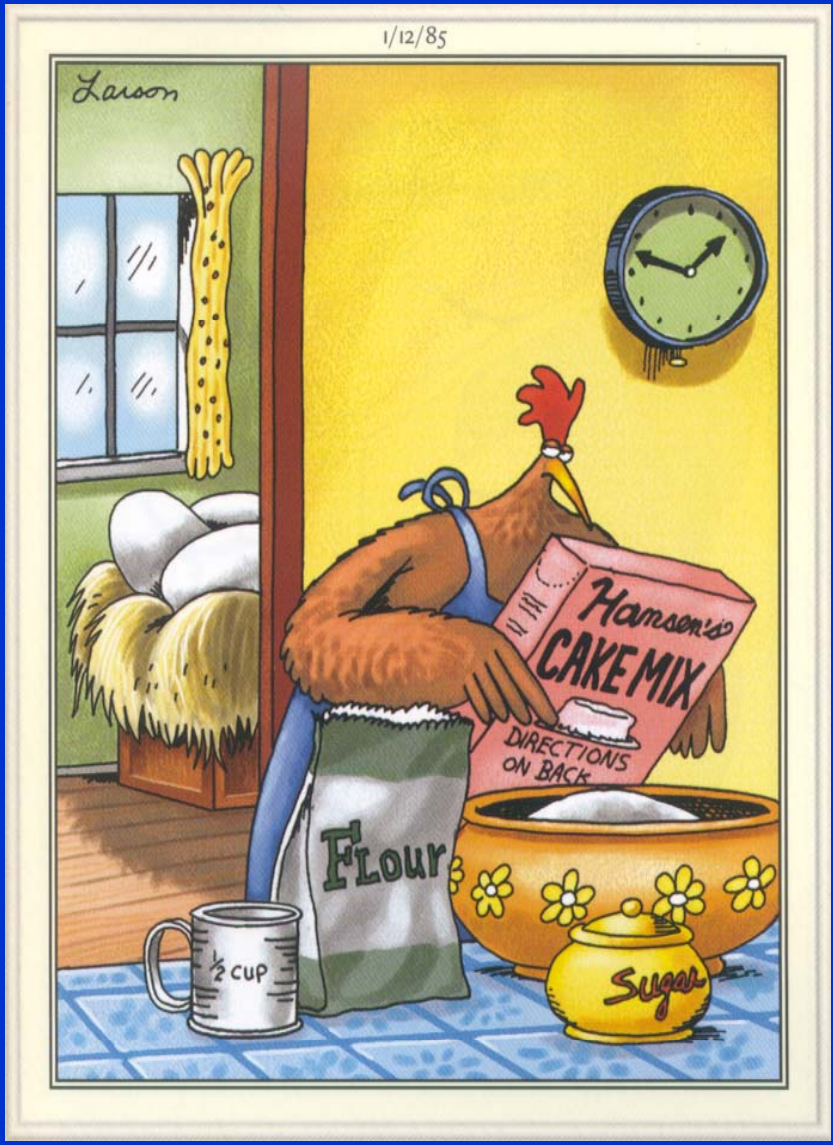
Genetics for Epidemiologists: Application of Human Genomics to Population Sciences

Thomas A. Pearson, MD, PhD
University of Rochester School of Medicine and Dentistry
Visiting Scientist, National Human Genome Research Institute

Teri A. Manolio, MD, PhD
Director, Office of Population Genomics
National Human Genome Research Institute

CONFLICT OF INTEREST DISCLOSURE

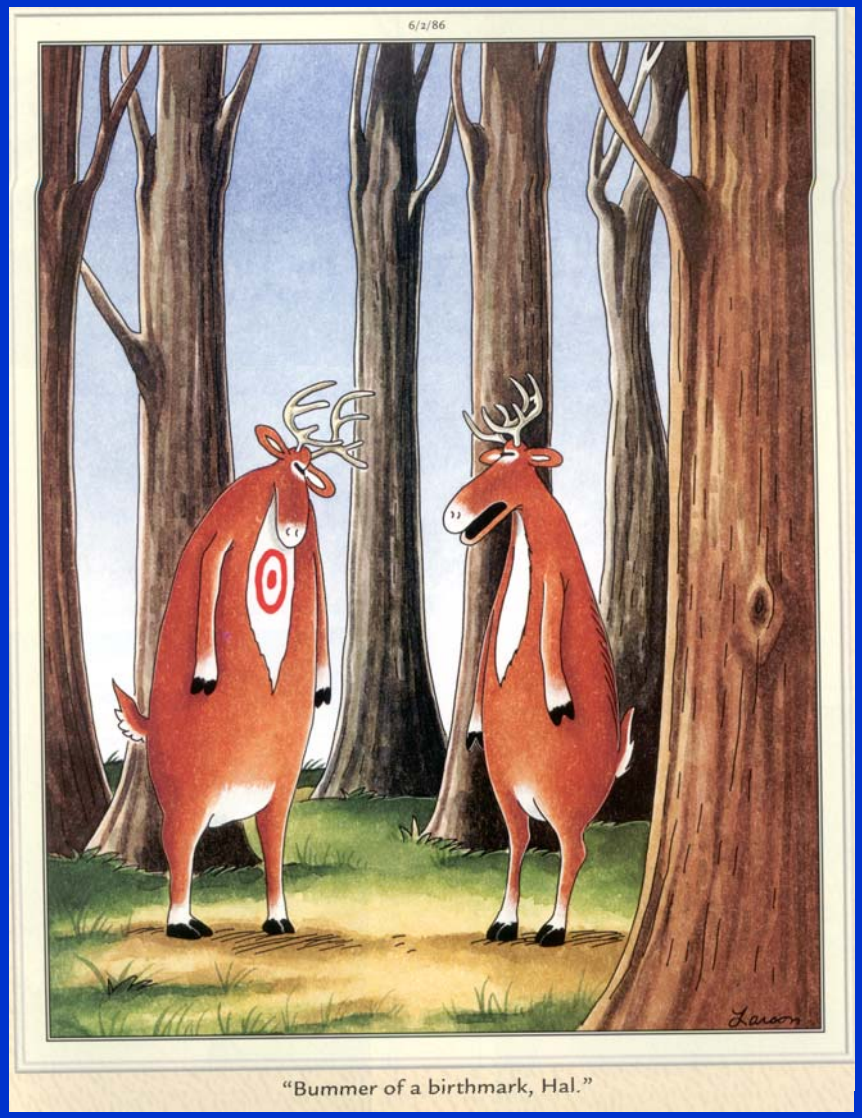


Larson, G. *The Complete Far Side*. 2003.

A Genome-wide Association Study of Crohn Disease*

- GWAS examining 300,000 SNPs in 547 patients with Crohn Disease and 928 controls.
- Confirmation of two genes previously associated with CD (*IL23R* on Chromosome 1, *CARD15* on Chromosome 16).
- Region of Chromosome 15 (p13.1) identified with multiple SNPs having significant associations with CD.
- Replication of 5p13.1 association in 1266 CD cases/559 controls and in 428 trios.
- SNPs located in gene desert.
  - Associated with expression of a prostaglandin receptor gene (*PTGER4*) located 270 kb proximally.
  - Speculated to control regulation of *PTGER4* gene.

*Libioulle C et al: PLoS Genetics 2007; 3: e58



Larson, G. *The Complete Far Side*. 2003.

Course Overview

- Goal: To familiarize epidemiologists and population-based investigators with recent developments in the theory and methods of human genetics and genomics.
- Learning Objectives
- Format
  - Eight lectures
  - Case studies
  - Discussion
  - Webcast

NHGRI Catalog of GWAS (www.genome.gov/gwastudies/)

- All publications reporting genome-wide association studies (beginning March, 2005)
  - Platforms with density of at least 100,000 SNPs
  - Identified by literature searches, media, HUGE Navigator
- Data Presented
  - Citation
  - Disease/trait
  - Sample sizes
  - Chromosomal region
  - Gene
  - Association
    - Strongest risk allele
    - Odds ratio per copy
    - Risk allele frequency
    - P value of association

Genetics for Epidemiologists

Lecture 1. The Biologic Basis for Analysis of Gene Variants

Thomas A. Pearson, MD, PhD
University of Rochester School of Medicine
Visiting Scientist, NHGRI

Lecture 1: Learning Objectives

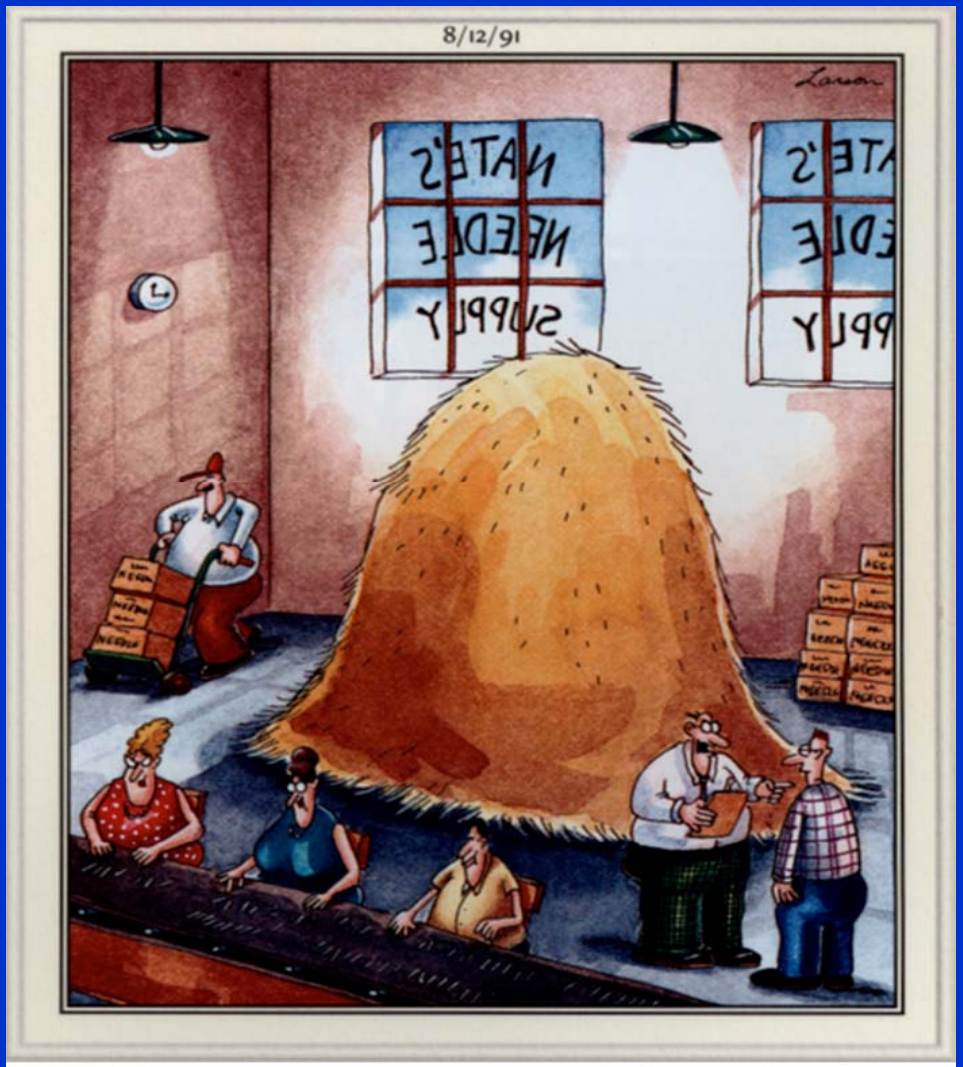
- Provide an overview for the course.
- Review the structure and function of the human genome.
- Discuss patterns of inheritance.
- Describe types of genetic variation as potential causes of disease.
- Introduce online informatics resources which describe genotype and phenotype of human genetic variants.

Genetics Vs. Genomics

Genetics:	The science of inheritance (Bateson, 1905)
Genetic Code:	DNA structure and function (Watson and Crick, 1953)
Genomics:	The field within genetics concerned with the structure and function of the entire DNA sequence of an individual or population (Roderick, 1986)

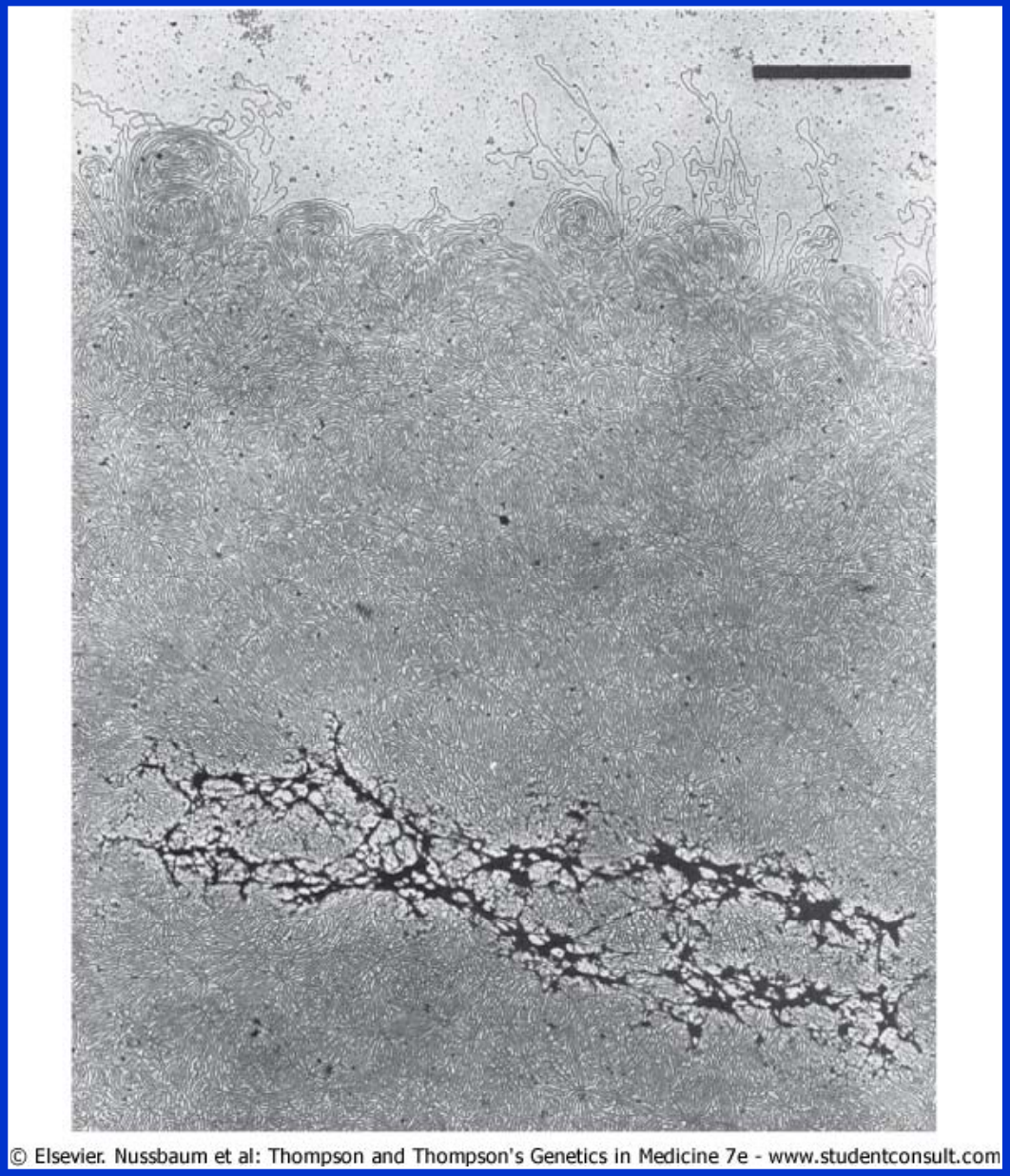
The Challenge: Finding Genetic Variants Affecting Human Health

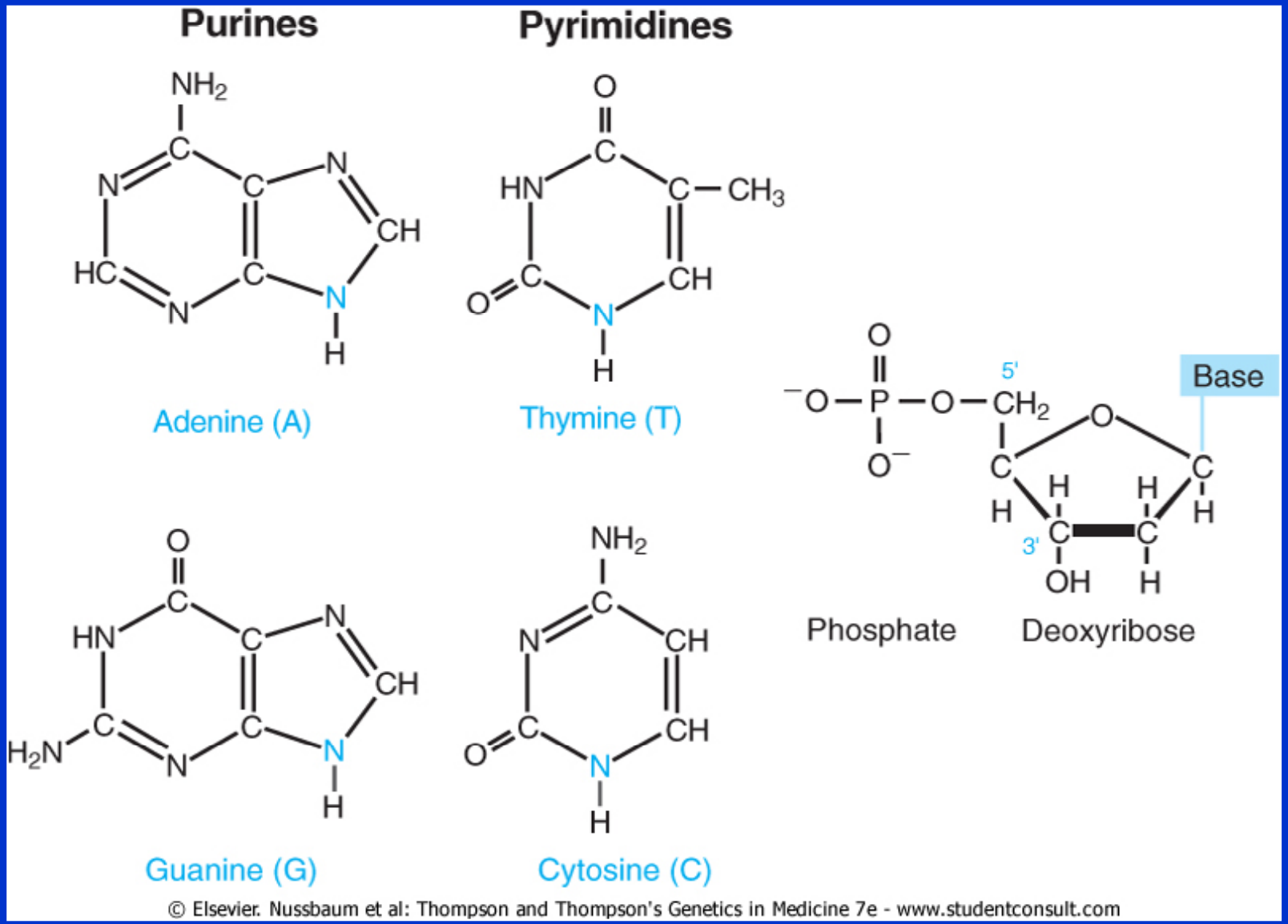
Chromosomes	46 (22 pairs of autosomes, X, Y)
Genes	20,000-25,000
Base-pairs	3,000,000 kilobases (haploid)
Variants	99.9% of bases are identical between all people

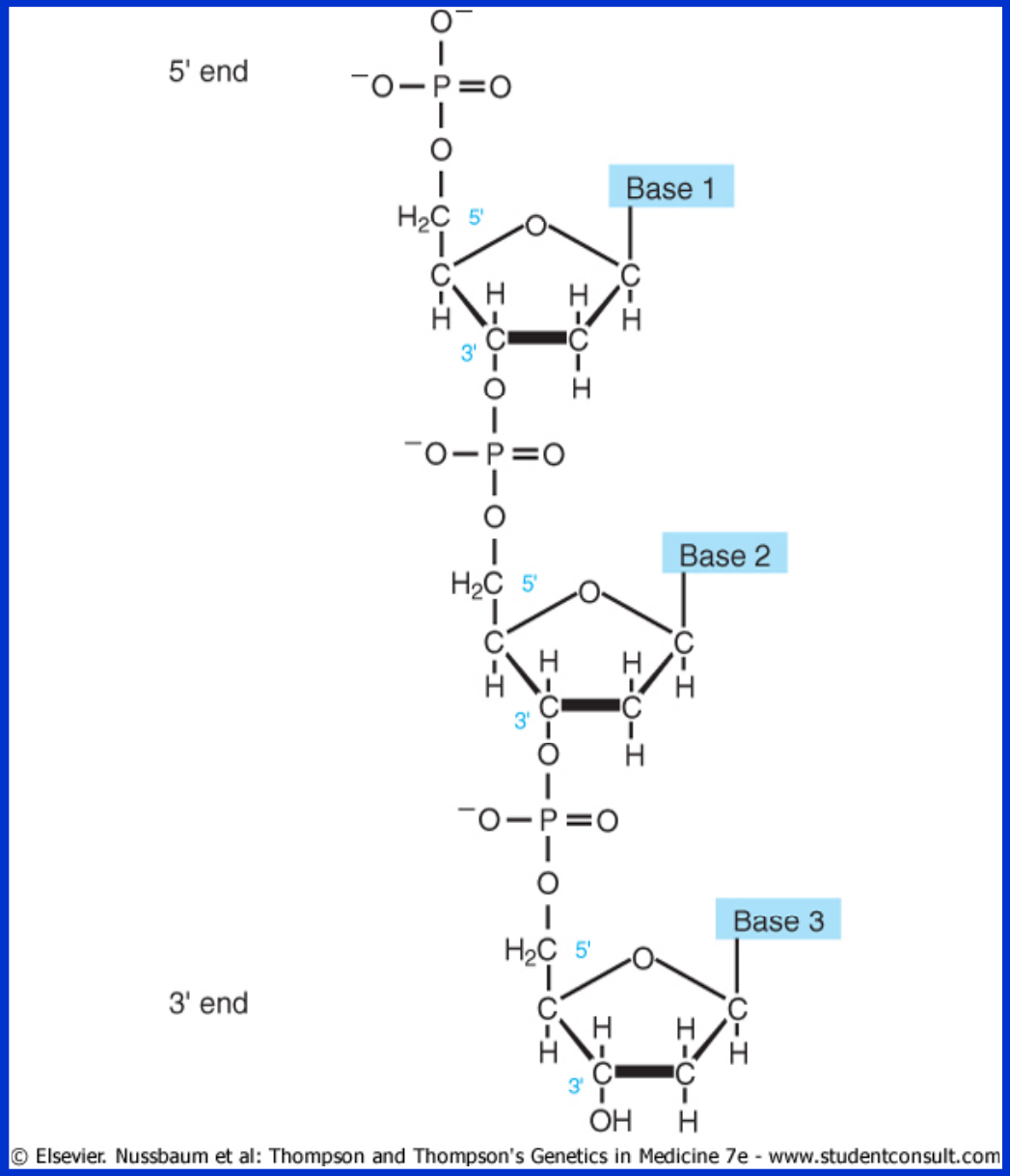


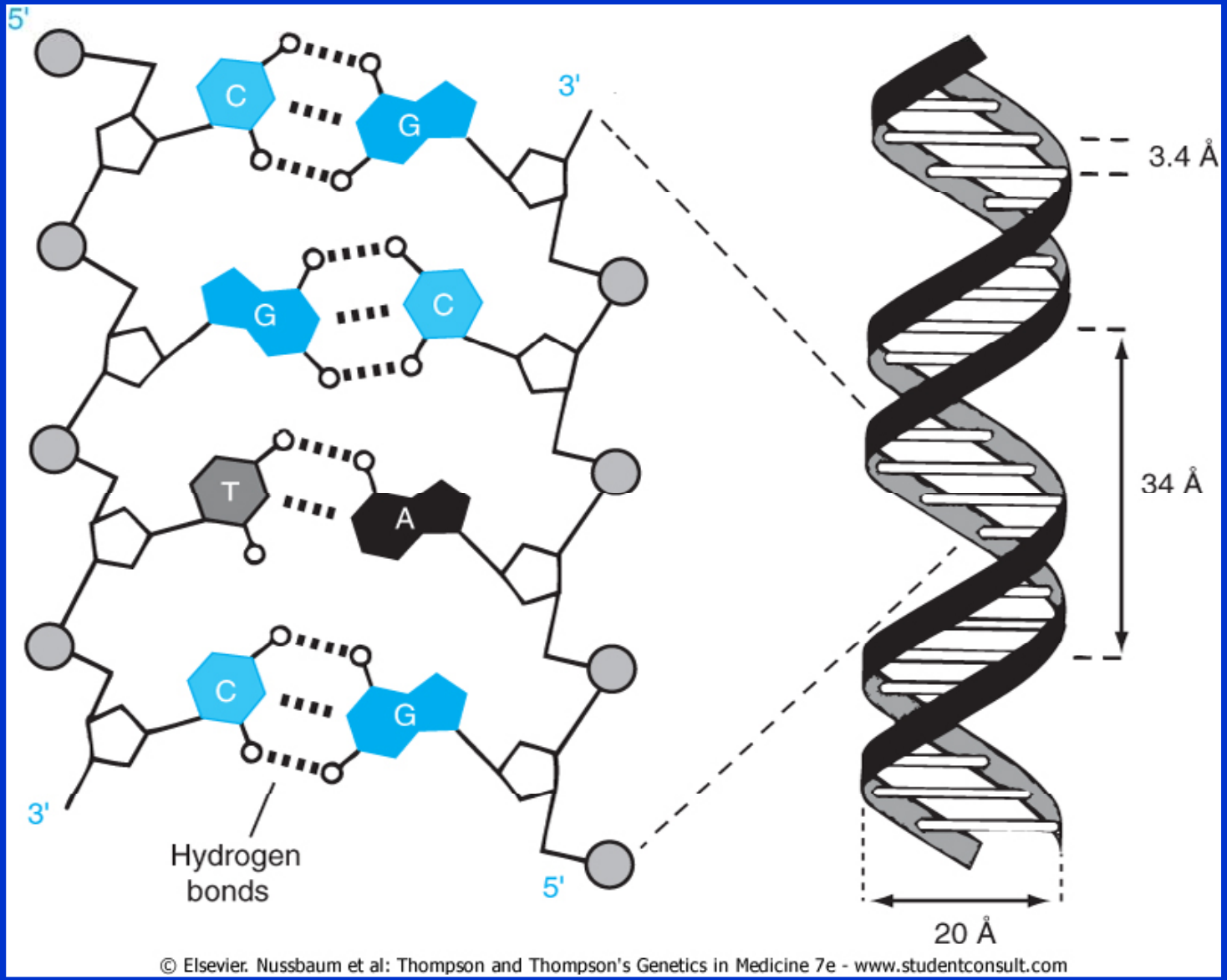
"According to these figures, Simmons, your department has lost another No. 2 Double N -- and I want you to find it!"

Larson, G. *The Complete Far Side*. 2003.





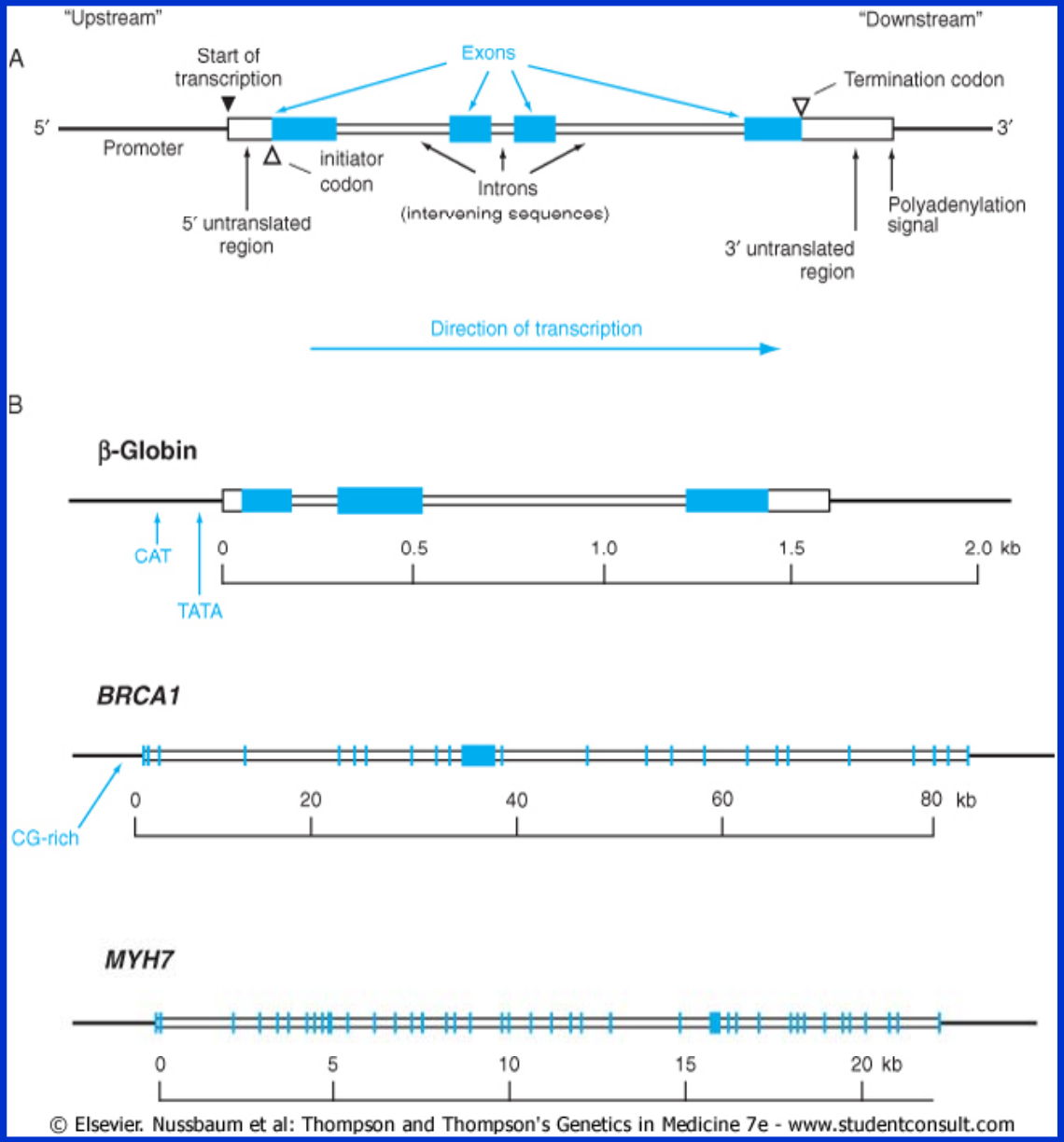


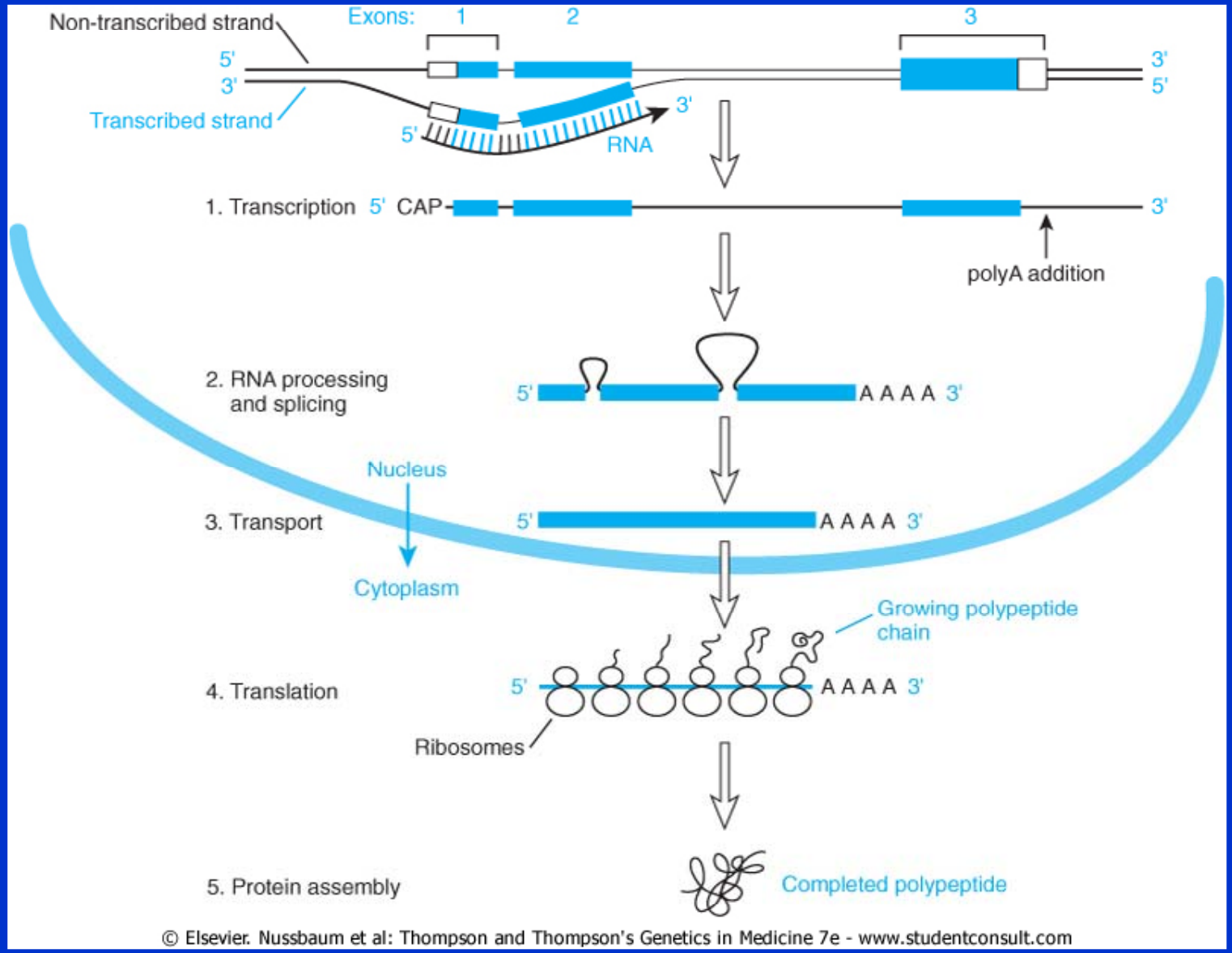


© Elsevier. Nussbaum et al: Thompson and Thompson's Genetics in Medicine 7e - www.studentconsult.com

Structure of Human Genes: Potential Sites of Gene Variation

- Exons
- Introns
- Regulatory Elements
  - Promoters
  - PolyA Tail
  - Enhancers
  - Silencers
  - Locus Control Regions





The Genetic Code

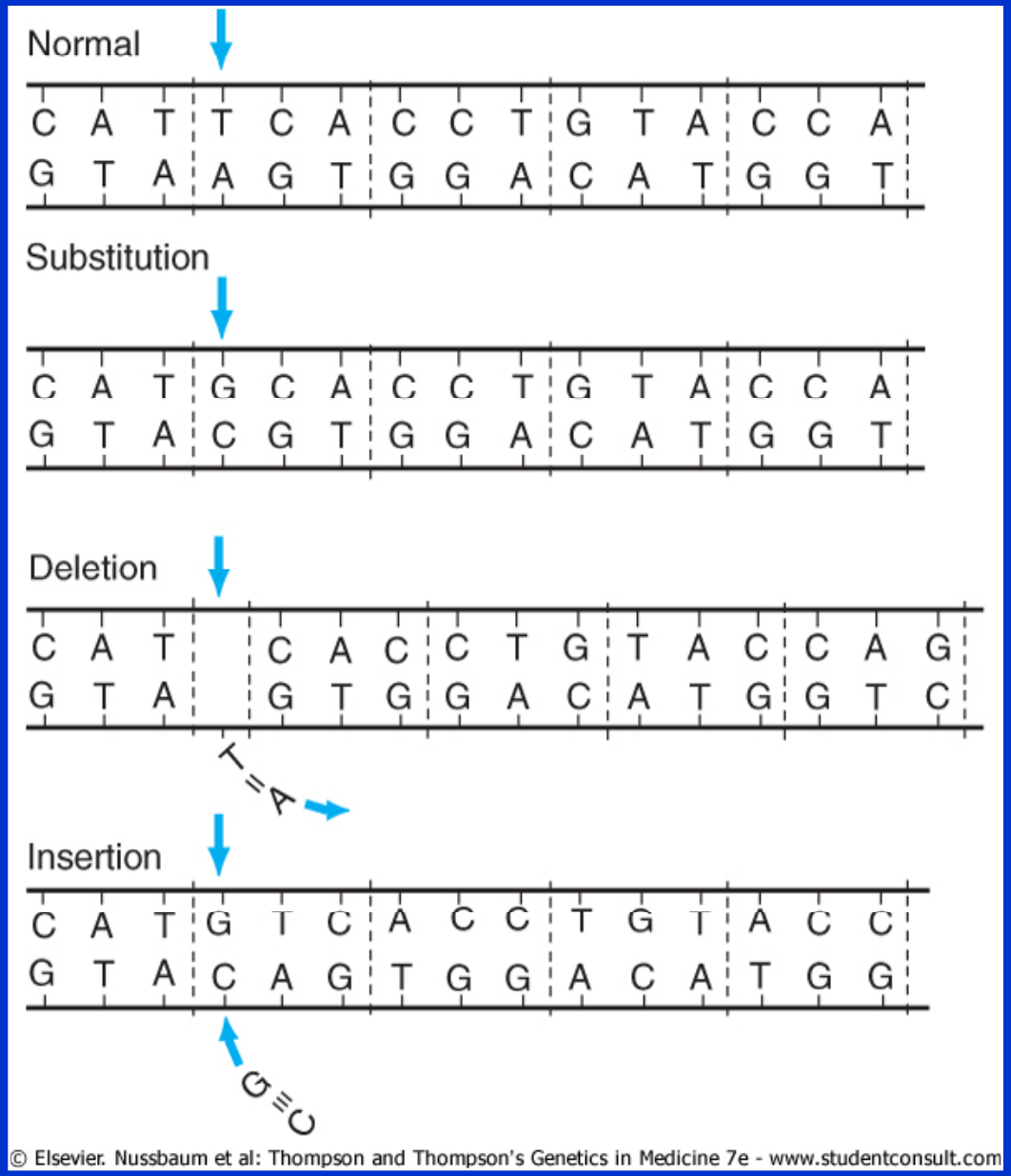
- Sequences of four DNA bases (A,T,C,U) are translated in triplets called codons, each encoding an amino acid.
- 64 codons (4X4X4) constitute the genetic code.
- Some of the 20 amino acids will be encoded by more than one codon.
- Three stop or nonsense codons designate termination of translation of mRNA.

Mutations

Class	Mechanism	Frequency	Example
Genome	Chromosome segregation	1/25-50 cell divs.	Trisomy 21
Chromo-some	Chromosome rearrangement	1/1700 cell divs.	Cancer Cells
Gene	Base-pair mutation	1/1000 base-pairs	SNPs

Types of Mutations: Single Nucleotide Substitutions

Silent (Synonymous)	No effect on amino acid sequence.
Missense (Nonsynonymous)	Alters amino acid sequence.
Nonsense	Encodes termination codon.
Termination	Destroys termination codon and affects adjacent gene.



Types of Mutations (Indels): Deletions and Insertions

- Simple: those involving a short segment of DNA and only 2 alleles.
- Short tandem repeat polymorphism: 2-4 nucleotide repeat units repeated 5-25 times, affecting many alleles.
- Variable number tandem repeats: 10-100 nucleotide repeat units repeated hundreds of times, affecting many alleles.
- Copy number variants: 200bp-1.5Mb segments of DNA affecting a few alleles.

Mendel's Principles of Inheritance

- Segregation: The pair of alleles for any given trait separate and only one allele passes from each parent to an offspring; which allele passed is random.
- Independent Assortment: Traits encoded by different pairs of alleles are inherited independent of each other, unless genetically linked.

Mendelian Traits Listed in Online Mendelian Inheritance of Man (OMIM), as of July, 2007

	Gene Known	Gene Unknown	Total
Autosomal	1932	1954	3386
X-Linked	177	133	310
Y-Linked	2	4	6
Mitochondrial	26	0	26

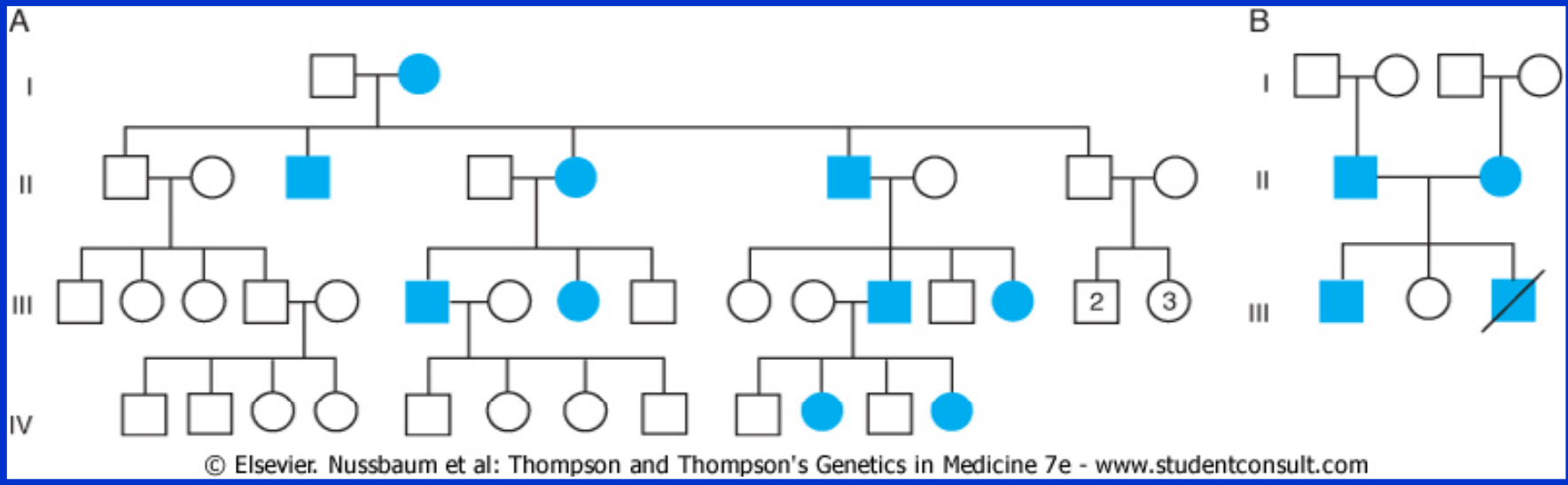
Patterns of Inheritance

- Mendelian Disease: Condition (phenotype) caused almost entirely by a single major gene, in which the disease is manifested in only 1 (recessive) or 2 (dominant) of the 3 possible genotype groups.
- Common disease, Common Variant:

  Common conditions (phenotype) attributable to a limited number of allelic variants which occur in 1-5% or more of the population.

Autosomal Dominant Inheritance

- Phenotype appears in every generation (unless new mutation).
- Any child of an affected person has a 50% risk of inheriting the trait.
- Male-to-male transmission is present.
- Phenotypically normal family members do not transmit the phenotype to children.
- Male:Female occurrence is usually equal.



Complete & Incomplete Dominance

Complete dominance: Phenotypes are indistinguishable in heterozygous or homozygous state.

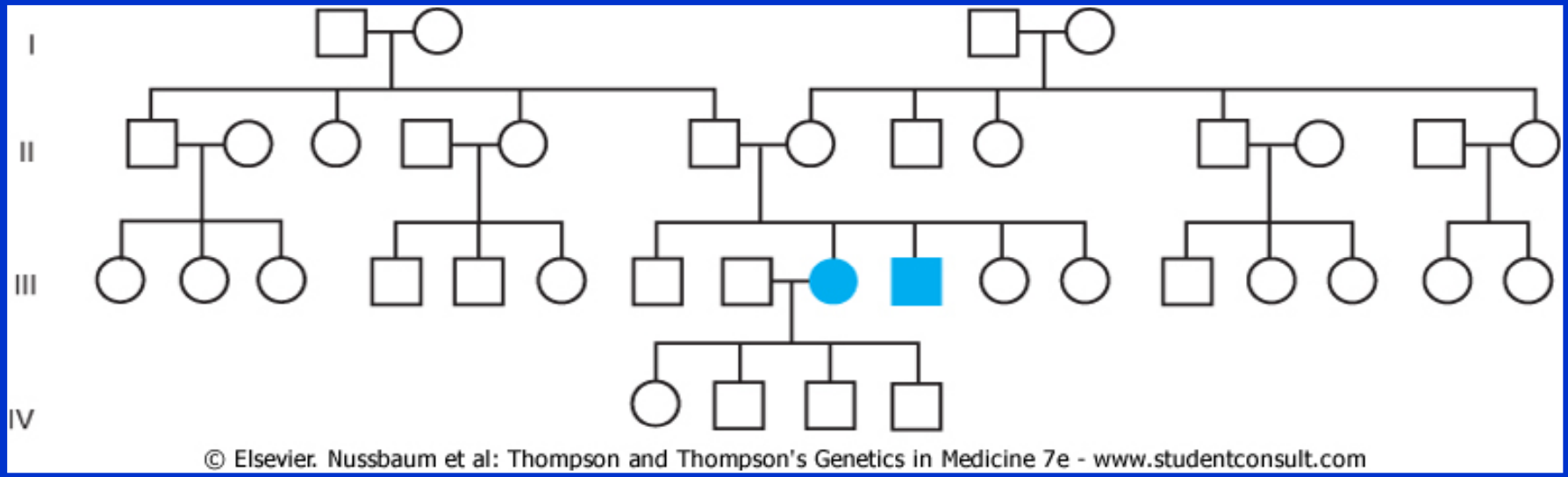
(Example: Huntington's Disease)

Incomplete dominance: Phenotypes are more severe in the homozygous than in the heterozygous state. (Example:

Familial Hypercholesterolemia)

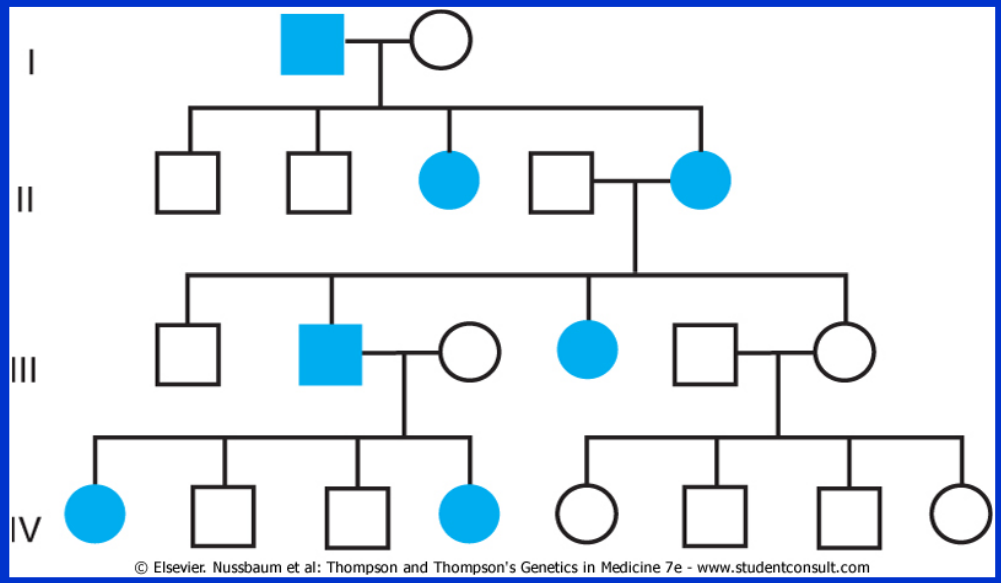
Autosomal Recessive Inheritance

- Phenotype, if it appears in one family member, is seen only in the sibship of the proband.
- Male:Female occurrence is usually equal.
- Parents of affected child are asymptomatic carriers of the gene variant.
- Consanguinity is increased in families with recessive disorders.
- Offspring of two heterozygous parents have the following risks:
  25% affected, 50% carrier, 25% noncarrier.



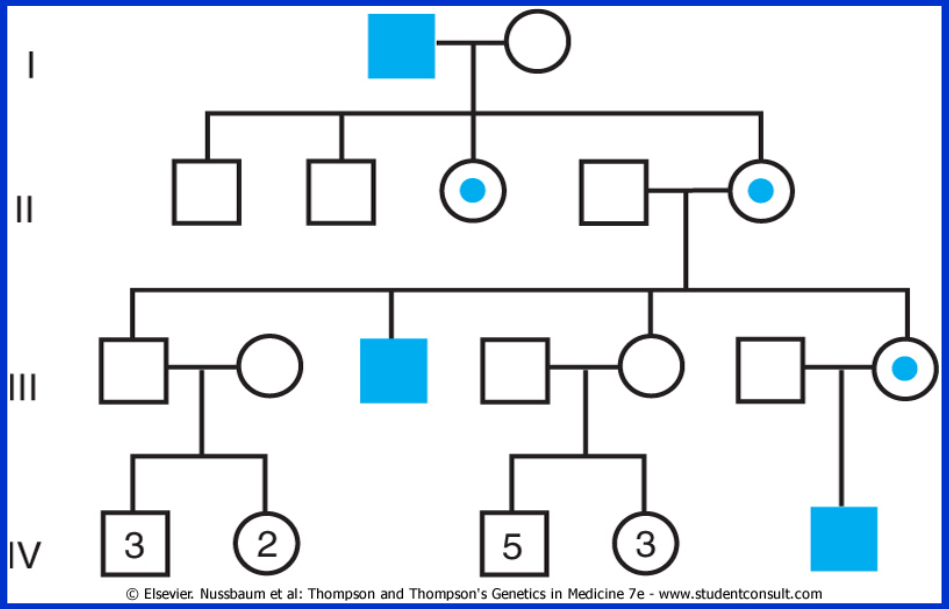
X-Linked Dominant Inheritance

- Both male and female offspring of female carriers have 50% risk of inheriting the phenotype.
- No male-to-male transmission.
- Number of females affected >number of males.
- All daughters of affected males are affected but none of their sons.
- Severity in females modified by X-inactivation (Lyonization)
- Example: Vitamin D resistant rickets



X-Linked Recessive Inheritance

- Only males are affected.
- Affected males are related through carrier females.
- No male-to-male transmission
- Unaffected males do not transmit.
- All daughters of affected males are carriers.
- Examples: Hemophilia A, red-green color blindness

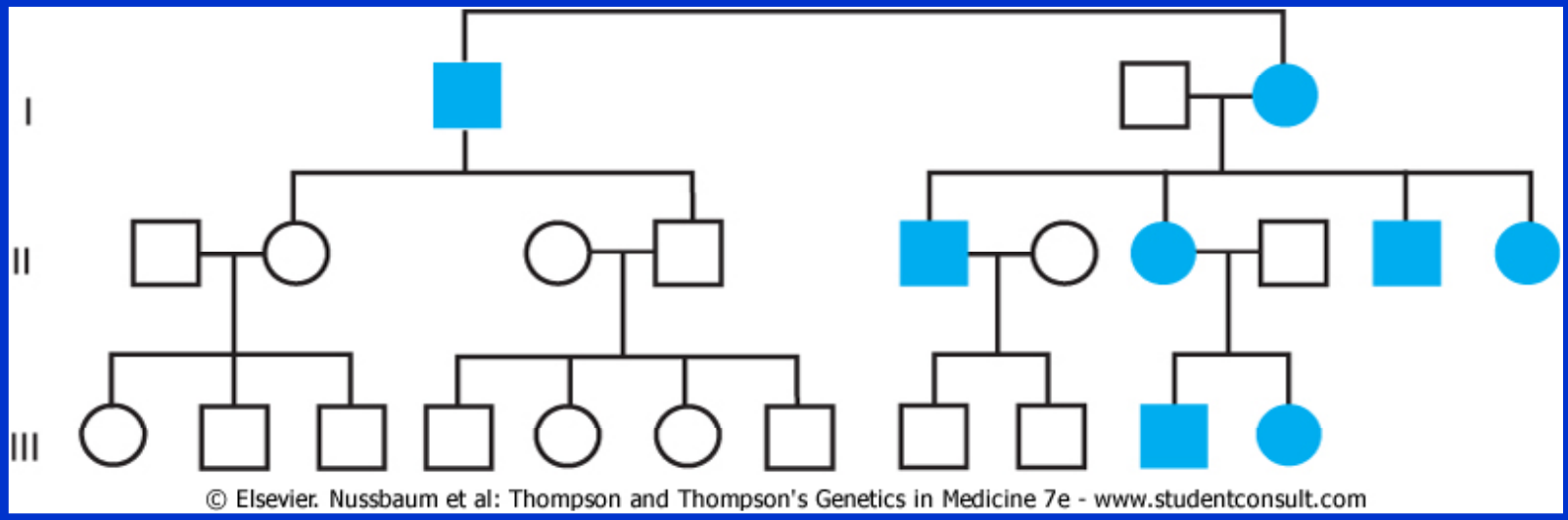


Y-Linked Inheritance

- Only males affected
- All sons but no daughters of affected men are affected.
- Not X-linked since male-to-male transmission occurs.
- Sex-limited, because trait does not pass through unaffected females.

Mitochondrial Inheritance

- Matrilineal inheritance
- All children of affected females are affected.
- Numbers of affected males and females are equal.



Exceptions to the Rules of Inheritance

- Sporadic mutations
- Genetic heterogeneity
- Nonpenetrance
- Variable expressivity
- Late onset conditions
- Sex-limited/sex-influenced phenotypes

The Common Disease-Common Variant Hypothesis

SNP1 in Exon1 of GeneA

SNP1 in Exon 3 of GeneA

SNP2 in Exon 3 of GeneA

SNP 2 in Exon 2 of GeneB

SNP, Reg. Element, GeneA

Environmental Exposures

Disease

Susceptibility Variants Associated with Systemic Lupus Erythematosus in Women*

- Case-control study of 720 women with SLE and 2337 control women.
- 317,501 SNPs assessed genome-wide
- Two replication studies with 1846 female cases and 1825 female controls.
- At least 17 SNPs associated with SLE at P<2xE-7

*International Consortium for SLE Genetics. Nature Genetics 1/20/08

Logistic Regression Model of Independent Contributions of Markers Associated with SLE*

Gene	Chromosome	OR	P
PKY	3p14.3	1.27	9.2E-07
HLA region	6p21.33	1.82	4.5E-17
HLA region	6p21.32	1.40	2.8E-12
IRF5/TNP03	7q32.1	1.61	1.7E-14
KIAA1542	11p15.5	0.78	1.3E-07
ITGAM	16p11.2	1.70	1.9E-18

Cstatistic=0.67;15% of heritability explained

*Int. Consort. for SLE Genetics.NatGen 1/20/08

Bioinformatics Resources in Human Genomics

- Nat. Center for Biotechnology Information, NLM: http://www.ncbi.nlm.gov

  OMIM: http://www.ncbi.nlm.gov/OMIM
  GenBank: http://www.ncbi.nlm/Genbank/
  RefSeq: http://www.ncbi.nlm.gov/RefSeq
  dbSNP: http://www.ncbi.nlm.gov/dbSNP
  dbGAP: http://www.ncbi.nlm.gov/dbGAP

Online Mendelian Inheritance in Man (OMIM)

- Catalog of human genes and genetic disorders.
- Concise information on most human conditions having a genetic basis.
- Pictures illustrating the condition or disorder
- Full citation information, linked to PubMed
- OMIM Numbering System
- OMIM Maps: cytogenetic location of genes and diseases.

GenBank
NIH Genetic Sequence Database

- Annotated collection of all publicly available DNA and protein sequences
  - >80 million sequence records
  - >80 billion bases
- Concise description of the sequence
  - Scientific name and taxonomy of organism
  - Bibliographic references (PubMed)
  - Features provided by submitter (e.g. biologic function, mutations and modifications, secondary or tertiary structure)

RefSeq: Database of Single Reference Sequences

- Each molecule in sequence (DNA, RNA, Protein).
- Nonredundant.
- Linked to nucleotide and protein sequences in GenBank.
- Updated by NCBI staff and collaborators to reflect current sequence data and biology.
- Review status indicated on each record.

dbSNP: Database of Single Nucleotide Polymorphisms

- 6.2 million validated human SNPs and counting
  - Up to 10 million SNPs likely exist
  - Over 200,000 SNPs within genes
- Relates genes to specific diseases.

Conclusions

- 1900-present: Human genetics, Mendelian disorders
- 1953-present: Molecular genetics, structure and function of the human genome.
- 1980-present: Identification of genetic variants and candidate gene studies
- 2003-present: Sequencing of the entire human genome and genome-wide association studies

Questions?

References

Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 2008; 299: 1335-1344 (Includes a glossary of terms frequently used in GWAS).

Nussbaum RL, McInnes RR, Willard HF. Thompson and Thompson Genetics in Medicine, 7th ed. Philadelphia, Elsevier, 2007.