Genetic basis of common human disease

Dan Evans, PhD, MPH

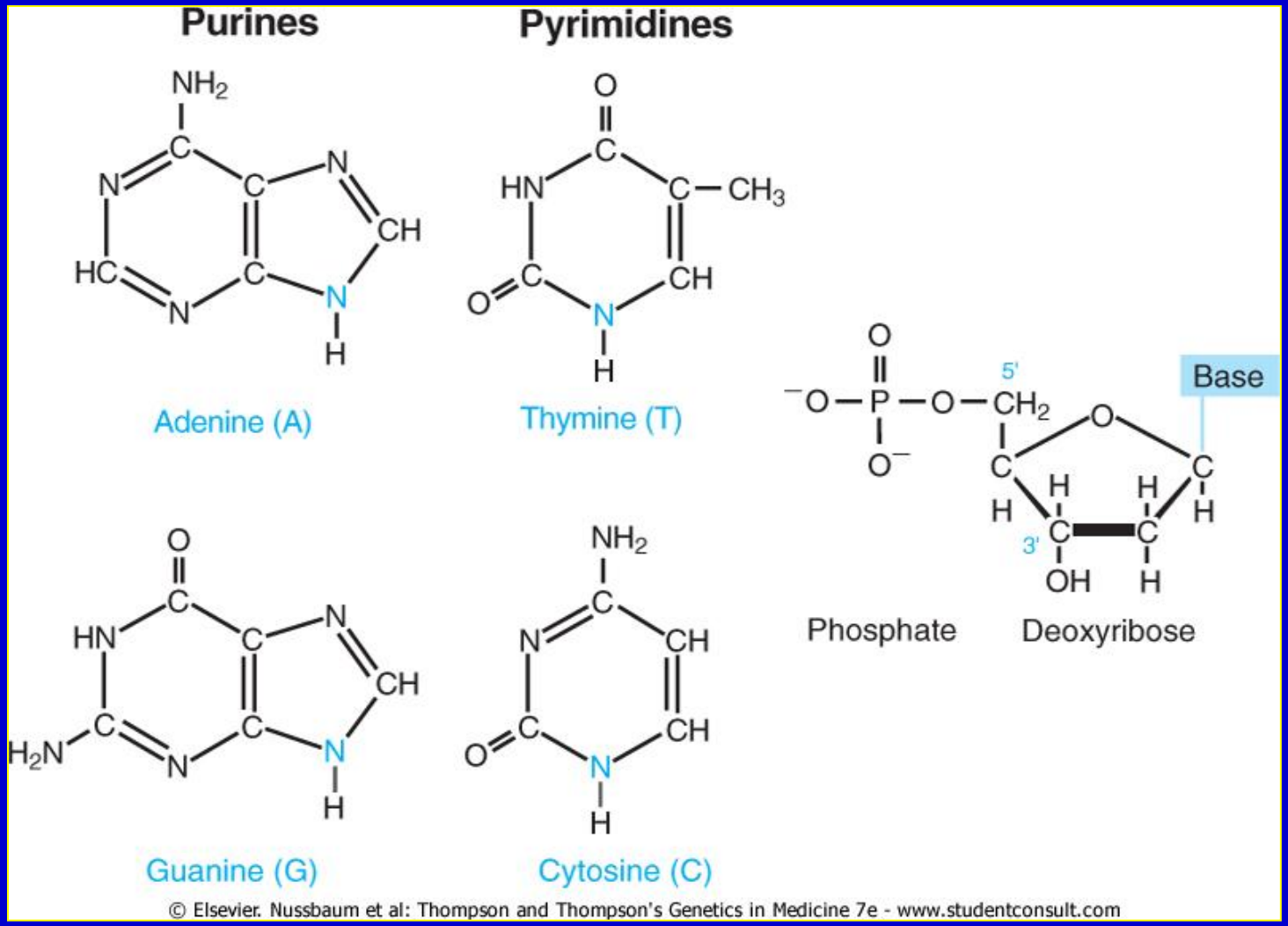
Research Scientist

California Pacific Medical Center Research Institute

Genetics of human disease

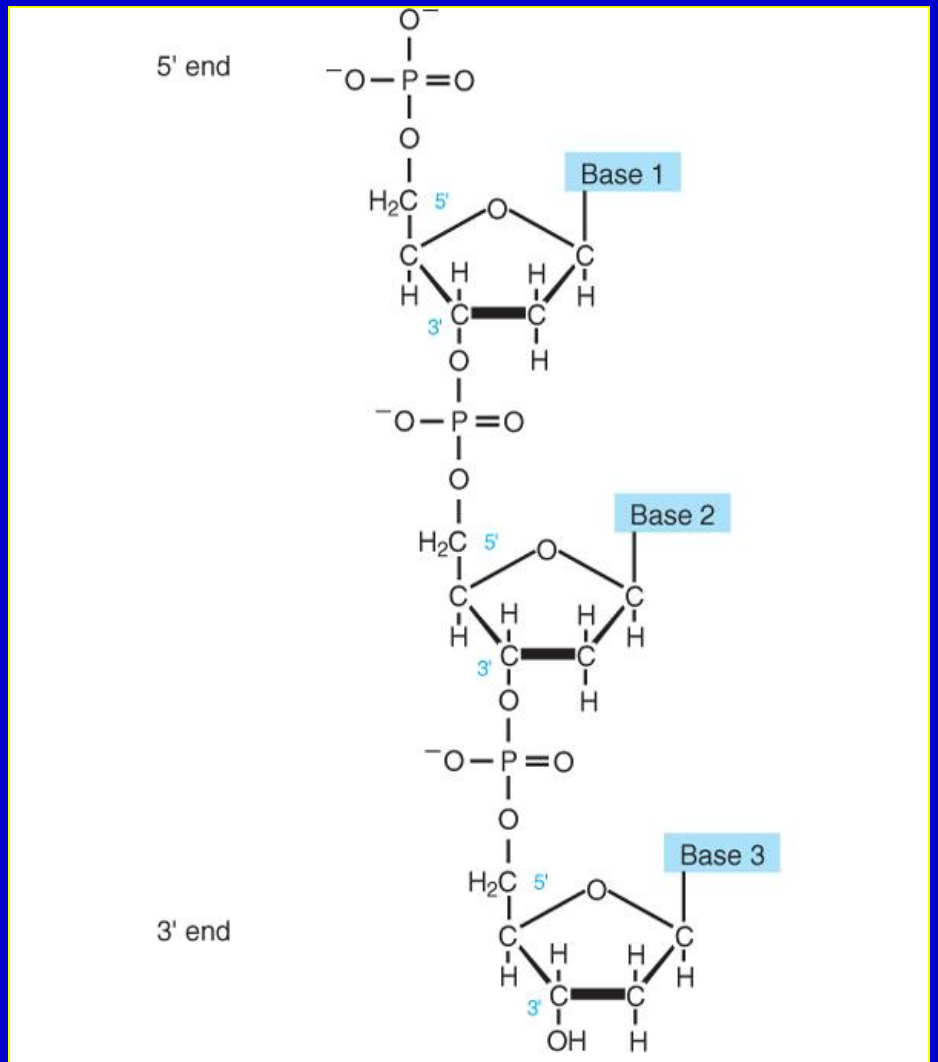
- Goal: identify genetic variation that contributes to human disease
- Diagnostic screening that could enable early/more effective intervention
- Understanding genetic/molecular basis of human disease, which could help to direct the development of therapeutic strategies

Deoxyribonucleic acid (DNA) is a polymer of nucleotides
5 carbon sugar, phosphate group, one of 4 nitrogenous bases

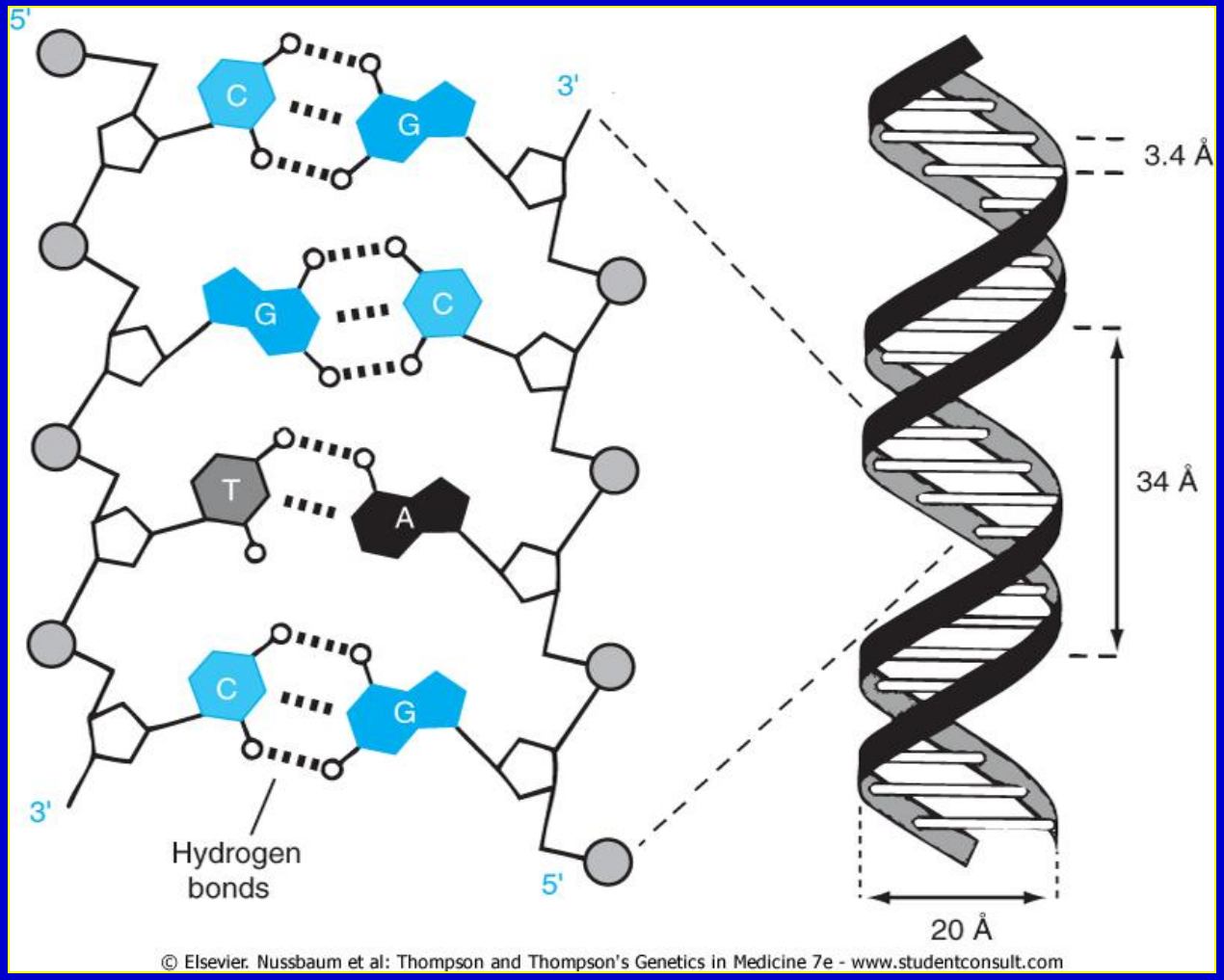


DNA

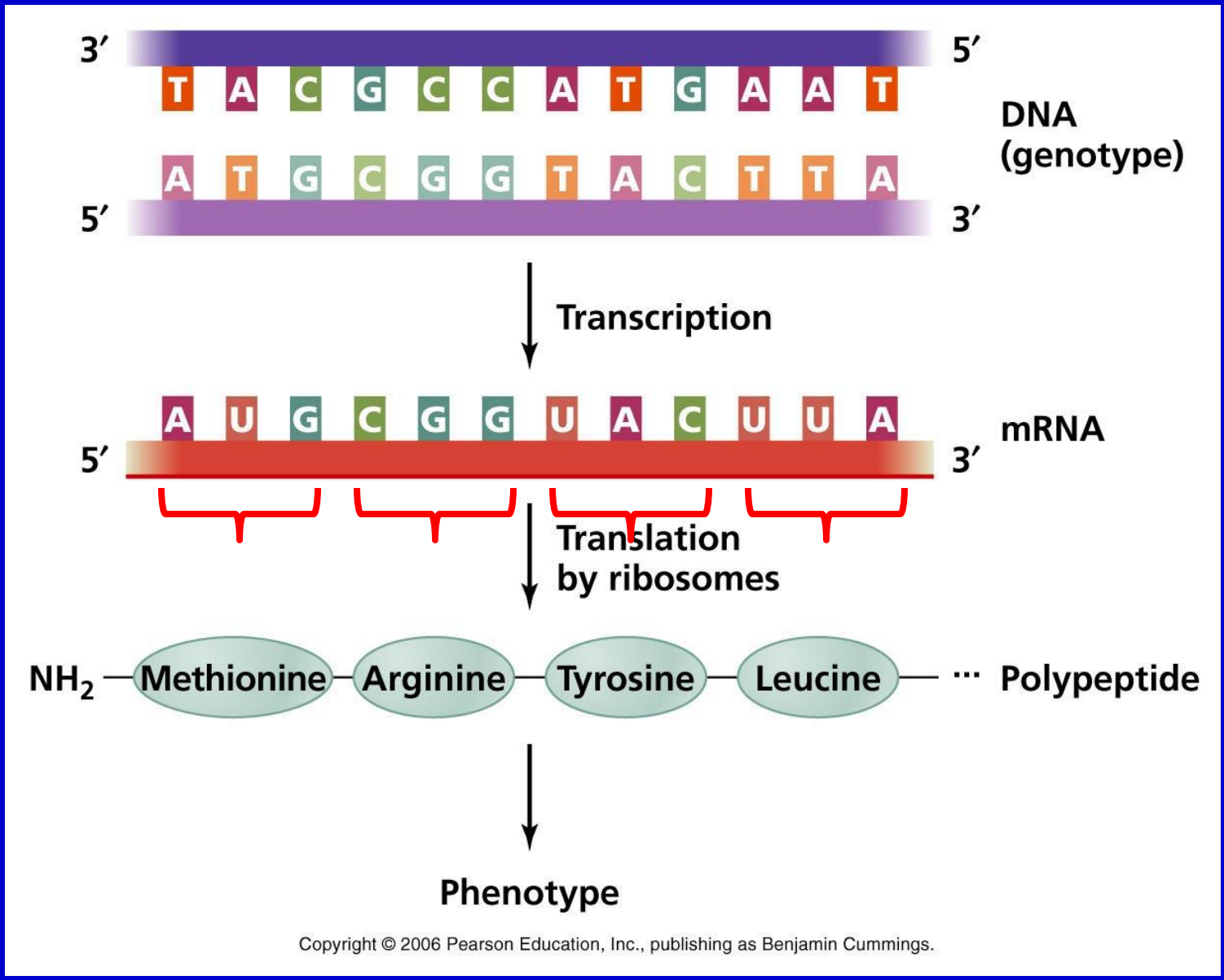
Polymer with sugar-phosphate backbone with nitrogenous bases



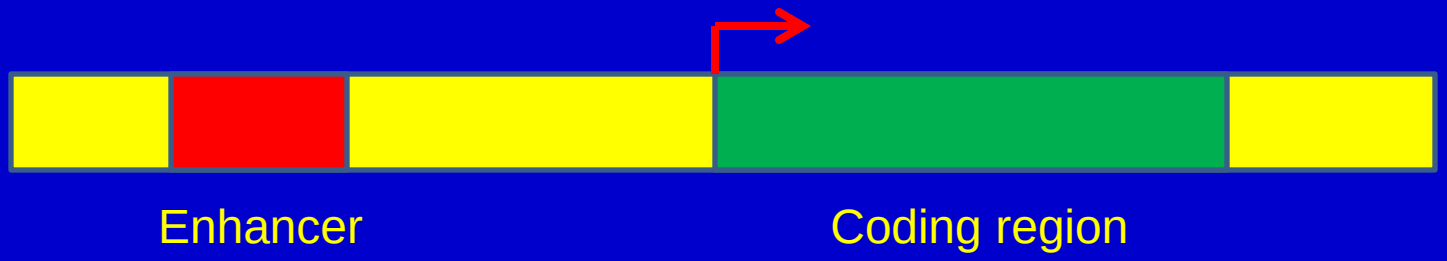
- DNA double helix with hydrogen bonds
- Hydrogen bonds create base pairs
  - G - C
  - A - T



DNA → RNA → Protein



DNA sequence is partitioned into functional units named genes

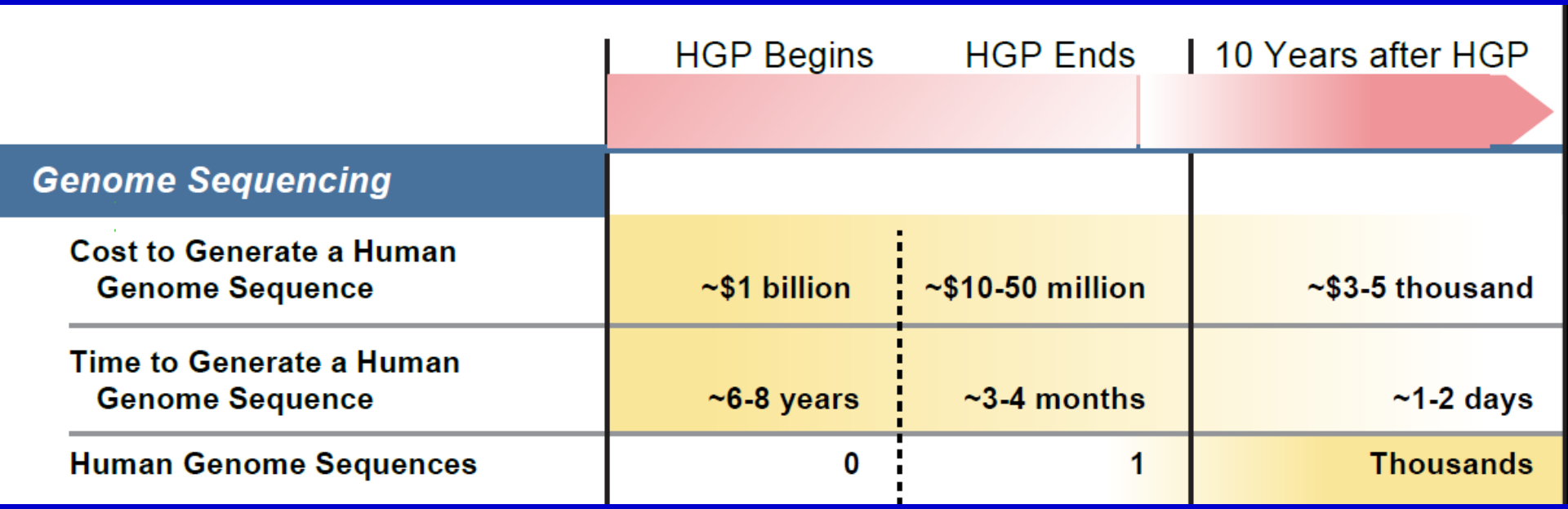


Human Genome

Chromosomes	46 (22 pairs of autosomes, X, Y)
Genes	20,000-25,000
Base-pairs	3,000,000,000 bases

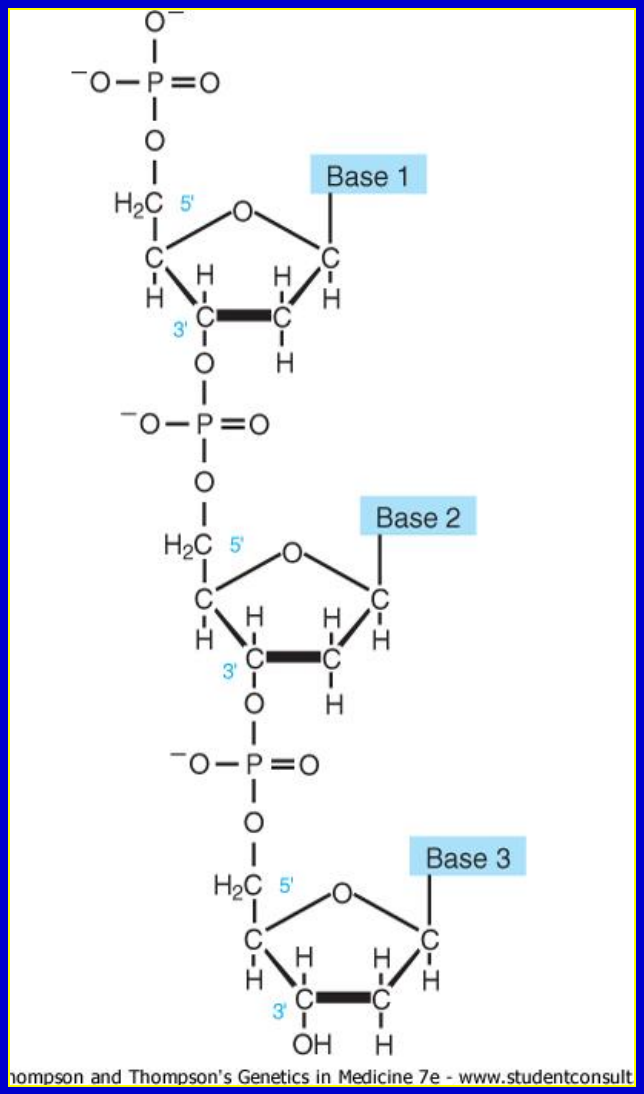
DNA sequencing becoming easier and cheaper

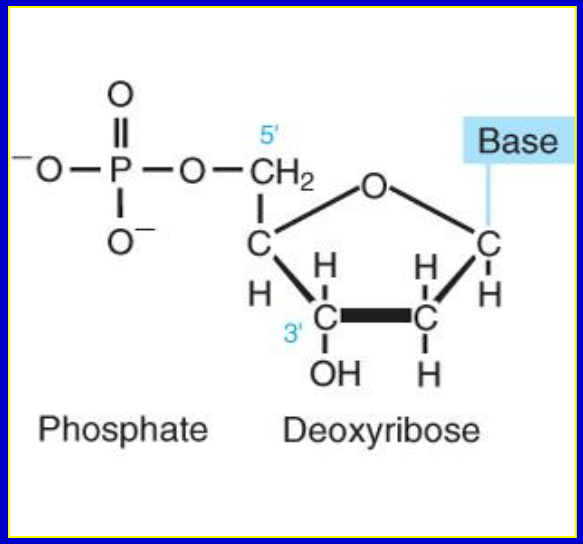
1990 2003



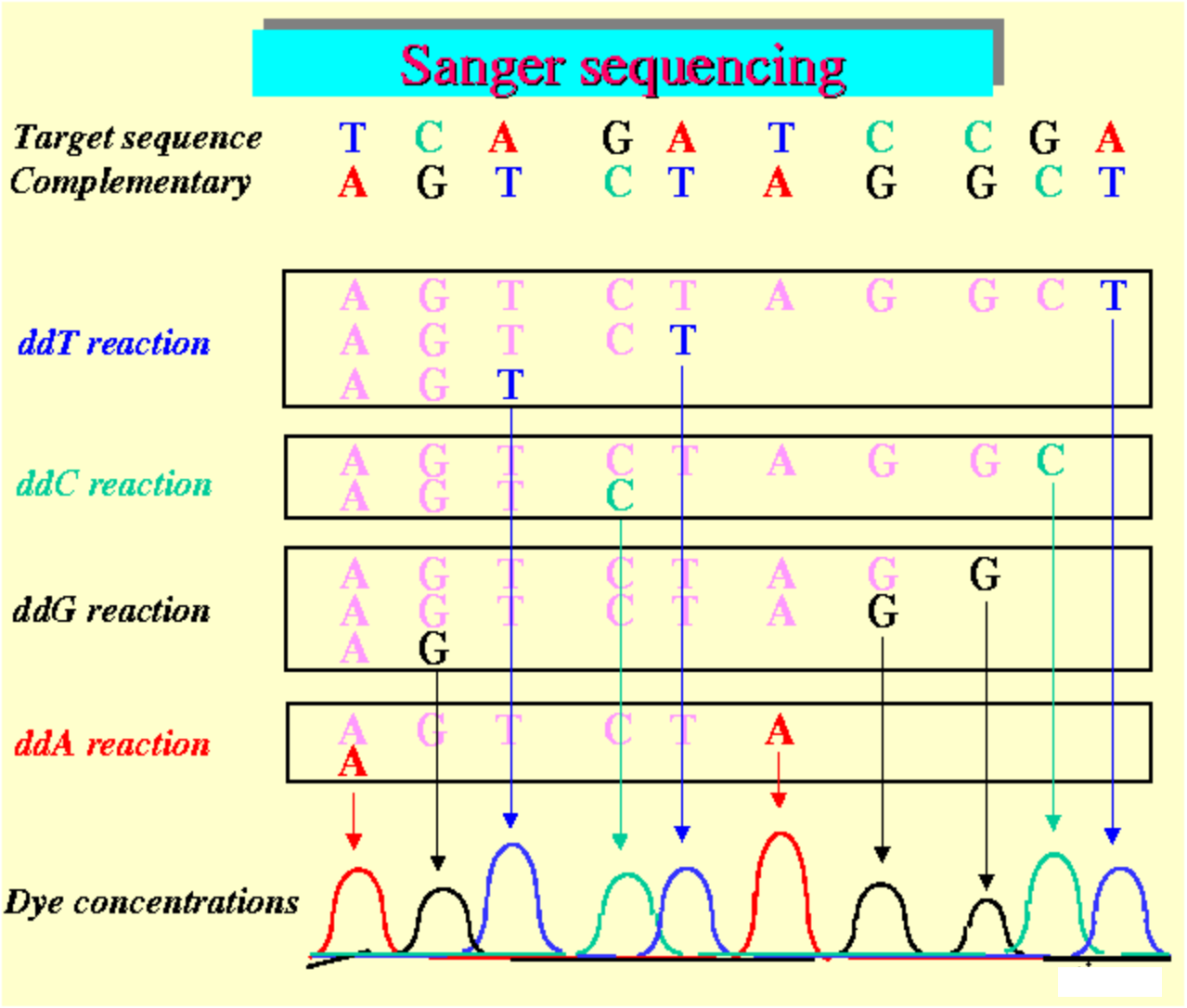
www.genome.gov
NHGRI

DNA elongation requires OH on 3' C





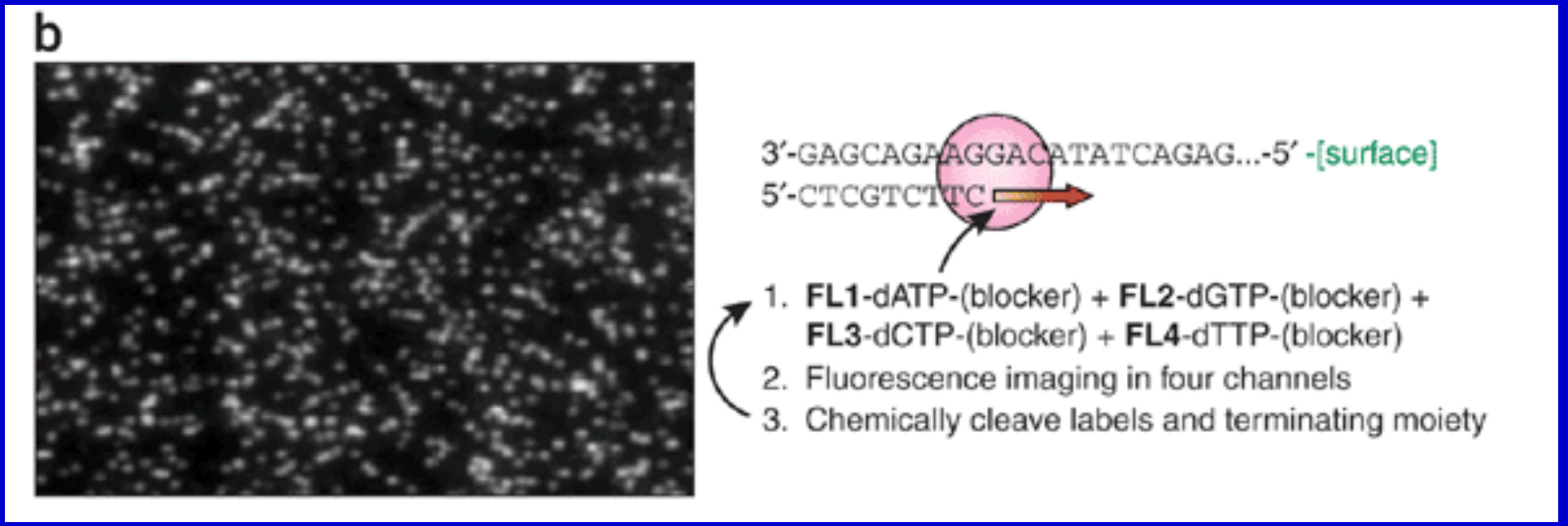
Sanger sequencing



http://stat.fsu.edu/~lilei/lilei/research/sanger-c.gif

Parallel sequencing (Next-gen)

- DNA fragmentation
- Immobilize fragments on a slide
- Copies of fragments spatially clustered on slide

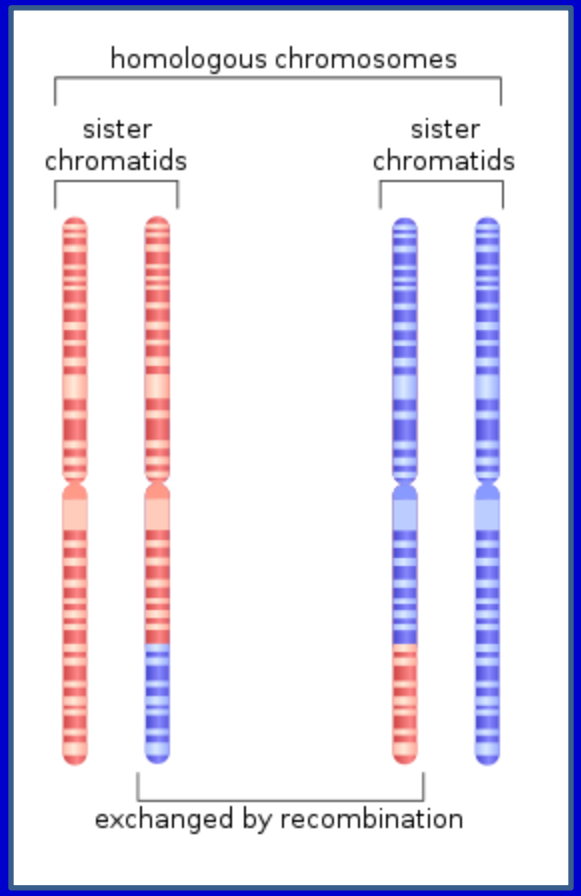


Shendure, 2008, Nature Biotechnology

Sequence to find variants

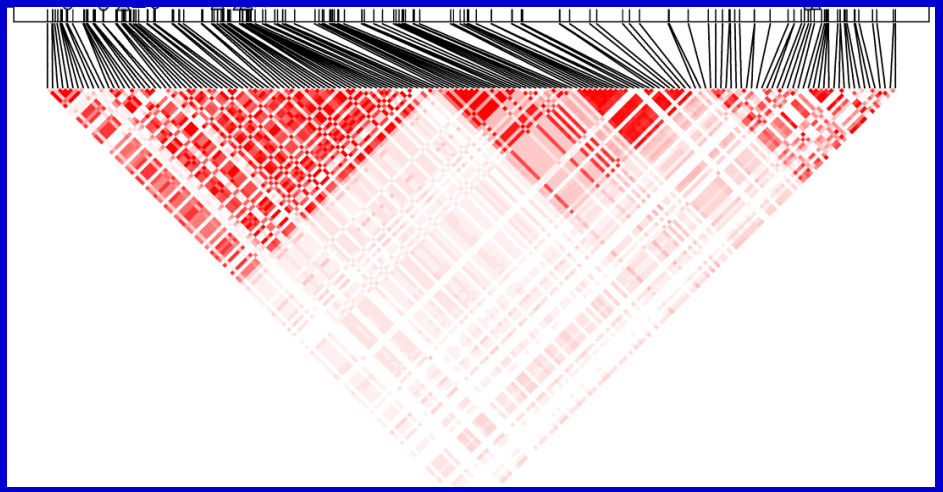
- International HapMap project
  - Sequenced 90 Northern and Western Europeans, 90 Africans (Yoruba), 45 Han Chinese in Beijing, 45 Japanese in Tokyo
  - 3,000,000 polymorphic sites (variants)
  - Most variants are single nucleotide polymorphisms (SNPs)
- 1000 genomes project
  - Sequenced 1000 people from 14 populations
  - 38,000,000 variants (~1% bases)

SNP linkage



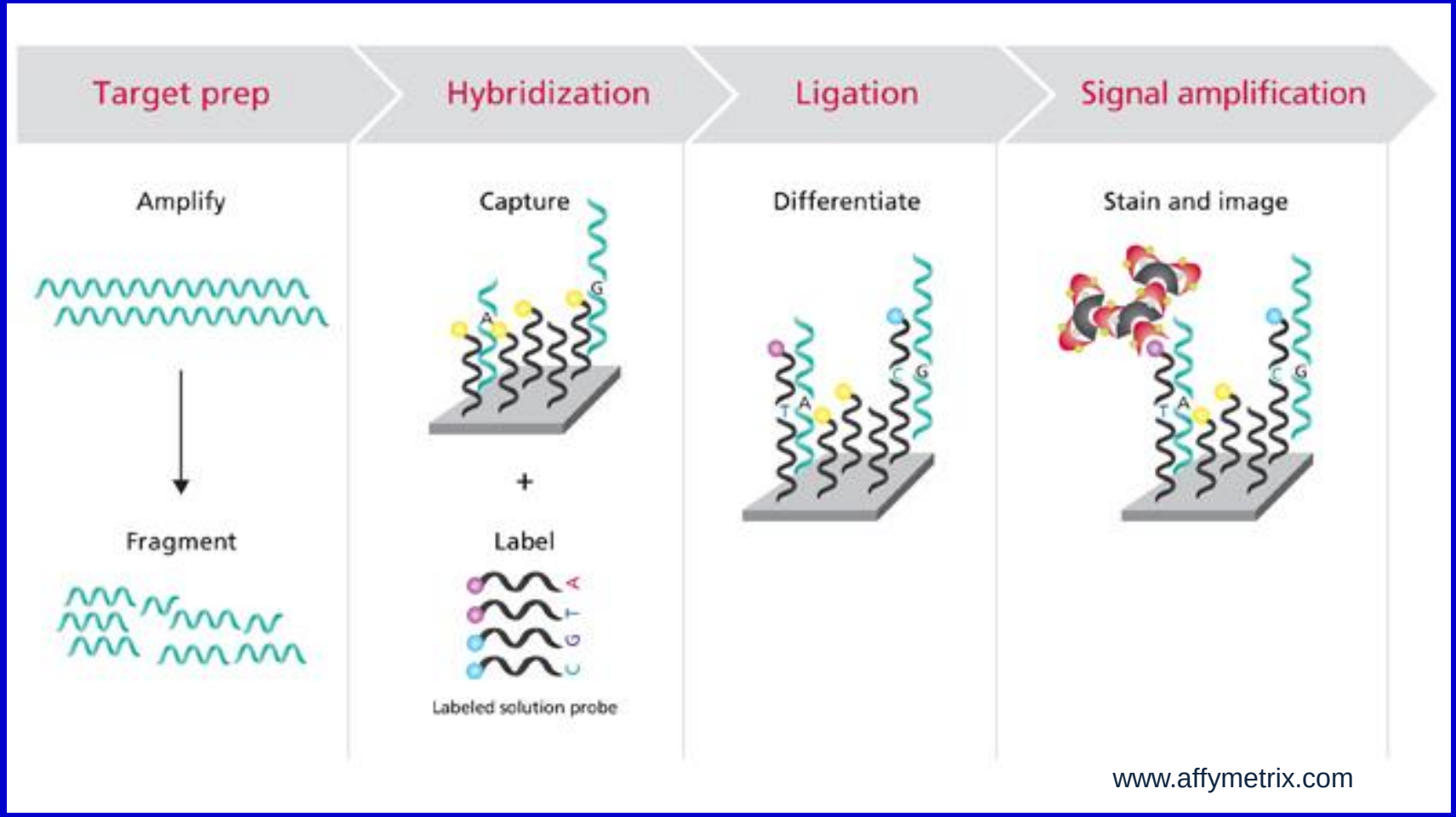
As distance between 2 SNPs decreases, frequency of recombination decreases

Results in blocks of correlated SNPs, named haplotypes



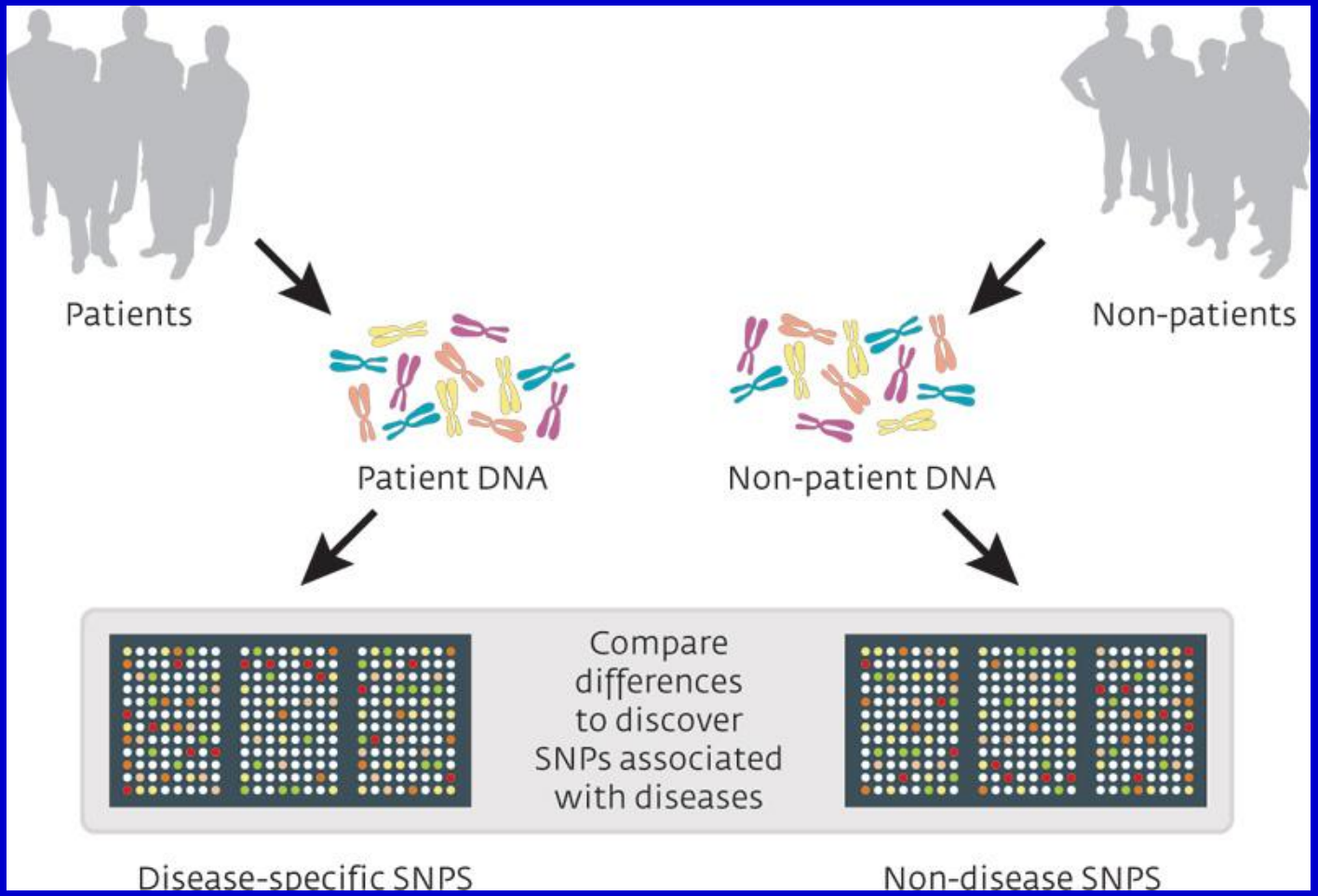
Adapted version of Figure 10-18 in Watson, JD et al (2003) *Molecular Biology of the Gene* (5th ed.), Pearson/Benjamin Cummings, p. 283

Genotype subset of variants



Our group genotyped 1 million SNPs in close to 10,000 people

Genome-wide association study (GWAS)



Multiple testing

- At α 0.05, 5% of results are expected to be false positives
- 5% of 1 million SNPs = 50,000 false positives
- Bonferroni correction adjusts α to keep family-wise error rate at 5%
  - Significance threshold 0.05/# tests
  - $0.05/1x10^{6} = 5x10^{-8}$

Latest genetic experiment