

SISG 2023 - Module 2

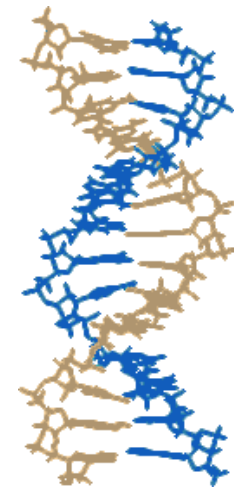
Introduction to Genetics and Genomics

What is a gene?

Block #1 – Monday, July 10

Joe Lachance and Greg Gibson

joseph.lachance@biology.gatech.edu



Date	Time (PDT)	Reading	Topic	Instructor
Monday, July 10	8:30am – 8:45am		Introductions	
	8:45am – 10:00am	Portin and Wilkins (Genetics, 2017)	What is a Gene ?	JL
	10:00am – 10:30am		Break	
	10:30am – 11:45am	Visscher et al. (Nat Rev Genet, 2008)	Heritability	GG
	11:45am – noon		Q&A Discussions	

Monday, July 10	1:30pm – 3:00pm	Goddard et al. (Proc Roy Soc B, 2016)	Quantitative Genetics	GG
	3:00pm – 3:30pm		Break	
	3:30pm – 4:45pm	Deichmann (Dev Biol, 2016)	Molecular Biology	JL
	4:45pm – 5:00pm		Q&A Discussions	

Tuesday, July 11	8:30am – 10:00am	Abdellaoui et al. (Am J Hum Genet, 2023)	Genome-Wide Association Studies	GG
	10:00am – 10:30am		Break	
	10:30am – 11:45am	Christmas et al. (Science, 2023)	Molecular Evolution	JL
	11:45am – noon		Q&A Discussions	

Tuesday, July 11	1:30pm – 3:00pm	Wang et al. (Nat Rev Genet, 2009)	Gene Expression Profiling	GG
	3:00pm – 3:30pm		Break	
	3:30pm – 4:45pm	Prohaska et al. (Cell, 2023)	Population Genetics	JL
	4:45pm – 5:00pm		Q&A Discussions	

Wednesday, July 12	8:30am – 10:00am	Ashley (Nat Rev Genet, 2016)	Genomic Medicine	GG
	10:00am – 10:30am		Break	
	10:30am – 11:45am	Bergström et al. (Nature, 2021)	Genetic Ancestry	JL
	11:45am – noon		Q&A Discussions	

What is a gene?



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

Terminology

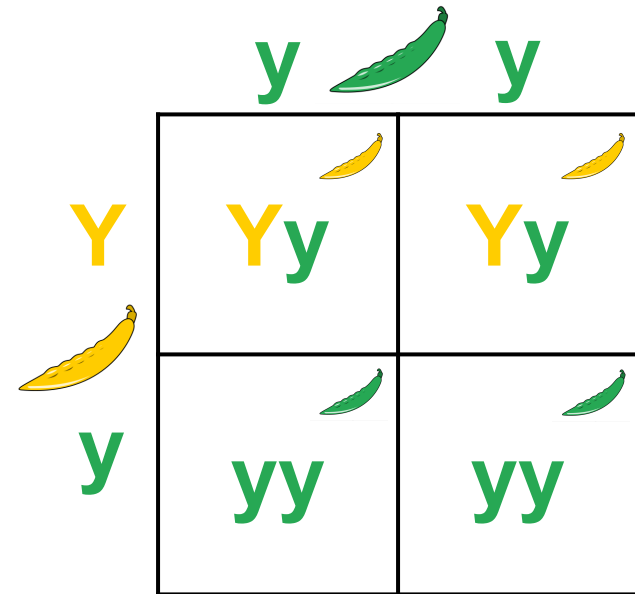
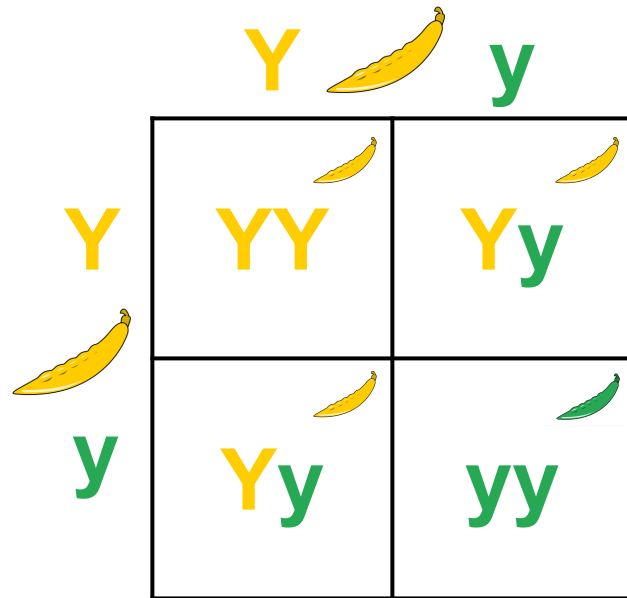
- **Genes:** DNA sequences that encode a functional protein or RNA molecule
- **Allele:** one of two or more alternative forms of a gene
- **Genotype:** the genetic makeup of an individual
- **Phenotype:** the observable characteristics and traits of an organism
- **Genome:** the complete set of genetic 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 (1st law)
 - Parental pairs of alleles separate during gamete formation
- Law of independent assortment (2nd law)
 - Pairs of alleles for different traits segregate independently
- Law of dominance (3rd 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

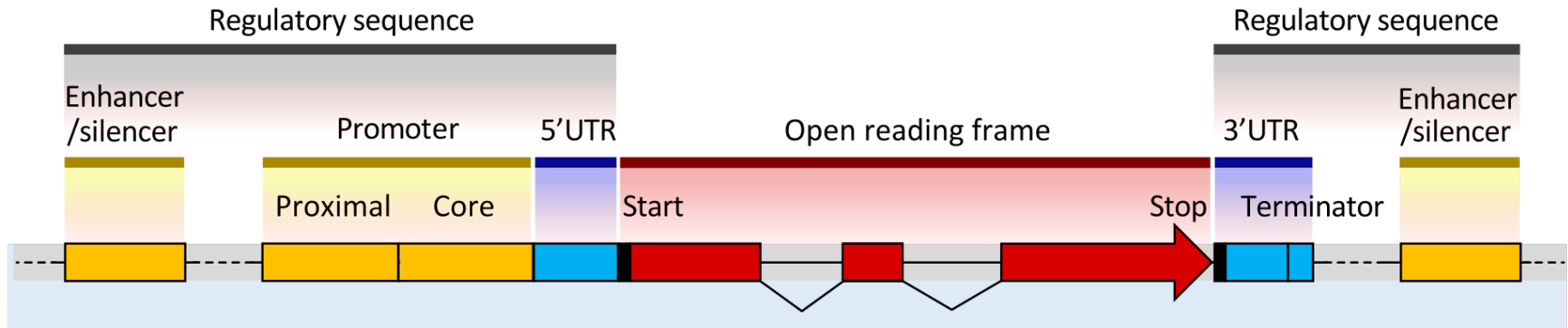
Usually diploid

Reproduction

Asexual (binary fission)

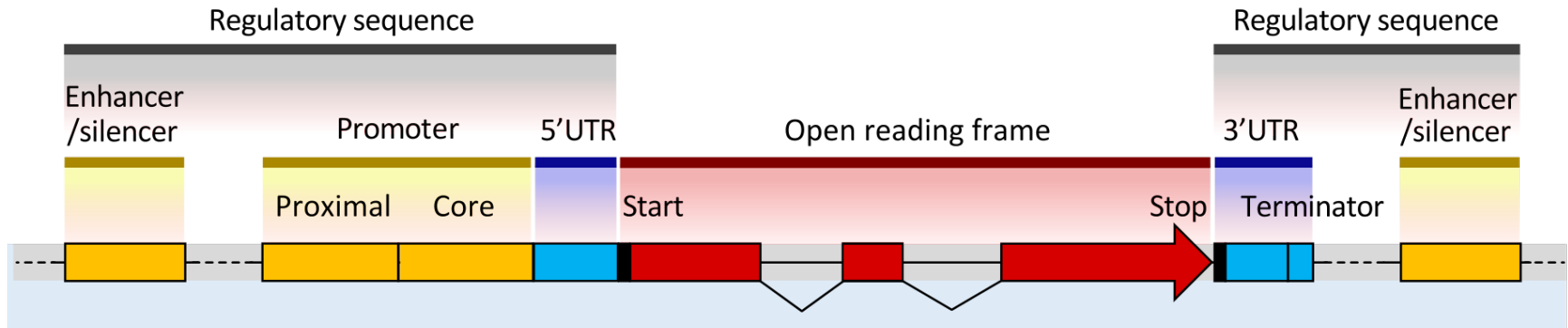
Asexual (mitosis) and
sexual (meiosis)

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

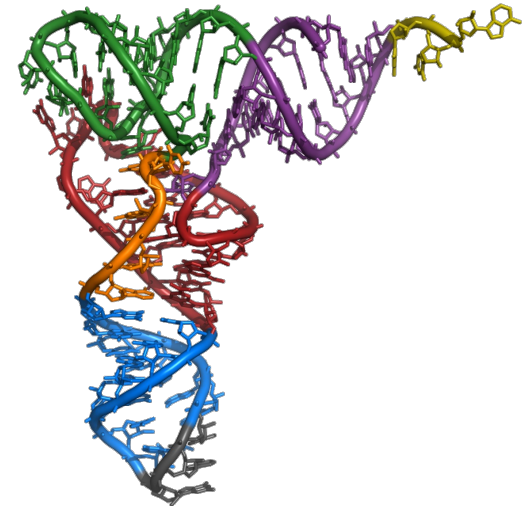


- Which parts would you consider to be part of a gene?



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?

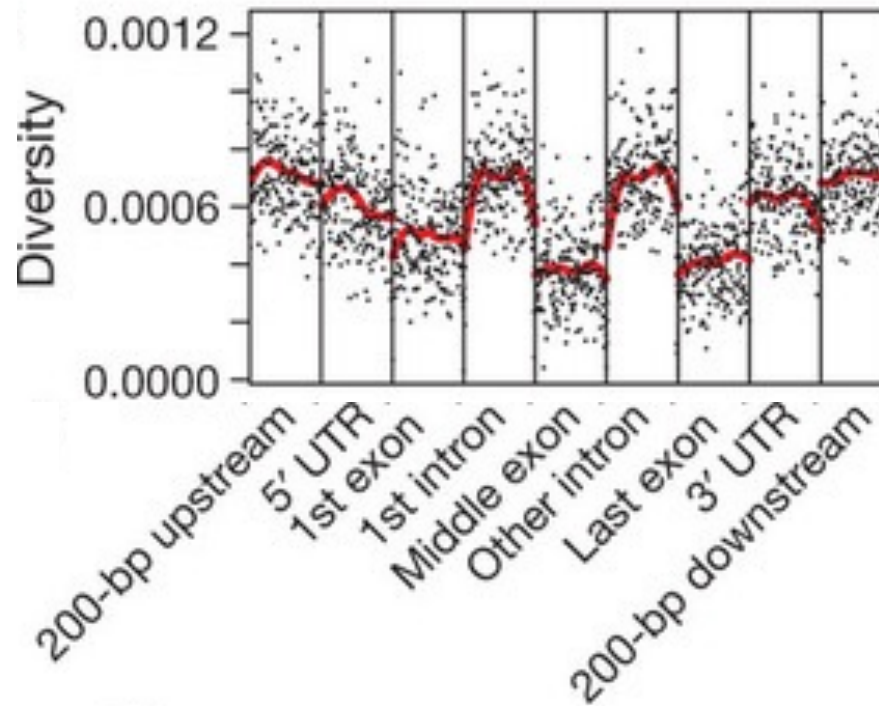
- Hopi Hoekstra and Jerry Coyne
 - Protein-coding DNA matters



- Sean Carroll
 - It's all about regulatory DNA



Polymorphism near genes



- Exons contain less genetic variation than non-coding DNA

GeneCards

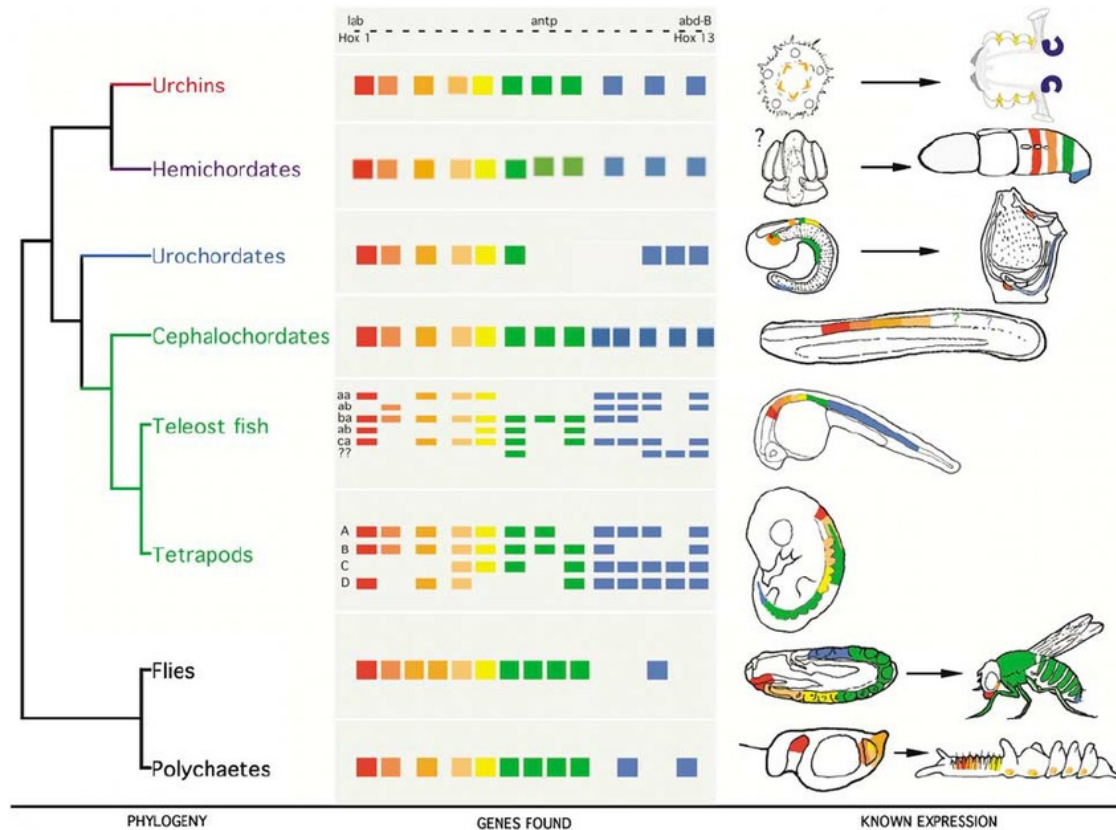
The screenshot displays the GeneCards website interface for the ACE2 gene. The page is titled "ACE2 Gene (Protein Coding) ★" and includes the following information:

- Gene Name:** ACE2 Gene (Protein Coding) ★
- Function:** Angiotensin I Converting Enzyme 2
- GCID:** GC0XM015494
- GIFTS:** 50
- Navigation:** Home, User Guide, Analysis Tools, News And Views, About, My Genes, Log In / Sign Up
- Jump to section:** Aliases, Disorders, Domains, Drugs, Expression, Function, Genomics, Localization, Orthologs, Paralogs, Pathways, Products, Proteins, Publications, Sources, Summaries, Transcripts, Variants
- Research Products:** Antibodies, Assays, Cell Lines, Clones, Primers, Inhib. RNA, Genotyping
- Service Providers:** ORIGENE (Proteins, Antibodies, Assays, Genes, shRNA, Primers, CRISPR, Lentiviral Particles), genomics (ORF Clones, CRISPR, Cloning Vectors, Lentiviral Vectors), BIO BASIC (Genes at \$0.09/bp, Proteins, Genes (Complex Genes)), GenScript (Genes, Peptides, Proteins, CRISPR)
- Aliases for ACE2 Gene:** Angiotensin I Converting Enzyme 2^{2 3 5}, Angiotensin I Converting Enzyme (Peptidyl-Dipeptidase A) 2^{2 3}, Angiotensin-Converting Enzyme Homolog^{3 4}, Angiotensin-Converting Enzyme 2^{3 4}, ACE-Related Carboxypeptidase^{3 4}, Metalloprotease MPROT15^{3 4}, Peptidyl-Dipeptidase A^{2 3}, ACEH^{3 4}, EC 3.4.17.23⁴, EC 3.4.17⁵²
- External IDs for ACE2 Gene:** HGNC: 13557, Entrez Gene: 59272, Ensembl: ENSG00000130234, OMIM: 300335, UniProtKB: Q9BYF1
- Previous GeneCards Identifiers for ACE2 Gene:** GC0XM015289, GC0XM014404, GC0XM014781, GC0XM014940, GC0XM015338, GC0XM015489, GC0XM013333, GC0XM015579
- Search aliases for ACE2 gene in PubMed and other databases**
- Summaries for ACE2 Gene**

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

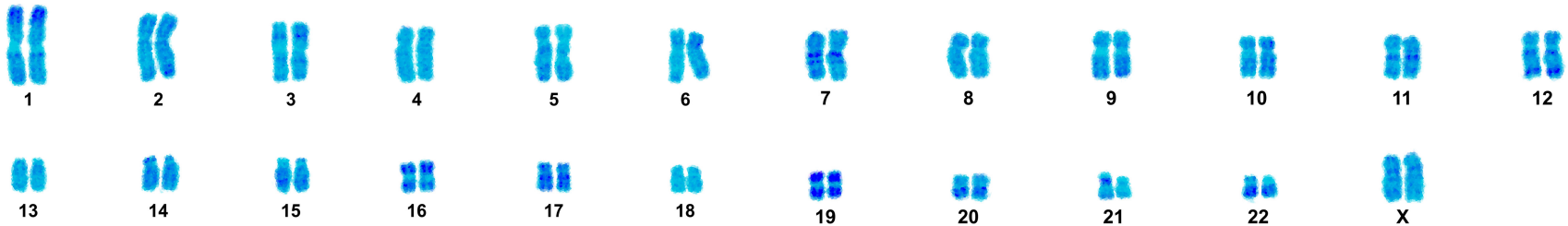


Different species share many of the same genes



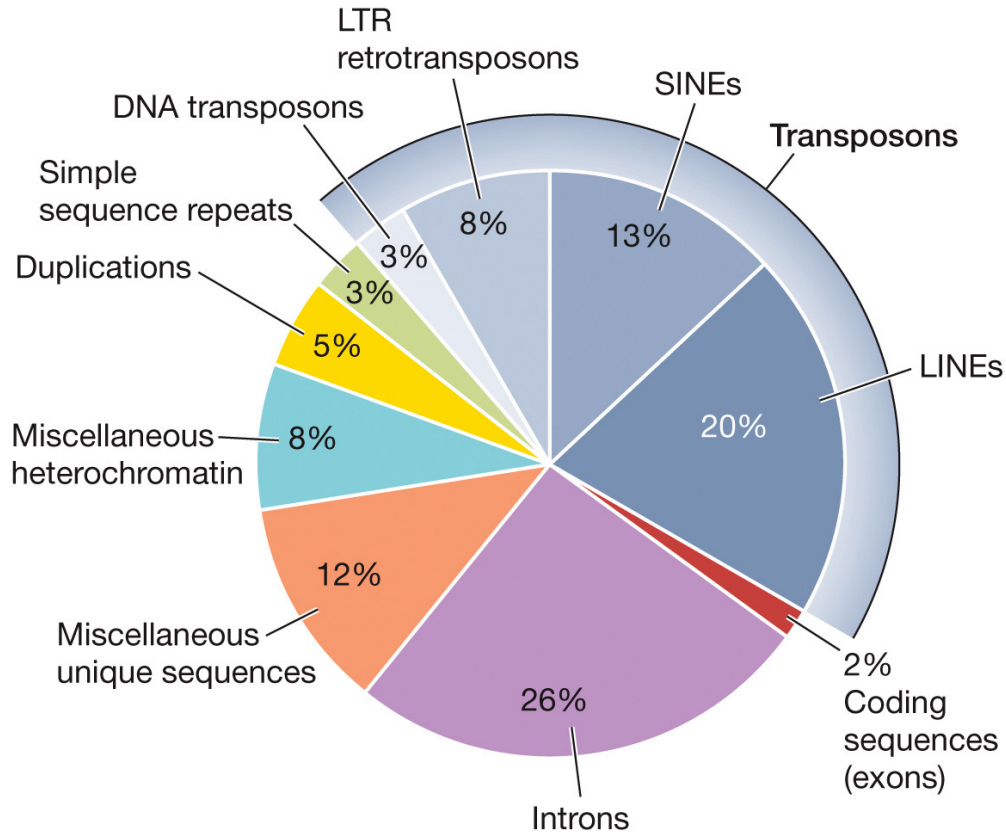
- Homologous genes are two or more genes that descend from a common ancestral DNA sequences

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)

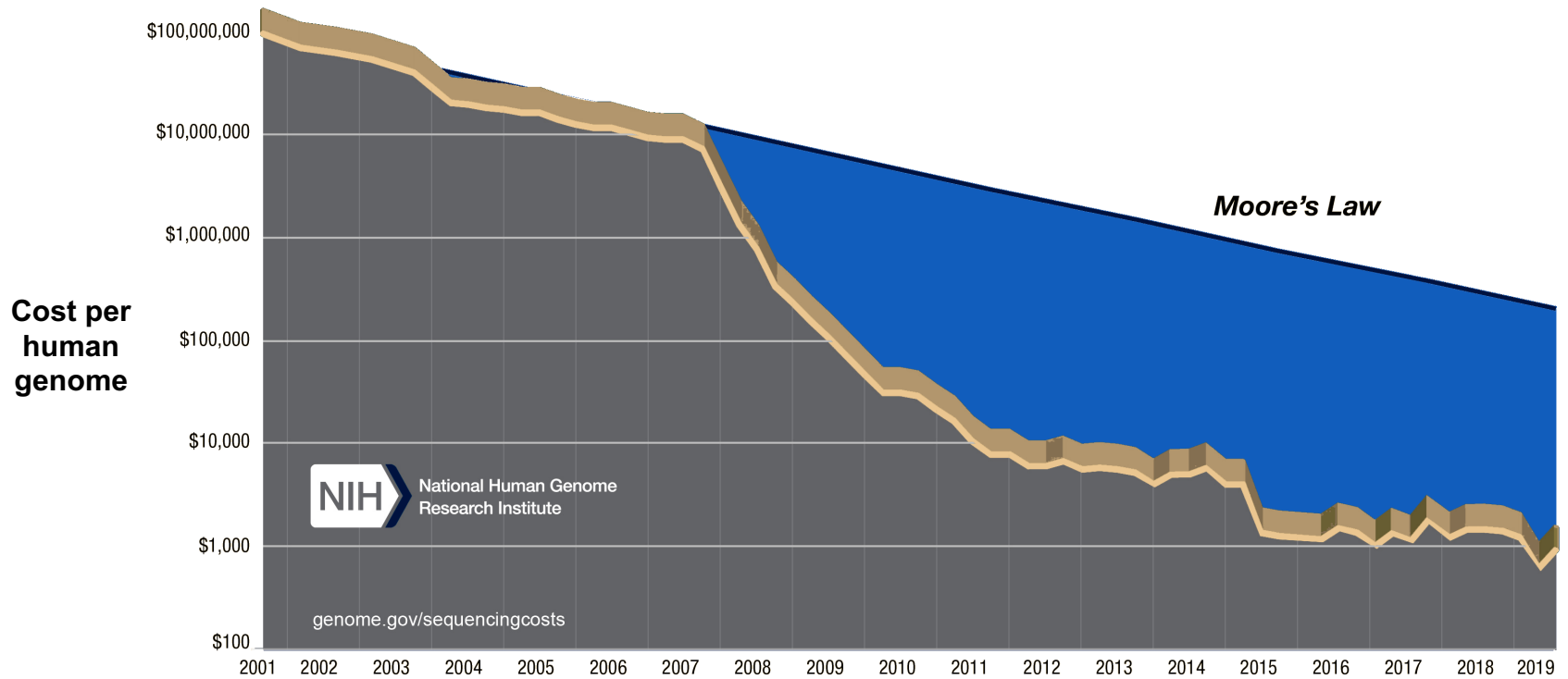
- WGS is sometimes called next-generation sequencing
- Depth of coverage: average number of reads per base pair in a genome (low coverage = 5-10X, high coverage: >30X)
- One error per 100,000 base pairs (high coverage)
- Relatively expensive, but getting cheaper



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

chrX: 15,450,000 15,500,000 15,550,000 15,600,000 15,650,000 15,700,000 15,750,000

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

SNPedia 2.11 SNPs (including empty pages)

PS297452 PS714205 PS4646174 PS4646156 PS2008556 PS2106899 PS1970241

100 vertebrates Basewise Conservation by PhyloP

Cons 100 Verts 4.00 0 -4.5

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.

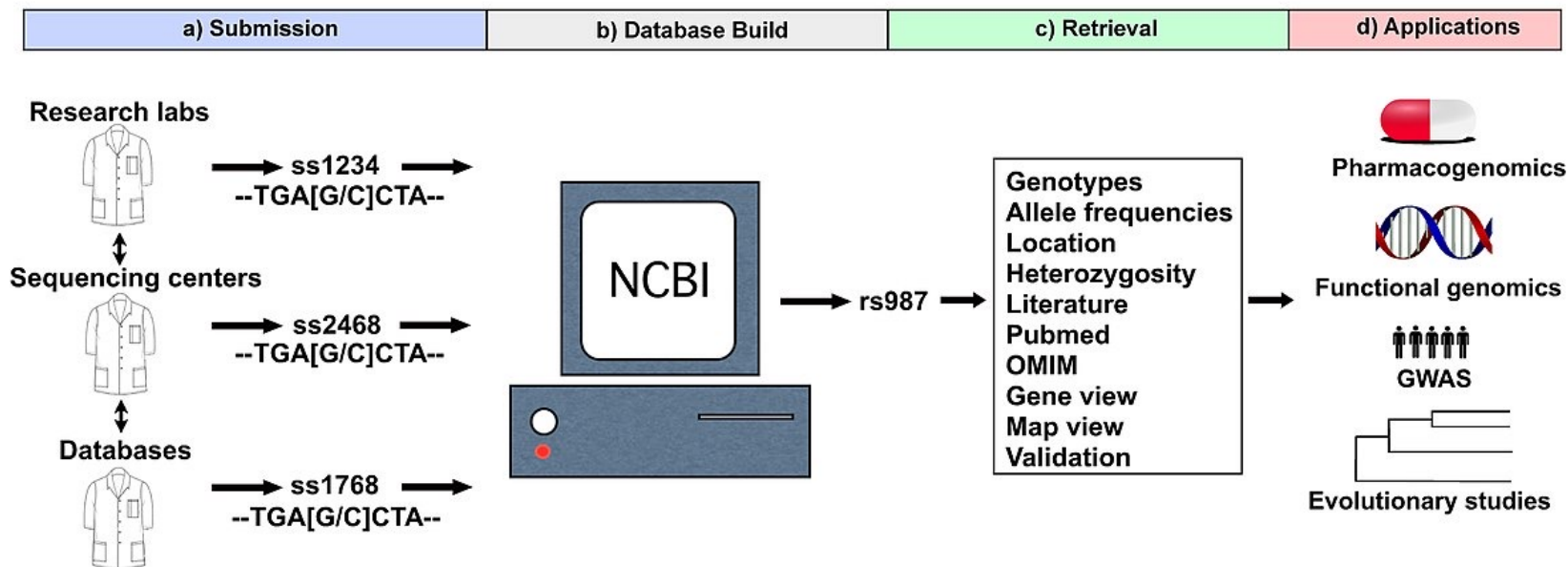
Mapping and Sequencing refresh

Base Position dense ▾	P12 Fix Patches hide ▾	P12 Alt Haplotypes hide ▾	P12 Assembly hide ▾	Centromeres hide ▾	P12 Chromosome Band full ▾
Clone Ends hide ▾	18 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>



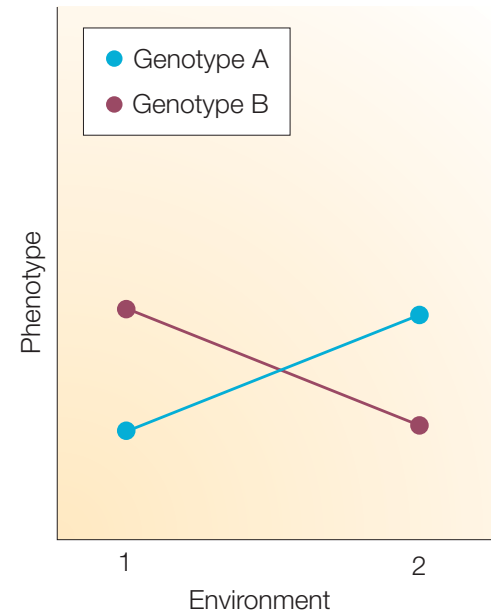
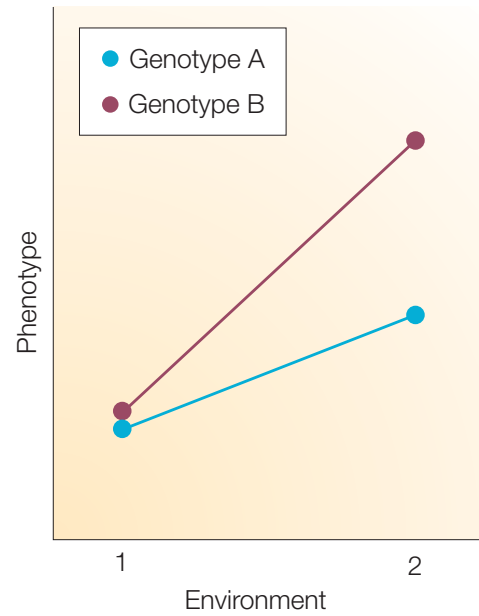
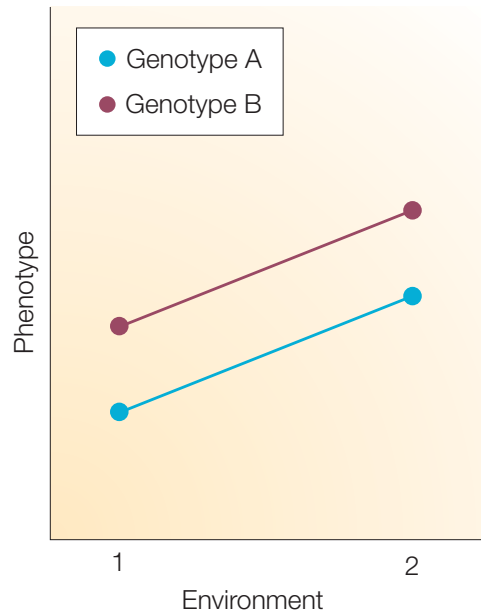
dbSNP



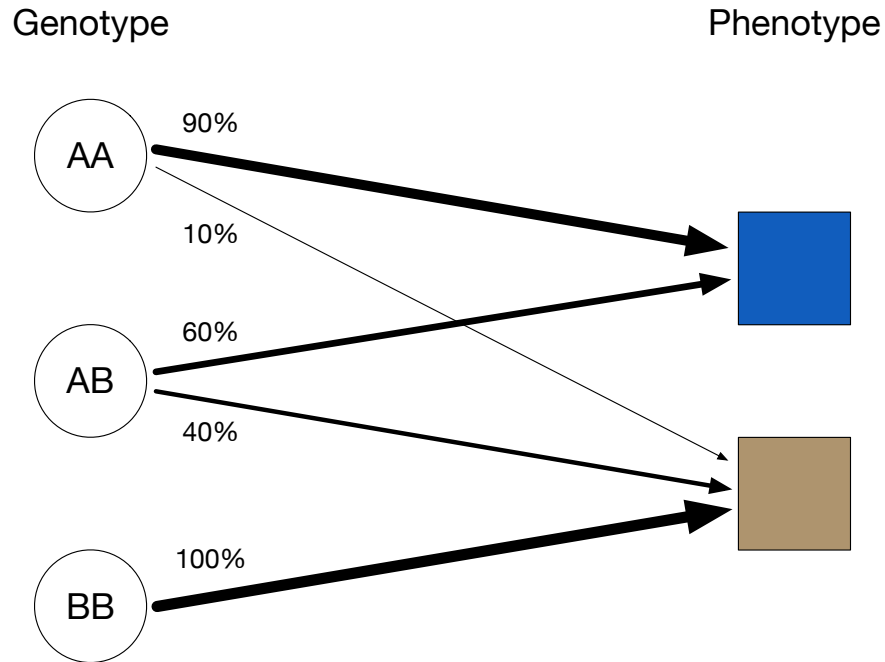
- A comprehensive resource of known polymorphisms
- rs numbers refer to specific genetic variants
- <https://www.ncbi.nlm.nih.gov/snp/>



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

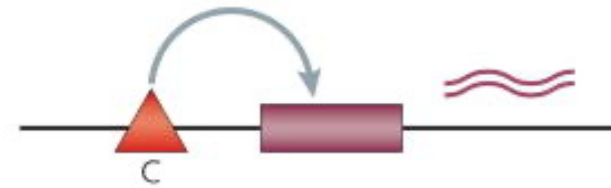
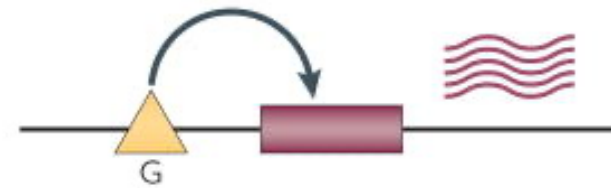
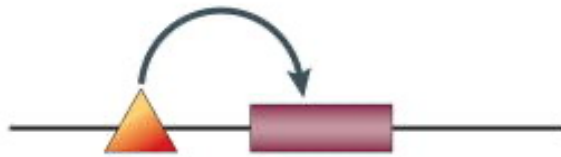
Pleiotropy



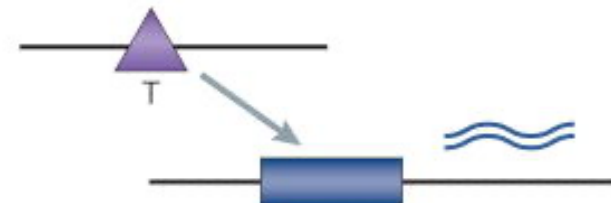
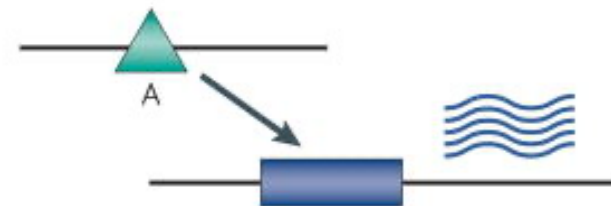
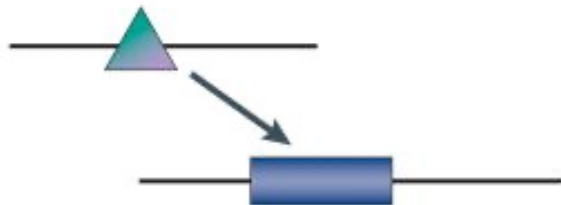
- Be careful to avoid terminology like "cancer gene" or "height gene"
- This is because genes often contribute to multiple phenotypes (i.e., they are **pleiotropic**)
- Example: A mutation in the *Frizzle* gene results in feathers that curve outward, fewer eggs laid, and high body temperatures

Cis and trans in genetics

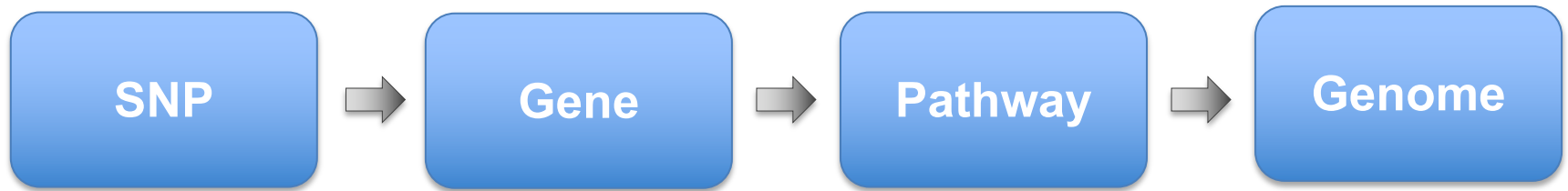
a *Cis* (local)



b *Trans* (distal)



Units of analysis in genetics



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

The mapping problem



- SNPs can affect more than one gene (or none)
- Gene to pathway mappings are hindered by ascertainment bias
- Mappings need not be one-to-one