open.michigan

Author(s): David Ginsburg, M.D., 2012

Attribution Key

for more information see: http://open.umich.edu/wiki/AttributionPolicy

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

PD-GOV **Public Domain – Government**: Works that are produced by the U.S. Government. (17 USC § 105)

PD-EXP **Public Domain – Expired**: Works that are no longer protected due to an expired copyright term.

PD-SELF **Public Domain – Self Dedicated**: Works that a copyright holder has dedicated to the public domain.

ZERO **Creative Commons – Zero Waiver**

BY **Creative Commons – Attribution License**

BY-SA **Creative Commons – Attribution Share Alike License**

BY-NC **Creative Commons – Attribution Noncommercial License**

BY-NC-SA **Creative Commons – Attribution Noncommercial Share Alike License**

GNU-FDL **GNU – Free Documentation License**

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

PD-INEL **Public Domain – Ineligible**: Works that are ineligible for copyright protection in the U.S. (17 USC § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

FAIR USE **Fair Use**: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (17 USC § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

DNA Sequence Variation

M1 Patients and Populations

David Ginsburg, MD

Fall 2012

University of Michigan
Medical School

Relationships with Industry

UMMS faculty often interact with pharmaceutical, device, and biotechnology companies to improve patient care, and develop new therapies. UMMS faculty disclose these relationships in order to promote an ethical & transparent culture in research, clinical care, and teaching.

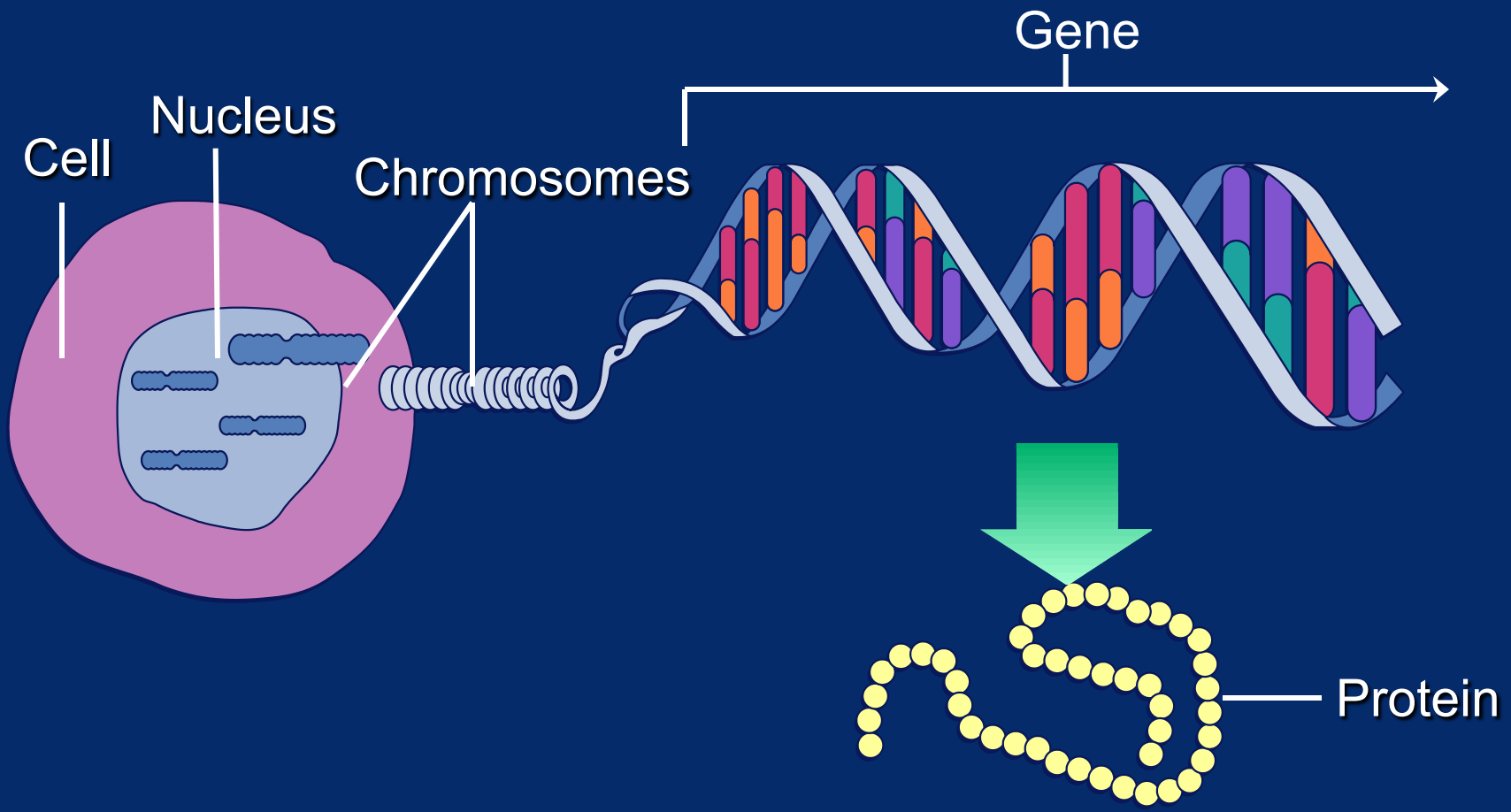
- I am a member of the Board of Directors for Shire plc.
- I am a member of the Scientific Advisory Boards for Portola Pharmaceuticals and Catalyst Biosciences.
- I benefit from license/patent royalty payments to Boston Children's Hospital (VWF) and the University of Michigan (ADAMTS13).

Disclosure required by the UMMS Policy on Faculty Disclosure of Industry Relationships to Students and Trainees.

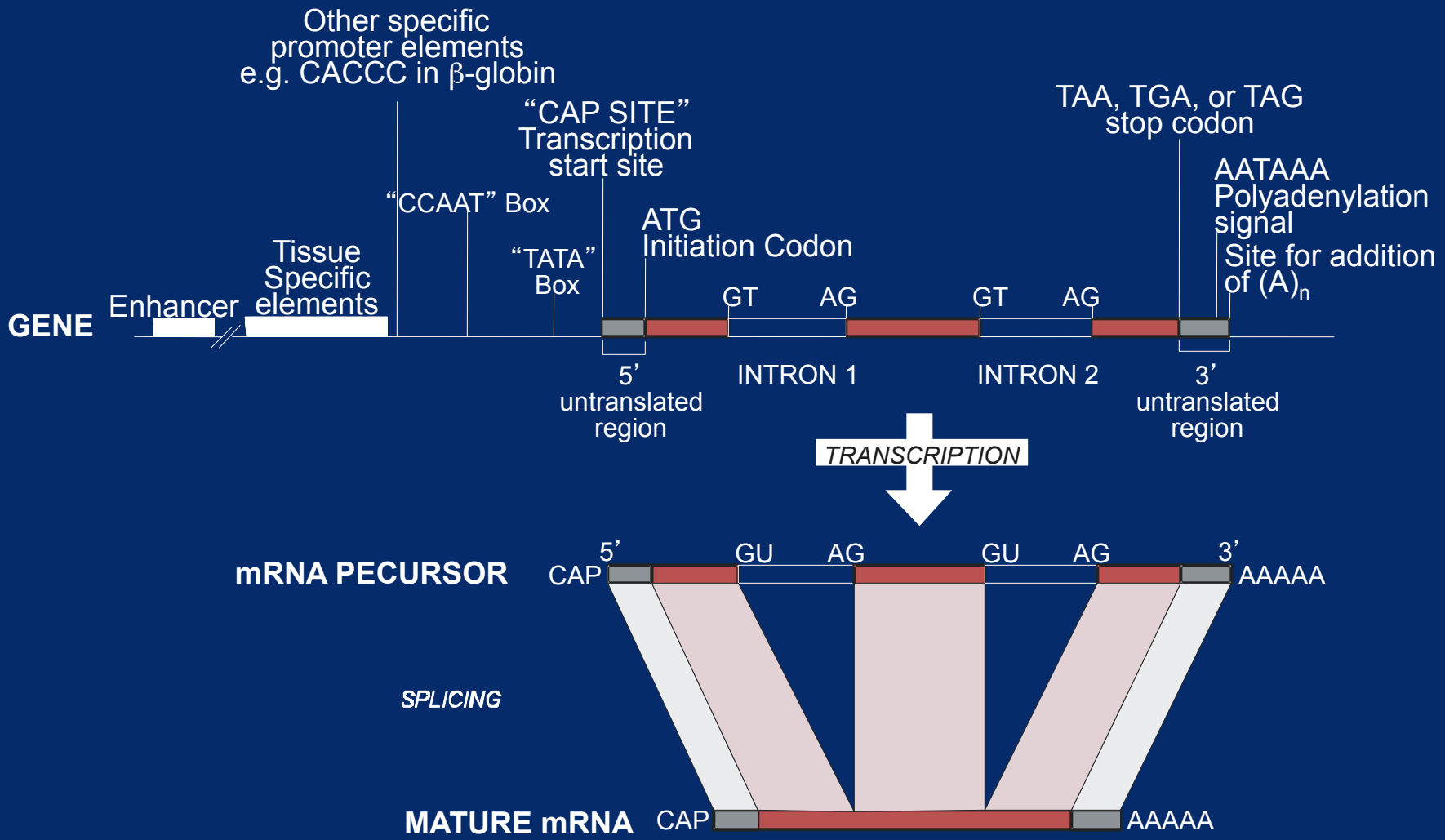
Learning Objectives

- Understand the meaning of DNA sequence and amino acid polymorphisms.
- Recognize the different types of DNA sequence polymorphisms:
  - STR, SNP, CNV
- Know the different classes of DNA mutation:
  - Point mutations (silent, missense, nonsense, frameshift, splicing, regulatory) insertion/deletions, rearrangements
- Understand how to distinguish a disease-causing mutation from a neutral DNA sequence variation

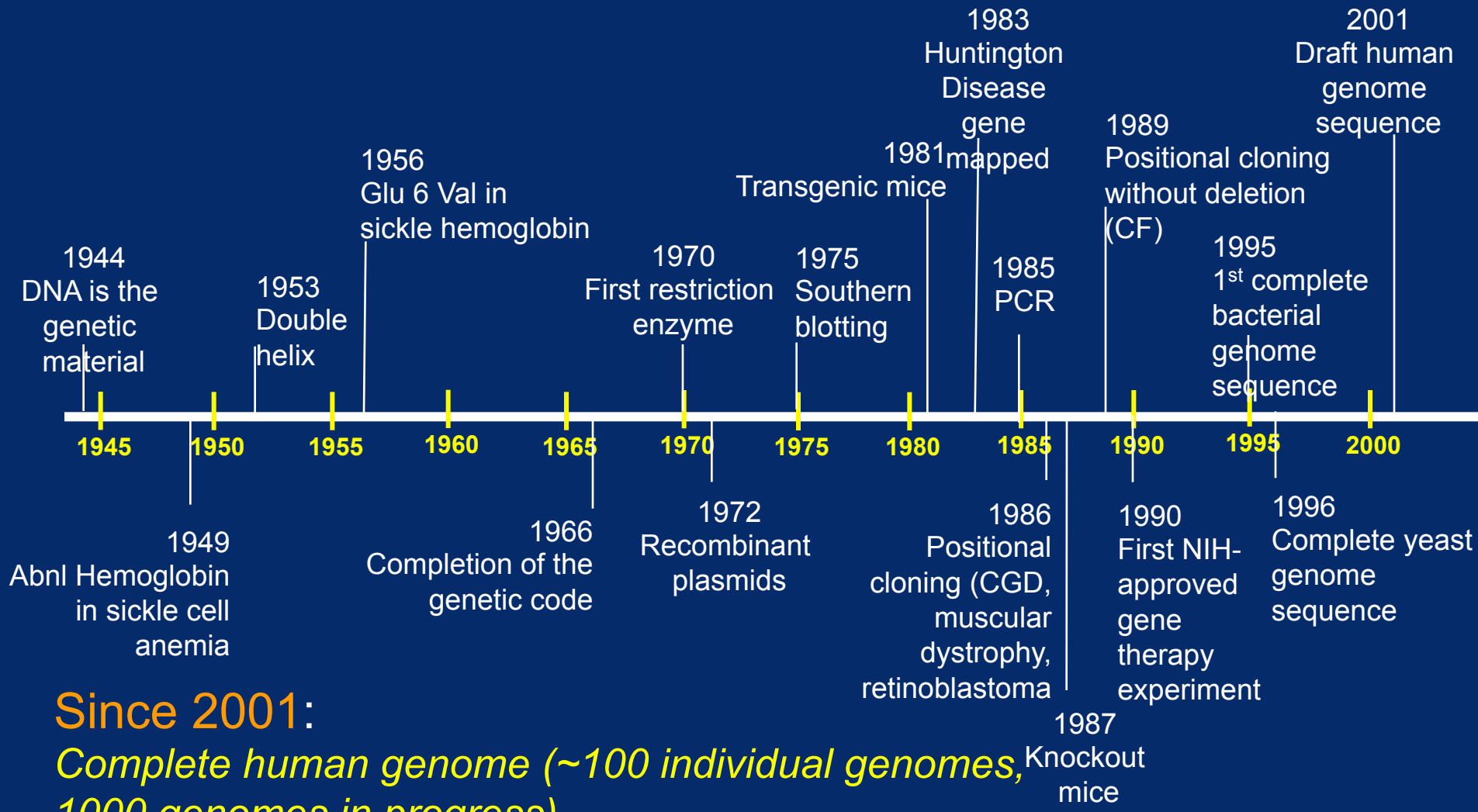
Chromosomes, DNA, and Genes



Adapted from *ASCO teaching slides*



 Gelehrter, Collins and Ginsburg: *Principles of Medical Genetics 2E;* Figure 5.1



Since 2001:

Complete human genome (~100 individual genomes, 1000 genomes in progress)

Growing index of human variation: human hapmap, dbSNP ; 1000 genomes

Complete genomes of >6500 other species

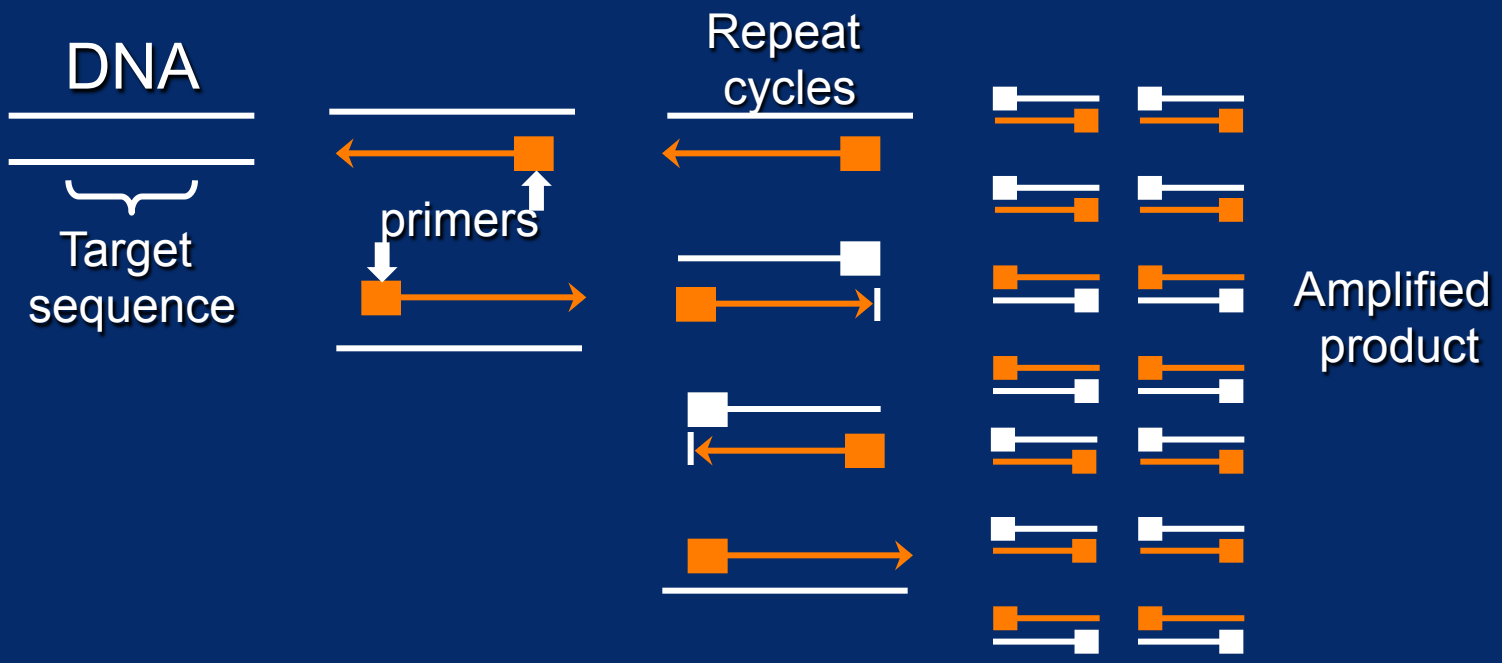
Gelehrter, Collins and Ginsburg: *Principles of Medical Genetics 2E;* Figure 5.4

DNA Samples and PCR

Blood sample

cheek swab
urine sample
Forensic specimens

DNA for analysis



Hybridization array: gene chip

DNA chip v6.0:

~1 million SNPs

RNA expression:

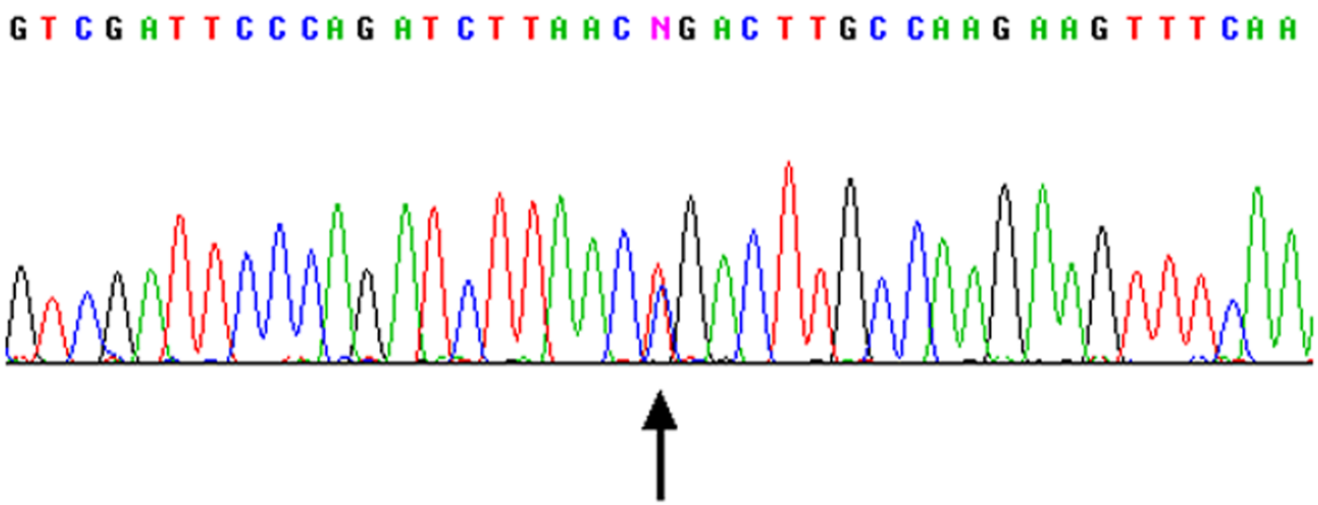
All ~ 20,000 genes

CGH*:*

Survey whole genome for large deletions, insertions, rearrangements


Image of DNA
sequencing
process removed

DNA Sequencing



New Sequencing Technologies

Searching for Cheaper Genome Sequencers

Company	Format	Read Length (bases)	Expected Throughput MB (million bases)/day
454 Life Sciences	Parallel bead array	100	96
Agencourt Bioscience	Sequencing by ligation	50	200
Applied Biosystems	Capillary electrophoresis	1000	3–4
Microchip Biotechnologies	Parallel bead array	850-1000	7
NimbleGen Systems	Map and survey microarray	30	100
Solexa	Parallel microchip	35	500
LI-COR	Electronic microchip	20,000	14,000
Network Biosystems	Biochip	800+	5
VisiGen Biotechnologies	Single molecule array	NA	1000

Generation next. Companies racing for the $1000 genome sequence strive simultaneously for low cost, high accuracy, the ability to read long stretches of DNA, and high throughput.

Learning Objectives

- Understand the meaning of DNA sequence and amino acid polymorphisms.
- Recognize the different types of DNA sequence polymorphisms:
  - STR, SNP, CNV
- Know the different classes of DNA mutation:
  - Point mutations (silent, missense, nonsense, frameshift, splicing, regulatory) insertion/deletions, rearrangements
- Understand how to distinguish a disease-causing mutation from a neutral DNA sequence variation

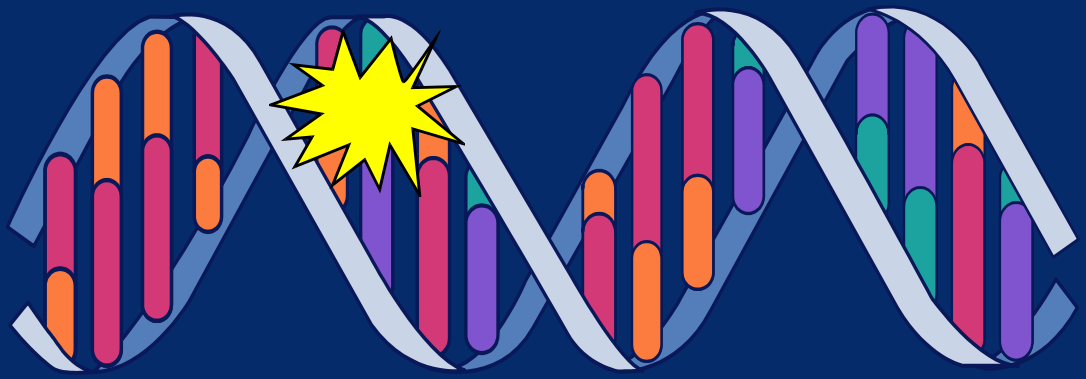
DNA Sequence Variation

- DNA Sequence Variation:
  - Human to human: ~0.1% (1:1000 bp)
    - Human genome = $3X10^9$ bp X 0.1% =~$3X10^6$ DNA common variants
  - Human to chimp: ~1-2%
  - More common in "junk" DNA: introns, intergenic regions

- **poly·mor·phism**
  Pronunciation: "päl-i-'mor-"fiz-&m
  Function: *noun*
  : the quality or state of existing in or assuming different forms: as **a (1) :** existence of a species in several forms independent of the variations of sex **(2) :** existence of a gene in several allelic forms (3) : existence of a molecule (as an enzyme) in several forms in a single species

Mutations

A mutation is a change in the “normal” base pair sequence



- Can be:
  - a single base pair substitution
  - a deletion or insertions of 1 or more base pairs (indel)
  - a larger deletion/insertion or rearrangement

Adapted from *ASCO teaching slides*

Polymorphisms and Mutations

- Genetic polymorphism:
  - Common variation in the population:
    - Phenotype (eye color, height, etc)
    - genotype (DNA sequence polymorphism)
  - Frequency of minor allele(s) $\geq$ 1%

- DNA (and amino acid) sequence variation:
  - Most common allele $\leq$ 0.99 = polymorphism (minor allele(s) $\geq$ 1%)
  - Variant alleles < 0.01 = rare variant

- Mutation-- any change in DNA sequence
  - Silent vs. amino acid substitution vs. other
  - neutral vs. disease-causing
  - $1X10^{-8}$/bp/generation (~70 new mutations/individual)

- balanced polymorphism= disease + polymorphism

- Common but incorrect usage:
  - "mutation vs. polymorphism"

All DNA sequence variation arises via mutation of an ancestral sequence

	< 1%	**≥ 1%**
"Normal"	Rare variant or "private" polymorphism	polymorphism
"Disease"	Disease mutation	*Example: Factor V Leiden (thrombosis) 5% allele frequency*

Common but incorrect usage:

"a disease-causing mutation" **OR** *"a polymorphism"*

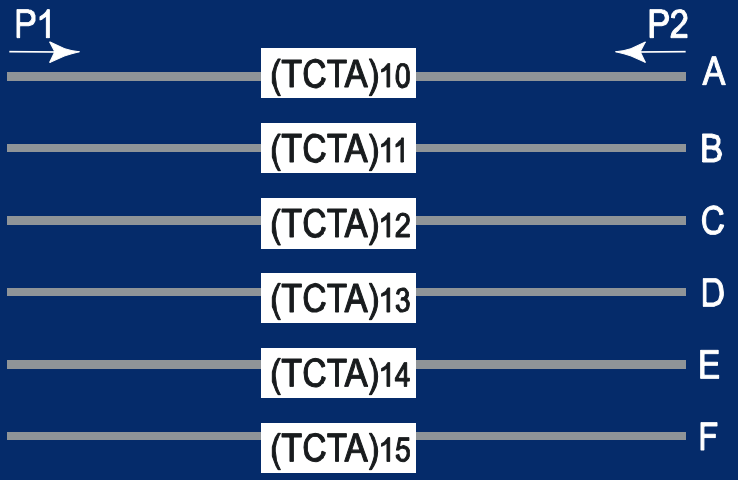
Learning Objectives

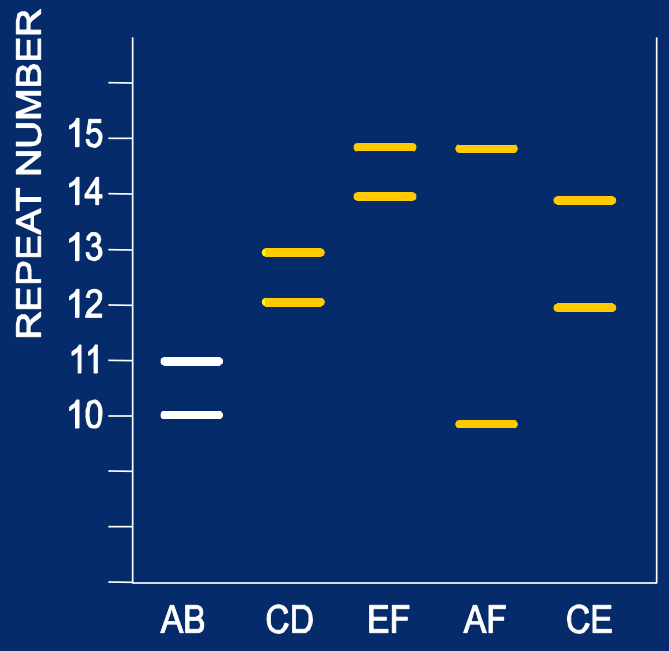
- Understand the meaning of DNA sequence and amino acid polymorphisms.
- Recognize the different types of DNA sequence polymorphisms:
  - STR, SNP, CNV
- Know the different classes of DNA mutation:
  - Point mutations (silent, missense, nonsense, frameshift, splicing, regulatory) insertion/deletions, rearrangements
- Understand how to distinguish a disease-causing mutation from a neutral DNA sequence variation

Types of DNA Sequence Variation

- RFLP: **R**estriction **F**ragment **L**ength **P**olymorphism
- VNTR: **V**ariable **N**umber of **T**andem **R**epeats
  - or minisatellite
  - ~10-100 bp core unit
- SSR : **S**imple **S**equence **R**epeat
  - or STR (simple tandem repeat)
  - or microsatellite
  - ~1-5 bp core unit
- SNP: **S**ingle **N**ucleotide **P**olymorphism
  - Commonly used to also include rare variants (SNVs)
- Insertions or deletions
  - INDEL – small (few nucleotides) insertion or deletion
- Rearrangement (inversion, duplication, complex rearrangement)
  - CNV: **C**opy **N**umber **V**ariation

STR



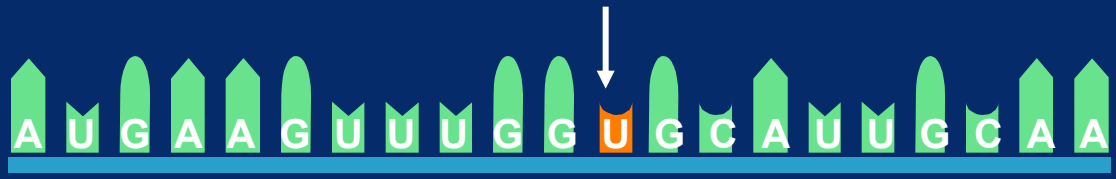


Gelehrter, Collins and Ginsburg: *Principles of Medical Genetics 2E;* Figure 5.22

SNP

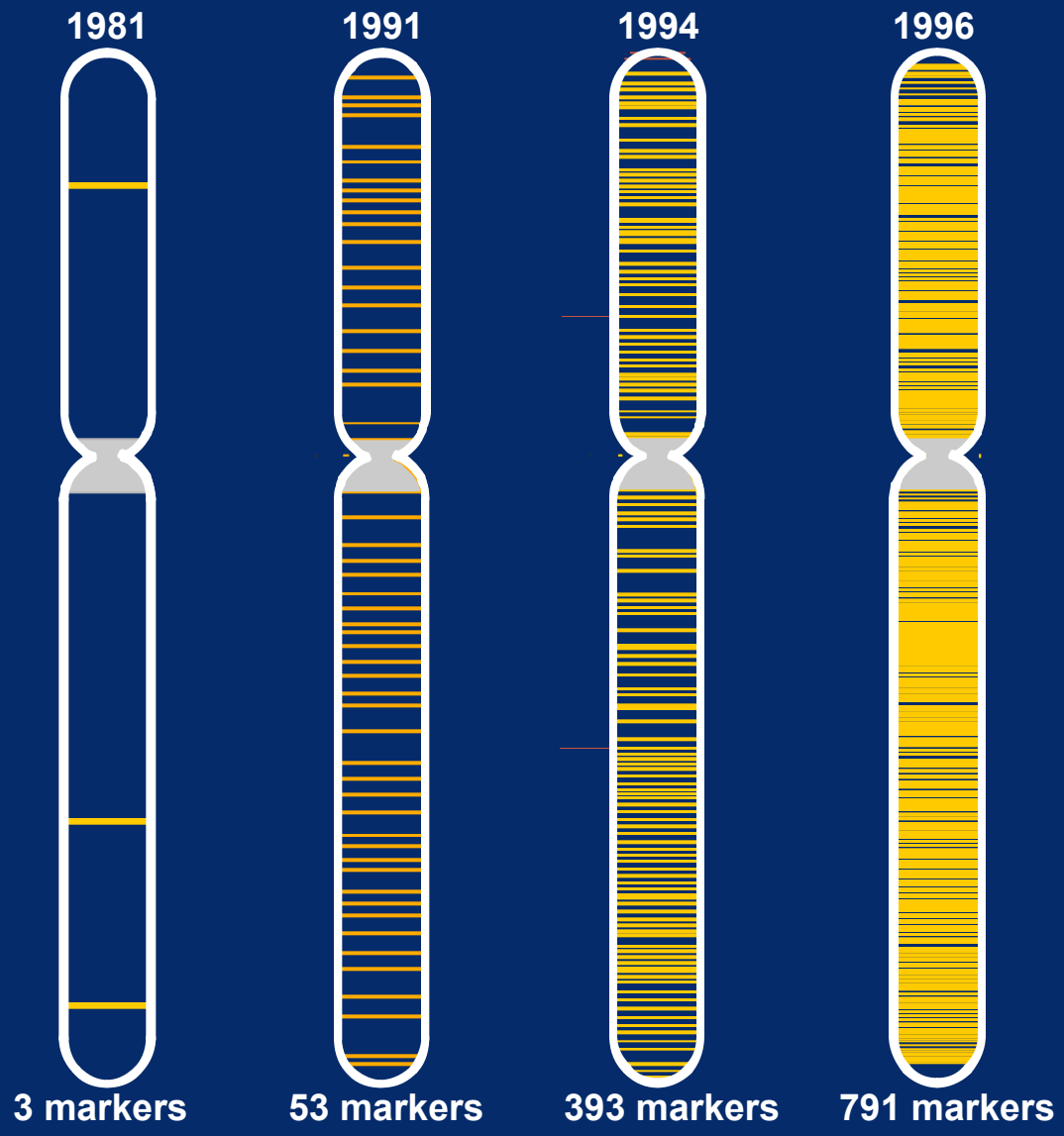
Allele 1

Allele 2



- Most are “silent”
- Intragenic
- Promoters and other regulatory sequences
- Introns
- Exons
  - 5’ and 3’ untranslated regions
  - Coding sequence (~1-2% of genome)

Human Chromosome 4



2010

- 23,653,737 total human entries in dbSNP

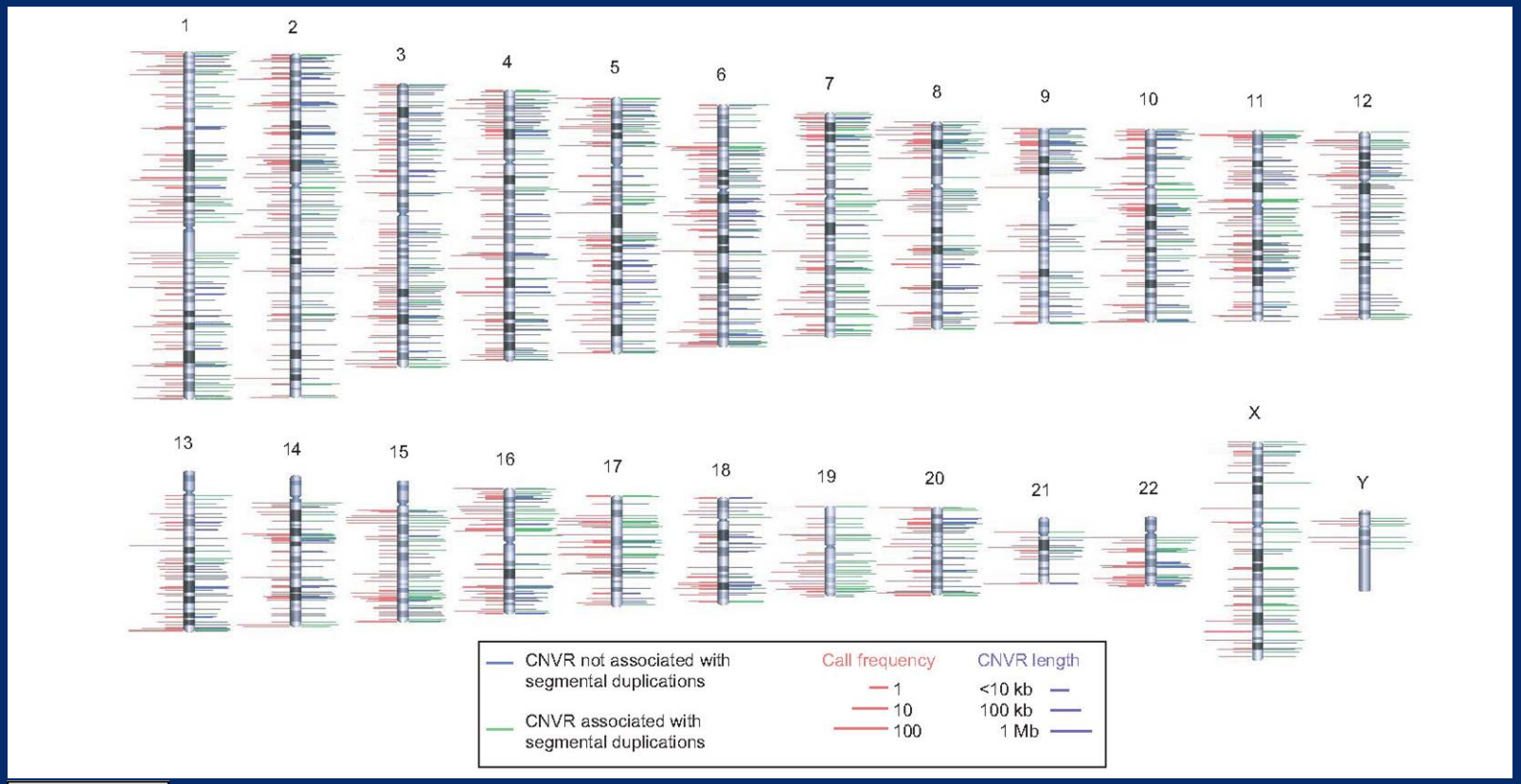
http://www.ncbi.nlm.nih.gov/projects/SNP/

- Chromosome 4
  - 4,311,728 SNPs
- ~1M SNP chip commercially available

Gelehrter, Collins and Ginsburg: *Principles of Medical Genetics 2E;* Figure 10.3

Copy Number Variation (CNV)

- Kb to Mb in size (average ~250 Kb)
- >>2000 known, affect ~12% of human genome
- ? ~100 / person
- ? Role in human disease/normal traits



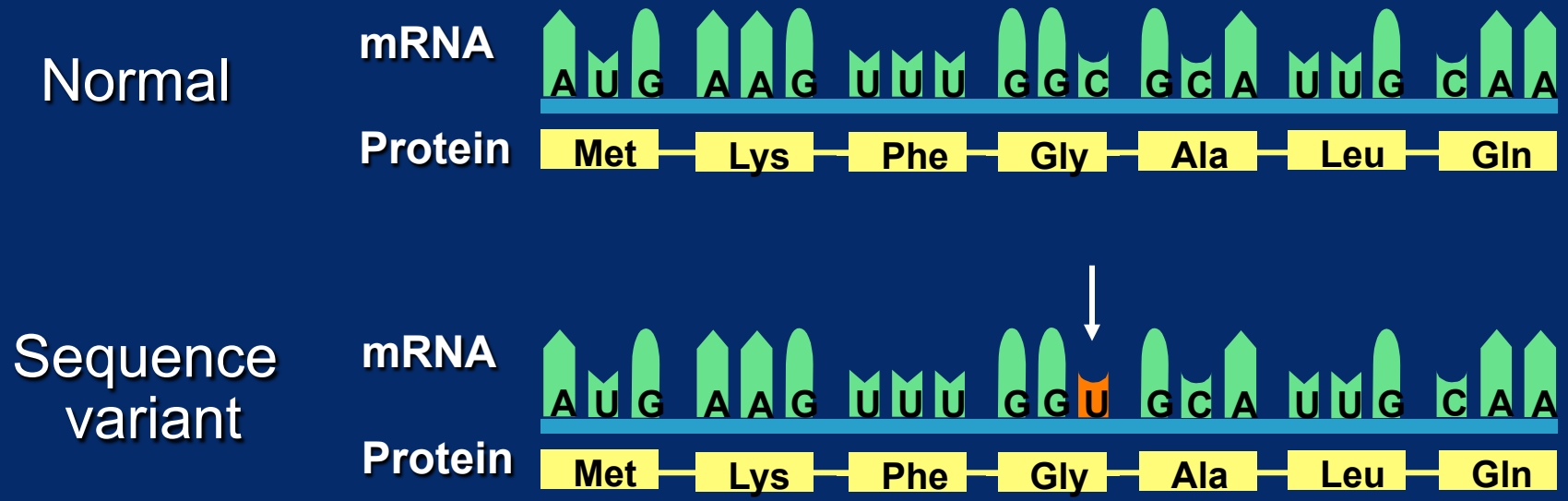
 Nature **444**:444, Nov 2006.

Learning Objectives

- Understand the meaning of DNA sequence and amino acid polymorphisms.
- Recognize the different types of DNA sequence polymorphisms:
  - STR, SNP, CNV
- Know the different classes of DNA mutation:
  - Point mutations (silent, missense, nonsense, frameshift, splicing, regulatory) insertion/deletions, rearrangements
- Understand how to distinguish a disease-causing mutation from a neutral DNA sequence variation

Silent Sequence Change

(Synonymous SNP)

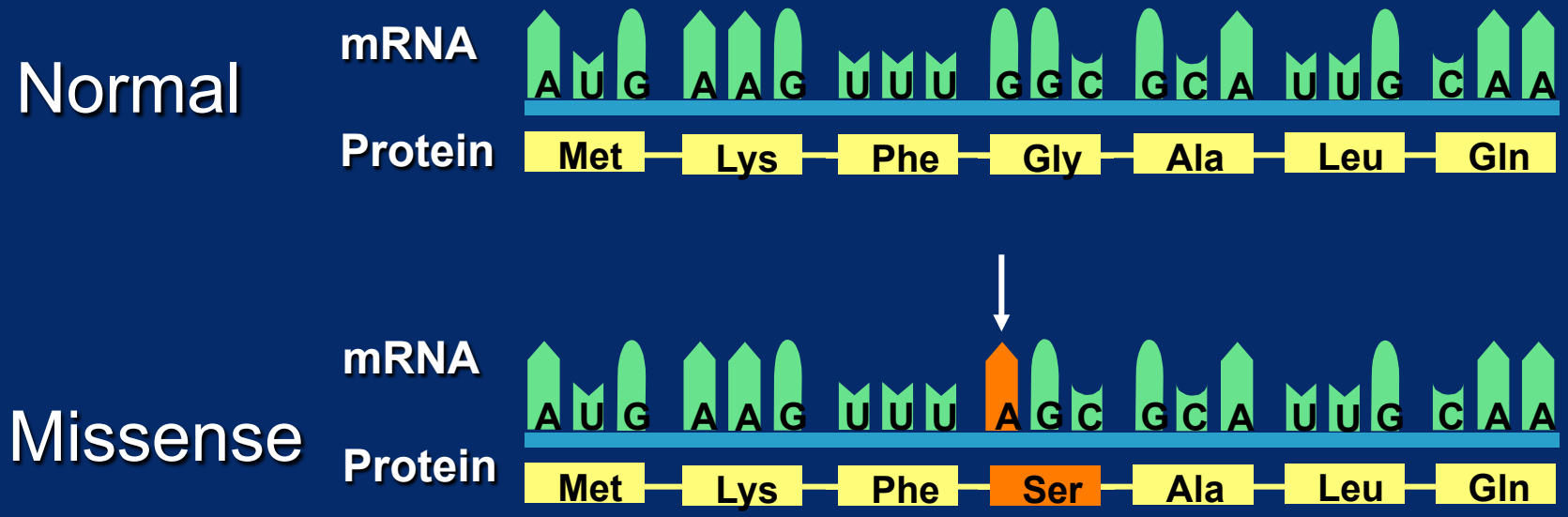


Changes that do not alter the encoded amino acid

Adapted from *ASCO teaching slides*

Missense Mutation

(Nonynonymous SNP)

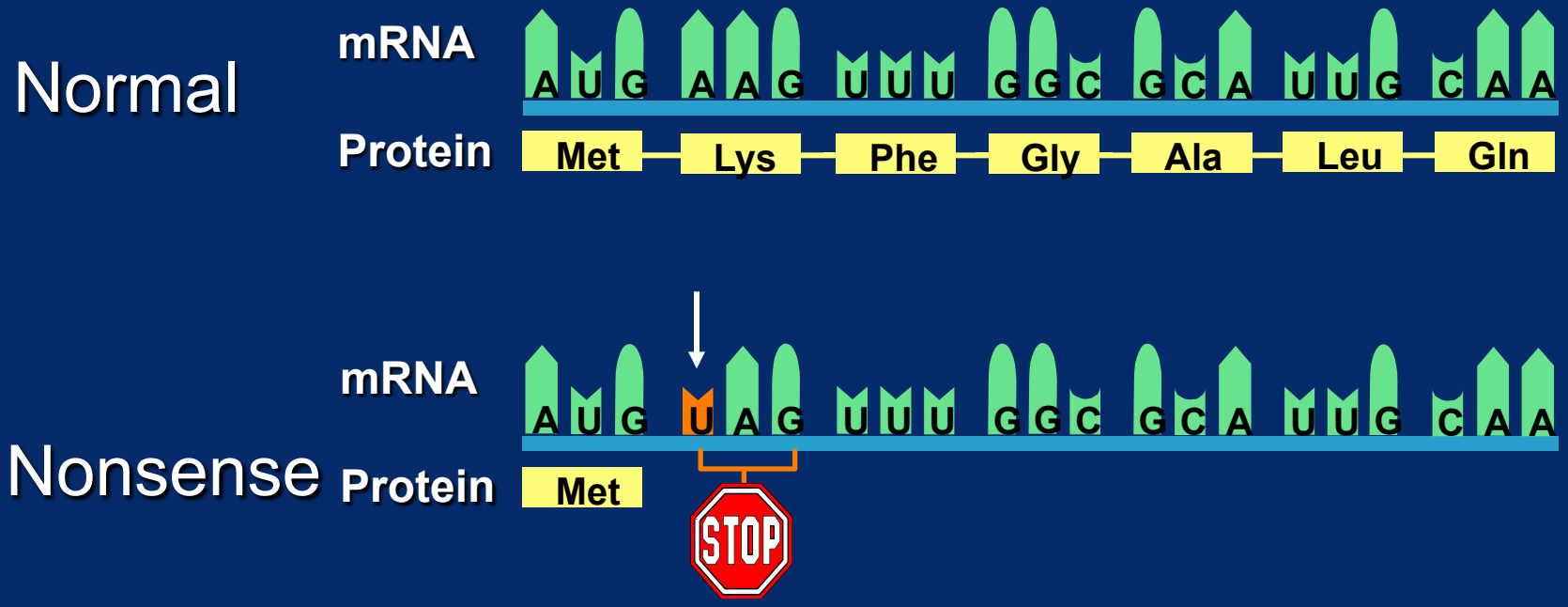


Missense: changes to a codon for another amino acid
(can be harmful mutation or neutral variant)

Adapted from *ASCO teaching slides*

Nonsense Mutation

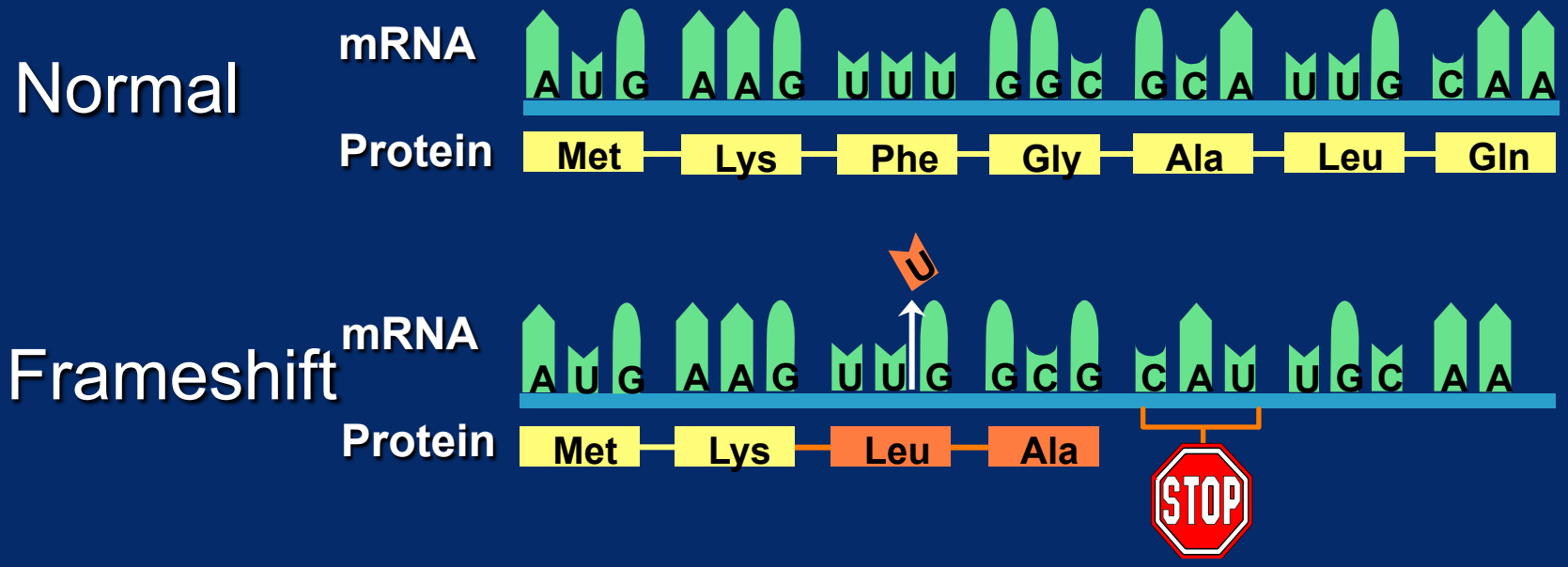
(Nonynonymous SNP)



Nonsense: change from an amino acid codon to a stop codon, producing a shortened protein

Adapted from *ASCO teaching slides*

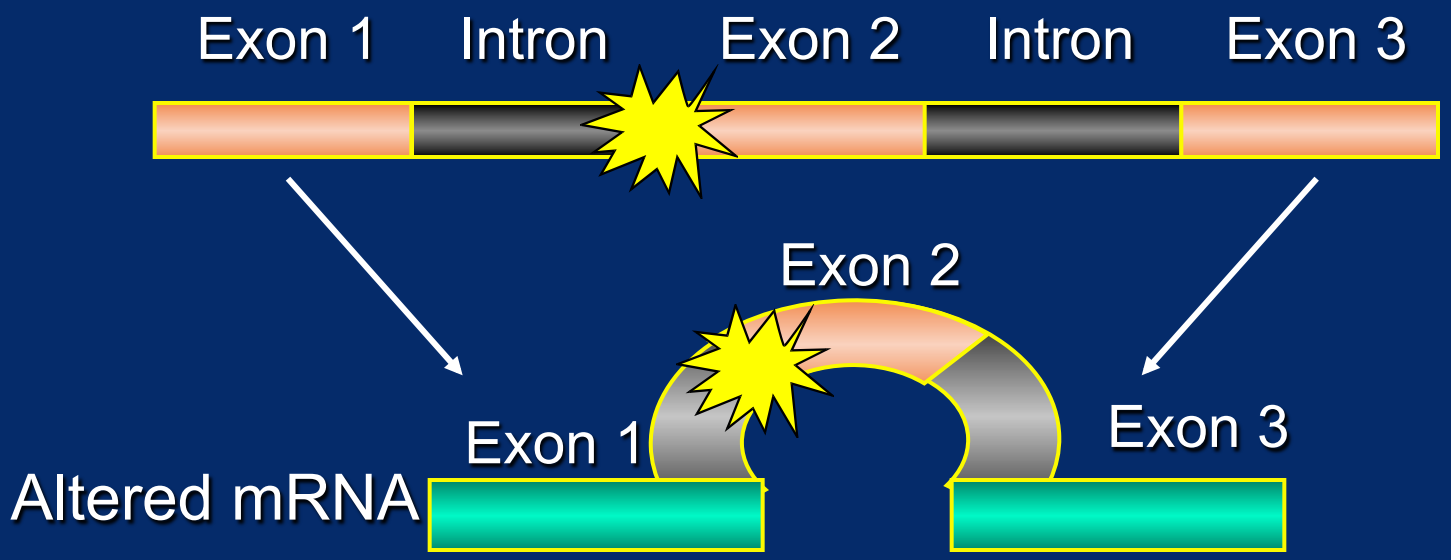
Frameshift Mutations



Frameshift: insertion or deletion of base pairs, producing a stop codon downstream and (usually) shortened protein

Adapted from *ASCO teaching slides*

Splice-site Mutations



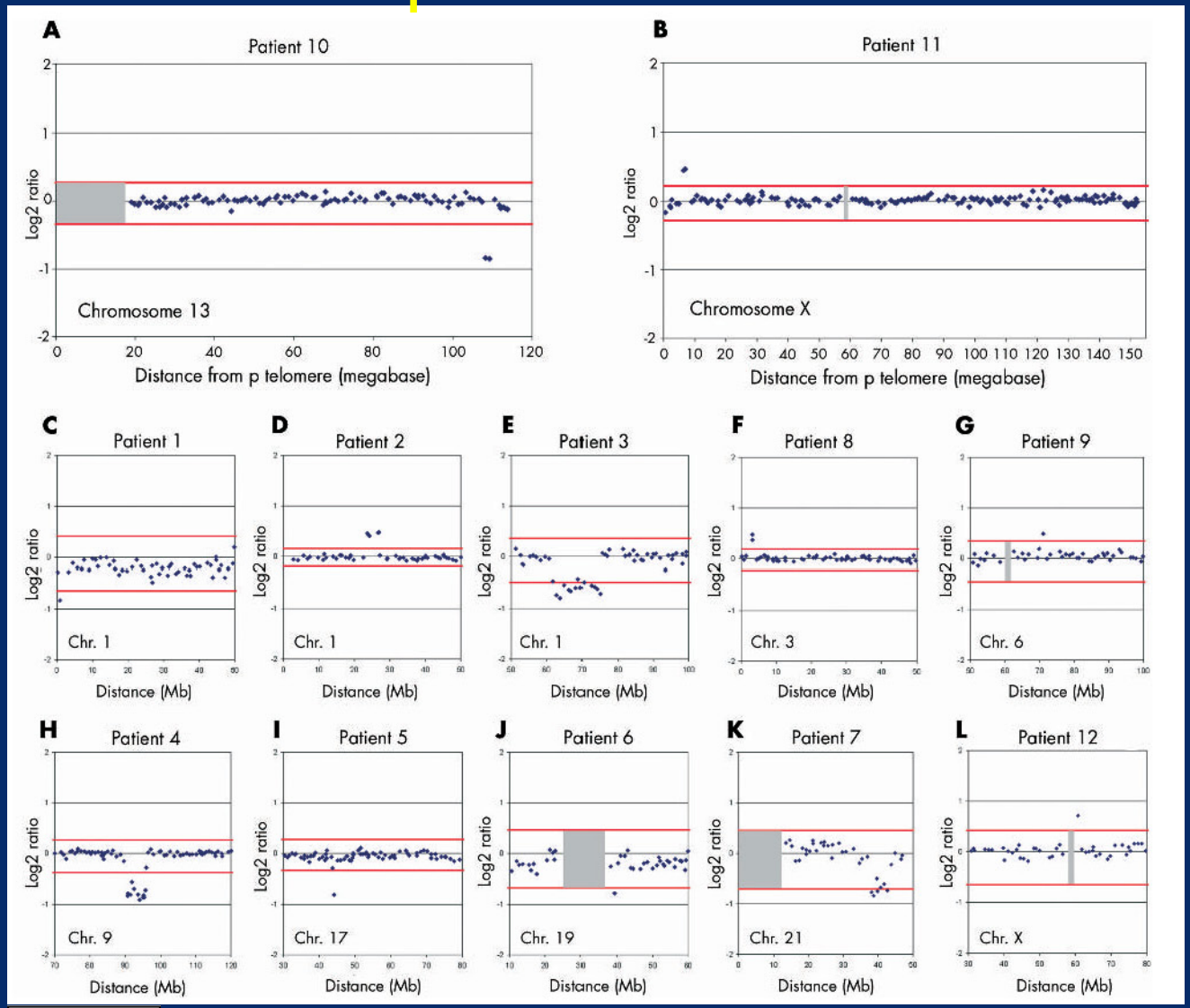
Splice-site mutation: a change that results in altered RNA sequence

Adapted from *ASCO teaching slides*

Other Types of Mutations

- Mutations in regulatory regions of the gene
- Large deletions or insertions
- Chromosomal translocations or inversions

Potentially pathogenic CNV detected in ~10-20% of unexplained intellectual disability



 J. Med. Genet. 2004, 41:241.

Learning Objectives

- Understand the meaning of DNA sequence and amino acid polymorphisms.
- Recognize the different types of DNA sequence polymorphisms:
  - STR, SNP, CNV
- Know the different classes of DNA mutation:
  - Point mutations (silent, missense, nonsense, frameshift, splicing, regulatory) insertion/deletions, rearrangements
- Understand how to distinguish a disease-causing mutation from a neutral DNA sequence variation

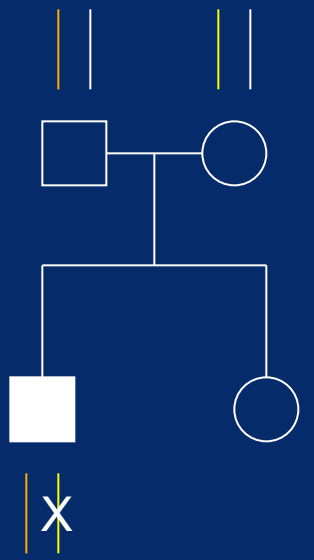
Tests to Detect Mutations

- Many methods/technologies
- Rapidly changing

- DNA sequencing
  - Most direct and informative
  - The gold standard
  - Targeted region (known mutation)
  - “Whole” gene (unknown mutation)
  - … Whole exome / whole genome

How do we distinguish a disease causing mutation from a silent sequence variation?

- Obvious disruption of gene
  - large deletion or rearrangement
  - frameshift
  - nonsense mutation
- Functional analysis of gene product
  - expression of recombinant protein
  - transgenic mice
- New mutation by phenotype and genotype



- Computer predictions
- Disease-specific mutation databases
  - Same/similar mutation in other patients, not in controls
- Rare disease-causing mutation vs. private "polymorphism" (rare variant)

Learning Objectives

- Understand the meaning of DNA sequence and amino acid polymorphisms.
- Recognize the different types of DNA sequence polymorphisms:
  - STR, SNP, CNV
- Know the different classes of DNA mutation:
  - Point mutations (silent, missense, nonsense, frameshift, splicing, regulatory) insertion/deletions, rearrangements
- Understand how to distinguish a disease-causing mutation from a neutral DNA sequence variation

Additional Source Information

for more information see: http://open.umich.edu/wiki/CitationPolicy

Slide 10: CC: BY-SA: Ricardipus (wikipedia) http://creativecommons.org/licenses/by-sa/2.0/deed.en