Genetics and Genomics in Clinical Research

An Immersion Course for Clinical Investigators at UAB

Introduction and Overview

Bruce R. Korf, MD, PhD

Goals

- Describe approaches to study of the genomic contributions to rare and common disorders

- Recognize ethical, legal, and social issues involved in design of a genomics research protocol

- Describe major bioinformatic approaches used in the analysis of genomic data

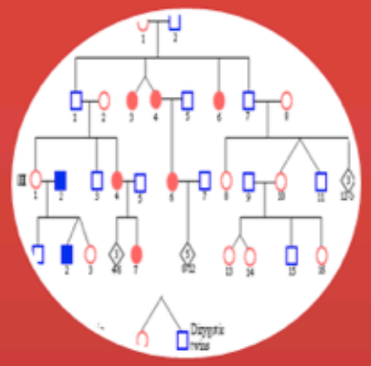
- Recognize opportunities to explore genomic aspects of medical problems and the resources available at UAB

Schedule

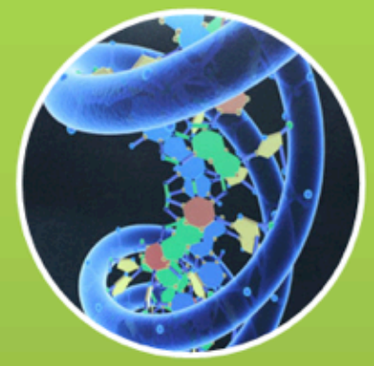
	Monday, Oct 12	Tuesday, Oct 13*	Wednesday, Oct 14	Thursday, Oct 15	Friday, Oct 16
8:30-9:00	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast
9:00-10:00	Review of Genetic Principles Dr. Bruce Korf	Copy number variants (CNVs) and chromosomal microarray technologies Dr. Fady Mikhail	Case Studies and Informed Consent Dr. Ashley Cannon	Approaches to Bioinformatics Data Analysis Dr. David Crossman	Genome Wide Linkage Analysis Dr. Hemant Tiwari
10:15-11:15	Overview of Genetics and Genomics in Clinical Research Dr. Bruce Korf	Genotyping Technologies Dr. Mike Crowley	Analytics of next-generation sequencing data Dr. Greg Cooper	Transcriptome and Pathway Analysis Dr. David Crossman	Genome Wide Association Analysis Dr. Hemant Tiwari
11:30-12:30	Human Population Genetics Dr. Greg Barsh	Next-Generation sequencing Dr. Mike Crowley	Genomic counseling Ms. Kelly East	Variant Analysis of Exome/ Genome Sequencing Data Dr. David Crossman	Genetics in Medicine Dr. Bruce Korf

Genetics

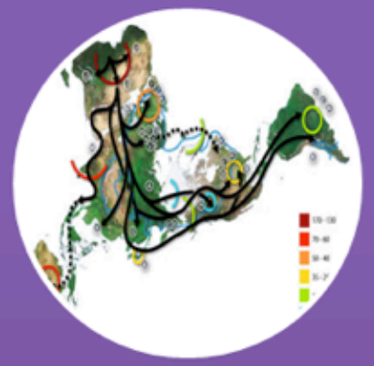
Scientific discipline that deals with the variability and transmission of biological traits.



Transmission through families



Flow of information in cell



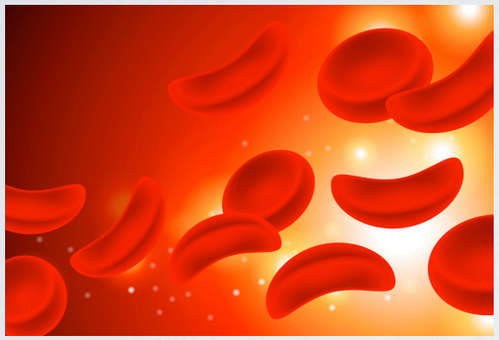
Population forces and evolution

Genomics

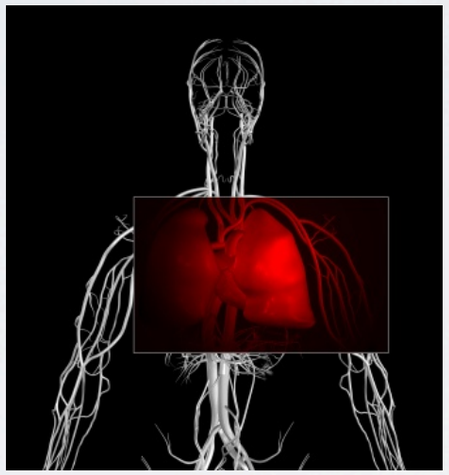
"For the newly developing discipline of mapping/sequencing (including analysis of the information) we have adopted the term GENOMICS. We are indebted to T. H. Roderick of the Jackson Laboratory, Bar Harbor, Maine, for suggesting the term. The new discipline is born from a marriage of molecular and cell biology with classical genetics and is fostered by computational science."

(Victor A. McKusick and Frank H. Ruddle. A new discipline, a new name, a new journal [editorial]. Genomics 1987 Sep; 1:1-2.)

Human “Phenome”

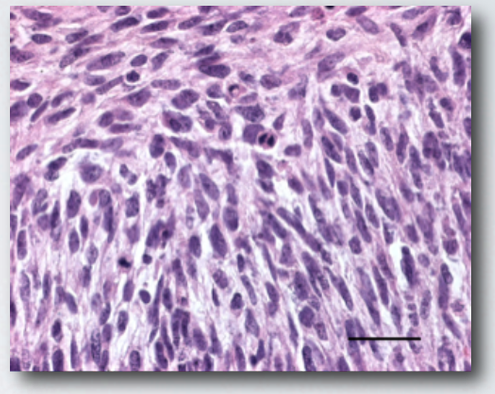


Single Gene



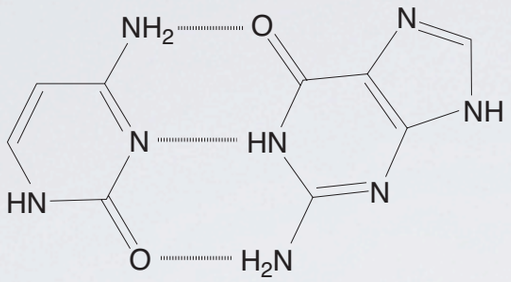
Multifactorial

Pharmacogenetic

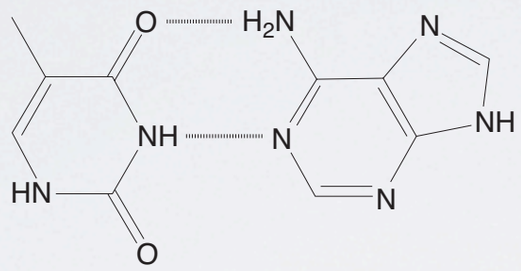


Cancer

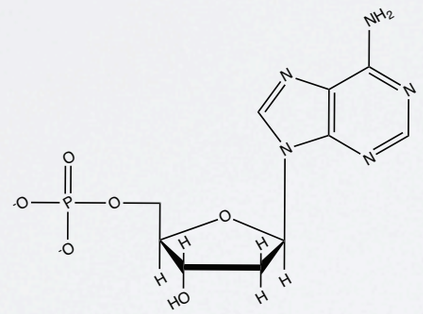
DNA and Genetic Information



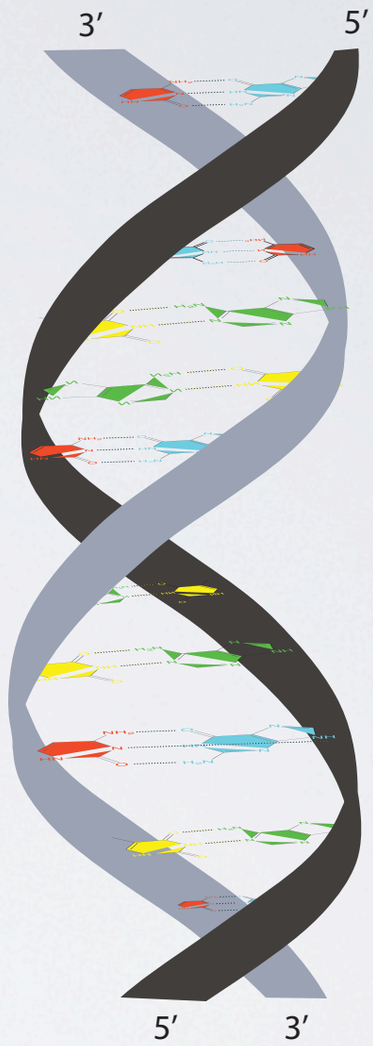
Cytosine Guanine



Thymine Adenine

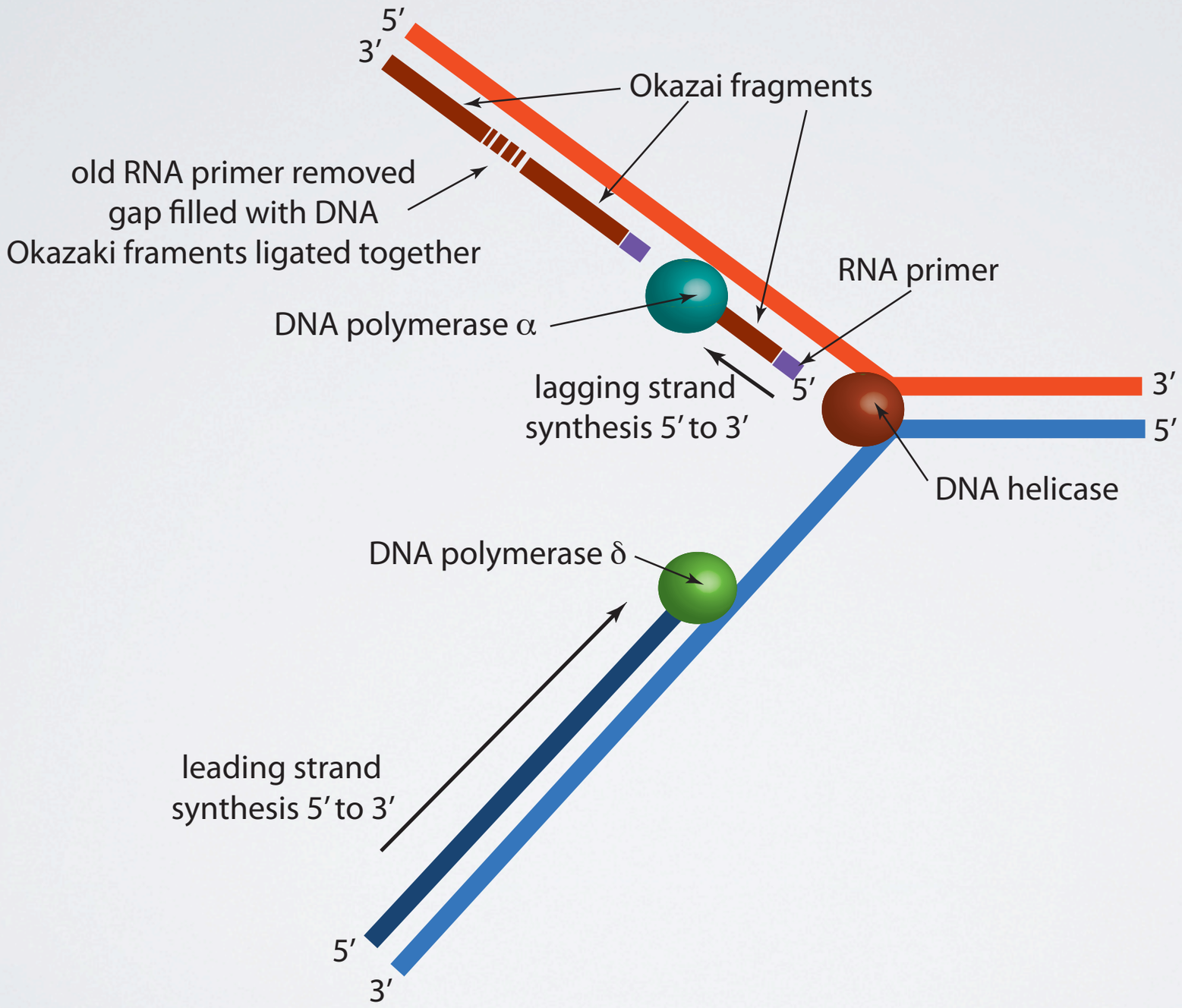


deoxyadenosine monophosphate

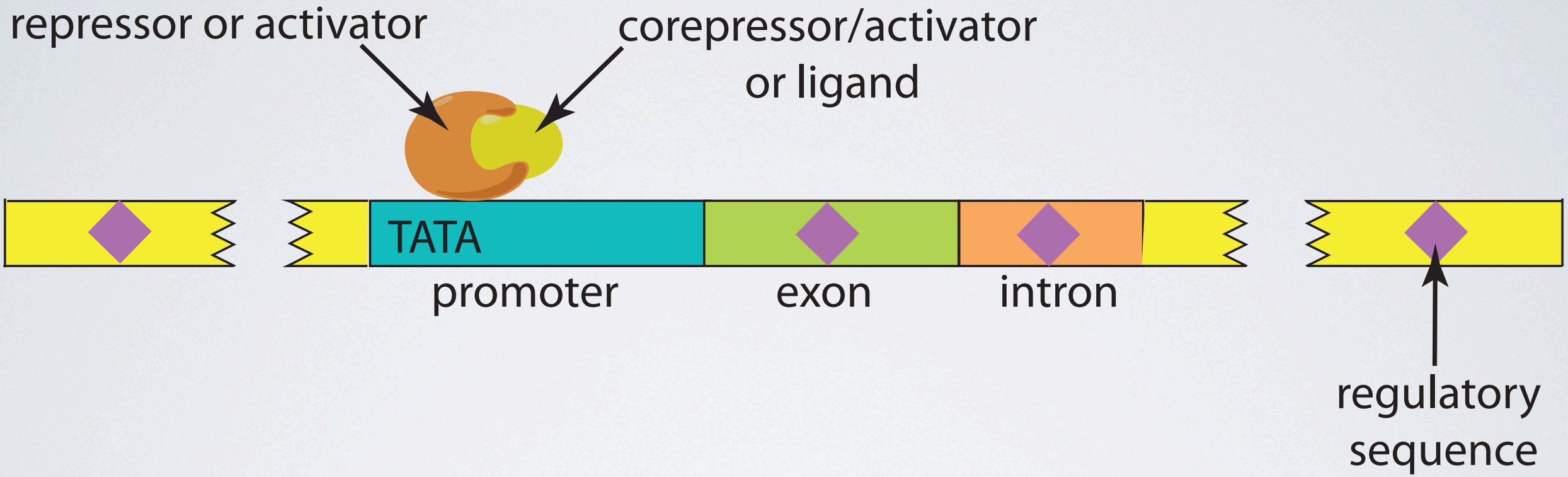


DNA → RNA → Protein

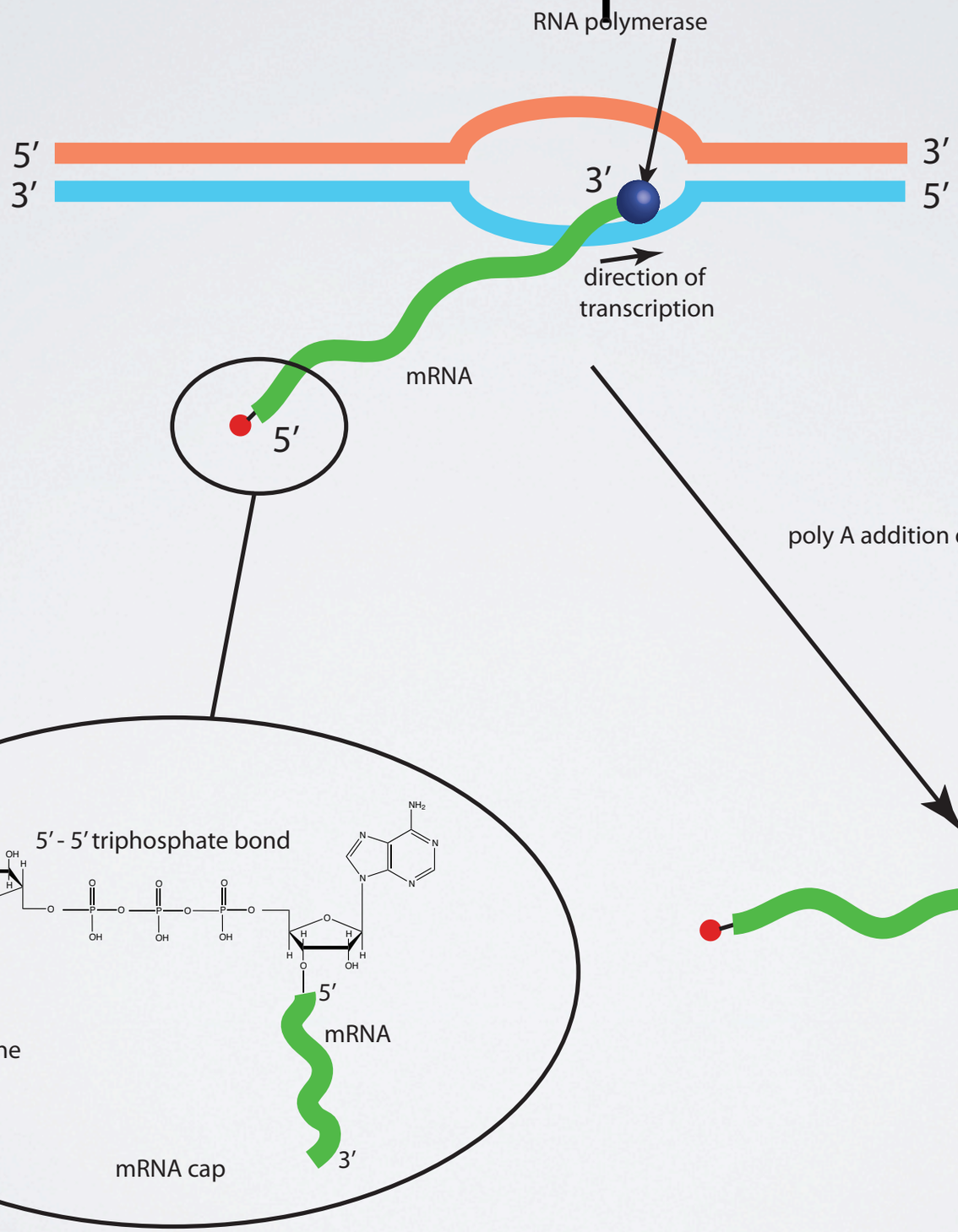
DNA Replication



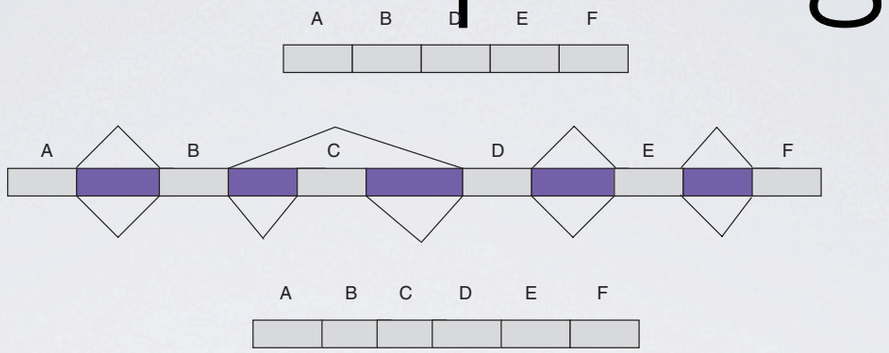
Gene Regulation

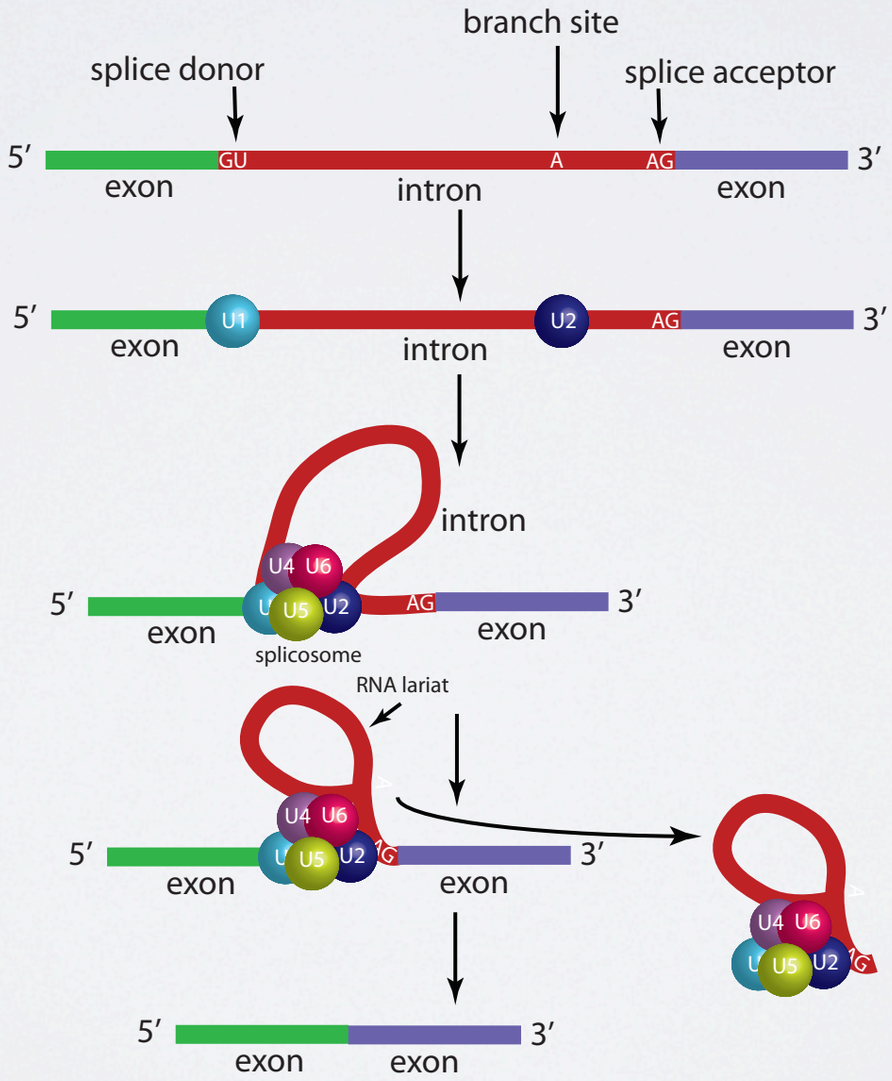


Transcription

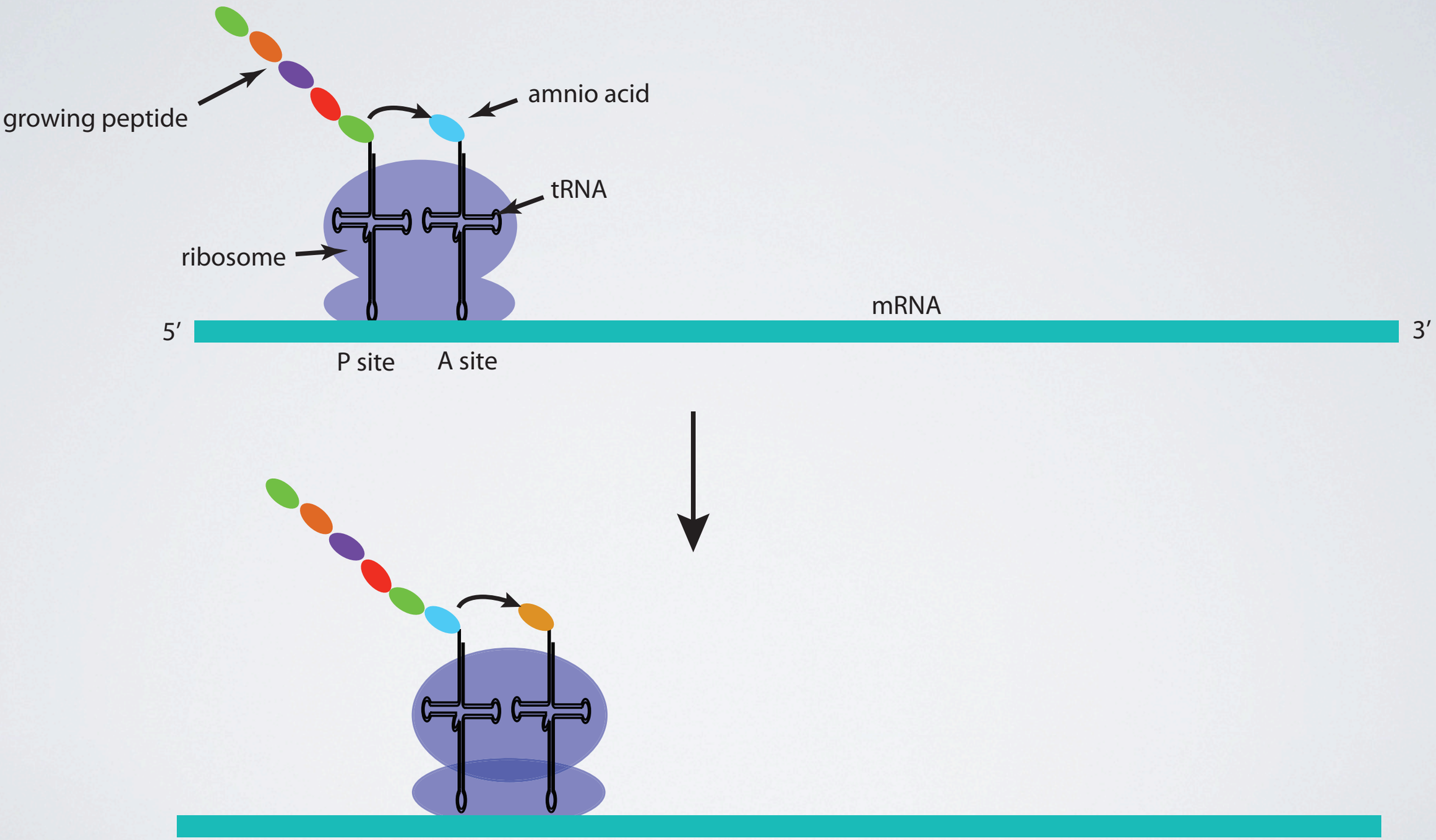


RNA Splicing

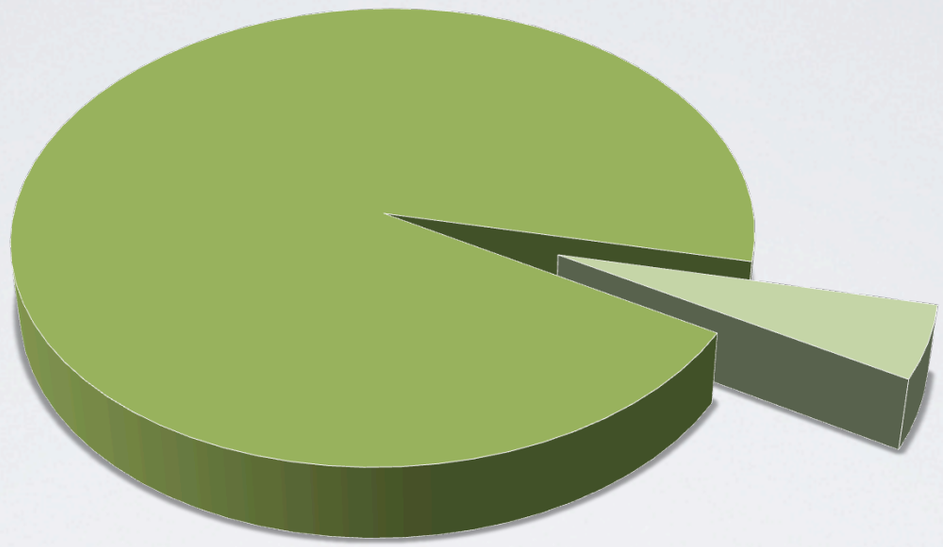




Translation



The Human Genome



Repeated and non-coding sequences

- Interspersed repeats
- Pseudogenes
- Simple sequence repeats
- Segmental duplications
- Blocks of repeated sequences

coding sequences

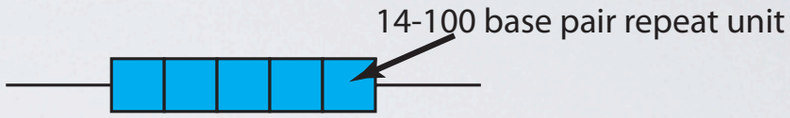
- non-coding RNA's
- tRNA
- rRNA
- snRNA
- protein-encoding RNA's

Repeated Sequences

simple sequence repeat

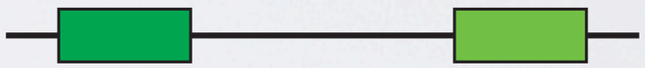
...GCGACACACACACACACAGT...

variable number tandem repeat



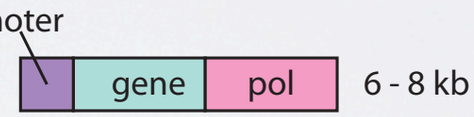
highly repeated sequences at
centromeric and subtelomeric regions

segmental duplications

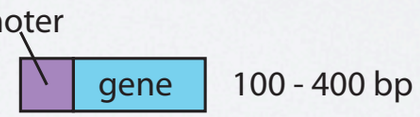


transposon-derived repeats

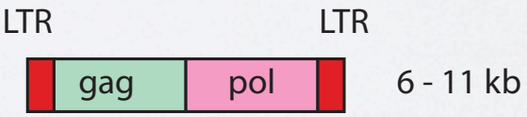
LINE



SINE

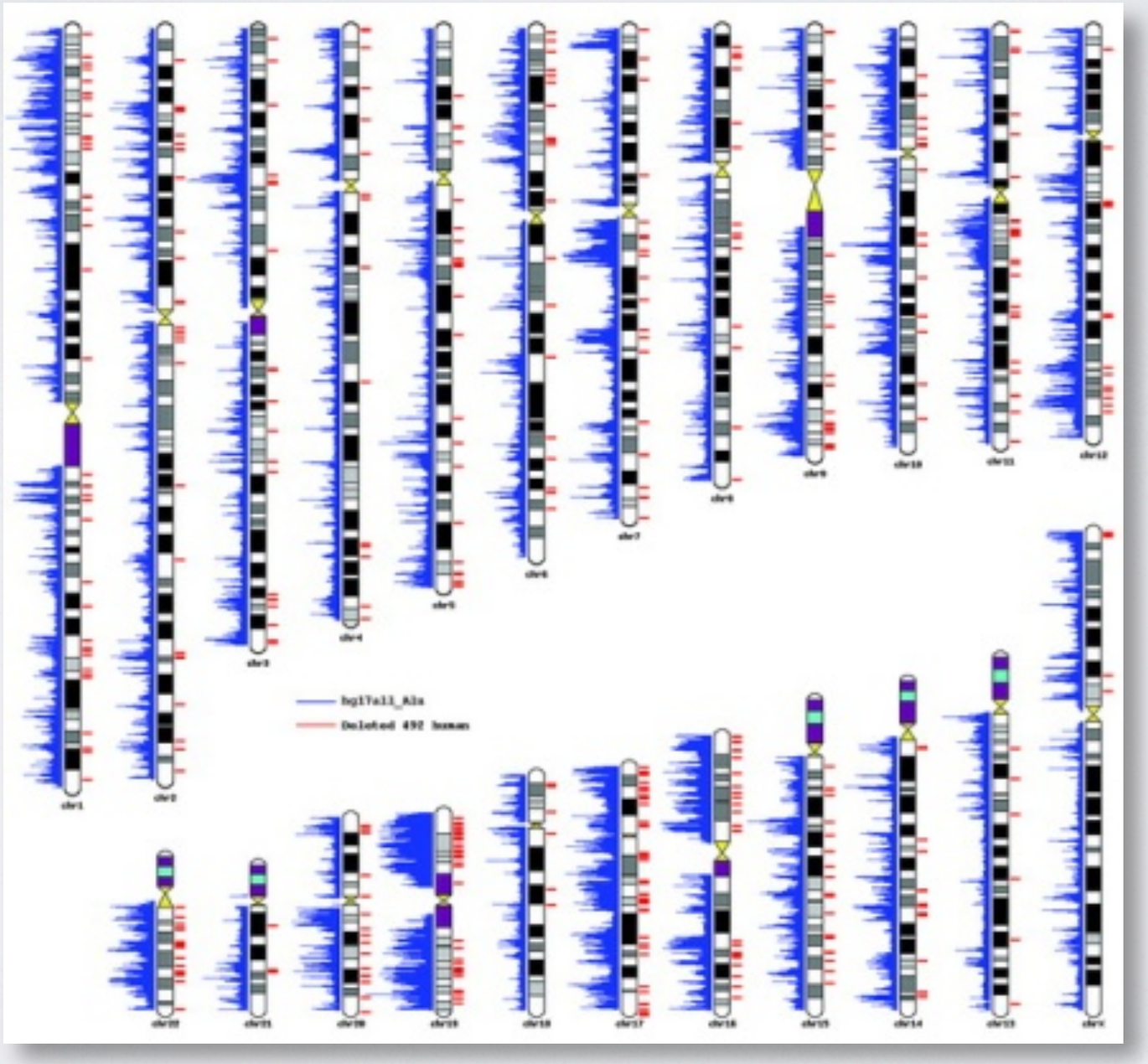


retroviral-like element



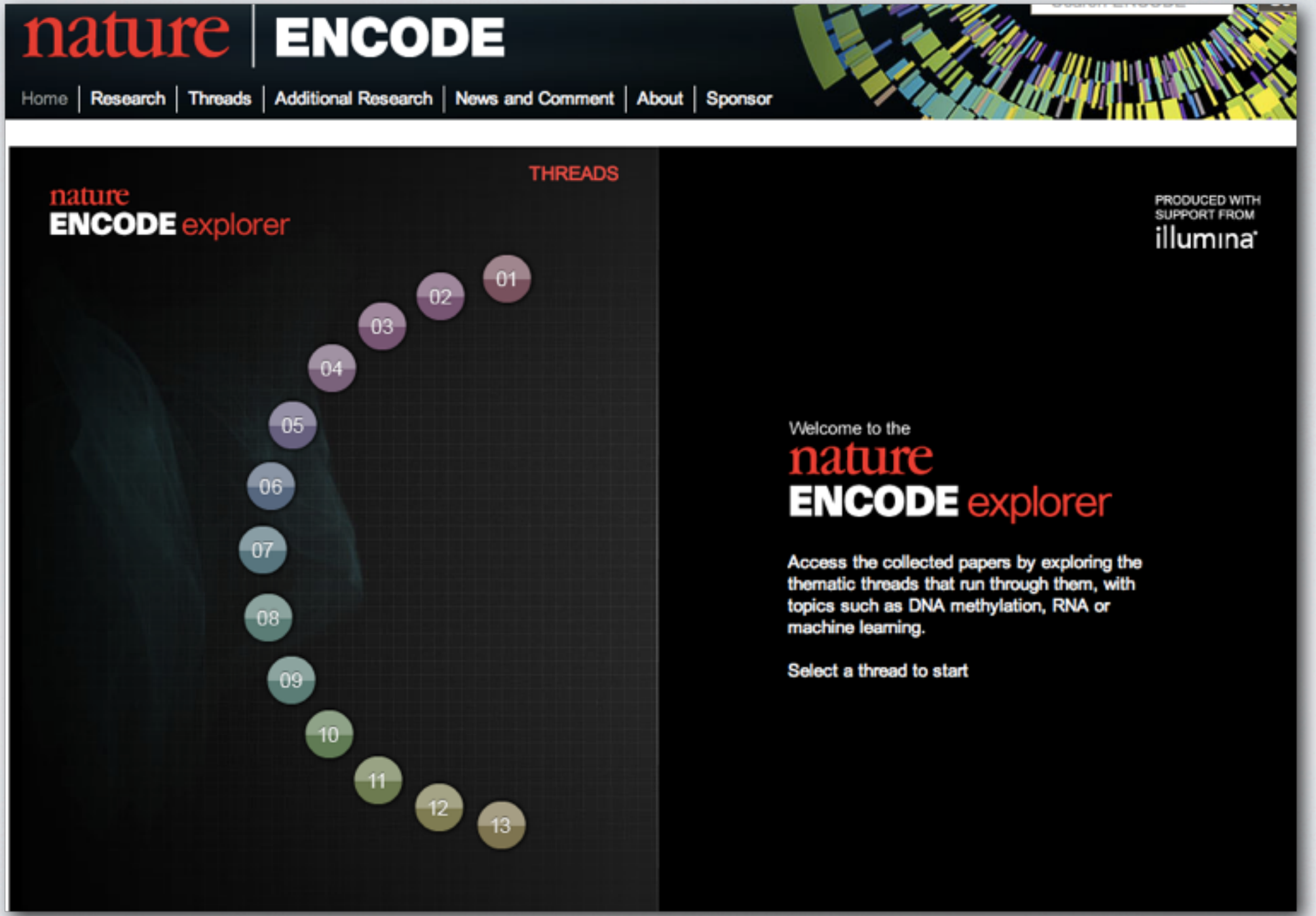
transposon

Alu Sequences



Sen SK, et al. Am. J. Hum. Genet., 79:41-53, 2006

ENCODE Project



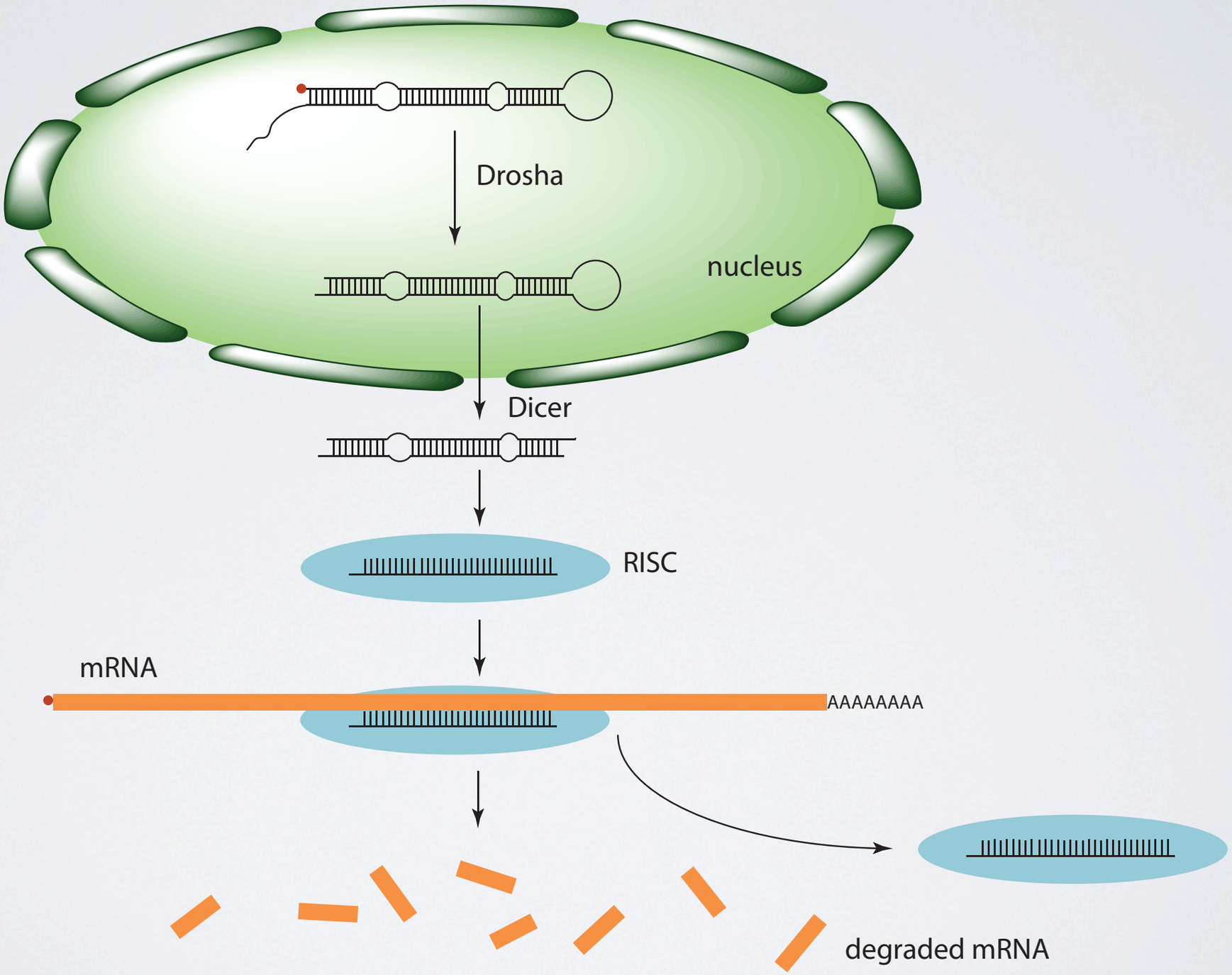
ENCODE Findings

- annotated 20,687 protein-encoding genes
- average 6.3 alternatively spliced isoforms per gene
- 8,801 small RNAs; 9,640 long non-coding transcripts
- >80% genome transcribed in some cell type
- >400,000 enhancers and 70,000 promoters

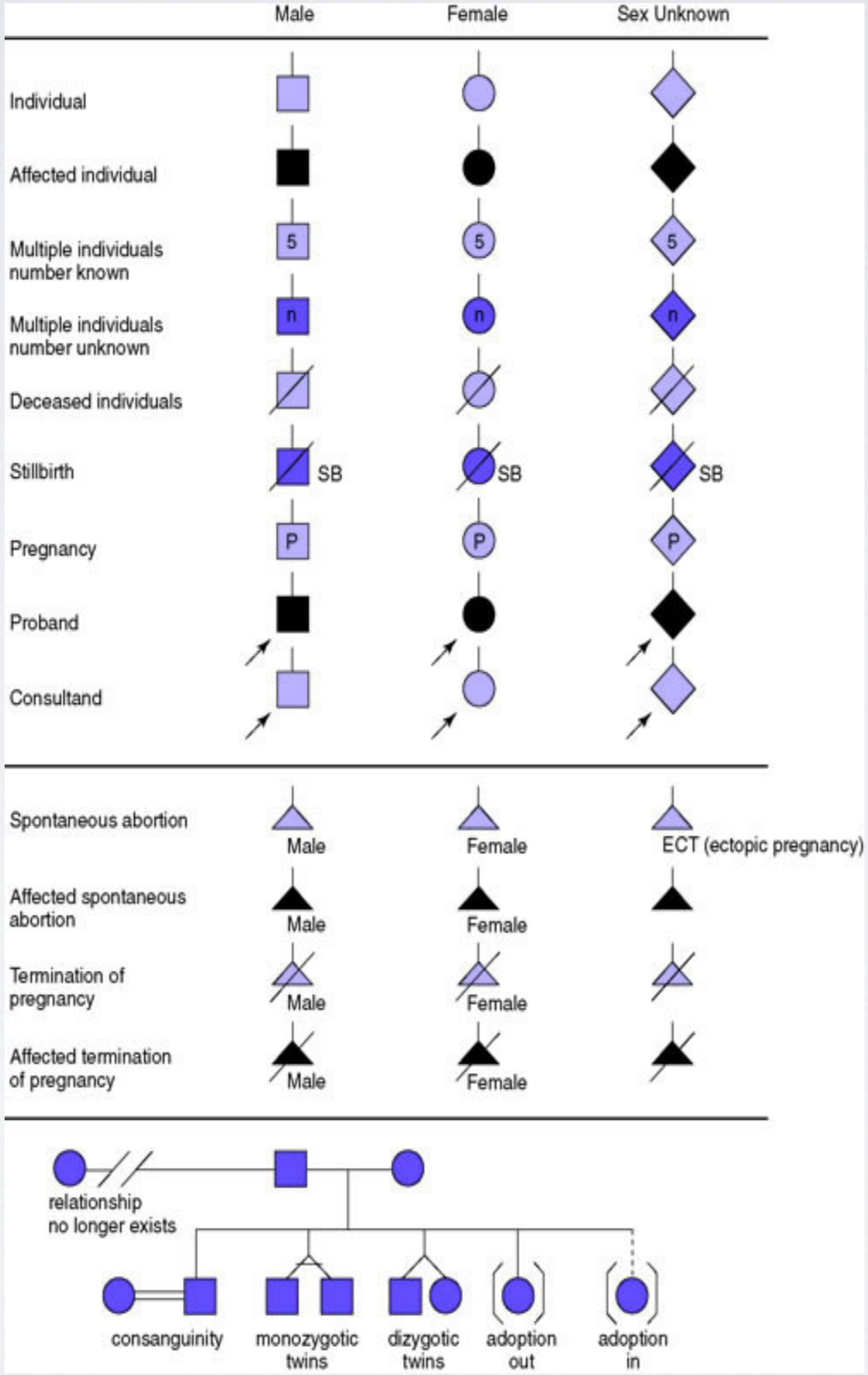
Non-coding RNAs

tRNA	transfer RNA	protein synthesis
rRNA	ribosomal RNA	protein synthesis
snRNA	small nuclear RNA	splicing
snoRNA	small nucleolar RNA	RNA modification
miRNA	micro RNA	gene regulation
siRNA	small interfering RNA	viral defense
lncRNA	long non-coding RNA	gene regulation/unknown

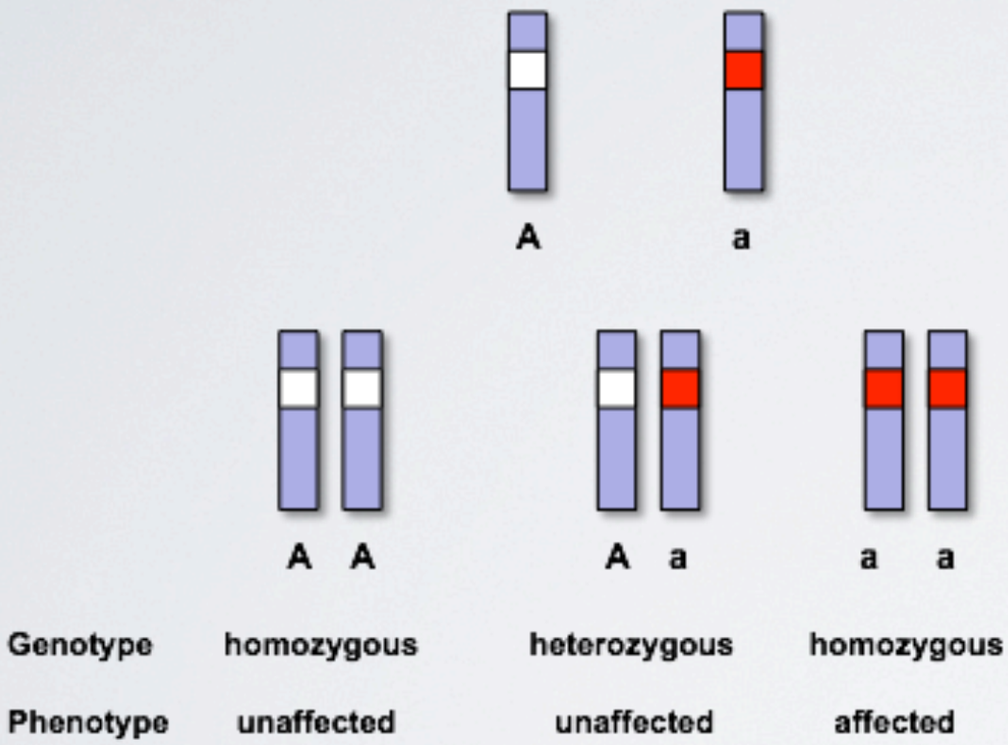
Micro RNA

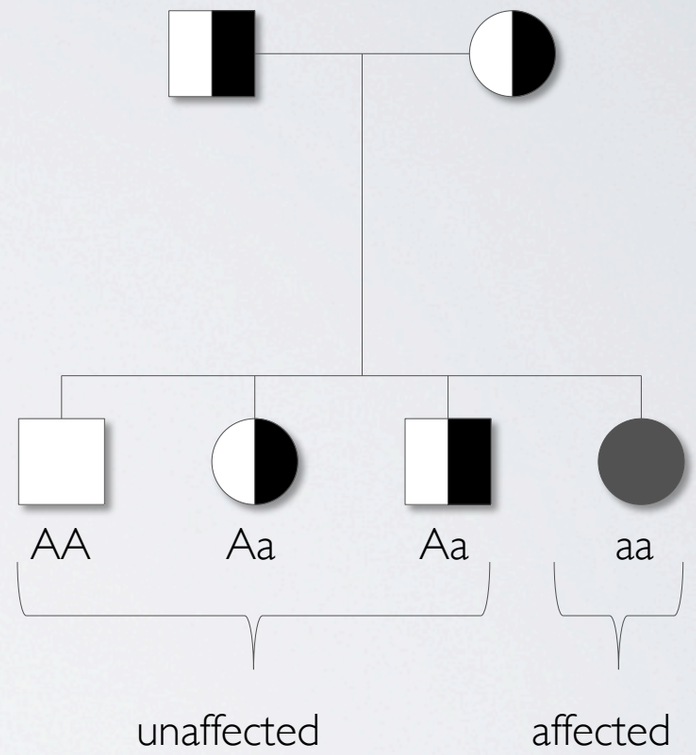


Pedigree Symbols

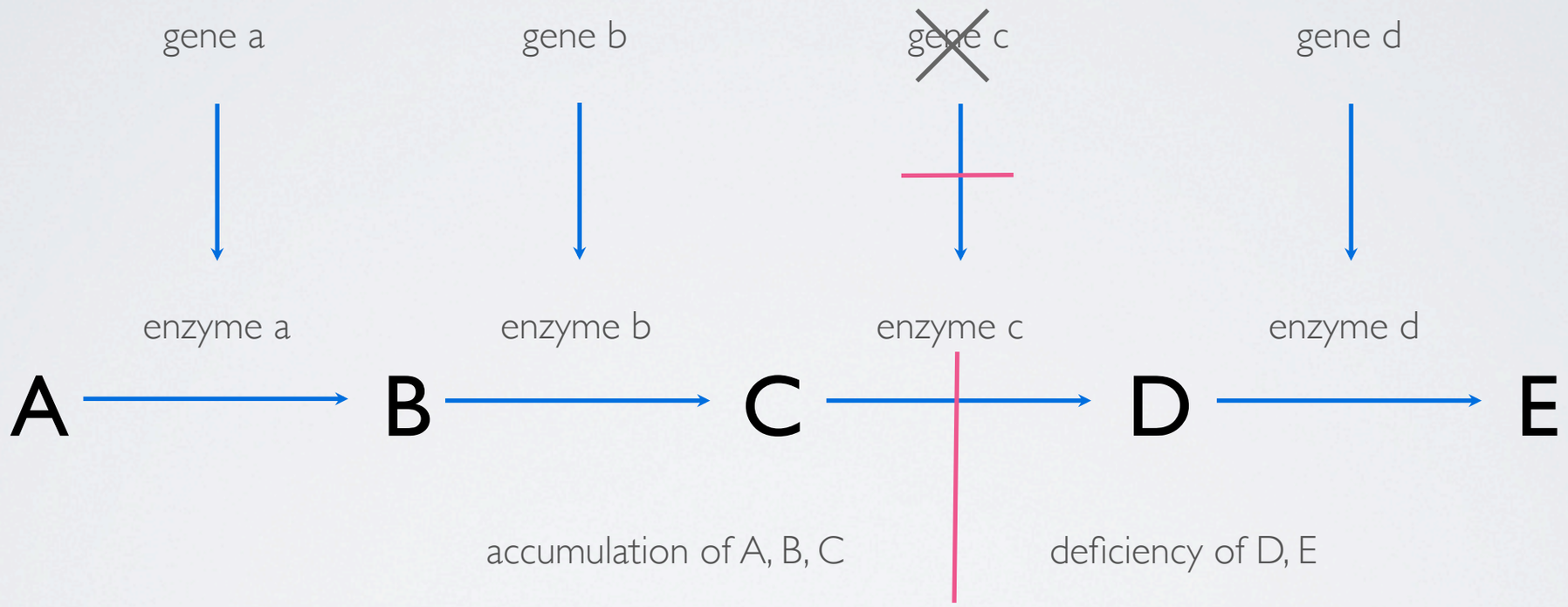


Autosomal Recessive





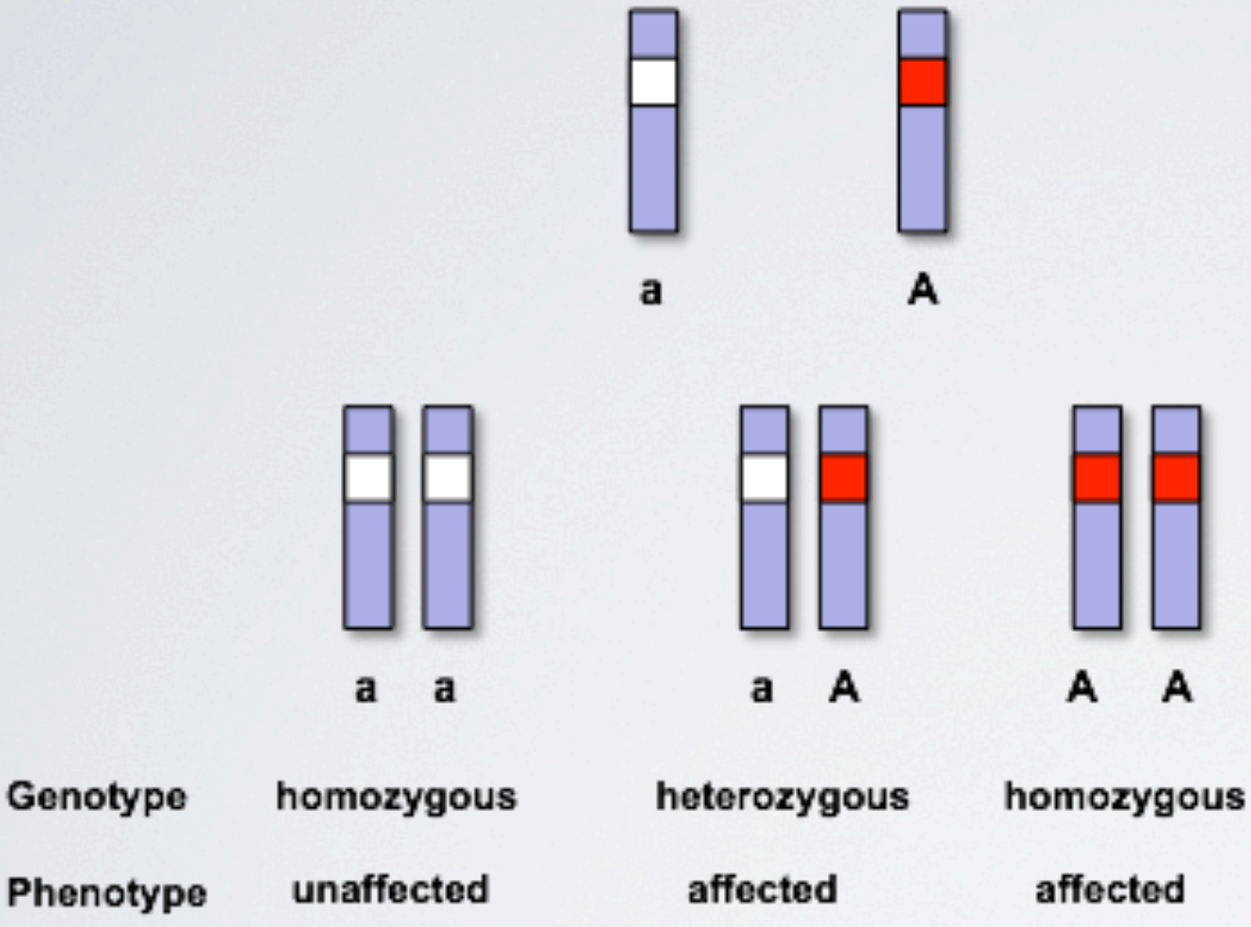
Metabolic Pathways

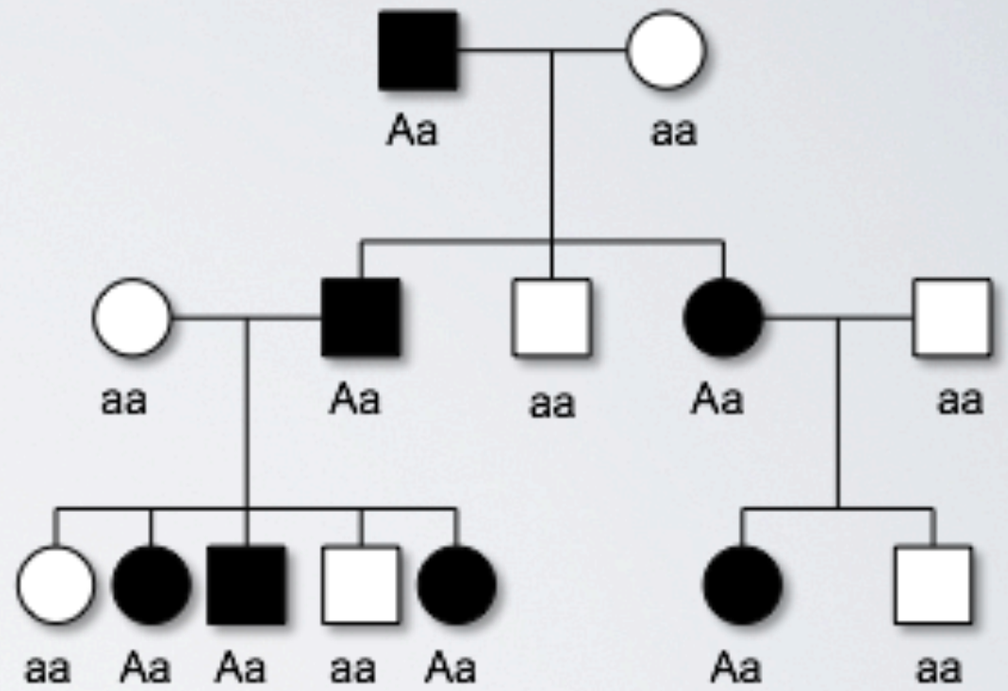


Recessive Mechanisms

Loss of function mutations	Excess capacity in system	Require biallelic mutation
• Deletion • Stop • Frameshift • Missense • Structural	• Enzyme • Membrane receptor • Ion channel	• Abnormal protein does not “poison” system

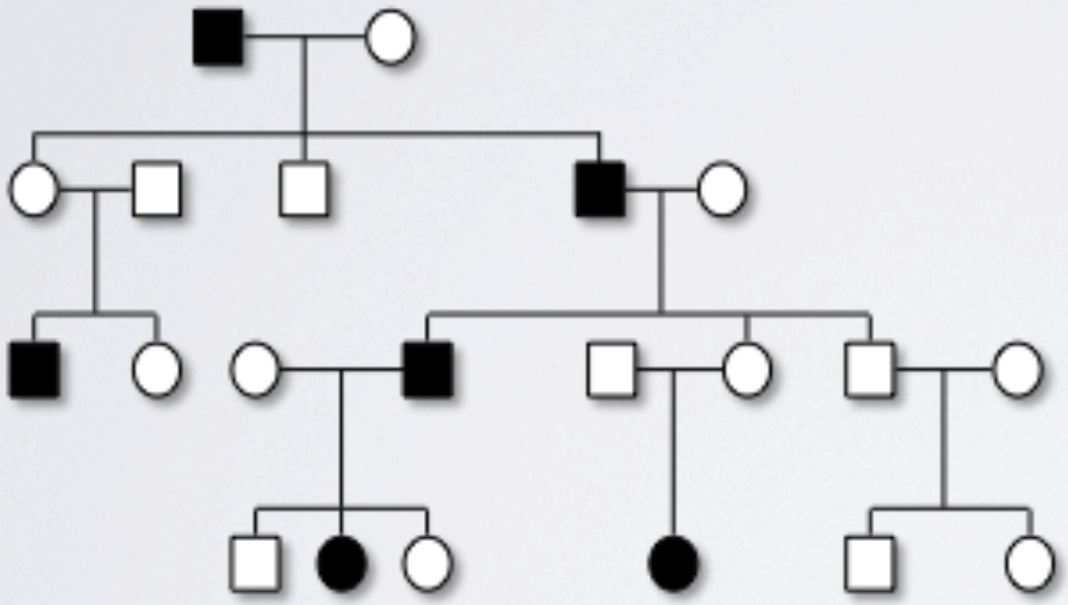
Autosomal Dominant

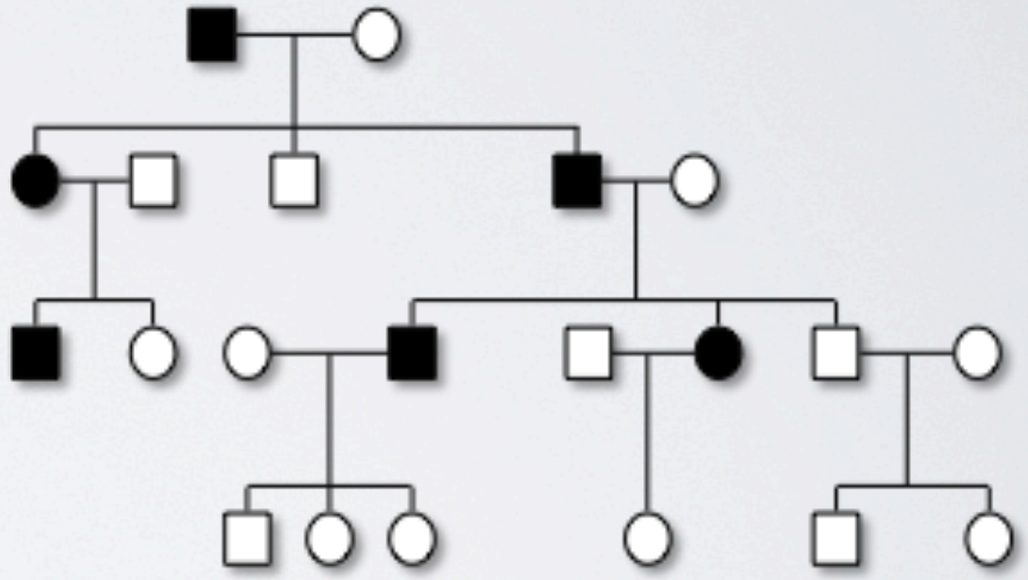




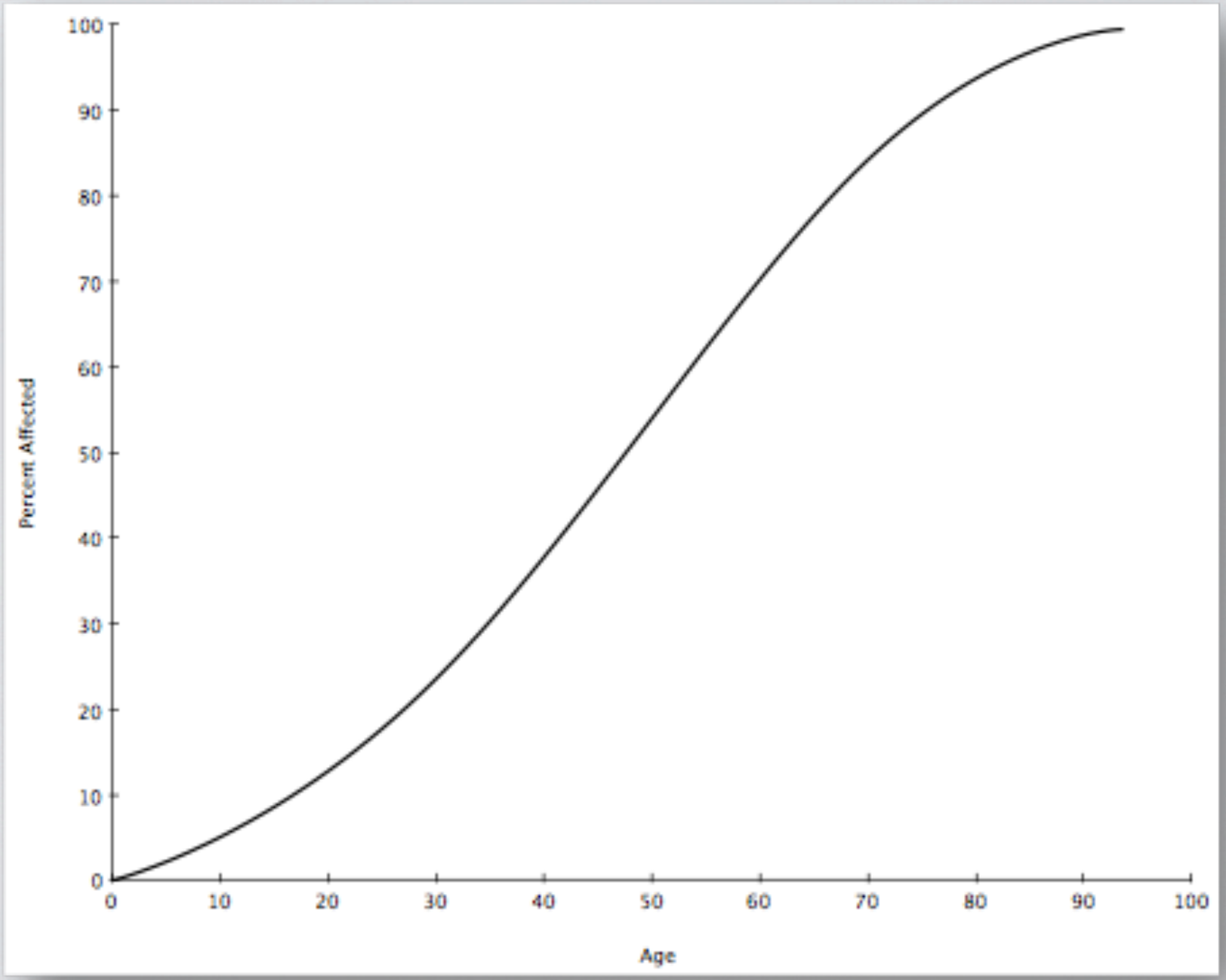
Penetrance

Fraction of individuals who carry a gene who manifest a specified phenotype



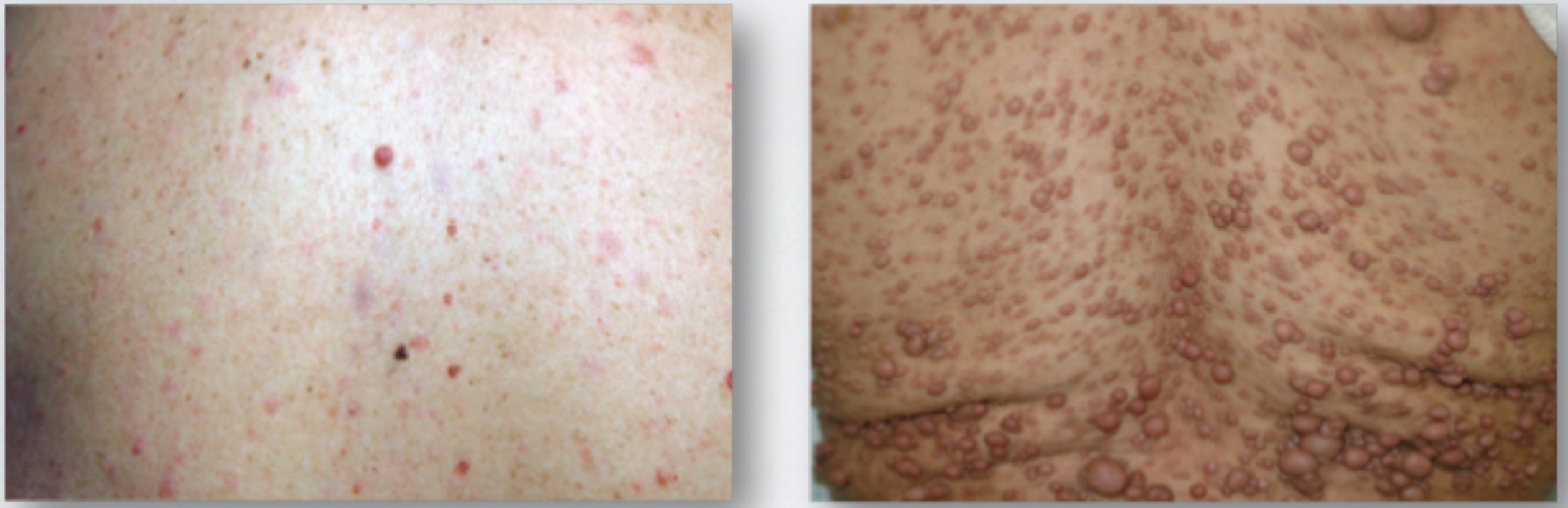


Age-Dependent Penetrance



Expressivity

- different modes or degrees of expression of trait in population



dermal neurofibromas in NF1

Dominant Mechanisms

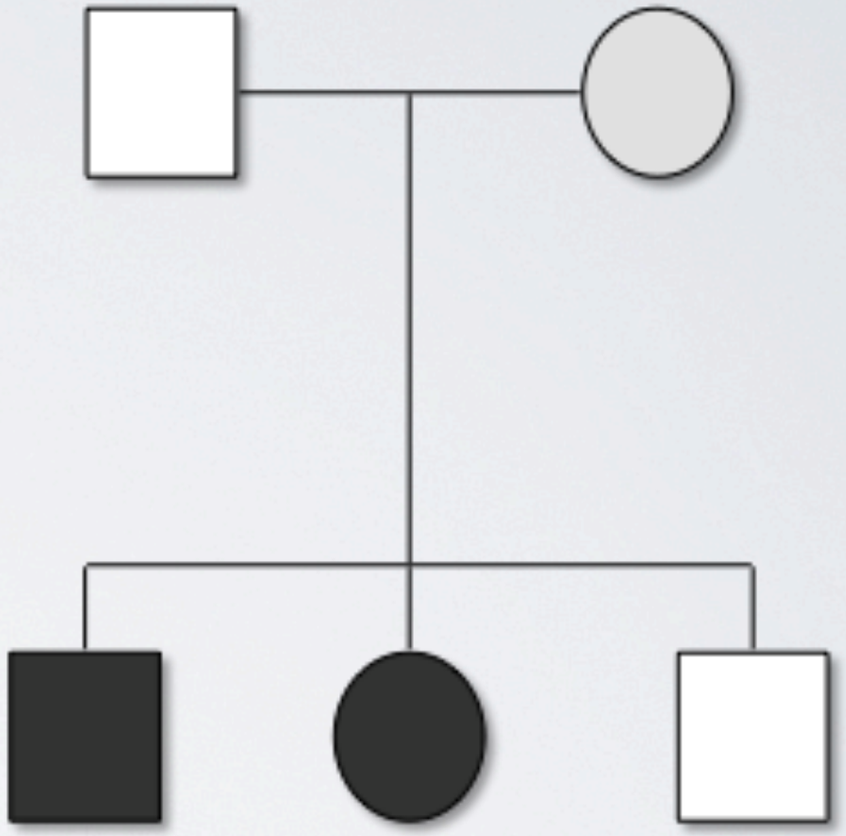
- **Deletion**
- **Stop**
- **Frameshift**
- **Missense**
- **Structural**

- **Haploinsufficiency**
- **Dominant negative**
- **Tumor suppressor**

- **Signaling pathway**

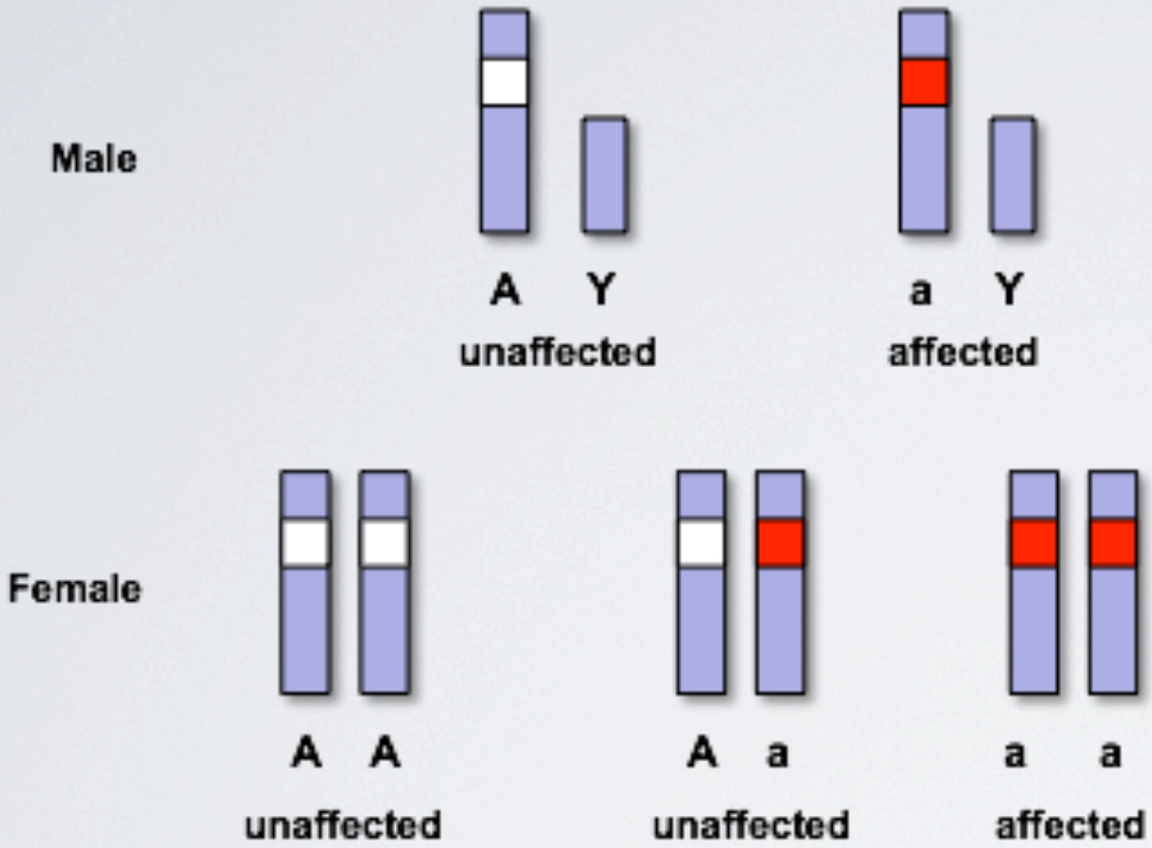
- **Missense**

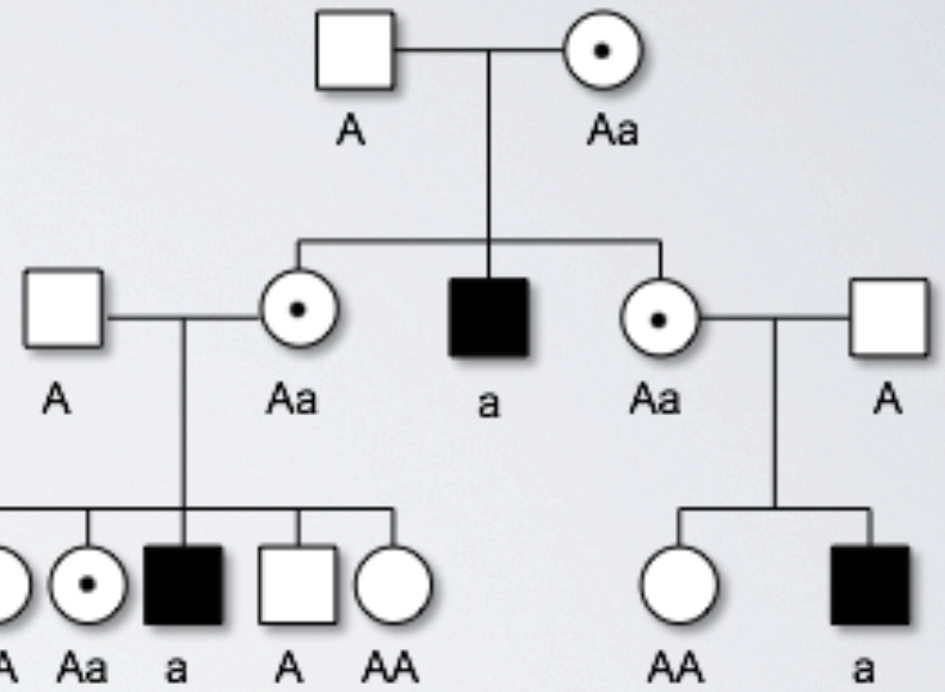
Mosaicism



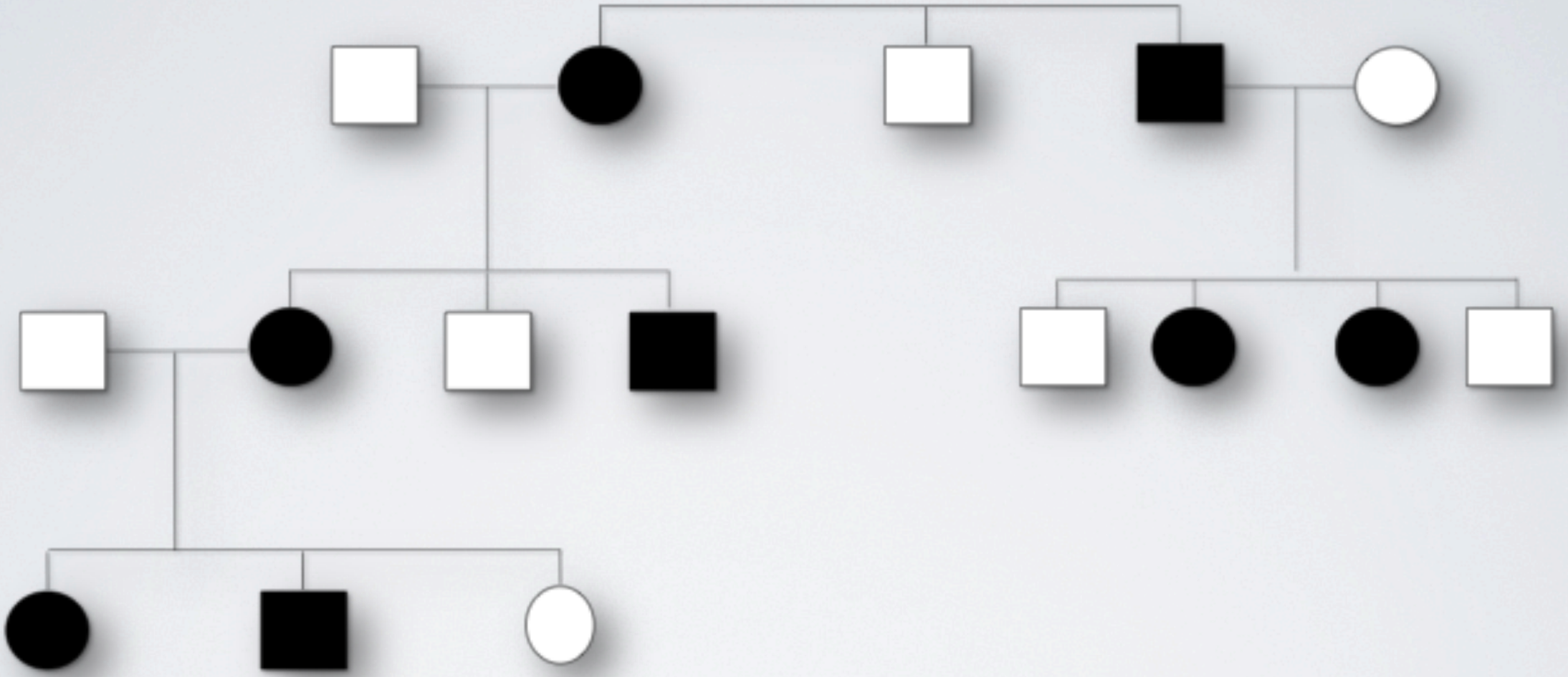
- Germ line
- Somatic

X-linked Recessive



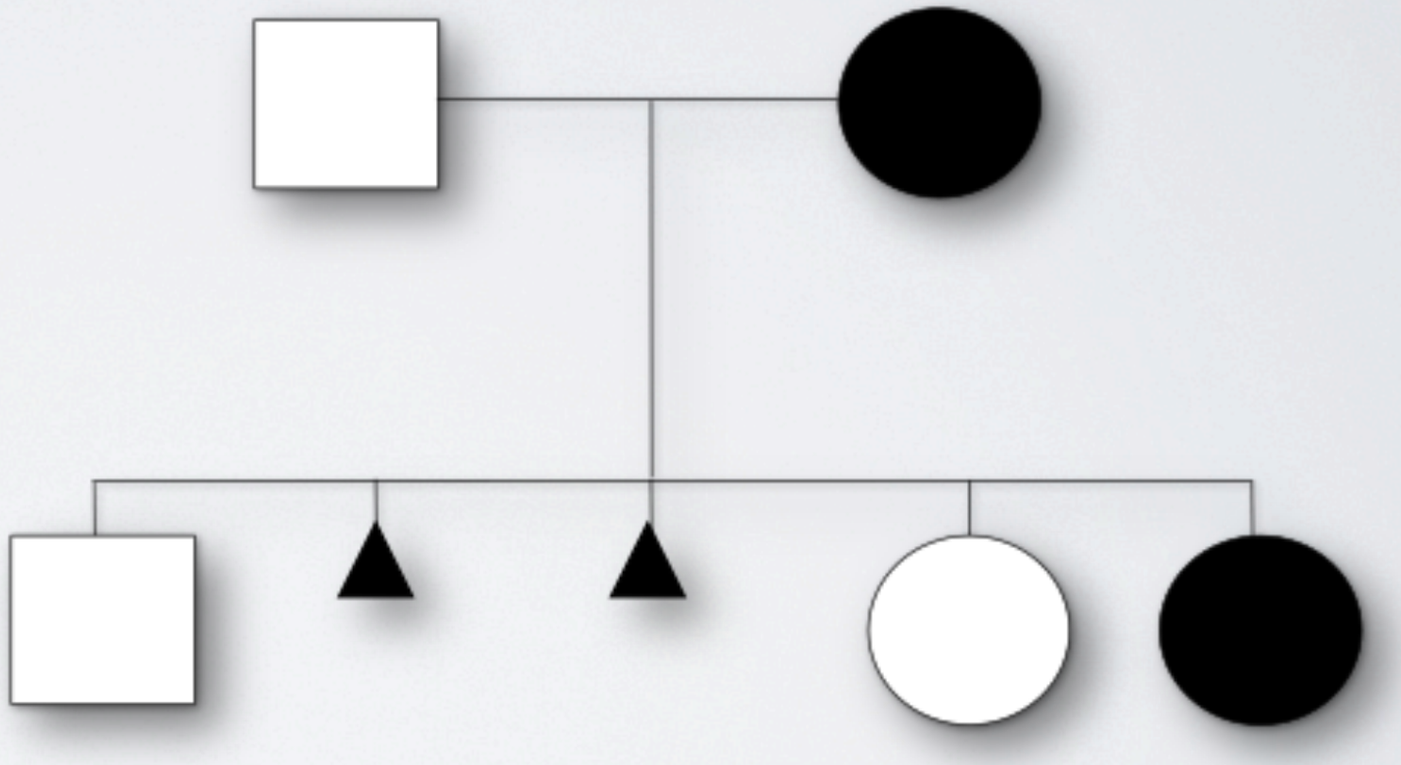


X-linked Dominant

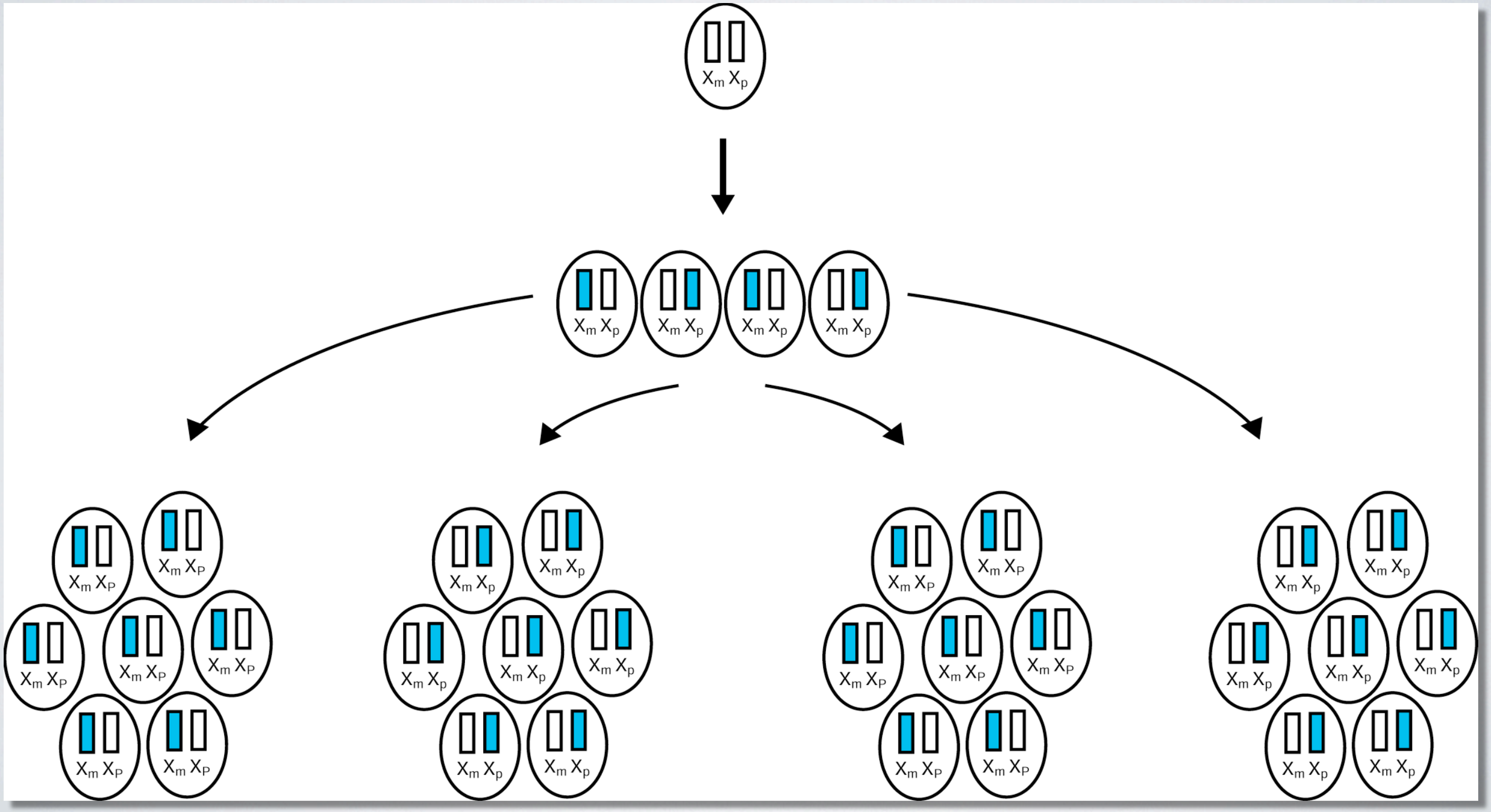


male transmits to all daughters, no sons

X-linked Dominant, Male Lethal

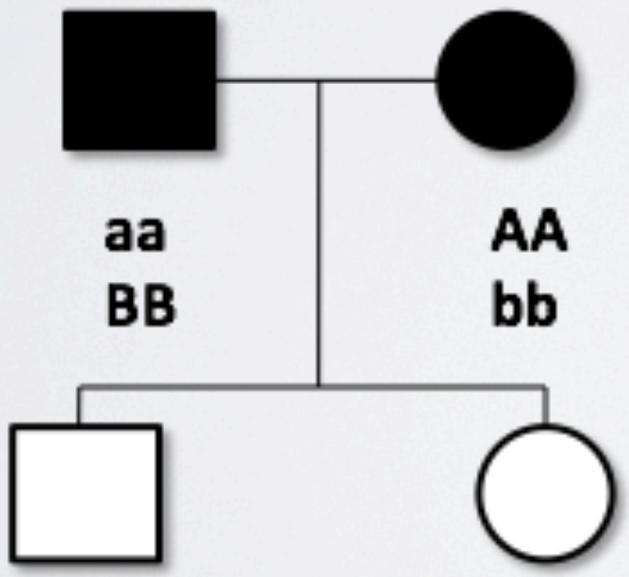


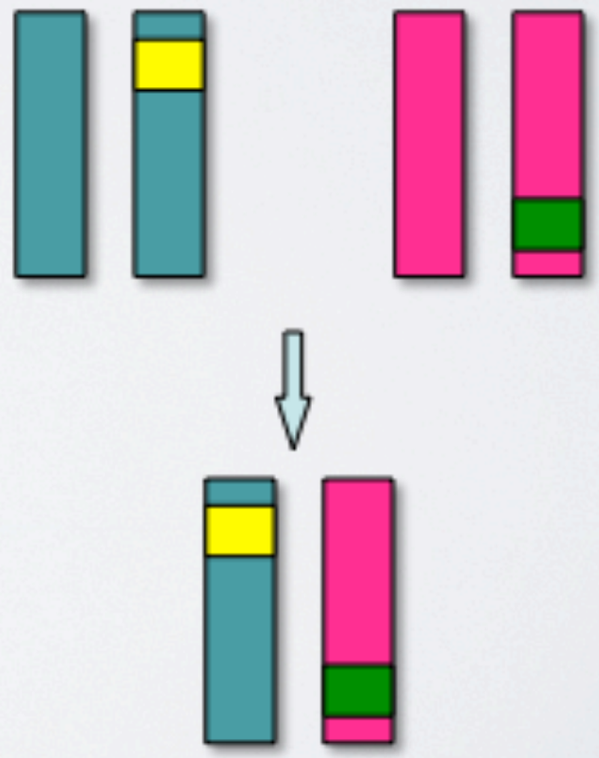
X-chromosome Inactivation



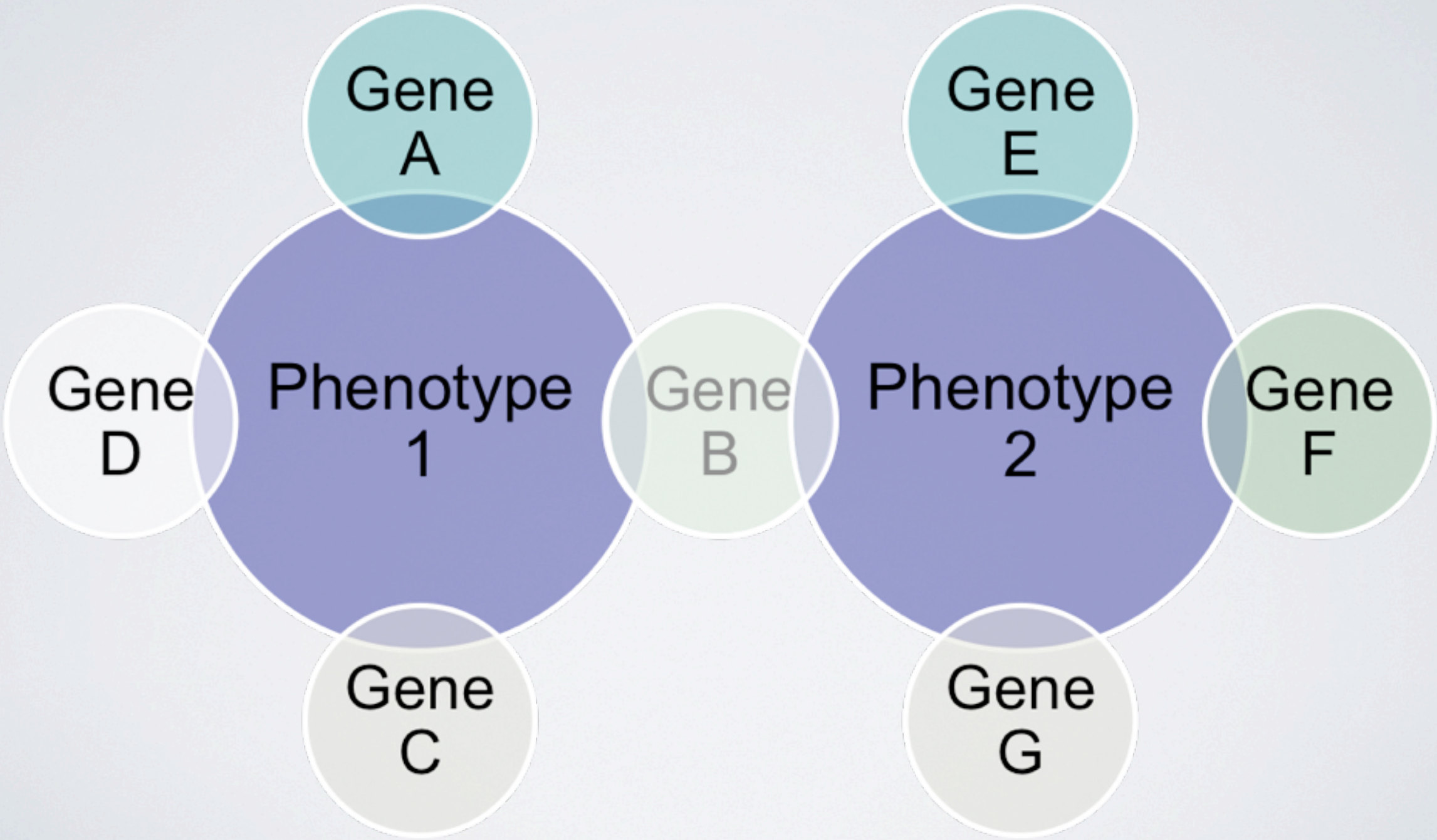
Genetic Heterogeneity

- Locus
  - Different loci responsible for similar disorder
- Allelic
  - Different types of mutations in same gene

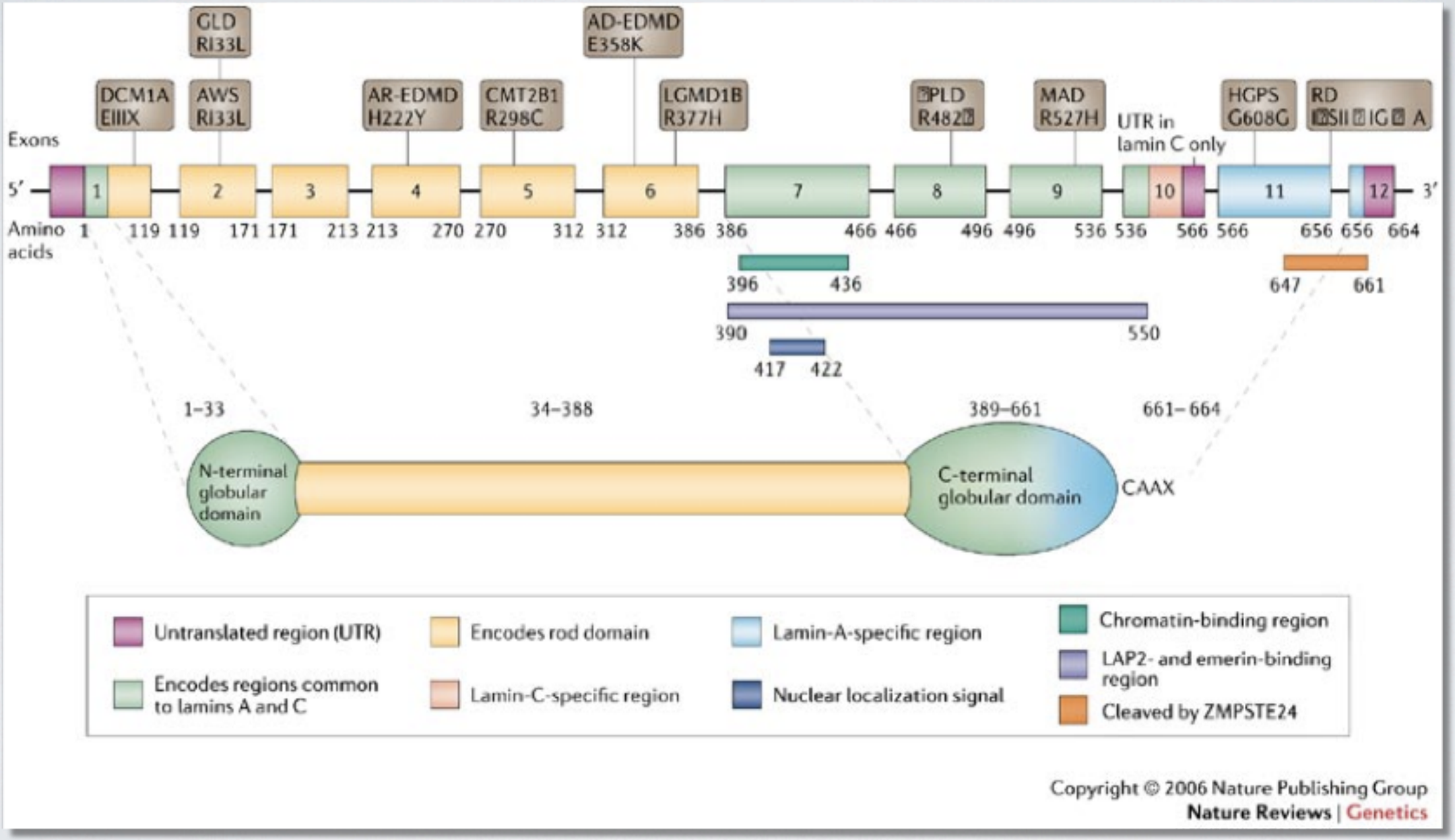




Genetic Heterogeneity

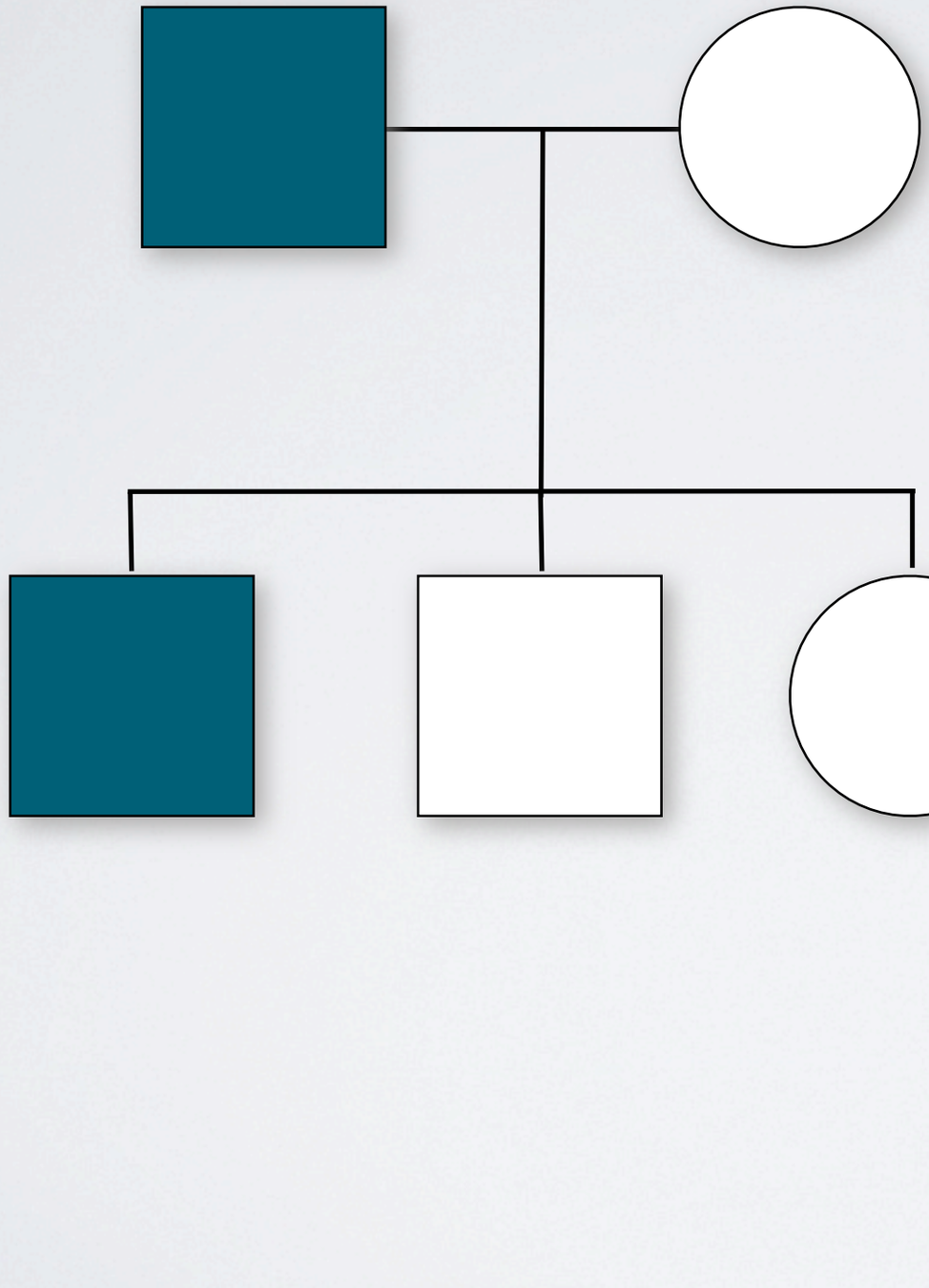


Allelic Heterogeneity



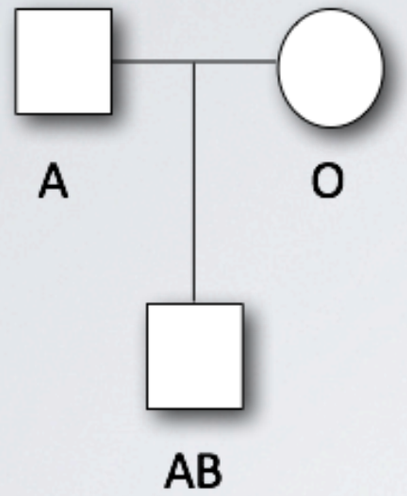
Brian C. Capell & Francis S. Collins
Nature Reviews Genetics **7**, 940-952 (December 2006)

Sex-Limited Expression

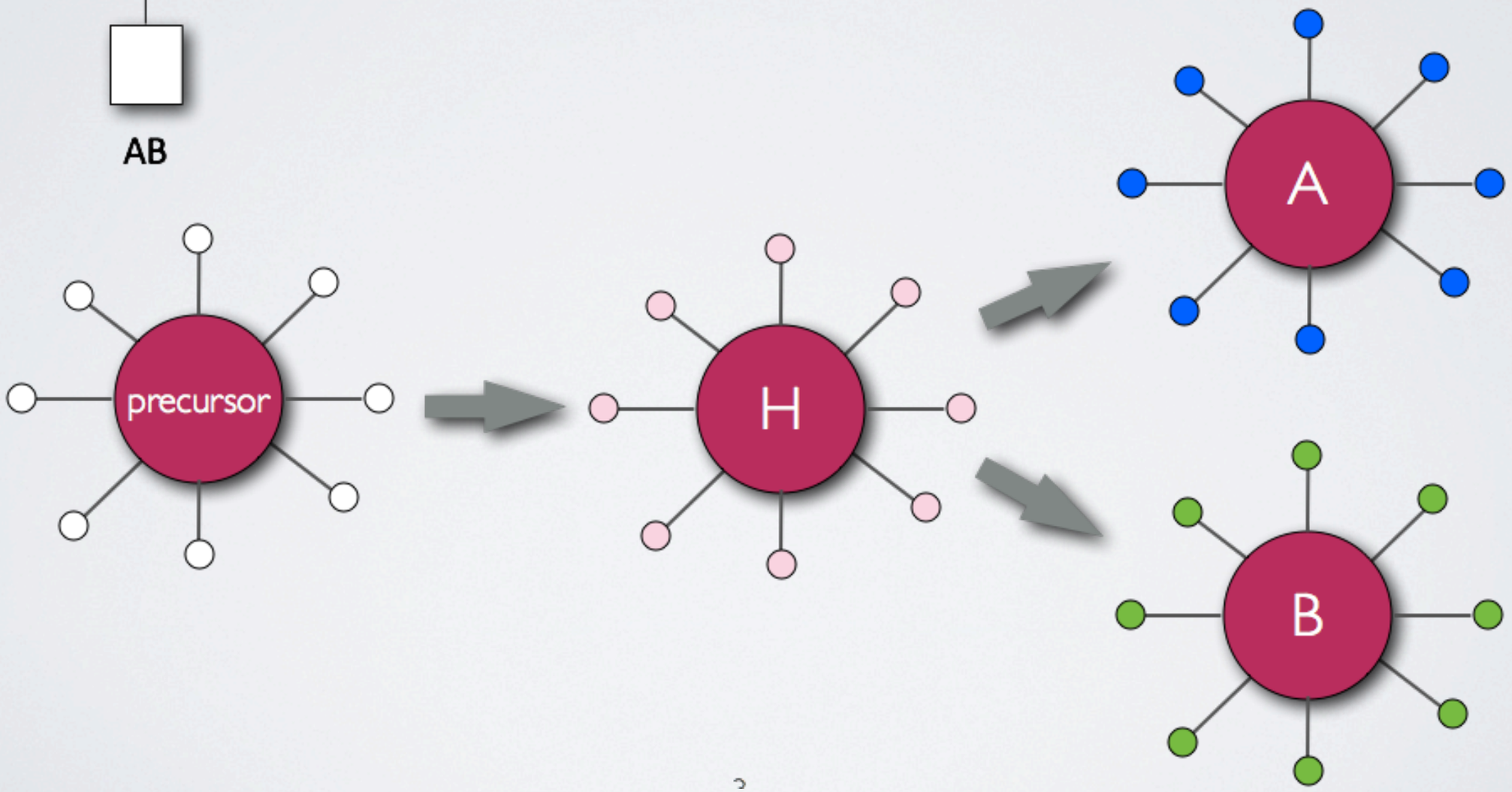


male pattern baldness

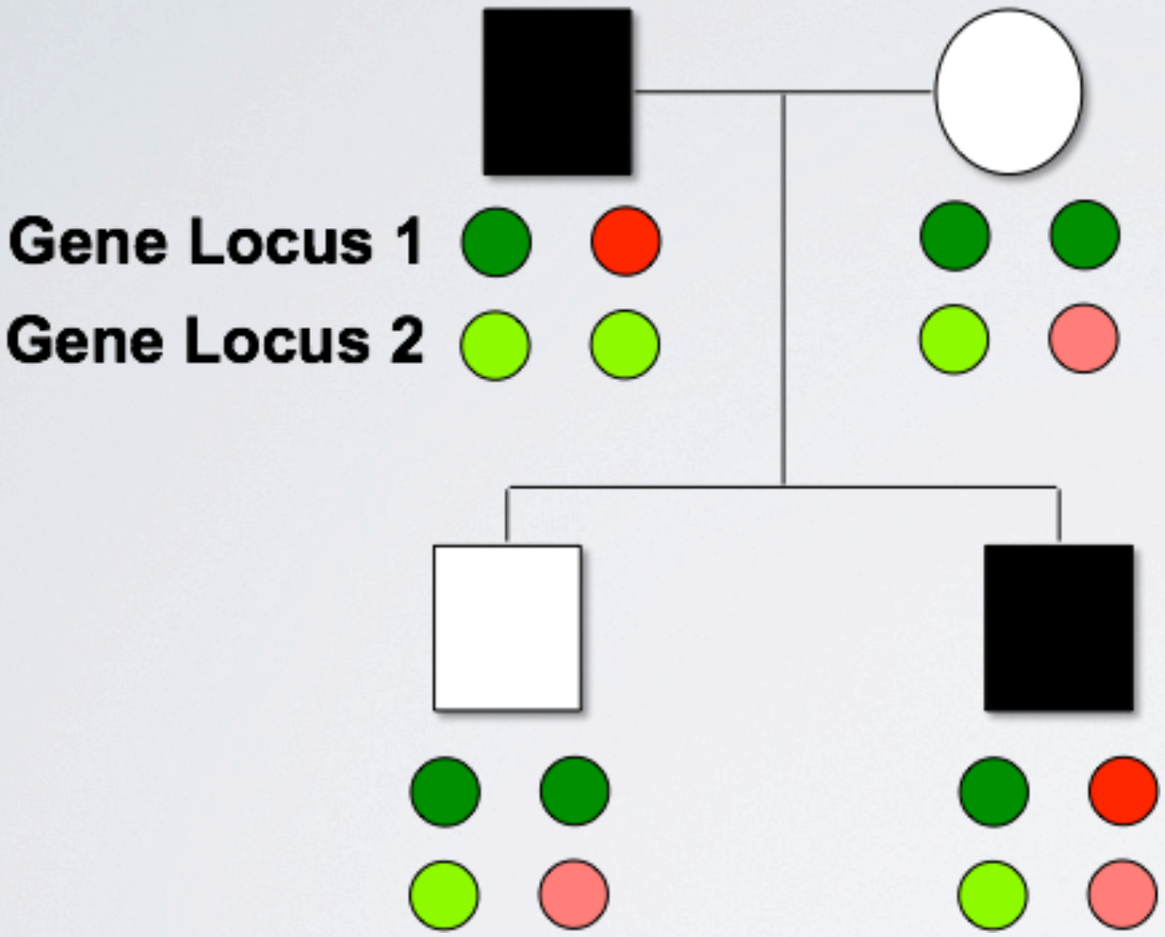
Epistasis



gene interaction results in modification of phenotype

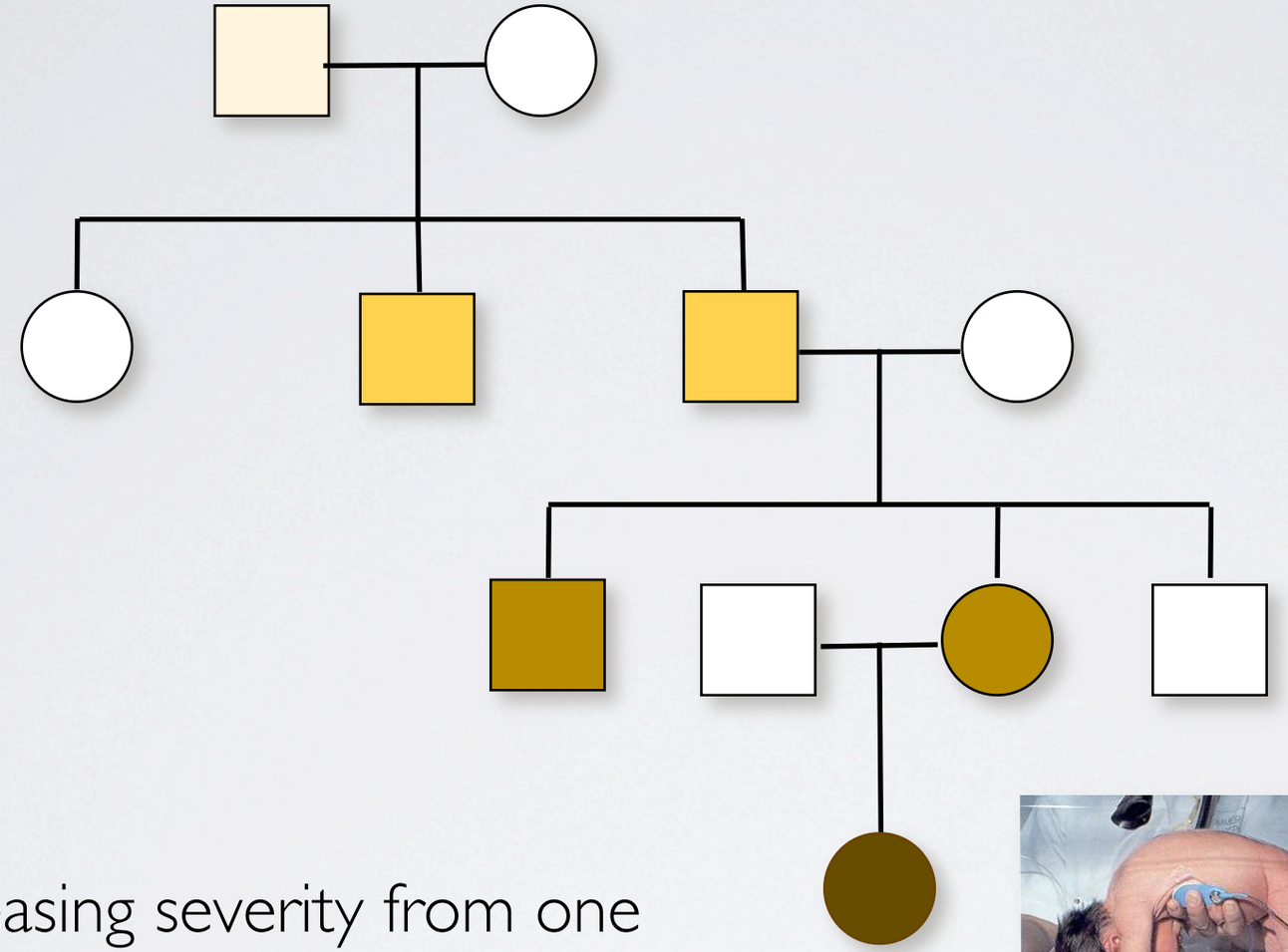


Digenic Inheritance

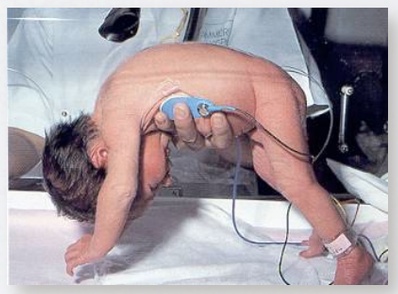


- retinitis pigmentosa
- heterozygosity for two genes required for phenotype
- neither alone is sufficient to produce phenotype
- encode proteins that interact with one another

Anticipation



- increasing severity from one generation to next
- commonly seen in triplet repeat disorders



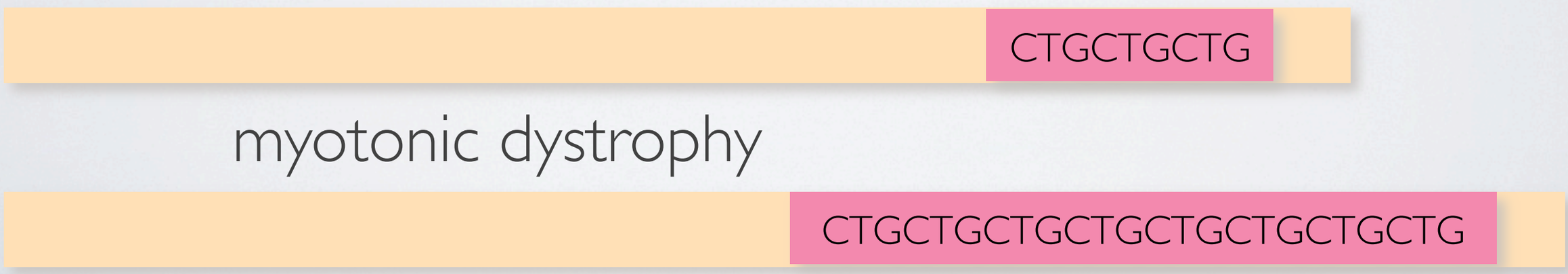
myotonic dystrophy

Triplet Repeat Expansion

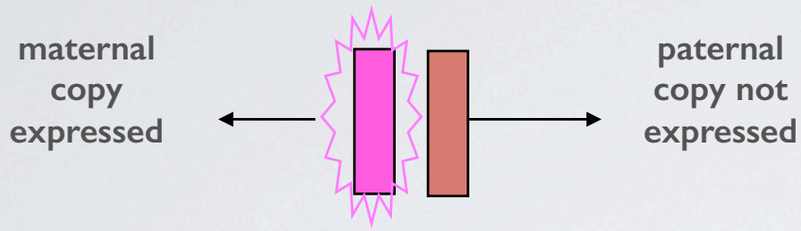
fragile X syndrome

Huntington disease
spinocerebellar ataxia

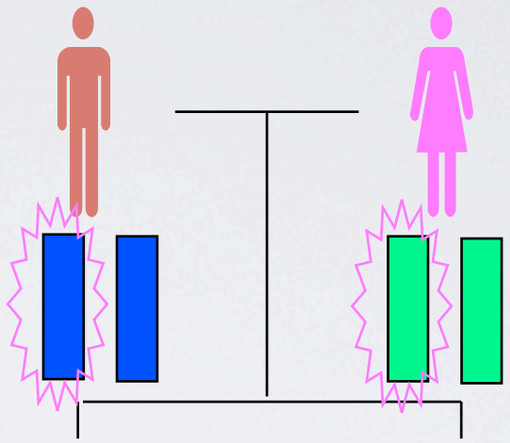
myotonic dystrophy

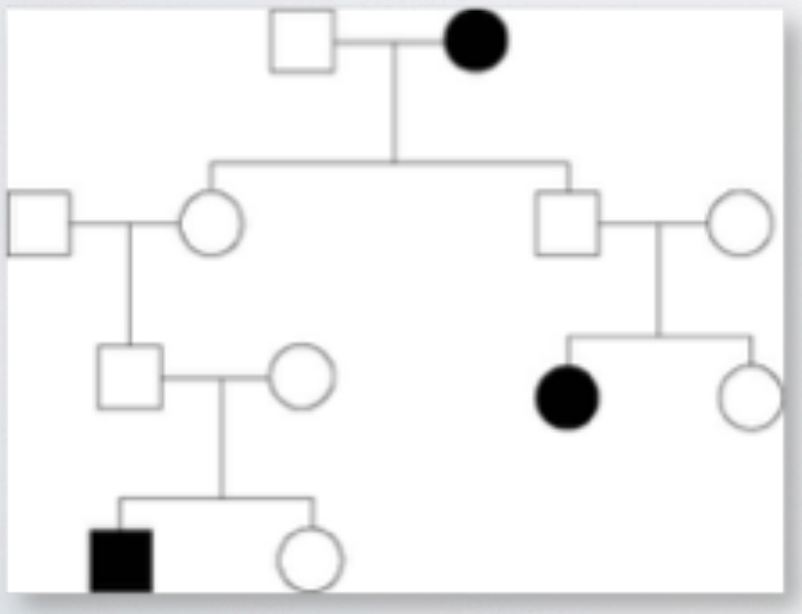


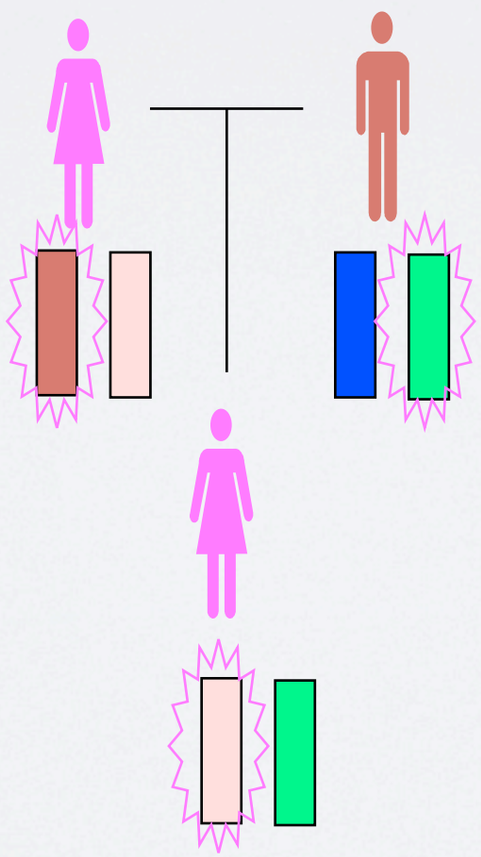
Genomic Imprinting

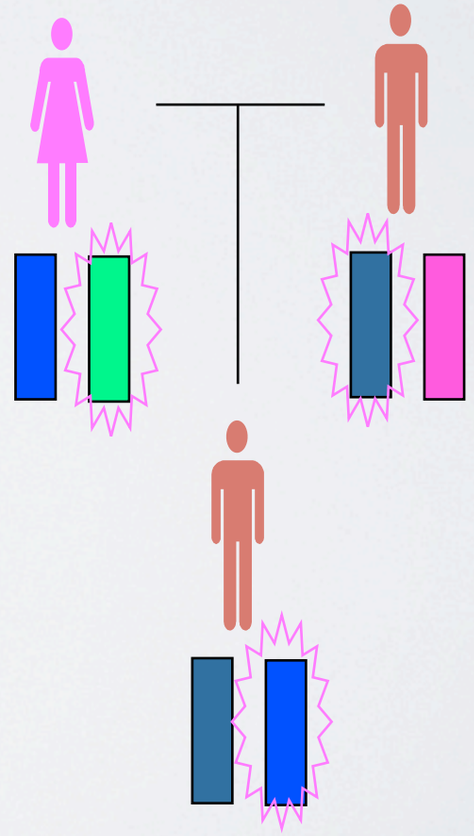


Imprinting: differential expression of maternal and paternal copy of a gene

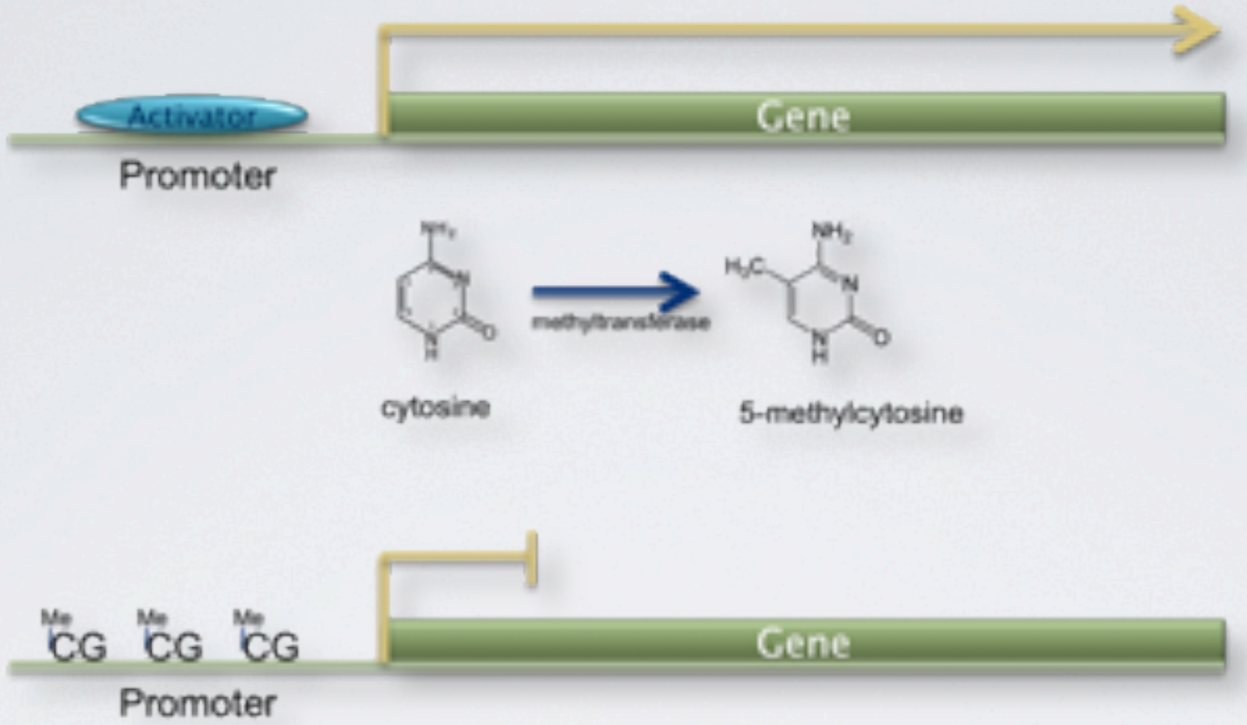


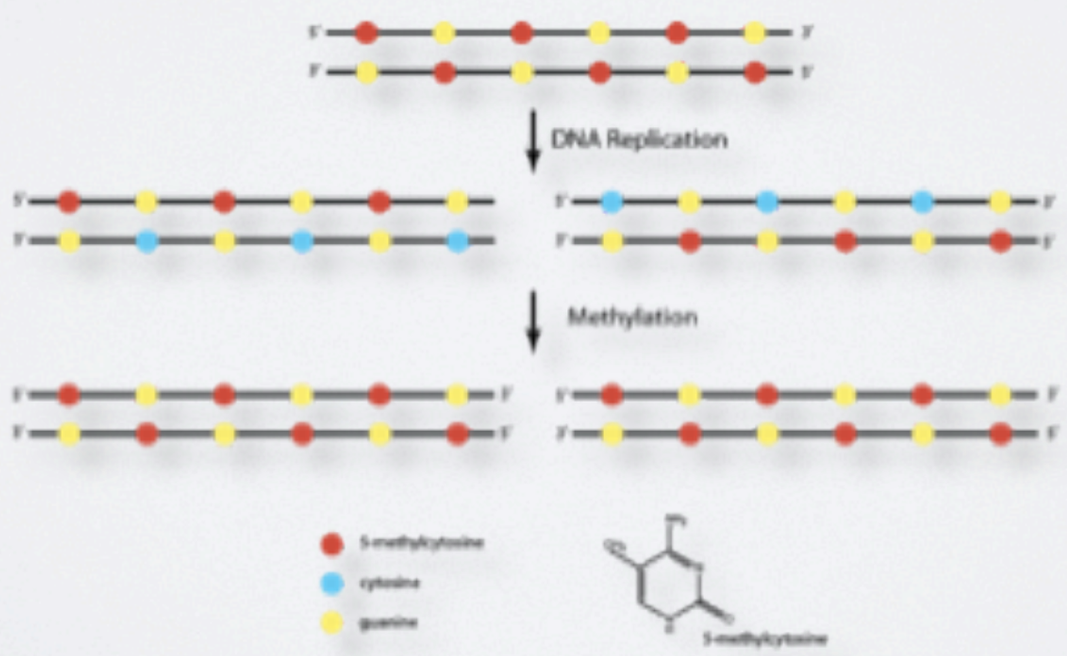




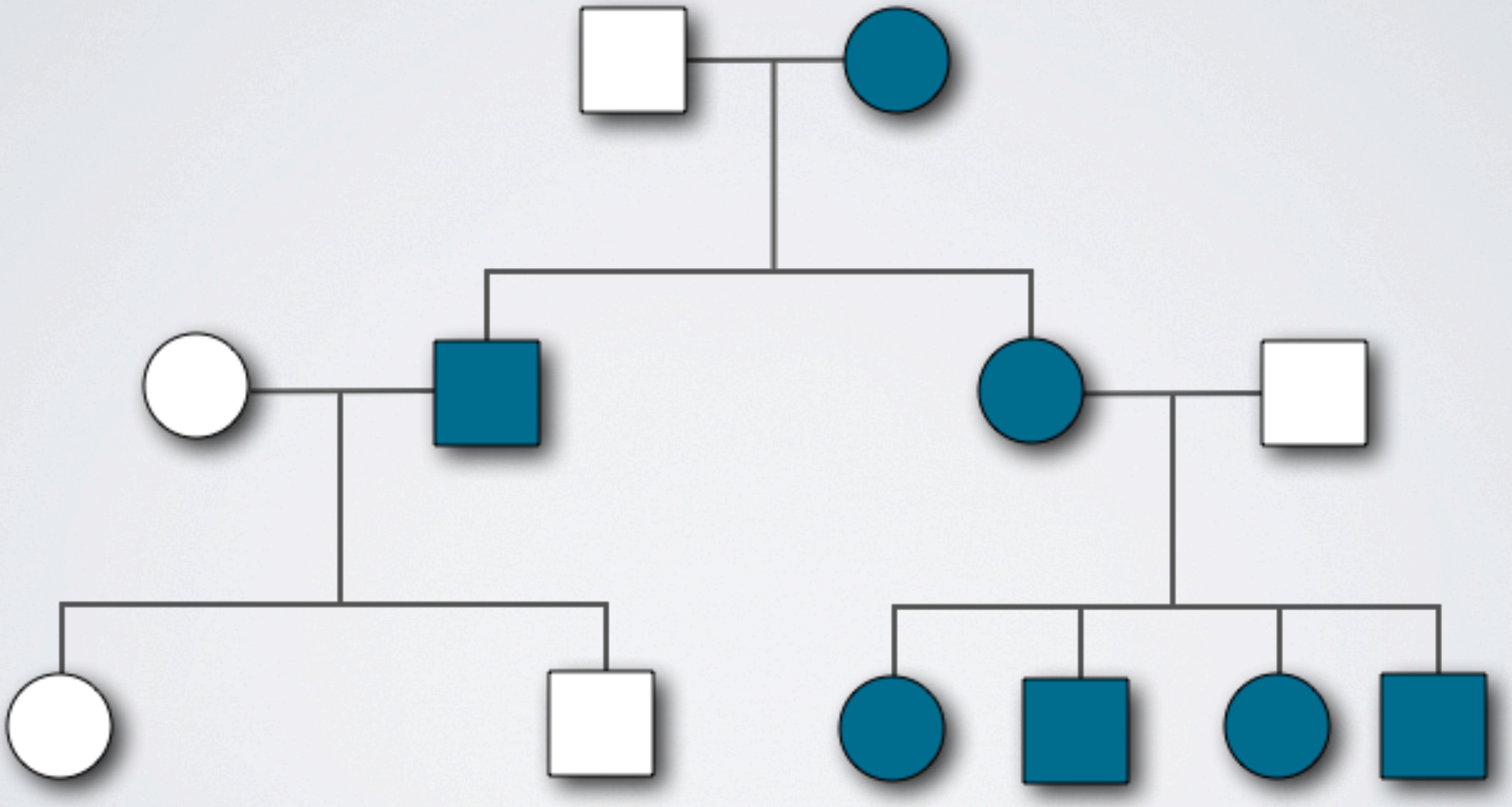


Epigenetics

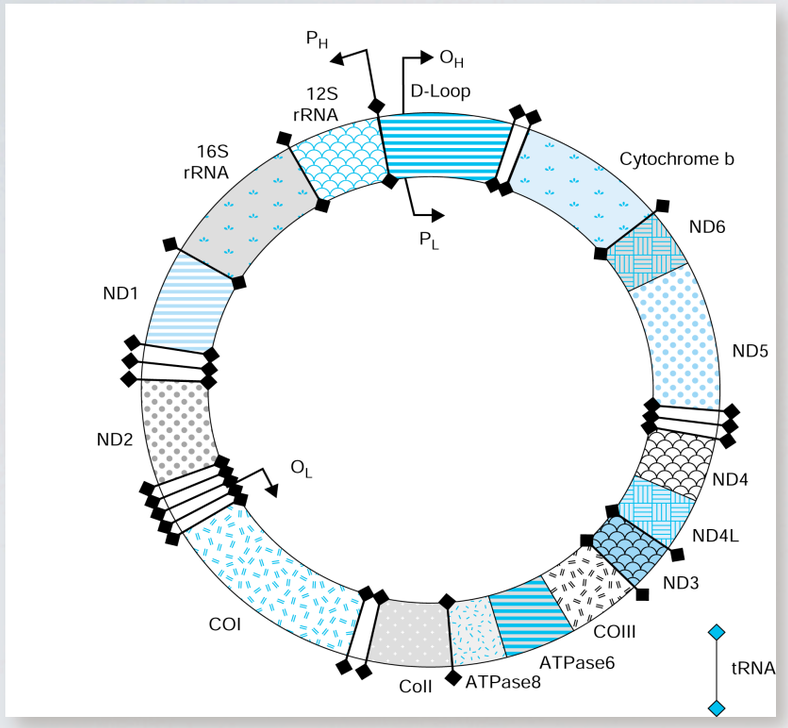




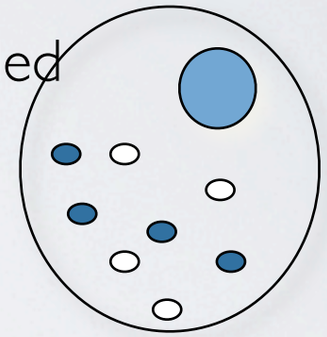
Maternal Inheritance

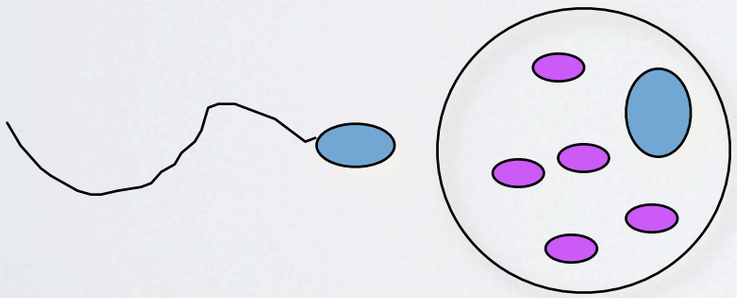


Mitochondrial Genetics

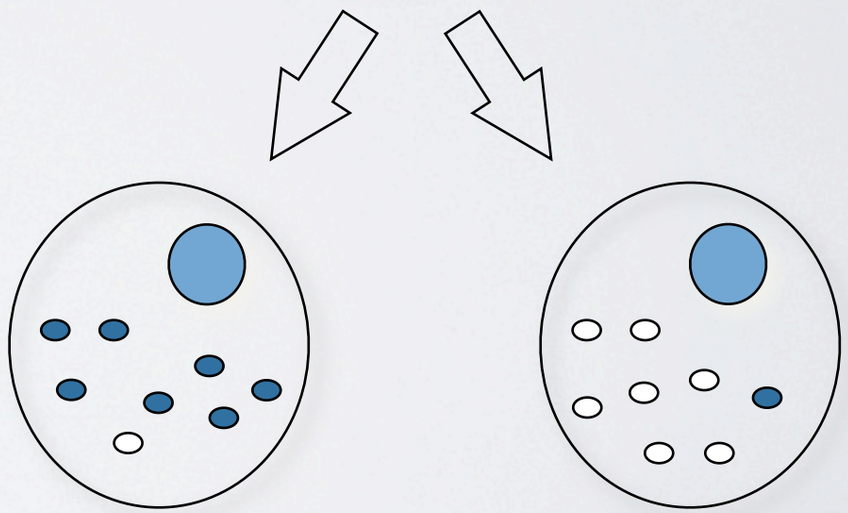


- 16 kb circular double-stranded DNA
- multiple copies per mitochondrion
- 13 subunits of mitochondrial proteins, tRNA's, rRNA's
- most mitochondrial proteins encoded in nucleus



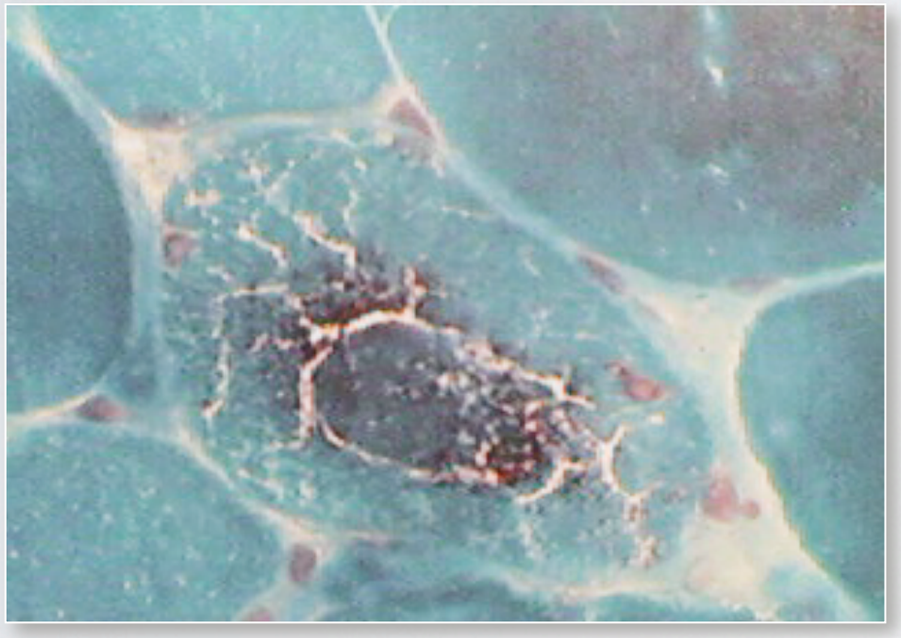


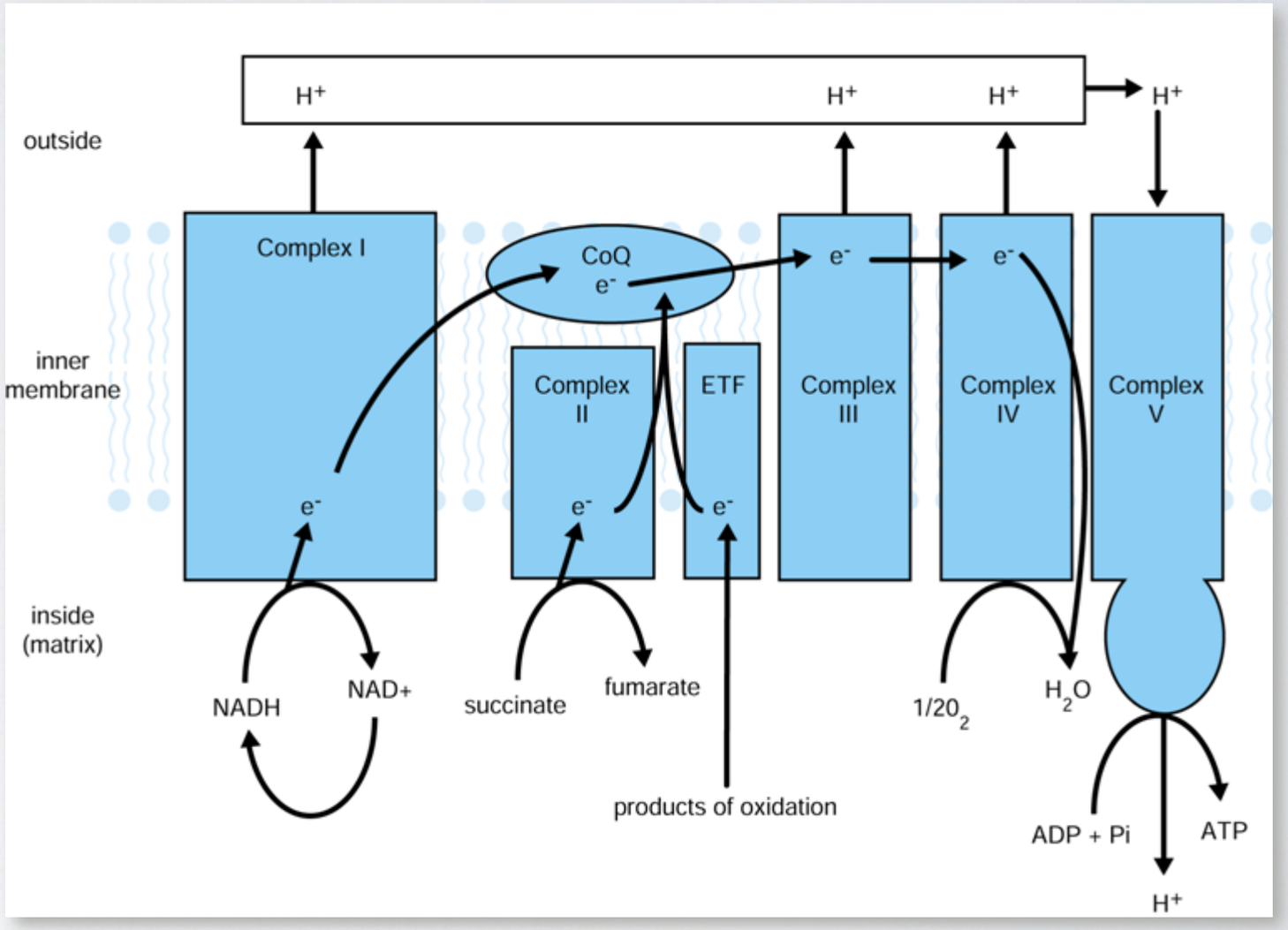
fertilization



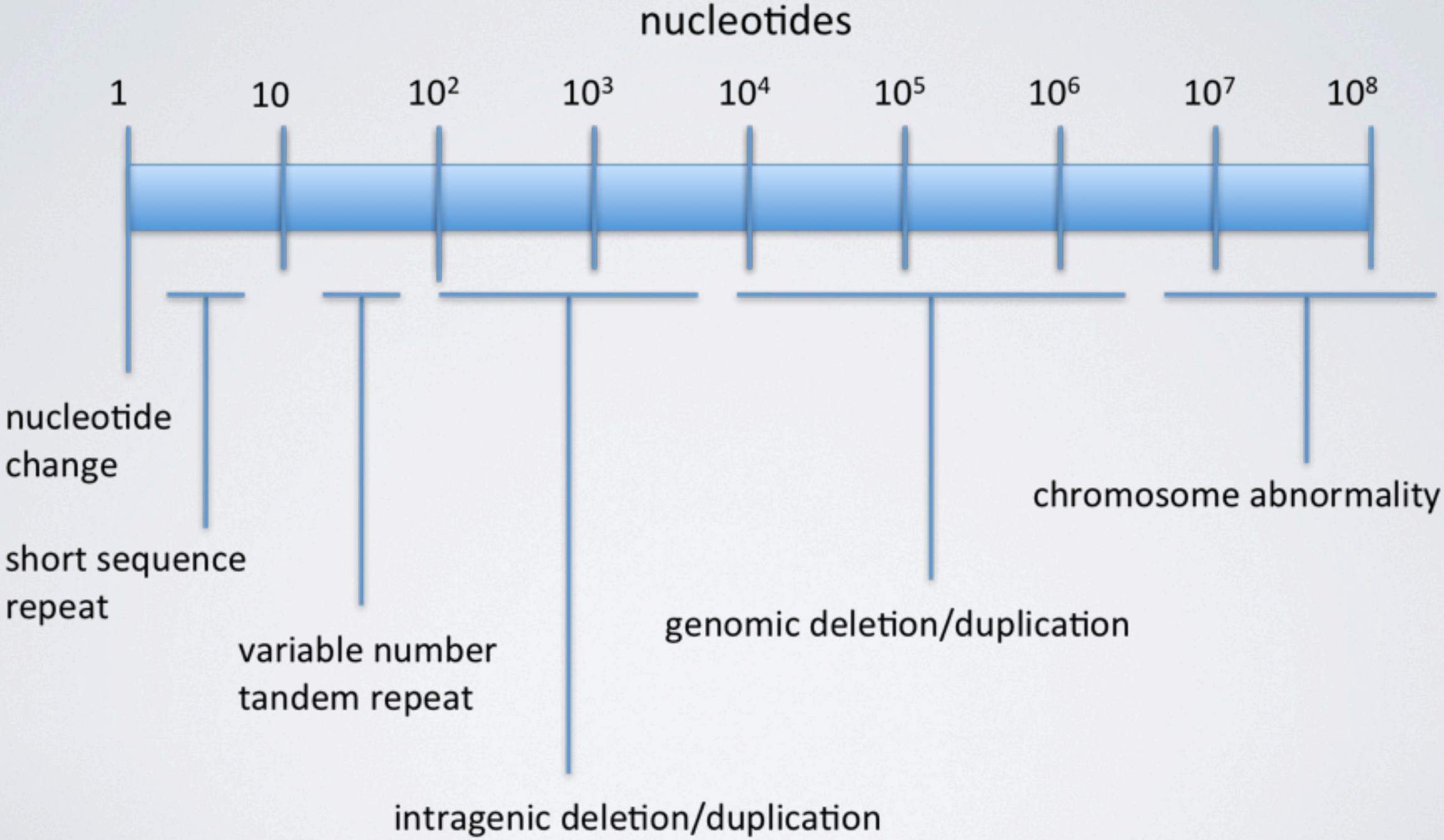
heteroplasmy

Energy Metabolism

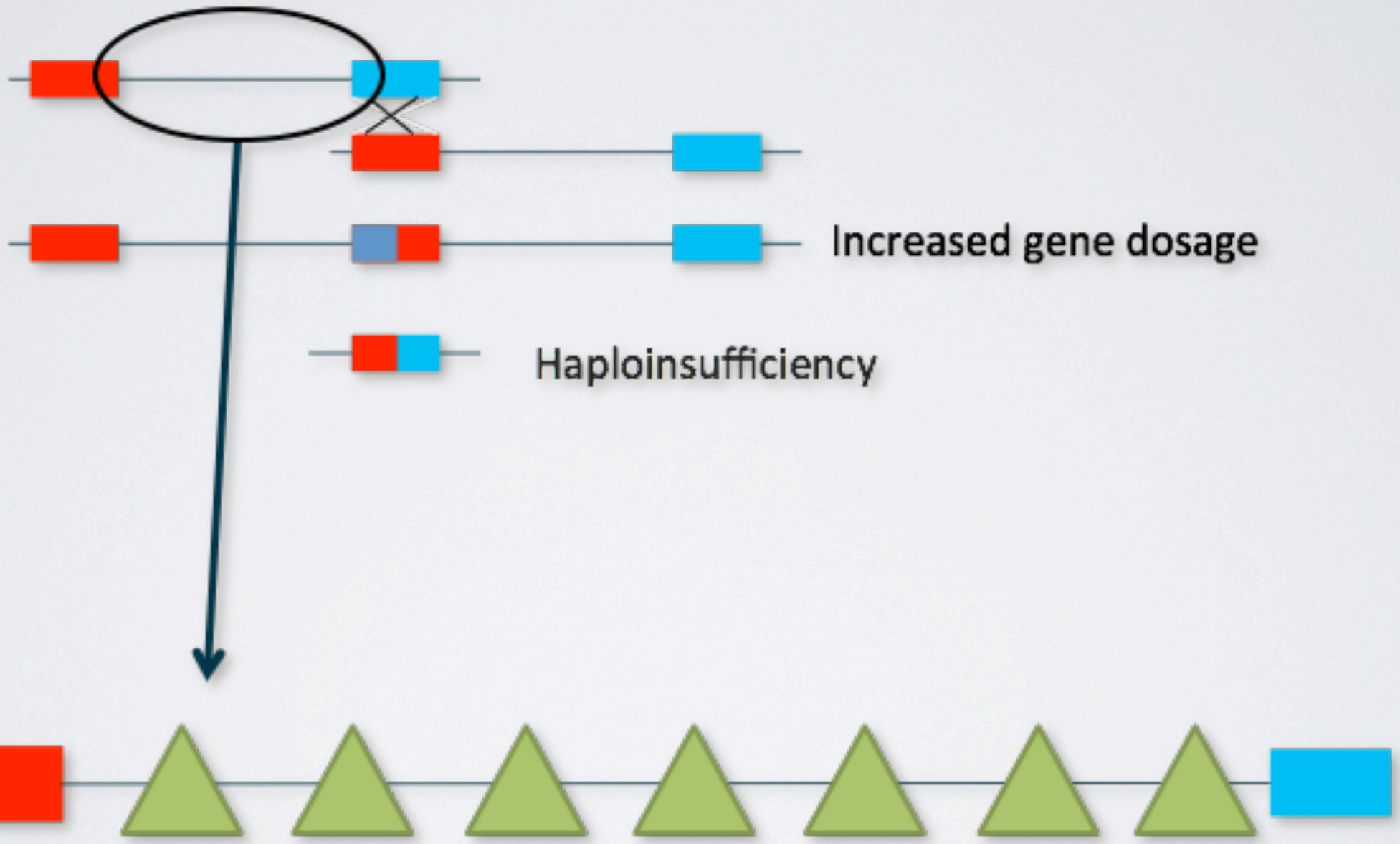




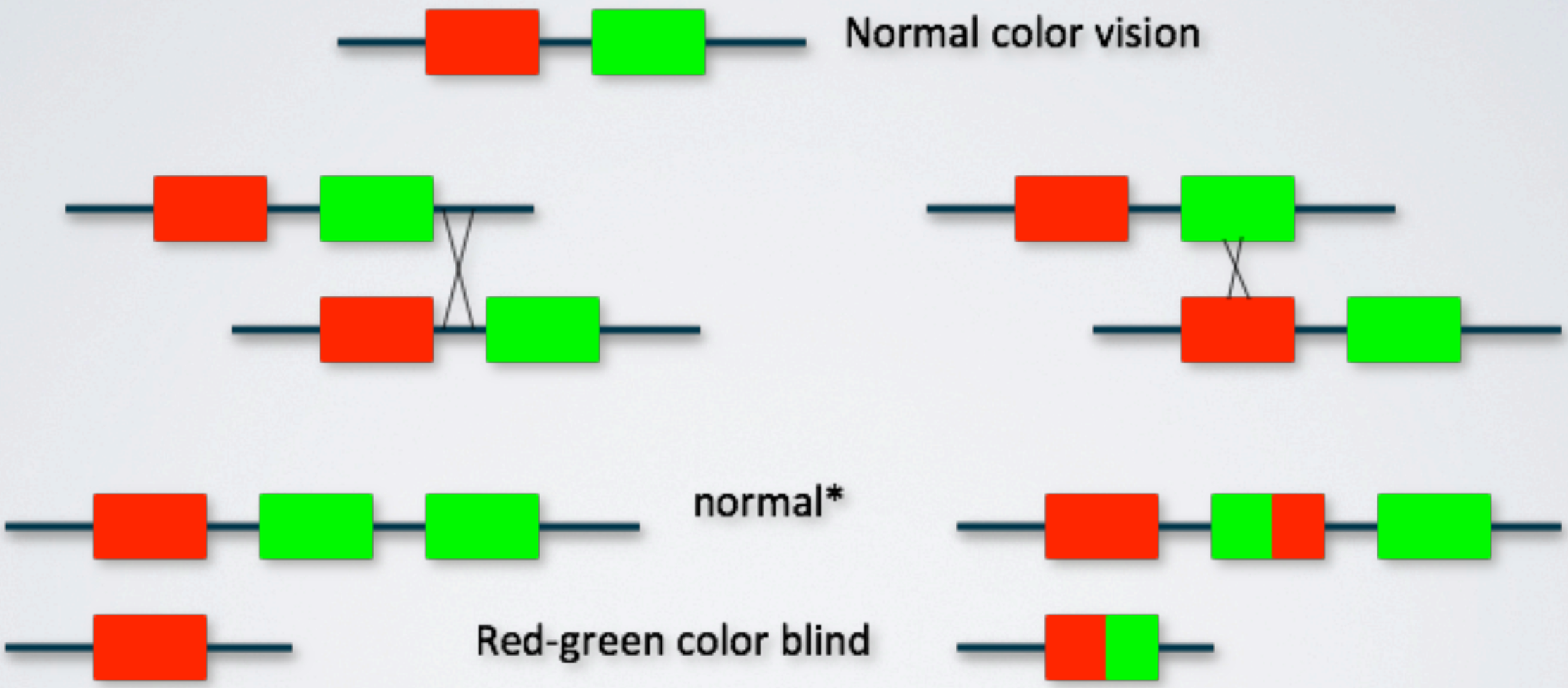
Gene Mutation



Copy Number Variation

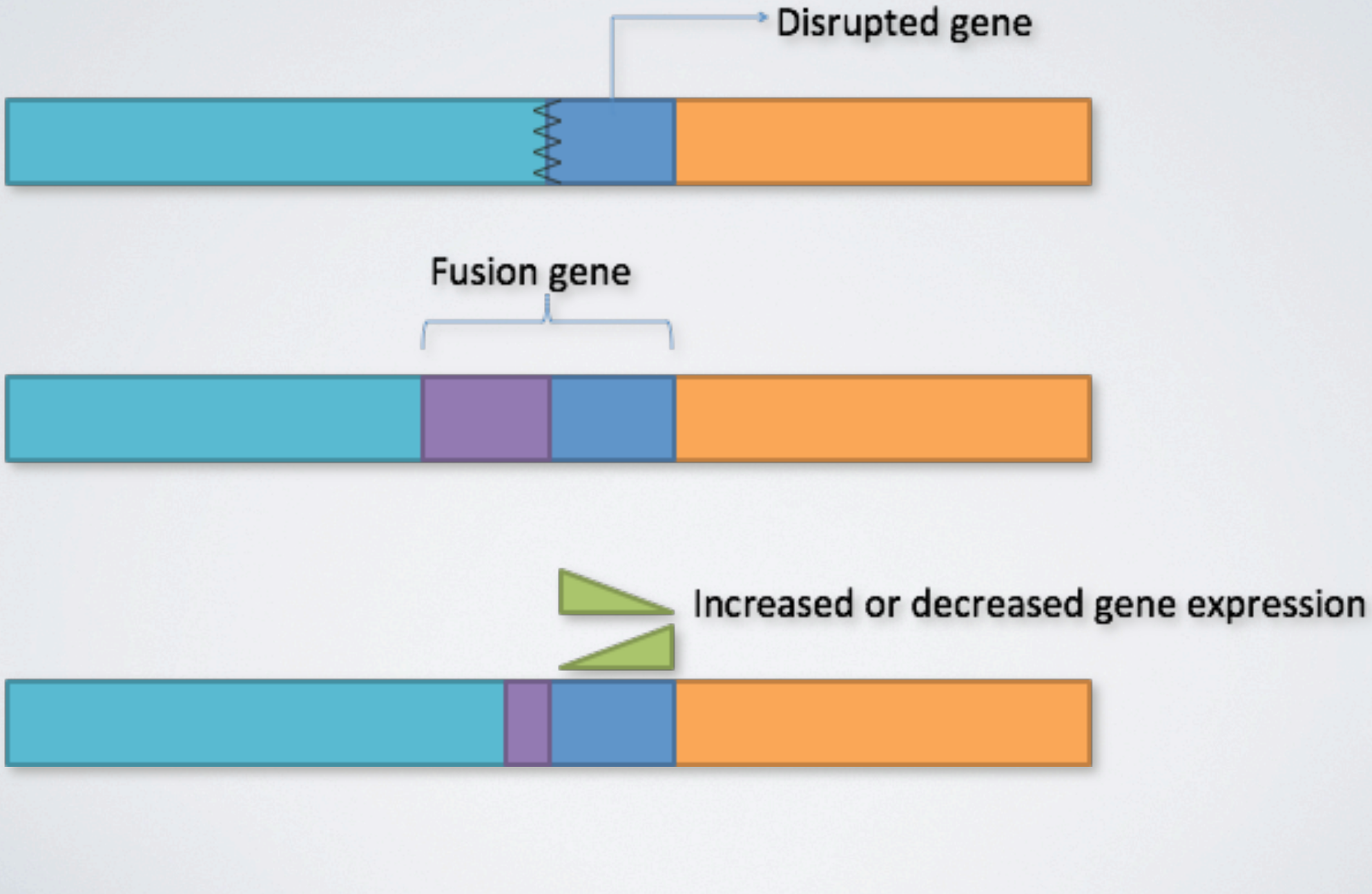


Color Blindness

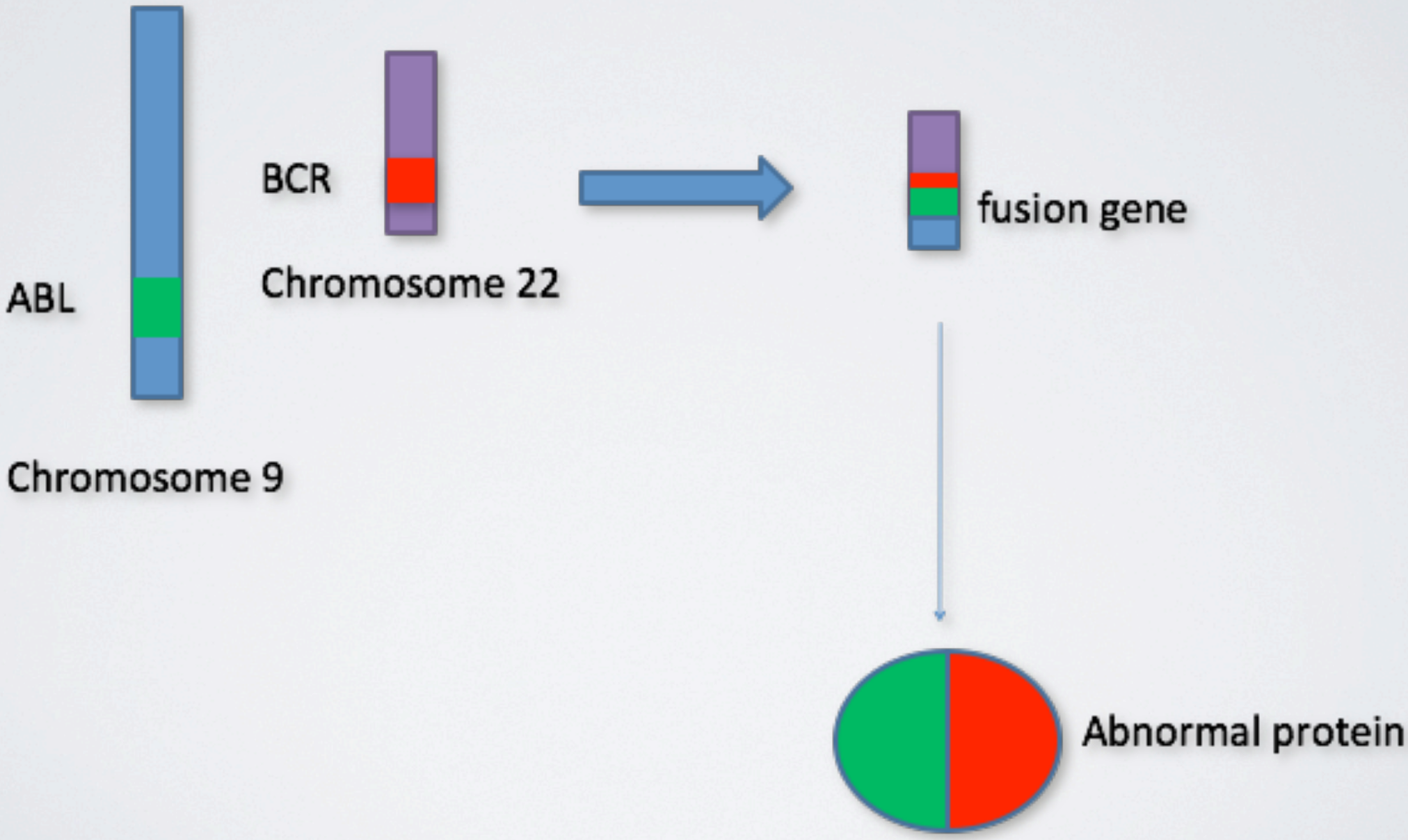


* Color vision may be abnormal if green gene not expressed

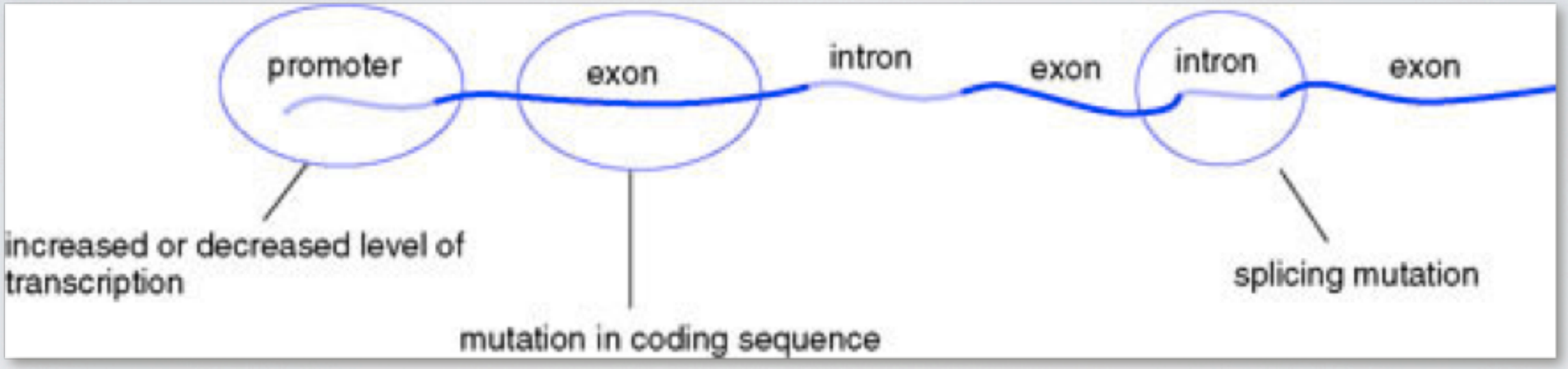
Gene Rearrangement



Philadelphia Chromosome



Gene Mutation



Point Mutations

TCC CAA ATC GTC CCT CGA GTT
ser gln ile val pro arg val

wild type sequence

TCC CAG ATC GTC CCT CGA GTT
ser gln ile val pro arg val

silent mutation

TCC CAA ATC CTC CCT CGA GTT
ser gln ile leu pro arg val

conservative mutation

TCC CAA ATC GTC GCT CGA GTT
ser gln ile val ala arg val

non-conservative mutation

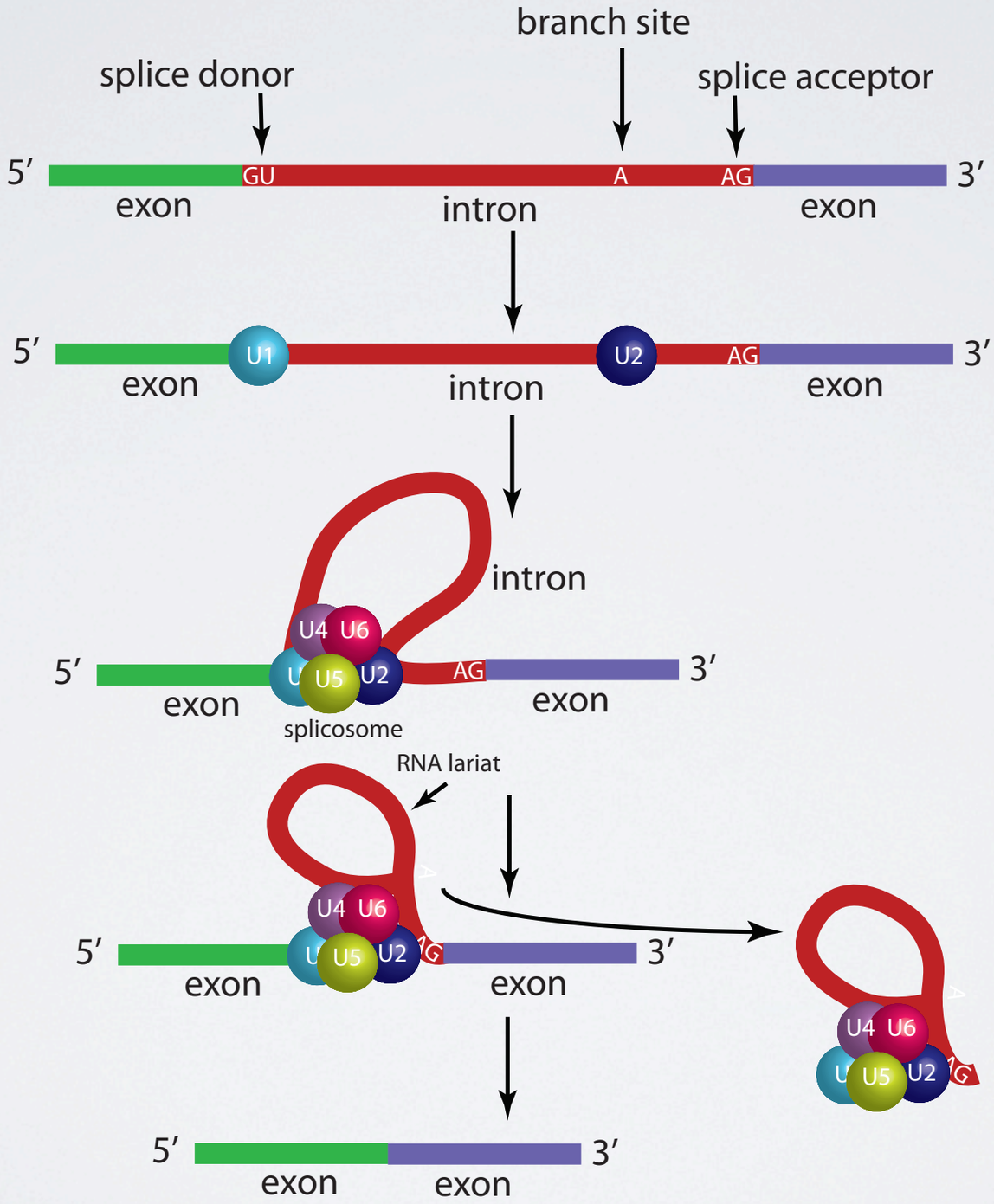
TCC CAA ATC GTC CCT TGA GTT
ser gln ile val pro stop

stop mutation

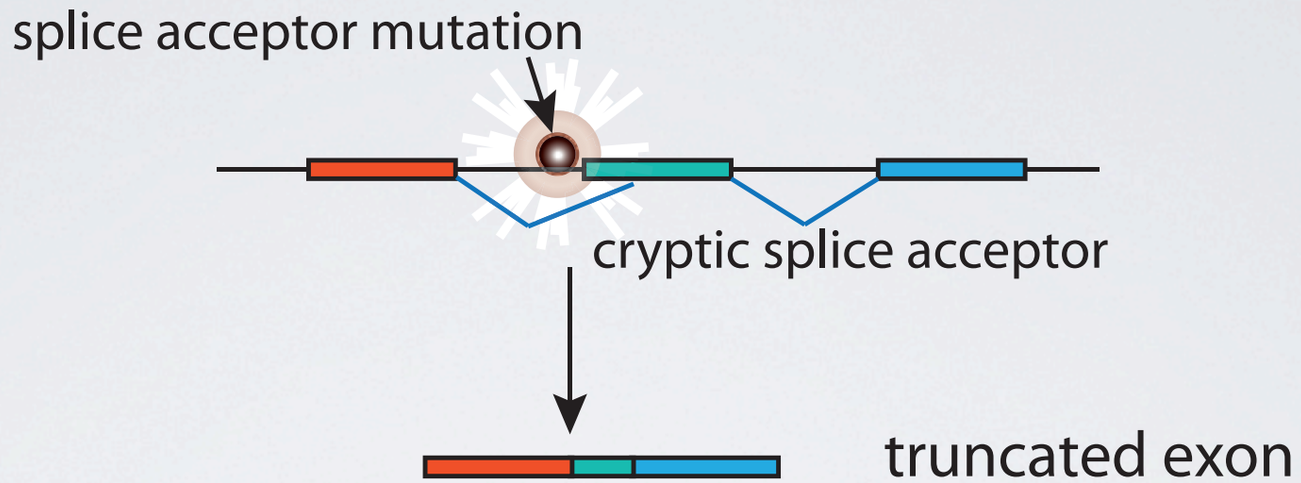
TCC CAG AAT CGT CCC TCG AGT T
ser gln asn arg pro ser ser

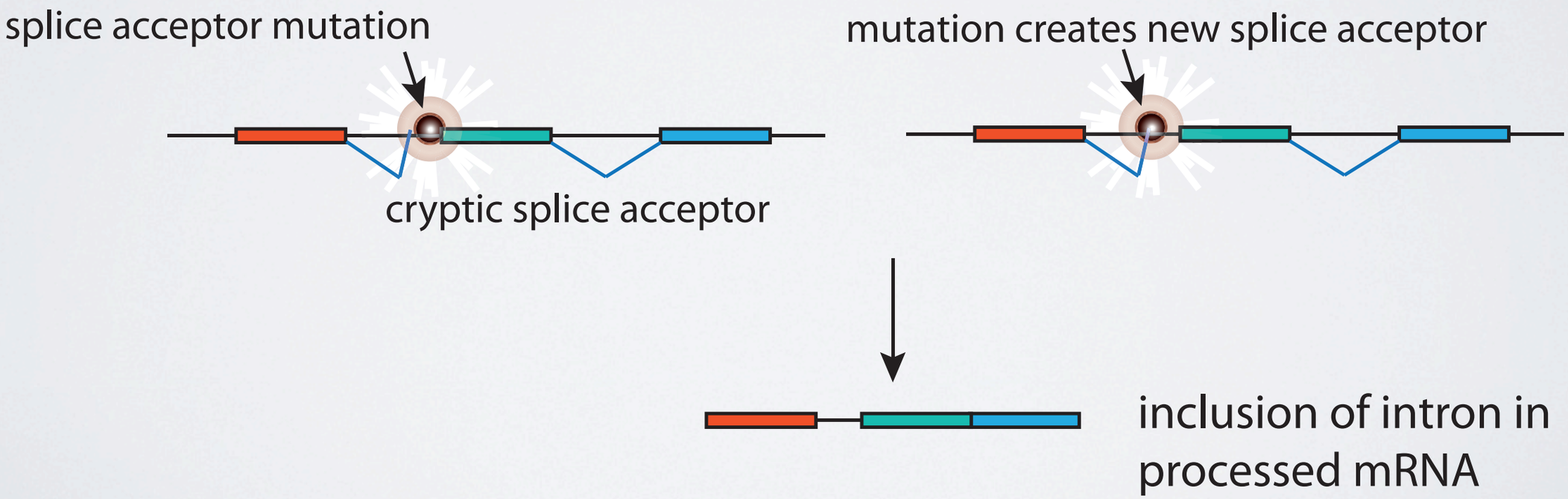
frameshift mutation

RNA Splicing



Splicing Mutation





Mutation Rate

doi:10.1038/nature09534

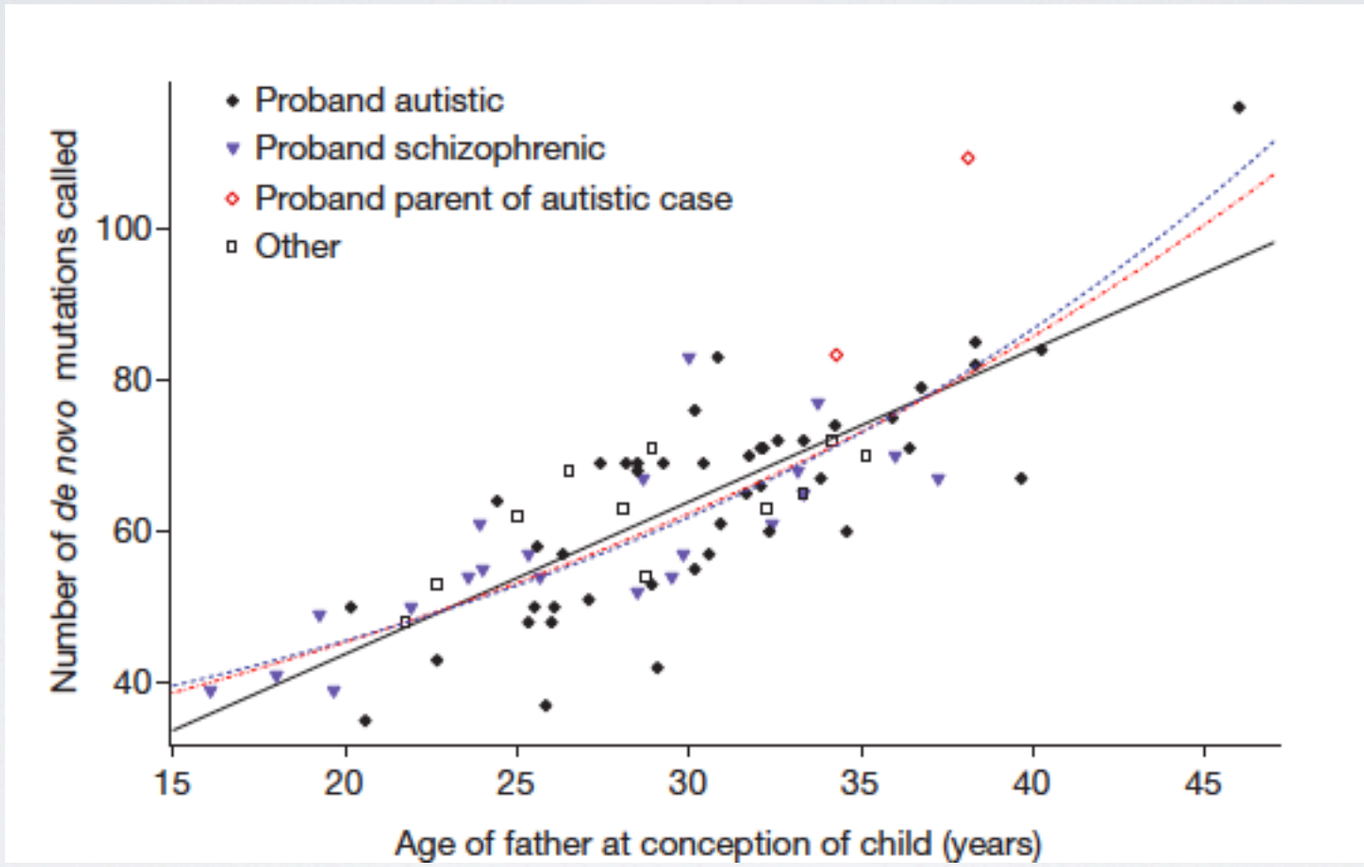
A map of human genome variation from population-scale sequencing

The 1000 Genomes Project Consortium*

The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype. Here we present results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. We undertook three projects: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother–father–child trios; and exon-targeted sequencing of 697 individuals from seven populations. We describe the location, allele frequency and local haplotype structure of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously undescribed. We show that, because we have catalogued the vast majority of common variation, over 95% of the currently accessible variants found in any individual are present in this data set. On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders. We demonstrate how these results can be used to inform association and functional studies. From the two trios, we directly estimate the rate of *de novo* germline base substitution mutations to be approximately 10^{-8} per base pair per generation. We explore the data with regard to signatures of natural selection, and identify a marked reduction of genetic variation in the neighbourhood of genes, due to selection at linked sites. These methods and public data will support the next phase of human genetic research.

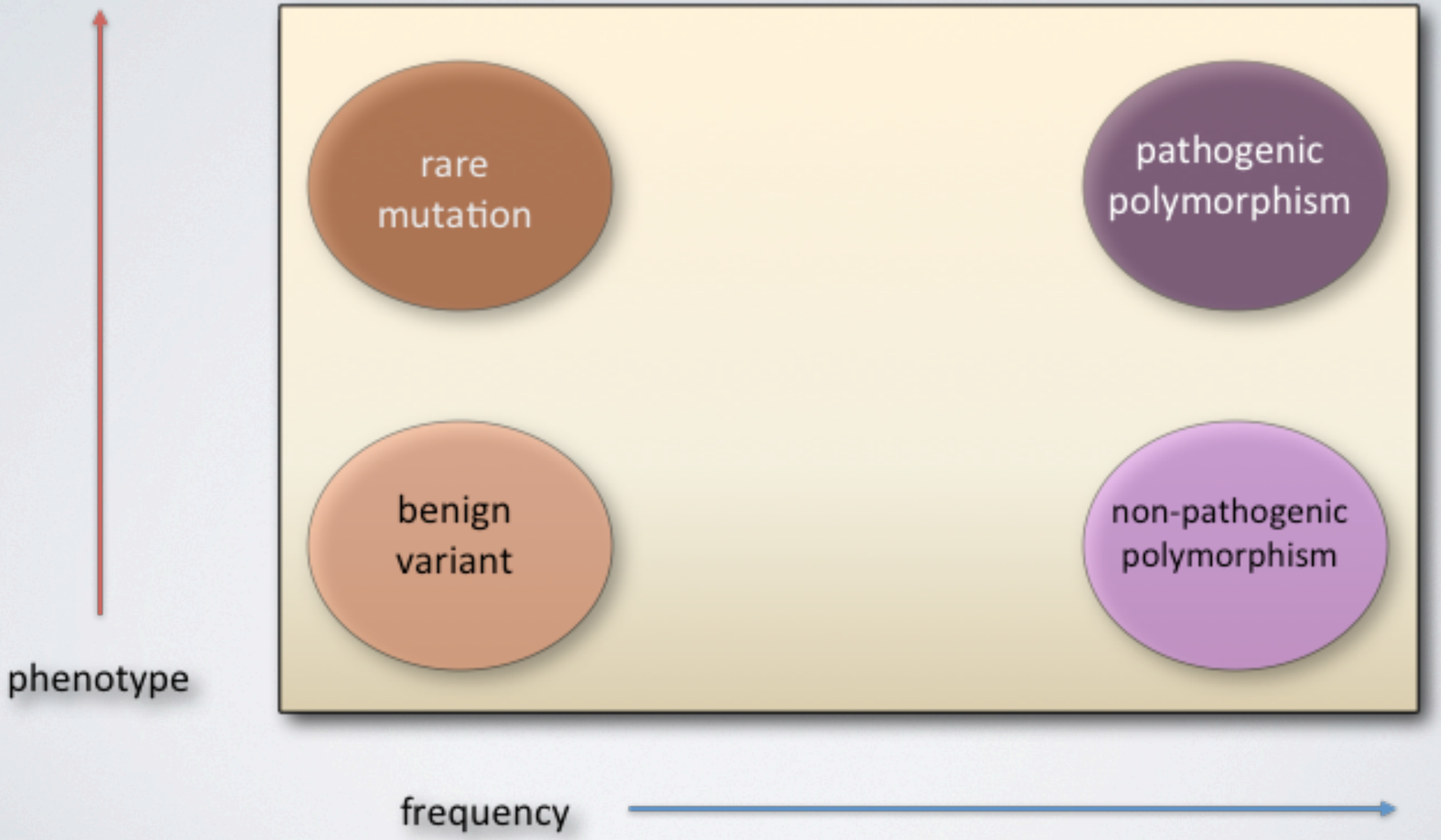
If there are 10^8 sperm per ejaculate, in principle every base could be mutated in at least one sperm cell and each germ cell has around 10 mutations

Paternal Age Effect

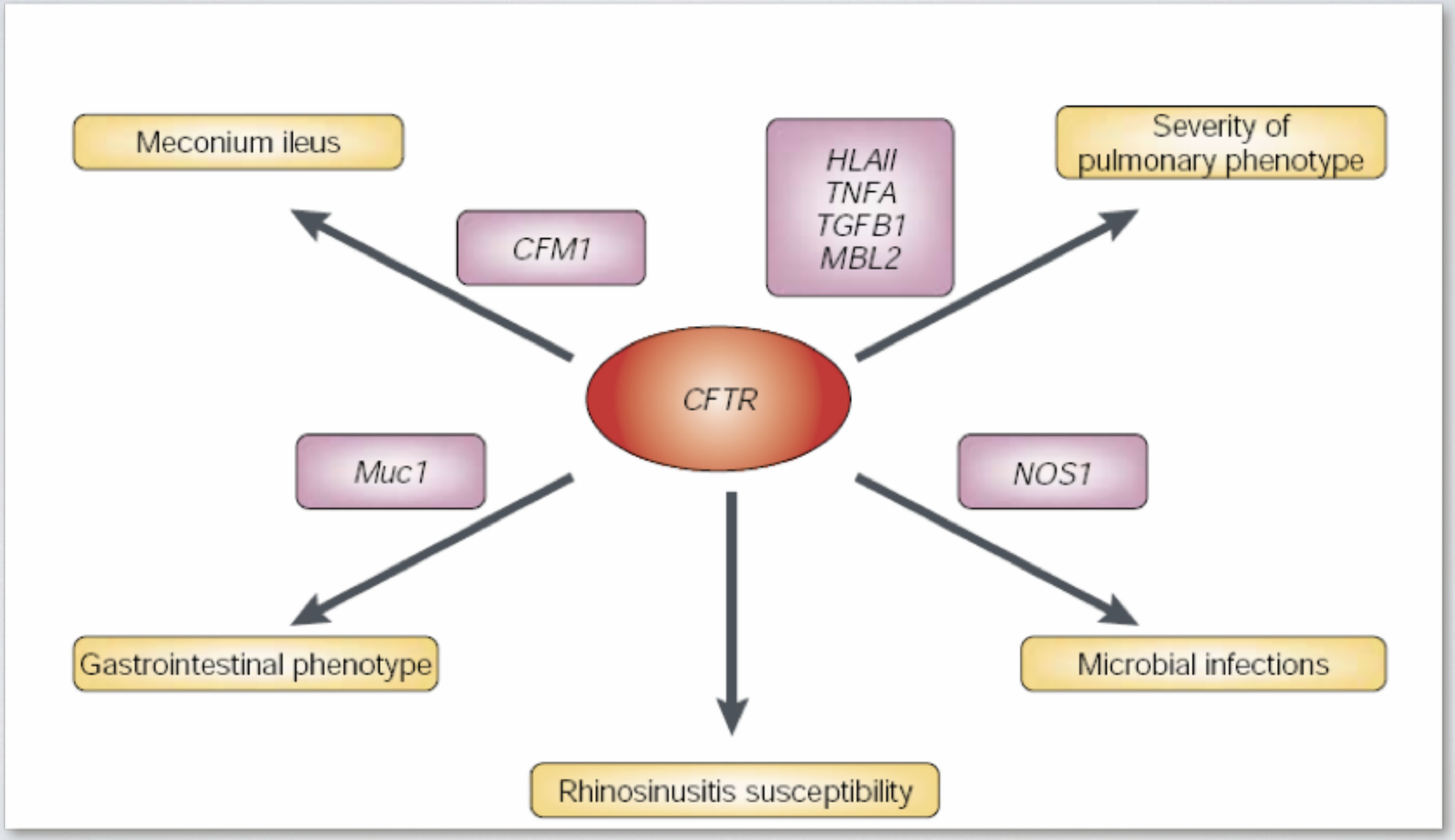


Kong et al. Nature 2012;488:41.

Polymorphism vs. Variant



Genetic Modifiers



Badano JL. *Nature Reviews Genetics* 2002;3:779-789.

Multifactorial Inheritance

- Tendency to recur in families
- Does not follow Mendelian genetics
- Combination of multiple genes and/or environmental factors
- Contribution to wide range of common disorders