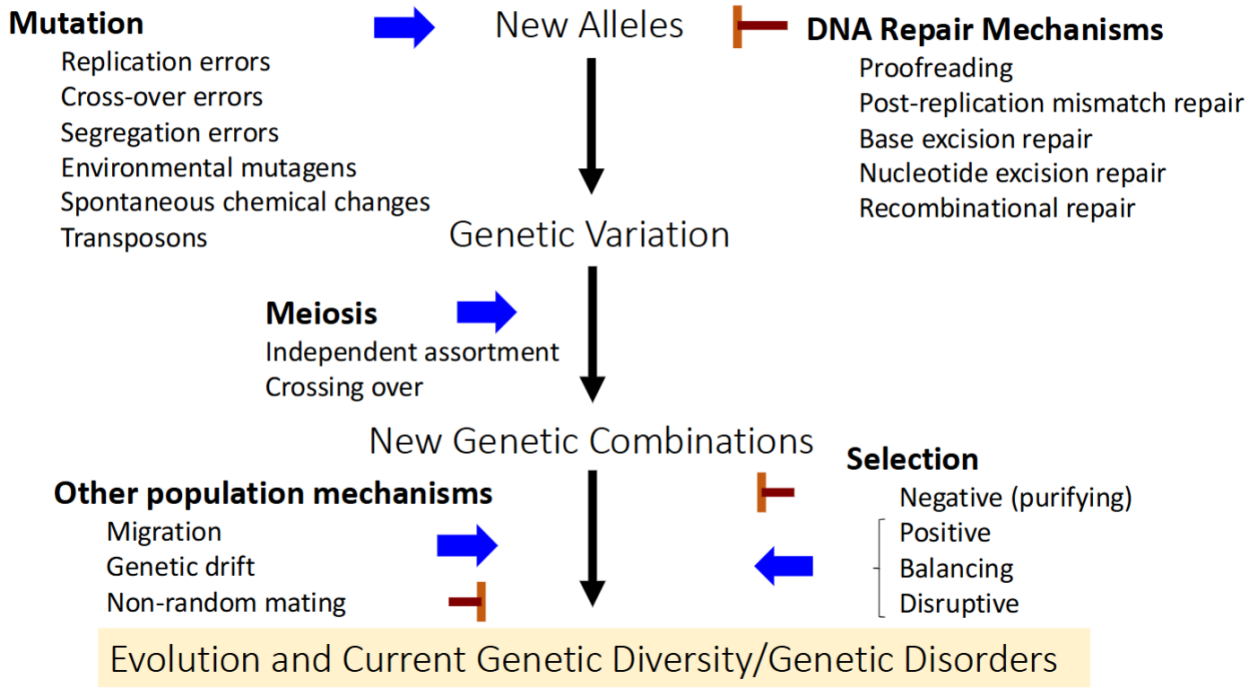
January 23, 2026

Analysis of Genetic Variation

What causes genetic variation in our genome?

- Base substitutions
  1. Errors during DNA replication
  2. Endogenous chemical damage to DNA
- Deletions, insertions, and duplications (indels)
  1. Errors during DNA replication
  2. Errors during crossing over
  3. Transposable elements

In Germline Cells



Causes of Base Substitutions

1. Errors during DNA replication
    - Nucleotide mis-incorporation by DNA polymerase
    - Frequency 10^{-3} (6000000 mistakes per genome replication)
    - A mis-incorporated nucleotide becomes a heritable mutation after a subsequent replication without repair, as in this case.
2. Endogenous Chemical Damage to DNA
    - 20000 - 100000 events per day in each of our cells.
    - E.g., **Hydrolysis** Deamination of Cytosine. (Cytosine hydrolyzes into Uracil)
    - E.g., **Oxidation** Guanine oxidizes into 8-oxoguanine, which results in a Hoogsteen base pair with Adenosine.

Causes of Deletions, Insertions, and Duplications

1. Errors during DNA replication
    - Slippage causes insertion or deletion
2. Errors during crossing over
    - **Inter**chromosomal. If crossing over happens right after the promotor of a gene and right before the terminator, then one chromosome inherits zero copies and the other inherits two copies of the gene.
    - **Intra**chromosomal. If a section in one chromosome crosses over with itself in the same direction, an entire gene may be cleaved from the DNA, resulting in a deletion.
    - **Intra**chromosomal, inversion. If a section in one chromosome crosses over with itself in the opposite direction, an entire gene may be inverted.

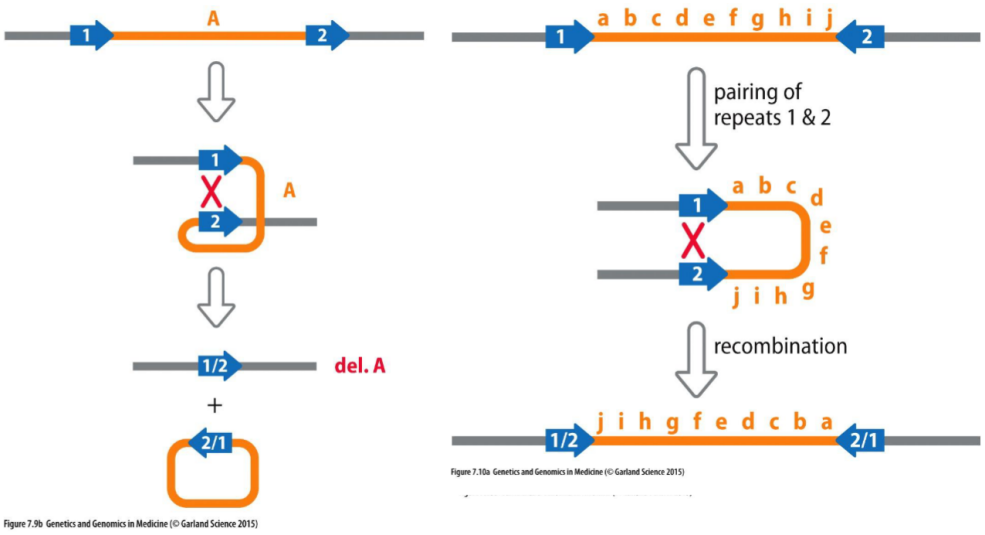


Figure 7.9b Genetics and Genomics in Medicine (© Garland Science 2015)

Figure 7.10a Genetics and Genomics in Medicine (© Garland Science 2015)

3. Transposable Elements
    - Can jump into and disrupt genes (large insertions)
    - Can mediate exon shuffling
    - Can serve as homologous sequences

Summary: Where do Mutations Come From?

Great majority arise from *endogenous* sources:

- Errors in normal cellular processes
  - DNA replication
  - Crossing over
  - Chromosome segregation
  - DNA damage repair
- Spontaneous chemical changes in DNA
- Movement of transposable elements

Occasionally from *external* sources

- UV light
- Ionizing radiation
- Chemicals

Mutations and Polymorphisms

- **Mutations** are heritable changes in DNA sequences (new variants, i.e., alleles)
- **Naturally occurring alleles in populations:**
  - **Common** sequence variants ($> 1\%$ frequency) are called **polymorphisms** (probably ancient mutations and either selected for or neutral).
  - **Rare** sequence variants ($< 1\%$ frequency) that are clearly different from a normal, functional allele (considered the "wild type") are often called **mutations** (probably recent origin and selected against)

How do we Study Human Genetic Variation

Approaches:

1. Classical studies of Mendelian traits and diseases
2. Personal and Population-Based Genomic Sequencing
   - Obtain and compared genome sequence from thousands of individuals made possible by post-genome next generation sequencing (NGS)

Goals:

1. Comprehensive catalog of normal human DNA variation
2. Catalog genomes in tumors and individuals with genetic disorders
3. Correlate DNA variation with phenotype to identify disease markers

The 1000 genomes project focused on whole-genome sequencing to catalog $> 99\%$ of common variation.

- 26 populations, 2504 individuals, and 88×10^6 variants.
- Millions of variant sites per individual genome
- Average "Singleton" variants per individual in the thousands.

The Wellcome Trust UK10K Project focuses on exome sequencing (just exons)

- 10000 people in the UK with closely monitored phenotypes including trios and twins
- 4000 non-disease controls
- 6000 cases of severe conditions

SNPs

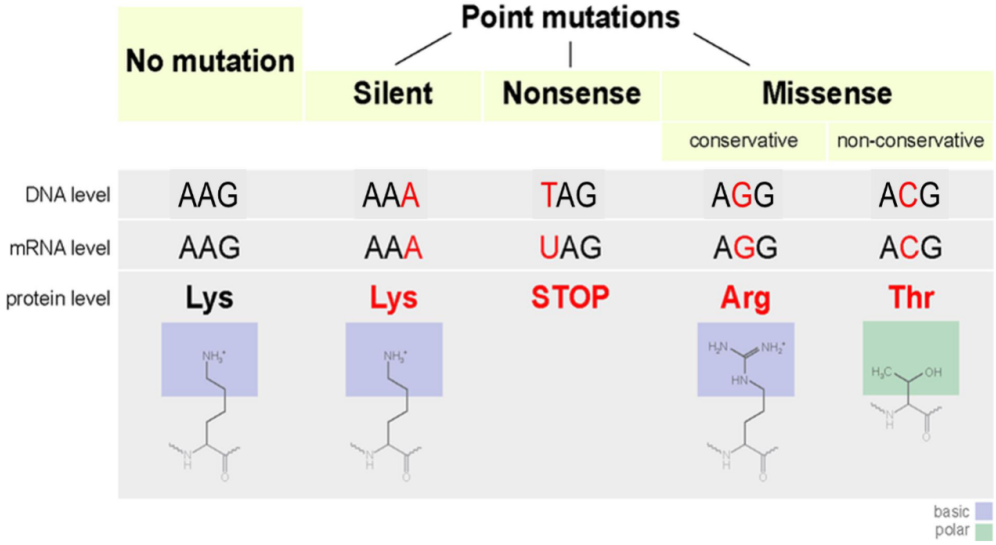
Single nucleotide variants and polymorphisms (SNPs) are the most common type of genetic variation in the human genome.

- A SNP is a change in one nucleotide. (e.g., substitution, insertion, deletion)

DNA is transcribed into proteins through codons (groups of three DNA base pairs), where every DNA sequence of length three represents a different protein.

- Therefore, a substitution mutates one amino acid.
- An insertion or deletion changes the *reading frame* during transcription, and results in every amino acid after the mutation being different. This is much more devastating than a substitution.

Types of Point Mutations



How Mutations Affect Gene Function