Human Genetic Variation

Genetics in Medicine - 0504321

2022-2023 Second Semester

Dr. Osama Alsmadi

Chromosomes, Genes & DNA

- Somatic cells are diploid - 46 chromosomes
  - 22 pairs autosomes; 1 pair sex chromosomes
- Each pair of autosomes is homologous
  - Contains the same genes in the same order
  - 1 is maternal, the other is paternal
- Chromosome are composed of deoxyribonucleic acid (DNA)
  - Genome contains 3 billion base pairs (haploid)
  - ~1% encode proteins
- Genes are located on chromosomes

Genomic Advancements

- The last 10-20 years has seen advances of ‘genomic technologies’

- Next Generation Sequencing (NGS)
  - Whole Genomes
  - Whole Exomes (WES)
  - Targeted Gene panel
- Whole Genome Sequencing (WGS)
  - Decrease in size of technology
  - Improvement in IT and bioinformatics
  - Decrease in genomic technology costs

What is Genetics?

Genetics refers to the study of individual genes and their roles in inheritance

There are 3 Billion DNA base pairs in the human genome

What is Genomics?

Genomics refers to the study of the entire genes, their interactions and functions

Your **genome** is one whole set of all your genes plus all the DNA between your genes.

There are around **20,000** genes in your **genome**

Germline Meanings:

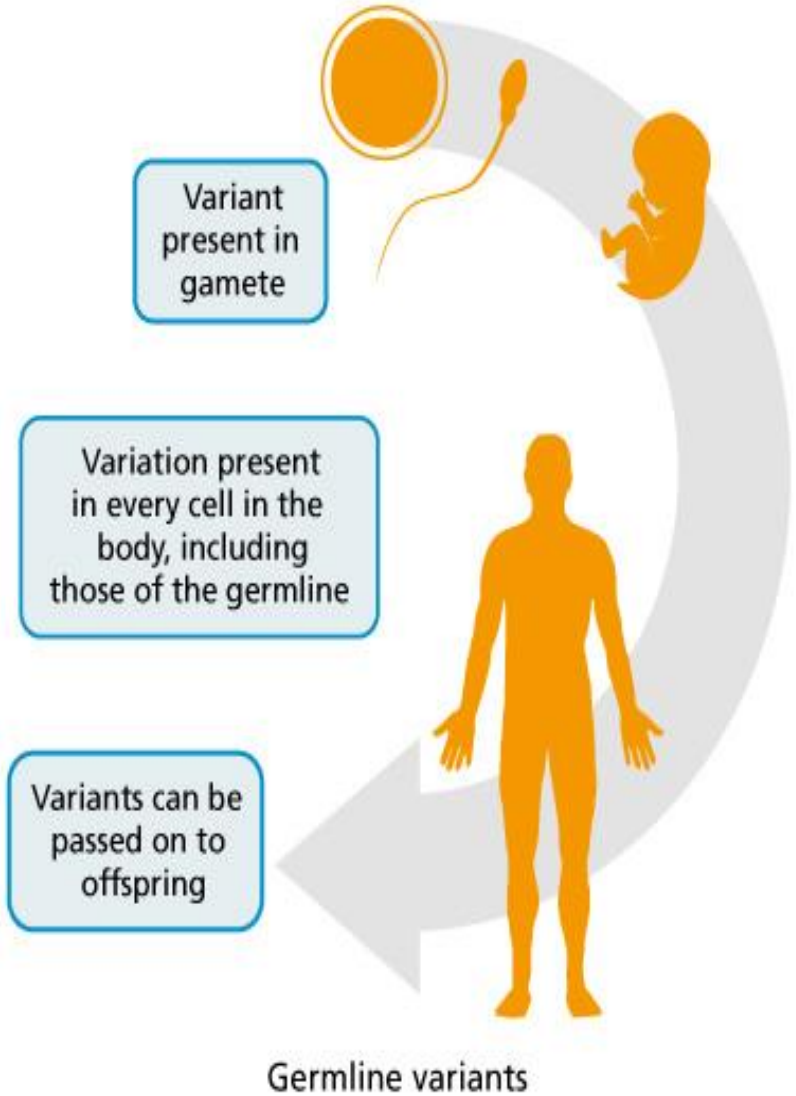
- In biology and genetics, the germline is the sequence of germ cells with their genetic material that get passed to the offsprings.

- Reproductive cells (sperm or the egg), are part of the germline.

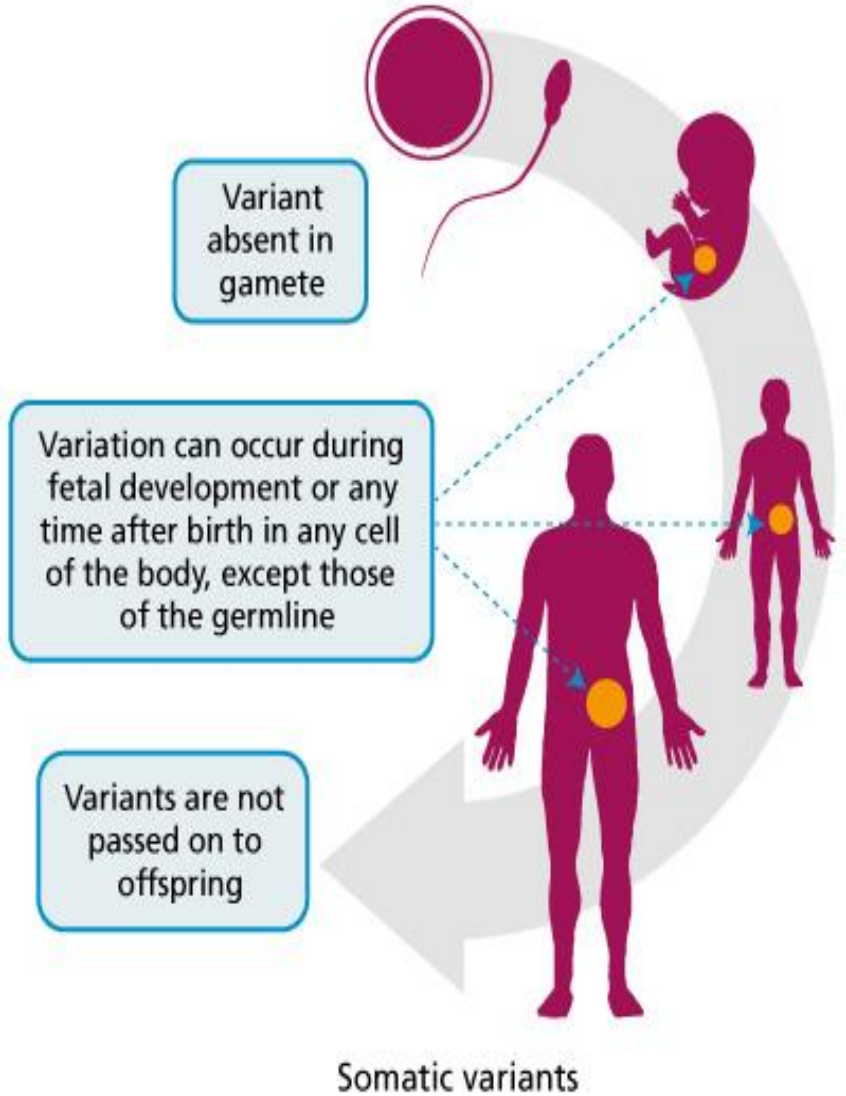
- Cells that are not in the germline are called **somatic cells**. Example cells of the liver/skin.

- If there is a mutation in the germline, it will be/can be passed to offspring, however changes in a somatic cell it won't.

Germline/Somatic Variants



Germline variants



Somatic variants

Genetic diseases Classification

Three Groups:

- 1. Multifactorial inheritance (polygenic)
- 2. Monogenic (Mendelian)
- 3. Chromosomal aberrations

Disorders with multifactorial inheritance (polygenic)

- influence of multiple genes plus interplay with environmental factors
- relatively frequent
- Diabetes mellitus
- Hypertension
- Certain congenital heart diseases
- Some types of cancer (ovarian, breast, colon)

- Can run in families-
  - 1st degree relatives about 5-10%;
  - 2nd degree relatives 0.5-1%

Monogenic (Mendelian) Disorders

- Mutation of one gene: Mendelian inheritance
- There are more than 5,000 diseases

- Autosomal dominant
- Autosomal recessive
- X-linked

Autosomal Dominant Disorders

- Both Homozygotes And Heterozygotes Are Affected
- Usually Heterozygotes (Inherited from one parent)
- Males And Females are Affected
- Transmitted From One Generation To The Other at 50% rate

Autosomal recessive

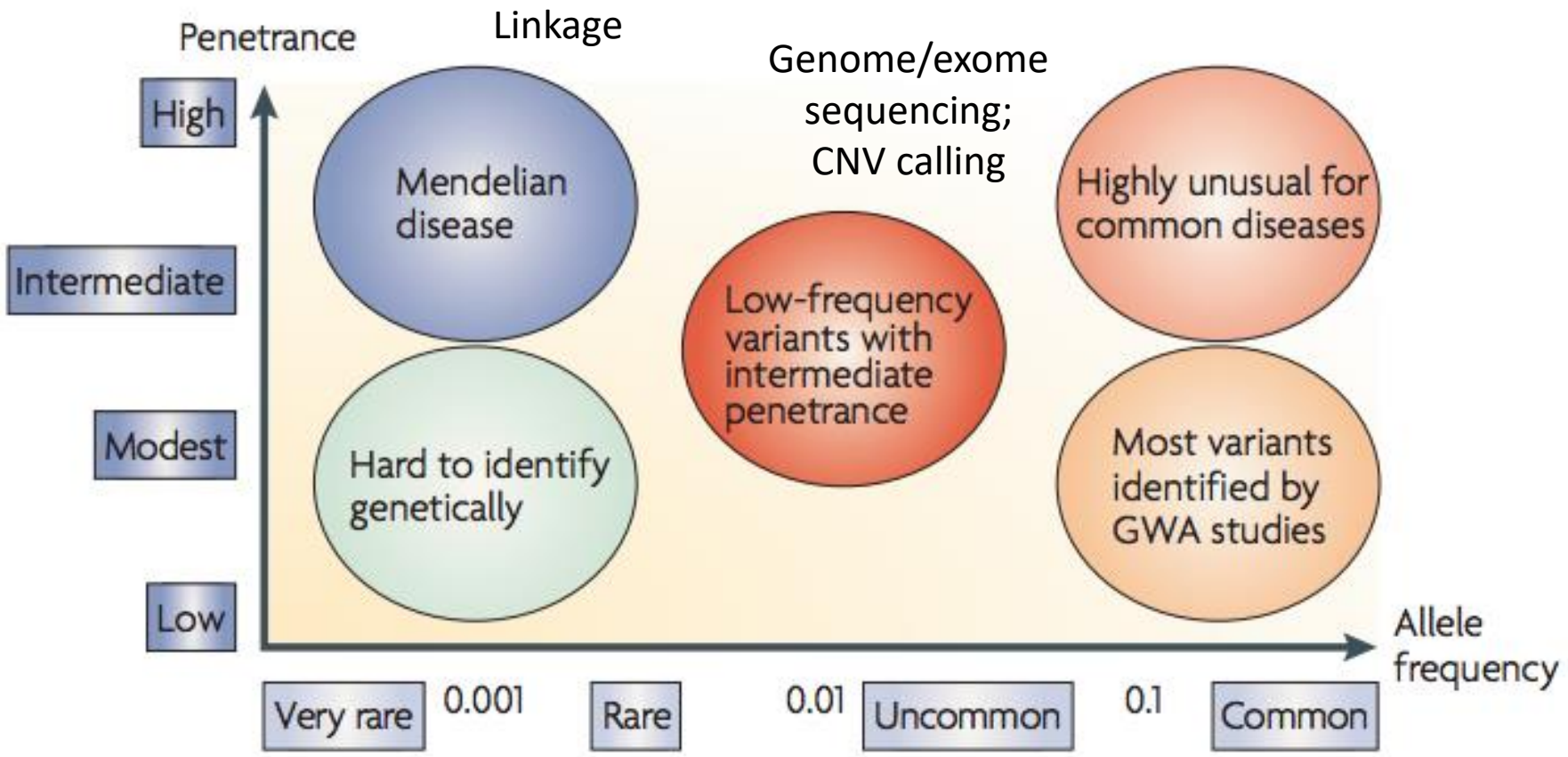
- The majority of Mendelian disorders
- Homozygotes are affected, heterozygotes (parents are carriers)
- 25% of descendants can be affected
- **if the mutant gene occurs with low frequency - high probability in consanguineous marriages**
- Onset of symptoms often in childhood
- Frequently enzymatic defect

X-linked diseases

- Transmitted as heterozygous from mother to sons
- Daughters: 50% carriers, 50% healthy
- Sons: 50% diseased & 50% healthy

- Hemophilia A (defect of Factor VIII)
- Hemophilia B (defect of Factor IX)
- Muscle dystrophy (Duchen disease)

Different Genotyping Methods for Different Types of Variants



(McCarthy, Nat Review Genet, 2008)

Where does Variation come from?

- Mutation
  - random changes to DNA
    - errors in mitosis & meiosis
    - environmental damage
- Reproduction
  - mixing of alleles
    - recombination of alleles
      - new arrangements in every offspring
    - new combinations = new phenotypes
  - spreads variation
    - offspring inherit traits from parent

Significance of Mutations

- Most mutations are neutral – have little or no effect on the expression of genes or function of proteins
- Harmful mutations – leads to defective proteins – disrupt normal biological functions
- Cause genetic disorders
- Associated with many types of cancer (somatic changes)

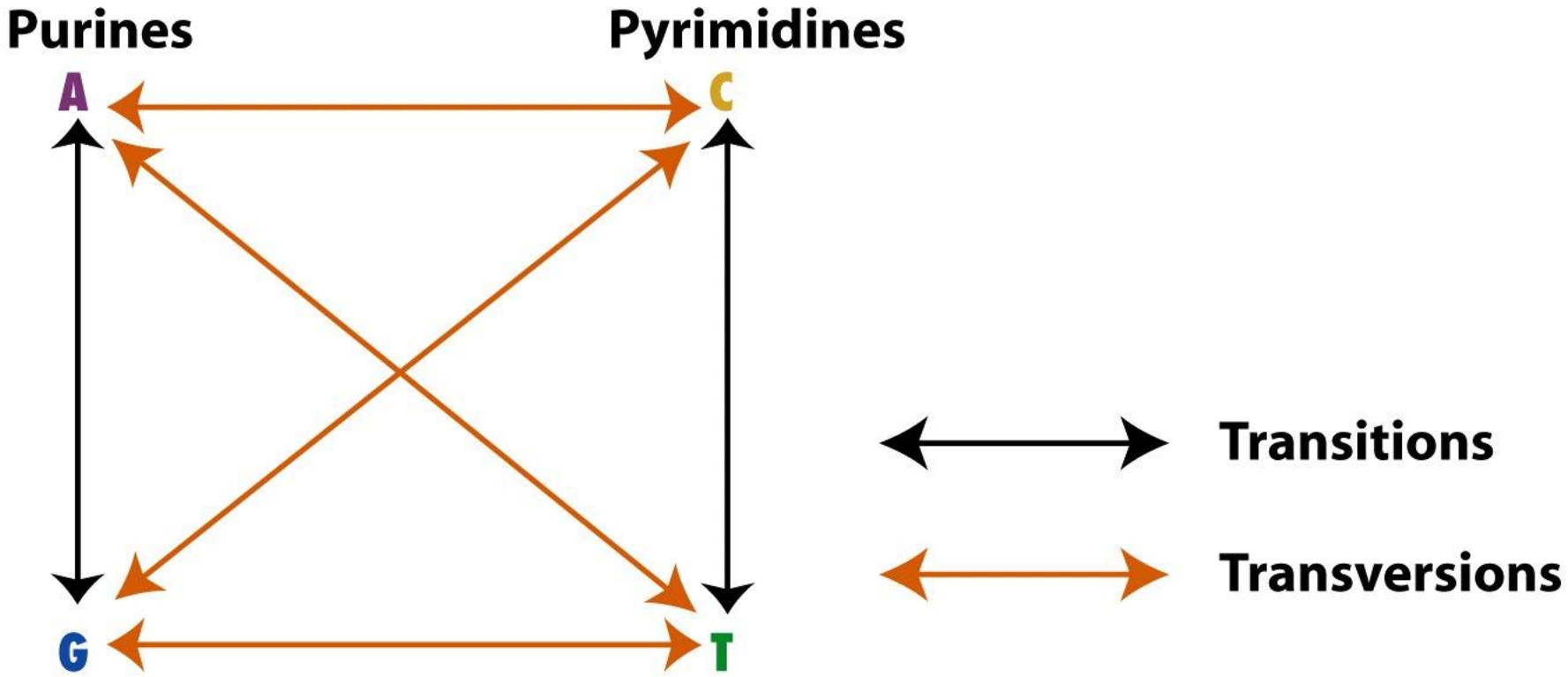


Figure 5-4 Evolutionary Analysis, 4/e

Transitions are more common than transversions because DNA repair enzymes can recognize wrong insertion representing a a transition better than a transversion

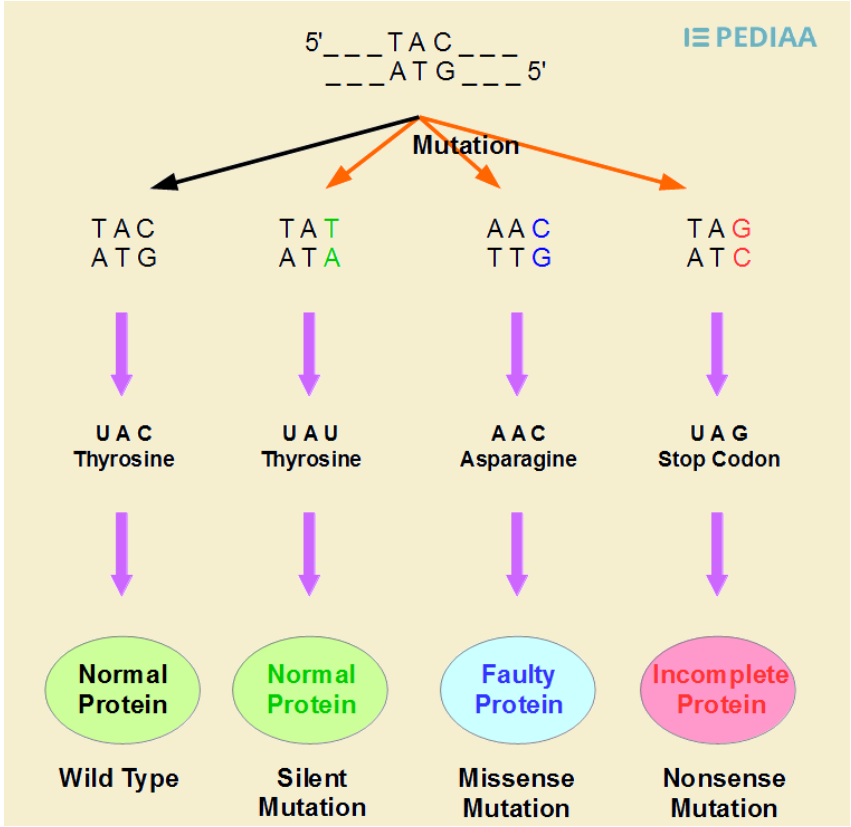
Gene Mutations create changes in DNA sequence, amino acid sequence, and changes in the encoded proteins

Point Mutations

- Mutations that affect one nucleotide are called point mutations because they occur at a single point in the DNA sequence

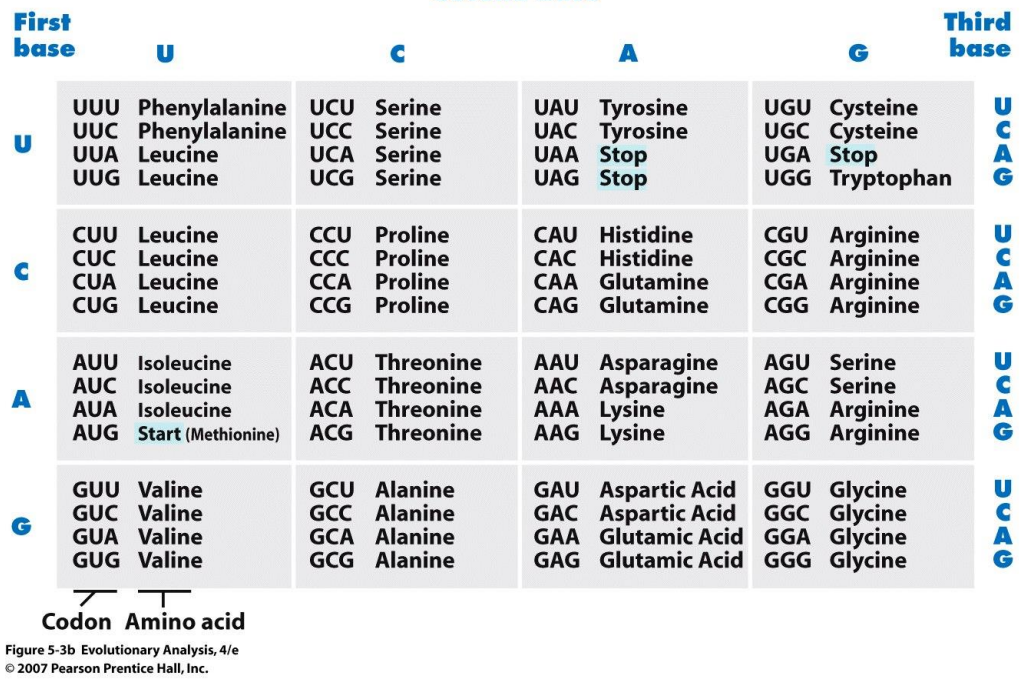
1) Substitutions

- ✓ substitute one nucleotide for another
- ✓ change one of the amino acids in a protein



From Genes to Proteins via mRNA

- Proteins consist of polypeptide chains made of amino acids
- There are 20 amino acids
- Genetic code
  - **64 combinations** of 3 bases called **codons**
  - There are three **stop codons**:
    - (UAA, UGA, UAG)
- Genetic code is degenerate
- Genetic code is universal



First base	Second base: U		Second base: C		Second base: A		Second base: G		Third base
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine	C
	UUA	Leucine	UCA	Serine	UAA	Stop	UGA	Stop	A
	UUG	Leucine	UCG	Serine	UAG	Stop	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine	C
	CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine	A
	CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine	G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine	C
	AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine	A
	AUG	Start (Methionine)	ACG	Threonine	AAG	Lysine	AGG	Arginine	G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic Acid	GGU	Glycine	U
	GUC	Valine	GCC	Alanine	GAC	Aspartic Acid	GGC	Glycine	C
	GUA	Valine	GCA	Alanine	GAA	Glutamic Acid	GGA	Glycine	A
	GUG	Valine	GCG	Alanine	GAG	Glutamic Acid	GGG	Glycine	G

Codon Amino acid

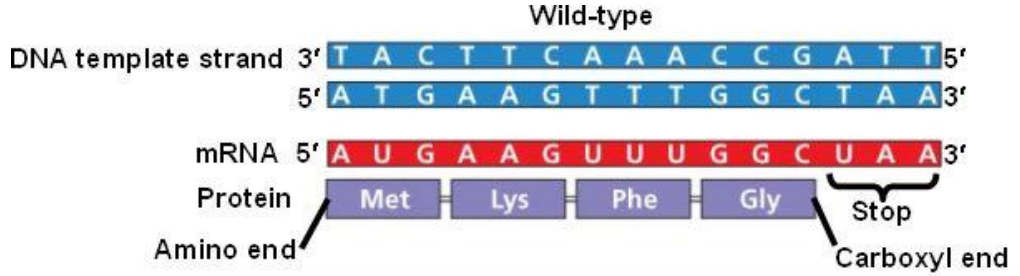
Figure 5-3b Evolutionary Analysis, 4/e

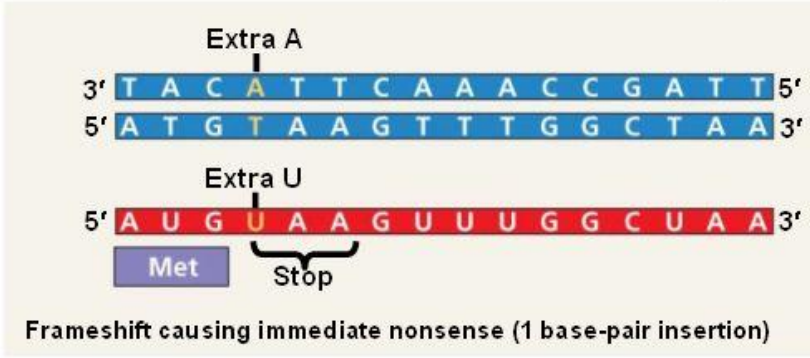
Frameshift Mutations

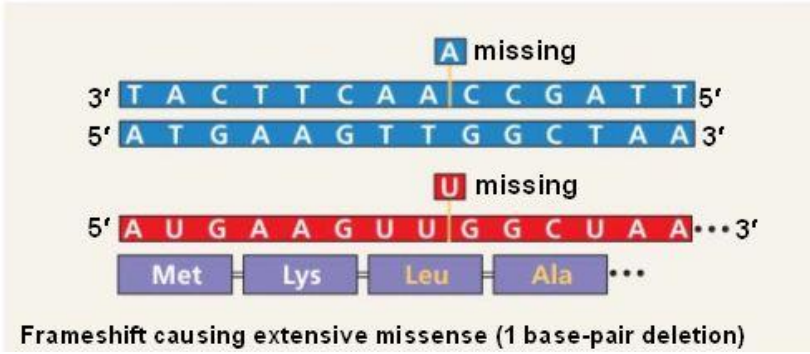
❖Insertions or deletions

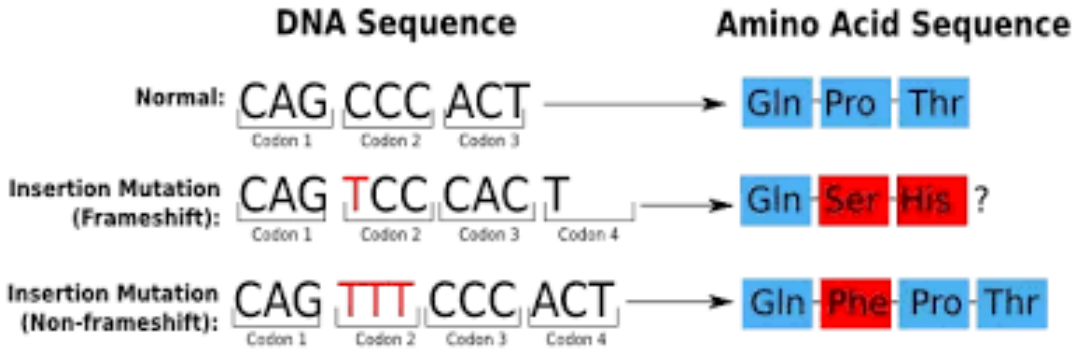
✓Shift the “reading frame” of genetic code.

✓May change every amino acid that follow the point of mutation altering the protein so it is unable to perform its normal function

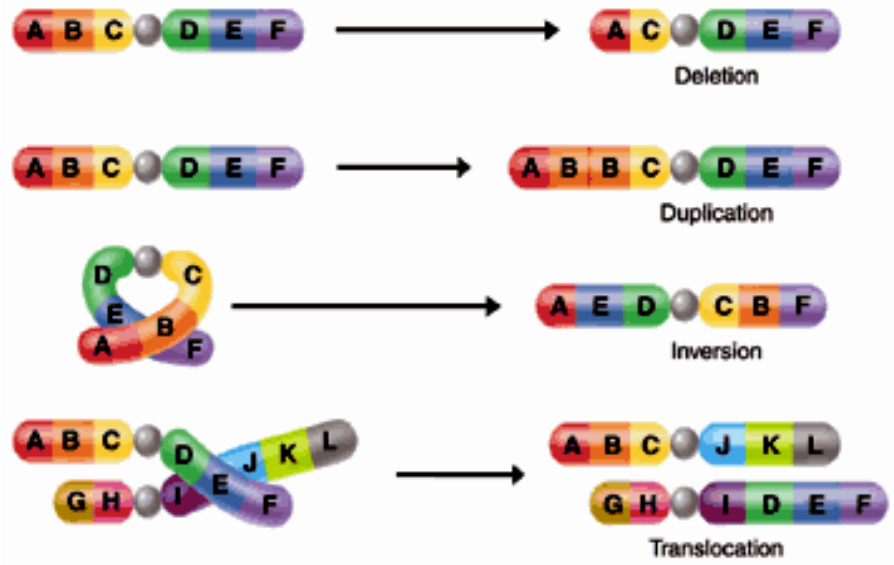








Chromosomal Mutations



1. **Deletions**

   Involve loss of all or part of a chromosome

2. **Duplications/Additions**

   Produce extra copies of parts of a chromosome

   e.g. Fragile X syndrome

3. **Inversions**

   Reverse direction of parts of chromosomes

4. **Translocations**

   When part of one chromosome breaks off and attaches to another

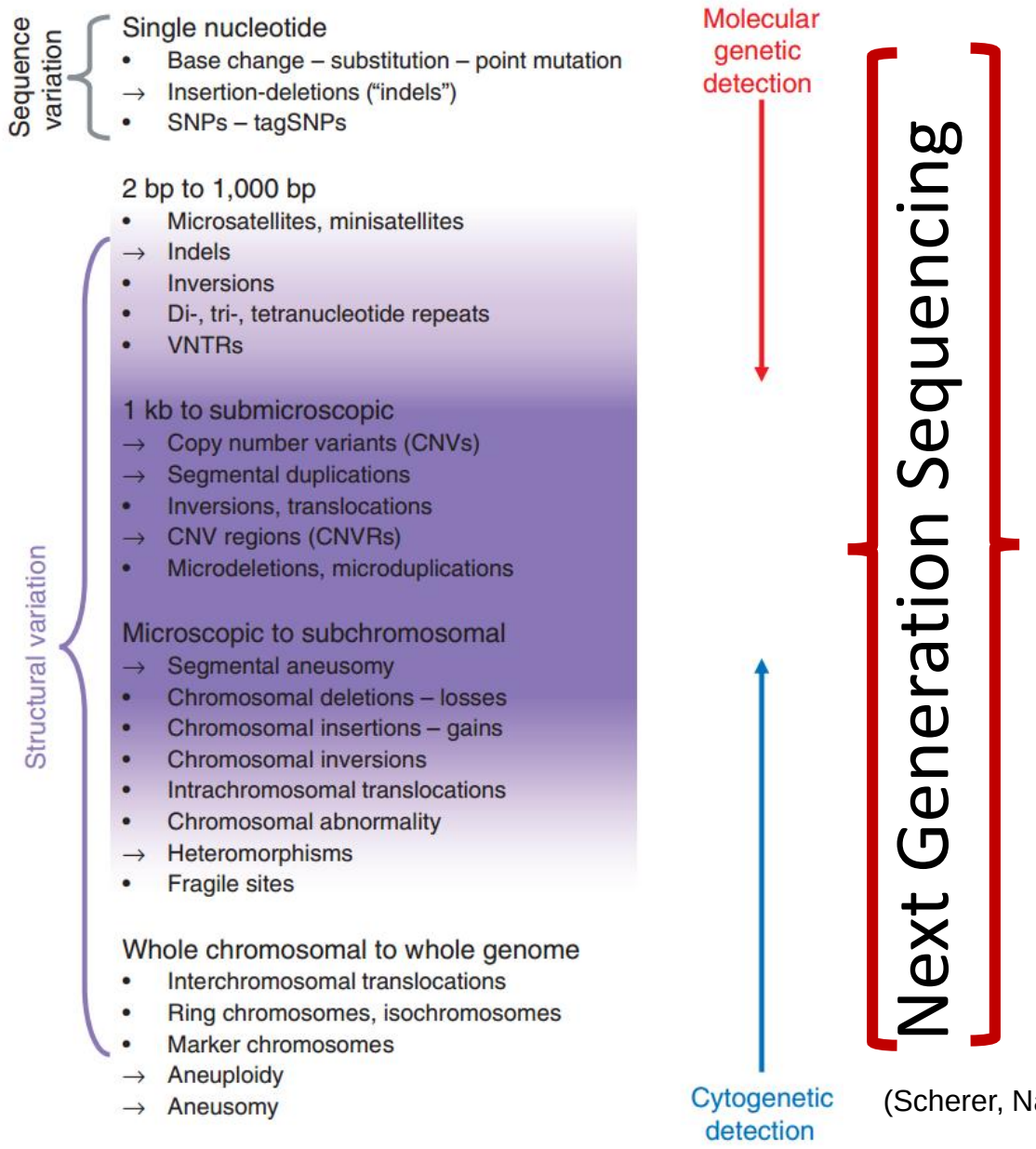
5. **Non Disjunction -** during meiosis, a pair of chromosomes do not separate & a gamete has one more chromosome while the other has one less

Human Genetic Variation

- Single nucleotide polymorphisms (SNPs)
- Tandem repeat Sequences
  - Microsatellites (<8 bp)
  - Minisatellites (VNTRs; 8-100 bp)
- Copy number variants (CNVs; 1Kb – 1Mb)
- Insertions – deletions (indels; 100bp – 1Kb)

- These size limitations are arbitrary

Genetic variation across size spectrum

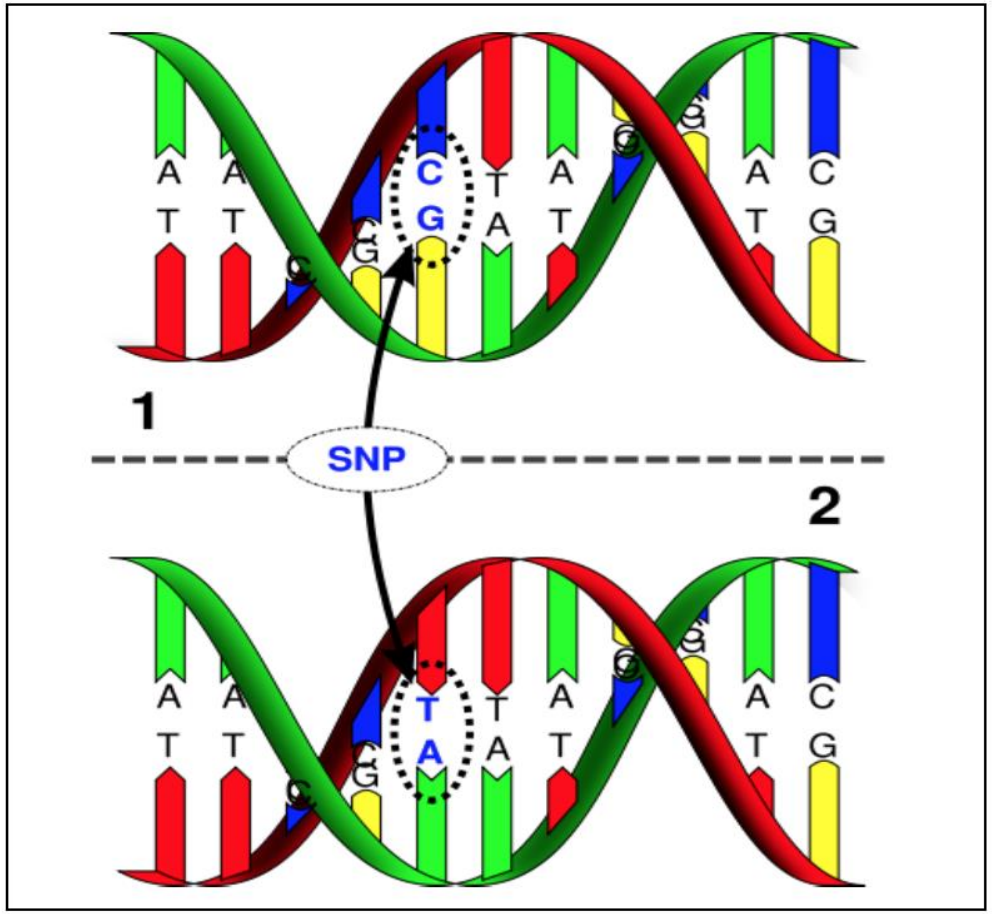


(Scherer, Nature Genetics, 2007)

SNPs

- ~10 million SNPs in nay given human genome & counting
- The most common type of genetic variation
- 2 alleles; e.g., A → T
- Occurs across the entire genome & in stable regions
- Many SNPs are in linkage disequilibrium (LD)
  - SNPs close together are more likely to travel together in a block
  - Can use 1 '**tag**' SNP per block – cost effective for **GWAS**

Terminology



C-allele: 70% frequency

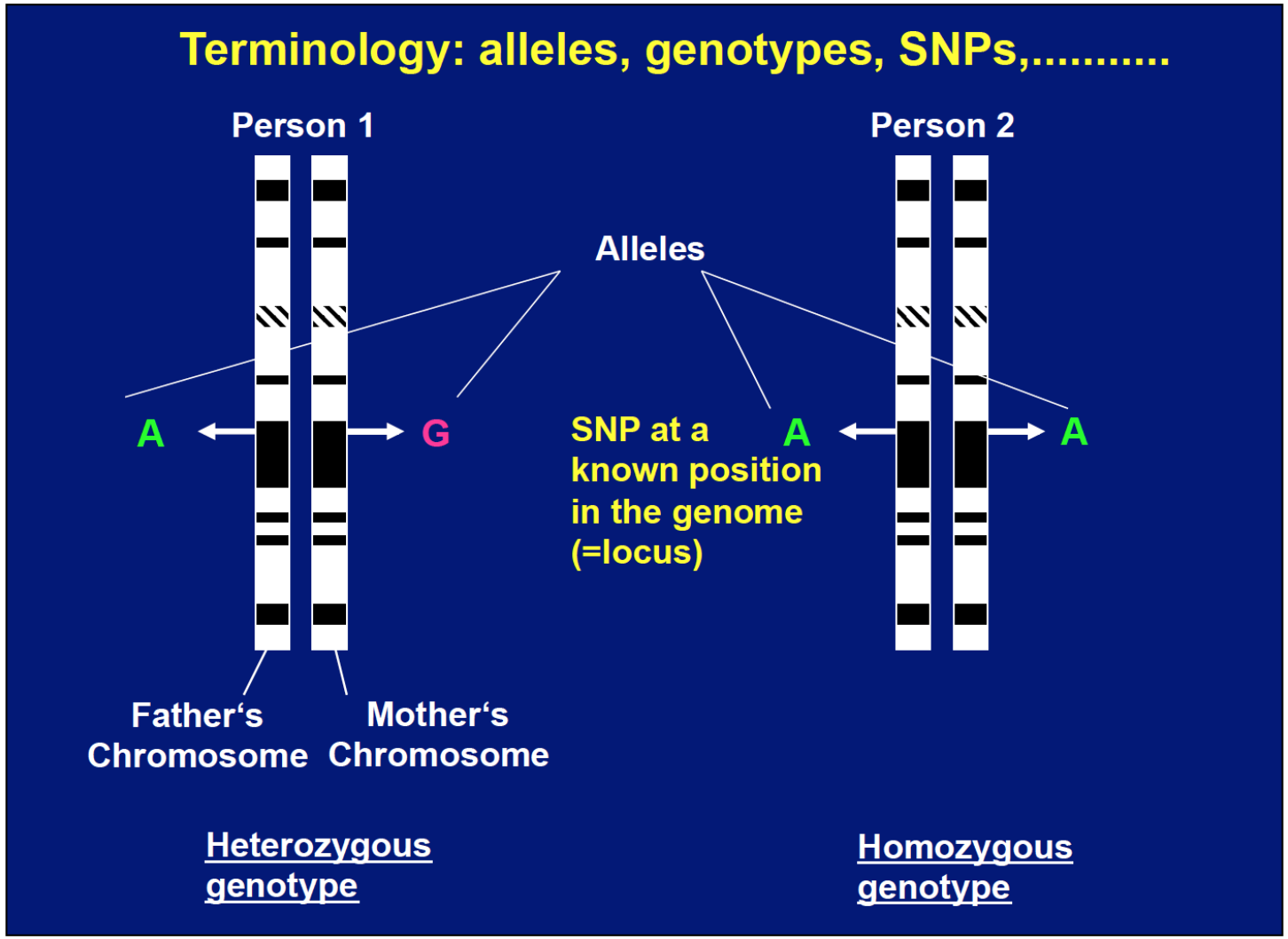
C = major allele

T-allele: 30% frequency

T = minor allele

Slide courtesy of Sven Cichon

Terminology



If allele **G** is associated with risk for disease, it is the **risk allele.**

Allele A the **protective** allele (WT)

Slide courtesy of Sven Cichon

Genetic Variation Databases

Database	Content	Address
dbSNP	SNPs covering the human genome	http://www.ncbi.nlm.nih.gov/projects/SNPs
HapMap	Catalog of variants from HapMap Project	http://hapmap.org
1000 Genome Project	Extension of HapMap – aim to catalog 95% of variants with 1% freq	www.1000genomes.org
UCSC Genome Bioinformatics	Reference human genome sequence with annotation	http://genome.ucsc.edu
Ensembl	Genome browser, annotation, comparative genomics	http://www.ensembl.org/index.html

Genetic Variation Databases

Database	Content	Address
GeneCards	Database of human genes linked to relevant databases	http://www.genecards.org
PharmGKB	SNPs involved in drug metabolism	http://www.pharmgkb.org
DGV	Database of Genomic Variants, including CNV	http://projects.tcag.ca/variation
SCAN	SNP & CNV annotation based on gene function & expression	http://www.scandb.org/newinterface
OMIM	Online Mendelian Inheritance in Man – over 12,000 genes	http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim

Genetic Variation Databases

Database	Content	Address
HuGE navigator	Human genome epidemiology knowledge base	http://hugenavigator.net/HuGENavigator/home.do

Best Pract Res Clin Endo Metab, 2012. 26:119.

Collecting DNA

- Sources of DNA
  - Blood samples
  - Buccal brushes
  - Saliva samples
  - Dried blood spots
- Depends on
  - Conditions at time of collection
  - Resources available to process samples
  - What other biological samples will be collected
  - Long & short term storage
  - Quality control

Saliva vs. Blood Samples

- Considerations
  - Lower cost
  - More convenient & acceptable to patients
  - Increases compliance
  - Lower mean yield of DNA with comparable quality
  - No difference in success from high throughput genotyping


Thank you!