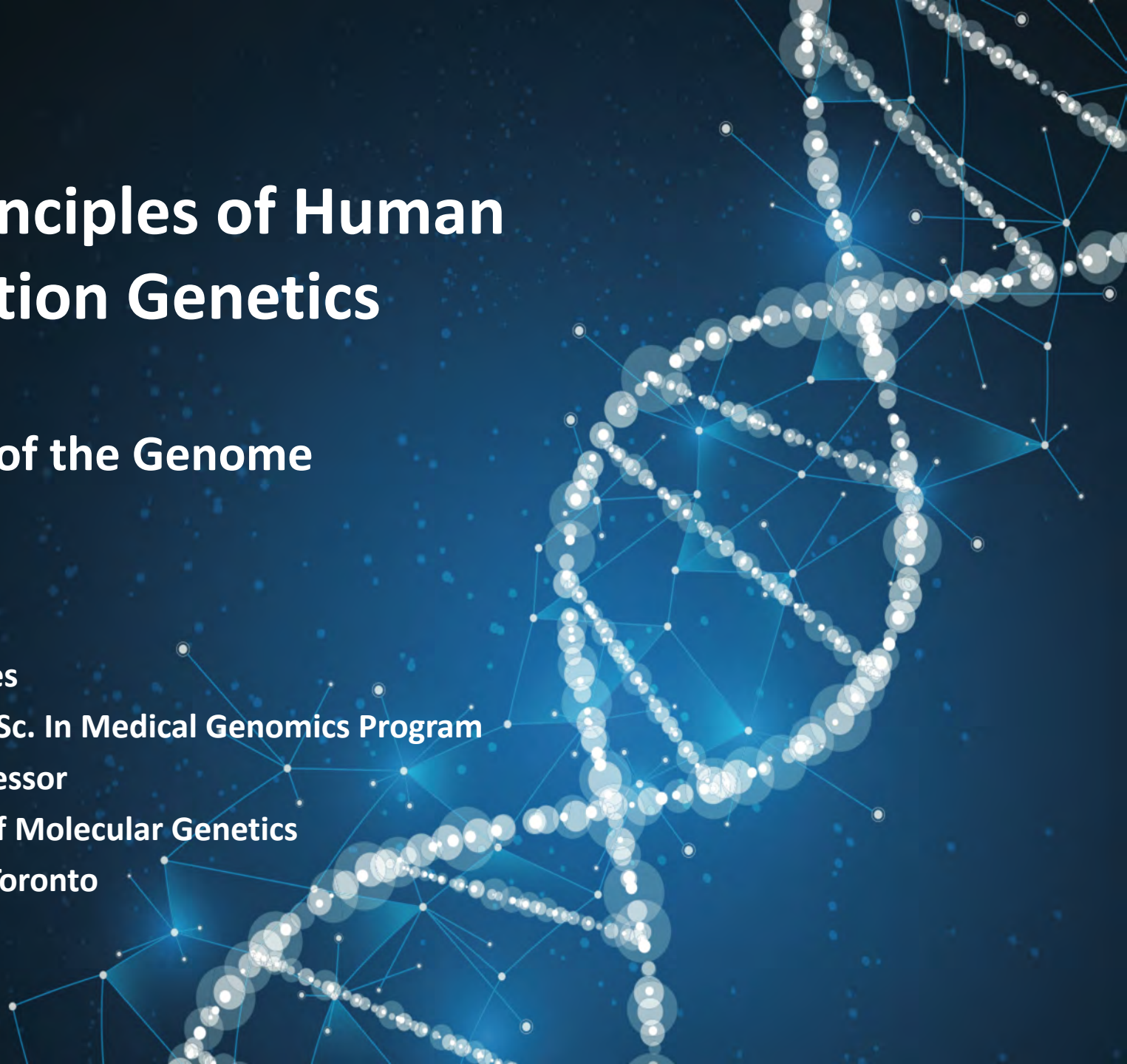
The Principles of Human Population Genetics

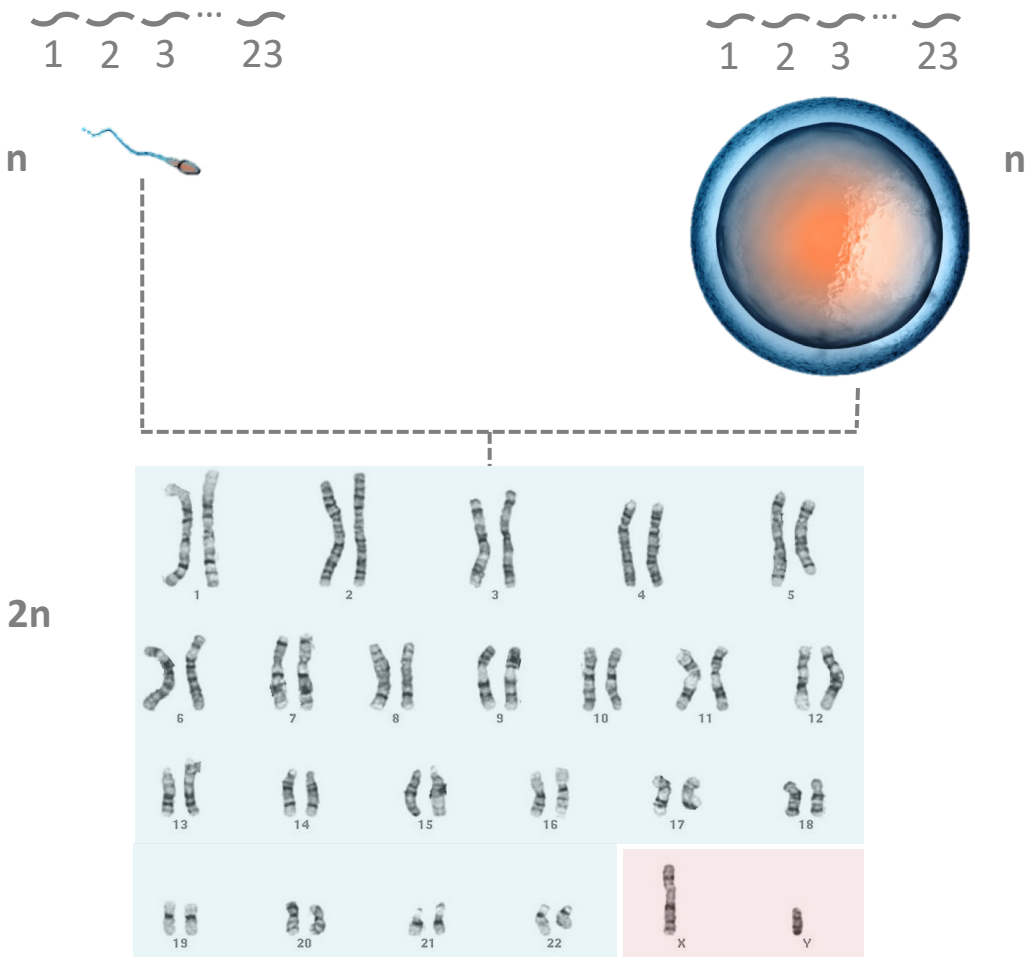
Lecture 1:

Anatomy of the Genome

Dr. Erin B Styles
Director, M.H.Sc. In Medical Genomics Program
Assistant Professor
Department of Molecular Genetics
University of Toronto



The Anatomy of The Human Genome: An Introduction



Haploid: containing a single copy (n) of each chromosome. In humans, n = 23

Diploid: containing two copies of each chromosome. In humans, 2n = 46

Image: Miami Edu; http://fig.cox.miami.edu/~cmallery/150/mendel/karyotype.htm

The Anatomy of The Human Genome: An Introduction

- Each chromosome carries a different subset of genes, which are arranged linearly along its DNA

Homologous Chromosomes (homologs) carry matching genetic information – at any specific **locus** they may be identical, or carry different **alleles**

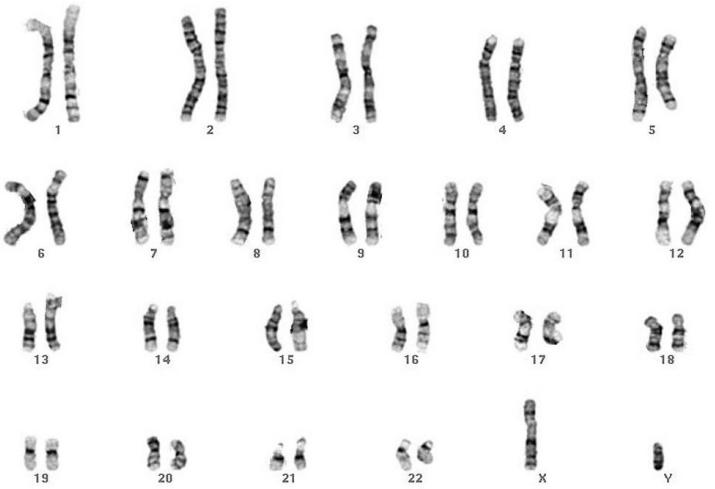
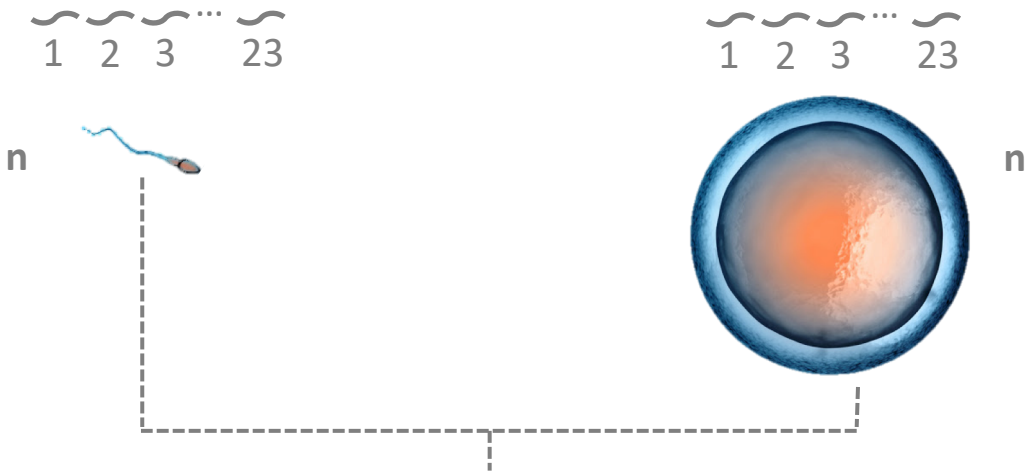


Image: Miami Edu; http://fig.cox.miami.edu/~cmallery/150/mendel/karyotype.htm

The Anatomy of The Human Genome: An Introduction



How much information is this?

1 human cell:
n = 3 feet of DNA
2n = 6 feet of DNA

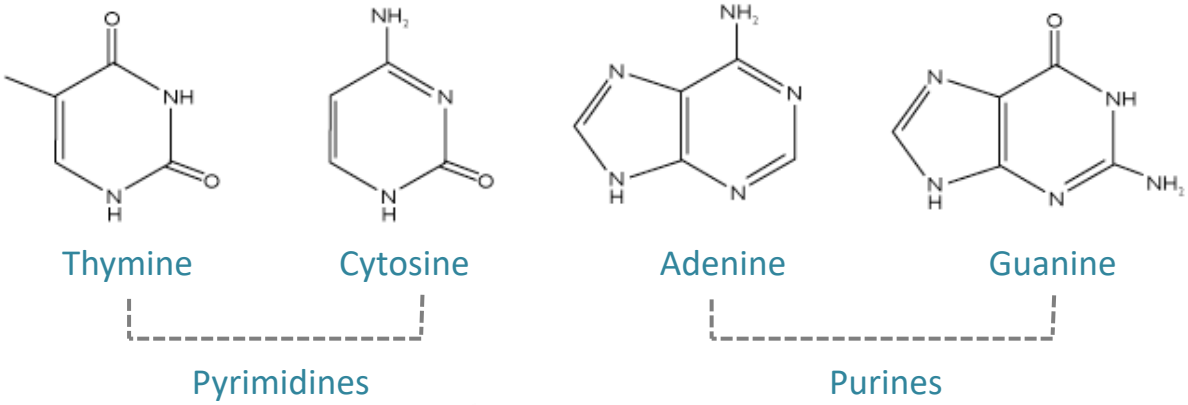
2n

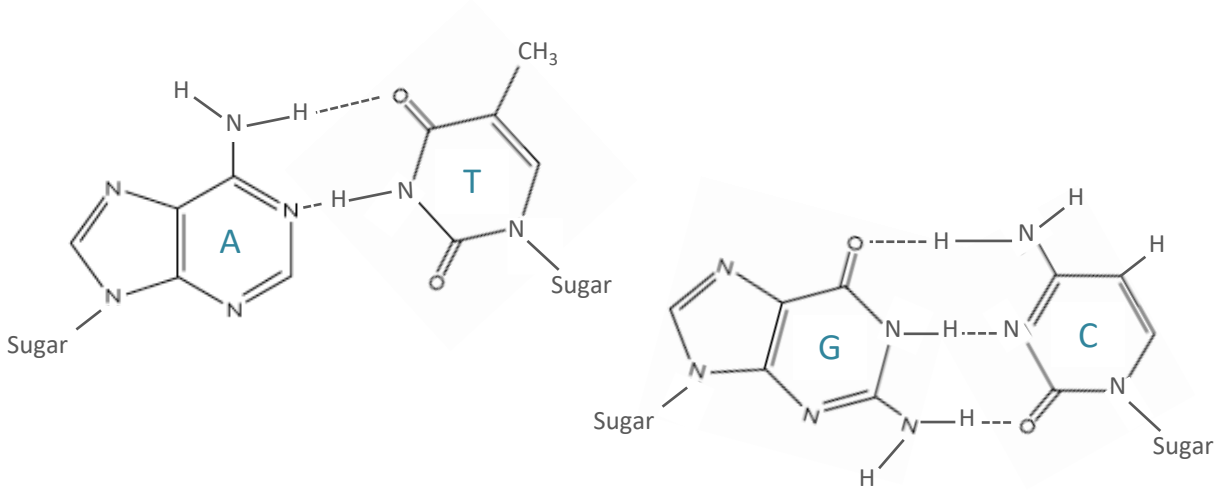
1 baby = ~3.7×10^{13} cells

~42 billion miles

The Anatomy of The Human Genome: DNA Structure

Albrecht Kossel won the first Nobel Prize in Physiology & Medicine in 1910, for his discovery of the nucleotide bases.

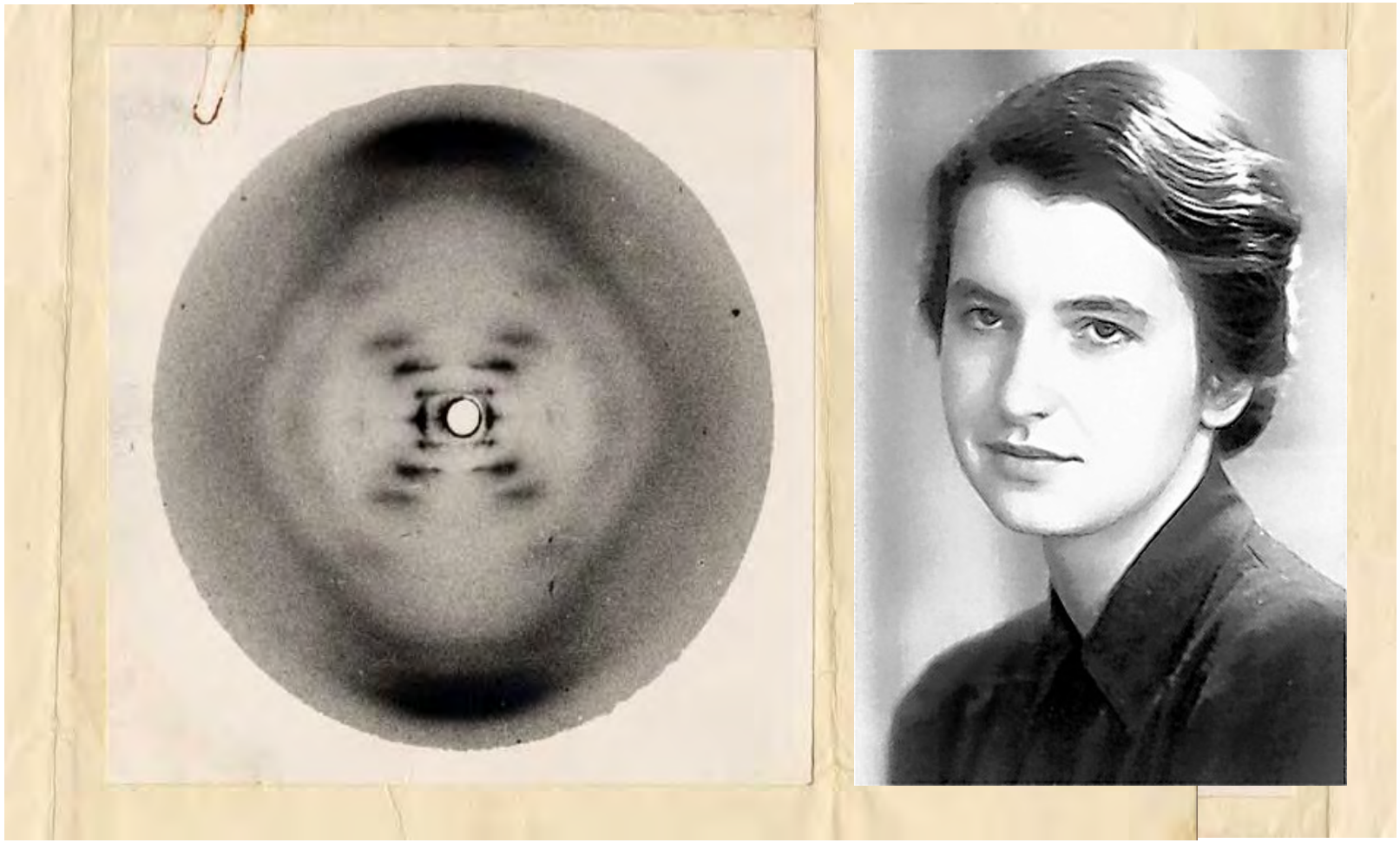




Each chromosome consists of a continuous chain of nucleotides

The nuclear genome consists of 46 linear DNA molecules

The Anatomy of The Human Genome: DNA Structure



Franklin & Gosling, *Nature* 171, 1953.

The Anatomy of The Human Genome: Expert Packaging

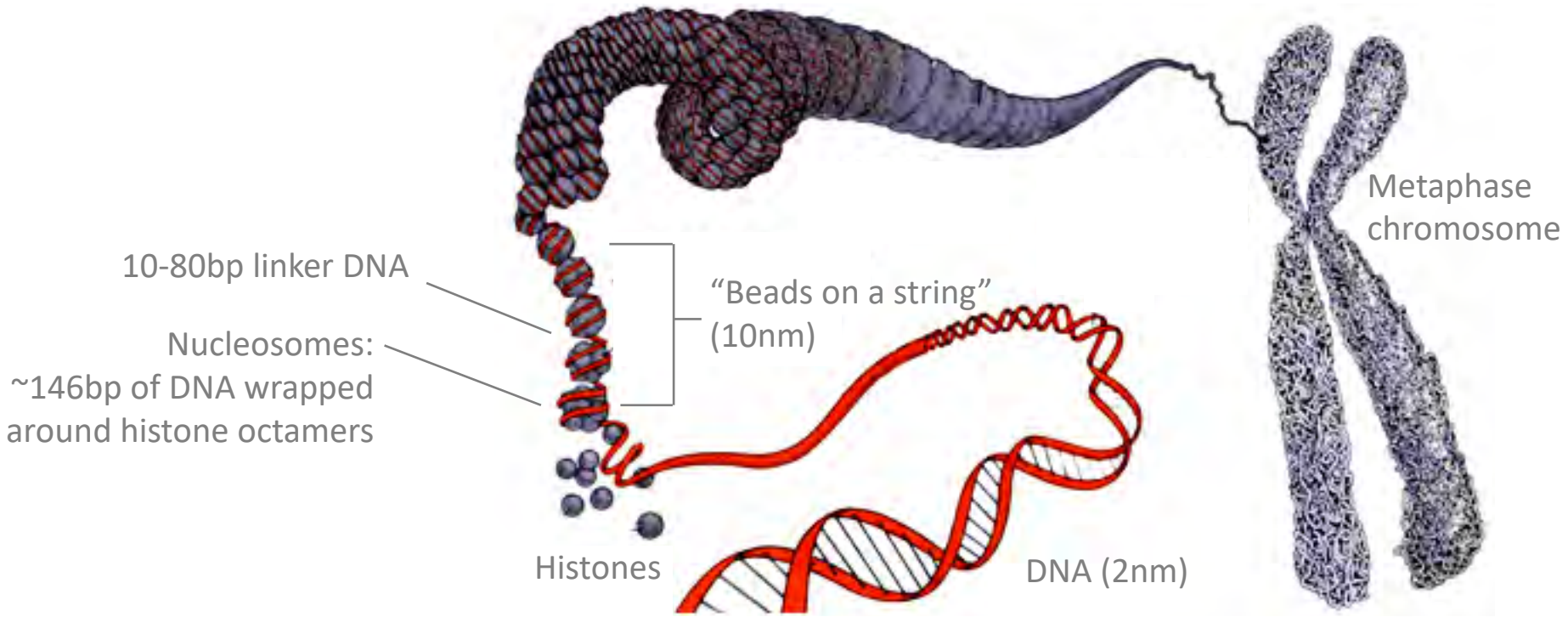


Image: Adapted from Figure 2, Creative Biomart, "*Chromosome Structure Proteins*"

articles

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

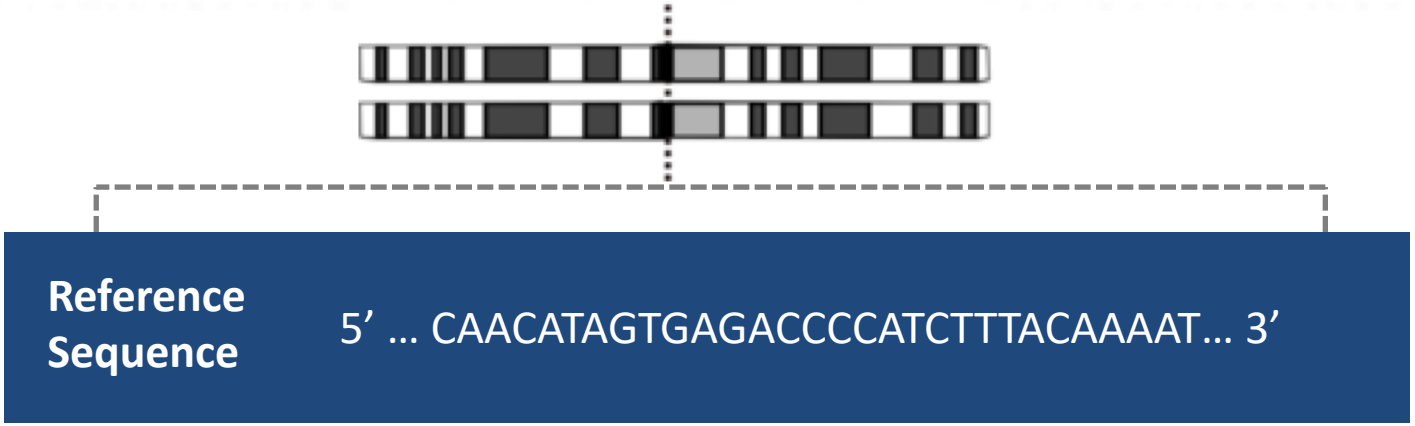
* *A list of authors and their affiliations appears in the* Supplementary Information

One human genome ≈
3.2 billion nucleotides

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

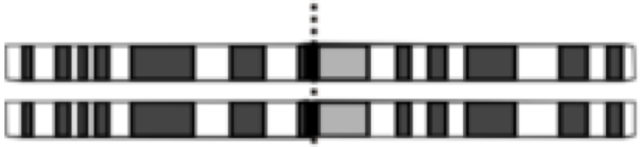
So, what's written in those >3 billion bases? What does the human genome encode? Did the Human Genome Project answer these questions?

The Anatomy of The Human Genome: Insights from The Human Genome Project



- **The Human Genome Project** determined the *sequence* of the majority of the bases in our genome.
- The **reference genome** is a composite – the DNA source material came from many anonymous donors, both male and female, and primarily from Buffalo, NY.

The Anatomy of The Human Genome: Insights from The Human Genome Project



Reference Sequence	5’ … CAACATAGTGAGACCCCATCTTTACAAAAT… 3’
Individual 1	5’ … CAACATAGTGAGACCCCATCTTTACAAAAT… 3’
Individual 2	5’ … CAACATAG**C**GAGACCCCATCTTTACAAAAT… 3’
Individual 3	5’ … CAACATAGTGAGACCCCATCTTTA**A**AAAAT… 3’
Individual 4	5’ … CAAC **--** AGTGAGACCCCATCTTTACAAAAT… 3’

The Anatomy of The Human Genome: Insights from The Human Genome Project

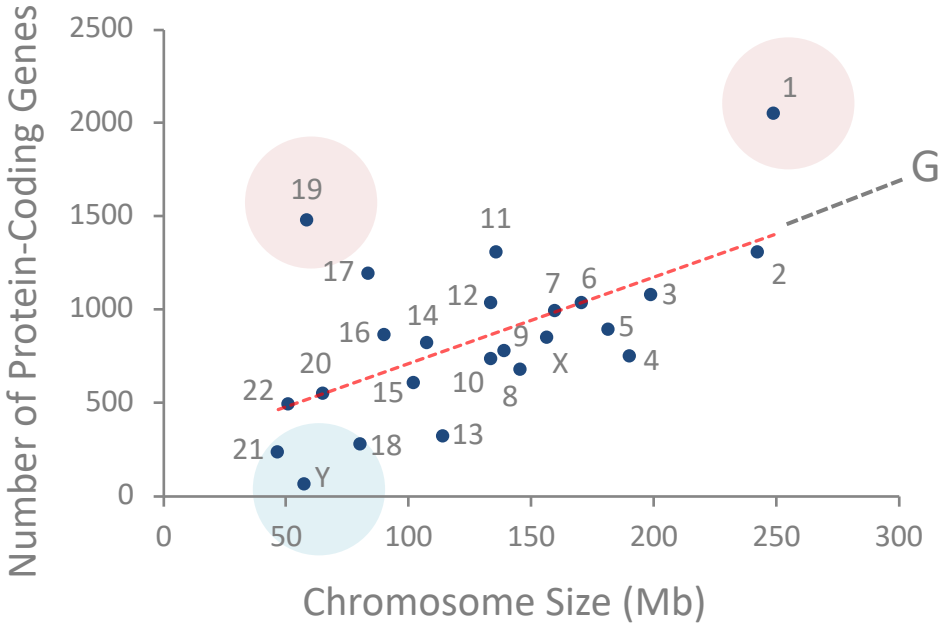
- **The Human Genome Project** determined the *approximate number of genes* in our genome

- Produced a refined list of predicted protein-coding genes (20 – 25 000), down from an earlier estimate of ~100 000

- The mammalian genome does **not contain many more genes** than the genomes of many other organisms

The Anatomy of The Human Genome: Insights from The Human Genome Project

Genes are not randomly distributed throughout the genome – there are chromosome regions (and even entire chromosomes) that are more **gene-rich** or **gene-poor** relative to the rest



Average gene – 27 000 bp
Smallest gene (tRNA genes) – 76 bp
Largest gene (DMD) – 2.4×10^6 bp

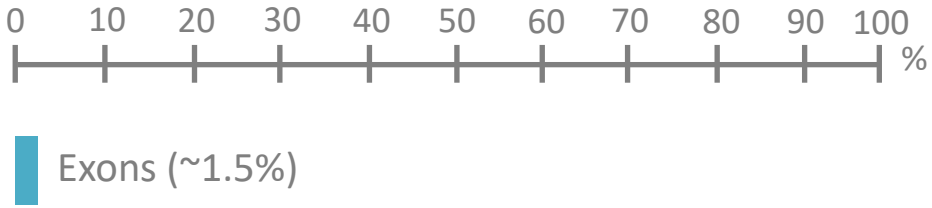
Data in graph from Ensembl Release 91

The Anatomy of The Human Genome: What's Left to Learn?

What did the HGP *not* tell us?

1. What is our genome comprised of? What do our 3.2 billion bps encode?
2. What kinds of variation can be present in the human genome?
3. How can this variation impact human health, and contribute to disease?

The Anatomy of The Human Genome: DNA Elements



Protein-coding genes

- Average intron size – 3 365 bp
- Average exon size – 145 bp

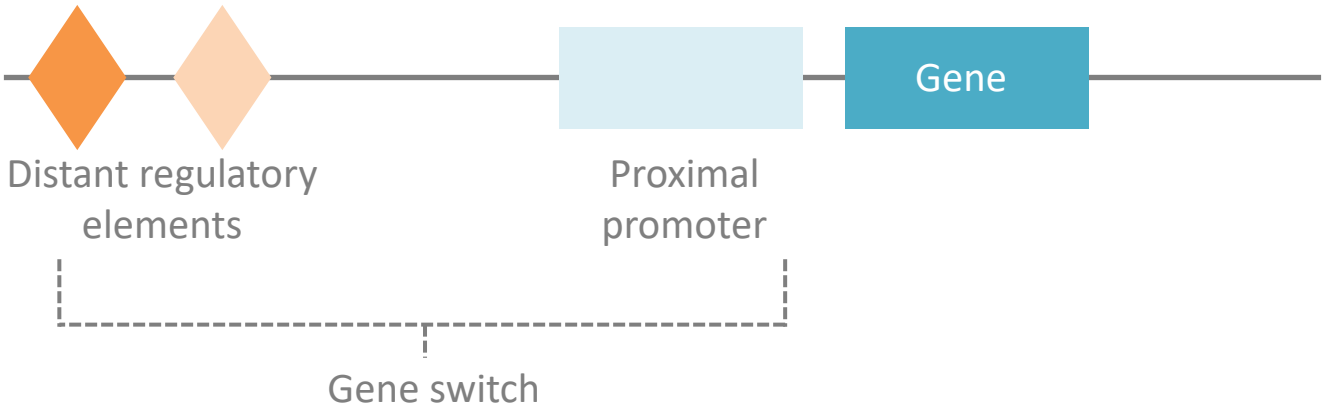
The Anatomy of The Human Genome: DNA Elements

Exons (~1.5%)

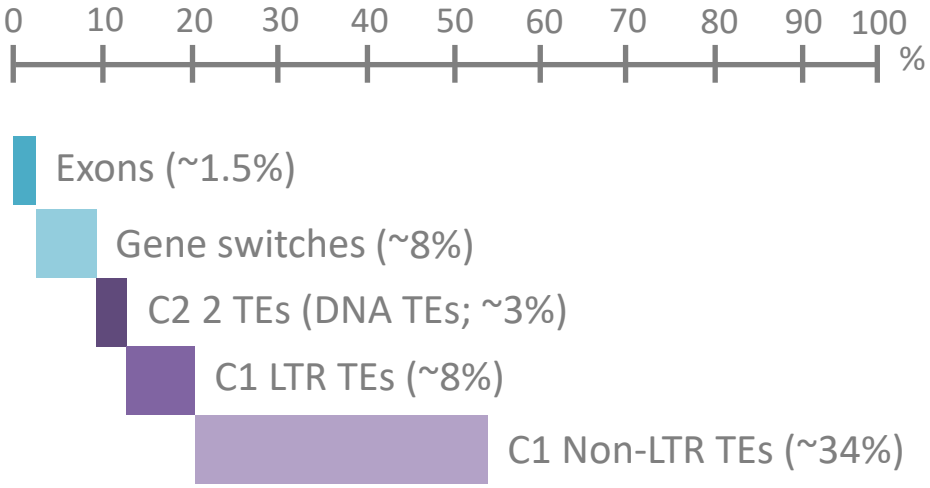
Gene switches (~8%)

<u>Genes switches</u> determine if a gene is expressed, & are comprised of:

- A proximal promoter
- Distant regulatory element(s)



The Anatomy of The Human Genome: DNA Elements



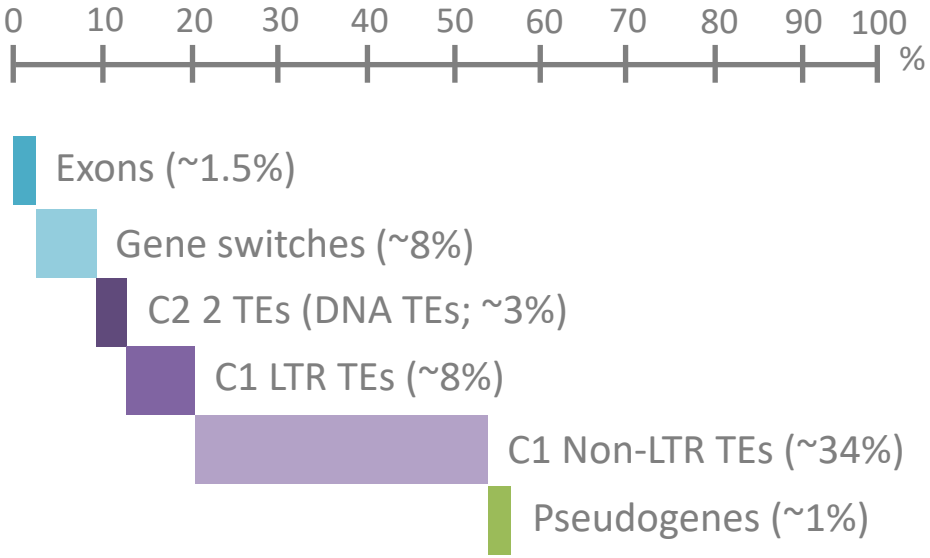
Transposable elements also called *transposons* and *jumping genes* are DNA sequences that move from one location in the genome to another.

There are two major types of TEs:

- Class 2 TEs
  - *DNA Transposons*
- Class 1 TEs
  - *LTR Retrotransposons*
  - *Non-LTR Retrotransposons*

- Class 2 TEs and Class 1 LTR TEs exist in the human genome exclusively as *fossils*
- Class 1 Non-LTR TEs are the *only* active TEs in the human genome!

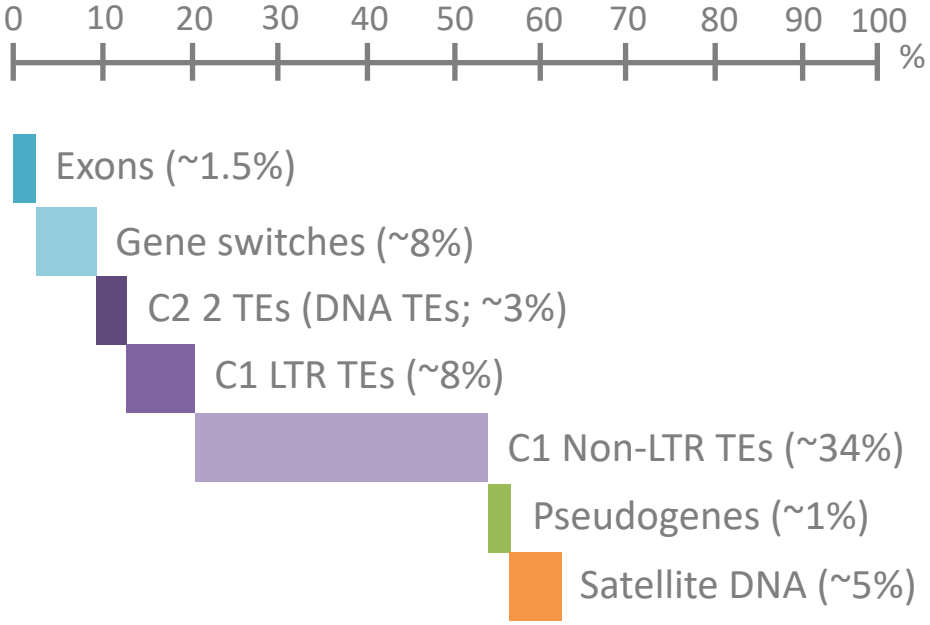
The Anatomy of The Human Genome: DNA Elements



Pseudogenes are DNA sequences that are *similar* to coding genes.

- Have sequence homology to a known gene + loss of some / all function
- There are ~14 600 pseudogenes in the human genome

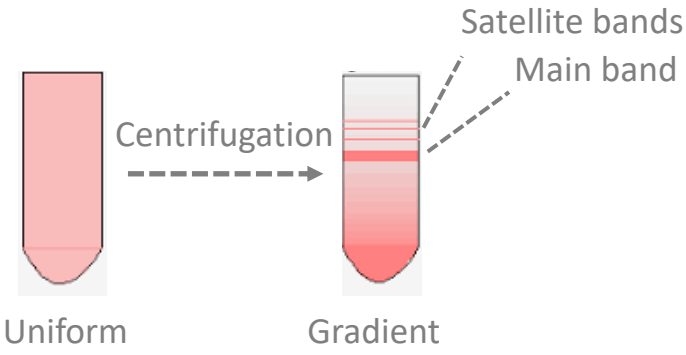
The Anatomy of The Human Genome: DNA Elements



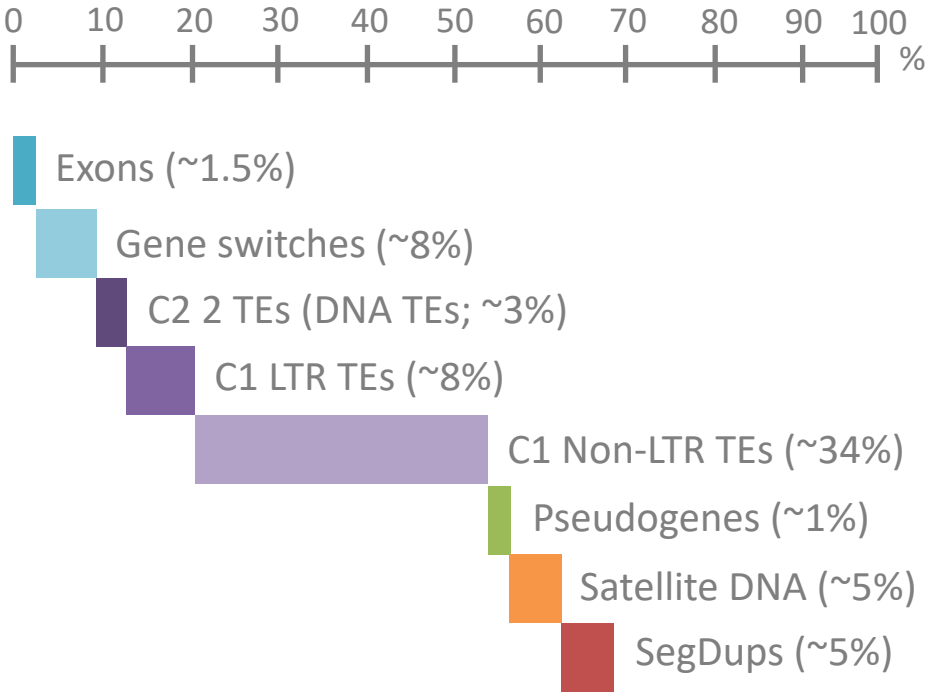
Satellite DNA are clustered tracts of repeated DNA motifs. They're classified into three main groups based on the length of the repeat:

- Microsatellites (1-9bp)
- Minisatellites (10-60bp)
- Macrosatellites (~65- several thousand bp)

- Comprised of a different proportion of A, C, T & G than bulk DNA, and therefore have a different density.



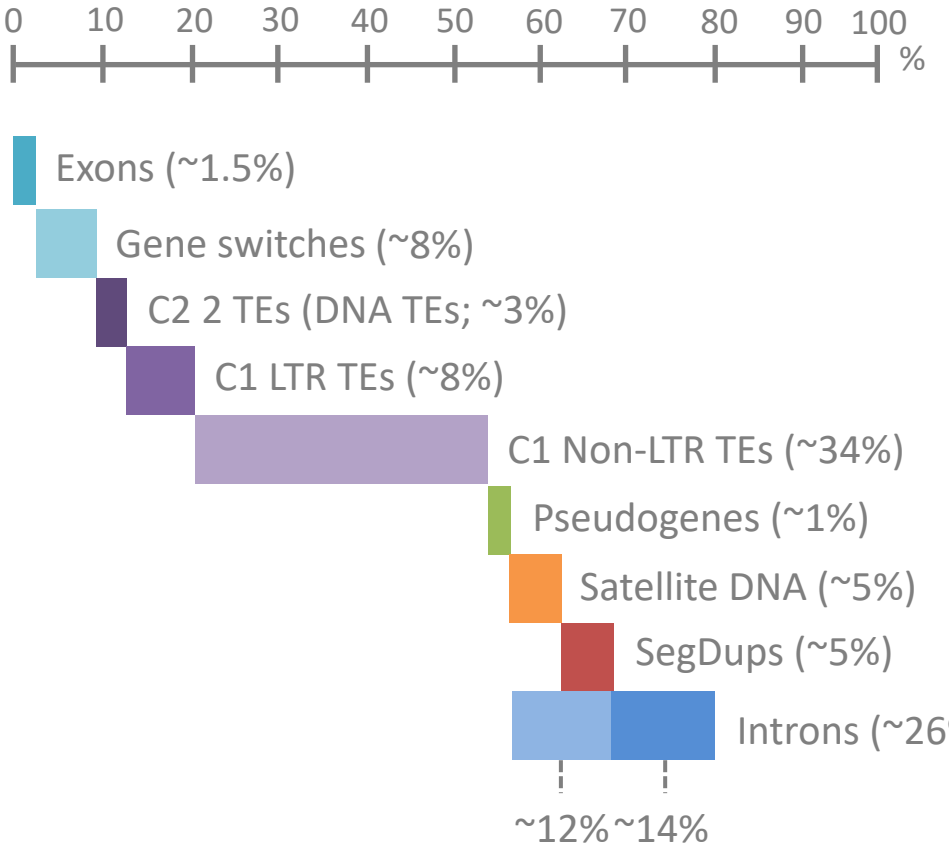
The Anatomy of The Human Genome: DNA Elements



Segmental Duplications, also called "low copy repeats," involve duplications of substantial segments of a chromosome (can span hundreds of kilobases).

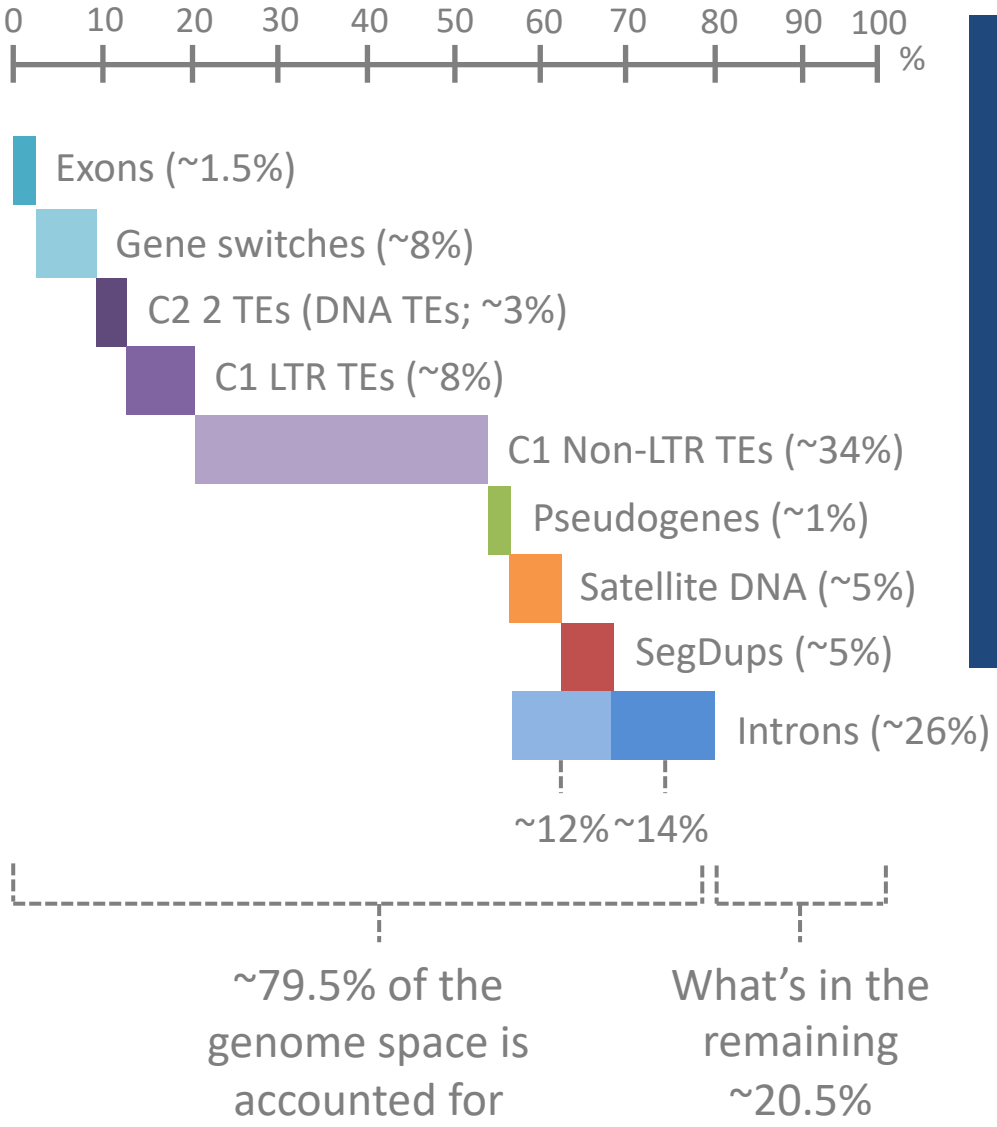
Segmental Duplications have been found on all human chromosomes and are not randomly distributed - some chromosomes (e.g. Y) have an exceptionally high proportion

The Anatomy of The Human Genome: DNA Elements



~56% of the genome is comprised of some form of repetitive DNA

The Anatomy of The Human Genome: DNA Elements



These categories are **not fixed** – as technology changes, new research modulates the proportion of the genome comprised of each of these elements.

For years, researchers thought all the DNA other than the ~1.5% of the genome consisting of protein-coding DNA was non-functional **junk DNA**, but we're rapidly proving ourselves wrong!

The Anatomy of The Human Genome: Why Are We Doing This?

"We shall not cease from exploration. And the end of all our exploring will be to arrive where we started, and know the place for the first time."

T. S. Eliot

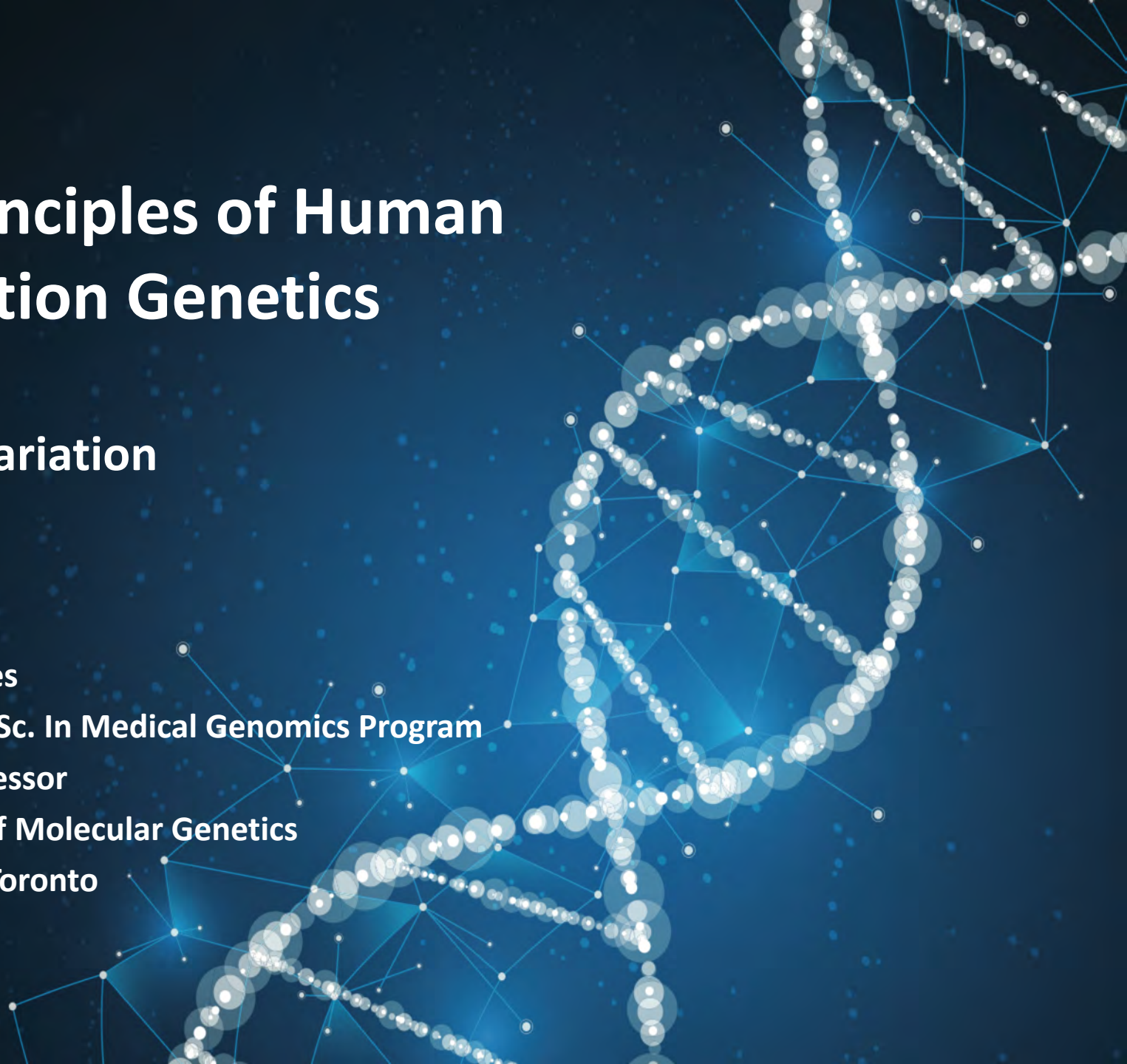
Around the time when the Human Genome was first being sequenced, Francis Collins, who was the then-director of the National Human Genome Research Institute (NHGRI), noted that the human genome could be likened to a book with multiple uses:

1. It's a history book
2. It's a shop manual and a blueprint
3. It's a medical textbook

The Principles of Human Population Genetics

Lecture 2:

Genetic Variation



Dr. Erin B Styles
Director, M.H.Sc. In Medical Genomics Program
Assistant Professor
Department of Molecular Genetics
University of Toronto

The Scale of Human Genetic Variation

Variation occurs by changes to our base DNA sequence, and can be roughly classified into two categories:

Changes that do not affect total DNA content

- A single nucleotide being **replaced** with another
- Multiple nucleotides at a time being **sent to another location**
- A series of nucleotides being **inverted**

Changes in copy number of a DNA sequence

- Abnormal chromosome segregation leading to **fewer or more chromosomes** than normal
- Altered numbers of **specific sequences**
- Deletion or insertion of a **single nucleotide**

Variants and Polymorphism

- Mutations result in alternative forms of DNA known as **variants**

If a variant is present in a population at a

>1% frequency in that population = **Polymorphism**

<1% frequency in that population = **Rare Variant**

Single Nucleotide Polymorphisms (SNPs)

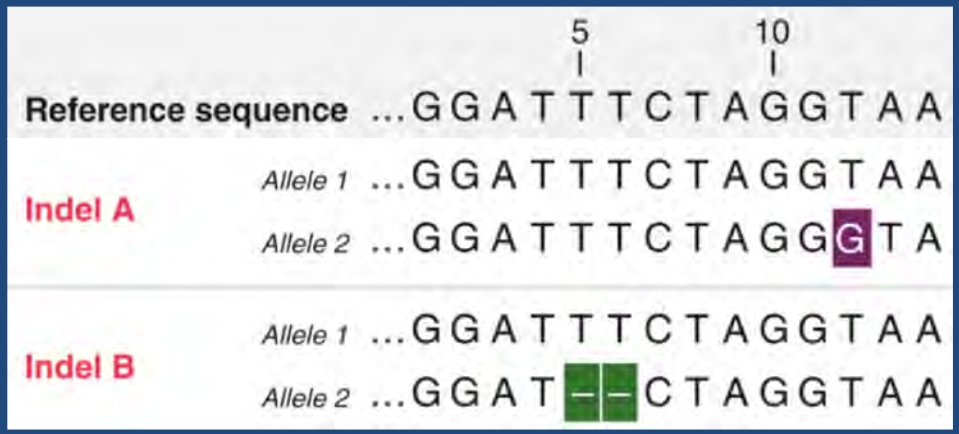
- ~1x every 1000 bp in the genome (more in noncoding regions than exons)
  - ~700 000 000 SNPs have been identified
  - >100 000 of these have been documented in exonic regions
- The health significance of *the vast majority* of these is unknown

Adapted from Figure 4.6, *Genetics and Genomics in Medicine*, 2015

Variants and Polymorphism

Insertion-Deletion Polymorphisms (in/dels or indels)

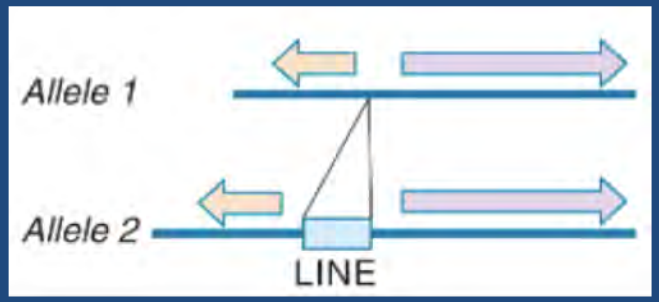
- 1 to ~1000 bp in size
- >1 million have been described
- >50% are *simple*
  - Only 2 alleles exist
- The rest are *multiallelic*
  - >2 alleles exist



Adapted from Figure 4-1, *Genetics in Medicine Eighth Edition,* 2015

Insertion-Deletion Polymorphisms (in/dels or indels) – Major Types

Mobile Element Insertion (MEI) Polymorphisms (simple)



- LINE and *Alu* elements are the most common
- ~10 000 MEI polymorphisms have been described
- At least 5000 have an insertion frequency >10% in various populations!

Adapted from Figure 4-2, *Genetics in Medicine Eighth Edition,* 2015

Insertion-Deletion Polymorphisms (in/dels or indels) – Major Types

Microsatellite Polymorphisms (multiallelic)

- Repeated between 1 and ~50x at a particular site
- Many tens of thousands of microsatellite polymorphic loci have been identified in the human genome

Adapted from Figure 4-2, *Genetics in Medicine Eighth Edition,* 2015

Copy Number Variants (CNVs)

- Related to indels, but *much* larger repeats
  - Variants >500 kb are found in 5-10% of the population
  - Variants > 1 Mb are found in 1-2%

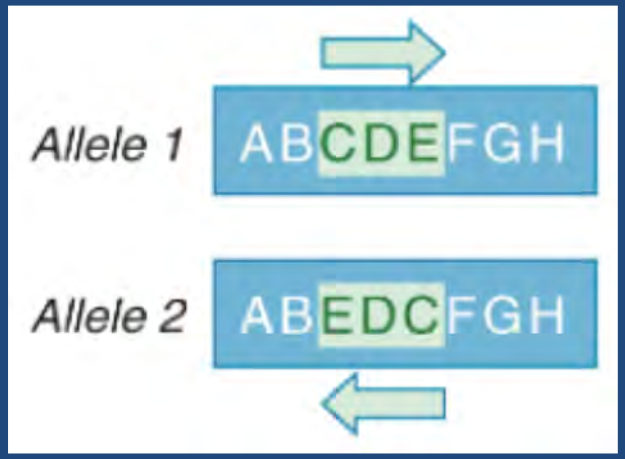
- Smaller CNVs tend to have only 2 alleles (the presence or absence of the segment), but larger CNVs tend to have multiple alleles

Adapted from Figure 4-1, *Genetics in Medicine Eighth Edition,* 2015

The content of any two human genomes can differ by as much as 50-100 Mb due to copy number differences at CNV loci!

Inversion Polymorphisms

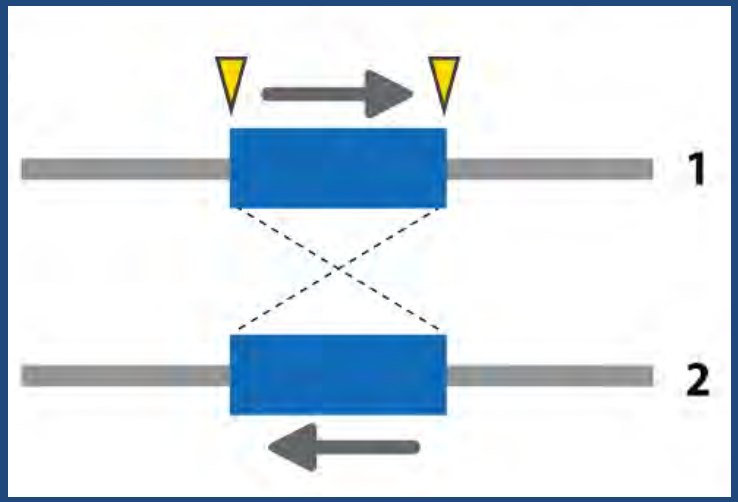
- Differ in size from a few base pairs to several Mb
- Simple polymorphisms (2 alleles possible)
- *Balanced* inversions don't involve a gain or loss of DNA, and can achieve very high frequencies in the general population (10-35%)

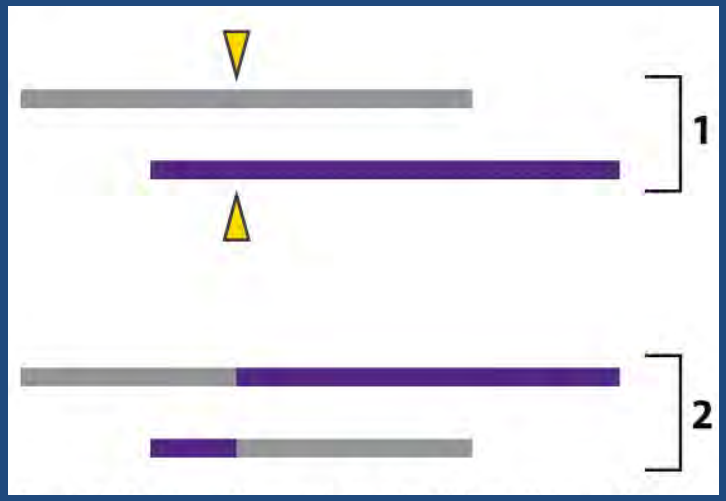


Adapted from Figure 4-1, *Genetics in Medicine Eighth Edition,* 2015

Variants and Polymorphism: Structural Variation

Balanced Structural Variation:



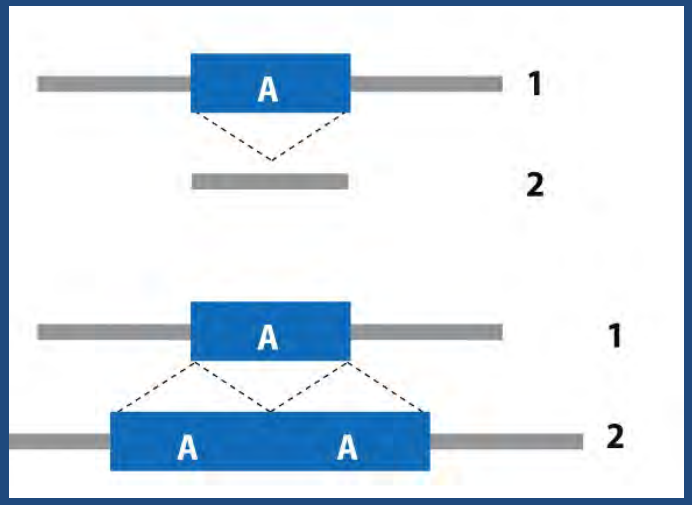


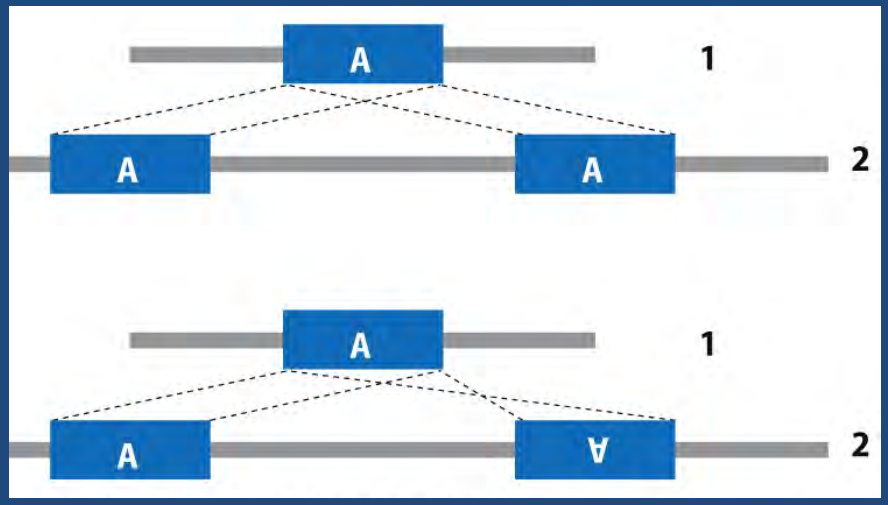
DNA variants have the same DNA content and differ in organization

Adapted from Figure 4-2, *Genetics in Medicine Eighth Edition,* 2015

Variants and Polymorphism: Structural Variation

Unbalanced Structural Variation:



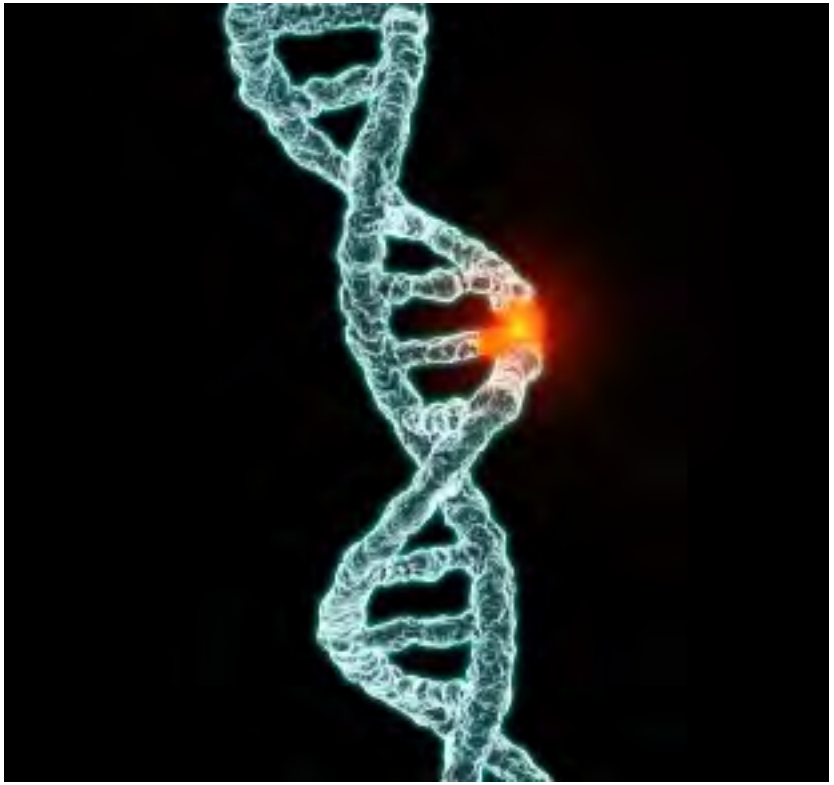


DNA variants differ in DNA content

Adapted from Figure 4-2, *Genetics in Medicine Eighth Edition,* 2015

Variants and Polymorphism: Consequences of Mutation

- Different kinds of mutations have different effects

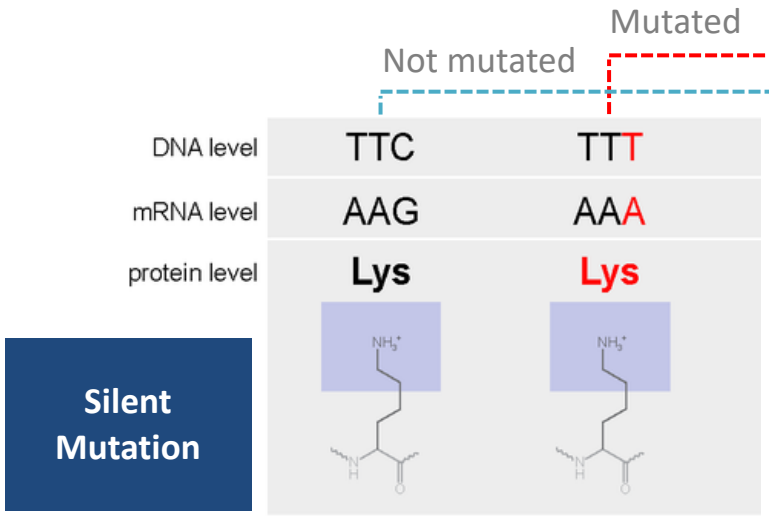


https://www.ashg.org/education/csertoolkit/nofindings.html

- Nucleotide substitutions:
  - Synonymous mutations
  - Nonsynonymous mutations
- Deletions, Insertions, and Rearrangements

Variants and Polymorphism: Consequences of Nucleotide Substitution Mutations

- **Synonymous** mutations fall into one of two categories:
  - Silent mutations
  - Mutations that affect RNA transcription, processing, and translation
- These mutations modify the DNA without altering the amino acid sequence encoded



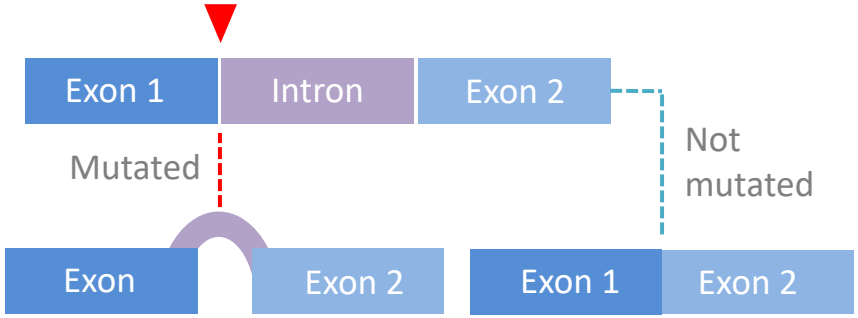
First Base	Second Base: U		C		A		G		Third Base
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC	phe	UCC	ser	UAC	tyr	UGC	cys	C
	UUA	leu	UCA	ser	UAA	stop	UGA	stop	A
	UUG	leu	UCG	ser	UAG	stop	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC	leu	CCC	pro	CAC	his	CGC	arg	C
	CUA	leu	CCA	pro	CAA	gln	CGA	arg	A
	CUG	leu	CCG	pro	CAG	gln	CGG	arg	G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC	ile	ACC	thr	AAC	asn	AGC	ser	C
	AUA	ile	ACA	thr	AAA	lys	AGA	arg	A
	AUG	met	ACG	thr	AAG	lys	AGG	arg	G
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC	val	GCC	ala	GAC	asp	GGC	gly	C
	GUA	val	GCA	ala	GAA	glu	GGA	gly	A
	GUG	val	GCG	ala	GAG	glu	GGG	gly	G

Table 3-1, *Genetics in Medicine Eighth Edition,* 2015

Adapted from User Jonsta247, *Wikimedia Commons,* 2013

Variants and Polymorphism: Consequences of Nucleotide Substitution Mutations

- **Synonymous** mutations fall into one of two categories:
  - Silent mutations
  - Mutations that affect RNA transcription, processing, and translation
- These mutations modify the DNA without altering the amino acid sequence encoded



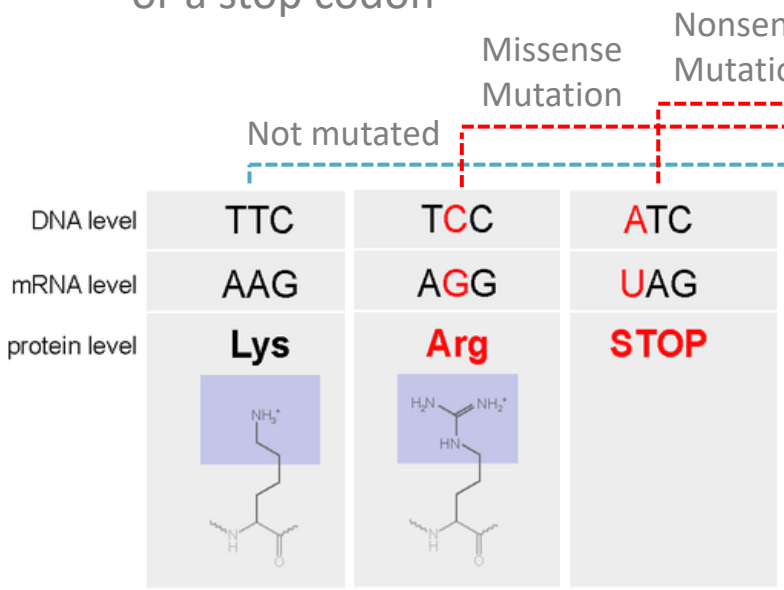
Synonymous, but *not* Silent

First Base	Second Base: U		C		A		G		Third Base
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC	phe	UCC	ser	UAC	tyr	UGC	cys	C
	UUA	leu	UCA	ser	UAA	stop	UGA	stop	A
	UUG	leu	UCG	ser	UAG	stop	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC	leu	CCC	pro	CAC	his	CGC	arg	C
	CUA	leu	CCA	pro	CAA	gln	CGA	arg	A
	CUG	leu	CCG	pro	CAG	gln	CGG	arg	G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC	ile	ACC	thr	AAC	asn	AGC	ser	C
	AUA	ile	ACA	thr	AAA	lys	AGA	arg	A
	AUG	met	ACG	thr	AAG	lys	AGG	arg	G
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC	val	GCC	ala	GAC	asp	GGC	gly	C
	GUA	val	GCA	ala	GAA	glu	GGA	gly	A
	GUG	val	GCG	ala	GAG	glu	GGG	gly	G

Table 3-1, *Genetics in Medicine Eighth Edition,* 2015

Variants and Polymorphism: Consequences of Nucleotide Substitution Mutations

- **Nonsynonymous** mutations fall into one of two categories:
  - Missense mutations
  - Nonsense mutations
- These mutations modify the DNA, which alters the amino acid sequence, inserting a different amino acid or a stop codon



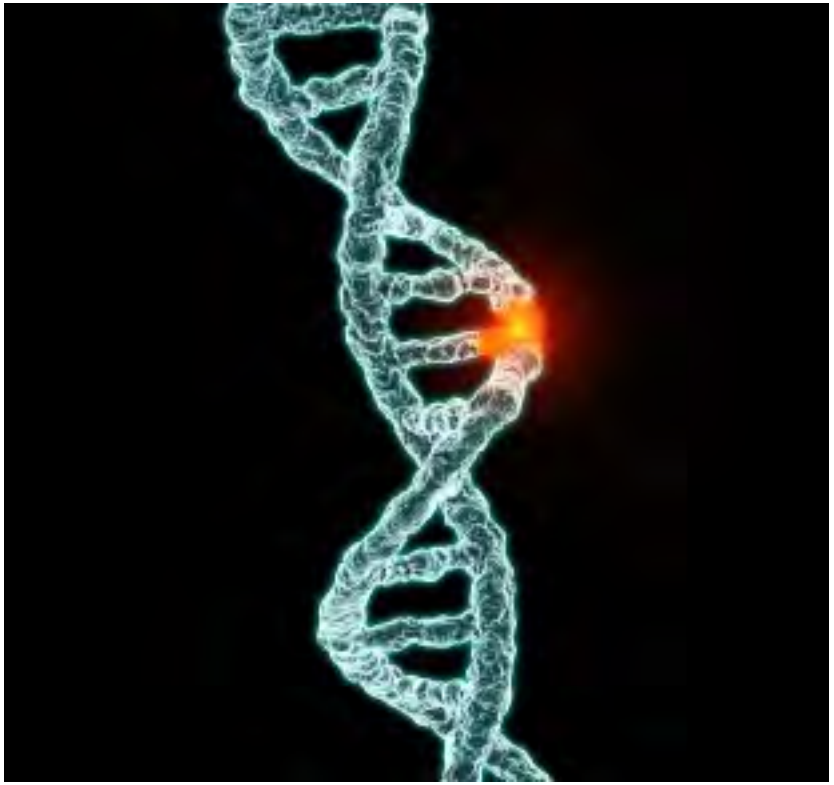
First Base	Second Base: U		C		A		G		Third Base
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC	phe	UCC	ser	UAC	tyr	UGC	cys	C
	UUA	leu	UCA	ser	UAA	stop	UGA	stop	A
	UUG	leu	UCG	ser	UAG	stop	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC	leu	CCC	pro	CAC	his	CGC	arg	C
	CUA	leu	CCA	pro	CAA	gln	CGA	arg	A
	CUG	leu	CCG	pro	CAG	gln	CGG	arg	G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC	ile	ACC	thr	AAC	asn	AGC	ser	C
	AUA	ile	ACA	thr	AAA	lys	AGA	arg	A
	AUG	met	ACG	thr	AAG	lys	AGG	arg	G
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC	val	GCC	ala	GAC	asp	GGC	gly	C
	GUA	val	GCA	ala	GAA	glu	GGA	gly	A
	GUG	val	GCG	ala	GAG	glu	GGG	gly	G

Table 3-1, *Genetics in Medicine Eighth Edition,* 2015

Adapted from User Jonsta247, *Wikimedia Commons,* 2013

Variants and Polymorphism: Consequences of Mutation

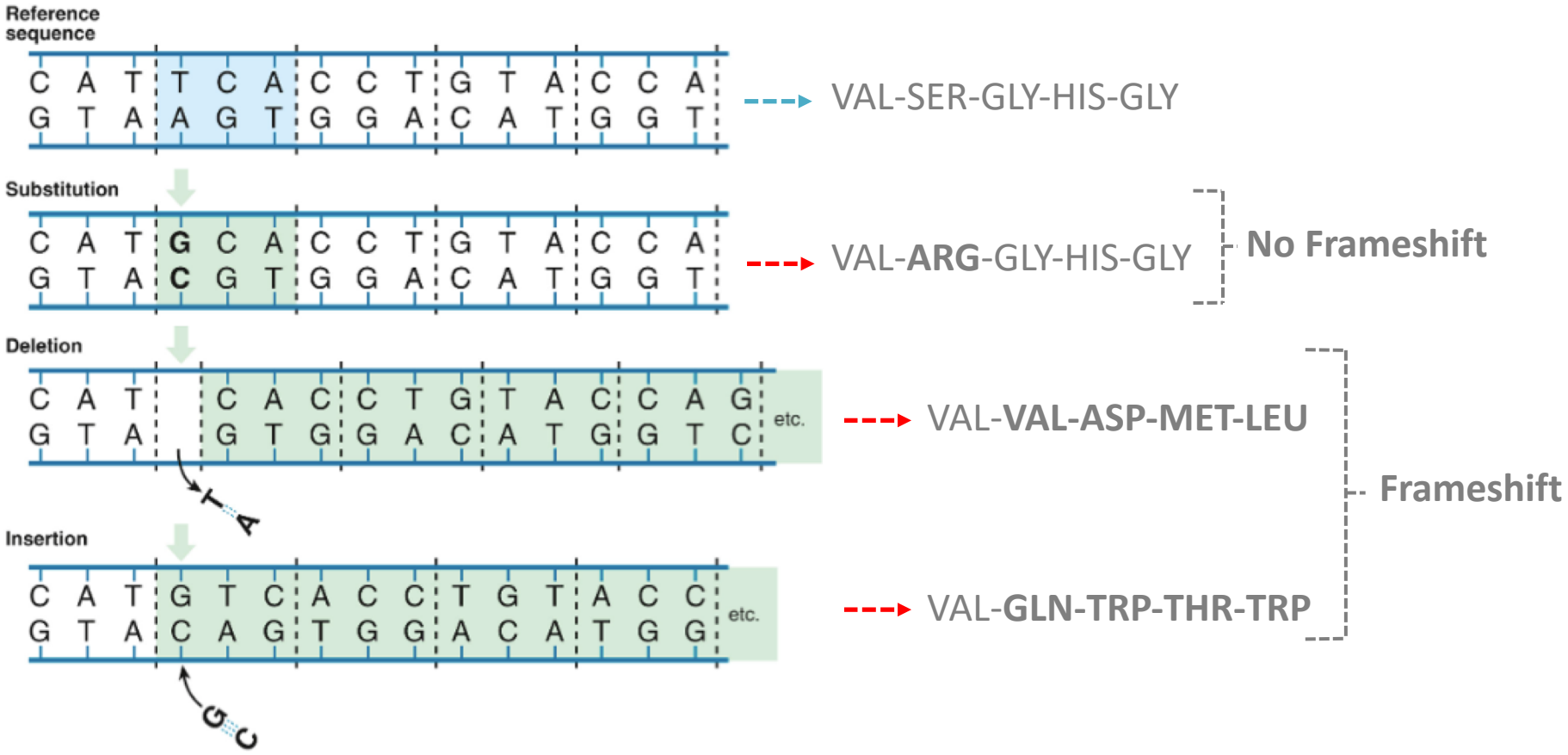
- Different kinds of mutations have different effects



https://www.ashg.org/education/csertoolkit/nofindings.html

- Nucleotide substitutions:
  - Synonymous mutations
  - Nonsynonymous mutations
- Deletions, Insertions, and Rearrangements

Variants and Polymorphism: Consequences of Indels & Rearrangements

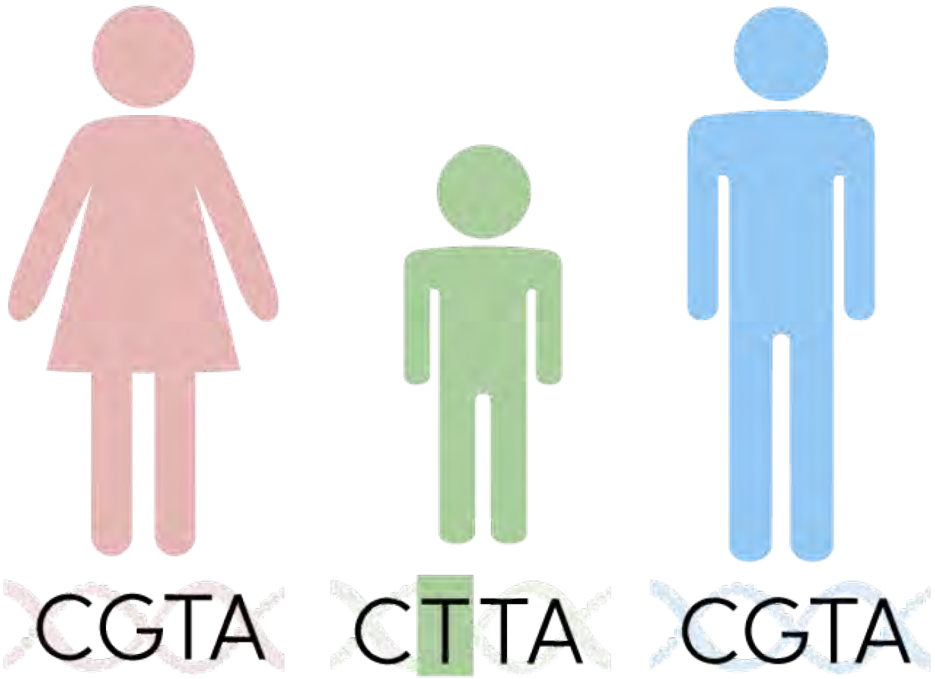


*Deletions, insertions and rearrangement mutations can cause a **frameshift** in the genetic code by altering the **reading frame***

Figure 4-4, *Genetics in Medicine Eighth Edition,* 2015

Overall Rate of Mutation

- The overall occurrence of new mutations can be accurately assessed by whole genome sequencing of **trios**



Overall rate of new mutations (averaged between gametes):

1.2×10^{-8} mutations per base pair, per generation

*Each person is likely to get **~75** new mutations in their genome*

https://www.ashg.org/education/csertoolkit/nofindings.html

Overall Rate of Mutation – How Different Are We?

Single nucleotide changes account for ~75% of DNA changes, and structural variation accounts for the remaining ~25%

Variation Detected in A Typical Human:

1. ~5-10 million SNPs
2. 25-50 thousand rare variants
3. ~75 new base pair mutations
4. 3-7 new CNVs
5. 200-500 thousand indels
6. 500-1000 larger deletions
7. ~150 in-frame indels
8. ~200-250 frameshift mutations
9. 10-12000 synonymous SNPs
10. 8-11000 nonsynonymous SNPs
11. 175-500 rare nonsynonymous variants
12. 1 new nonsynonymous mutation
13. ~100 nonsense mutations
14. 40-50 splice site-disrupting mutations
15. 250-300 genes with likely loss-of-function mutations
16. ~25 genes predicted to be completely inactivated

~640-1280 million different bases, between any two people

*Any two humans on the planet are nonetheless likely to be between **99.8-99.9% genetically identical***

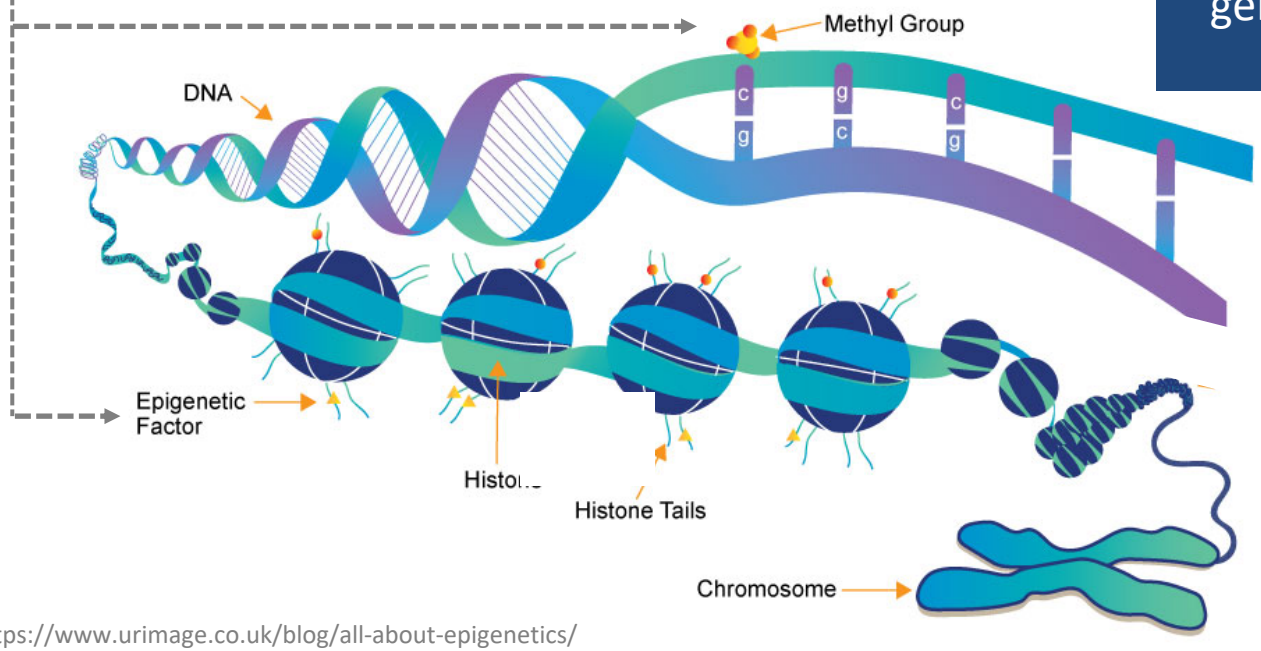
Non-DNA Variation: Epigenetics

Epigenetics is the study of heritable changes in gene expression that *do not* involve changes to the underlying DNA.

A change in phenotype without a change in genotype

"Epi" / επι = Greek for "over, outside of, around"

Epigenetic changes are "on top of," or in addition to the genomic nucleotide sequence.



https://www.urimage.co.uk/blog/all-about-epigenetics/

Epigenetics: Types of Modification

Epigenetic modifications come primarily in three basic types:

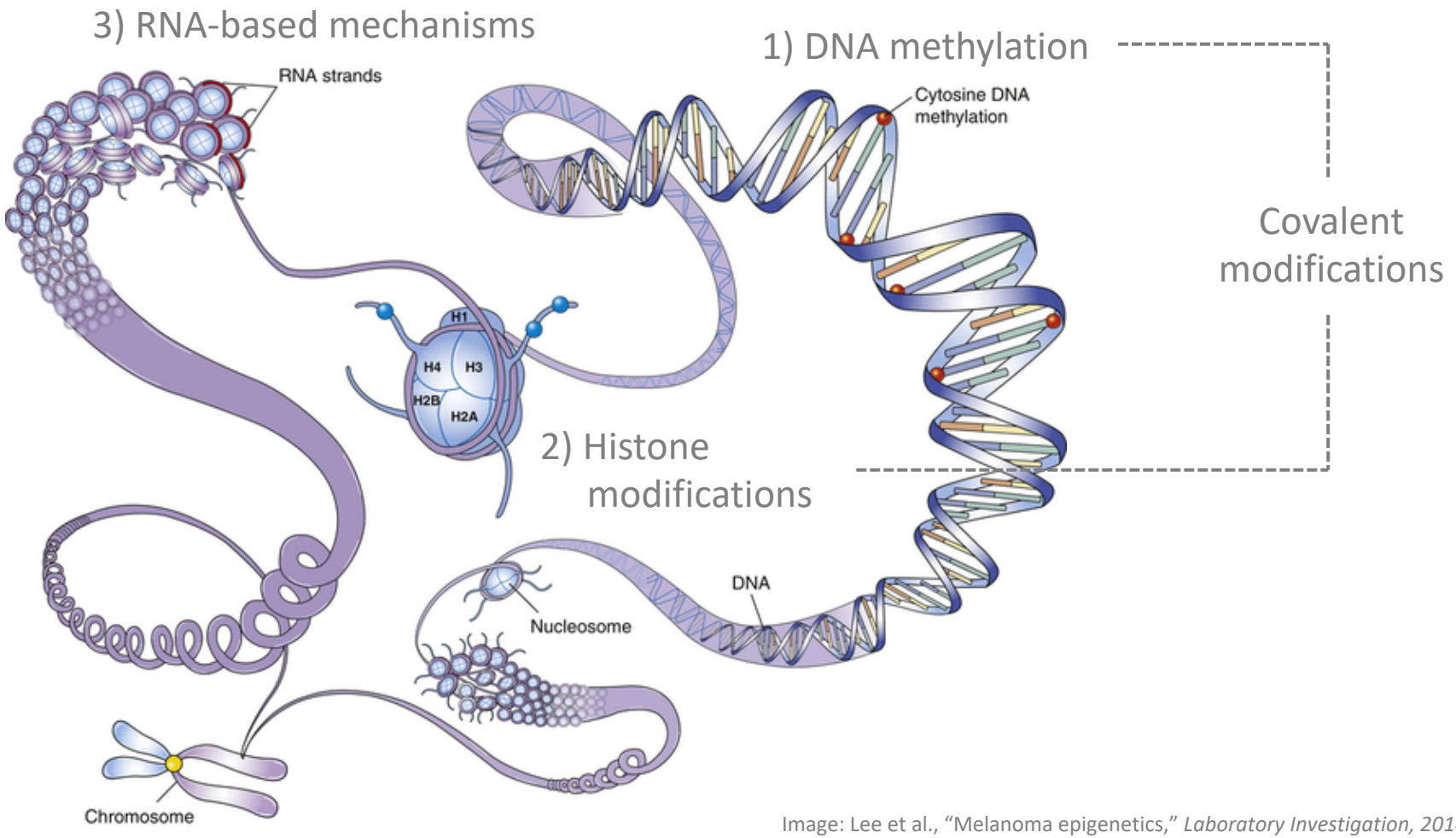


Image: Lee et al., "Melanoma epigenetics," *Laboratory Investigation, 2014.*

Epigenetics: Covalent Modifications

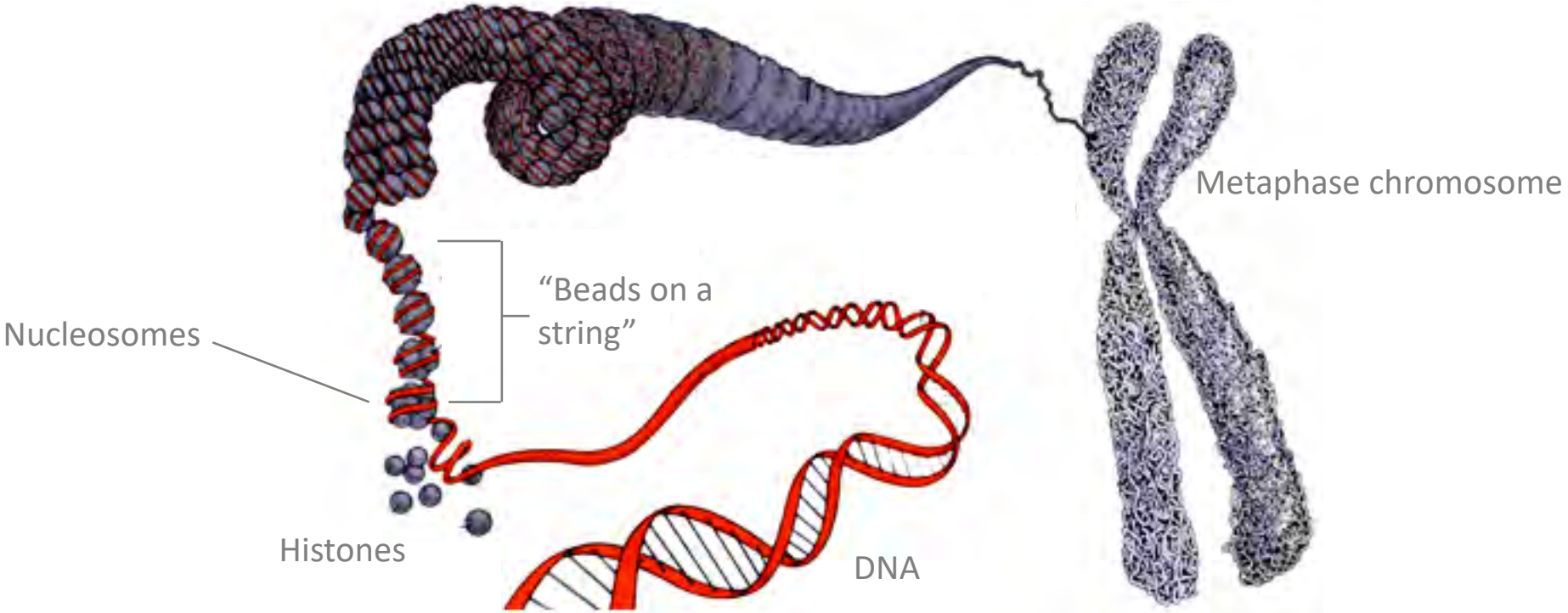


Image: Adapted from Figure 2, Creative Biomart, "*Chromosome Structure Proteins*"

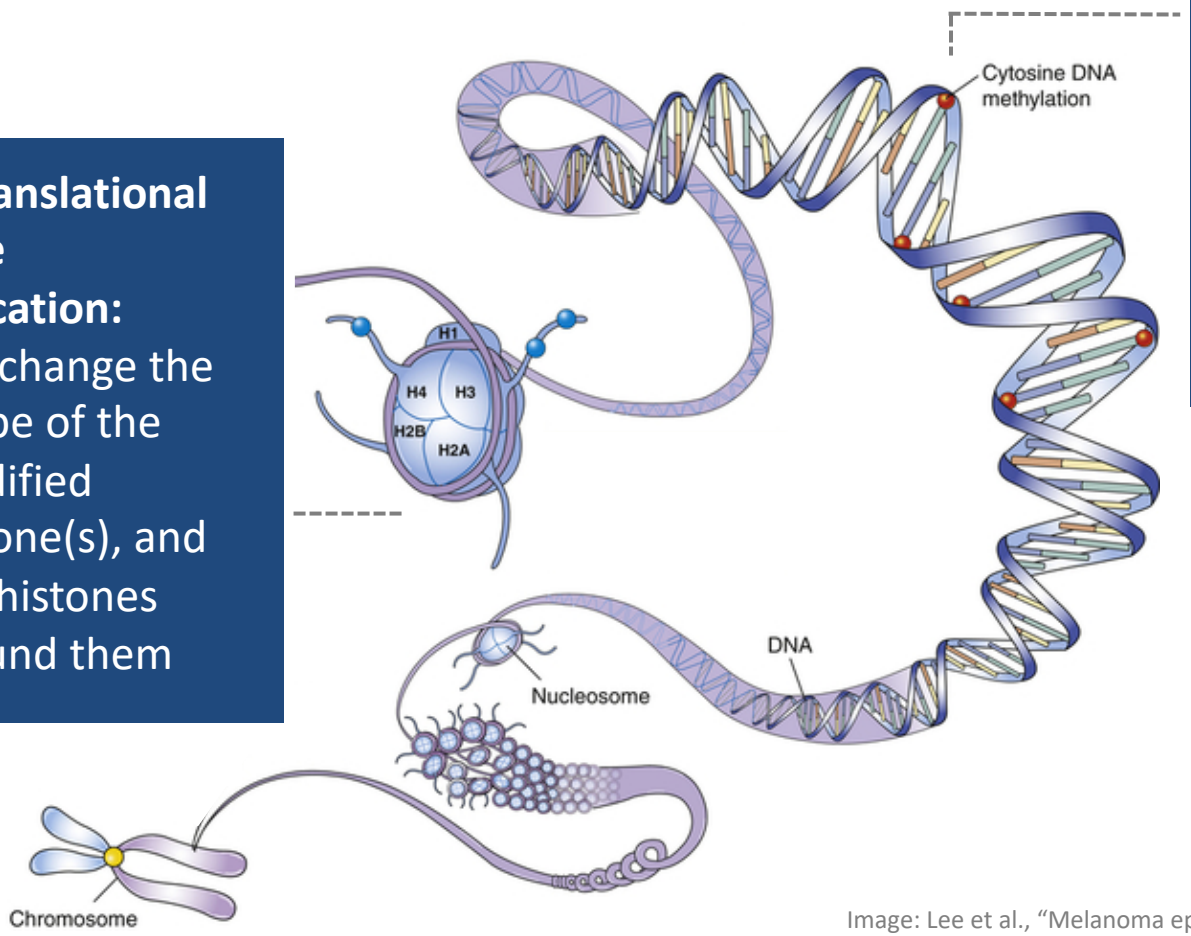
A **major premise** of epigenetic regulation: If the way DNA wraps around the histones changes, gene expression can change too.

Epigenetics: Covalent Modifications

Both covalent mechanisms of epigenetic modification can remodel chromatin.

Post-translational histone modification:

- can change the shape of the modified histone(s), and the histones around them

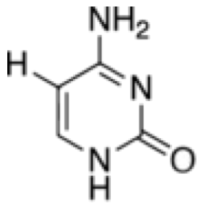


DNA methylation:

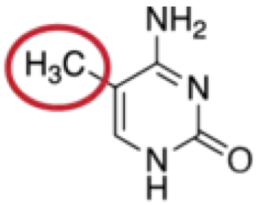
- More attractive to histone-modifying enzymes
- Less transcribed

Image: Lee et al., "Melanoma epigenetics," *Laboratory Investigation, 2014.*

Epigenetics: Genomic Locations of DNA Methylation



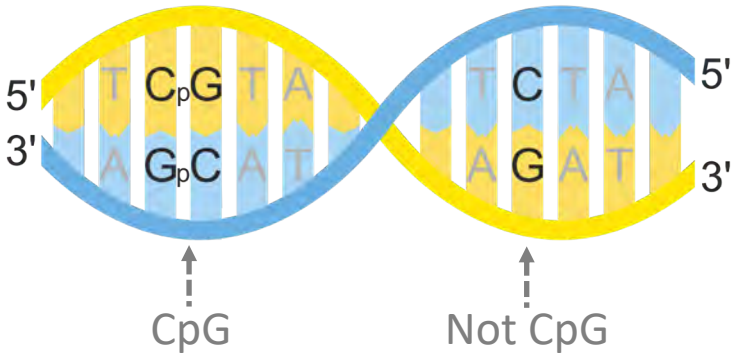
Cytosine



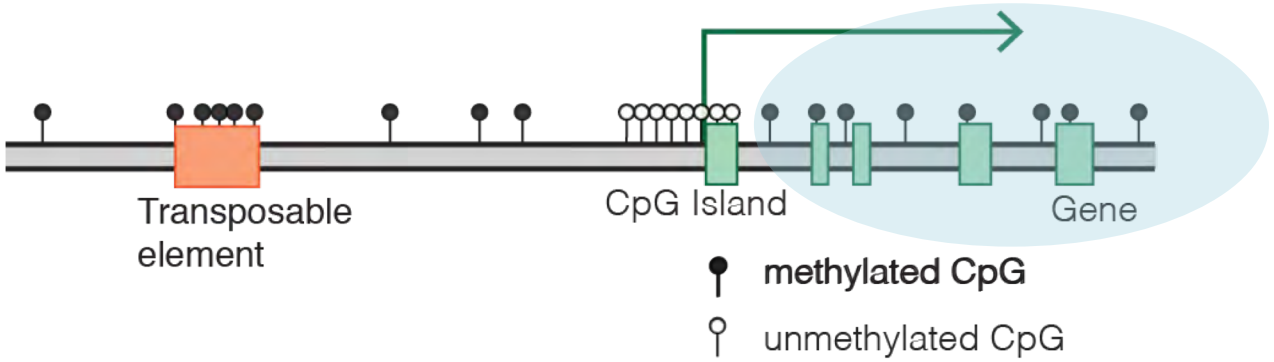
5-methylcytosine / 5-mC

- Occurs by the addition of a methyl group (CH_3) to DNA. Two bases (cytosine & adenine), can be methylated.
- The most widely characterized is the covalent addition of CH_3 to the 5 carbon of the cytosine ring

- 5-mC methylation is seen across ~1.5% of the human genome, mostly in paired symmetrical methylation of a **CpG site**
  - *Exception: embryonic stem cells*



Epigenetics: Genomic Locations of DNA Methylation



CpG Island:
Gene promoter, 1/10

CATTCCGCCTTCTCTCCCGAGGTGGCGCGTGGGA
GGTGTTTTGCTCGGGTTCTGTAAGAATAGGCCAGG
CAGCTTCCCGCCGGGATGCGCTCATCCCCTCTCGGG
GGTTCCGCTCCCACCGCGCCGCGTTCGGCCGGTT
CCGCCTGCGAGATGTTTTCCGACGGACAATGATTC
CACTCTCGGCGCCTCCCATGTTGATCCCAGCTCCT
CTGCGGGCGTCAGGACCCCTGGGCCCCGCCCCGG
CTCCACTCAGTCAATCTTTTGTCCCCGTATAAGGCG
GATTATCGGGGTGGCTGGGGGCGGCTGATTCCGA
CGAATGCCCTTGGGGGTCACCCGGGAGGGAACTC
CGGGCTCCGGCTTTGGCCAGCCCGCACCCCTGGT
TGAGCCGGCCCGAGGGCCACCAGGGGGCGCTCG
ATGTTCCTGCAGCCCCCCGCAGCAGCCCCACTCC
CCGGCTCACCCTACGATTGGCTGGCCGCCCGGAG
CTCTGTGCTGTGATTGGTCACAGCCCGTGTCCGTC
GCGGGCGCCGGGGCGGATACGAGGTGACGCGCA
GAGGCCCAGCTCGGGGCGGTGTCCCGCGCCGGC
GACTGCGGCGGAGTTTCGCGAGGGCCGAAGCG
GGGCAGTGTGACGGCAGCGGTCCTGGGAGGCGC
CCGCGCGCGTCGGAGCAGCTCCCCGTCCTCCGCA
GCCGTCACCGCCGGCGGTCGCCGCGCCCTGGCC
TCCCGCACTCGCGCACTCCTGTCCGCCGCCCACC
GCCCACCTCCCACCTCGATGCGGTGCCGGCTGC

Non-CpG Island:
Generic sequence, 1/100

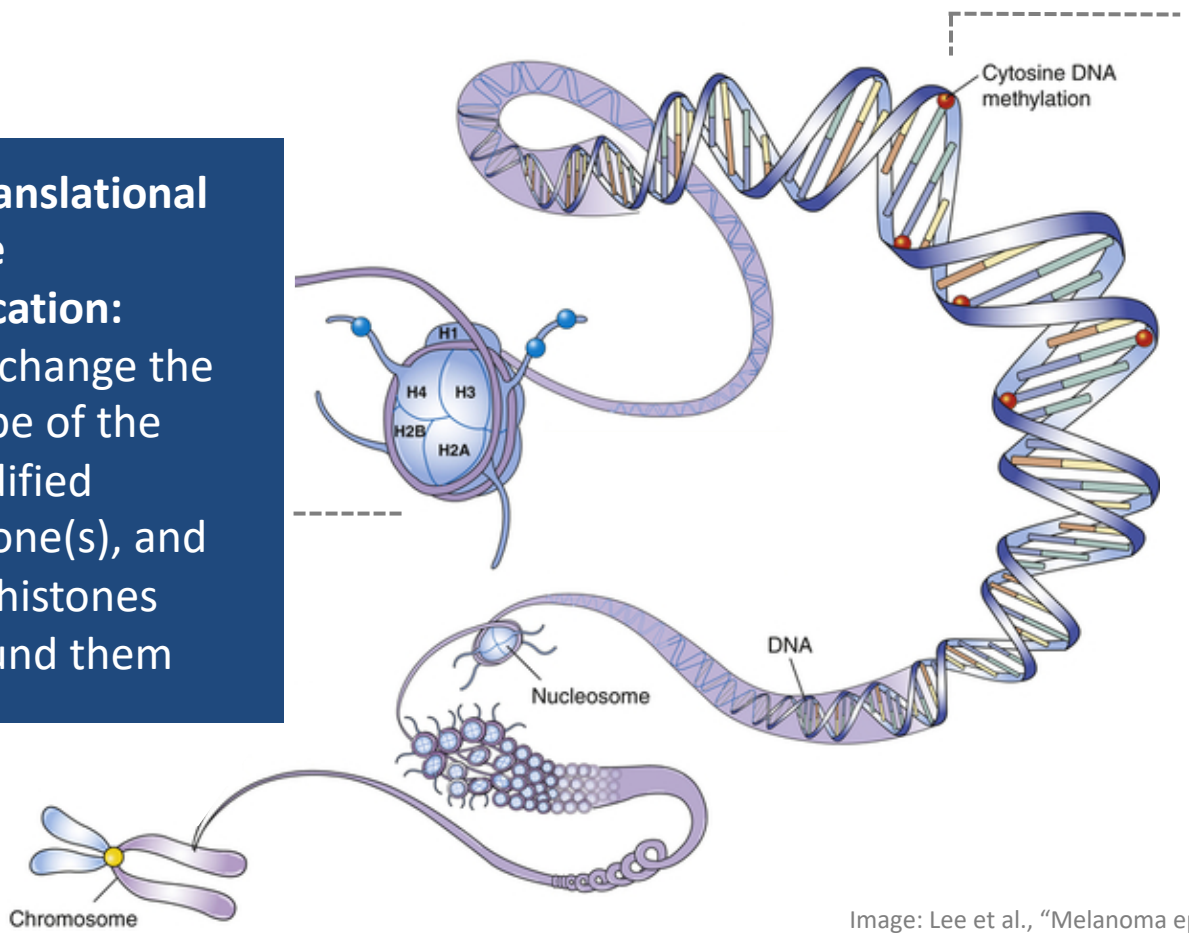
CTCTTAGTTTTGGGTGCATTTGTCTGGTCTTCCAAA
CTAGATTGAAAGCTCTGAAAAAAAAAAACTATCTTGT
GTTTCTATCTGTTGAGCTCATAGTAGGTATCCAGGA
AGTAGTAGGGTTGACTGCATTGATTTGGGACTACAC
TGGGAGTTTTCTTCGCCATCTCCCTTTAGTTTTCCT
TTTTTTCTTTCTTTCTTTTCTTTTTTTTCTTTTTTTTT
TTGAGATGTCGTCTTGCTCAGTCCCCCAGGCTGGA
GTGCAGTGGTGCGATCTTGGCTCACTGTAGCCTCC
ACCTCCCAGGTTCAAGCAATTCTACTGCCTTAGCCT
CCCGAGTAGCTGGGATTACAAGCACCCGCCACCAT
TCCTGGCTAATTTTTTTTTTTGTATTTTTAGTTGAGA
CAGGGTTTCACCATGTTGGTGATGCTGGTCTCAGA
CTCCTGGGGCCTAGCGATCCCCCTGCCTCAGCCT
CCCAGAGTGTTAGGATTACAGGCATGAGCCACTGT
ACCCGGCCTCTCTCCAGTTTCCAGTTGGAATCCAA
GGGAAGTAAGTTTAAGATAAAGTTACGATTTTGAAAT
CTTTGGATTCAGAAGAATTTGTCACCTTTAACACCT
AGAGTTGAACGTTCATACCTGGAGAGCCTTAACATT
AAGCCCTAGCCAGCCTCCAGCAAGTGGACATTGGT
CAGGTTTGGCAGGATTCGTCCCCTGAAGTGGACT
GAGAGCCACACCCTGGCCTGTCACCATACCCATCC
CCTATCCTTAGTGAAGCAAAACTCCTTTGTTCCCTT
CTCCTTCTCCTAGTGACAGGAAATATTGTGATCCTA

Epigenetics: Covalent Modifications

Both covalent mechanisms of epigenetic modification can remodel chromatin.

Post-translational histone modification:

- can change the shape of the modified histone(s), and the histones around them



DNA methylation:

- More attractive to histone-modifying enzymes
- Less transcribed

Image: Lee et al., "Melanoma epigenetics," *Laboratory Investigation, 2014.*

Epigenetics: Histone and Chromatin Modification

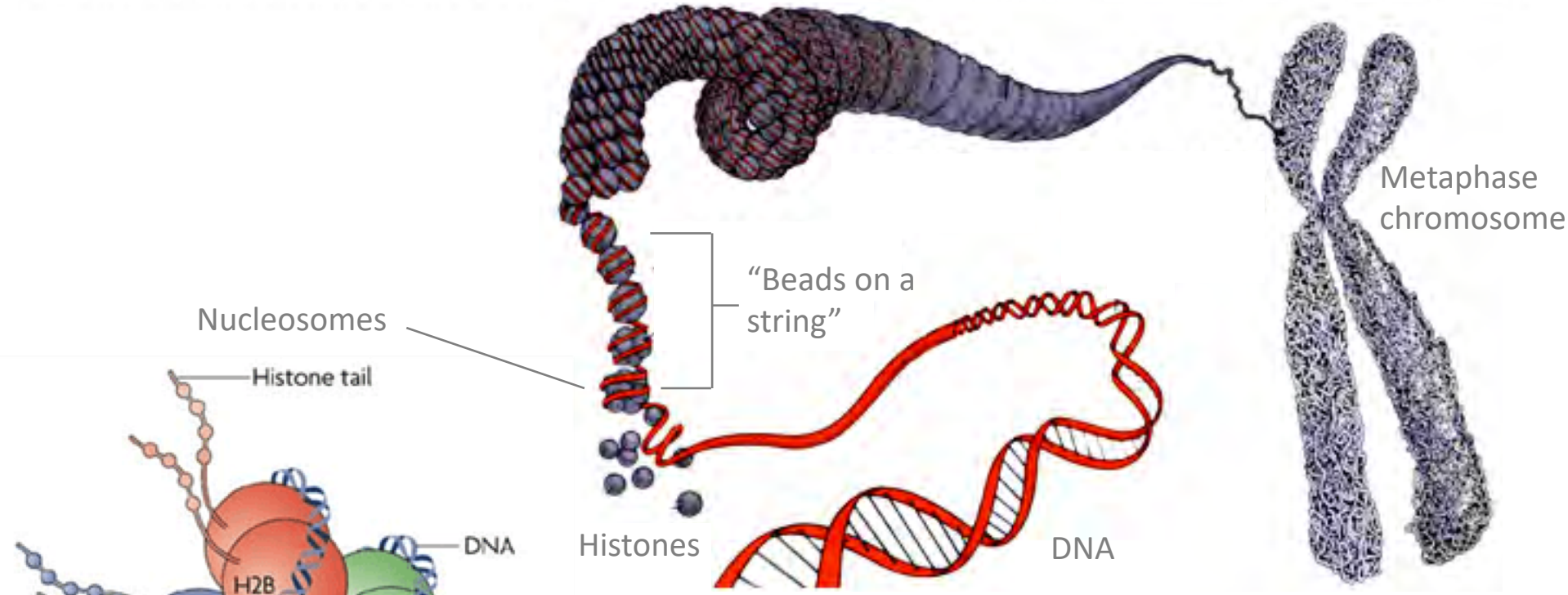


Image: Adapted from Figure 2, Creative Biomart, "*Chromosome Structure Proteins*"

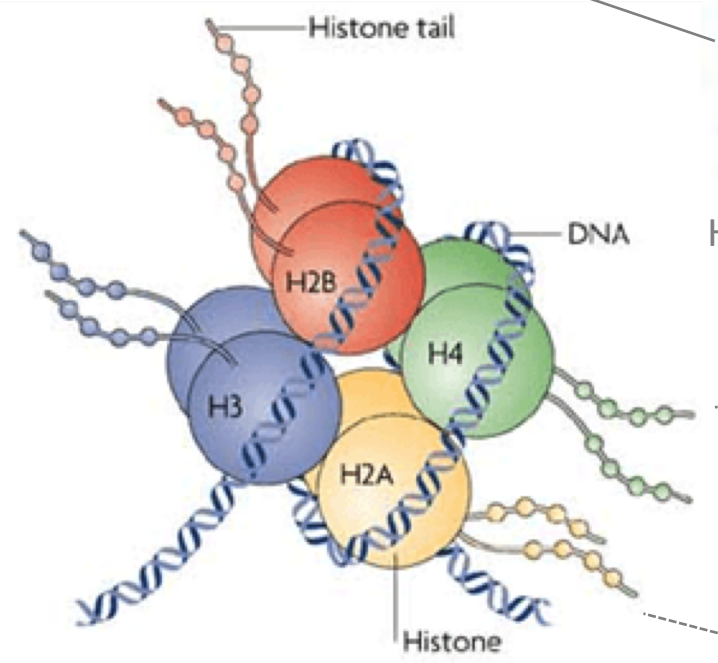
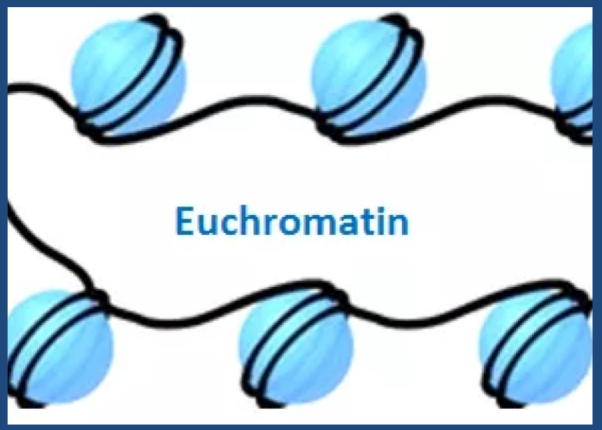


Image: Adapted from IRB Barcelona, "*Understanding the molecular origin of epigenetic markers*"

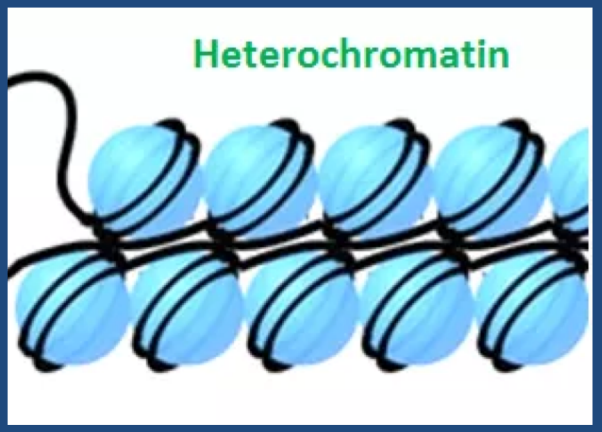
- Methylation
- Acetylation
- Phosphorylation
- Ubiquitination
- Sumoylation
- Ribosylation
- Citrullination

Epigenetics: The Histone Code

Common Histone Modifications:



------- *Tri-methylation* of histone H3 at lysines 4 (K4), 36 (K36), and 79 (K79), and a high level of histone *acetylation*



------- *Tri-methylation* of histone H3 lysine residues K9, K20, and K27

Image: Adapted from Viva Differences, "*Euchromatin Vs. Heterochromatin: What is the difference?*"

Epigenetics: RNA-Based Modifications

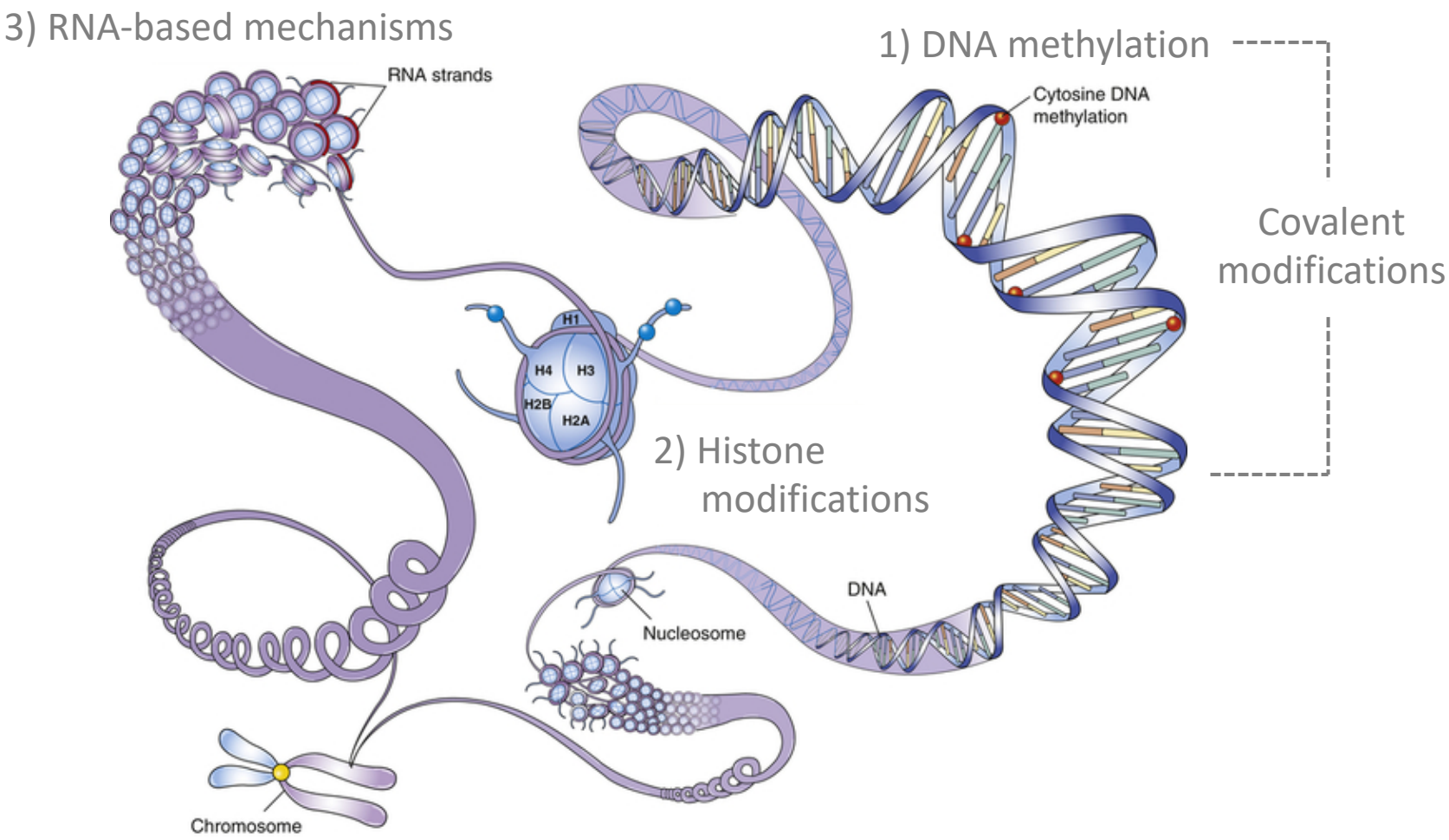


Image: Lee et al., "Melanoma epigenetics," *Laboratory Investigation, 2014.*

Epigenetics: Non-coding RNAs

Non-coding gene

- AKA: non-coding RNA (ncRNA)
- ~22 500 in humans

microRNA	– miRNA
piwi-interacting RNA	– piRNA
short interfering RNA	– siRNA
long non-coding RNA	– lncRNA

Short (<30ntds)

Long (>200ntds)

Coding gene

- AKA: Protein-coding genes
- ~20 000 in humans

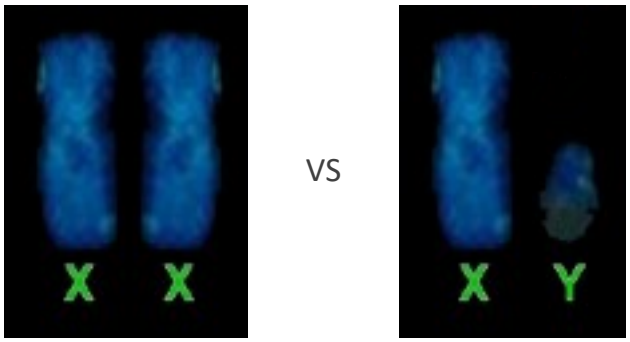
Transcription

mRNA

Translation

Protein

Epigenetics: Long ncRNAs – X Inactivation



The **X Inactivation Centre** (XIC) encodes 4 lncRNAs:

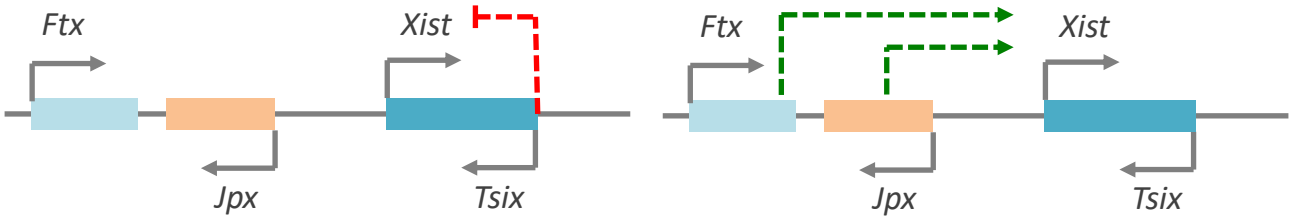
- Xist (X-inactive specific transcript)
- Tsix (the antisense transcript of Xist)
- Jpx (Just proximal to Xist)
- Ftx (Five prime to Xist)

Xa (active)

- Tsix

Xi (inactive)

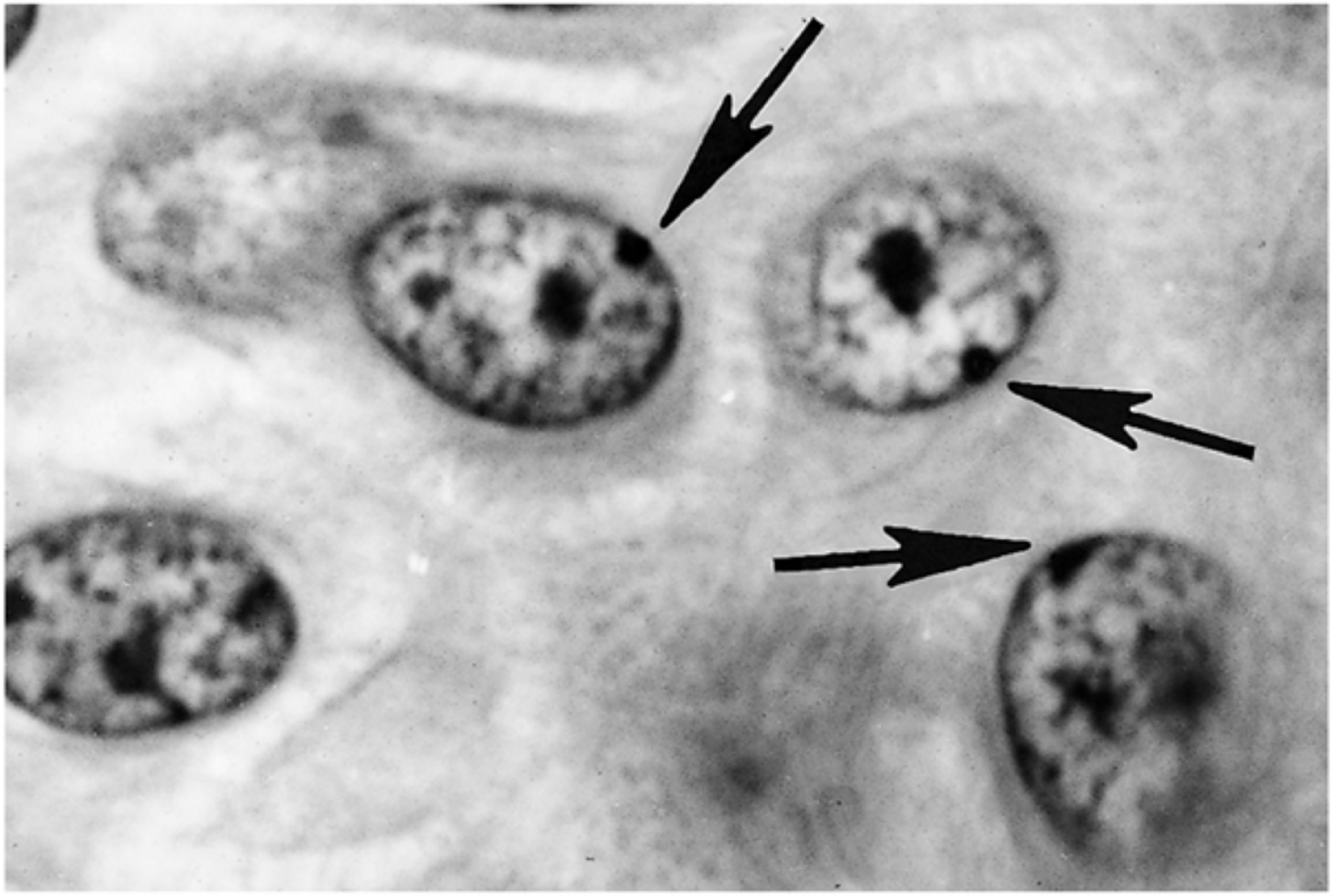
- Xist
- Jpx
- Ftx



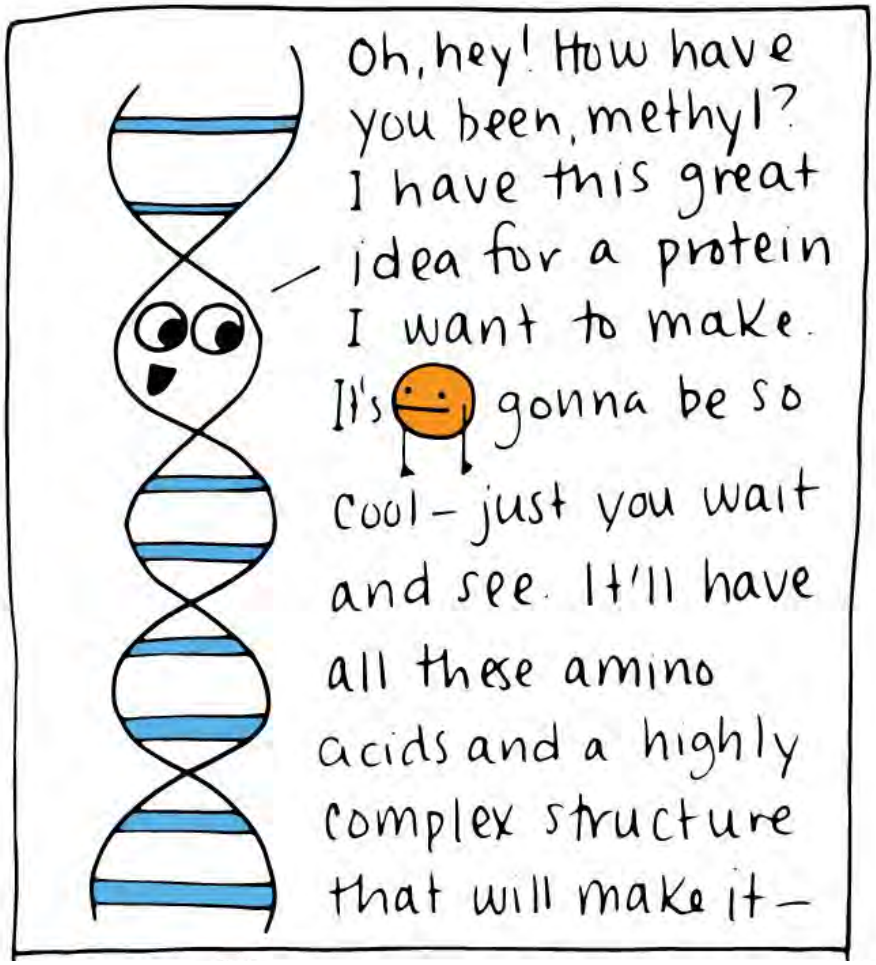
- Xist is **up-regulated** by Jpx & Ftx on Xi
- Xist is **down-regulated** by Tsix on Xa

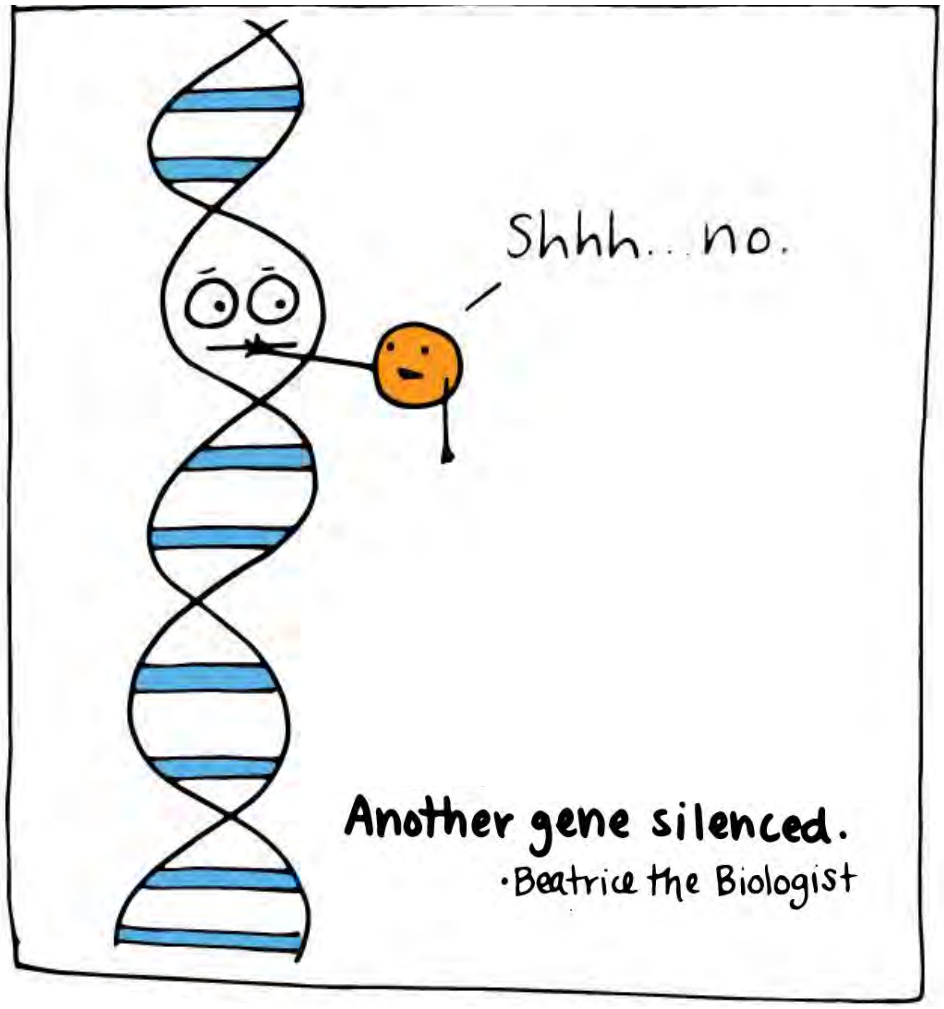
Epigenetics: Long ncRNAs – X Inactivation

Arrows – Xi Chromosomes / Barr Bodies



From N. Ason/DeHaan, Figure 9.8, *Biological World*, 1973


Oh, hey! How have you been, methyl? I have this great idea for a protein I want to make. It's gonna be so cool - just you wait and see. It'll have all these amino acids and a highly complex structure that will make it -

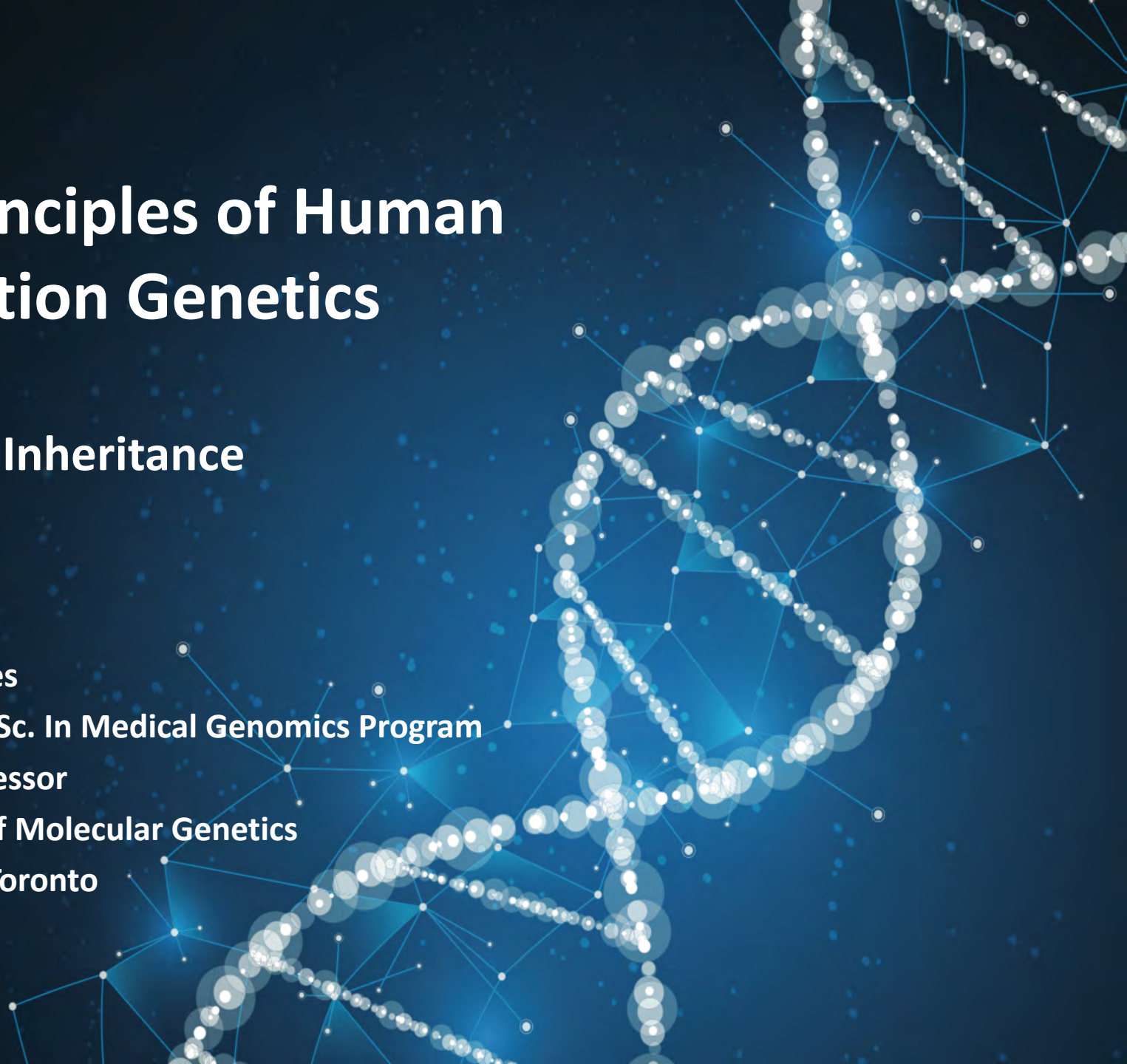

Shhh.. no.
Another gene silenced.
·Beatrice the Biologist

The Principles of Human Population Genetics

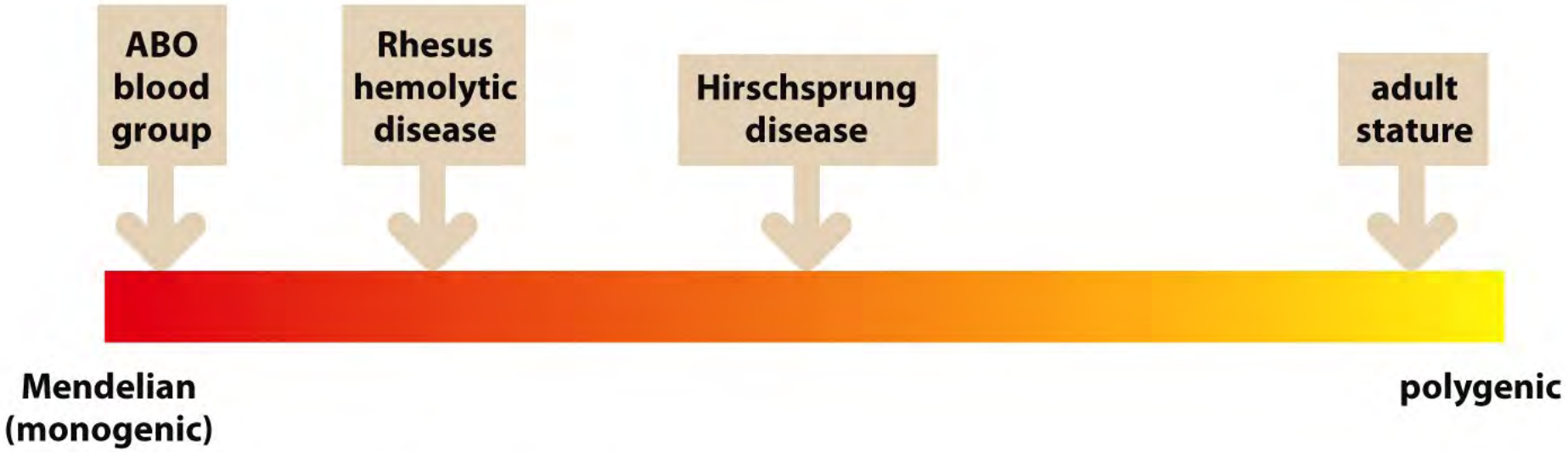
Lecture 3:

Modes of Inheritance

Dr. Erin B Styles
Director, M.H.Sc. In Medical Genomics Program
Assistant Professor
Department of Molecular Genetics
University of Toronto



Inheritance – Monogenic Versus Multifactorial



- The majority of human genetic characters are **not** Mendelian / monogenic

Figure 3.1, *Human Molecular Genetics, Fourth Edition,* 2011

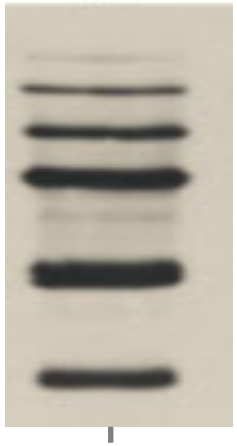
Inheritance – Monogenic Versus Multifactorial

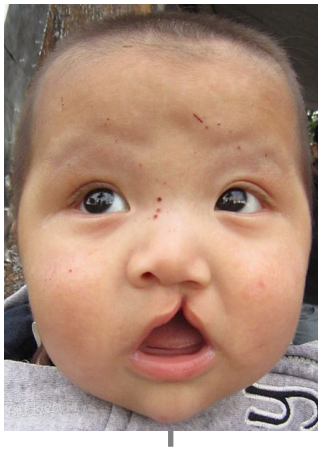
5’ … CAACATAGTGAG**A**CCCCATCTTTACAAAAT… 3’

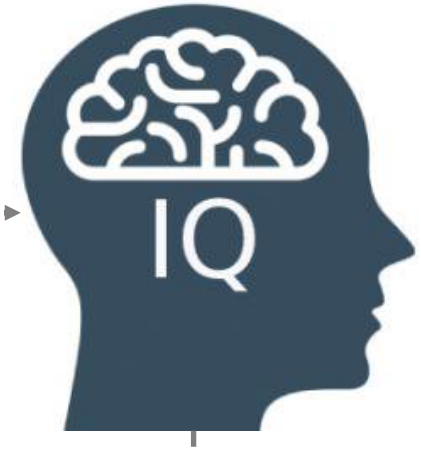
Almost always inherited in a cleanly Mendelian pattern

Small proteins

Big proteins







Usually display Mendelian pedigree patterns

Unlikely to display a Mendelian pedigree pattern

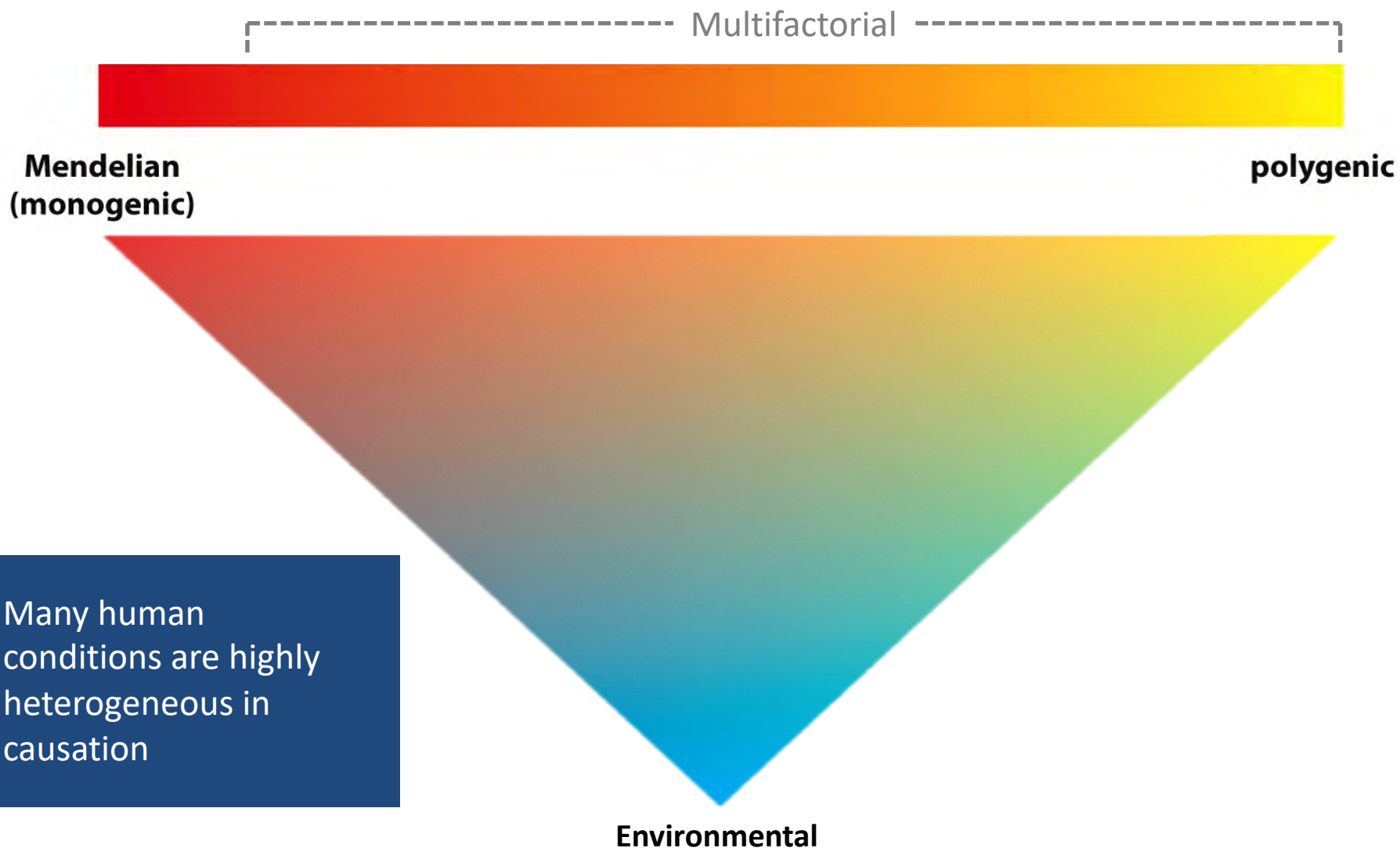
Even less likely to be Mendelian

http://www.abcam.com

User Doc James, *Wikimedia Commons,* 2010

www.proprofs.com

Inheritance – Monogenic Versus Multifactorial



Many human conditions are highly heterogeneous in causation

Adapted from Figures 3.1 and 3.15, *Human Molecular Genetics, Fourth Edition,* 2011

Mendelian Inheritance – The Basics

There are adult onset Mendelian disorders:

- <10% manifest after puberty
- ~1% occur after the reproductive period

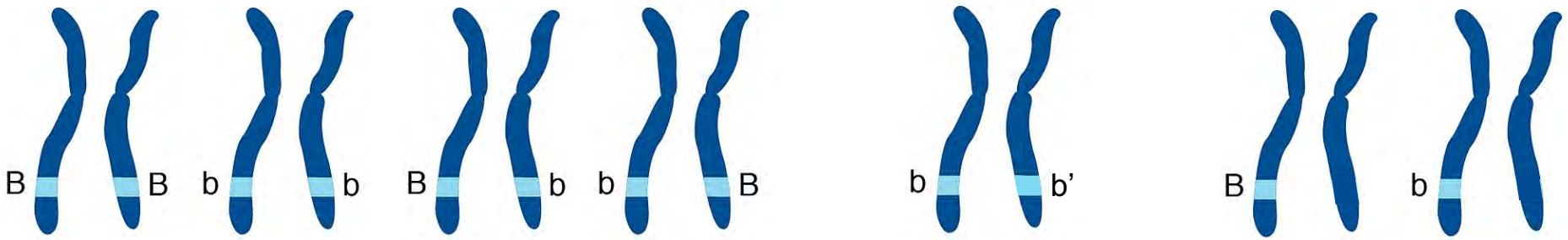
~200 Mendelian disorders have phenotypes that include common adult illnesses

Single gene disorders affect children disproportionately:

- 1 in 300 neonates
- ~7% of pediatric hospitalizations

Image: geneticdisordersuk.org

Mendelian Inheritance – Important Terminology



Homozygous:

Both alleles are the same at a given locus

Heterozygous:

Different alleles (one mutant, one wild-type) are present at a given locus

Compound Heterozygous:

Two different mutant alleles are present at a given locus

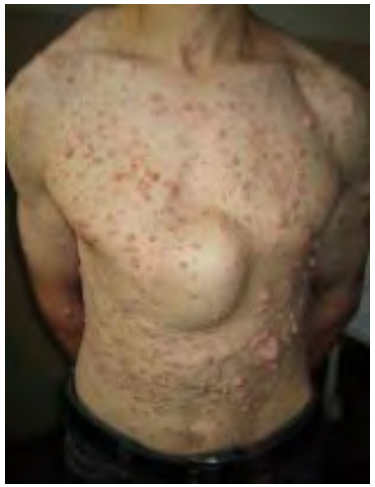
Hemizygous:

Only one allele is present at a given locus

Adapted from User Was a bee, *Wikimedia Commons,* 2016

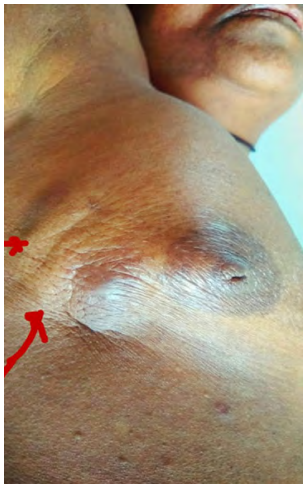
Mendelian Inheritance – Important Terminology

Penetrance: The proportion of individuals with a specific genotype who manifest that genotype at the phenotypic level.



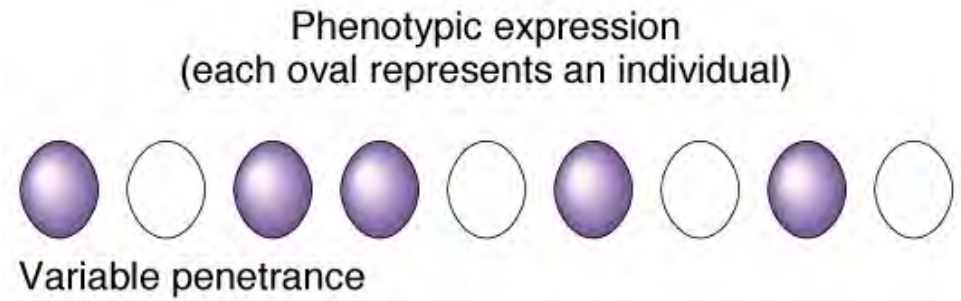
Neurofibromatosis type 1 (NF-1):

Caused by microdeletion in the *NF1* gene, exhibits 100% penetrance – all individuals with this mutation will show symptoms.



Familial breast cancer:

Can be caused by mutation in the *BRCA1* gene, exhibits reduced (80%) penetrance.

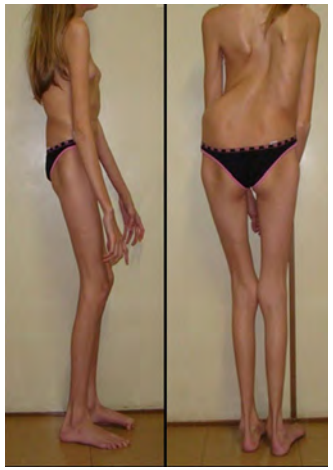


Left: Mika et al., *Am J Case Rep*, 2010; Right: Adapted from S. Moolchandani, healthline.com;
Bottom: Adapted from Figure 6-22, *Introduction to Genetic Analysis, Tenth Edition* 2012

Mendelian Inheritance – Important Terminology

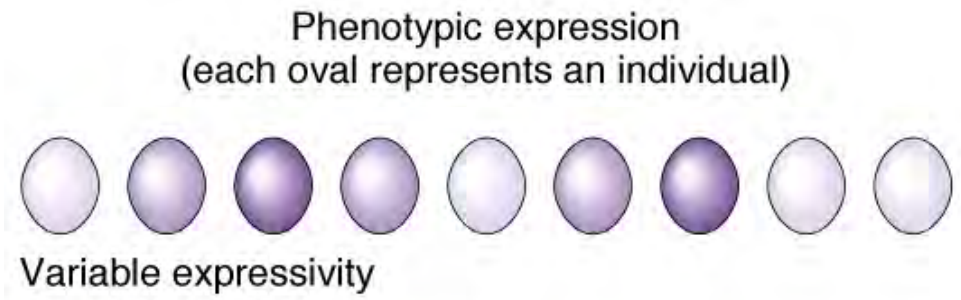
Penetrance: The proportion of individuals with a specific genotype who manifest that genotype at the phenotypic level.

Expressivity: The degree to which a particular genotype is expressed in the phenotype of a particular individual.



Marfan Syndrome:

Highly variable expressivity – people carrying a mutation in the *FBN1* gene may exhibit severe scoliosis, life threatening organ and heart defects, and severe ocular defects, *or* they may simply be very tall and thin, with elongated limbs and digits.



Left: Adapted from User CFCF, *Wikimedia Commons,* 2015; Right: The Marfan Foundation, 2014;
Bottom: Adapted from Figure 6-22*, Introduction to Genetic Analysis, Tenth Edition* 2012

Mendelian Inheritance – Important Terminology

Penetrance *describes what happens in a population.*
Expressivity *describes what happens in a particular individual.*

Treacher-Collins Syndrome:

Caused by a mutation in the *TCOF1* gene. In its most obvious form (*the child in the image*), the facial features are easy to recognize:

- Small mandible, palpebral fissures slant down, ear microtia, lower eyelid defect, hearing impairment (not visible)

However, the child's mother also harbors the *TCOF1* mutation – it would be challenging to say whether or not she is demonstrating *non-penetrance*, or *very low expressivity.*

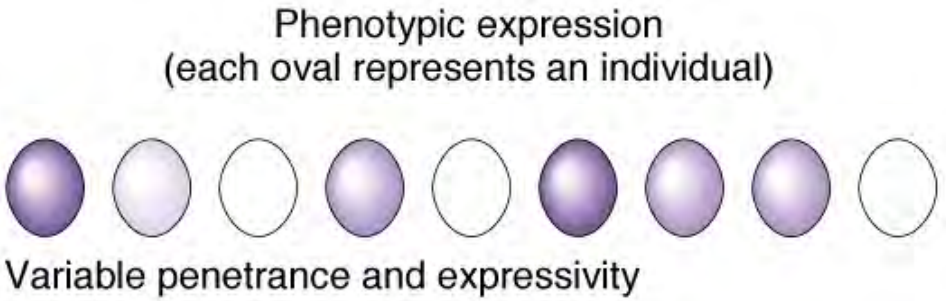
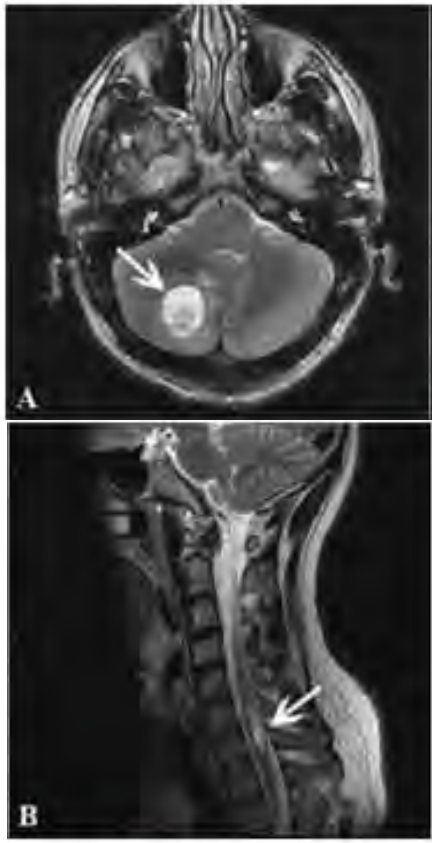


Figure 6.7, *Emery's Elements of Medical Genetics, Fifteenth Edition,* 2017;
Bottom: Adapted from Figure 6-22, *Introduction to Genetic Analysis, Tenth Edition* 2012

Mendelian Inheritance – Important Terminology

Pleiotropy: A single gene / gene pair often produces multiple, diverse phenotypic effects in multiple organ systems, with a variety of signs and symptoms occurring at different points during the life span



Von Hippel-Lindau Syndrome:

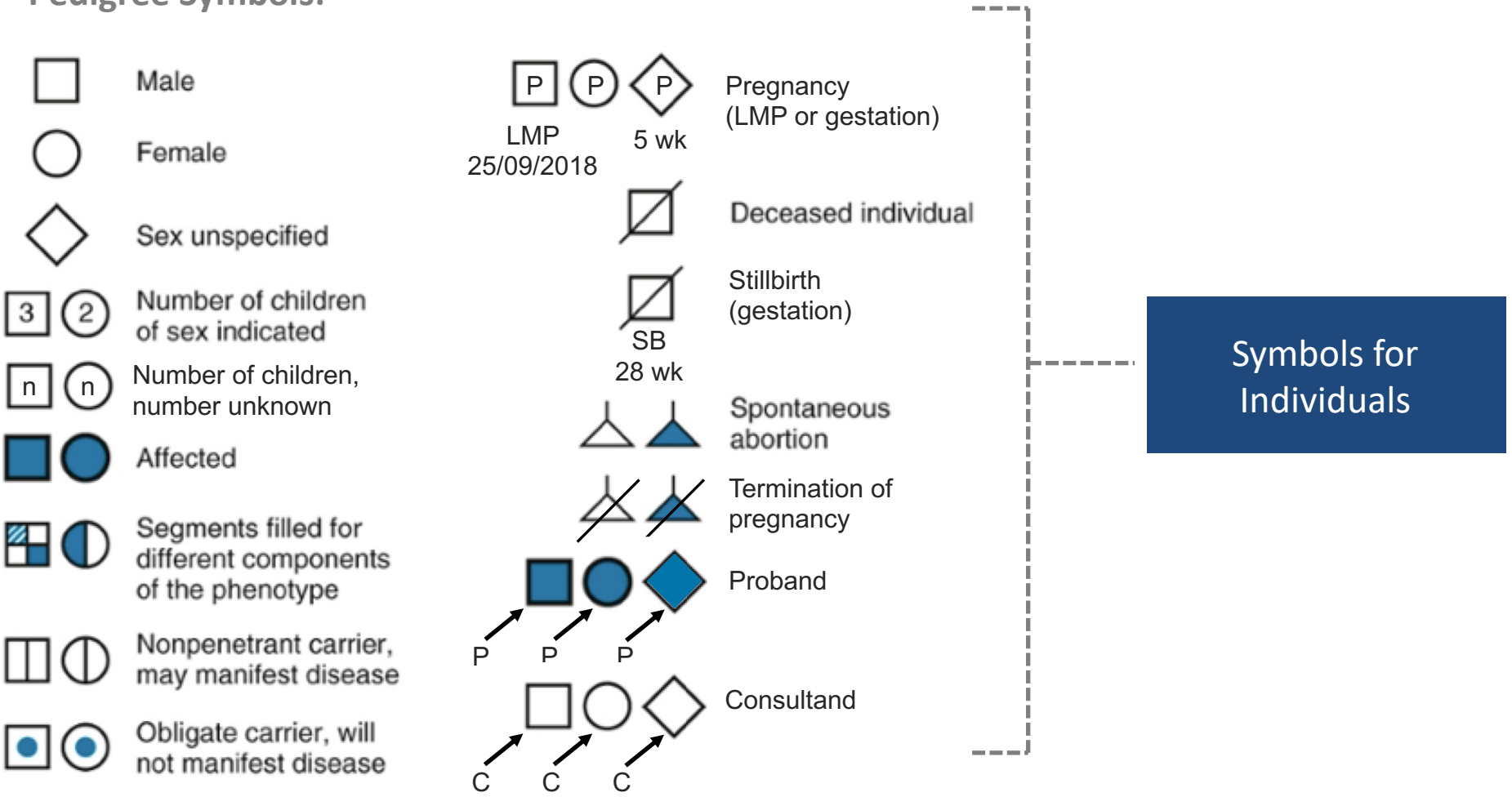
Caused by mutations in the *VHL* gene, and exhibits a high degree of phenotypic pleiotropy. Individuals with VHL syndrome may have:

- Hemangioblastomas of the brain (*top panel, A*), spinal cord (*bottom panel, B*) and / or retina
- Endolymphatic tumors of the inner ear
- Tumors of the epididymis (male) / broad ligament of the uterus (female)
- Renal and / or pancreatic cysts
- Renal cell carcinoma
- Pheochromocytoma

Adapted from J. Cochrane Miller, Radiology Rounds, Massachusetts General Hospital, Dept. of Radiology, 2009

Mendelian Inheritance – Pedigree Symbols and Annotation

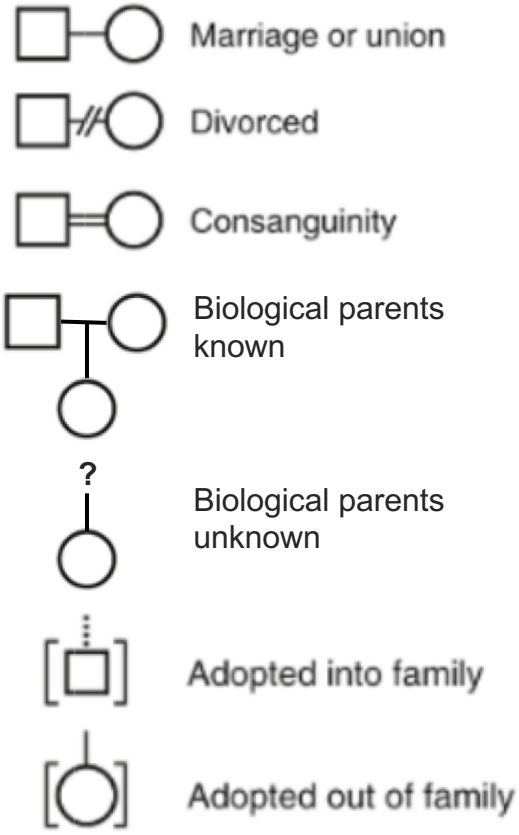
Pedigree Symbols:

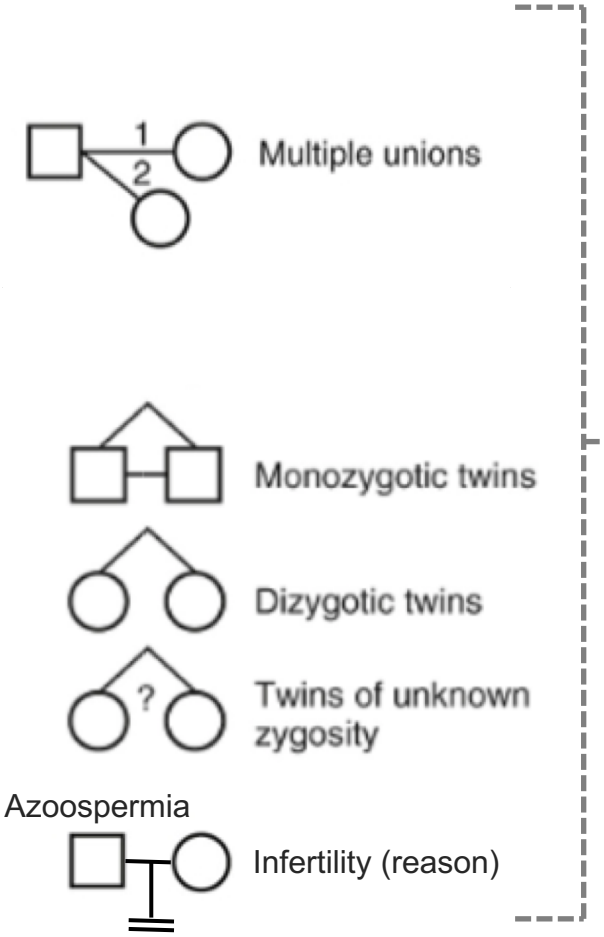


Adapted from Figure 3.2, *Human Molecular Genetics, 4th Edition,* 2011

Mendelian Inheritance – Pedigree Symbols and Annotation

Pedigree Symbols:

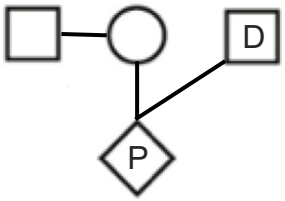




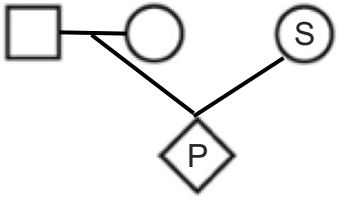
Symbols for Relationships

Mendelian Inheritance – Pedigree Symbols and Annotation

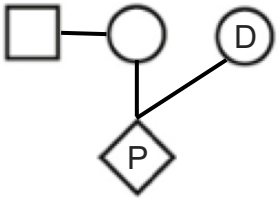
Pedigree Symbols:



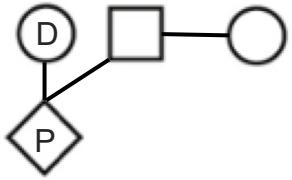
Sperm donation



Surrogate mother



Ovum donation



Surrogate ovum donation

Symbols for Assisted Reproductive Scenarios

Mendelian Inheritance – Types of Inheritance

There are **FIVE** archetypal Mendelian pedigree patterns:

1. Autosomal dominant
2. Autosomal recessive
3. X-linked dominant
4. X-linked recessive
5. Y-linked

Dominant and Recessive:

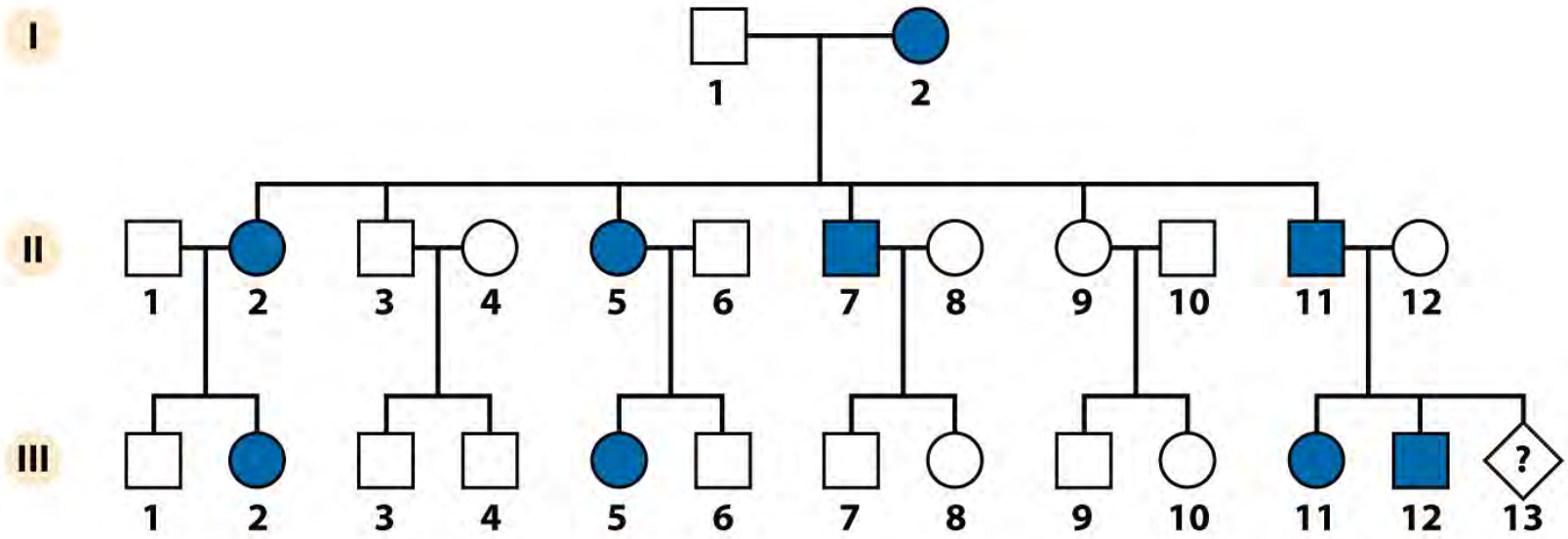
- If a character is manifested in a *heterozygous individual*, it is dominant
- If a character is only manifested in an *homozygous individual*, it is recessive

Dominance and recessiveness are properties of genetic characters, NOT genes.

Example:

- Sickle-cell anemia is a recessive disorder, because only Hb^S homozygotes manifest it
- Red blood cell sickling trait is a dominant trait, since Hb^S heterozygotes exhibit this phenotype

Mendelian Inheritance – Autosomal Dominant Inheritance



- An affected individual usually has at least one affected parent
- A child with only one affected parent has a 50% chance of being affected
- It equally affects either sex

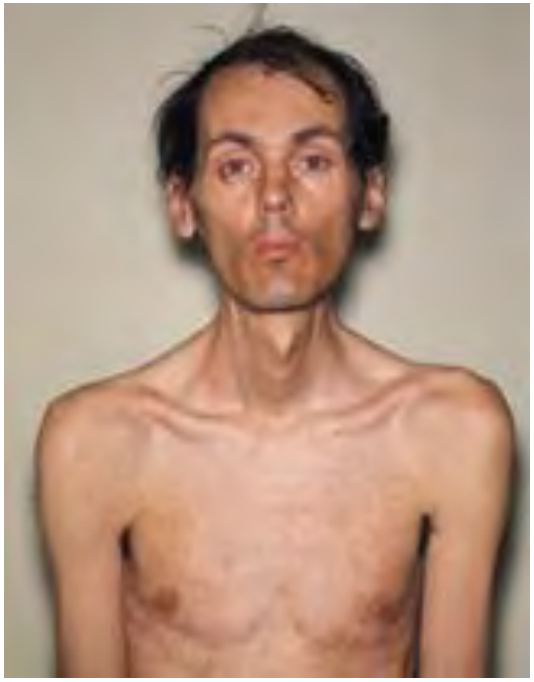
What is the likelihood that Individual III_{13} will be affected?

1 in 2, or 50%

Figure 3.3, *Human Molecular Genetics, Fourth Edition,* 2011

Mendelian Inheritance – Autosomal Dominant Inheritance

Example:



Myotonic Dystrophy

- There are two types of MD:
  - MD1, caused by a mutation in the *DMPK* gene
  - MD2, caused by a mutation in the *CNBP* gene
- MD1 has a frequency of:
  - ~1 in 8000 in the general population
  - ~1 in 550 in certain regions of Northeastern Quebec, in Canada

User Filip em, *Wikimedia Commons,* 2008

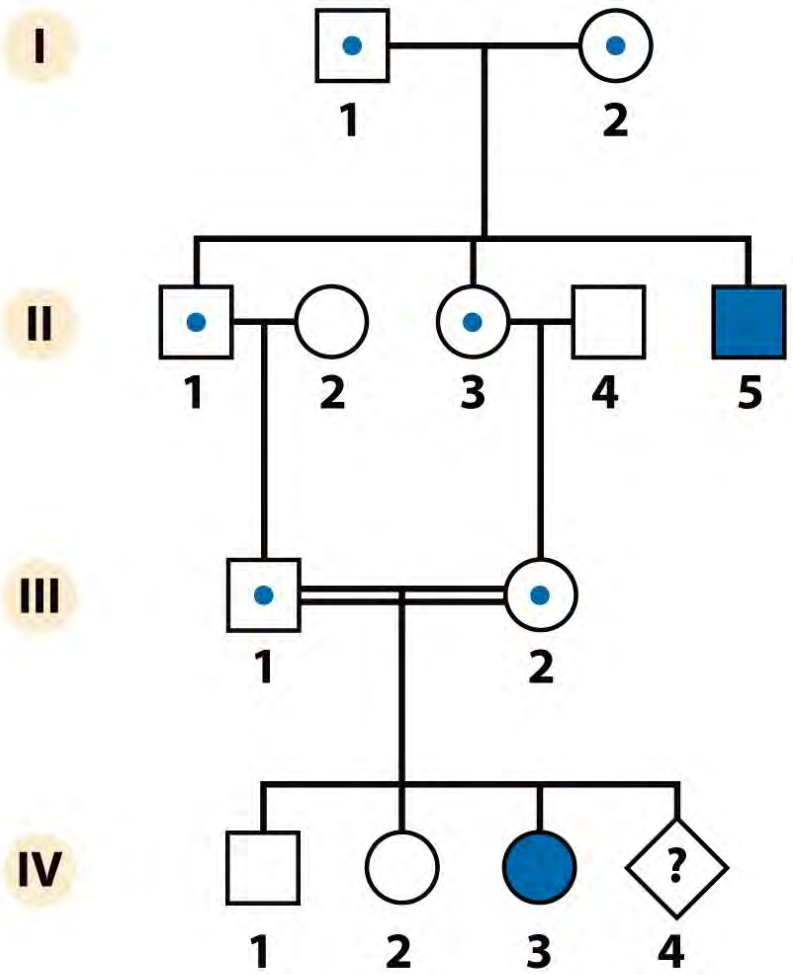
Mendelian Inheritance – Autosomal Dominant Inheritance

Affected x Unaffected		Parent 2 Genotype d/d Gametes		Risk for Disease
		d	d	
Parent 1 Genotype D/d Gametes	D	D/d	D/d	½ Affected (D/d) ½ Unaffected (d/d)
	d	d/d	d/d	

Affected x Affected		Parent 2 Genotype D/d Gametes		Risk for Disease
		D	d	
Parent 1 Genotype D/d Gametes	D	D/D	D/d	¾ Affected (D/D or D/d) ¼ Unaffected (d/d)
	d	D/d	d/d	

Adapted from Chapter 7, *Genetic Medicine, Eighth Edition,* 2016

Mendelian Inheritance – Autosomal Recessive Inheritance



- Affected people are often born to unaffected parents
- Parents of affected people are usually asymptomatic carriers
- There is an increased incidence of parental consanguinity
- It equally affects either sex
- After the birth of an affected child, each subsequent child has a 25% chance of being affected

What is the likelihood that Individual IV_4 will be affected?

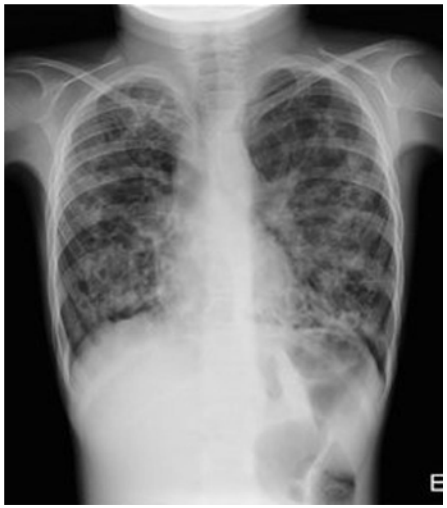
1 in 4, or 25%

Figure 3.4, *Human Molecular Genetics, Fourth Edition,* 2011

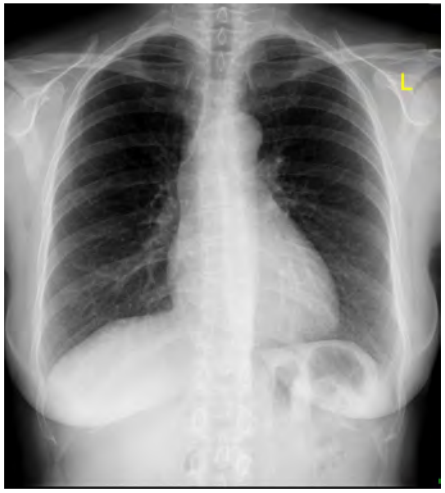
Mendelian Inheritance – Autosomal Recessive Inheritance

Example:

CF lungs, x-ray



Norma lungs, x-ray



Cystic Fibrosis (AKA Mucoviscidosis)

- Caused by any of a number of mutations in the *CFTR* gene (there are currently >1500 known *CFTR* mutations that can contribute to CF)
  - The most common autosomal recessive disorder in white children (~1 in 2000)
  - ~1 in 23 white individuals is a silent carrier
- Symptoms:
  - Meconium ileus
  - Liver disease
  - Pancreatic insufficiency
  - Pulmonary disease
  - Infertility
  - Carcinoma

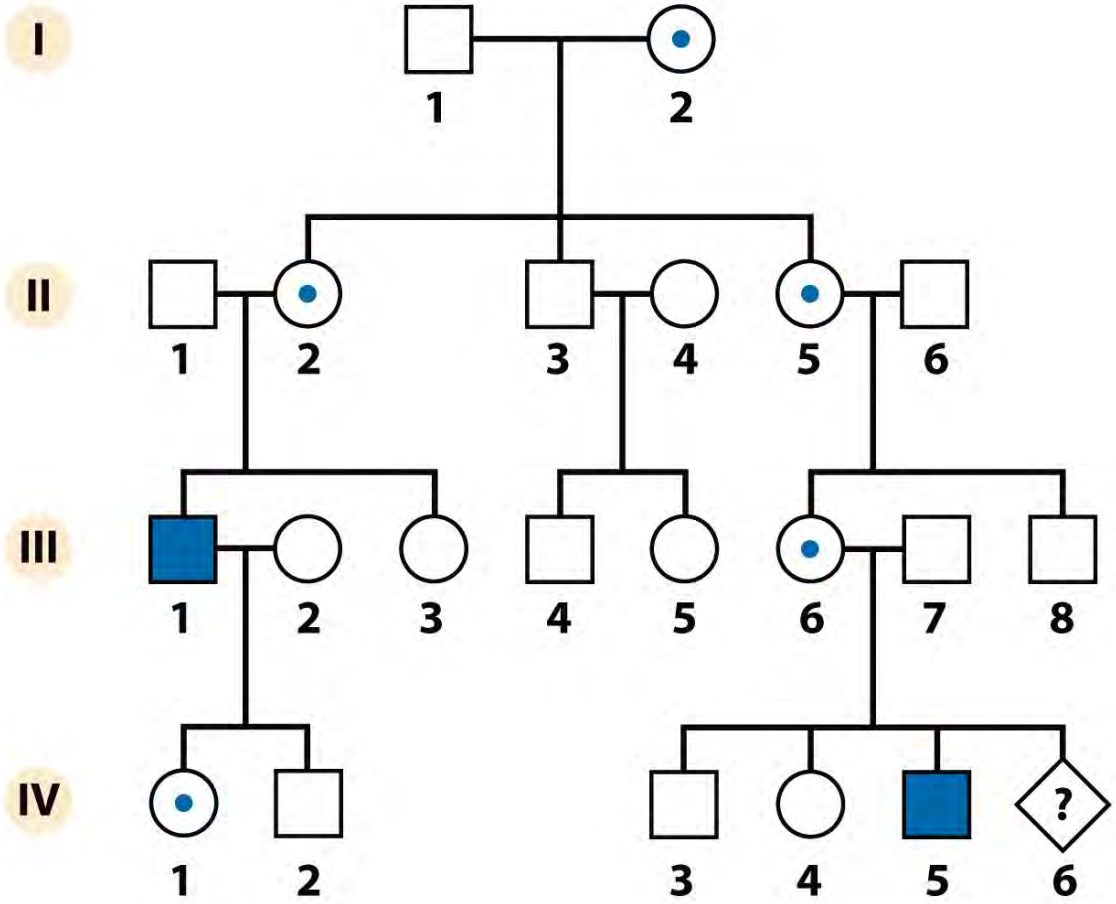
Top: Case courtesy of Dr Henry Knipe, Radiopaedia.org, rID: 31525; Bottom: Case courtesy of Dr Charudutt Jayant Sambhaji, Radiopaedia.org, rID: 5922

Carrier x Carrier		Parent 2 Genotype C/c Gametes		Risk for Disease
		C	c	
Parent 1 Genotype C/c Gametes	C	C/C	C/c	¼ Unaffected (C/C) ½ Unaffected carriers (C/c) ¼ Affected (c/c)
	c	C/c	c/c	

Carrier x Affected		Parent 2 Genotype c/c Gametes		Risk for Disease
		c	c	
Parent 1 Genotype C/c Gametes	C	C/c	C/c	½ Unaffected carriers (C/c) ½ Affected (c/c)
	c	c/c	c/c	

Affected x Affected		Parent 2 Genotype c/c Gametes		Risk for Disease
		c	c	
Parent 1 Genotype c/c Gametes	c	c/c	c/c	All Affected (c/c)
	c	c/c	c/c	

Mendelian Inheritance – X-Linked Recessive Inheritance



- Affects mainly males
- Affected males are usually born to unaffected parents (the mother is usually an asymptomatic carrier, who may have affected relatives)
- Affected females usually have an affected father and a mother who is an asymptomatic carrier
- There is no male-to-male transmission in the pedigree

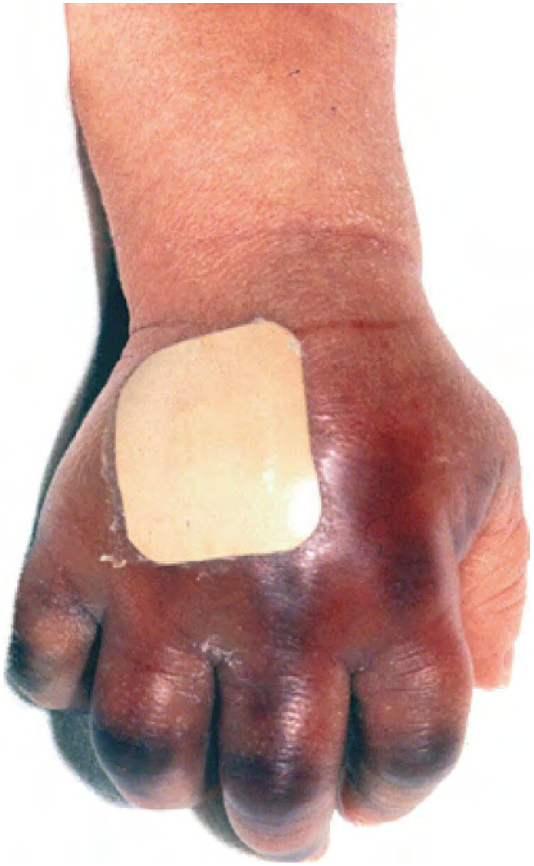
What is the likelihood that Individual IV_6 will be affected if male? If female?

1 in 2, or 50% if male; ~0% if female

Figure 3.5, *Human Molecular Genetics, Fourth Edition,* 2011

Mendelian Inheritance – X-Linked Recessive Inheritance

Example:



Hemophilia

- There are two types of Hemophilia that have an X-linked recessive mode of inheritance:
  - A (Classic), caused by mutation in the *F8* gene
  - B (Christmas disease), caused by mutation in the *F9* gene
- Symptoms:
  - Blood clotting impairment, caused by a deficiency in coagulation factor VIII (Type A) / IX (Type B)
  - Joint swelling / pain
  - Muscle hemorrhage
  - Hematuria
  - Intracranial hemorrhage

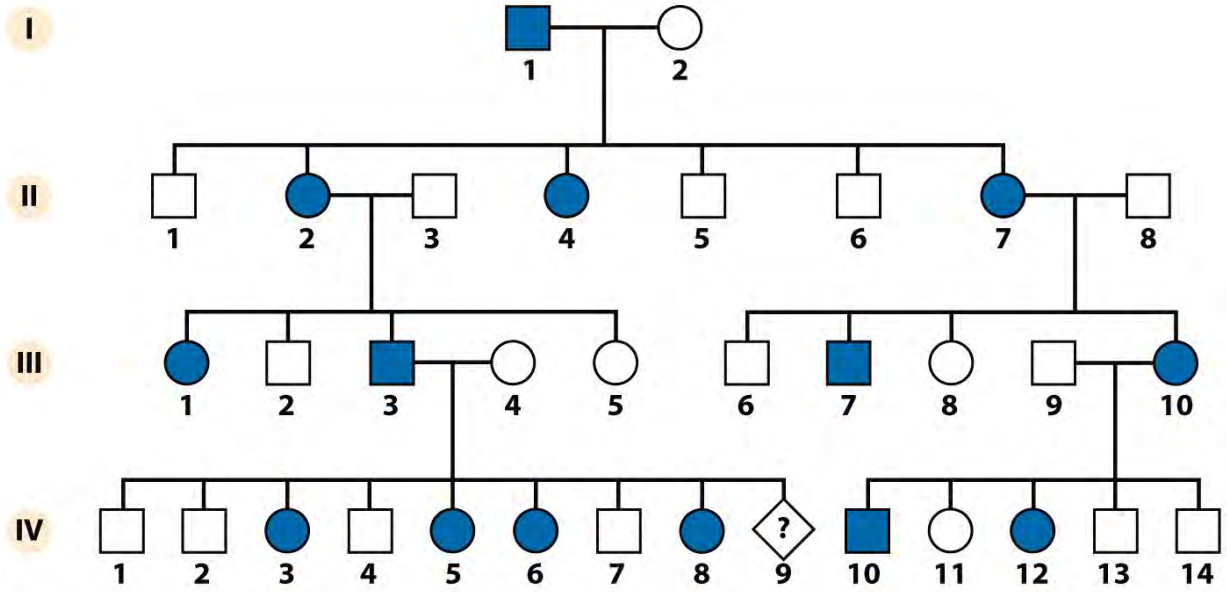
Chalmers, *Blood Reviews*, 2004

Mendelian Inheritance – X-Linked Recessive Inheritance

Affected Male x Non-carrier Female		Female Genotype X_H/X_H Gametes		Risk for Disease
		X_H	X_H	
Male Genotype X_h/Y Gametes	X_h	X_H/X_h	X_H/X_h	All females are carriers (X_H/X_h) All males unaffected (X_H/Y)
	Y	X_H/Y	X_H/Y	
Unaffected Male x Carrier Female		**Female Genotype X_H/X_h Gametes**		**Risk for Disease**
		X_H	X_h	
Male Genotype X_H/Y Gametes	X_H	X_H/X_H	X_H/X_h	¼ Non-carrier female (X_H/X_H) ¼ Carrier females (X_H/X_h) ¼ Unaffected male (X_H/Y) ¼ Affected male (X_h/Y)
	Y	X_H/Y	X_h/Y	

Adapted from Chapter 7, *Genetic Medicine, Eighth Edition,* 2016

Mendelian Inheritance – X-Linked Dominant Inheritance



What is the likelihood that Individual IV_9 will be affected if male? If female?

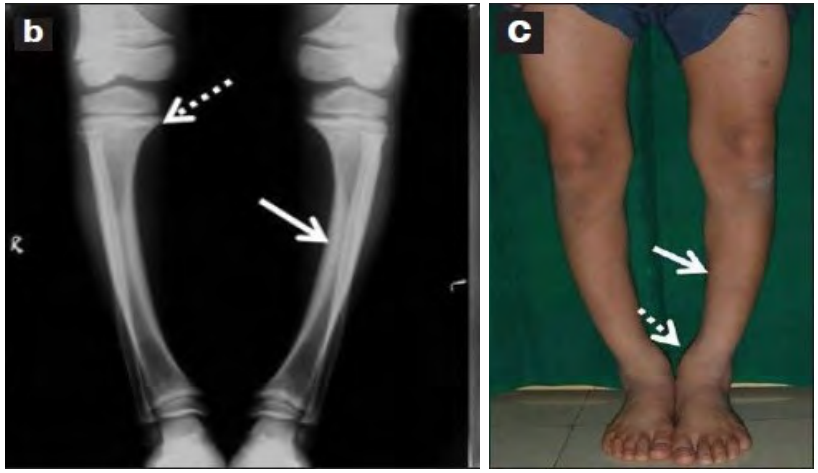
0% if male;
100% if female

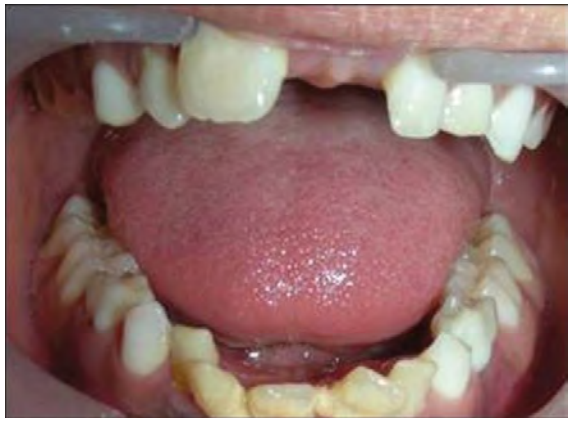
- It affects either sex, but more females than males
- An affected individual usually has at least one affected parent
- Females are usually affected more mildly and more variably than males
- The child of an affected female has a 50% chance of being affected
- All daughters and no sons of an affected male will be affected

Figure 3.6, *Human Molecular Genetics, Fourth Edition,* 2011

Mendelian Inheritance – X-Linked Dominant Inheritance

Example:





Hypophosphatemic Rickets (Vitamin D-Resistant Rickets)

- Caused by mutation in the *PHEX* gene, which impairs the ability of the kidney tubules to reabsorb filtered phosphate
- Symptoms:
  - Rickets with bone deformities
  - Short stature
  - Dental abnormalities

Sattur et al., *J Indian Soc Pedod Prev Dent*, 2010

Mendelian Inheritance – X-Linked Dominant Inheritance

Unaffected Male x Affected Female		Female Genotype X_R/X_r Gametes		Risk for Disease
		X_R	X_r	
Male Genotype X_r/Y Gametes	X_r	X_R/X_r	X_r/X_r	¼ Affected female (X_R/X_r) ¼ Unaffected female (X_r/X_r) ¼ Affected male (X_R/Y) ¼ Unaffected male (X_r/Y)
	Y	X_R/Y	X_r/Y	

Affected Male x Non-carrier Female		Female Genotype X_r/X_r Gametes		Risk for Disease
		X_r	X_r	
Male Genotype X_R/Y Gametes	X_R	X_R/X_r	X_R/X_r	All females affected (X_R/X_r) All males unaffected (X_r/Y)
	Y	X_r/Y	X_r/Y	

Adapted from Chapter 7, *Genetic Medicine, Eighth Edition,* 2016

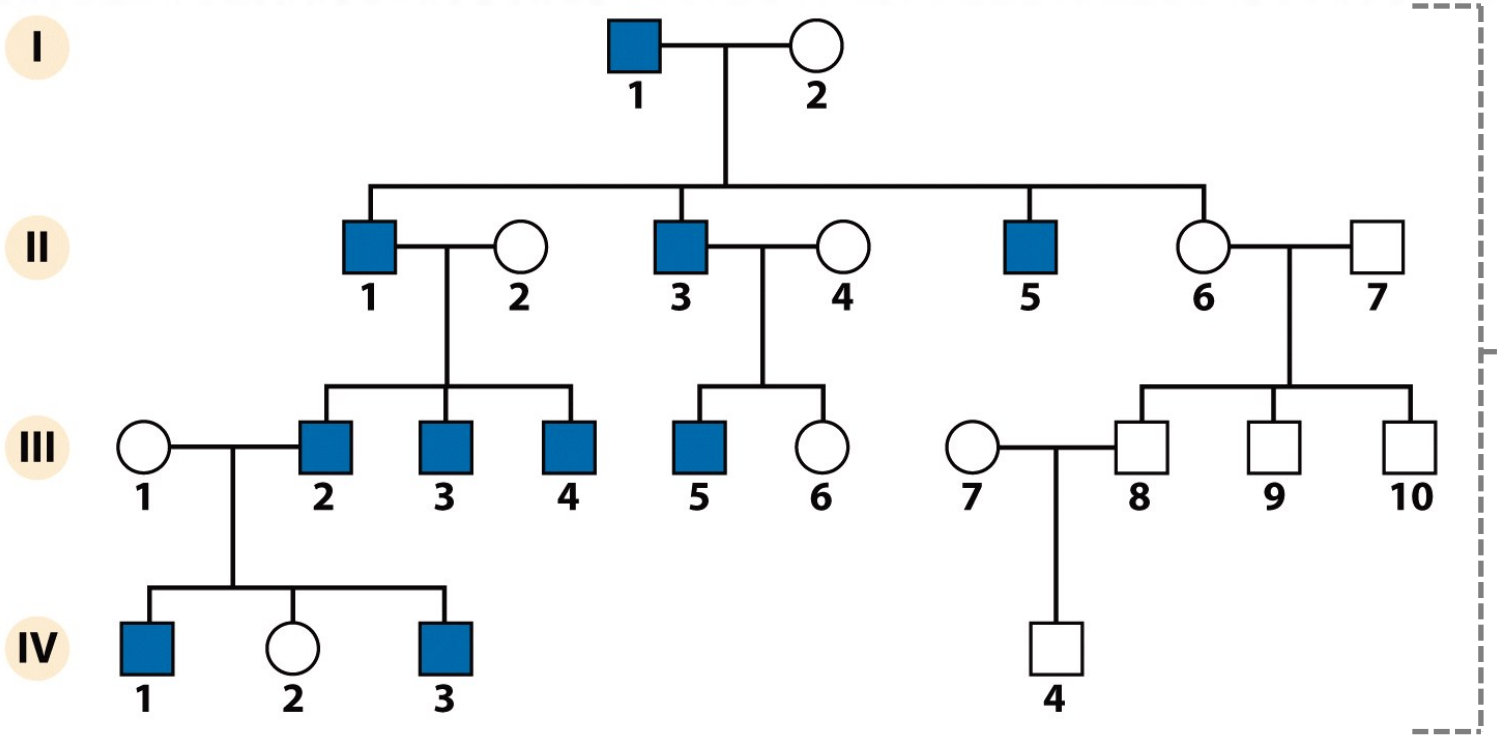
Mendelian Inheritance – A Note About X-Linked Disorders

- ~1/3 of X-linked disorders are penetrant in *some* but *not all* female heterozygotes
- Many disorders that *can* be classified as dominant or recessive show incomplete penetrance that varies due to patterns of random X-inactivation

Clinical expression of an X-linked condition therefore does not depend strictly on the particular gene involved – some geneticists have recommended dispensing with the terms **recessive** and **dominant** with respect to X-linked conditions altogether.

However, these terms are still widely used in clinical practice and in literature, so we will continue to use them, understanding that they describe extremes on a spectrum of penetrance and expressivity in female carriers of X-linked diseases.

Mendelian Inheritance – Y-Linked Inheritance



- It affects only males
- Affected males always have an affected father
- All sons of an affected man are affected

Figure 3.7, *Human Molecular Genetics, Fourth Edition,* 2011

Mendelian Inheritance – Y-Linked Inheritance

There are currently no known human conditions that give a stereotypical Y-linked pedigree.

Ichthyosis Hystrix Gravior, Lambert Type – Porcupine Man Disorder

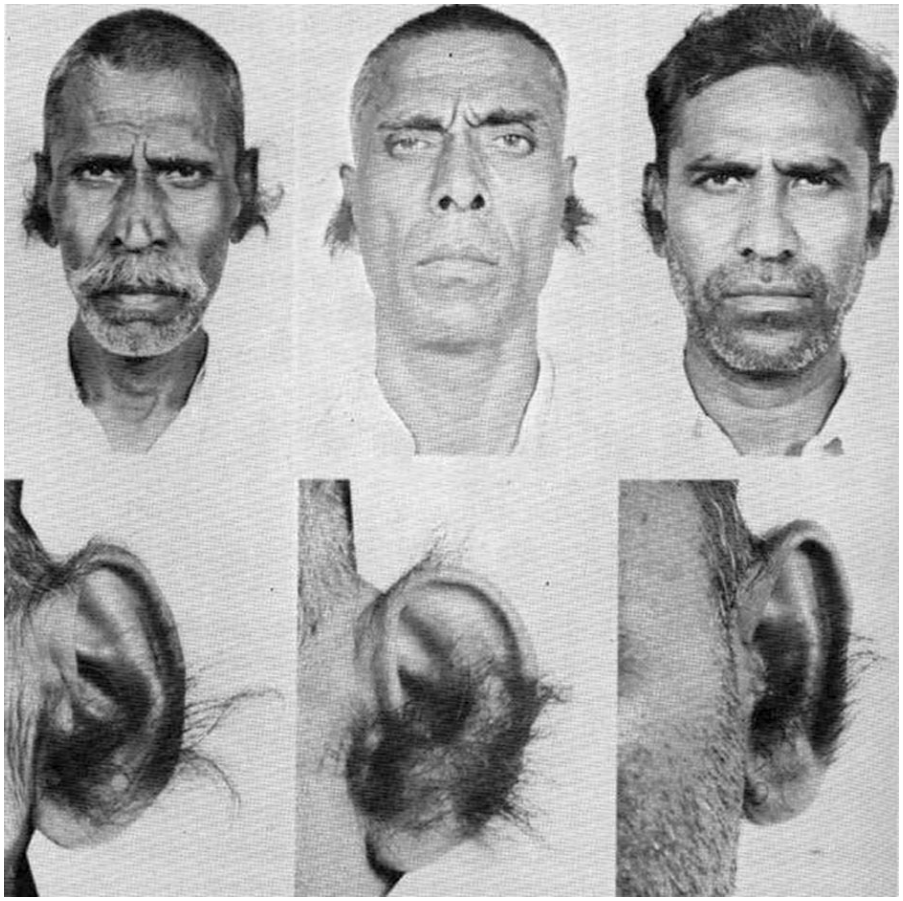
- Originally suggested to be Y-linked
- Based on a famous case featuring Edward Lambert (the original Porcupine Man), and eleven of his descendants with the same clinical presentation
- No cases of this disease are currently known

This is now thought to be an autosomal dominant disorder

User Filip em, *Wikimedia Commons,* 2007

Mendelian Inheritance – Y-Linked Inheritance

There are currently no known human conditions that give a stereotypical Y-linked pedigree.



Al Aboud, *Our Dermatology Onine*, 2014

Hypertrichosis Pinnae Auris –
Hairy Ears

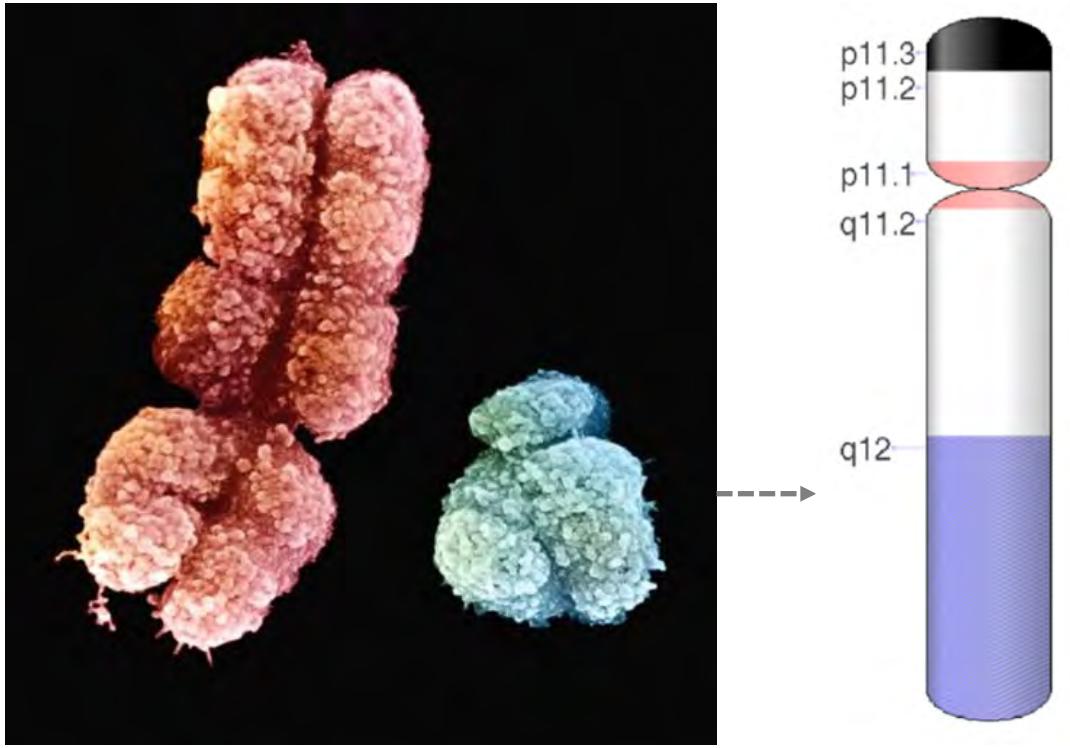
- Suggested to be Y-linked
- Y-chromosomal DNA studies in a Southern Indian population indicate that phenotype is not Y-linked in this group

It's therefore unlikely to be Y-linked in any population

Mendelian Inheritance – Y-Linked Inheritance

There are currently no known human conditions that give a stereotypical Y-linked pedigree.

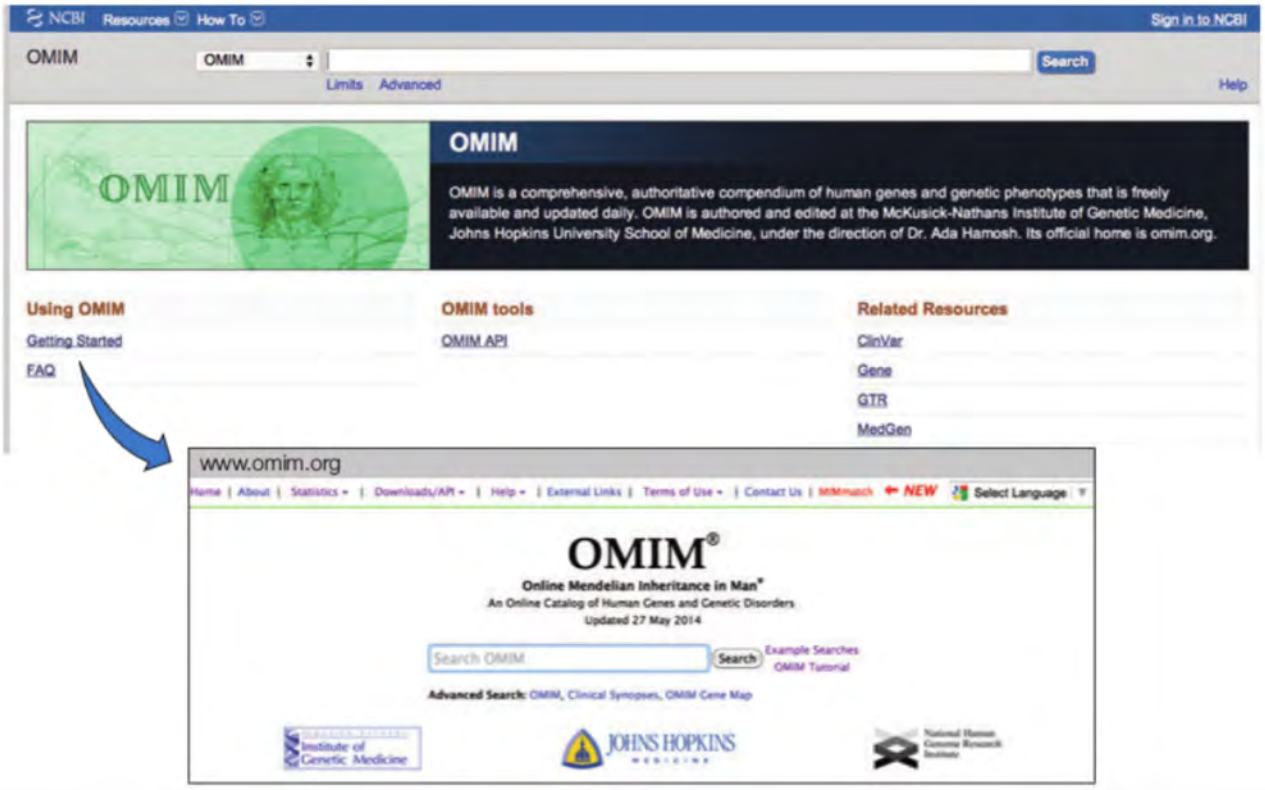
- Genetically typical females lack Y-linked genes, and such genes must code either for non-essential characters, or for male-specific functions.



Left: http://misjaciaza.blogspot.ca/2014/05/niedoceniany-chromosom-y.html;
Right: Adapted from User Was a bee, *Wikimedia Commons,* 2015

Mendelian Inheritance – OMIM

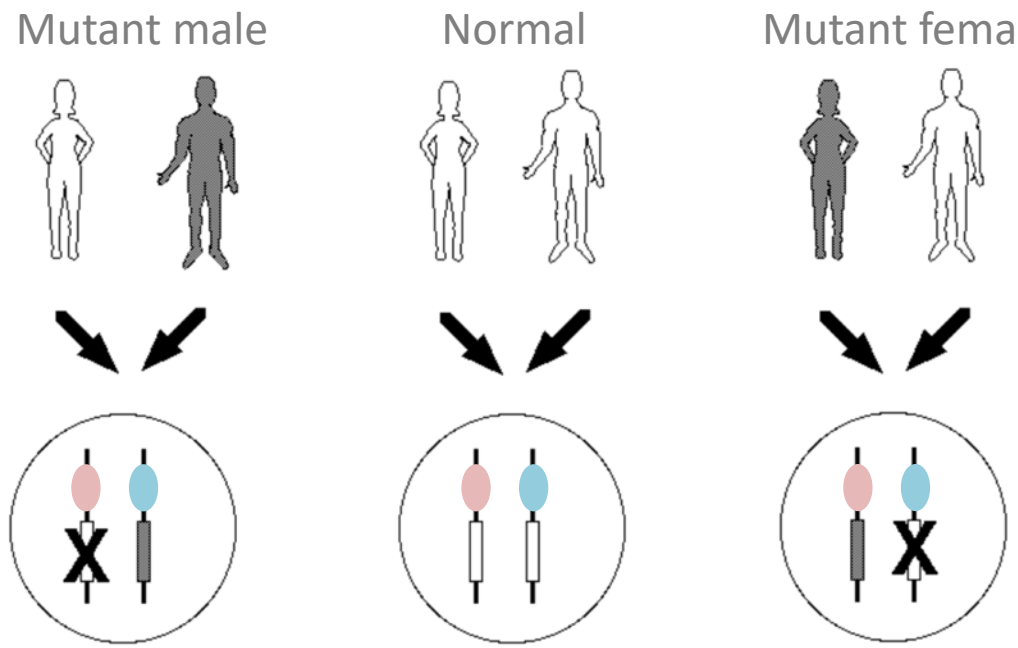
The MIM (Mendelian Inheritance in Man) was created by Dr. Victor A. McKusick as a catalogue of mendelian traits and disorders.



It's now an online (OMIM) database: **www.omim.org**

Epigenetic Imprinting

Imprinted genes are not equally expressed – they exhibit parent-specific monoallelic expression



Mutant allele **imprinted**: No phenotype

One allele **active**, one allele **imprinted**: No phenotype

Wildtype allele **imprinted**: Aberrant phenotype

When a gene at a **maternally imprinted locus** is expressed, the copy of the imprinted gene from the maternally inherited chromosome is always "off," while the copy from the paternally inherited chromosome is always "on"

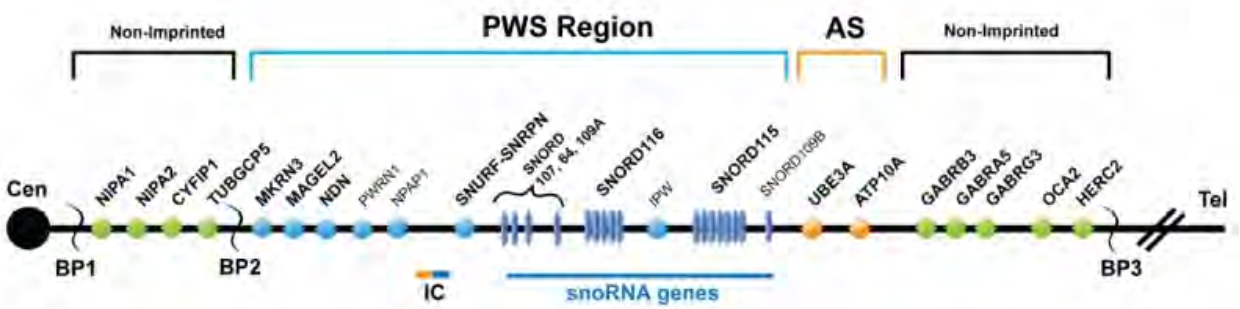
Adapted from Figure 2, C. C. Martin, "Genomic Imprinting," *Dynamic Development,* 1998

Epigenetic Imprinting

Adapted from Driscoll DJ, Miller JL, Schwartz S, et al. 1998

Maternally imprinted

Paternally imprinted



Prader-Willi Syndrome:
Loss of PWS region on paternal chromosome

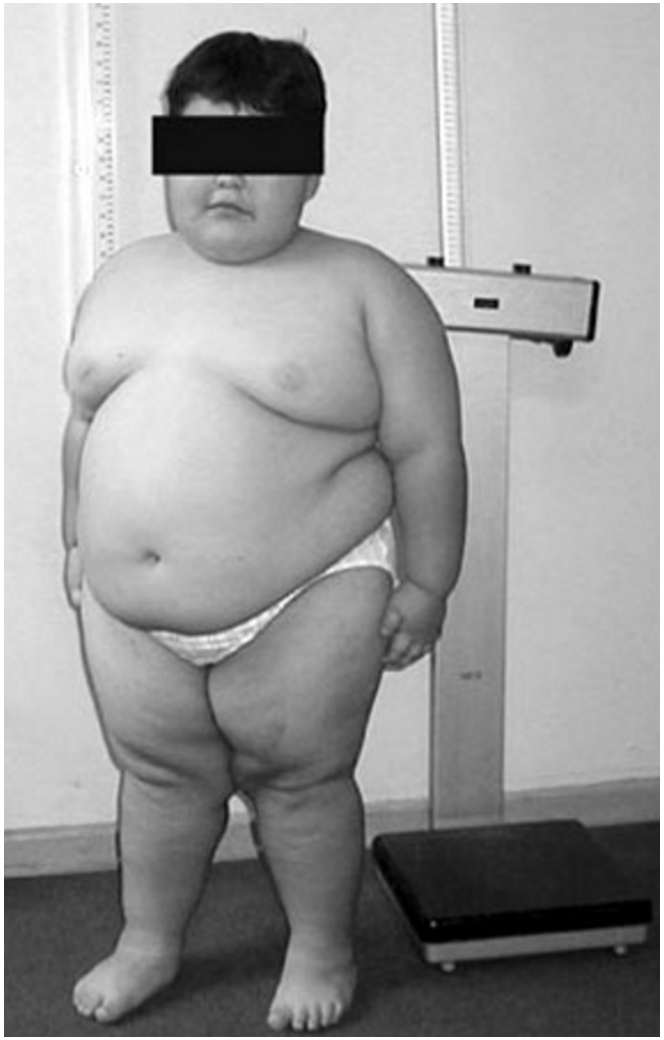
Angelman Syndrome:
Loss of AS region on maternal chromosome

Paternal Maternal

Deletion of the q11-13 region of chromosome 15 will lead to either Prader-Willi or Angelman Syndrome…

Depending on **which** parental chromosome carries the deletion

Epigenetic Imprinting

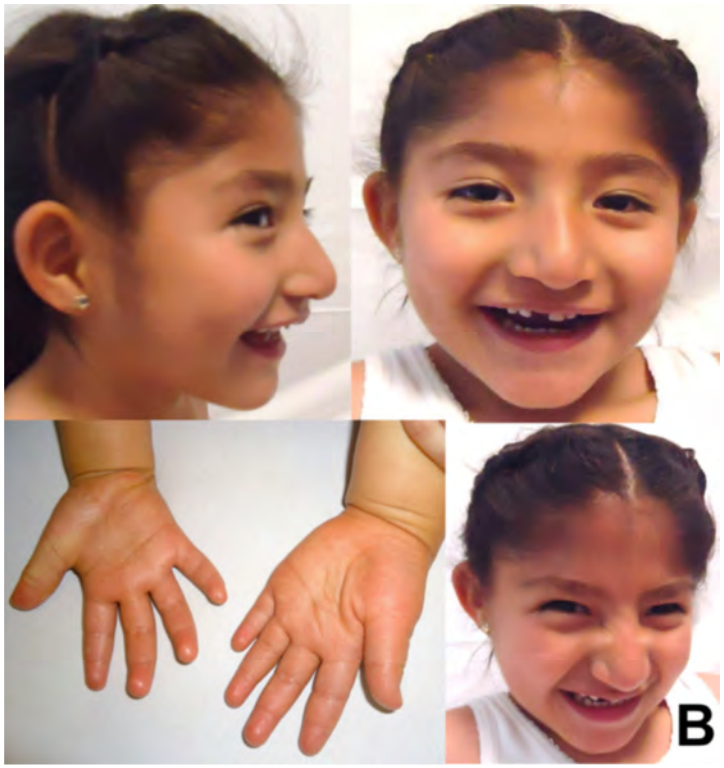


Cortés et al., *Revista Médica de Chile*, 2005

Prader-Willi Syndrome

- Caused by loss of the maternally imprinted PWS region on Chromosome 15
- Symptoms:
  - Hypotonia
  - Strabismus (crossed eyes)
  - Hyperphagia (and downstream obesity)
  - Hypogonadism
  - Mild to moderate intellectual delay

Epigenetic Imprinting



Yokoyama-Rebollar et al., *Molecular Cytogenetics*, 2015

Angelman Syndrome

- Caused by loss of the paternally imprinted AS region on Chromosome 15
- Symptoms:
  - Microcephaly
  - Tapered fingers and abnormal creases
  - Seizures
  - Severe developmental delay
  - Speech impairment
  - Movement or balance disorder
  - Atypical frequent laughter and happy demeanor

Epigenetic Imprinting – Pedigree Patterns

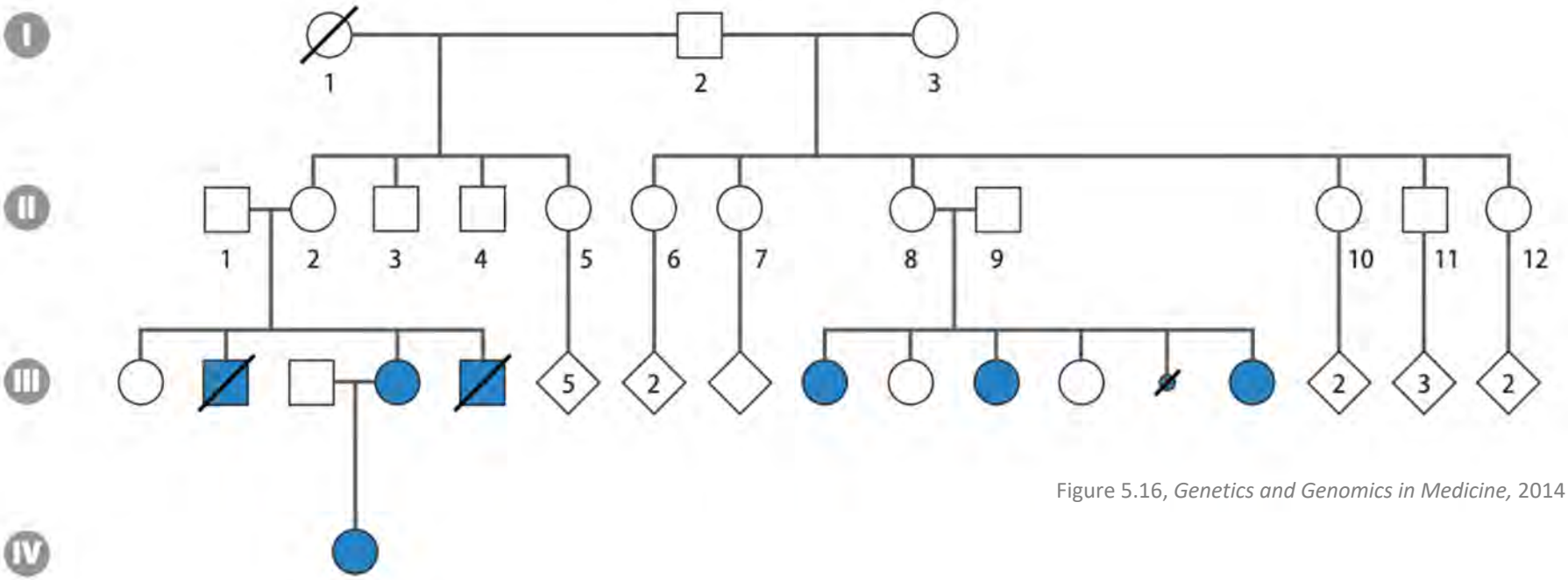


Figure 5.16, *Genetics and Genomics in Medicine,* 2014

Autosomal dominant disorder with paternally imprinted parent-of-origin effect:

- Phenotype manifests only when mutated allele is maternally inherited

Dr. Erin B Styles
She / Her
Director, M.H.Sc. In Medical Genomics Program
Assistant Professor
Department of Molecular Genetics
University of Toronto
www.moleculargenetics.utoronto.ca/medicalgenomics