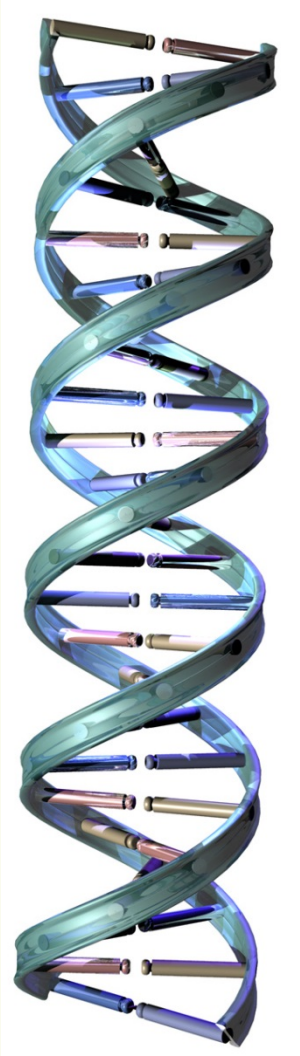
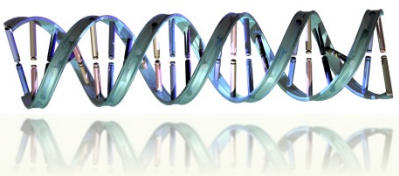
National Human Genome Research Institute



Genetics and Genomics Primer

2016 NHGRI Summer Short Course

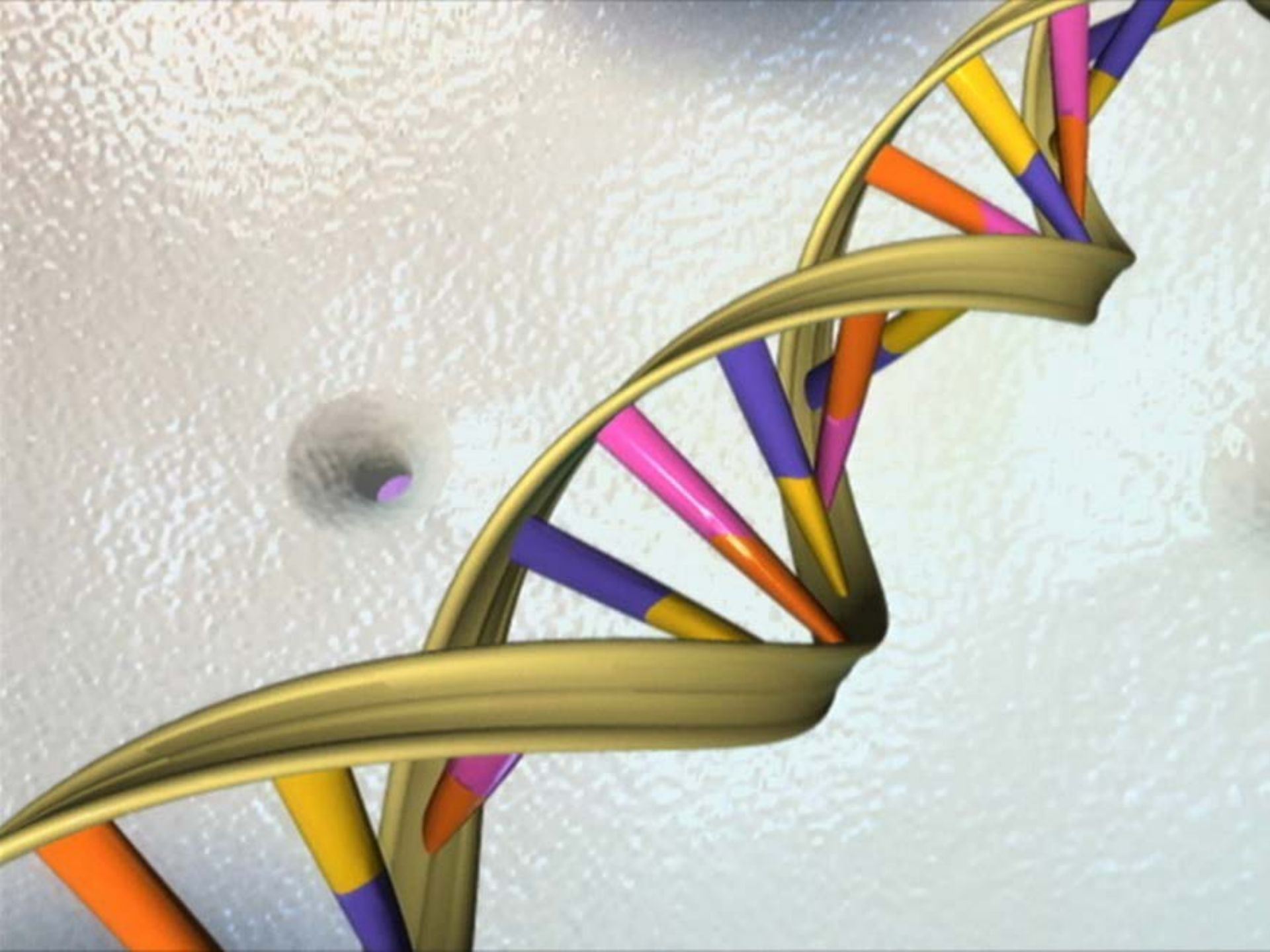
Bob Wildin, MD
Chief, Genomic Healthcare Branch
Division of Policy Communications and Education
NHGRI



The plan

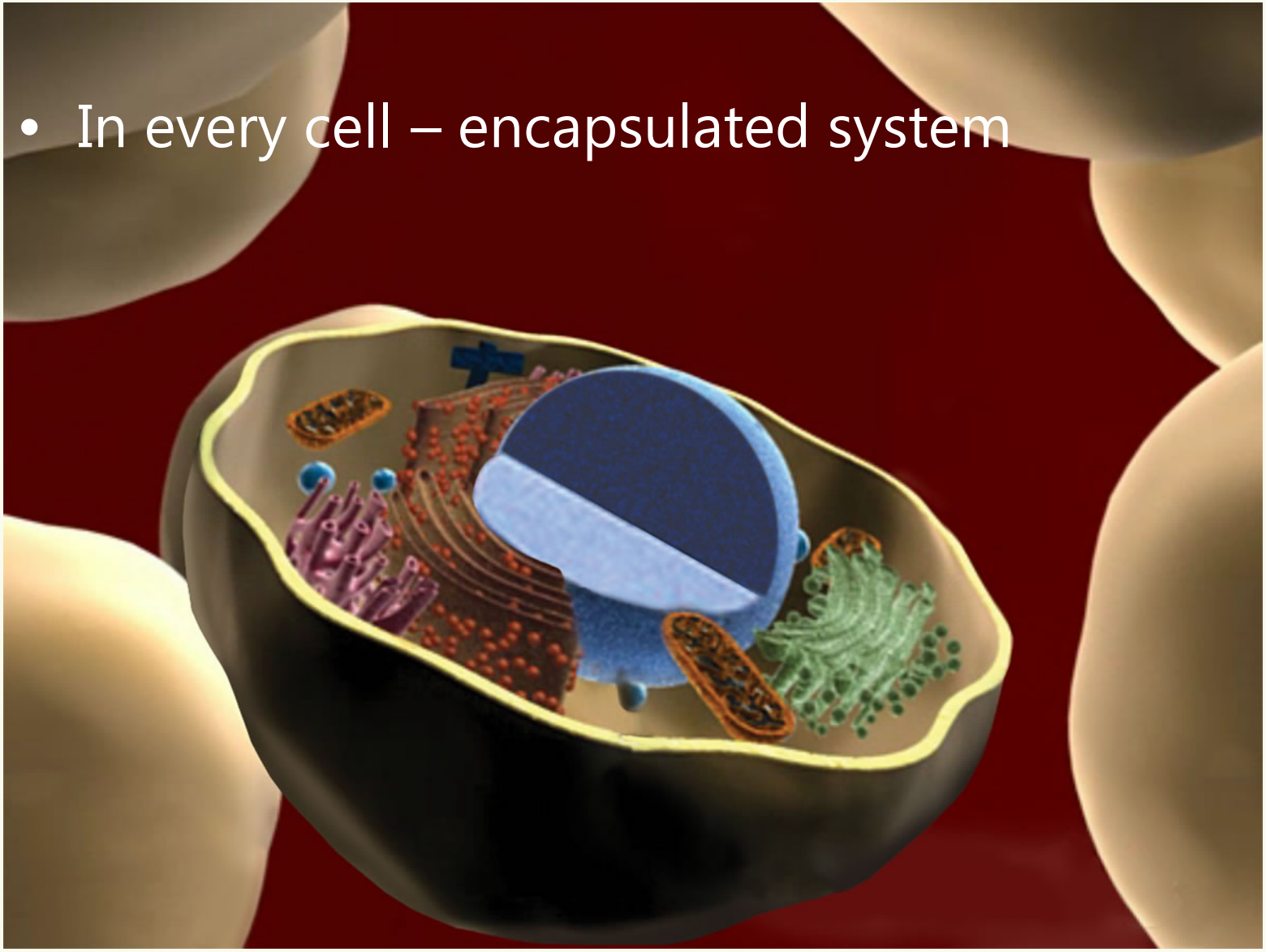
- DNA
- Genome
- Replication
- Variation
- Gene
- Proteins

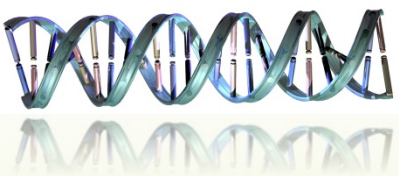
- Function
- Inheritance
- Balance
- Chance and Risk
- Bayes Theorem



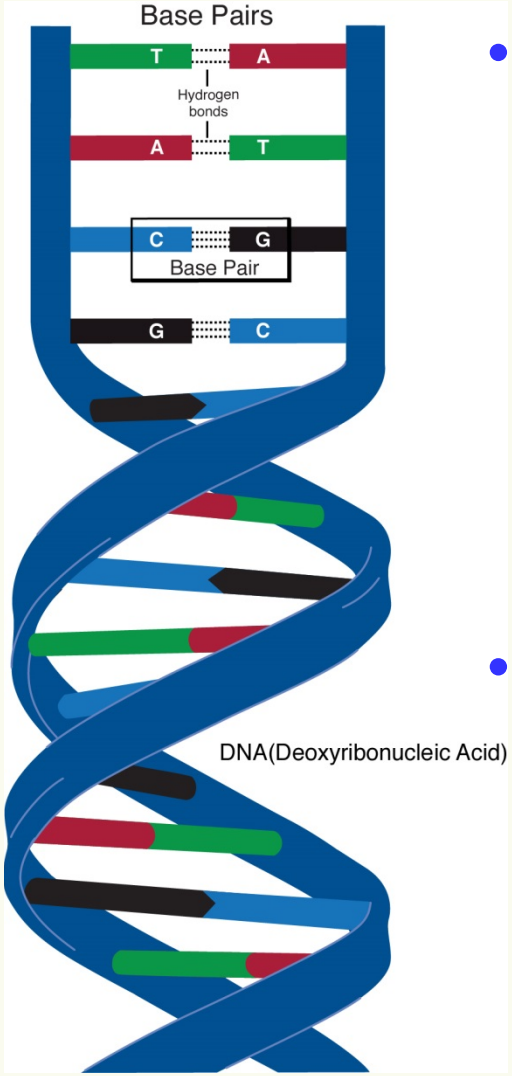
DNA: Where?

- In every cell – encapsulated system

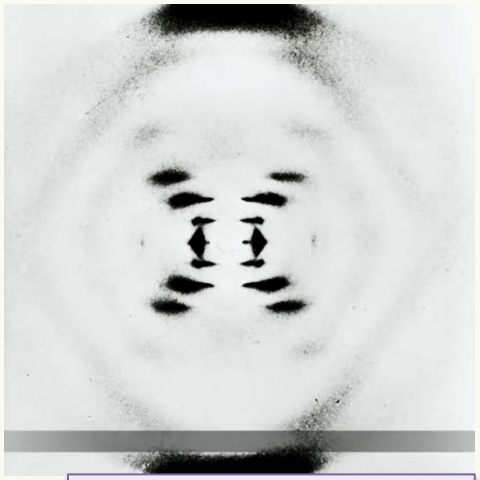




What is DNA?



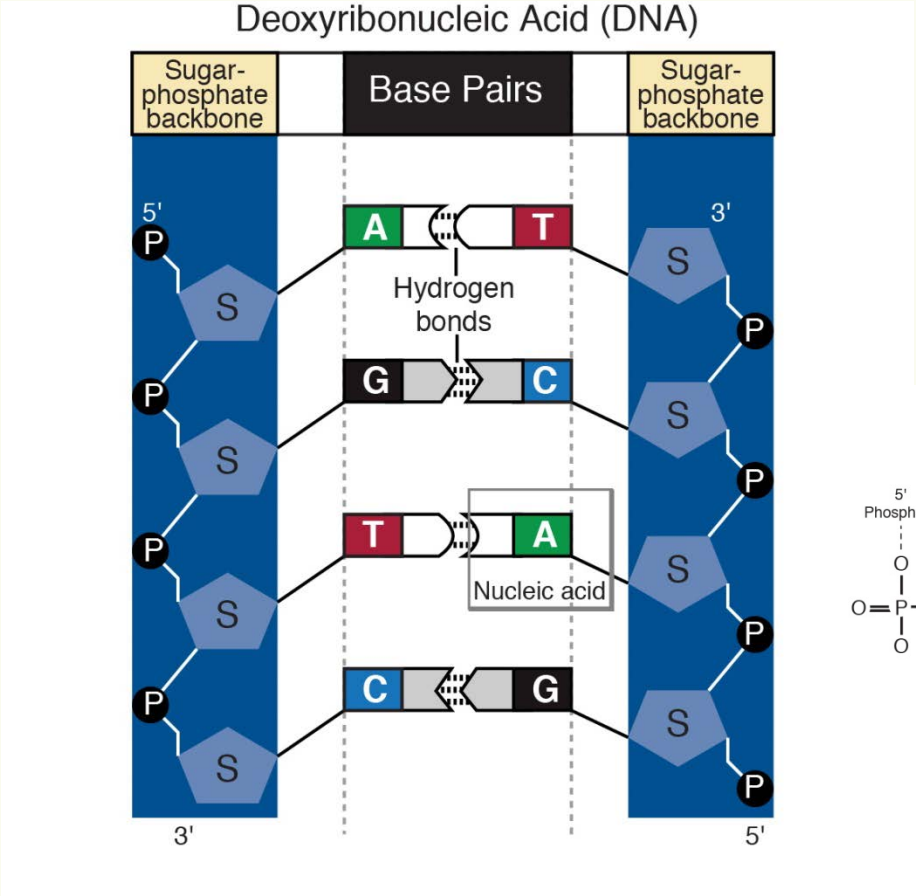
- DNA
  - Chemical
    - Deoxy-Ribonucleic Acid
  - Structure
  - Information
    - Sequence – order of bases
    - Genetic Code – protein sequence
  - How to get it (Isolation, extraction)
    - *DNA from strawberries video*
  - How fragile is it?
- Confusing points
  - Double Helix – two strands
  - Strands have opposite direction
  - Double stranded vs. single stranded
  - Two gene copies vs. one gene copy
    - (diploid vs. haploid)

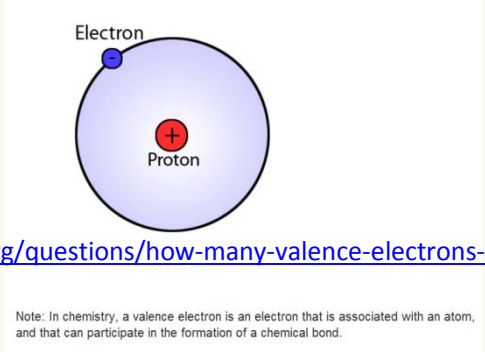


atmroinofni

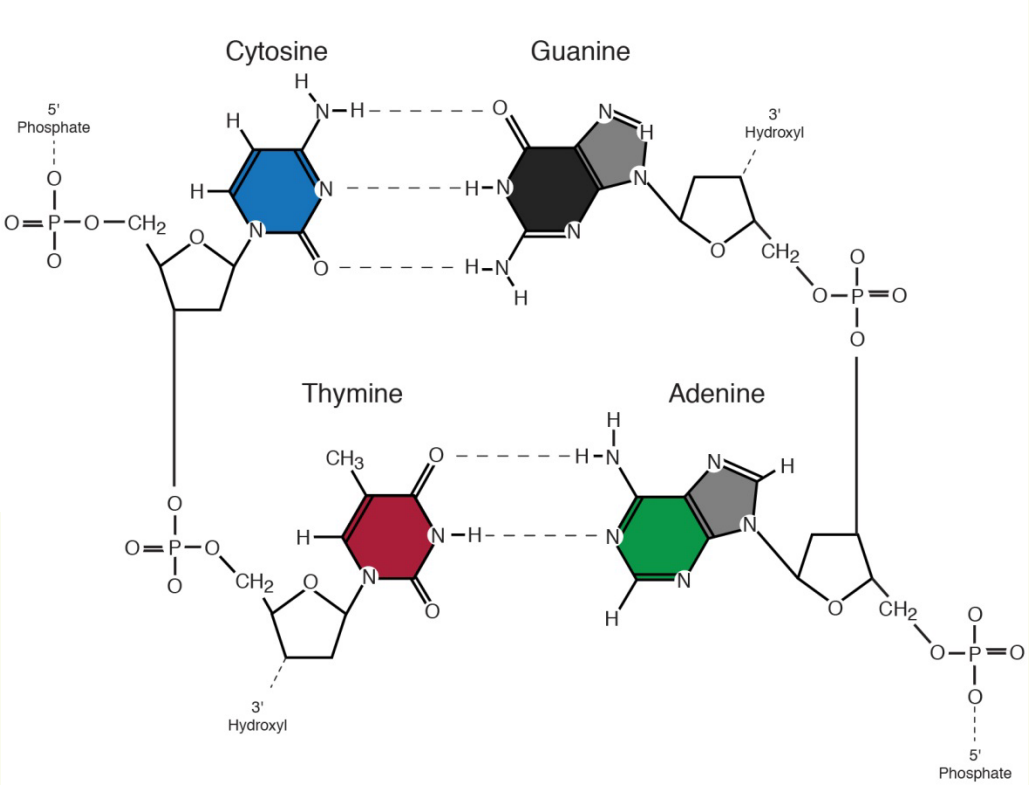
information

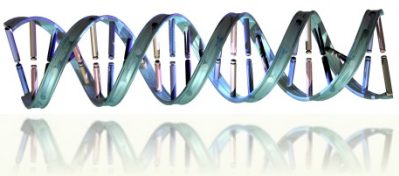
Double Stranded DNA- Base Pairing



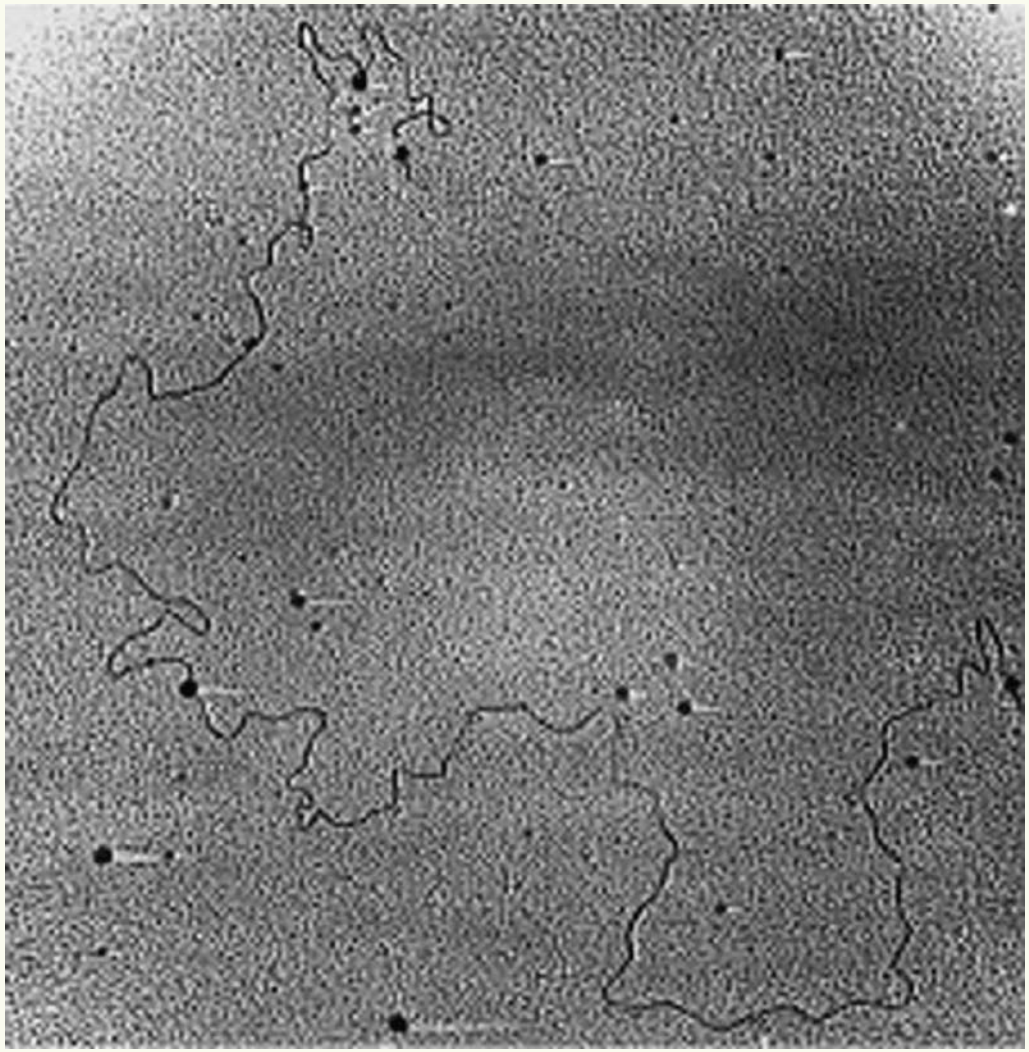


http://socratic.org/questions/how-many-valence-electrons-does-hydrogen-have

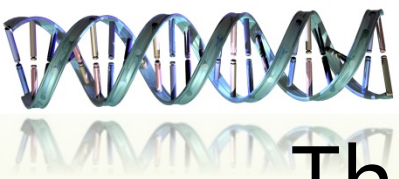




What is a Genome?



http://femsle.oxfordjournals.org/content/227/1/87



The Human Genome: Your DNA

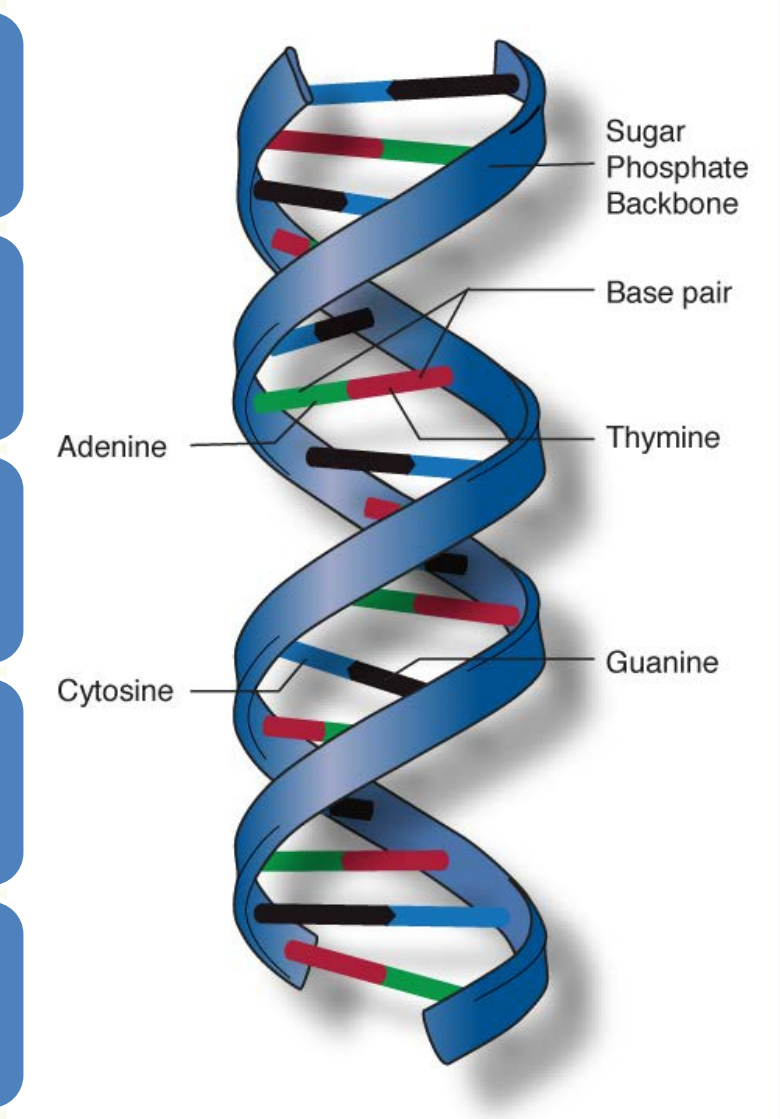
All the genetic material in the nucleus, plus the mitochondrial genome

Molecules of DNA that contain the coded instructions for how to build, maintain, and replicate a human being *Interview (1st 33 seconds)*

Is not identical in anyone but identical twins

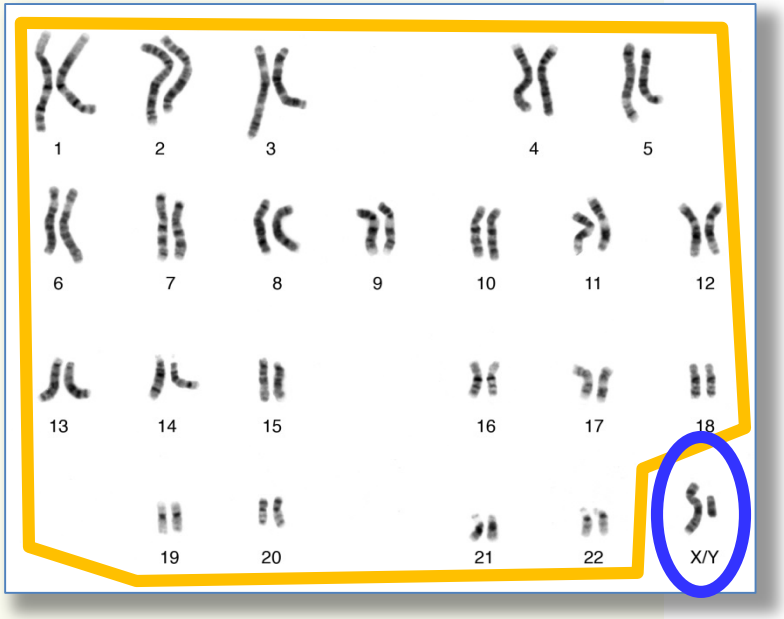
Always contains both benign variation and variation that can cause or contribute to disease(s)

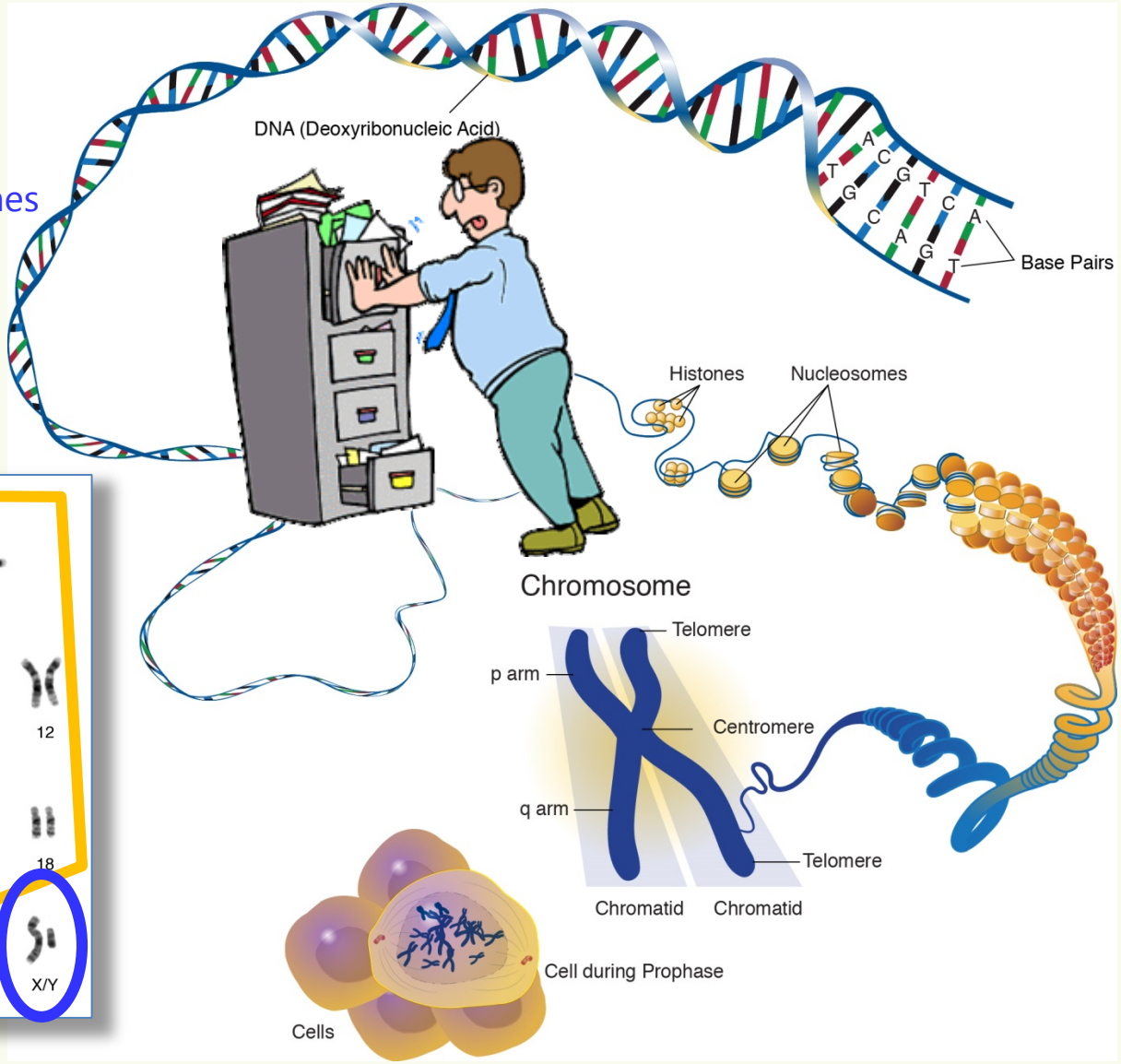
Its big! 3,300,000,000 base pairs (x2)

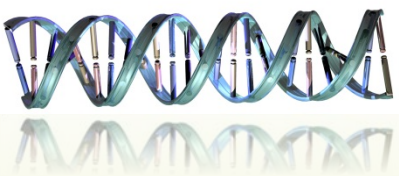


Genome Structure: Chromosomes

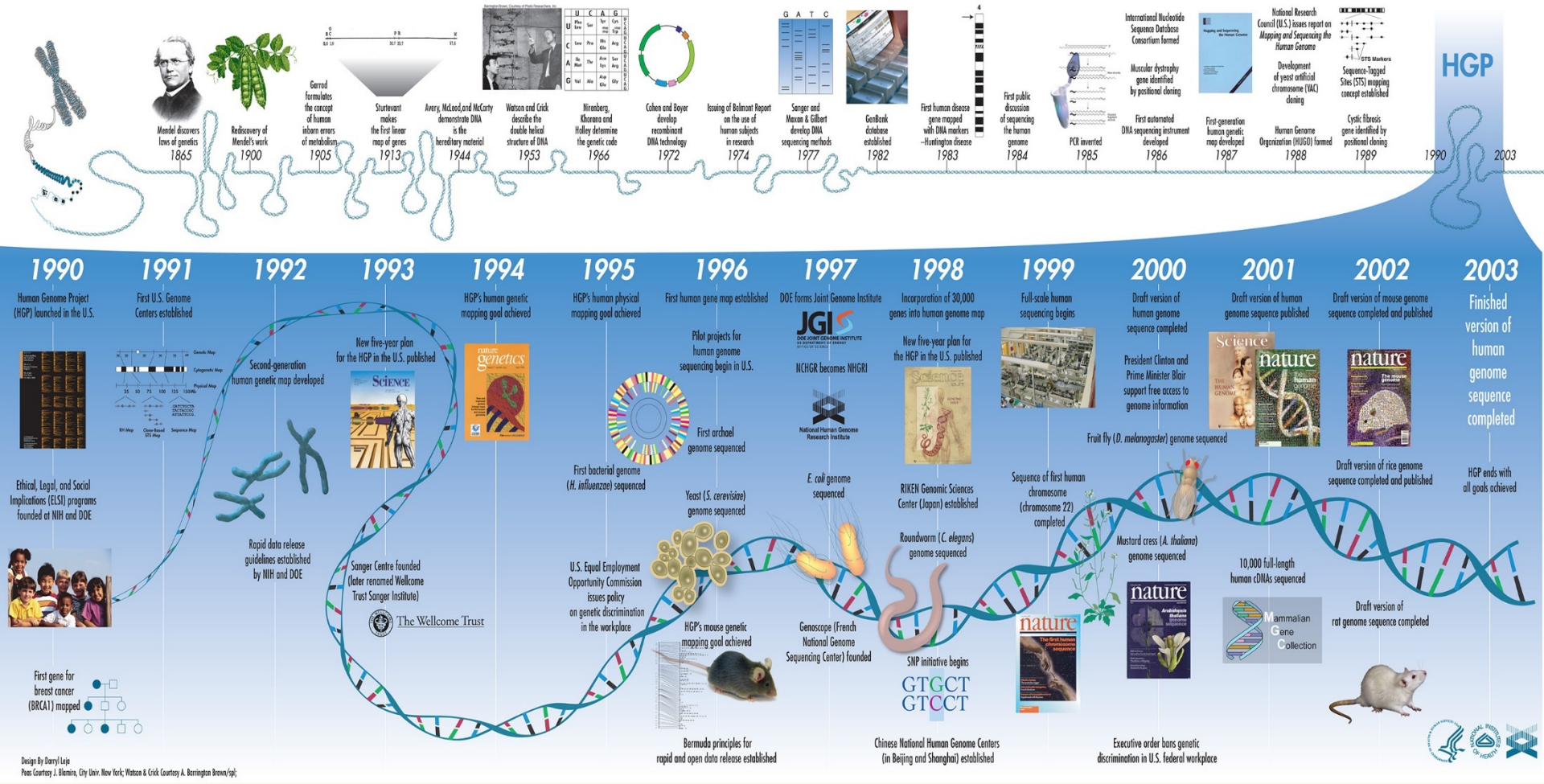
- 23 pairs (pairs!)
  - 22 pairs of autosomes
  - 1 pair of sex chromosomes
- Packages of DNA
- Consistent structure







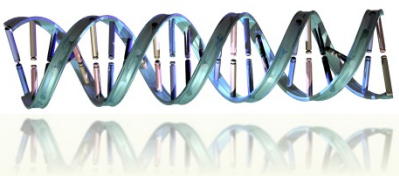
Human Genome Project Timeline



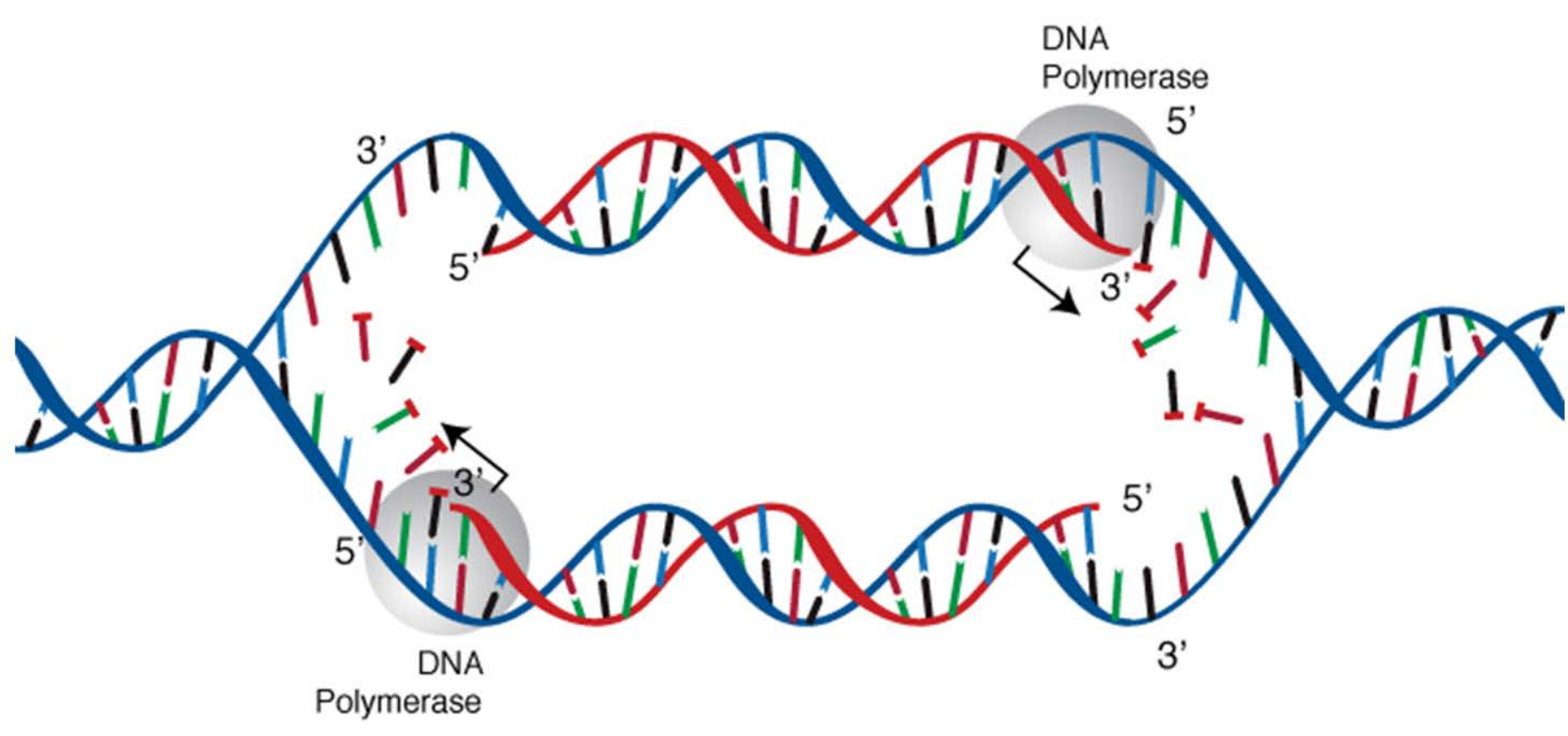
Storing DNA Sequence Information

```
>gi|568815575:c49269727-49250436 Homo sapiens chromosome X, GRCh38.p2 Primary Assembly
CAGAGGGCCAATGGGCAGCCATGCACGGATGACGTACAGCGAATTGAGAGCCGGAGAAGCACGCAGCCAA
TGGGGATGCCCGGGAGTGGGCGGTGTAGGGGGCGGGATGATCGAGGAATCAATGGAGAGGCCGGAATGGG
GGTGGTGCGAAATCCATGATGAAGAAAACAGAAGCAGACAGCTAATGCGGAAGTCTGAGCGAGAGAGGCA
TGTGGGCGGTGGGAAAAGTGGAATGGCTACAGGCGATGAGGGCAGCCGGGAACGAACGCAGAGGCCCAAA
AAGTGGGAGATCCCGCGGGTTGCGTAGGCGGTGGAACGGTAGCTAGAGAGAGAGAACAGCCAATGGAGAG
CCACGCGTGGGATGACGACAATGGCTGCAGCGAGCCAATGGGGAGGCTTGTGAGCGTGCACTGCGGCTTC
GAAGGGGTGGTGCCAAGGTGGAGCTAGCCAATACGAAGGCTAAAAGGAGACCAAGAGGCGGGTGTGTAGG
GAGCCGAGCCAATGGAGGTGGAGGCTGAAGGCGGGAGCGAGGACGAGAGCTGGCGGATAGCCAATGAGGG
AGCCAGGGTGAAGGCGTAGAGACCTCAAAGGACAGTACTAGTATGGGATGAGCCAATGCCGAGGCTAGAT
AGTGCAGCGGGACGGAGGGAGGACTTGGAGGCAAGATGGGGAAAAAAGCAGAGCCGACGGGGGCAGTGTA
GGTGGTGTGAAGGGAGGCATGAGTAGCCAATAAGGAGGCTGGAGAGGAGGACGCTGTGTAGACAGAGTCT
GAGGTATTGGGTGGCCGGACAGCAGAACCATAGAACGCAGGCAGCCATTAGGAAAAATGTGCTTCTCAGG
GATGGGGTTACCCTCTCCCTAACCCTCCGTGACTCCCCATTGCCCCAGGAATACCCAGAGCCCCTGAAGG
CCTAGACCACACCTTGGTGTGTGAACTAGGAAAGTGAAAATGACTGTCCTGTTGCAGTTAGTTACTTACT
GGGTCCTCCCTGCCCAACTACAACCTTTTCTTGAGCAGTGCCACAAGCCAAGAATGCCCCTGATTGTTAG
CCGCAAGCACTCTCCAGGACTTGATACTATTGTTGAATGAATGGCCGAATAGGAAATTTAACTTTCTGAA
GGATTCACTTTTTTAAATTTCAGCTCAGCAAACACTGACCTGACTGTGTGCTACCAGAAGTTGATTGGGC
ATATGGGTAGCTTCTGACTAATGGAATCACAATCAAAGCCCTGTCCTTCTTGCCCAACCTGACACTCCAC
ACCCCACAGGGGCCTCACCTCCCTTCTTTCTGTCCCCCCATCTCTGCTTCTACTGTTCCCTCTCCCAAGA
GTGCTGTTCCTTCTGCTGCTTCCATCTAGTAAACTCCTTTCGCTCTCCCTTTGGGGTCTCACCTTCTGTG
GGAACTCAACTGCAGCTTTTGTCCGCTGAACACATCCATTCATACATTCCTTCTCCATCCTGGGAAGATT
TACTGAACACCTACTGAGTGCCAGGTTCTGAGCATGCTGTTATCTCATTAGAACCTCACAGCATCCTAGG
AGGGAGGTGATCCTTCCAGCAGCCCCTCTCCCCAGGTCTCCATGTTCAGGATGTCCATGAGTTAGAAATG
AATATGTTCAGCTGCAAGTAACAGAAAAATCATAAAATAATGGGGGCTTAAGCAAGACAGAACTTTATTT
TTTTCTCACCTAAAAGTTCAAGCTGGCAAATAGTCCAGGGACTCCTTCCTGCCCACCTGGTTTGGGCATT
GTCTGTCCCGATCTGGCTCCAAGACTCTCAGTCTTCAACAAATCCTTTTGGCTGTCTATTAAAAATATTT
ACCAAATCCATCCTCTTCTCACCACCTCCACTGCCAGCACTATCACCAGTCACCTAGACTACTGCAGTAG
CCAGCTCCTTGCTGGTCCCCTGCTTCCACATTCAGGCCCTAGACCCCATTGTACACTCAGCATTCAGGGG
CTTCCTGTTAAATCCTAACACGGTTTGTGTTTCTCCTCTGCTTGCAACCCTGTCATGCCTTCCTCCTCAC
TCAGCGTTGAAGCCAAAGTCCTTACCATGGCCTACAAGGTCCTGCGTGGTCTAACCTCCGAGTCCCCTTT
CTGACCTCACCTCCAGCTTCTCTTCTTGCTCTCTCCACACTAACCACACCAGCCCCCTTGTAGACCTTGA
ACAAATCAGGTACATTCCCACCTCAGGACCTTTGCCCTGGCTGTTTCCTTTGGAGTTCCTCAGATATCTA
GCTGGCTCCCTCCCTCATCTCCTTCAAGTCTTTGTCAGAGGCCCCTGTCTCTGGGGACTTTTCCTGACCC
GTTTAATCTTGCCTACTCACCCCCAACTCCCAGCCCTCCCAGTCTCCTTCCCGAGAAGCCCTCCATGAGA
GAGGATATAGTCCAGTATAGAGGGGAGGTAGCGACCTCTACCCCTCTCTGCCCTCCTGACCCCAGCCCCC
ATTCTGCAAGGATCTTGGCTTTCAGCAGCCGTAGCATTTCCCTCTCACACCCTGGGCCTGCCCAAGGGGC
CAGCTCACATTCTCGGCCACCATCCTCACTTCGTTTGCCTCCTATATCTCTGGCCACTCTCTTTCATGGA
GTGAGTGACAGTAAGGGACAGATGGTCTGGGGAAGGTATAAGGAGAATGCAGCCCACACCTCTGATCCTG
```

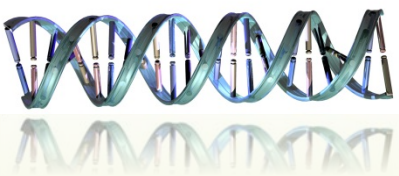
Genome data storage size discussion (Reid Robison):
https://medium.com/precision-medicine/how-big-is-the-human-genome-e90caa3409b0#.9p7kf0k9g



DNA Replication



- DNA Polymerase + Nucleotides + Template
  - 5'→ 3' growth
- Information preserved

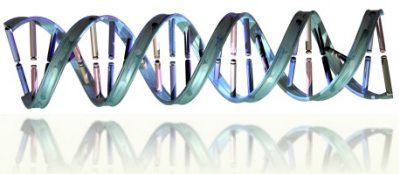


DNA Replication Video

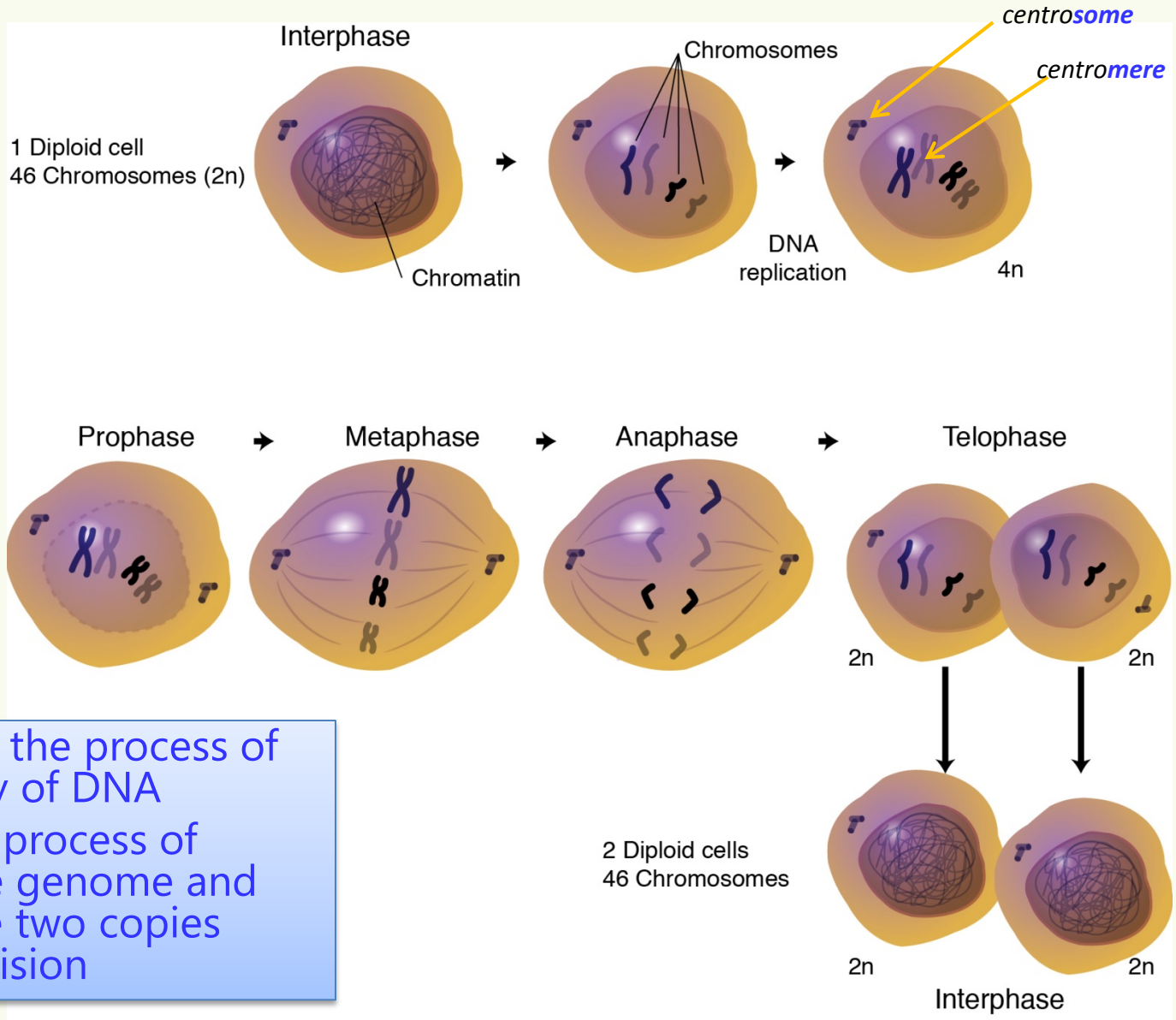
https://www.hhmi.org/biointeractive/dna-replication-basic-detail

How do we use DNA Replication?

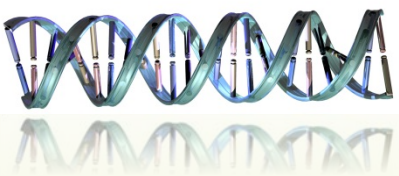
- Tech!
- Basis for
  - most DNA sequencing technologies
  - PCR DNA amplification
  - Mutation Detection techniques
  - DNA Diagnostics
  - Reconstructing ancient genomes



Replication vs. Mitosis



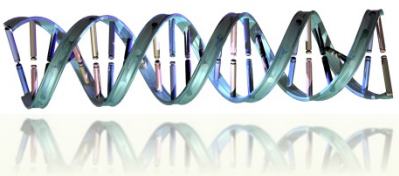
- **Replication** is the process of making a copy of DNA
- **Mitosis** is the process of replicating the genome and separating the two copies during cell division



Replication Errors

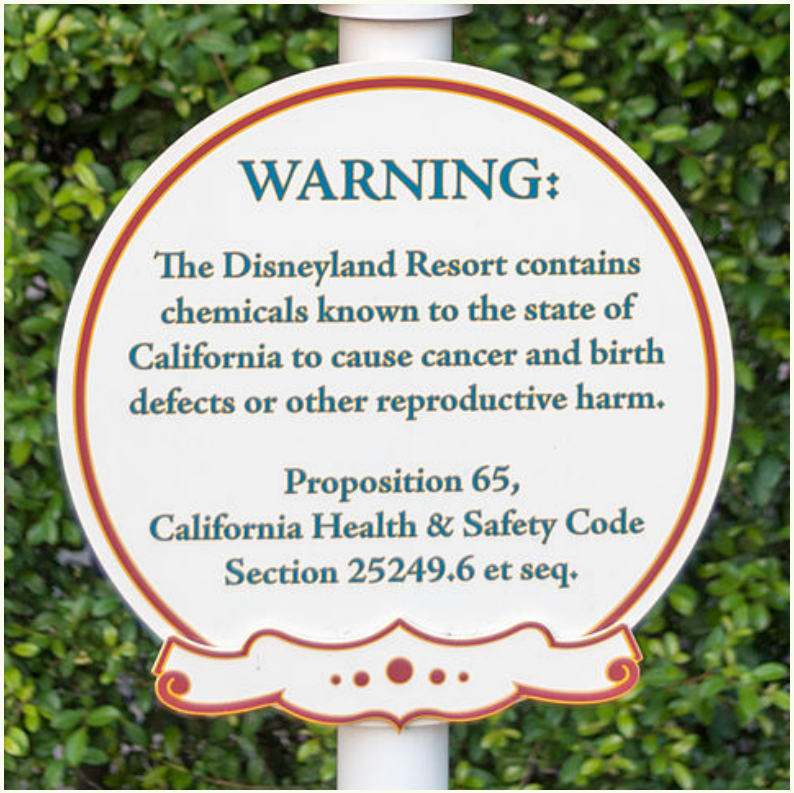
- Proofreading

- Nobody's perfect!



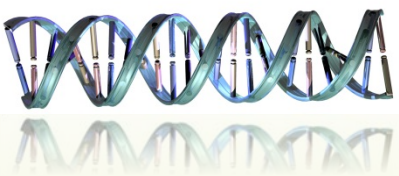
Other fidelity problems

- Chemicals (mutagens)
- Ultraviolet light
- Ionizing radiation
- DNA Repeat instability
  - e.g. triplet repeats



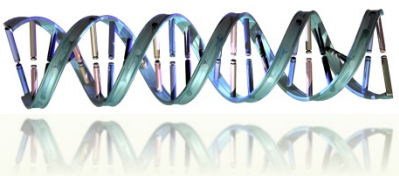
www.Wikipedia.org

The same but different



Variation

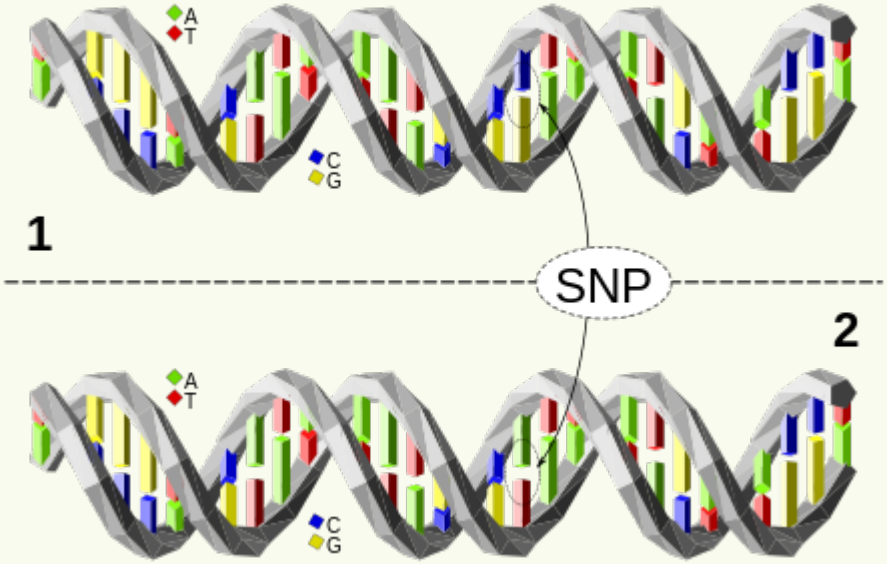
- "It is found to vary"
- Is defined *by reference to a "standard,"* which may or may not be "normal"
- Occurs normally without regard to functional consequence
- Is subject to selective pressures
- Originates in one member of population. Some descendants will inherit that and pass it on. Results in frequency differences in different population lineages

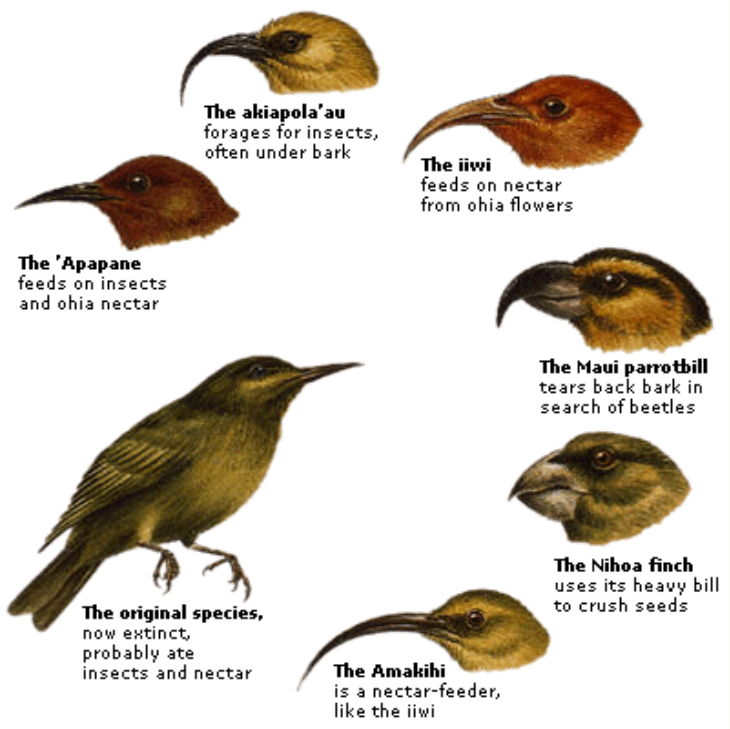


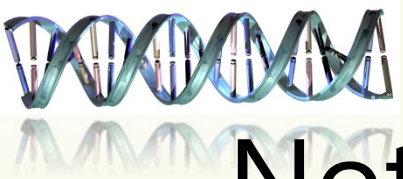
Variation

- Consequences

None	Worse
Better	Environment-dependent
Evolution	Phenotype variation
Disease	







Not Synonyms of "Variation"

Mutation

- The molecular and chemical processes that result in new variation
- Otherwise avoid!

Polymorphism

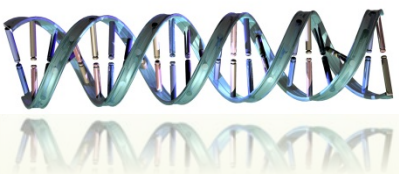
- Variation that exists in "the (normal) population" at 1% frequency or higher
- Assumed (not proven) to have no disease significance
- E.g. Single Nucleotide Polymorphism (SNP)

***de novo* (New)**

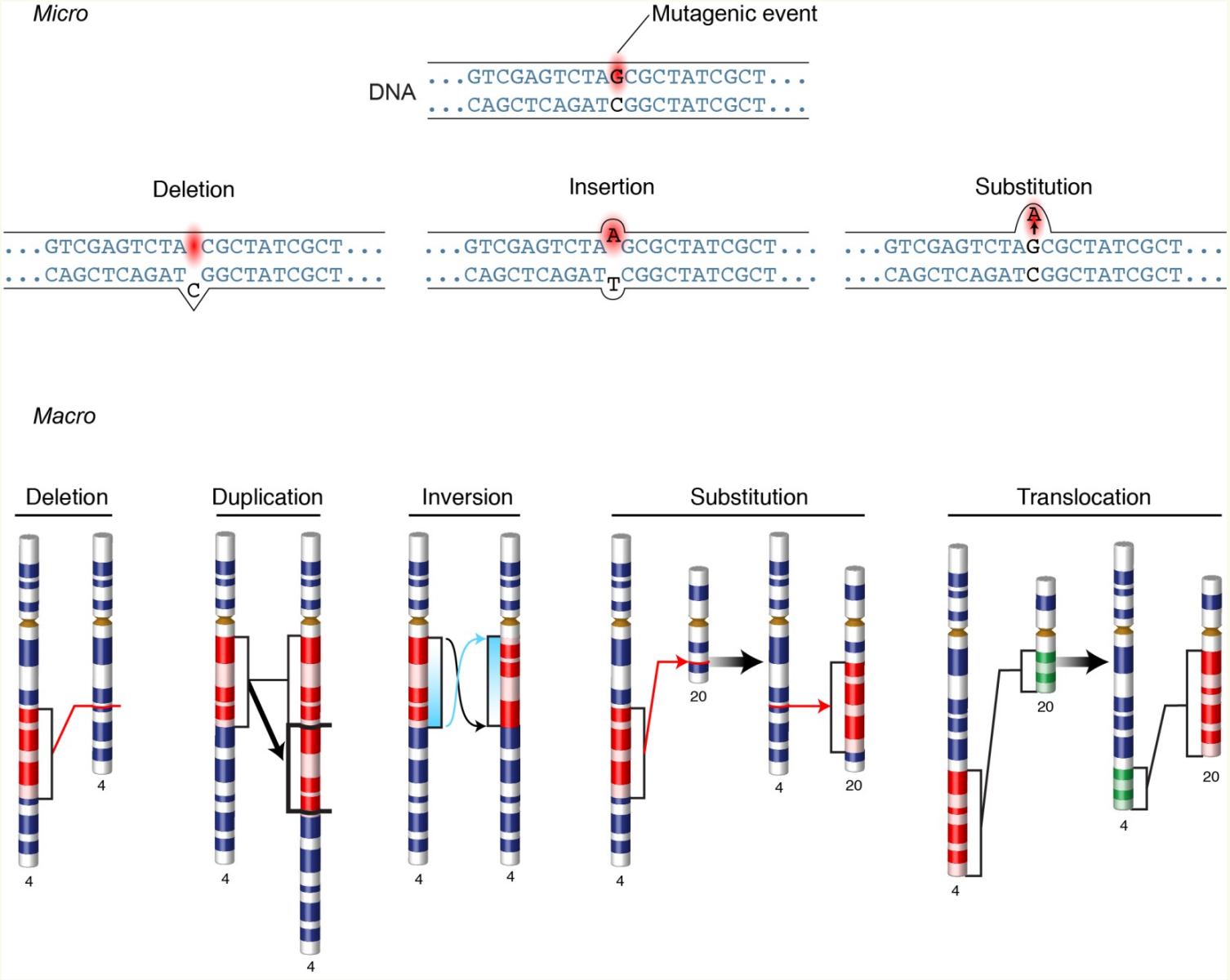
- new variation (from mutation)
- Altered between the last generation and this one

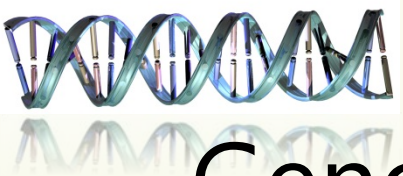
Marker

- Variation that is used to trace inheritance of DNA segments, or suggest linkage to disease genes
- E.g. Genome-Wide Association Studies (GWAS)
- E.g Linkage analysis



Mutation


Micro
Mutagenic event
DNA
...GTCGAGTCTAGCGCTATCGCT...
...CAGCTCAGATCGGCTATCGCT...
Deletion
...GTCGAGTCTA CGCTATCGCT...
...CAGCTCAGAT GGCTATCGCT...
C
Insertion
...GTCGAGTCTA A GCGCTATCGCT...
...CAGCTCAGAT T CGGCTATCGCT...
Substitution
A
...GTCGAGTCTAGCGCTATCGCT...
...CAGCTCAGATCGGCTATCGCT...
Macro
Deletion
Duplication
Inversion
Substitution
Translocation
4
20



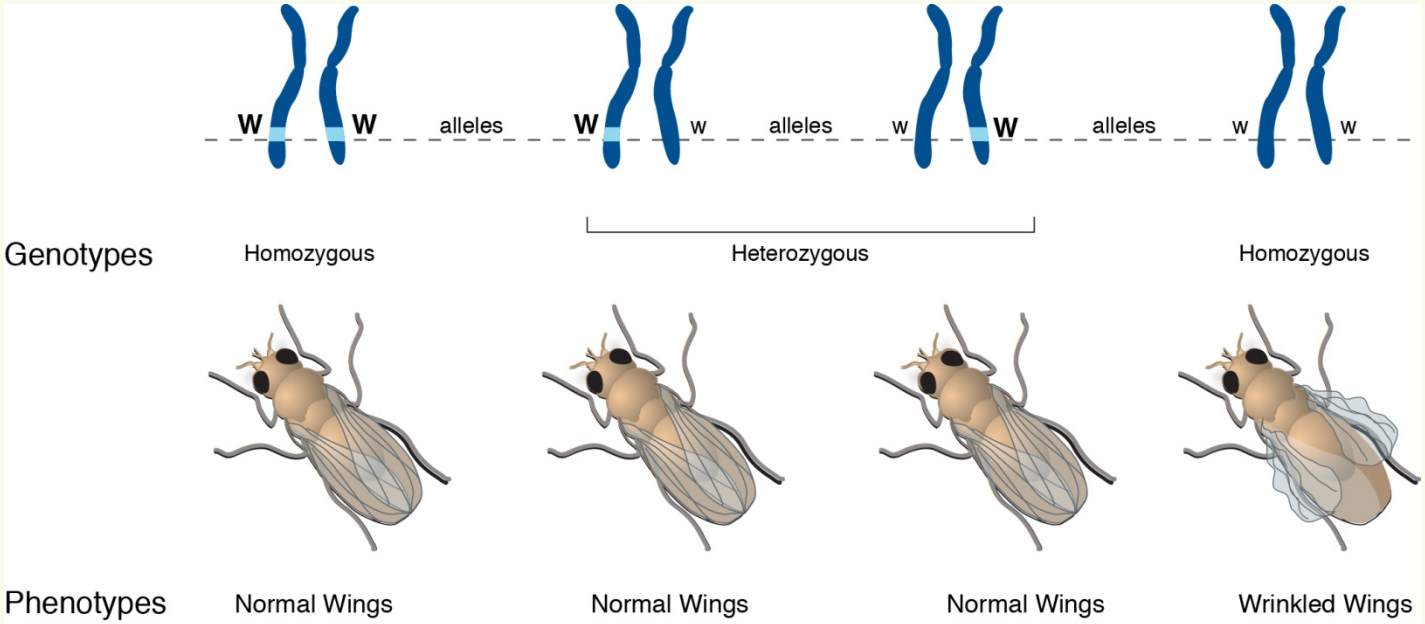
Genotype "codes for" Phenotype

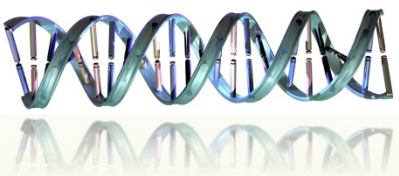
Genotype

- The *genetic code* describing an individual

Phenotype

- The *physical manifestations* of genotype in an individual

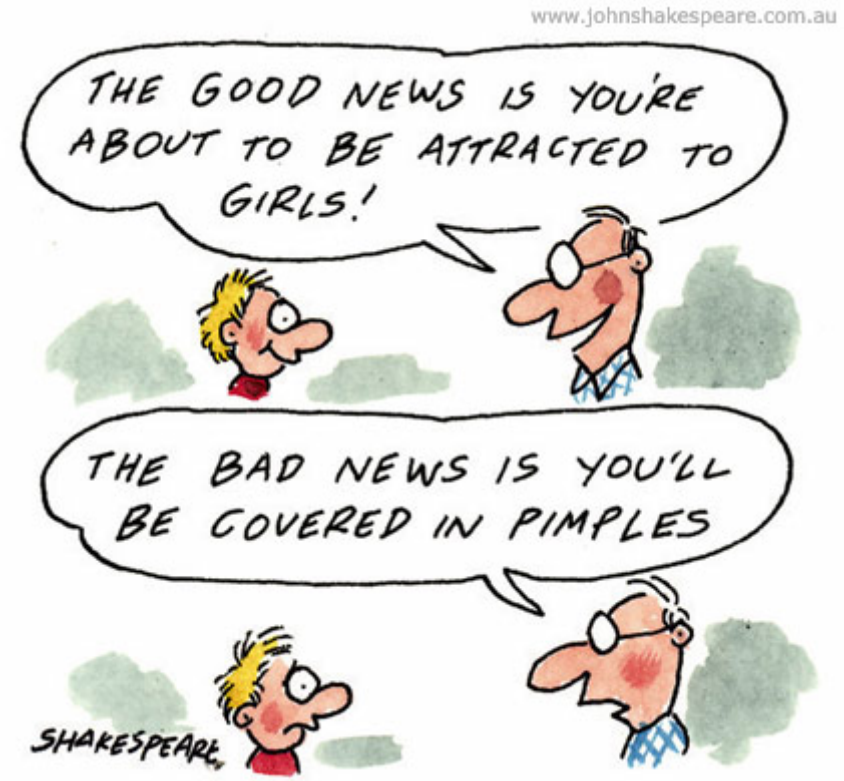


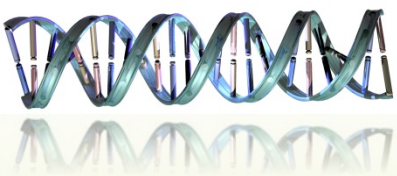


Stuff Happens

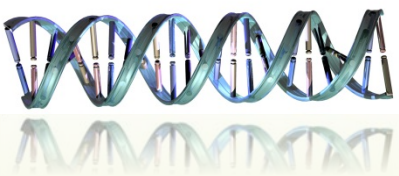
- Variation
  - Is essential
  - Is risky
  - Is relevant
  - Is relentless

… Is like puberty…





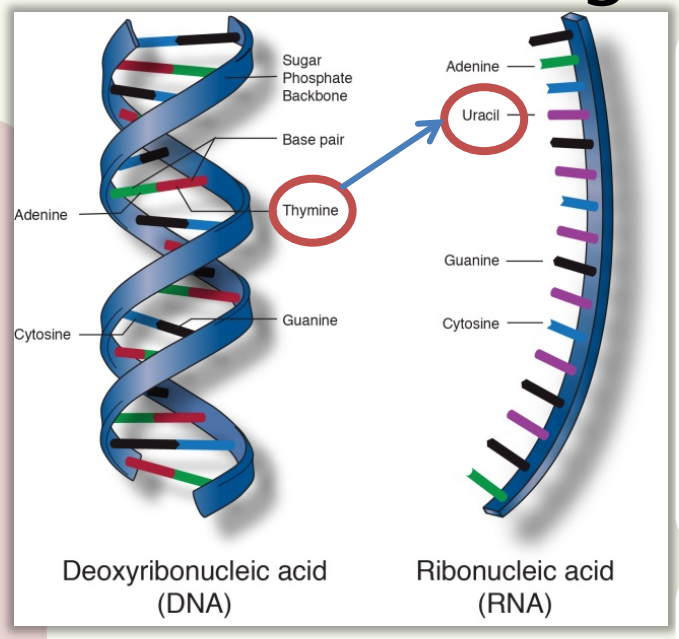
https://en.wikipedia.org/wiki/File:Gene_Wilder_1970.JPG

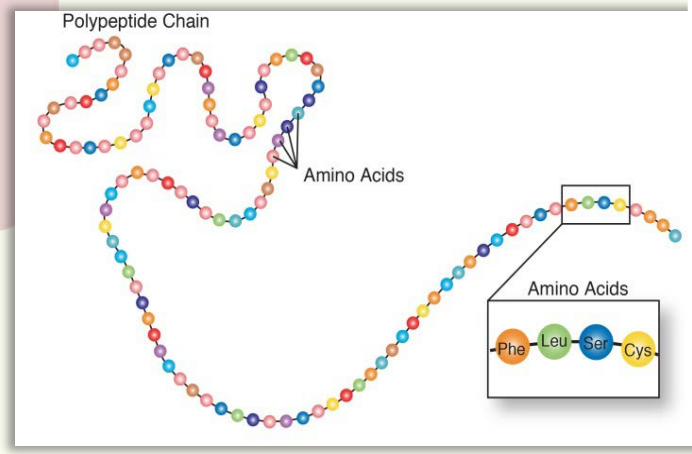


Basics of the Information-containing Molecules

DNA

- *Deoxy*nucleotide units ("bases")
- Double-stranded
- Bi-directional
  - 3' and 5' ends
  - Strands run anti-parallel
- 4 possible bases A,C,G,**T**
- "base pairs"
  - Complementary
  - A pairs with T
  - C pairs with G



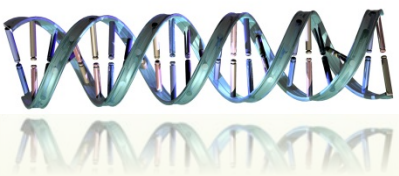


RNA

- Nucleotide units
- Single-stranded
- 4 possible bases A,C,G,**U**
- Multiple uses
- Messenger (mRNA)
  - Template for "complementary DNA" = cDNA

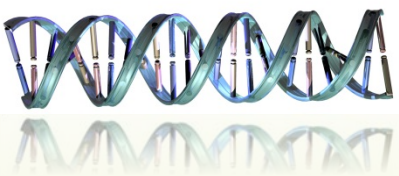
Protein

- Strand of amino acids
- 20 possible amino acids
- Sequence encoded by three-base (triplet) codons in DNA/mRNA

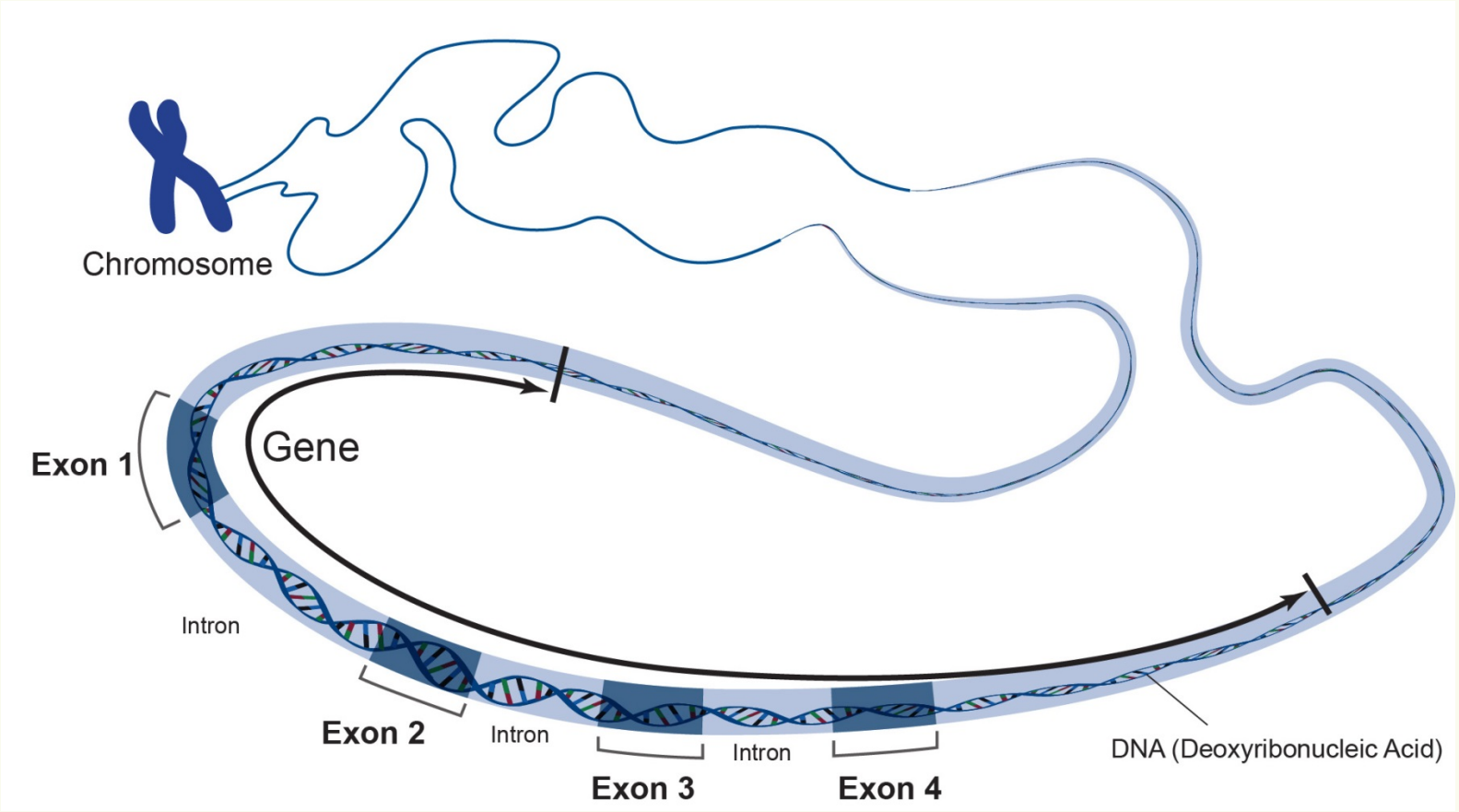


Gene

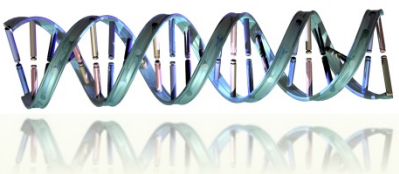
- Definition
  - A “thing”, which when inherited, corresponds to a trait
  - Encodes a functional unit, from a single transcription unit
  - *Carlos Bustamante Video*
- What do genes “do”?
  - Code/blueprint/detailed plan/orchestra
  - Regulation
  - Transcription/RNA
  - Translation/protein
  - Modification and other complex stuff



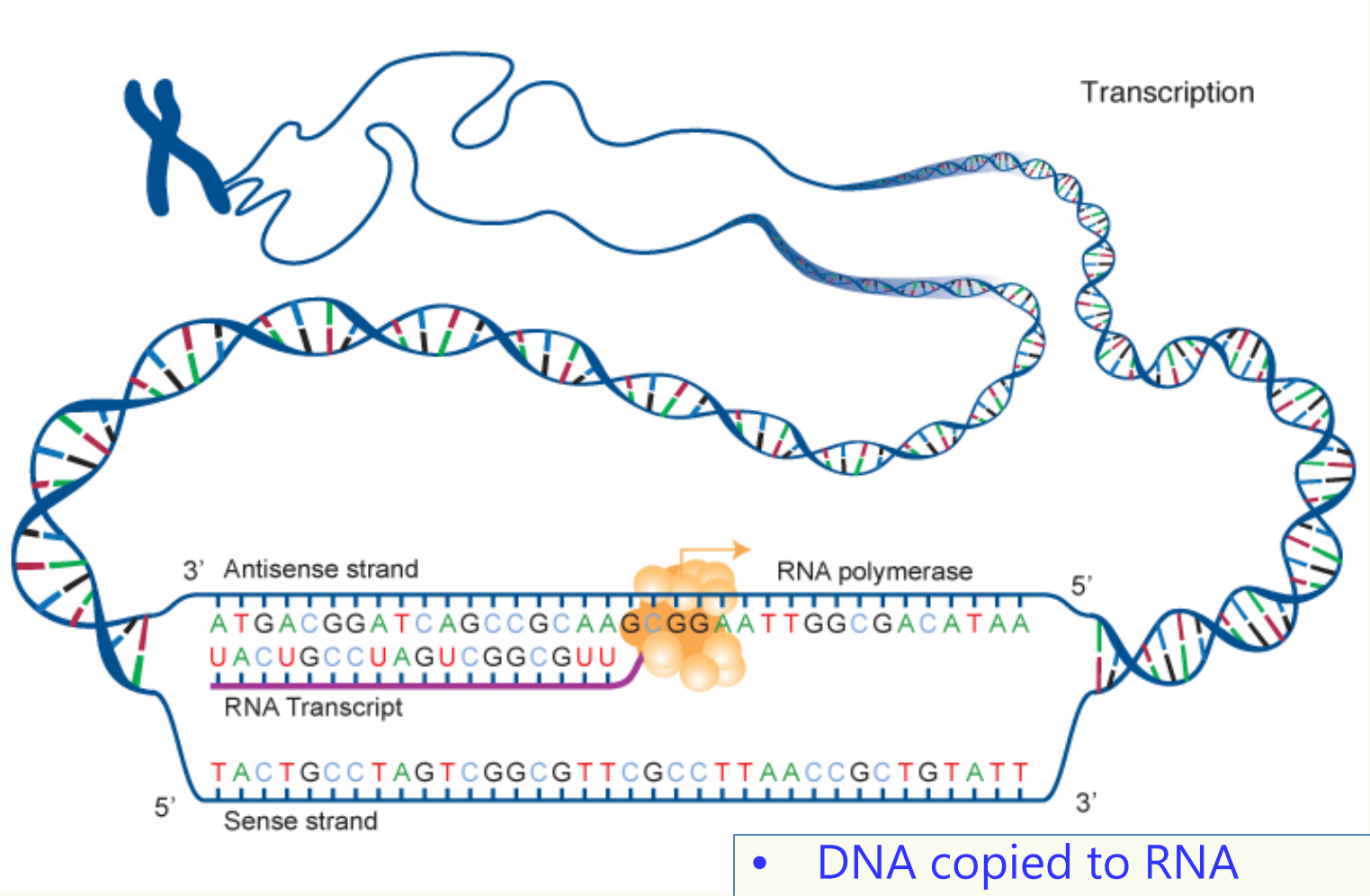
Basic Gene Structure: **Exons** and *Introns*



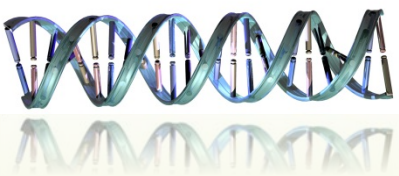
- **Exons** are segments of genes that
  - Make up the messenger RNA
  - Mostly code for proteins
- *Introns* are spacers that get cut out after transcription
- Gene coding regions are about 1% of the genome



Transcription



- DNA copied to RNA
- "Sense" strand
- Then → Introns are spliced out!



The Genetic Code

RNA codon table

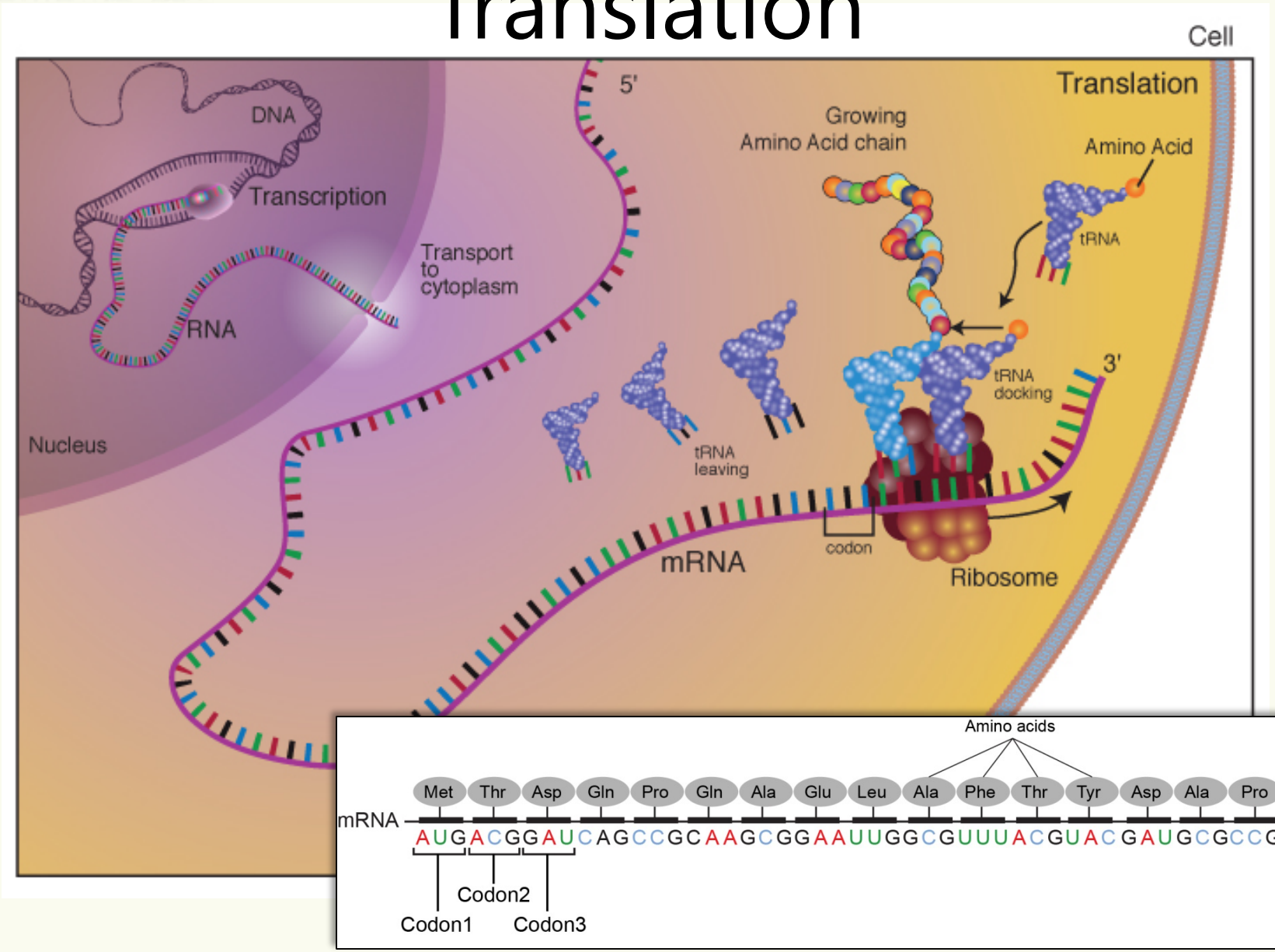
1st position	2nd position: U	2nd position: C	2nd position: A	2nd position: G	3rd position
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	stop	stop	A
	Leu	Ser	stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

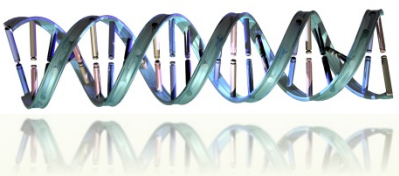
Amino Acids

Ala: Alanine
Arg: Arginine
Asn: Asparagine
Asp:Aspartic acid
Cys:Cysteine
Gln: Glutamine
Glu: Glutamic acid
Gly: Glycine
His: Histidine
Ile: Isoleucine
Leu: Leucine
Lys: Lysine
Met: Methionine
Phe: Phenylalanine
Pro: Proline
Ser: Serine
Thr: Threonine
Trp: Tryptophane
Tyr: Tyrosisne
Val: Valine

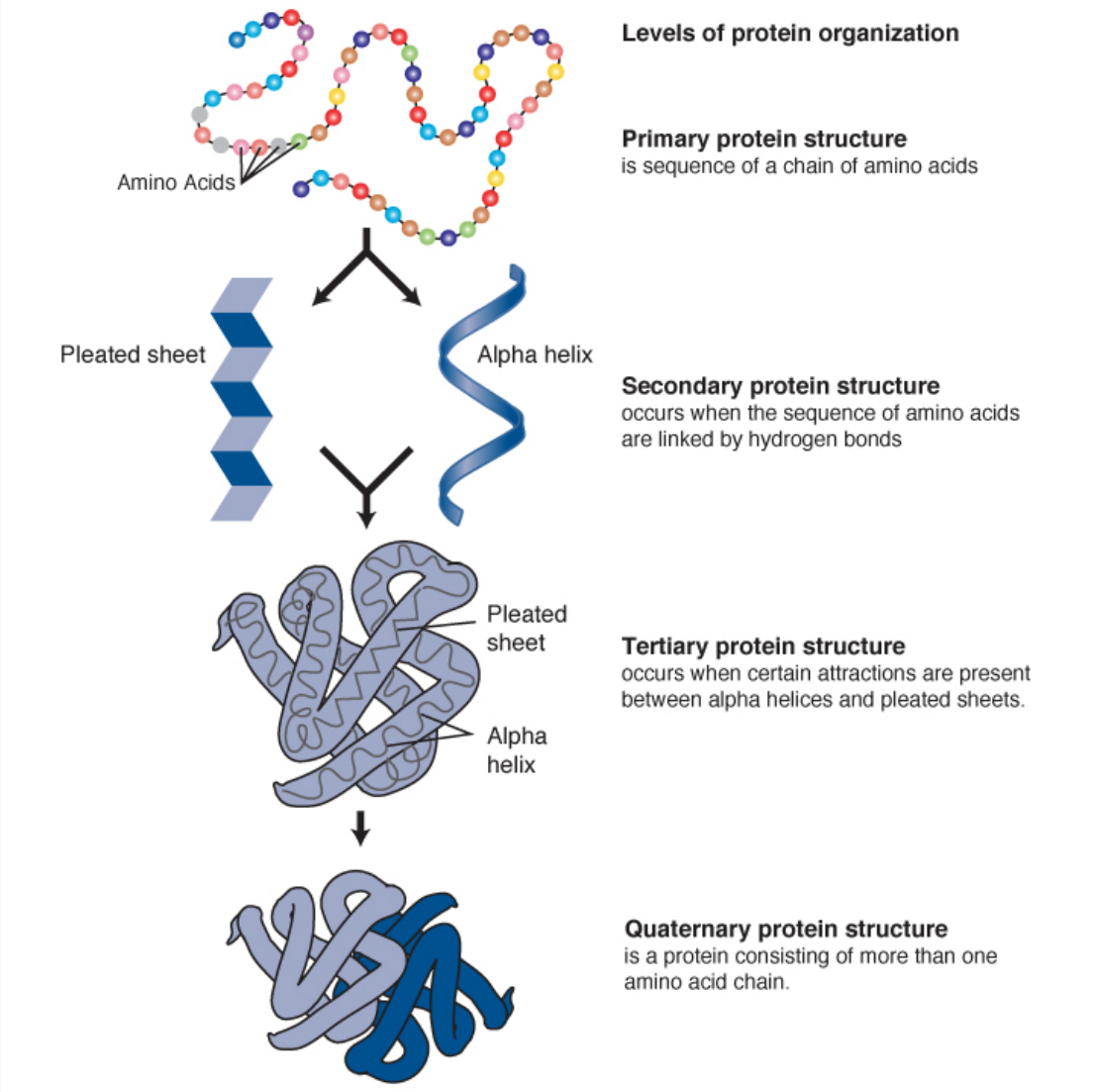
- Universal
- Three bases => 1 amino acid, or termination
- Degenerate
  - some base changes don't result in amino acid changes, they are **synonymous**
- Translation is reading-frame dependent
  - Insert/delete can shift triplet frame → translated differently

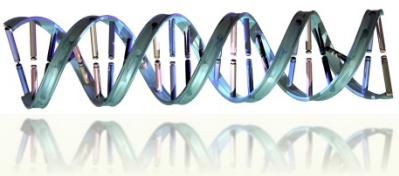
Translation





Proteins

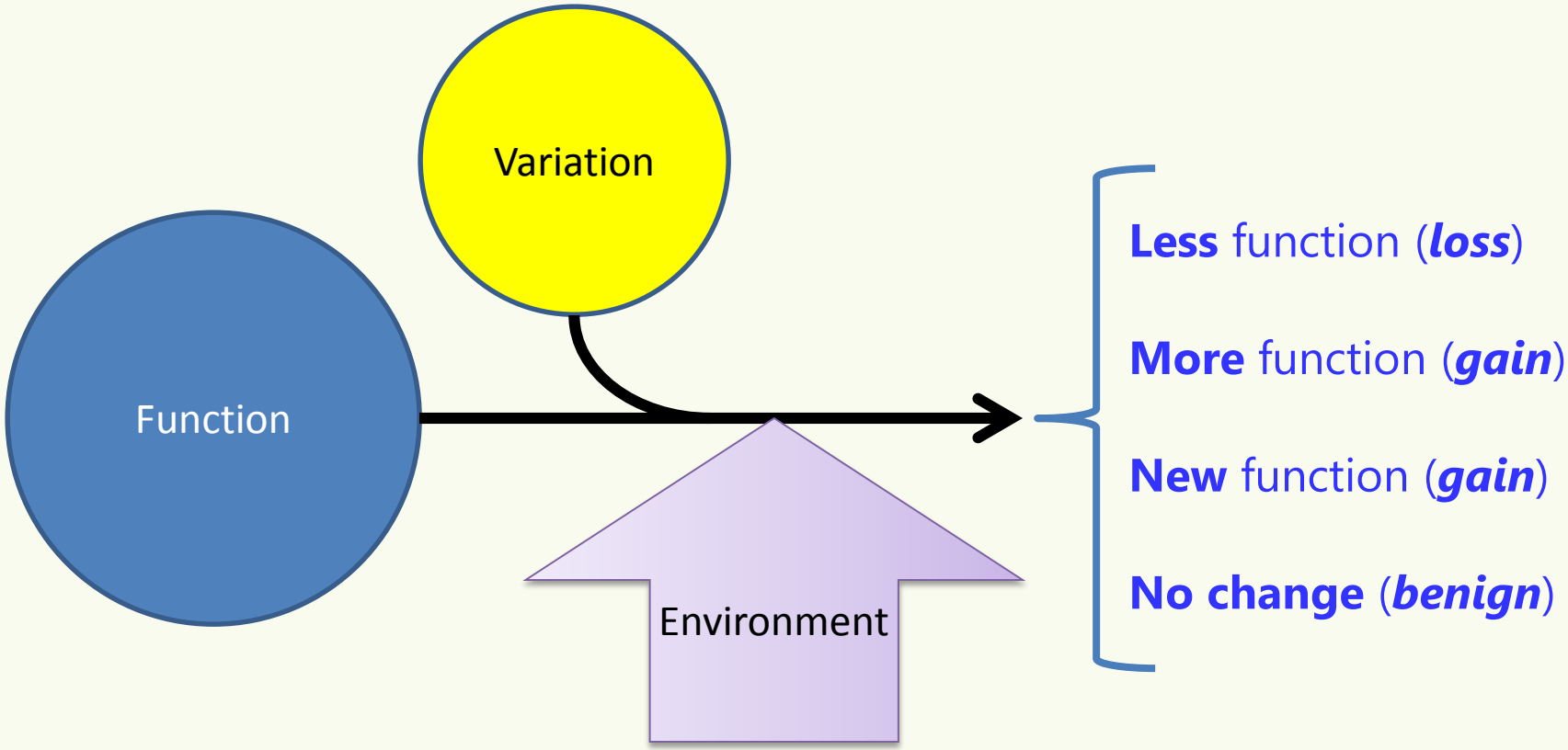

Levels of protein organization
Amino Acids
Primary protein structure
is sequence of a chain of amino acids
Pleated sheet
Alpha helix
Secondary protein structure
occurs when the sequence of amino acids are linked by hydrogen bonds
Pleated sheet
Alpha helix
Tertiary protein structure
occurs when certain attractions are present between alpha helices and pleated sheets.
Quaternary protein structure
is a protein consisting of more than one amino acid chain.

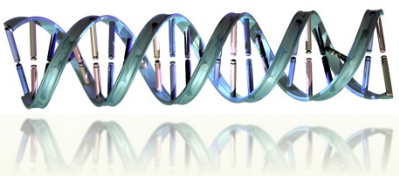


What proteins "do"

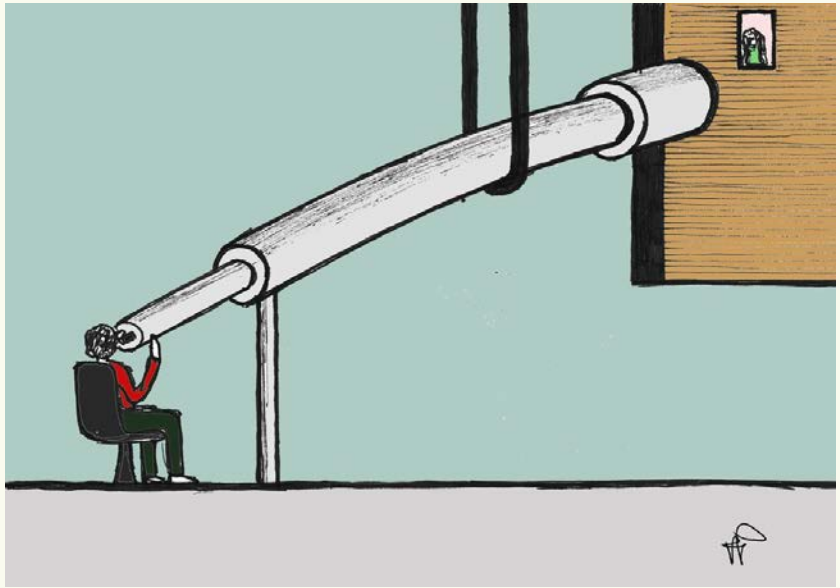
- Hold
- Digest/cut
- Build/convert
- Take up space
- Send messages
- Glue/attach
- Specify/Identify
- Logic
- Respond and regulate
- Contract/expand

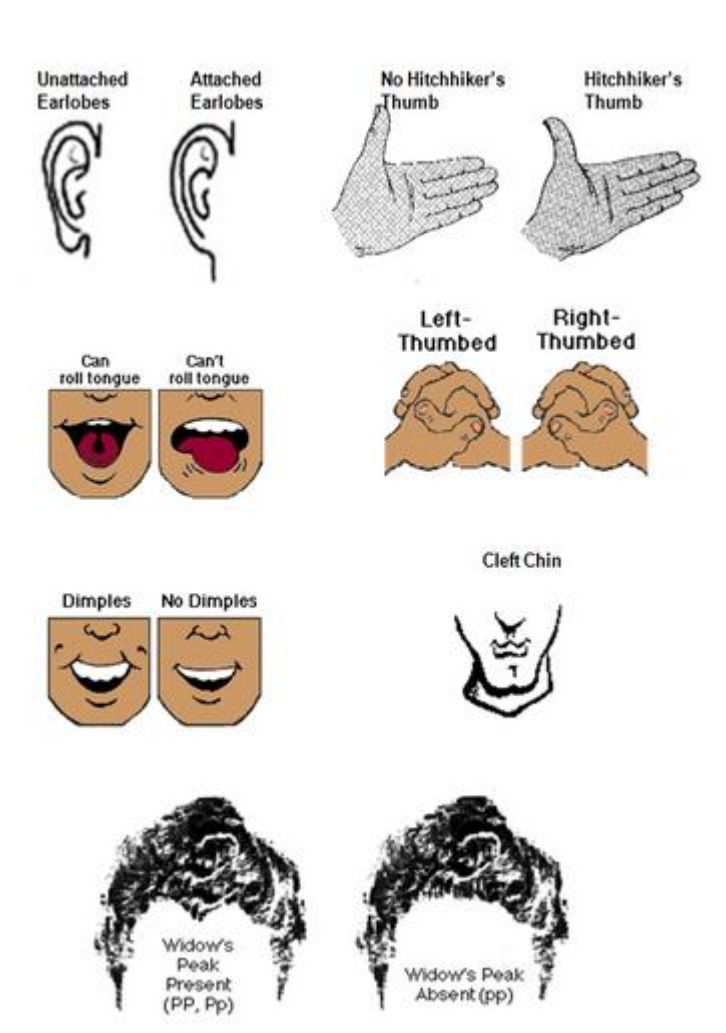
Functional Consequence of Variation

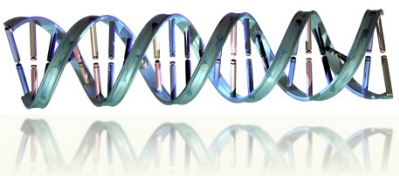




Inheritance

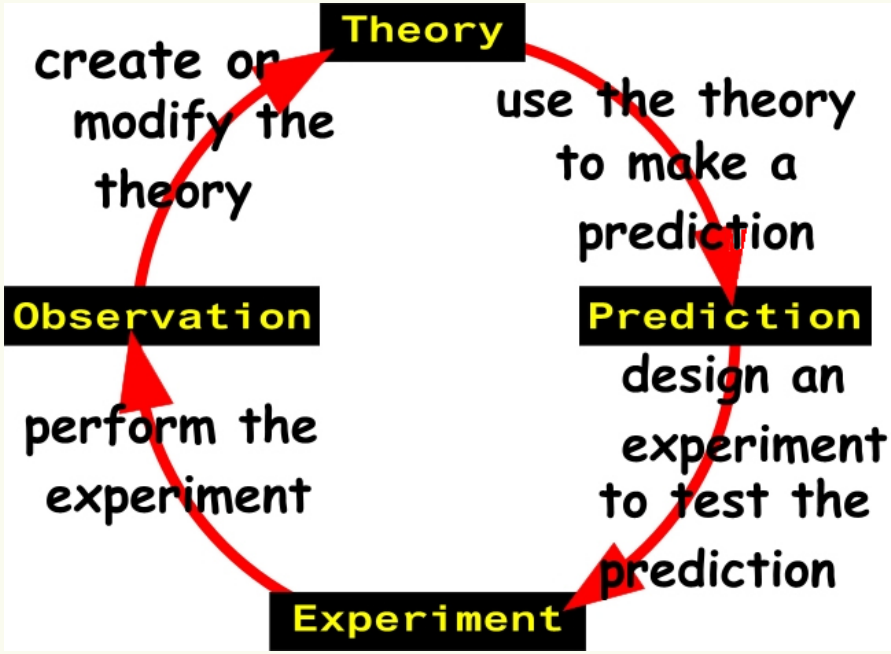


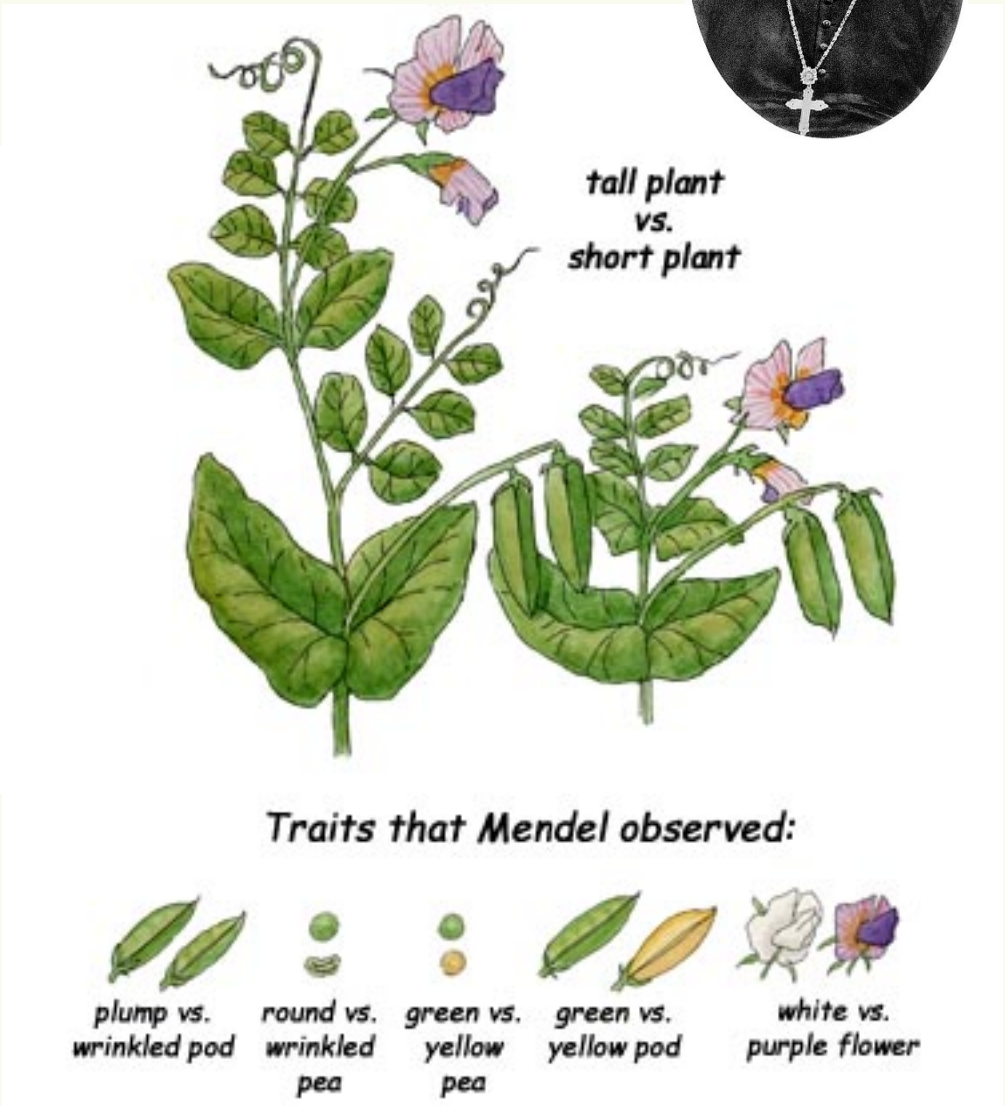




https://en.wikipedia.org/wiki/File:Gregor_Mendel_oval.jpg

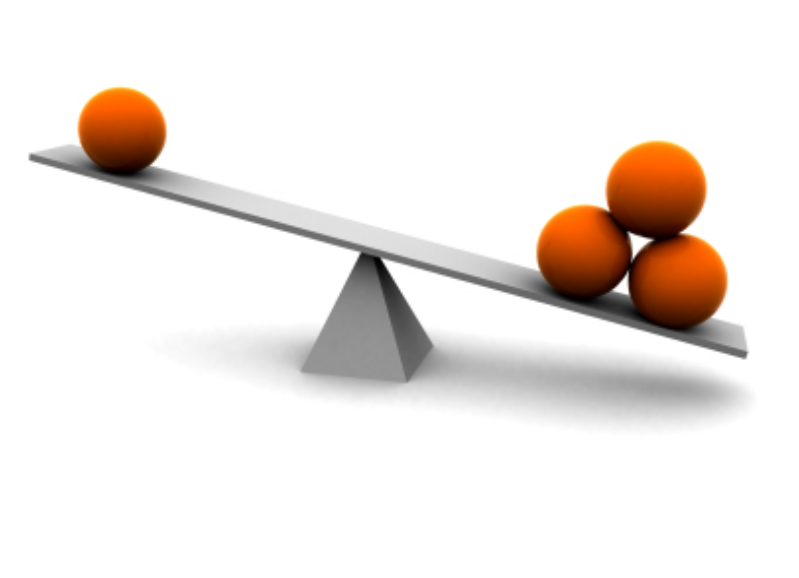
Gregor Mendel

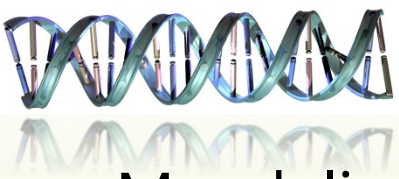




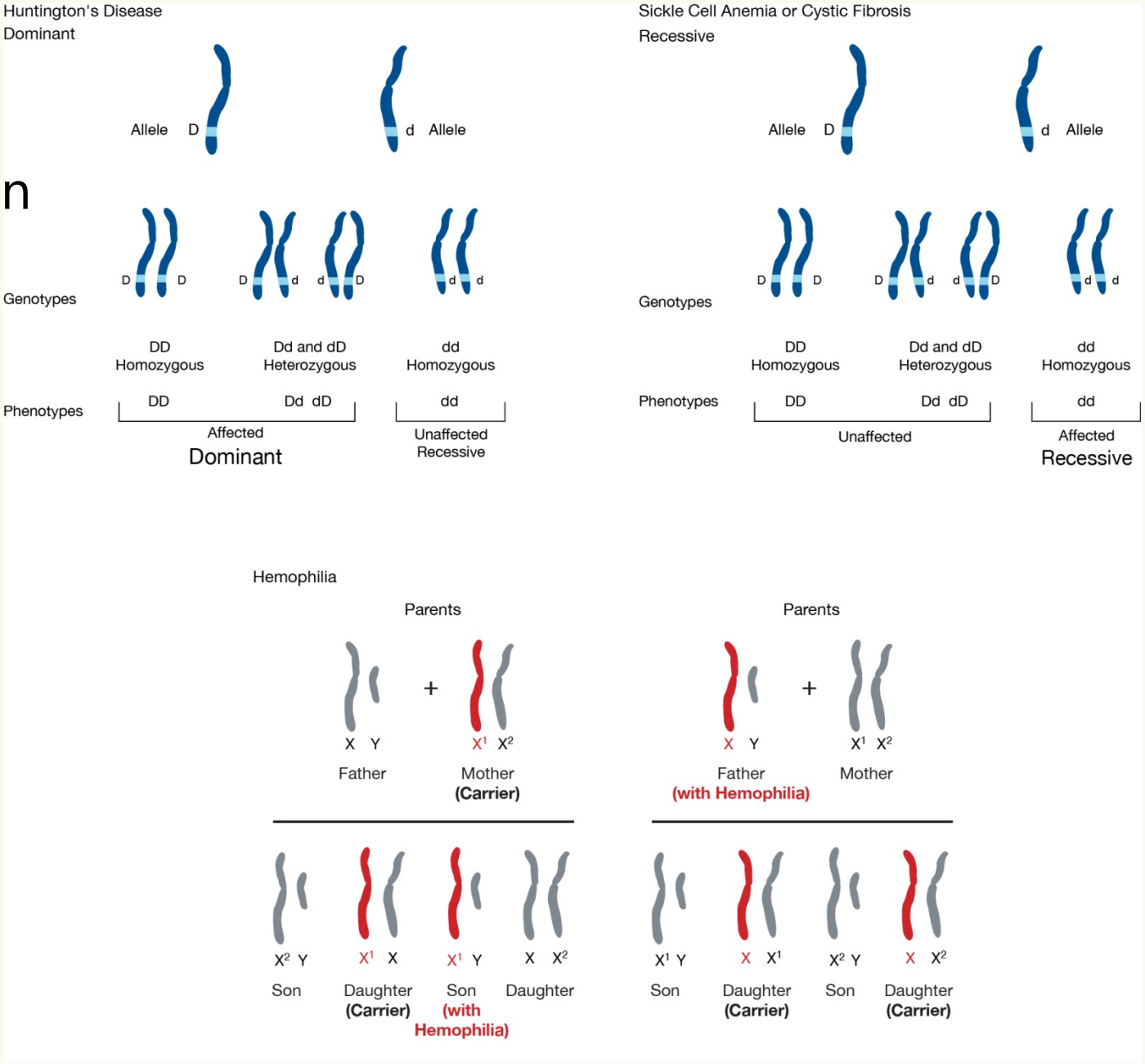
http://tomatosphere.org/teachers/guide/principal-investigation/scientific-method

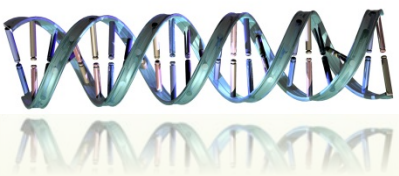
Balance and Imbalance



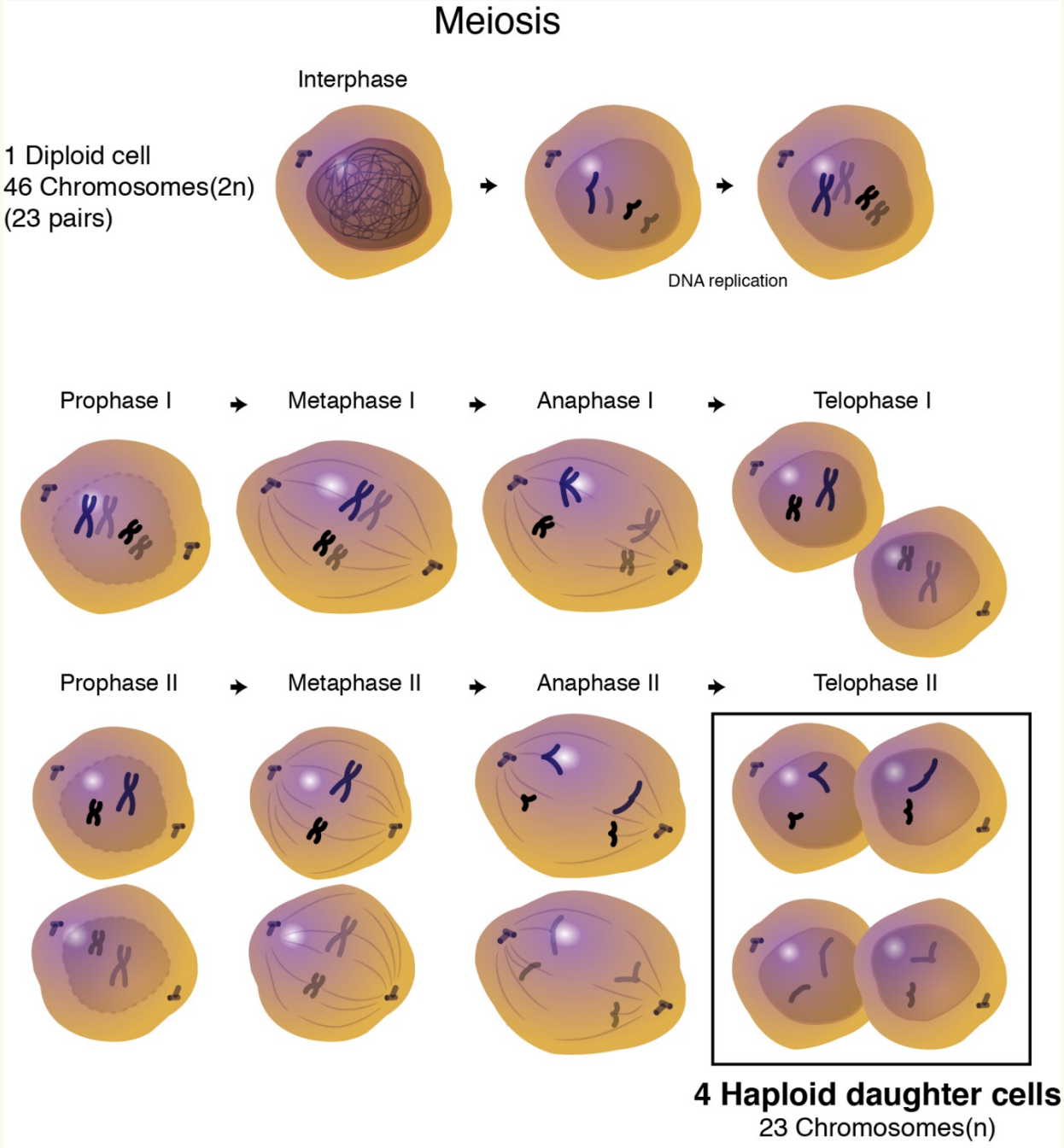


Mendelian Inheritance in Humans





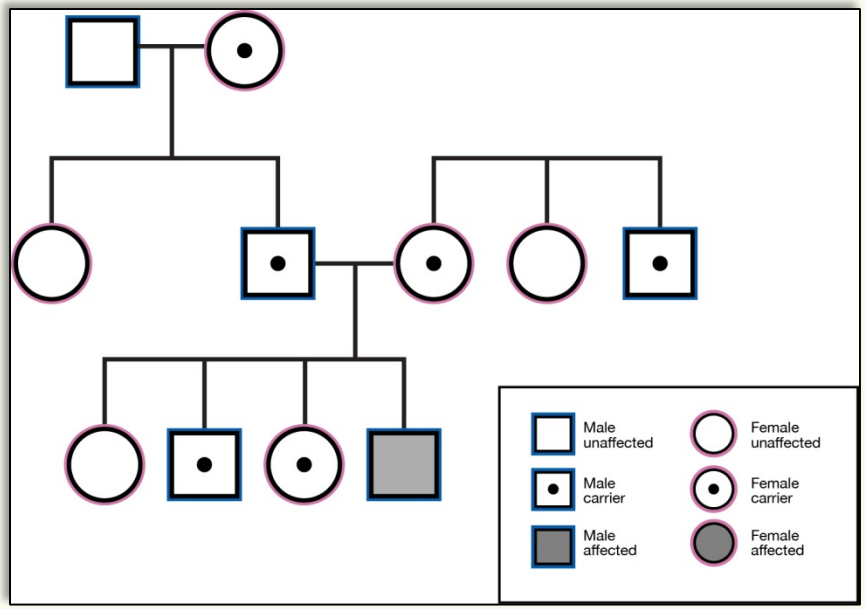
Meiosis

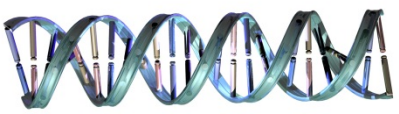

Meiosis
Interphase
1 Diploid cell
46 Chromosomes(2n)
(23 pairs)
DNA replication
Prophase I
Metaphase I
Anaphase I
Telophase I
Prophase II
Metaphase II
Anaphase II
Telophase II
4 Haploid daughter cells
23 Chromosomes(n)

Recessive

Autosomal Recessive Inheritance

Affected	Unaffected	"Carrier"	*Both parents* of an affected *are carriers* (or affected)
• No normal copy	• At least one normal copy	• Unaffected • Transmits 50-50	• An affected parent creates pseudo-dominant inheritance

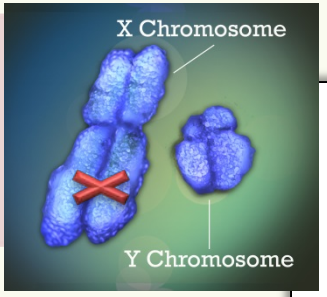




X-linked Recessive Inheritance

Affected

- No normal copy
- Males
- All daughters are carriers
- All sons are unaffected
- Rare females



Unaffected

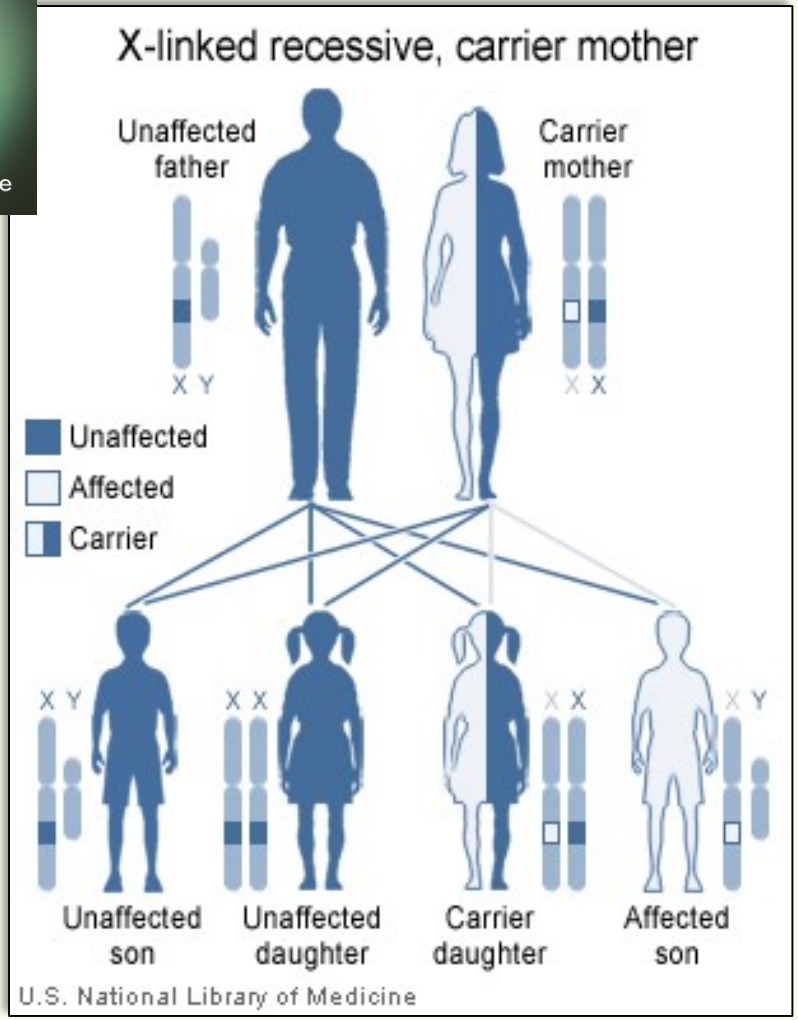
- At least one normal copy
- Non-carrier males
- Most females

"Carrier"

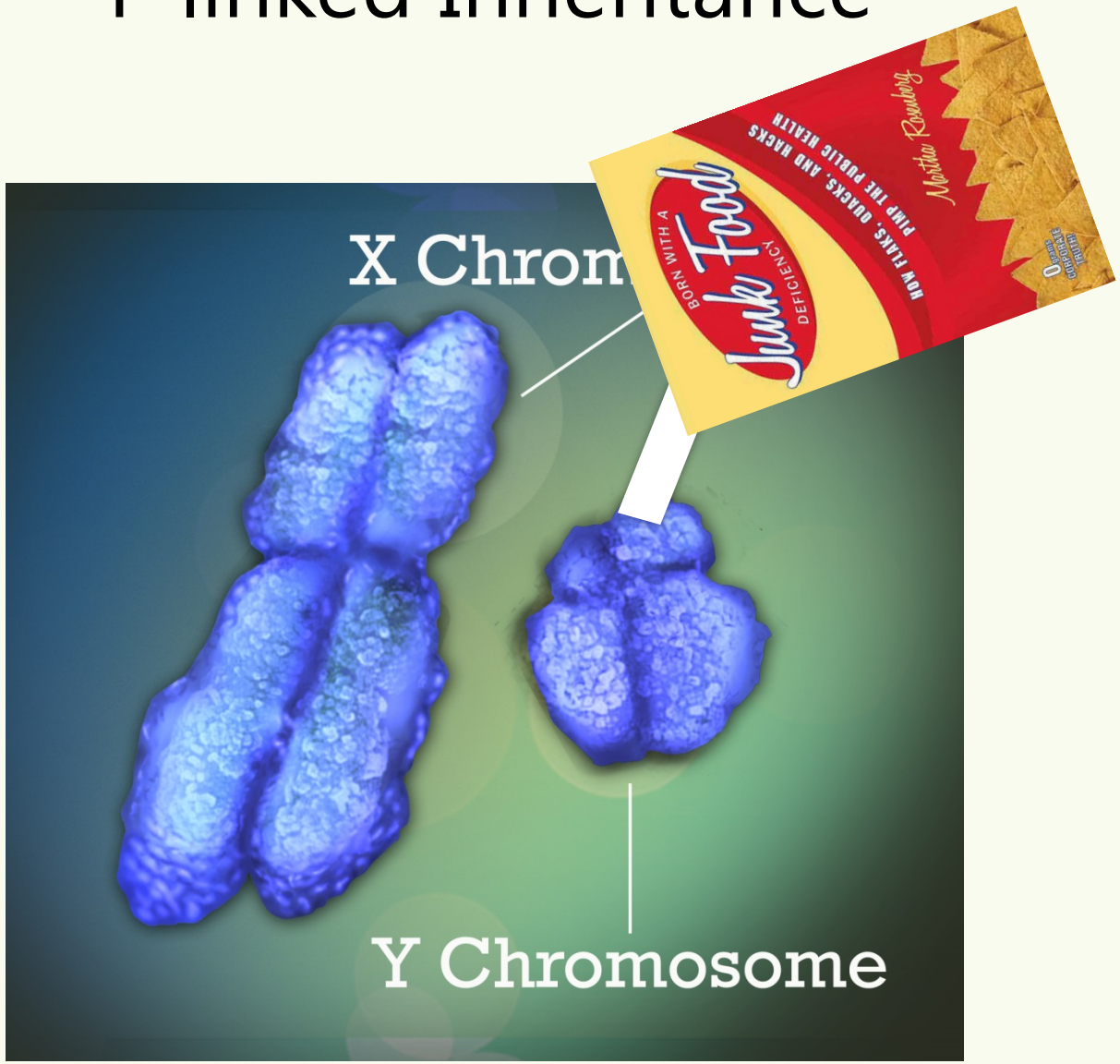
- Unaffected
- Females (and XXY males)
- Transmits 50-50

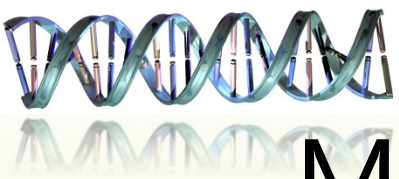
Mother* of an affected *is a carrier

- Always - for benign condition
- 2 out of 3 - when affected males can't reproduce
  - 1 out of 3 is de novo



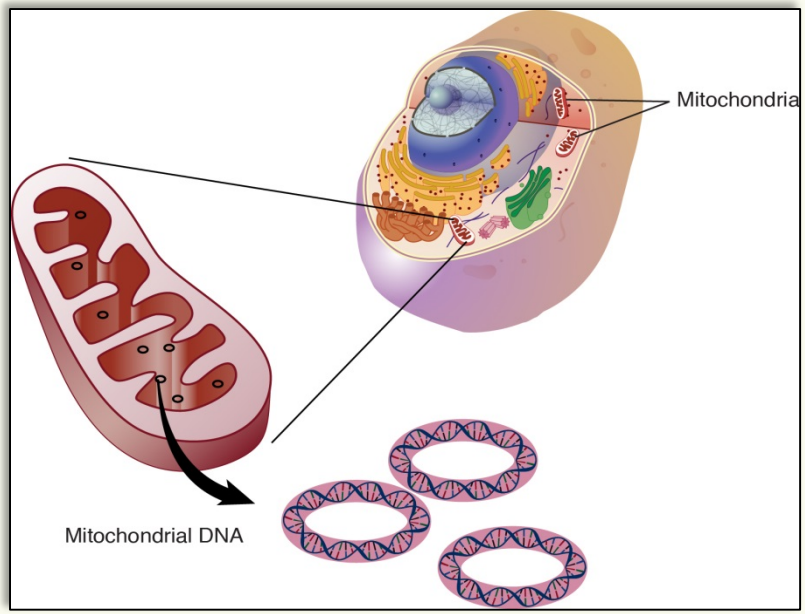
Y-linked Inheritance

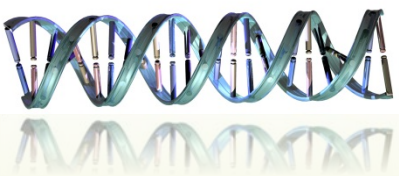




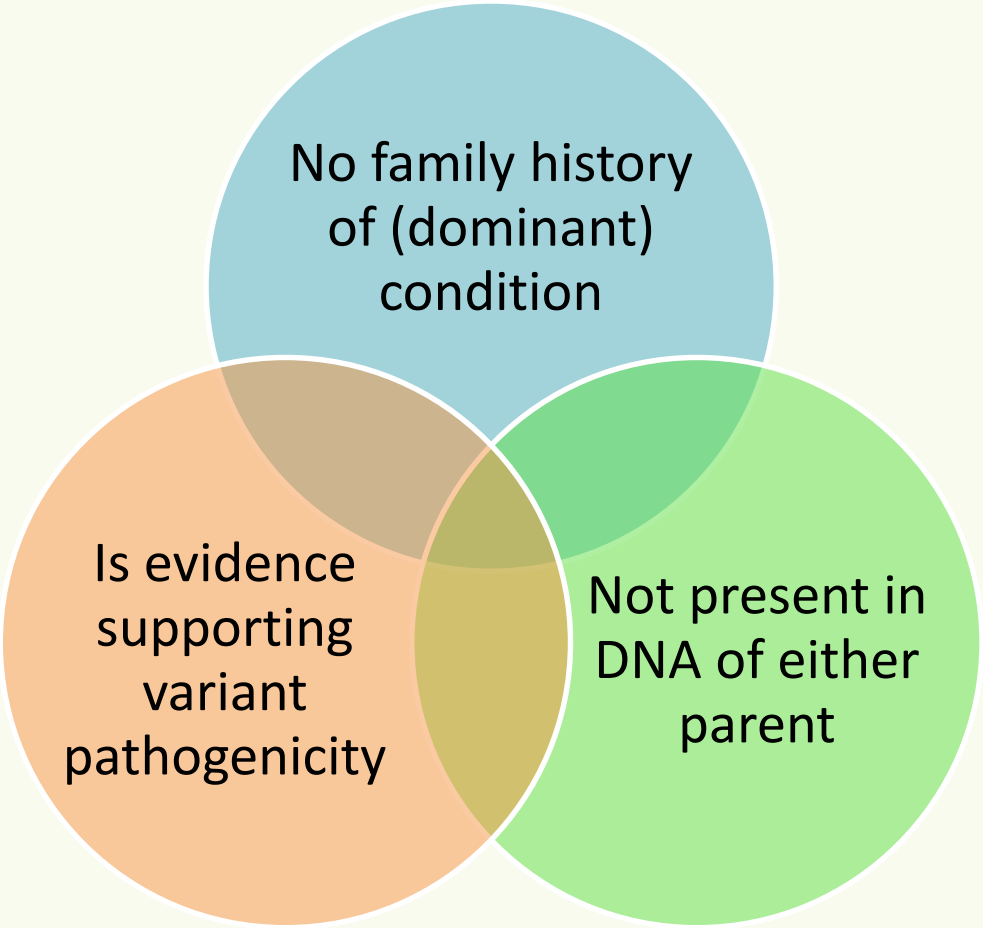
Mitochondrial Inheritance

Both sexes affected	Vertical transmission	Energy-intensive organs
• Variable expression	• Variable chance • *Maternal lineage* only • *No* transmission from males	• Brain • Muscle • Liver • …

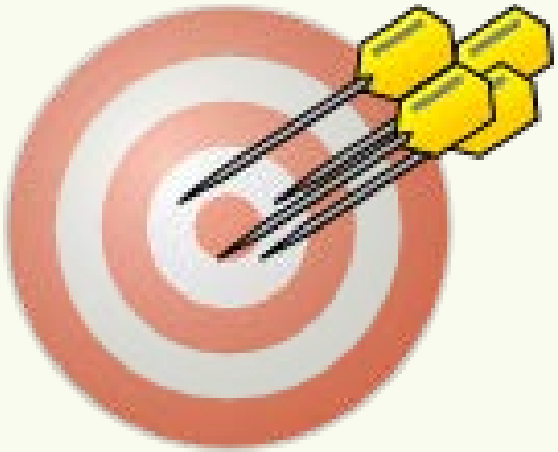




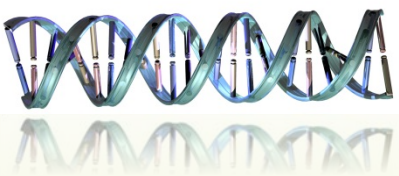
De Novo (New Mutation)



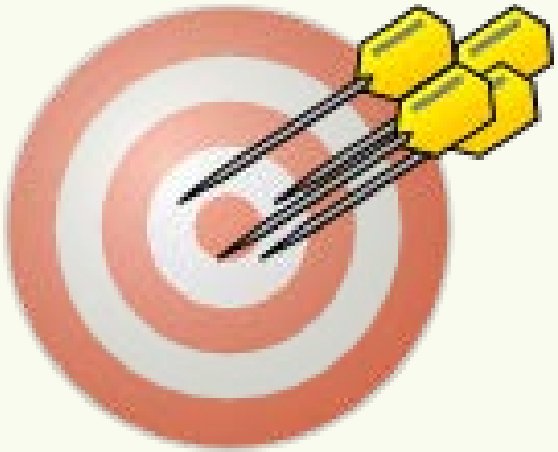
Chance and Risk



- "It happened by chance.."
  - Unplanned, not orchestrated, Apparently completely random
- Likelihood
  - Empirical
  - Predictable process with random step
  - Chance and Risk



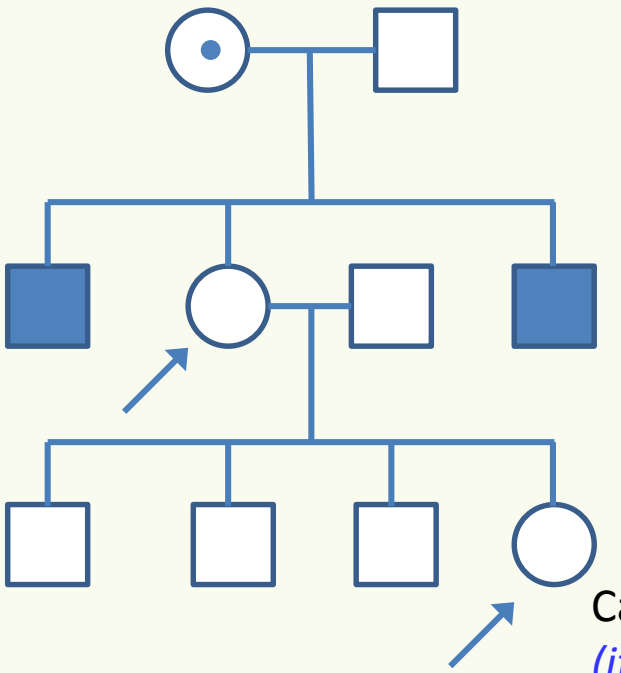
Chance in Genetics



- New genetic change (mutation)

- Inheritance of a genetic trait or disorder

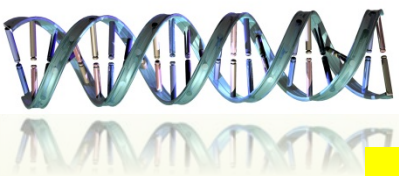
Visual Bayes Theorem: XLR Example



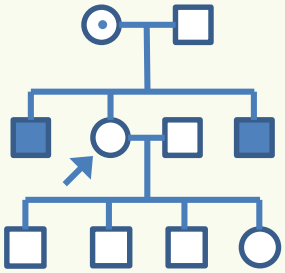
Carrier risk =½

(if her mother is carrier)

Hemophilia A



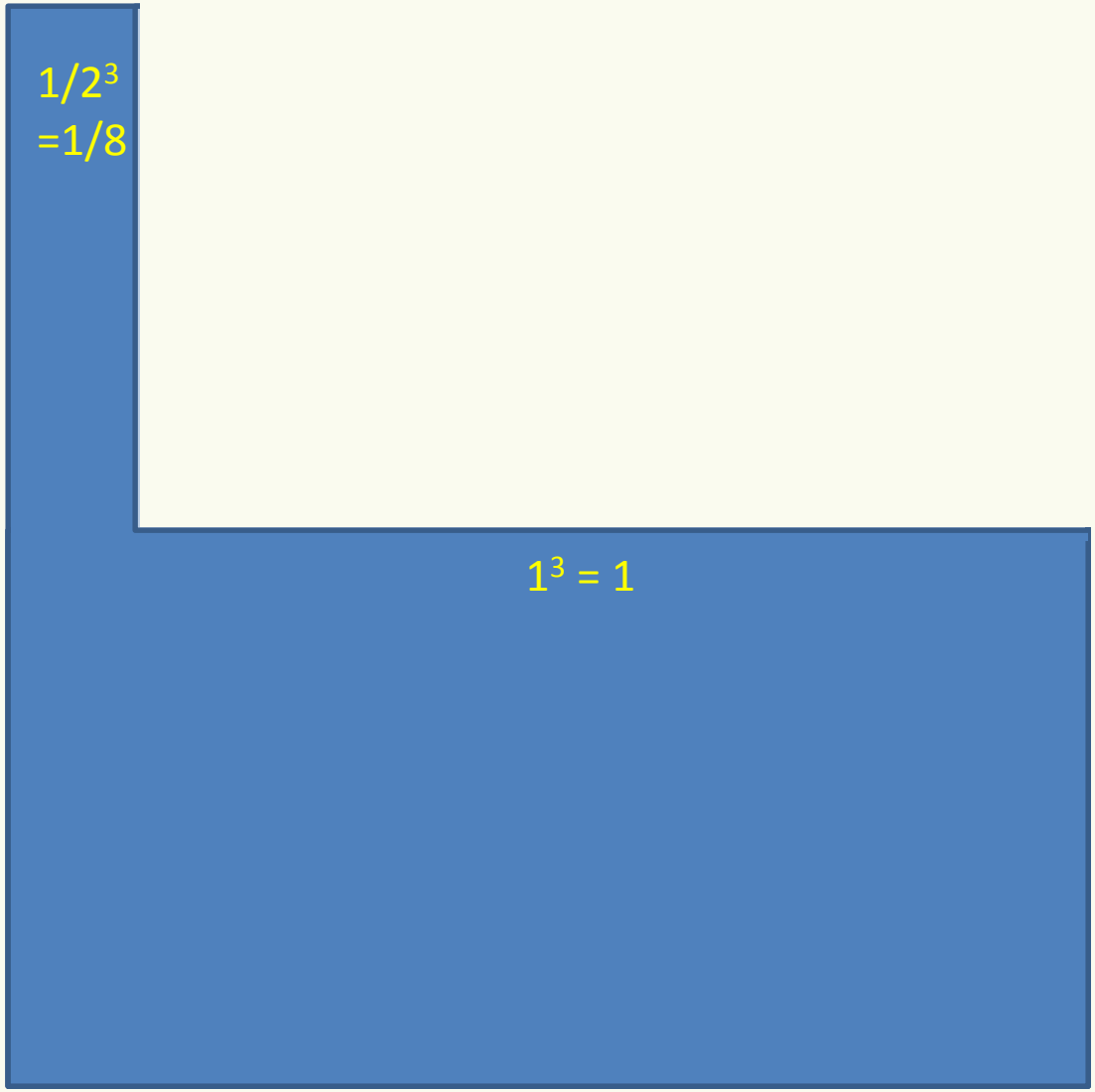
Bayes Theorem



Carrier XLR

probability of
3 normal sons

1



Probability of carrier

$$\frac{1/8 \times 1/2}{(1/8 \times 1/2) + (1 \times 1/2)}$$

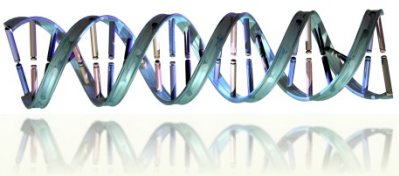
$$\frac{1/16}{1/16 + 1/2}$$

$$\frac{1/16}{9/16}$$

$$\frac{\mathbf{1}}{\mathbf{9}}$$

½ carrier

½ not carrier



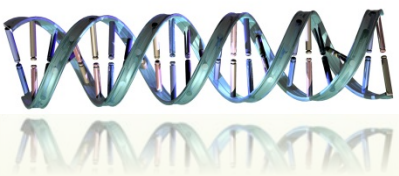
Resources

www.Genome.gov

http://www.genome.gov/GenomeTV/

- Education and training videos
  http://www.genome.gov/27552687

http://www.genome.gov/dmd/
Digital Media Database



THANK YOU!