

# SISG 2022 - Module 2

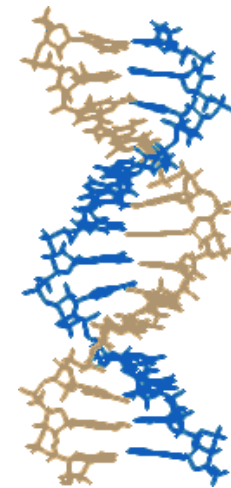
## Introduction to Genetics and Genomics

### What is a gene?

11:20am EDT, Monday, July 11<sup>th</sup>

Joe Lachance and Greg Gibson

[joseph.lachance@biology.gatech.edu](mailto:joseph.lachance@biology.gatech.edu)



# Instructors



Greg Gibson



Joe Lachance

Date	Time (PDT)	Time (EDT)	Topic	Instructor
Monday, July 11	8:00 – 8:20	11:00 – 11:20	Introductions	
	8:20 – 9:05	11:20 – 12:05	What is a Gene ?	JL
	9:05 – 9:20	12:05 – 12:20	Break	
	9:20 – 10:15	12:20 – 1:15	Heritability	GG
	10:15 – 10:45	1:15 – 1:45	Q&A Discussions	

Monday, July 11	11:45 – 12:45	2:45 – 3:45	Quantitative Genetics	GG
	12:45 – 1:00	3:45 – 4:00	Break	
	1:00 – 2:00	4:00 – 5:00	Molecular Biology	JL
	2:00 – 2:30	5:00 – 5:30	Q&A Discussions	

Tuesday, July 12	8:00 – 9:00	11:00 – 12:00	Genome-Wide Association Studies	GG
	9:00 – 9:15	12:00 – 12:15	Break	
	9:15 – 10:15	12:15 – 1:15	Molecular Evolution	JL
	10:15 – 10:45	1:15 – 1:45	Q&A Discussions	

Tuesday, July 12	11:45 – 12:45	2:45 – 3:45	Gene Expression Profiling	GG
	12:45 – 1:00	3:45 – 4:00	Break	
	1:00 – 2:00	4:00 – 5:00	Population Genetics	JL
	2:00 – 2:30	5:00 – 5:30	Q&A Discussions	

Wednesday, July 13	8:00 – 9:00	11:00 – 12:00	Genomic Medicine	GG
	9:00 – 9:15	12:00 – 12:15	Break	
	9:15 – 10:15	12:15 – 1:15	Genetic Ancestry	JL
	10:15 – 10:45	1:15 – 1:45	Q&A Discussions	

# Format

- Lectures will be recorded and posted online. Please remind us we accidentally forget to hit the record button on Zoom!
- Student microphones will be muted during each lecture, but please use the chat feature – the more questions you ask, the more you will learn
- Every few minutes we check to see if there are any questions in chat
- After each block of lectures, we will form small breakout rooms so that classmates can get to know each other via chat or video
- We will also hold open question and answer sessions (these informal "office hours" will not be recorded)

# What is a gene?



- How would you define a gene?
- Which matters more, structure or function?

# Terminology

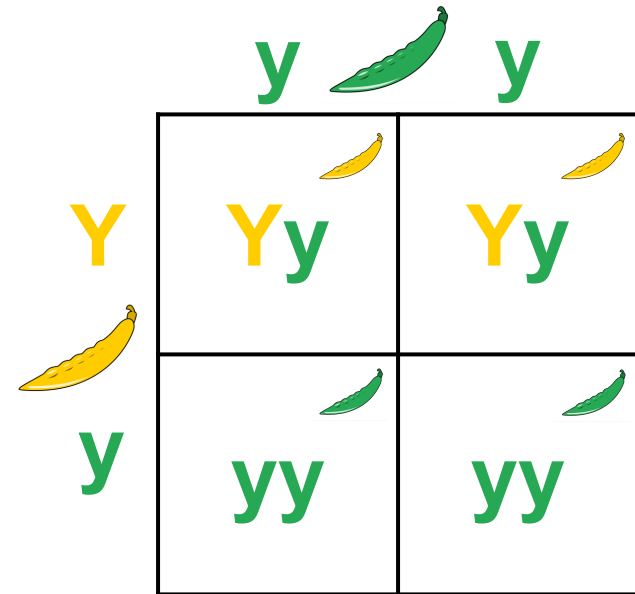
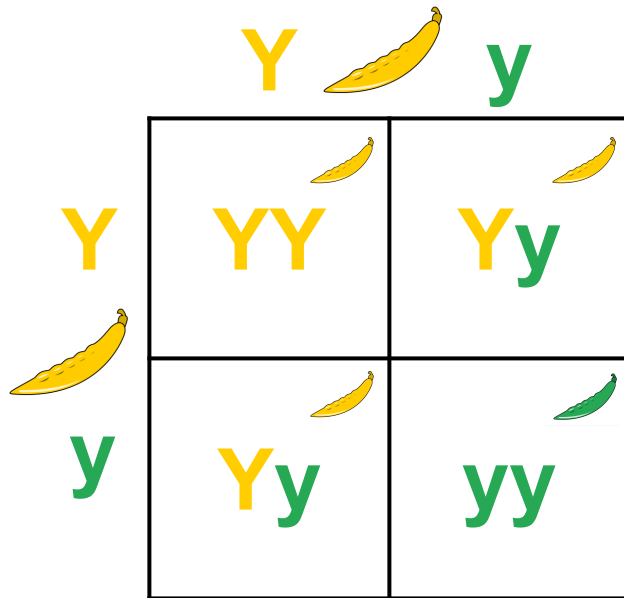
- **Genes:** DNA sequences that encode a functional protein or RNA molecule
- **Allele:** one of two or more alternative forms of a gene
- **Genotype:** the genetic makeup of an individual
- **Phenotype:** the observable characteristics and traits of an organism
- **Genome:** the complete set of genetic material in a cell or organism
- **Haplotype:** closely linked DNA sequences on the same chromosome that are co-inherited

# Mendel's laws of inheritance

- Law of segregation (1<sup>st</sup> law)
  - Parental pairs of alleles separate during gamete formation
- Law of independent assortment (2<sup>nd</sup> law)
  - Pairs of alleles for different traits segregate independently
- Law of dominance (3<sup>rd</sup> law)
  - Heterozygotes manifest the trait associated with the dominant allele
- *These rules are often broken!*



# Mendelian ratios



- Punnett squares can be used to predict the products of breeding
- Genetics is not always this simple!



## Prokaryotes

## Eukaryotes

Internal structures

No organelles

Organelles

DNA

Circular  
No introns  
DNA in cytoplasm

Linear  
Introns  
DNA in nucleus

Genome size

Tend to be < 5Mb

10Mb-100,000Mb

Chromatin

No histones

Histones

Ploidy

Haploid

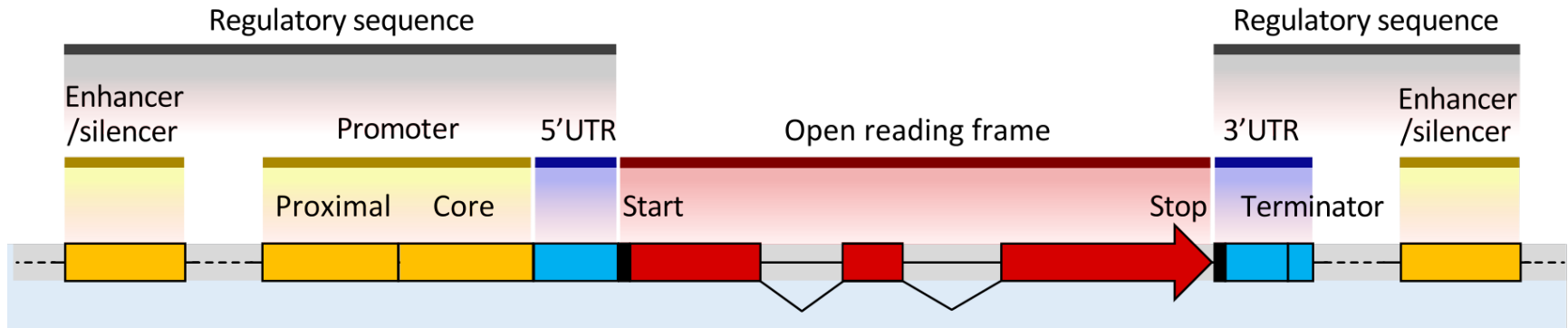
Usually diploid

Reproduction

Asexual (binary fission)

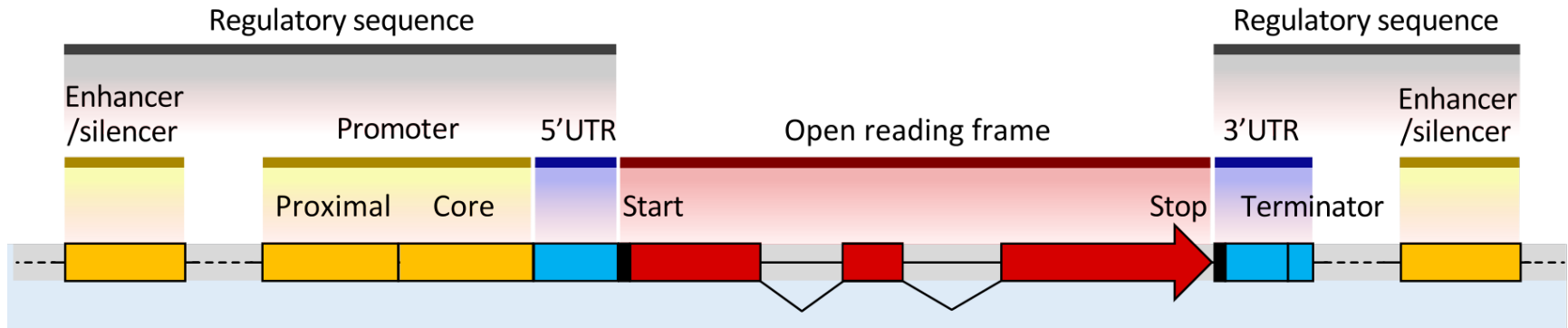
Asexual (mitosis) and  
sexual (meiosis)

# The structure of (protein coding) genes

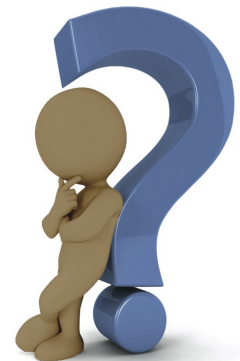


- Exons: nucleotide sequence not removed by splicing (coding DNA)
- Introns: nucleotide sequence removed by splicing (noncoding DNA)
- Cis-regulatory elements
  - Enhancers: increase the likelihood of transcription when bound to activators
  - Silencers: decrease likelihood of transcription when bound to repressors
  - Promoters: region of DNA where transcription is initiated
- UTRs: untranslated regions

# The structure of (protein coding) genes

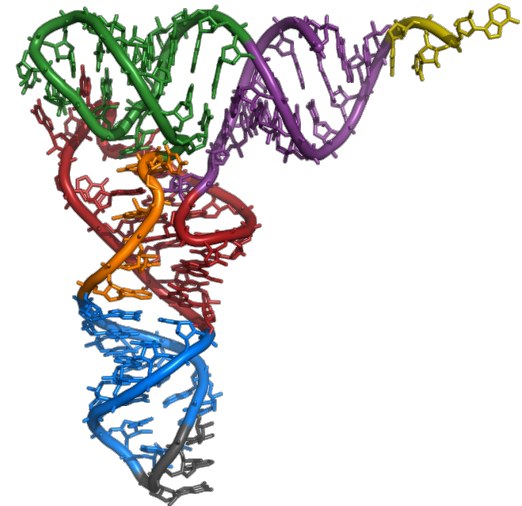


- Which parts would you consider to be part of a gene?



# RNA genes

- DNA sequences which encode function non-coding RNA are called RNA genes
- Transfer RNAs (tRNAs)
- Ribosomal RNAs (rRNAs)
- Different types of small RNAs (e.g., microRNAs, siRNAs)



# The locus of evolution

- What sort of genetic changes underlie morphological adaptations?

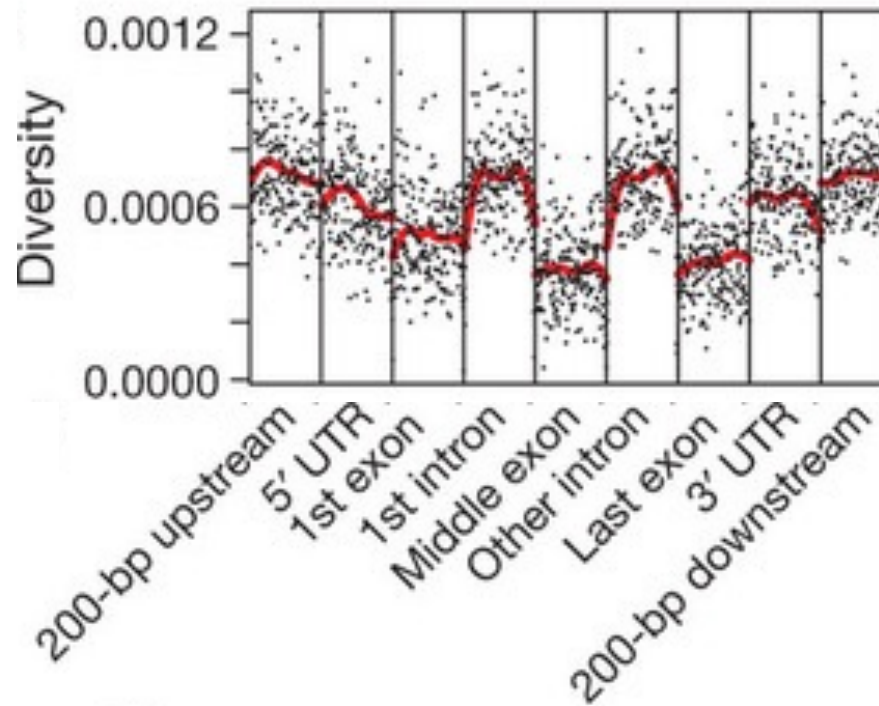
- Hopi Hoekstra and Jerry Coyne
  - Protein-coding DNA matters



- Sean Carroll
  - It's all about regulatory DNA



# Polymorphism near genes



- Exons contain less genetic variation than non-coding DNA

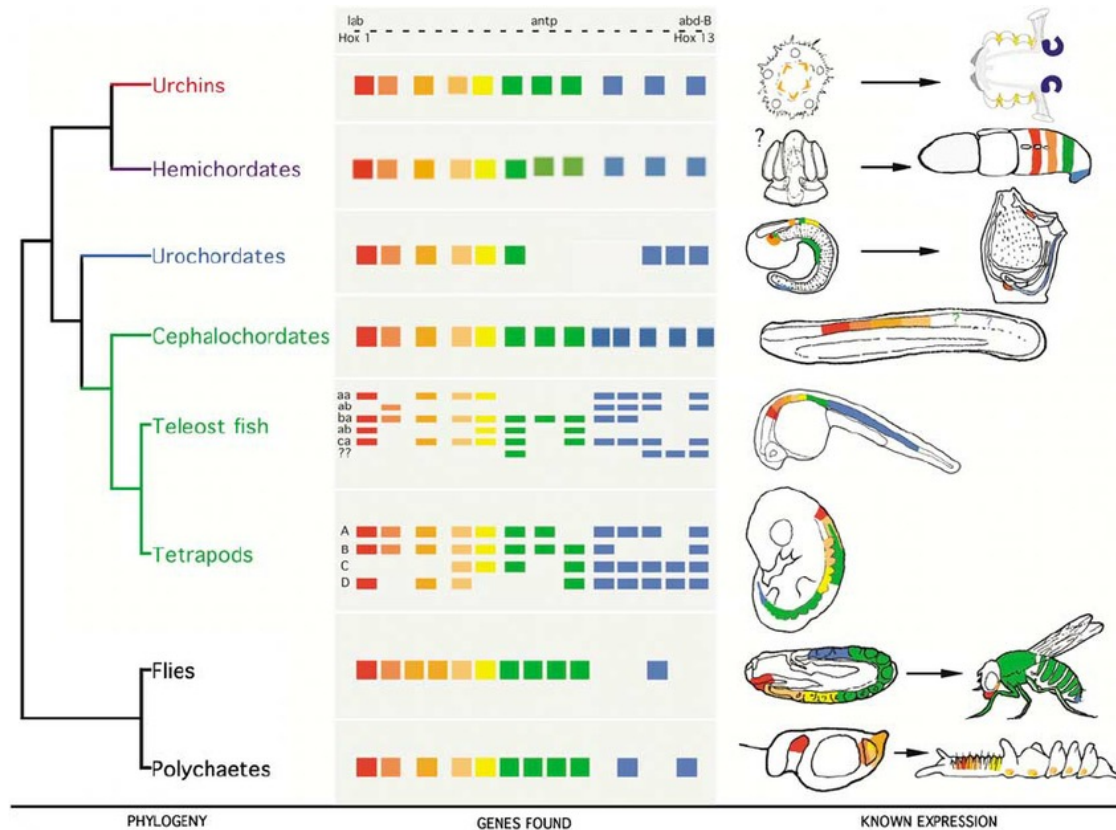
# GeneCards

The screenshot displays the GeneCards website interface. At the top, there is a navigation bar with various tools and services like MalaCards, LifeMap Discovery, PathCards, etc. The main header features the GeneCards logo and a search bar. Below the header, the page is dedicated to the ACE2 gene, showing its protein-coding nature and function as Angiotensin I Converting Enzyme 2. A table of navigation options is provided, including Aliases, Disorders, Domains, Drugs, Expression, Function, Genomics, Localization, and Orthologs. Below this, there are advertisements for ORIGENE, genomics, BIO BASIC, and GenScript. The main content area lists aliases for ACE2 Gene, such as Angiotensin I Converting Enzyme 2, and provides external IDs and previous GeneCards identifiers.

- How to find more information about a particular gene?
- <https://www.genecards.org>



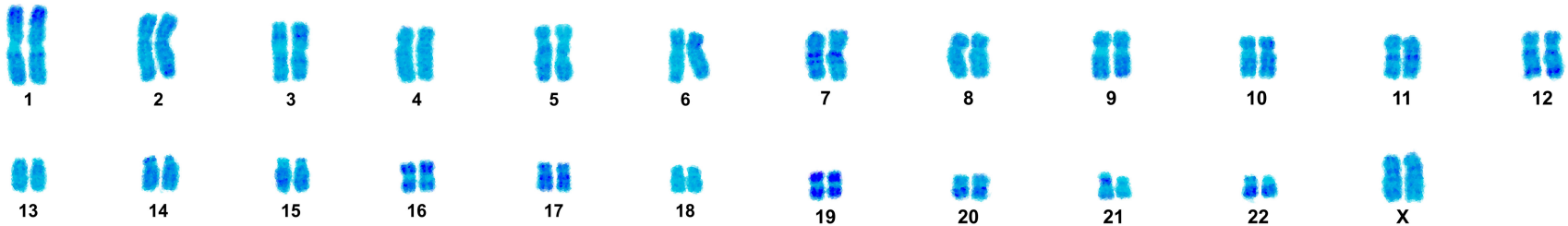
# Different species share many of the same genes



- Homologous genes are two or more genes that descend from a common ancestral DNA sequences

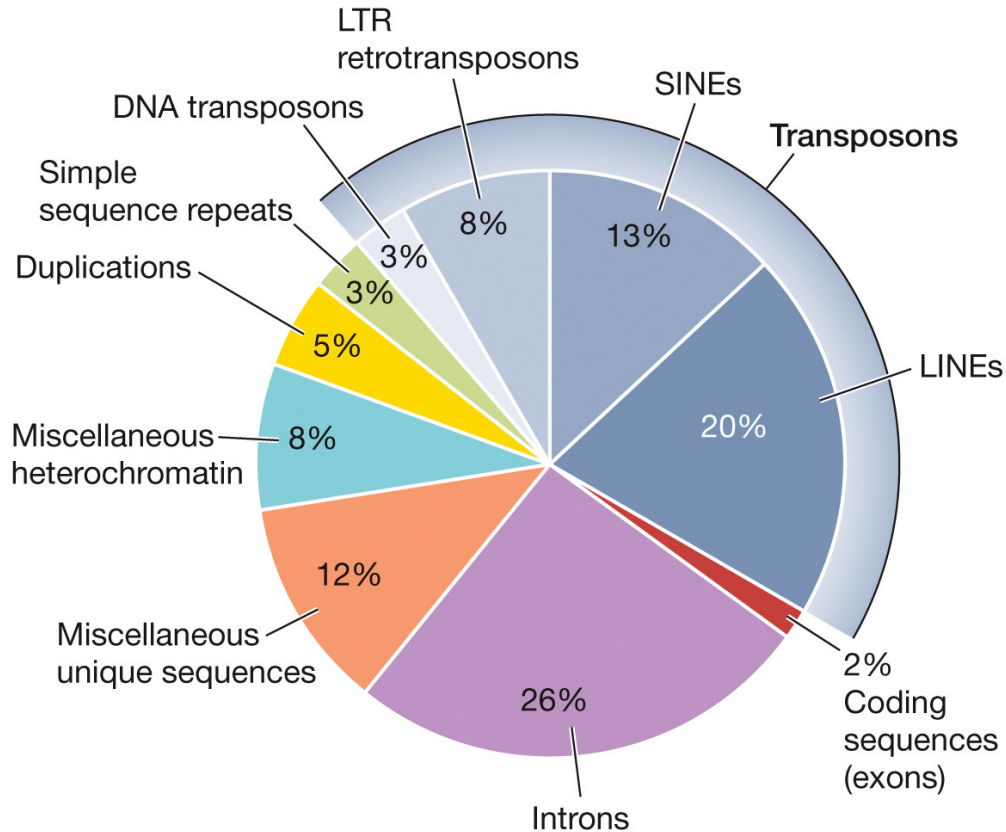


# Genes do not exist in isolation



- Genes are found on chromosomes
- DNA sequences that are close together on the same chromosome tend to be inherited together (**linkage**)
- Genes can interact with other genes (**epistasis**)

# Genomic fractions



*EVOLUTION 4e*, Figure 14.14  
© 2017 Sinauer Associates, Inc.

- Most of the human genome is non-coding
- Transposons (selfish DNA) make up a large % of the human genome

# Whole genome sequencing (WGS)

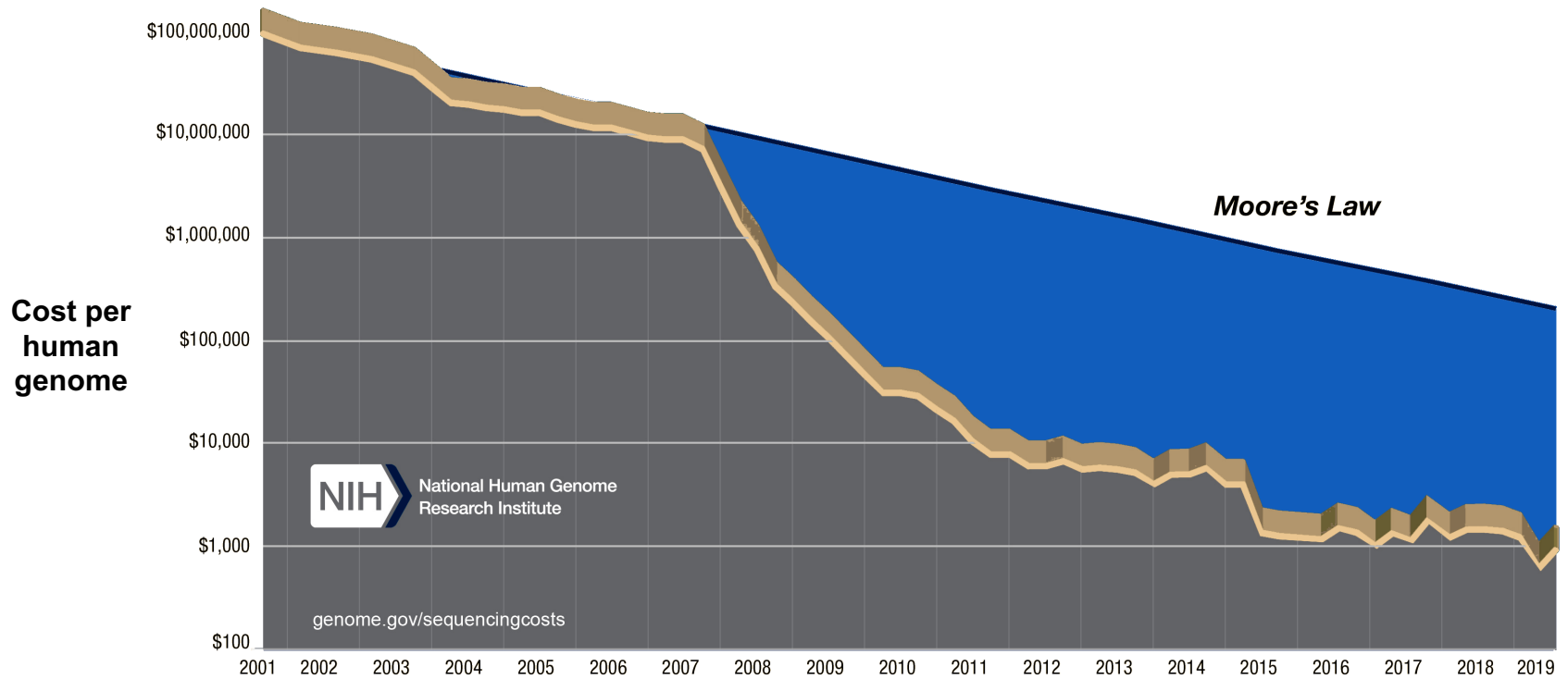
- WGS is sometimes called next-generation sequencing
- Depth of coverage: average number of reads per base pair in a genome (low coverage = 5-10X, high coverage: >30X)
- One error per 100,000 base pairs (high coverage)
- Relatively expensive, but getting cheaper



NovaSeq 6000

illumina

# Declining sequencing costs



# UCSC Genome Browser

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chrX:15,376,366-15,786,735 410,370 bp. enter position, gene symbol, HGVS or search terms go

chrX (p22.2) hg38

Scale 100 kb hg38

Chromosome Bands Localized by FISH Mapping Clones

GENCODE v25 Comprehensive Transcript Set (only basic displayed by default)

EMX CLTRN CHS8

NHORI-EBI Catalog of Published Genome-wide Association Studies

Cons 100 Verts

100 vertebrates Basewise Conservation by PhyloP

Multi-z alignments of 100 Vertebrates

Rhesus Mouse Dog E.tesart Chicken X\_Tropical Zebrafish Lamprey

Short Genetic Variants from dbSNP release 153

Common dbSNP(153)

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

move start < 2.0 > move end < 2.0 >

track search default tracks default order hide all add custom tracks track hubs configure multi-region reverse resize refresh

collapse all expand all

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

refresh

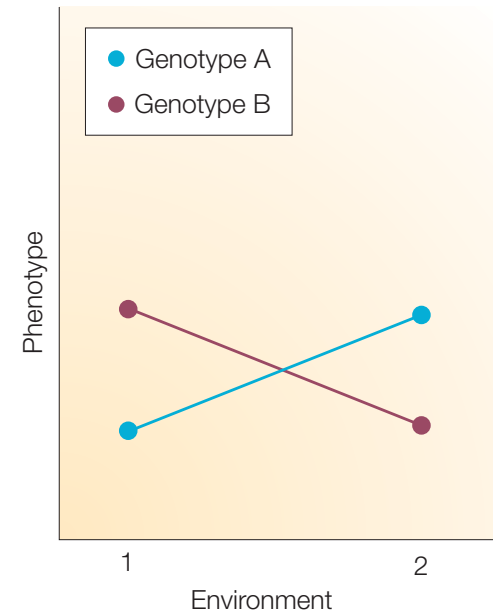
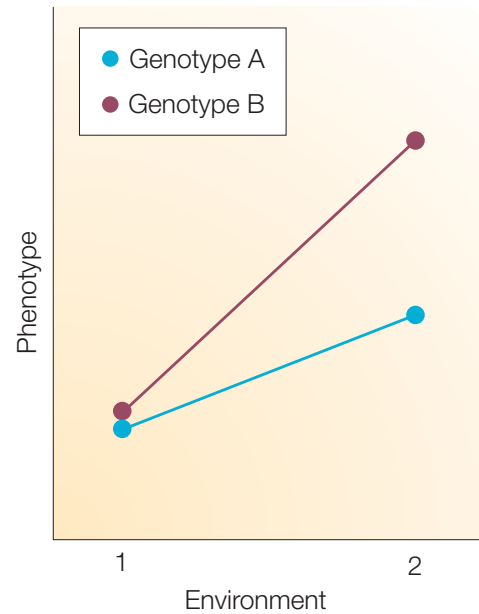
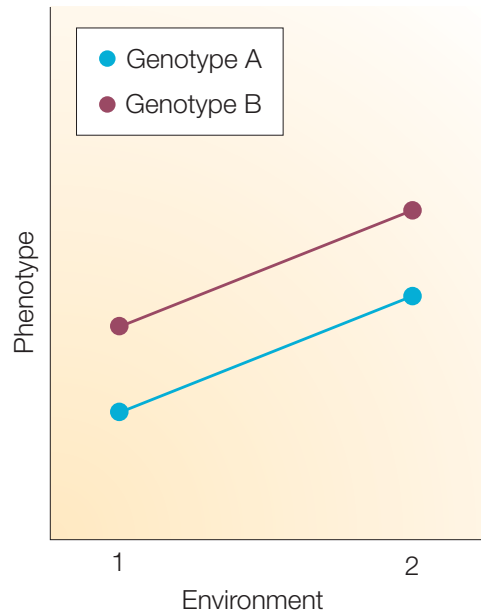
Mapping and Sequencing

Base Position dense ▾	P12 Fix Patches hide ▾	P12 Alt Haplotypes hide ▾	P12 Assembly hide ▾	Centromeres hide ▾	P12 Chromosome Band full ▾
Clone Ends hide ▾	FISH Clones hide ▾	P12 Gap hide ▾	P12 GC Percent hide ▾	GRC Contigs hide ▾	GRC Incident hide ▾
Hg19 Diff hide ▾	P12 INSDC hide ▾	LRG Regions hide ▾	Mappability... hide ▾	P12 RefSeq Acc hide ▾	Restr Enzymes hide ▾
Scaffolds hide ▾	Short Match hide ▾	STS Markers hide ▾			

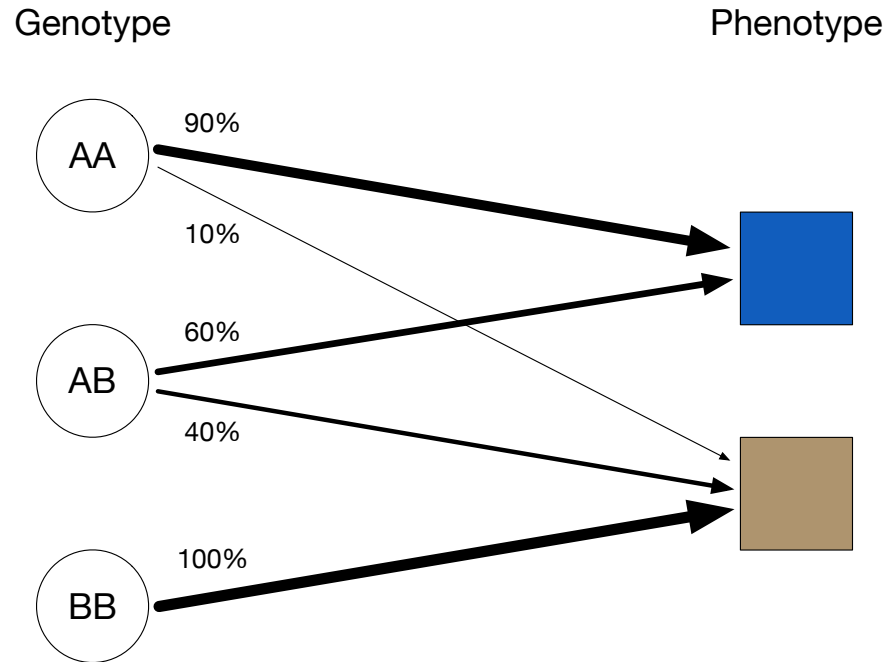
- An online resource for exploring the human genome
- <https://genome.ucsc.edu>



# Environmental context matters



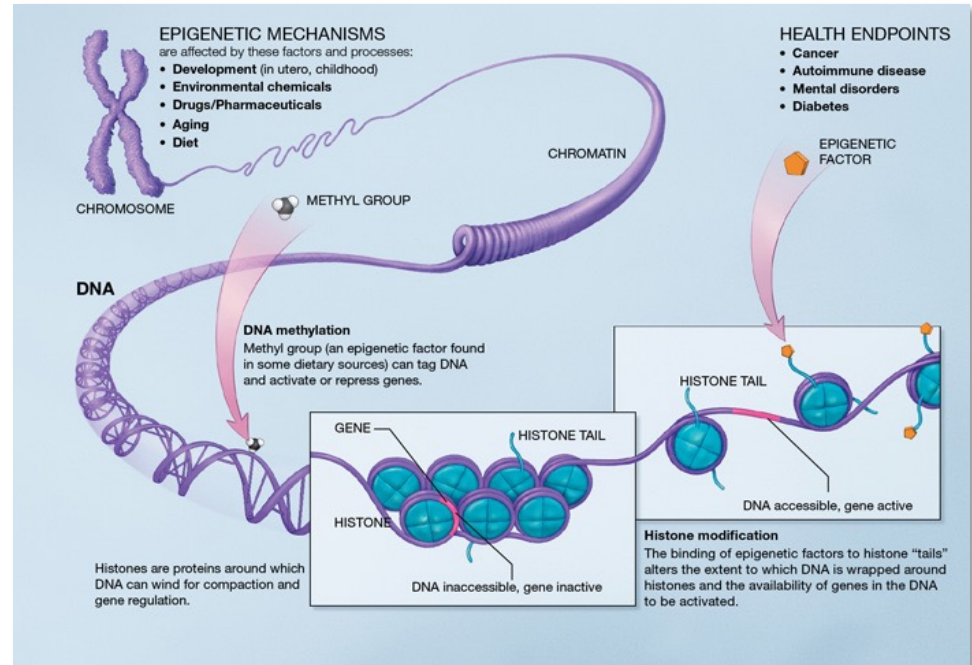
# Incomplete penetrance



- Genotype-phenotype maps are not always one-to-one (e.g., some alleles increase your chances of getting hypertension)
- **Penetrance** refers to the proportion of individuals with a given genotype that show the expected phenotype

# Epigenetics

- DNA methylation (methylated CpGs)
- Histone modification
- X-inactivation
- Genomic imprinting
- Different people have different epigenetic marks
- Most of these epigenetic marks are erased each generation



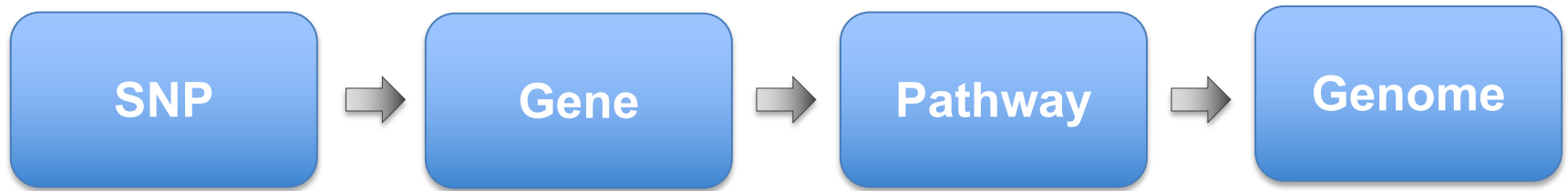


# Pleiotropy



- Be careful to avoid terminology like "**cancer gene**" or "**height gene**"
- This is because genes often contribute to multiple phenotypes (i.e., they are **pleiotropic**)
- Example: A mutation in the *Frizzle* gene results in feathers that curve outward, fewer eggs laid, and high body temperatures

# Units of analysis in genetics



- Genetic data be analyzed on **population** as well as **individual** scales
- Sometimes we are more focused on **traits**...