

MAPPING GENES TO TRAITS IN DOGS USING SNPs

OVERVIEW

This instructor guide provides support for a genetic mapping activity, based on the dog genome project data. 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 was created using actual DNA data from dog saliva that was collected and submitted by students who attended HHMI's 2013 Holiday Lectures on Science. The data were obtained and analyzed by Dr. Karlsson and colleagues at the Broad.

This 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.

KEY CONCEPTS AND LEARNING OBJECTIVES

- Comparing DNA sequences of many individuals reveals common variations across the genome.
- Some DNA variations may occur more frequently in individuals with one form of a trait than another.
- Variations correlated with a trait can point to the location of the gene (or genes) responsible for that trait.

After participating in this activity, students will be able to:

- Describe the key concepts in the article (and the video) in their own words.
- Connect the information in the article to the real data contained in the *SNP Cards*.
- Identify patterns in genomic data.
- Participate in a discussion of their interpretation of the genomic data.

CURRICULUM CONNECTIONS

Curriculum	Standards
NGSS (April 2013)	HS-LS-1, HS-LS3-3, Science and Engineering Practices 2 and 4, Crosscutting Concepts 1, 2, and 6
AP (2012–13)	Big Idea 3, EU 3.A.3, 3.C.1, and SP 2, 5, and 6
IB (2016)	3.2, 3.5, Aims 1, 4, and 5

KEY TERMS

nucleotide, base pair, allele, single-nucleotide polymorphism (SNP), genotype, phenotype, genome sequence, genome-wide association study (GWAS)

TIME REQUIREMENTS

The introductory section on the student worksheet, the NIH press release, and the six reading-comprehension questions that follow can be assigned as homework and should require about 30 to 40 minutes to complete, depending on the student's reading ability. The optional Karlsson video can also be assigned as homework to supplement and enhance the introduction to the topic (key segment is 12 minutes). Analyzing the SNP data will require roughly 20 minutes of in-class time, depending on the amount of class discussion. The optional chi-square statistical analysis will require an additional 20 minutes.

SUGGESTED AUDIENCE

This activity was designed for all levels of high school biology and biology-related electives (i.e., genetics and biotechnology). It is a great resource for connecting students with the current state of genetic research and could lead to a study of evolution as well, since dogs are an excellent example of artificial selection. In a college classroom, this activity would be appropriate for an introductory biology or genetics course.

PRIOR KNOWLEDGE

In order to get the most out of this activity, students should know what genes and mutations are and should be familiar with the relationship between genotype and phenotype. It would be helpful for students to understand the relationship between a single gene and the genome. The activity includes an optional chi-square analysis for students with some background or who need further practice with statistical methods.

MATERIALS

- *Variants in Three Genes Account for Most Dog Coat Differences* – NIH press release. <https://www.nih.gov/news-events/news-releases/variants-three-genes-account-most-dog-coat-differences>
- *Medicine in the Genomic Era: Dog Genomics and Dogs as Model Organisms*, video lecture with Elinor Karlsson (<http://www.hhmi.org/biointeractive/dog-genomics-and->

dogs-model-organisms). The key segment is from 1:56 to 12:15 minutes of the lecture.

- A set of *SNP Cards* displaying SNP alleles on chromosome 32 in dogs with short and long hair.
- A set of *SNP Cards* displaying SNP alleles on chromosome 27 in dogs with curly and noncurly hair.

PROCEDURE AND ANSWERS TO QUESTIONS

Part 1: Read and Answer Questions

Distribute the student worksheet to students. Ask students to read the NIH press release, *Variants in Three Genes Account for Most Dog Coat Differences*, and answer the associated questions before class. (Students will return to this article at the end of the activity and analyze some of the data reported in the article.)

1. *How many genes account for the wide variety of coat textures in dogs?* **Three.**
2. *In two or three sentences, describe how researchers went about identifying these genes.*
Answers will vary. Researchers conducted genome-wide scans of specific SNPs in thousands of dogs and a variety of breeds. When they analyzed the results, they found that just three genetic variants accounted for the majority of dog coat types.
3. *In this reading, why are SNPs referred to as “genetic markers” or “signposts”?* **A single-nucleotide polymorphism (SNP) is a DNA sequence variation that occurs commonly within a population. For example, a SNP may replace the nucleotide adenine (A) with the nucleotide cytosine (C) in a particular stretch of DNA. Researchers know the precise location of that SNP in the genome. Once a SNP is correlated with a particular trait (meaning that it occurs more frequently in individuals with that trait), researchers focus on that region of the genome and sequence it. In most cases, the SNP itself does not contribute to the trait, but it is close to the trait-causing stretch of DNA, serving as a marker or signpost for the location.**
4. *Why do you think it is important to analyze the DNA of many dogs when doing this research?* **An individual dog will have many variations compared to any other dog. Researchers analyze the DNA of many dogs to find variations that are common in one group compared to another. A variation is not a useful marker if it only occurs in one or a few dogs. Students may also mention that researchers need to look at many dogs to find SNPs associated with one trait and not another.**
5. *Do humans have SNPs?* **Yes.**
6. *How might the dog genome project benefit humans?* **Student answers will vary. The most likely answer will indicate that the human genome is very similar to the dog genome. If we can find the genes responsible for all coat types in dogs, we can use the same approach to identify genes involved in complex diseases in humans.**

Part 2: Introduction to a GWAS









Discuss the article and address student questions or points of confusion. Suggested discussion

questions:

- What is the relationship between a complete genome, a chromosome, and a gene?
- How are SNPs used in a GWAS?
- How do GWAS allow researchers to find genes that cause disease?
- What is the difference between a SNP and a mutation?

This part of the activity begins with a simple table exercise to introduce students to a GWAS. In the student worksheet, students examine Table 1, which displays SNPs at 17 different positions in the genome sequences of eight different dogs (four with white fur and four with black fur). Be sure to point out that, for simplicity's sake, the table only displays one nucleotide per SNP (one allele). In reality, each SNP locus has two nucleotides, one from each chromosome. To determine whether the SNPs in the table are correlated with white fur, students compare the SNPs of the black and white dogs, search for patterns in the data, and answer questions.

Table 1. Nucleotides at 17 Different Loci in Two Groups of Dogs.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	C	A	C	G	G	T	T	T	G	A	T	G	A	A	C	G	T
	C	T	C	C	A	T	T	C	G	A	T	T	G	A	C	T	A
	C	A	A	G	A	C	T	C	G	A	T	G	G	G	C	G	T
	C	A	C	G	A	T	T	C	G	G	T	G	G	A	C	G	T
	C	A	C	G	A	T	A	C	G	A	T	G	G	A	C	G	T
	C	T	A	G	A	T	A	T	G	A	T	G	A	A	C	G	A
	C	T	C	C	A	T	A	C	G	A	T	T	A	A	C	T	T
	C	A	C	G	G	T	A	C	G	G	T	G	A	A	C	G	T

1. What is the locus number (or numbers) for the 100% correlated SNP(s)? 7
2. What is the locus number (or numbers) for the somewhat correlated SNP(s)? The somewhat correlated SNP is at locus 13. Students may also include loci 2, 6, and 14, which are more weakly correlated.
3. What is the locus number (or numbers) for the uncorrelated SNP(s)? Any of the loci not mentioned above.
4. What are two possible explanations for why a SNP is correlated with a phenotype like coat color? A SNP can be correlated with a phenotype because it causes that phenotype (e.g., white fur) or because it is linked to, or located close to, the allele that causes the phenotype.

Part 3: Identify Correlations Using Real Data

Following that warm-up, students proceed to the card activity to identify most correlated

SNPs using two different sets of SNP Cards: one set for coat length (short/long) and one for coat type (curly/straight). Students can work individually or in groups.

Distribute the coat-length cards.

Give a set of coat-length SNP Cards to each student or group. The data displayed on the coat-length card shows SNPs at seven loci on chromosome 32 in 12 dogs with either short or long hair. Note that this time the SNPs are shown as two nucleotides for each locus. The chromosome number (32), followed by a seven-digit nucleotide number, indicates the SNP loci.

Ask students to shuffle the cards around and organize them to find correlations between certain SNPs and coat-length trait. Give students enough time to examine the cards and write down their answers.

Students then follow the directions in their student documents to identify SNPs correlated with coat length. Their data should look as follows:

Chr32 7420804

Allele	Short Coat	Long Coat	Difference
T	4	4	0
C	8	8	0
		Total number of differences	0

Chr32 7472206

Allele	Short Coat	Long Coat	Difference
A	9	12	3
G	3	0	3
		Total number of differences	6

Chr32 7473337

Allele	Short Coat	Long Coat	Difference
T	3	12	9
G	9	0	9
		Total number of differences	18

Chr32 7479580

Allele	Short Coat	Long Coat	Difference
T	9	12	3
C	3	0	3
		Total number of differences	6

Chr32 7482867

Allele	Short Coat	Long Coat	Difference
A	5	0	5
G	7	12	5
		Total number of differences	10

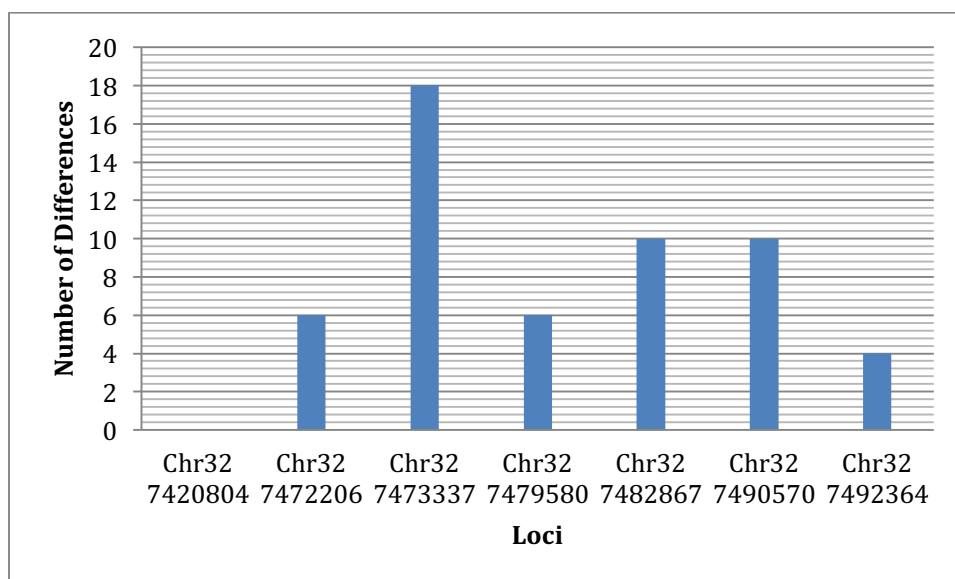
Chr32 7490570

Allele	Short Coat	Long Coat	Difference
--------	------------	-----------	------------

T	6	11	5
C	6	1	5
		Total number of differences	10

Chr32 7492364

Allele	Short Coat	Long Coat	Difference
C	3	1	2
G	9	11	2
		Total number of differences	4



1. Which SNP (or SNPs) is most strongly correlated with coat length? **Chr32 7473337**
2. Which SNP (or SNPs) is least strongly correlated with coat length? **Chr32 7420804**

Distribute the coat-texture cards.

Give a set of coat-texture *SNP Cards* to each student or group. The data displayed on the coat-texture card shows SNPs at six loci on chromosome 27 in 10 dogs with either curly or straight hair. Students follow the directions in their Student Worksheet, record their tallies in the charts provided, and determine the strongest correlation to coat type. Their results should look as shown on the following page.

Chr27 5525002

Allele	Curly Coat	Straight Coat	Difference
C	4	5	1
T	6	5	1
		Total number of differences	2

Chr27 5541113

Allele	Curly Coat	Straight Coat	Difference
C	7	4	3
T	3	6	3
		Total number of differences	6

Chr27 5542806

Allele	Curly Coat	Straight Coat	Difference
C	4	10	6
T	6	0	6
		Total number of differences	12

Chr27 5545082

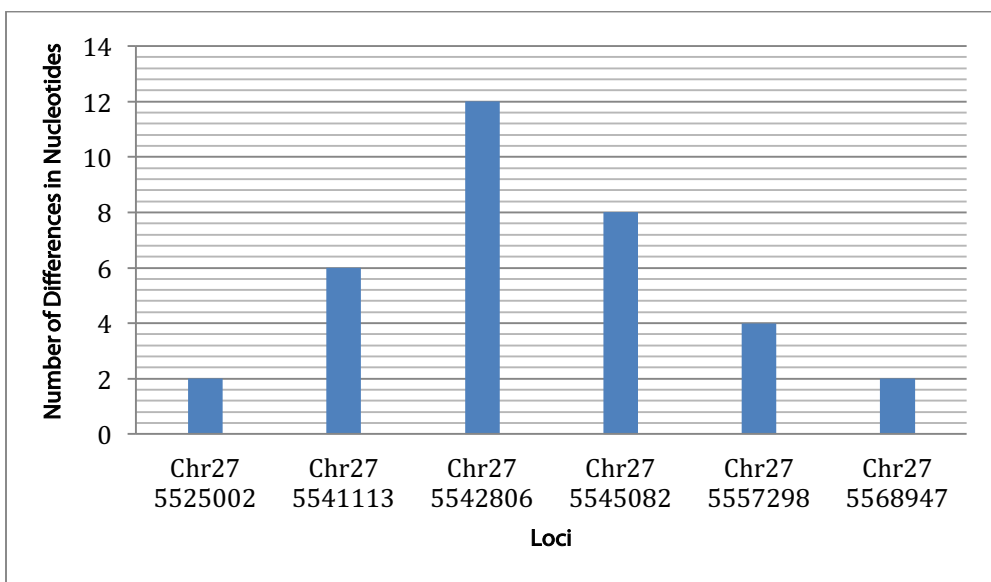
Allele	Curly Coat	Straight Coat	Difference
A	4	8	4
G	6	2	4
		Total number of differences	8

Chr27 5557298

Allele	Curly Coat	Straight Coat	Difference
C	4	2	2
T	6	8	2
		Total number of differences	4

Chr27 5568947

Allele	Curly Coat	Straight Coat	Difference
C	5	6	1
T	5	4	1
		Total number of differences	2



3. Which SNP (or SNPs) shows the strongest correlation to coat texture? **Chr27 5542806**
4. If you were to look for a gene involved in coat length, on which region of the genome would you focus your analysis? What about a gene for coat texture? Write your answer below.

Students should indicate that they would look for a gene involved in coat length on chromosome 32, somewhere around nucleotide 7473337. For the gene involved in coat texture, they would look on chromosome 27 near nucleotide 5542806.

Part 4: Chi-Square Analysis (Optional)

It is possible that the frequency of alleles between the two groups of dogs could be different just by chance. On the other hand, if these differences did not occur by chance, the correlations might indicate that the SNPs mark the positions of genes involved in coat length or texture. To determine whether the correlations in this activity are statistically significant, students conduct a chi-square analysis. This part of the activity is optional and requires students to be familiar with the chi-square test.

The results are as follows:

SNP locus: **Chr32 7473337**

Total number of G alleles = **9**

Expected number of G alleles = **4.5**

Total number of T alleles = **15**

Expected number of T alleles = **7.5**

Allele	Short Coat Observed	Long Coat Observed	Expected
G	9	0	4.5
T	3	12	7.5

4. *Chi-square value calculation:*

$$(4.5)^2/4.5 + (4.5)^2/7.5 + (4.5)^2/4.5 + (4.5)^2/7.5 = 4.5 + 2.7 + 4.5 + 2.7 = 14.4$$

5. *Record the p-value for the SNP with the strongest correlation:* **0.001**

6. *What does this p-value tell you?* **The correlation is statistically significant.**

7. *Which SNP has the strongest correlation for curly versus noncurly coat?* **Chr27 5542806**

8. *Is this correlation statistically significant?* **Yes**

9. *What is the evidence for your choice?* **The chi-square value for this locus is 8.56, which is much larger than 3.841.**

Part 5: What Are the Genes?

Students examine an additional table summarizing data collected on three genes responsible for seven major coat types of purebred dogs. Using the table, students answer questions that guide interpretation of the data. This research example relates to the article that students read at the beginning of the activity.

1. *What phenotype is the ancestral allele of the KRT71 gene associated with in dogs?* **The ancestral allele of KRT71 is associated with the straight phenotype because the dogs with the - allele all have straight fur and with the + allele curly fur.**

2. *Which breed of dog shown above exhibits the phenotype for a variant allele for FGF5 but an ancestral allele for KRT71?* **The Golden Retriever and Bearded Collie have a variant FGF5 allele, different from wolves, and the ancestral allele of the KRT71 gene.**

TEACHING TIPS

- The lecture by Dr. Karlsson (<http://www.hhmi.org/biointeractive/dog-genomics-and-dogs-model-organisms>) is an excellent introduction to GWAS in dogs. You could assign a segment of it as homework or consider showing chapters 2 to 7 (time 1:56 to 12:15 min.) in class, prior to the SNP activity.
- Students can work individually or in groups.
- You will need to print a complete set of SNP cards for each student or small group of students. You could laminate the cards for repeat use.
- Consider showing students how GWAS data is plotted; an example can be found at http://www.broadinstitute.org/igv/viewing_gwas. GWAS data is often shown as a "Manhattan plot" with SNPs represented as dots, color-coded by chromosome. The x-axis shows the locations of the SNPs in the genome. The larger the dot and the higher it is on the scale, the more significant the association with a trait.

- You may want to discuss with students that, although the human genome was obtained by pooling sequence data from several individuals, the reference dog genome consists of data from just one dog (Tasha). Scientists also sequenced dogs from other breeds and put all the differences they found compared to Tasha's genome in the public database called dbSNP. [<http://www.ncbi.nlm.nih.gov/SNP/>]
- Although this activity only shows a few SNPs on two chromosomes, a GWAS uses millions of SNPs across the entire genome and a computer sifts through the data to find correlations. This activity uses simple examples that can be done without a computer to demonstrate the concepts involved.
- When describing the difference between a SNP that causes a trait (causative SNP) and a correlated SNP that does not cause a trait, take the opportunity to discuss genetic linkage, which is typically taught during the Mendelian genetics unit. Even if a SNP in a GWAS shows strong correlation with a trait, it takes further investigation to see if it is a causative SNP (i.e., in the noncoding or coding region of a gene) or if it's simply closely linked to the causative SNP.
- After introducing GWAS, you could share scientific papers with students that describe using GWAS in a human genomics study to identify genes that cause disease. You can also point out additional areas of research in which GWAS are used, such as in agriculture.
- This activity could serve as a lead-in to evolution since dogs are an excellent example of artificial selection.

SOCIAL MEDIA SUGGESTIONS

- Students could write an analysis of the activity as a blog post (on a class blog or individual student blogs). Once posted, the comments feature could be used to facilitate peer review and discussion.
- As an alternative optional extension project, students could create a "Dog Genotype and Phenotype" infographic, intended for a veterinarian's office. Free, online infographic creation tools ([Easel.ly](#) or [Piktochart](#)) make this sort of brochure or handout easy to create.

AUTHORS

This activity was written by Melissa Csikari, Colonial Forge High School, Stafford, VA; Elinor Karlsson, Ph.D., Broad Institute of MIT, Cambridge, MA; Laura Bonetta, Ph.D., and Eriko Clements, Ph.D., HHMI.

They were edited by Laura Bonetta, Ph.D., HHMI and Robin Heyden, consultant; copyedited by Linda Felaco.

Reviewers: David McDonald, Ph.D., North Carolina Central University; Vince Buonaccorsi, Ph.D., Juniata College; Elinor Karlsson, Ph.D., Broad Institute.