Genomics and Disease

With an example classroom activity on GWAS

Andrew Webb. Temple University

June 27, 2017

Concept Review

- Gene
- Alleles
- Genotype
- Protein
- Translation
- Phenotype
- Mutations (i.e. SNPs)

What is a Genetic Disease?

Definition: A disease attributed to an abnormality in one's genome.

Genetic Diseases may be:

- **Monogenic**: caused by a single gene
- **Polygenic**: caused by multiple genes

Prominent Examples we will discuss:

- **Sickle Cell Disease (Anemia)**
- **Huntington's Disease**

What is the cause of Genetic Disease?

Definition: A disease attributed to an abnormality in one's genome.

But... What does that mean?

Example: **Hemoglobin β Chain (i.e. β-globin)**

Fragment of β-globin DNA Sequence

`5'- GTG - CAC - CTC - ACT - CCT - GAA - GAG -3'`

Associated Amino Acid (i.e. Protein) sequence

`Val - His - Leu - Thr - Pro - Glu - Glu`

What is the cause of Genetic Disease?

Definition: A disease attributed to an abnormality in one's genome.

But... What does that mean?

Example: **Hemoglobin β Chain (i.e. β-globin)**

Fragment of β-globin DNA Sequence

`5'- GTG - CAC - CTC - ACT - CCT - GTA - GAG -3'`

Associated Amino Acid (i.e. Protein) sequence

`Val - His - Leu - Thr - Pro - Val - Glu`

What is the cause of Genetic Disease?

This single mutation (or SNP) in β-globin is the cause of Sickle Cell Disease.

Note: β-globin is comprised of 147 amino acids (i.e. 441 nucleotides).

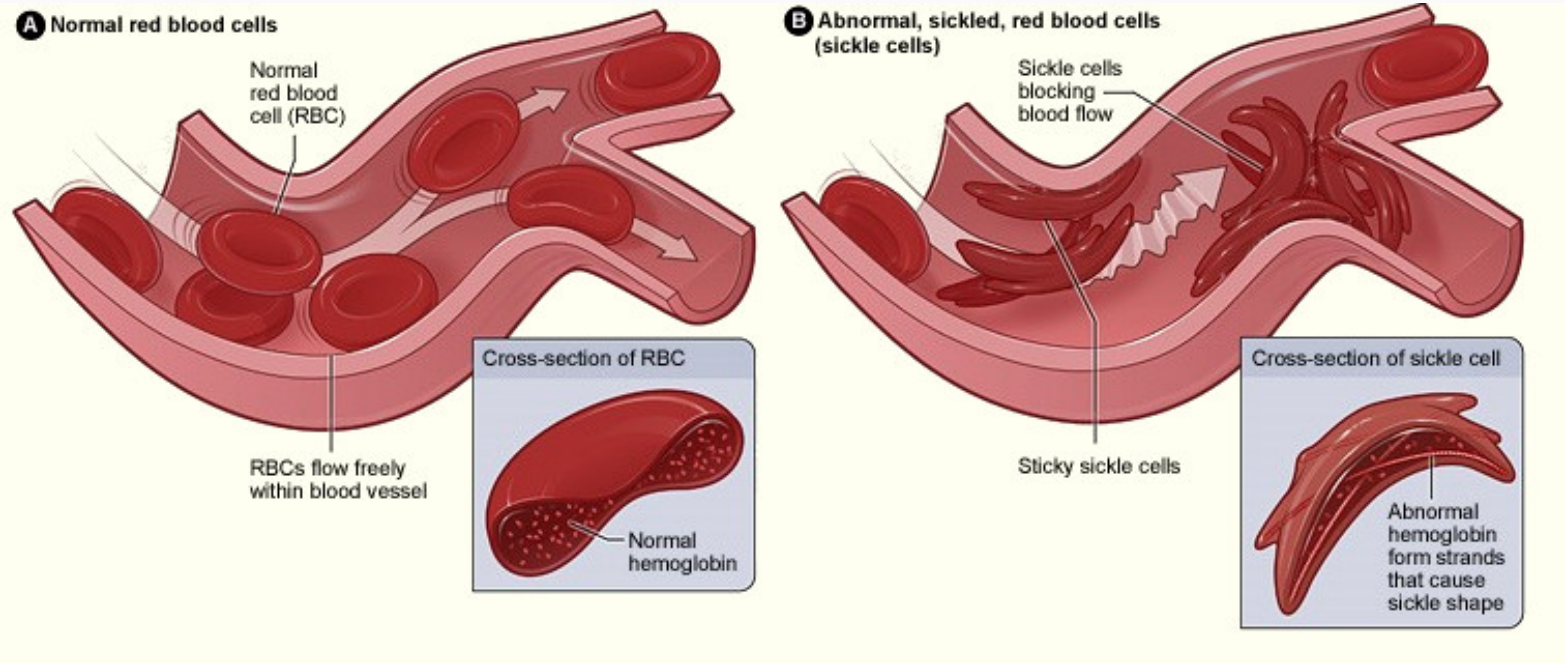
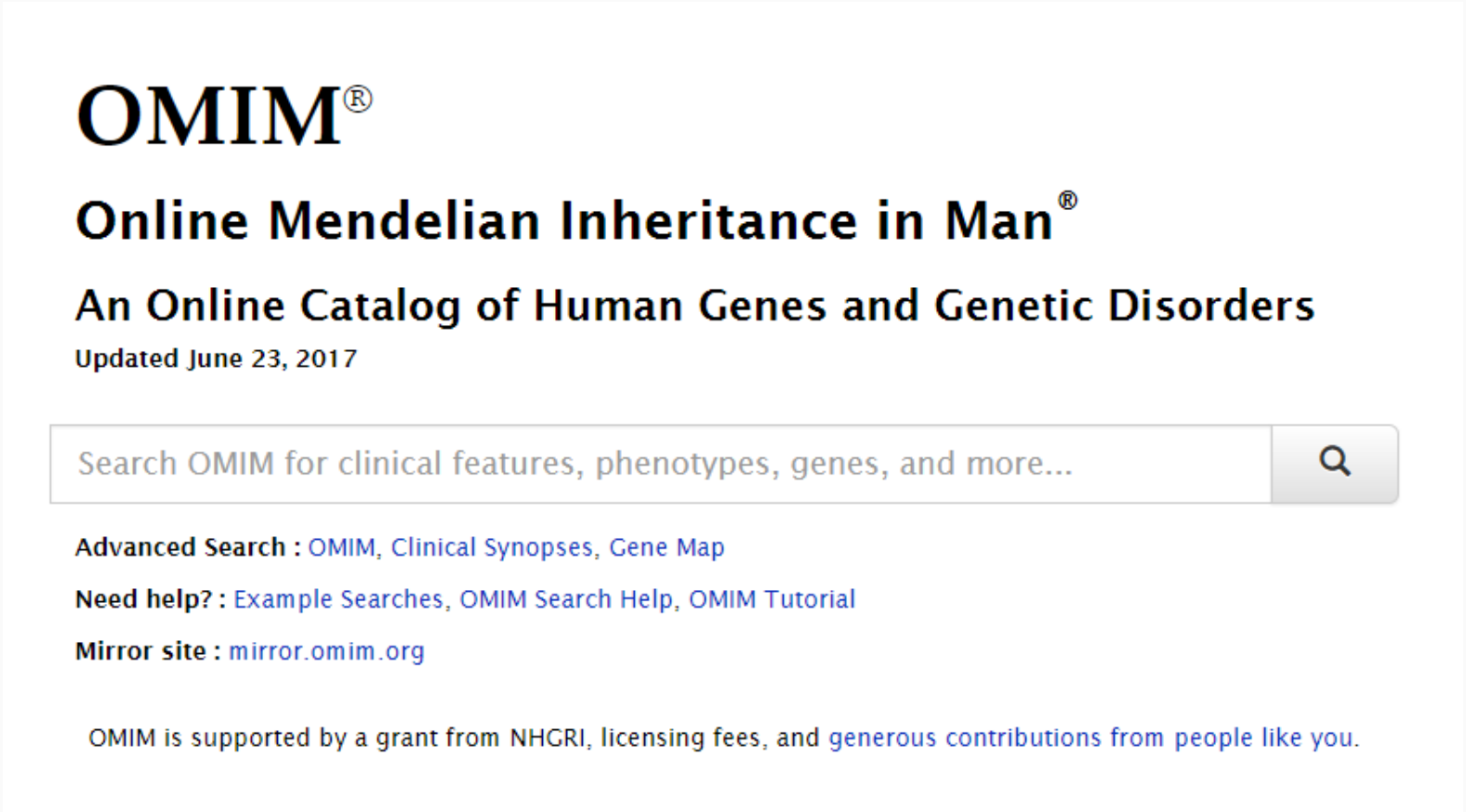


Image adapted from: https://www.nhlbi.nih.gov/health/health-topics/topics/sca

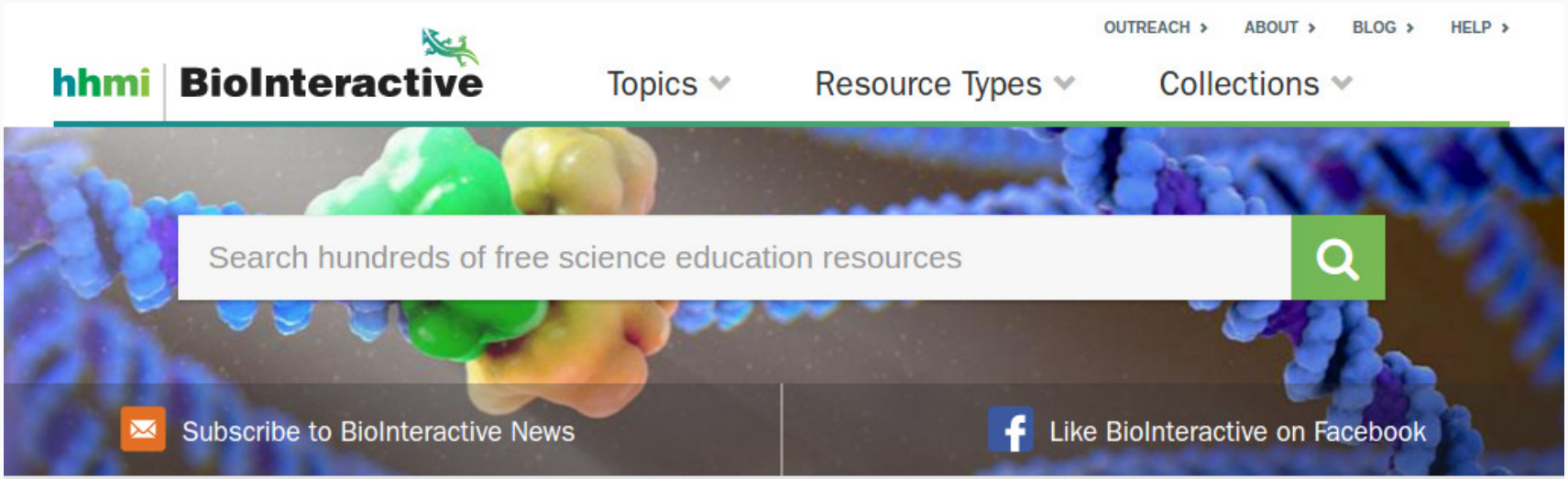
How do we connect students to these concepts?

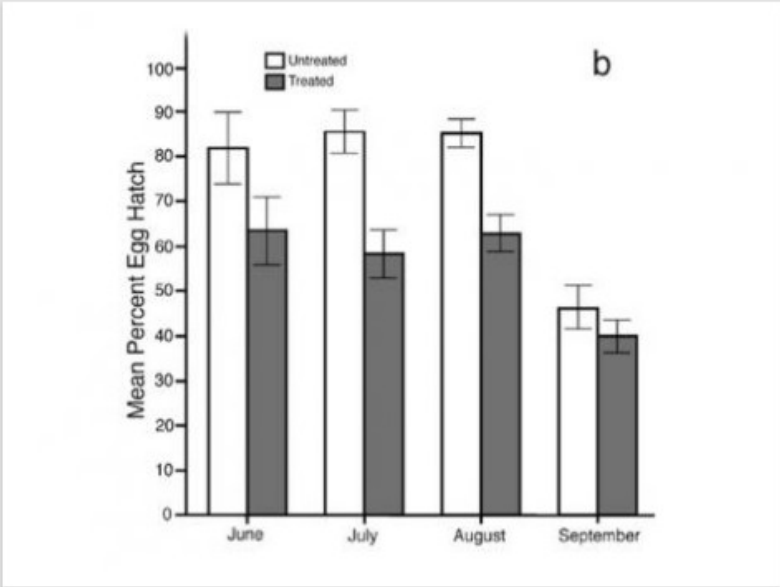
- Personalize the examples
  - In fun and safe ways
  - Example: Ability to taste Phenylthiocarbamide (PTC)
- This is sometimes difficult
  - Many sensitive topics
  - Especially genetic diseases
- Many free resources

Genetic Disease Resource

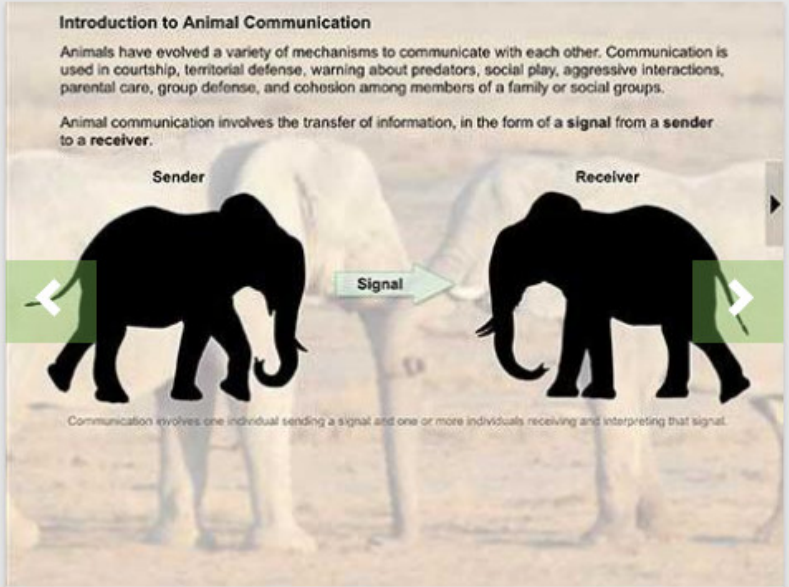


Classrom Activity Resource





DATA POINT

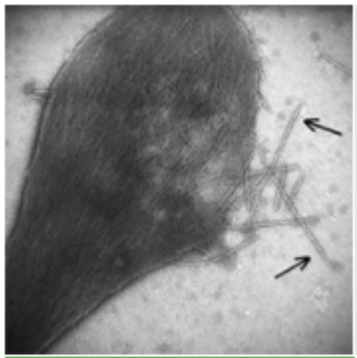


CLICK & LEARN

Activity on the Cause of Genetic Disease

How Do Fibers Form?

THIS IS PART OF: The Making of the Fittest: Natural Selection in Humans



Summary

A hands-on activity in which students construct models of sickle-cell hemoglobin fibers inside red blood cells to illustrate how changes in the structure of a protein can affect cell shape. Students are then asked to relate these changes to disease symptoms. Also available in Spanish.

Recommend

16 other people found this useful

SHARE THIS

Spanish versions of these resources are also available for download (en Español).

Appropriate for: high school biology (all levels), introductory college biology.

This hands-on activity serves as an extension to the Howard Hughes Medical Institute short film *The Making of the Fittest: Natural Selection in Humans*. Students will construct models of sickle cell hemoglobin inside red blood cells to illustrate how changes in the structure of a protein can affect cell shape.

Downloads

- Student Handout (PDF) 388K
- Teacher Materials (PDF) 233K
- Student Handout - Español (PDF) 455 KB
- Teacher Materials - Español (PDF) 330 KB

By downloading, you agree to the permissions to use this file.

The Challenges of Genetic Diseases

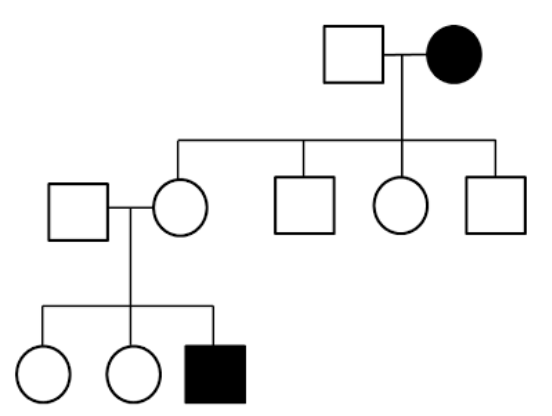
What are some of the major challenges posed by genetic diseases?

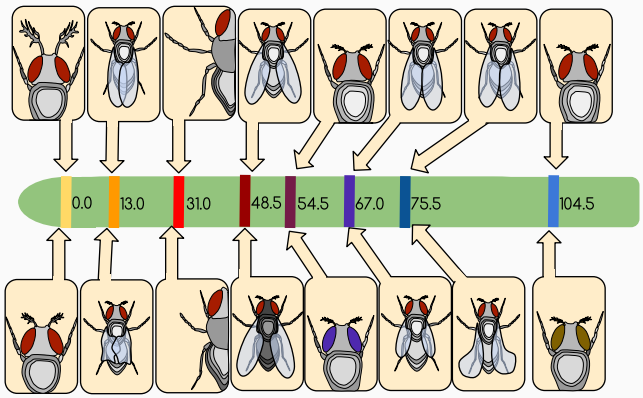
- May be caused by small changes to the genome
  - Human Genome: 3.2Gbp (i.e. 3.2 billion nucleotides)
- Disease variants may be caused by different mutations or genes
- May be caused by multiple genes
  - And potentially by multiple gene combinations
- The causative gene(s) may be completely unknown
  - Human Genes (protein-coding): 20,310

Question: How do we identify the mutations (SNPs) that associate with a genetic disease?

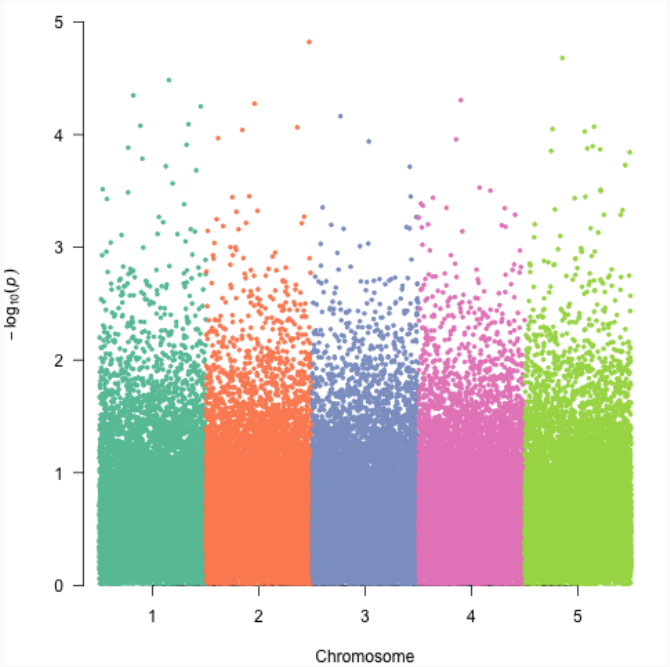
How do we identify these disease-associated SNPs?

In the Past - ”Classic” Methods





In the Present - Bioinformatic Methods



In the Past - ”Classic” Methods

Prior to the availability of genomes, the identification of these SNPs required a combination of:

- A detailed pedigree of a family (or families) with the genetic disease in question
  - Identify individuals homozygous for the genetic disease
- The creation of linkage maps (with known markers)
  - To locate the locus housing the SNPs

Example: This approach was used to understand the genetics of Huntington’s Disease

In the Present - Bioinformatic Methods

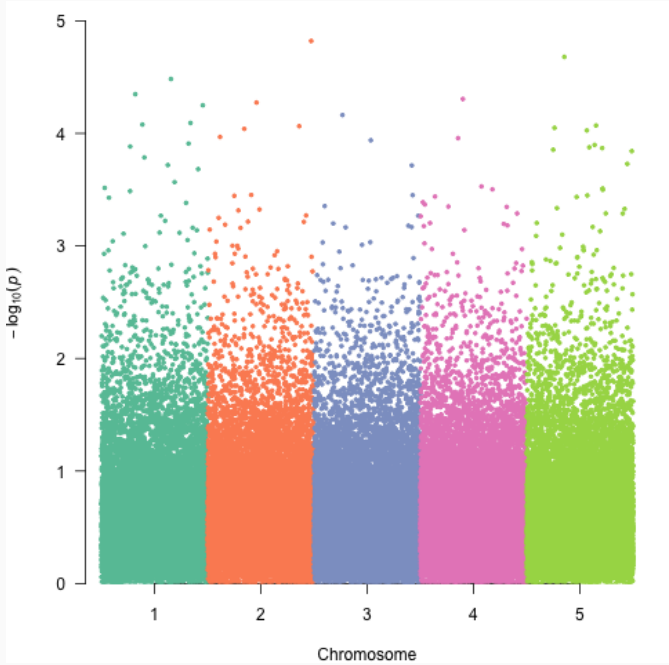
Following the release of a genome, the techniques used in the identification of disease-associated SNPs changed greatly:

- Obtain sequence information from individuals with the genetic disease
  - SNP Chips/Arrays
    - Collection of sites known to be polymorphic
  - Genome sequencing
- Association study
  - Simple SNP-based genetic linkage
  - Genome-wide association study (GWAS)

Genome-wide association study (GWAS)

Definition: GWAS compares the DNA sequence of individuals with various phenotypes for a trait or disease of interest.

- Used for mapping complex (polygenic) traits or diseases
- Used to identify SNPs within the genome that are associated with a trait or disease
- Only an association, cannot directly identify the causative gene(s).
- Due to the polygenic nature of the trait or disease, SNPs usually only have a small effect

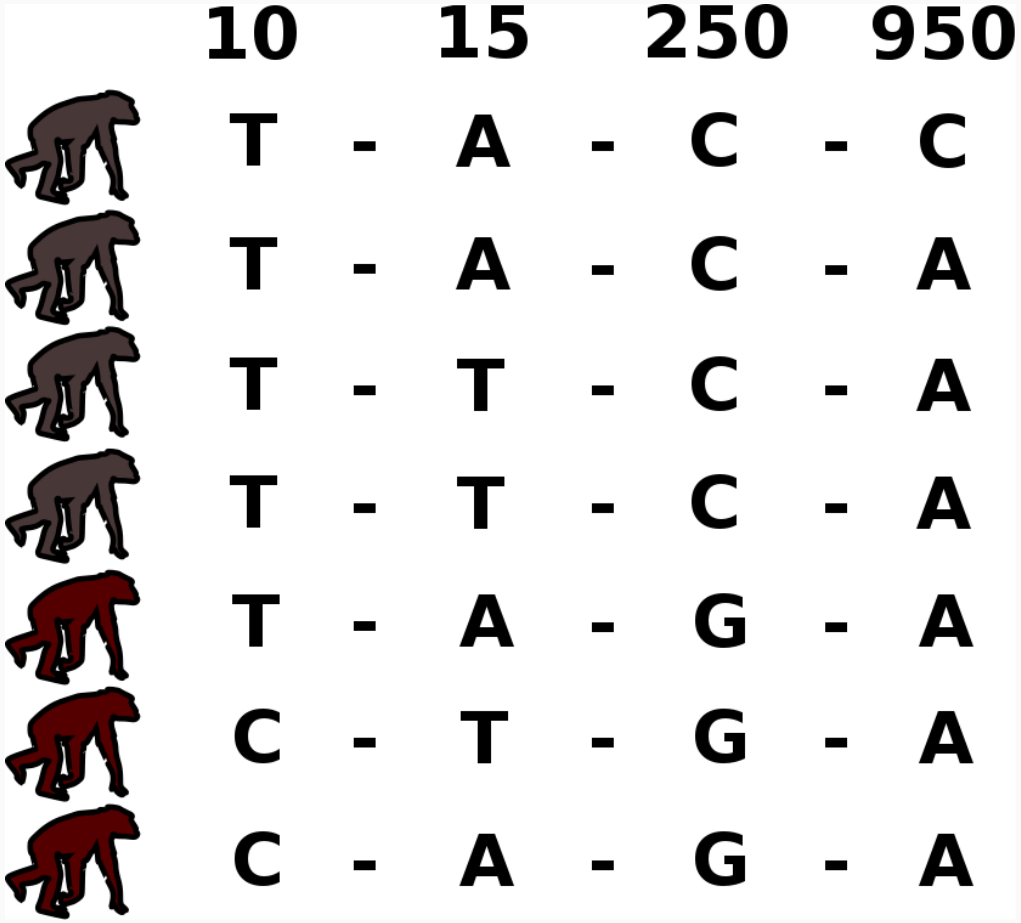


GWAS Basics

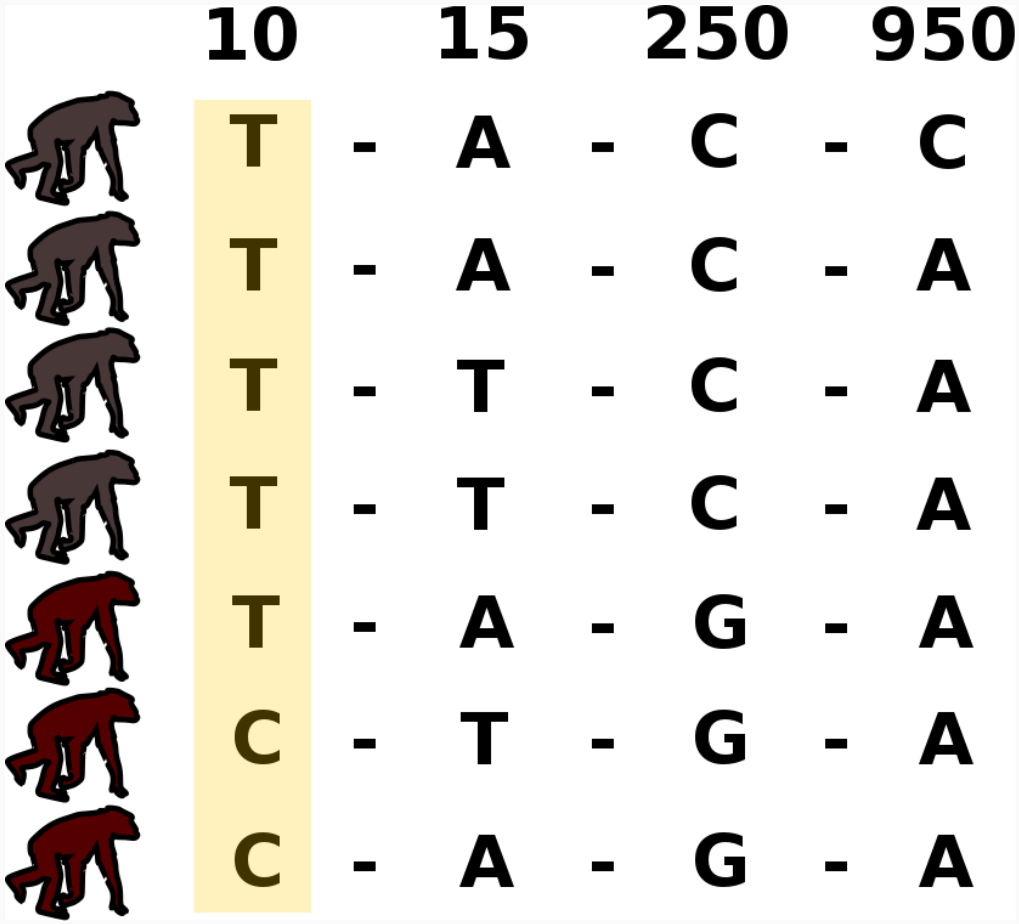
How is a GWAS performed?

- Sequence two groups: Case and Control
- Compare DNA sequence of the two groups
  - **Allele Counts**
  - Genotype Counts
- Use the counts to check for possible association
- Correct for multiple testing

Lets break down what that all means

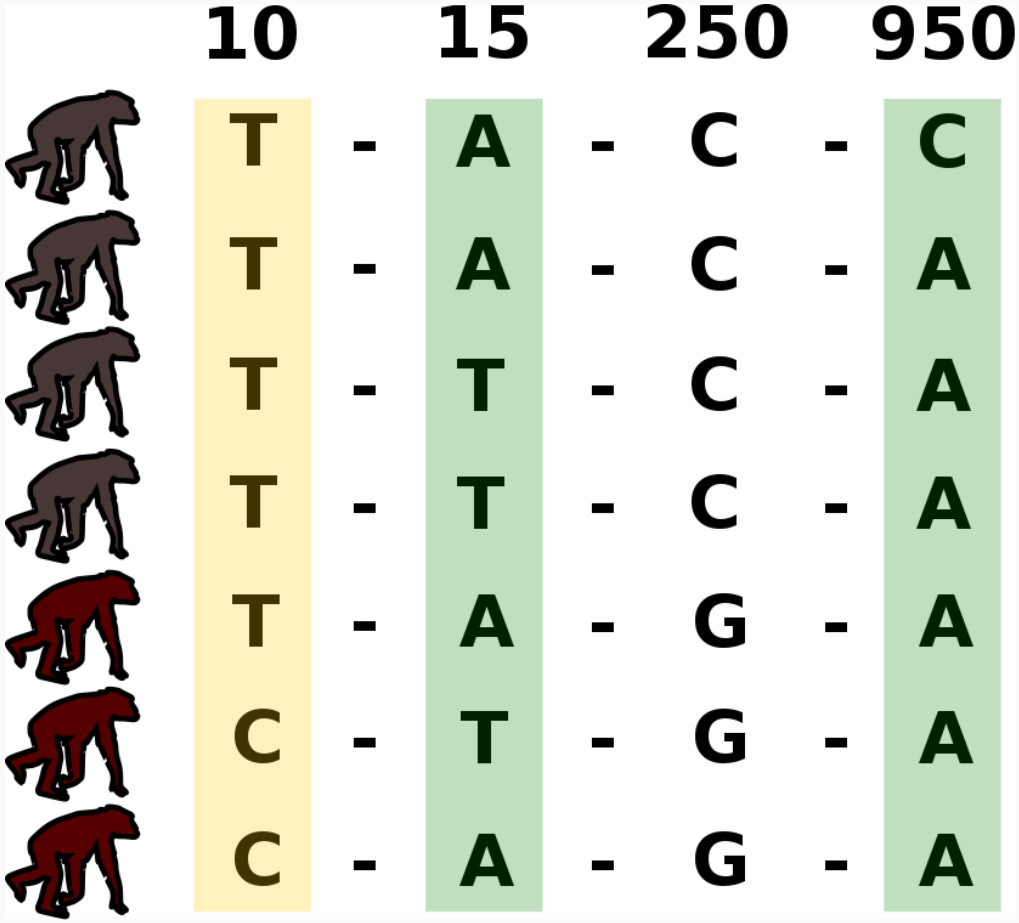


Lets break down what that all means



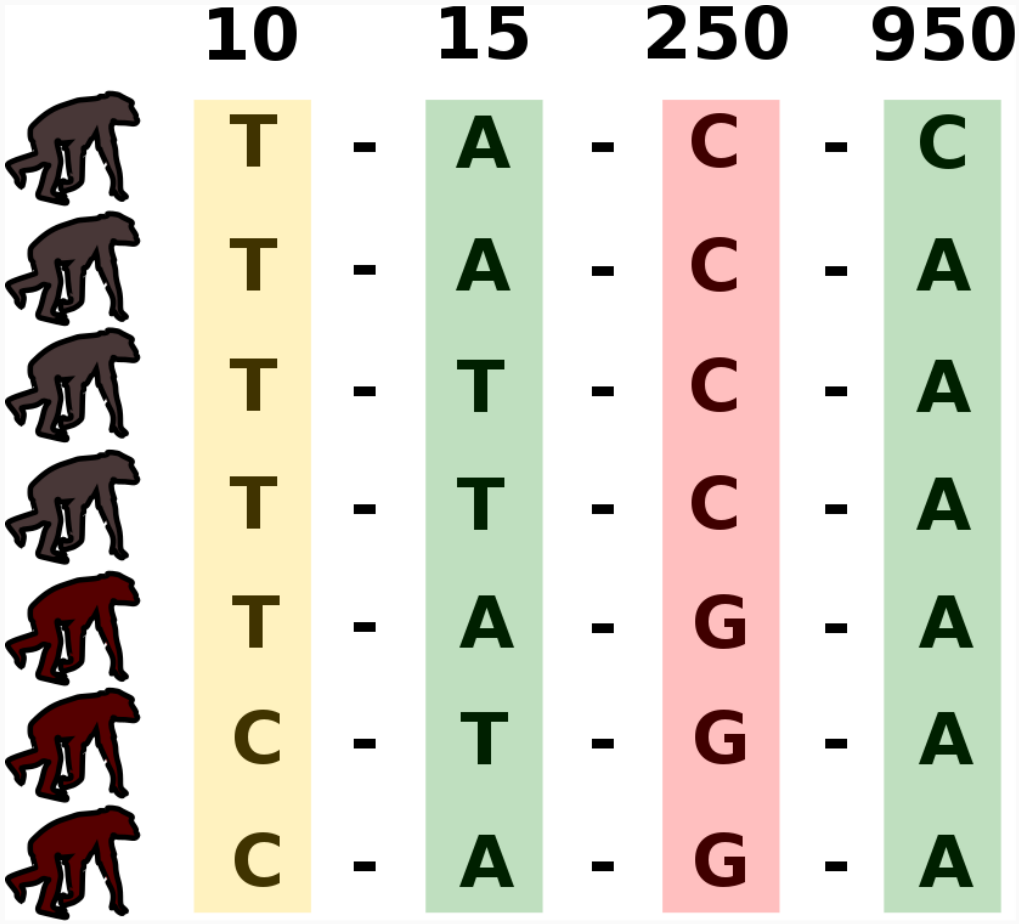
10. Normal T: 4, C: 0. Mutant T: 1, C: 2

Lets break down what that all means



15. Normal A: 2, T: 2. Mutant A: 2, T: 1

Lets break down what that all means

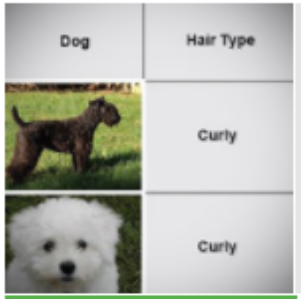


250. Normal C: 4, G: 0. Mutant C: 0, G: 3

Activity on Genome-wide association study (GWAS)

Mapping Genes to Traits in Dogs Using SNPs

THIS IS PART OF: **Medicine in the Genomic Era**



Summary

In this hands-on genetic mapping activity students identify single nucleotide polymorphisms (SNPs) correlated with different traits in dogs.

Recommend

198 other people found this useful

SHARE THIS

The activity complements a 29-minute lecture by Dr. Elinor Karlsson of the Broad Institute in Cambridge, MA (http://www.hhmi.org/biointeractive/dog-genomics-and-dogs-model-organisms), in which she discusses genome-wide association studies (GWAS). The activity consists of:

- A before-class assignment in which students read a press release from the National Institutes of Health (NIH) and answer questions on their student handout.
- An optional video lecture by Dr. Elinor Karlsson, explaining GWAS.
- An in-class component that explores how GWAS work by comparing single-nucleotide polymorphisms (SNPs) of dogs with different phenotypes and identifying correlations. This part of the activity has a hands-on component that uses SNP Cards.
- An optional statistical analysis activity that uses chi-square analysis.

Downloads

- Teacher Document (PDF) 670 KB
- Student Document (PDF) 5.6 MB
- SNP Cards (curly-straight) (PDF) 7.2 MB
- SNP Cards (short-long) (PDF) 8.0 MB
- Figure 2 (PPT) 123 KB
- Figure 3 (PPT) 193 KB

By downloading, you agree to the permissions to use this file.

Activity Genome-wide association study (GWAS)

	Chr11: 5,078,021	Chr11: 5,105,623	Chr11: 5,225,464	Chr11: 5,227,543	Chr11: 5,229,395	Chr11: 5,354,399	Chr11: 5,657,091
Case	AC	AA	AT	AT	AC	TT	GG
Case	AC	AA	AA	TT	CC	CC	GG
Case	CC	TA	AT	TT	AC	CT	GG
Case	AC	TA	AA	AT	AA	TT	CG
Case	CC	TA	AA	TT	AC	CT	CG
Case	AC	AA	AT	TT	CC	CC	CG
Control	CC	AA	TT	AA	CC	TT	GG
Control	AC	AA	TT	AA	CC	TT	CG
Control	AC	TA	TT	TA	CC	TT	GG
Control	CC	AA	TT	AA	CC	CT	CG
Control	AA	AA	TT	AA	CC	TT	GG
Control	CC	AA	TT	AA	CC	TT	GG

Activity Genome-wide association study (GWAS)

	Chr11: 5,078,021	Chr11: 5,105,623	Chr11: 5,225,464	Chr11: 5,227,543	Chr11: 5,229,395	Chr11: 5,354,399	Chr11: 5,657,091
Case	AC	AA	AT	AT	AC	TT	GG
Case	AC	AA	AA	TT	CC	CC	GG
Case	CC	TA	AT	TT	AC	CT	GG
Case	AC	TA	AA	AT	AA	TT	CG
Case	CC	TA	AA	TT	AC	CT	CG
Case	AC	AA	AT	TT	CC	CC	CG
Control	CC	AA	TT	AA	CC	TT	GG
Control	AC	AA	TT	AA	CC	TT	CG
Control	AC	TA	TT	TA	CC	TT	GG
Control	CC	AA	TT	AA	CC	CT	CG
Control	AA	AA	TT	AA	CC	TT	GG
Control	CC	AA	TT	AA	CC	TT	GG

Case	A: 4	A: 9	A: 9	A: 2	A: 5	T: 6	C: 3
Control	A: 4	A: 11	A: 0	A: 11	A: 0	T: 11	C: 2
Diff.	0	2	9	9	5	5	1

Case	C: 8	T: 3	T: 3	T: 10	C: 7	C: 6	G: 9
Control	C: 8	T: 1	T: 12	T: 1	C: 12	C: 1	G: 10
Diff.	0	2	9	9	5	5	1

Bioinformatics Example

Our GWAS identified two SNPs that are highly correlated with sickle cell disease.

How can we use this information to better understand the cause of sickle cell disease?

BLAST

In bioinformatics, **BLAST** for **B**asic **L**ocal **A**lignment **S**earch **T**ool is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

Different types of BLASTs are available according to the query sequences. For example, following the discovery of a previously unknown gene in the mouse, a scientist will typically perform a BLAST search of the human genome to see if humans carry a similar gene; BLAST will identify sequences in the human genome that resemble the mouse gene based on similarity of sequence. The BLAST algorithm and program were designed by Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers, and David J. Lipman at the National Institutes of Health and was published in the *Journal of Molecular Biology* in 1990 and cited over 50,000 times.[1]

How do we know if a difference is by chance

It might be possible that the differences we observed are due to random chance.

Is there a way to support our claim of association?

Couple methods: *Chi*^{2} and Fisher's Exact Test

	Case	Control	*Marginal Row Totals*
A	100 (92.5) [0.61]	85 (92.5) [0.61]	185
C	50 (57.5) [0.98]	65 (57.5) [0.98]	115
Marginal Column Totals	150	150	300 (Grand Total)

The chi-square statistic is 3.1727. The p-value is .074877. This result is *not* significant at $p < .05$.

Impact of GWAS (and other methods)

Scientists have successfully used GWAS (and other methods) to identify numerous disease-associated SNPs within the genome.

Condition	Gene	Variant(s)
ARSACS	SACS gene	1
Agenesis of the Corpus Callosum with Peripheral Neuropathy	SLC12A6 gene	1
Autosomal Recessive Polycystic Kidney Disease	PKHD1 gene	3
Beta Thalassemia and Related Hemoglobinopathies	HBB gene	10
Bloom Syndrome	BLM gene	1
Canavan Disease	ASPA gene	3
Congenital Disorder of Glycosylation Type 1a (PMM2-CDG)	PMM2 gene	2
Cystic Fibrosis	CFTR gene	28

Perspective

Question: How might the study of genetic diseases change in the near-future?

Advances, new technologies, and new analyses often bring change

- Cost of sequencing and advances in computing
  - Allows for large-scale analyses with higher resolution
  - UK10K
- CRISPR-CAS9
  - A bacterial immune defense from foreign genetic elements
  - Genome editing technology

Coding Activity

We will be exploring how to write a script to compare SNPs between control and case groups.