INTRODUCTION
Meet Tasha, a boxer dog (Figure 1). In 2005, scientists used Tasha’s DNA to obtain the first complete dog genome sequence, which contains 2.4 billion pairs of nucleotides over 39 pairs of chromosomes! Genome sequence data like this is useful for many reasons. In this activity, you’ll see how this data is used to find genes associated with different traits, such as fur color in dogs. The approach you’ll learn about is called a genome-wide association study (GWAS), and it can be used to learn about the genes of any organism, not just dogs.

PART 1: Introduction to GWAS
Read the information below to learn more about GWAS and its applications, then answer the questions that follow.

What Is GWAS?
GWAS is a method for identifying the genes associated with an organism’s collection of traits, or phenotype. It involves searching the genomes of many individuals — for example, many different dogs — to find DNA differences, or variations, associated with particular traits and phenotypes. After sequencing Tasha’s genome, for example, scientists sequenced and compared the genomes of many dogs from a variety of breeds. They found millions of common variations among these genomes. Some of these variations were associated with the color, length, or texture of a dog’s fur. You’ll learn more about these variations later in this activity.

1. In general, why do you think GWAS is useful? What kinds of problems could GWAS be used to solve?

What Are SNPs?
The variations found in GWAS are usually common variations in DNA sequences called single nucleotide polymorphisms, or SNPs (pronounced “snips”). A SNP is a variation in a single nucleotide, at a particular position in the genome, that occurs in over 1% of the population. Different variations of a SNP may be called SNP variants or alleles. Figure 2 shows two alleles for a common SNP in dogs.

As shown in Figure 2, SNPs are labeled based on the chromosome they are found on (e.g., chromosome 37, which is abbreviated as chr37) and their “nucleotide position” (e.g., 25,734,258), which is determined by counting from one end of the chromosome to the other. For the SNP shown in Figure 2, dogs may have either a C or an A. These two versions of the SNP are called the C allele and the A allele.

![Figure 2. DNA sequences from chromosome 37 (chr37) in three dogs. Two sequences are shown for each dog, one from each parental chromosome. (Complementary DNA strands are not shown.)](image)

The red letters and arrow indicate a common SNP at nucleotide position 25,734,258. Most dogs have either a C or an A at this position.
2. List the three combinations of alleles (C and A) that a dog could have for the SNP shown in Figure 2.

Using SNPs to Find Genes Associated with Particular Traits

The locations of the SNPs in a genome serve as “markers” or “signposts” for the locations of genes associated with particular traits. A GWAS uses associations between a particular SNP and a trait of interest to predict how close that SNP is to a gene responsible for that trait. How does this work? In general, the closer two DNA sequences are to one another on the same chromosome, the more likely they are to be inherited together. So, if a SNP is close to a gene, it is likely to be inherited with that gene — and will thus be associated with the gene’s trait.

One way to find SNPs associated with a certain trait is by comparing groups with different versions of that trait. In a GWAS looking for genes that affect dog fur color, for example, we could compare the SNPs of two groups: dogs with black fur and dogs with white fur. We would then determine which SNPs are significantly more common in dogs with black fur compared to dogs with white fur. These SNPs are “markers” for regions of the dog genome that contain genes affecting fur color.

3. Why do you think SNPs are referred to as “markers” or “signposts”?

Figure 3 shows several possibilities for why a SNP is associated with a certain trait. The SNP may be in the gene that causes the trait or in a regulatory area for that gene. If so, the SNP could directly affect the gene’s function and the resulting trait. However, some SNPs in or near a gene may have no effect on the gene or its trait.

4. Consider the different types of SNPs shown in Figure 3: associated, unassociated, and causative (including both noncoding and coding).
   a. Which types of SNPs affect protein production or function for the gene of interest?
   b. Which types of SNPs might be identified in a GWAS?
GWAS in the News

Read the following news release, which describes a GWAS study with dogs. Note that a dog’s coat refers to its fur or hair.

Variants in Three Genes Account for Most Dog Coat Differences

Variants in just three genes acting in different combinations account for the wide range of coat textures seen in dogs — from the poodle’s tight curls to the beagle’s stick-straight fur. A team led by researchers from the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health, reports these findings today in the advance online issue of the journal Science.

“This study is an elegant example of using genomic techniques to unravel the genetic basis of biological diversity,” said NHGRI Scientific Director Eric Green, M.D., Ph.D. “Genomics continues to gain new insights from the amazing morphological differences seen across the canine species, including many that give clues about human biology and disease.”

Until now, relatively little was known about the genes influencing the length, growth pattern and texture of the coats of dogs. The researchers performed a genome-wide scan of specific signposts of DNA variation, called single nucleotide polymorphisms, in 1,000 individual dogs representing 80 breeds. These data were compared with descriptions of various coat types. Three distinct genetic variants emerged to explain, in combination, virtually all dog hair types.

“What’s important for human health is the way we found the genes involved in dog coats and figured out how they work together, rather than the genes themselves,” said Elaine A. Ostrander, Ph.D., chief of the Cancer Genetics Branch in NHGRI’s Division of Intramural Research. “We think this approach will help pinpoint multiple genes involved in complex human conditions, such as cancer, heart disease, diabetes and obesity.”

Artificial selection, at the heart of breeding for desirable traits in domesticated animals, has yielded rapid change in a short span of canine history. While researchers estimate that modern dog breeds diverged from wolves some 15,000 years ago, the genetic changes in the dog genome that create multiple coat types are more likely to have been pursued by breeders in just the past 200 years. In fact, short-haired breeds, such as the beagle, display the original, more wolf-like versions of the three genes identified in the study.

Modern dog breeds are part of a unique population structure, having been selectively bred for many years. Based on this structure, the researchers were able to break down a complex phenotype — coat — into possible genetic variations. “When we put these genetic variants back together in different combinations, we found that we could create most of the coat varieties seen in what is among the most diverse species in the world — the dog,” Dr. Ostrander said. “If we can decipher the genetic basis for a complex trait such as the dog’s coat, we believe that we can do it as well with complex diseases.”


Answer the following questions to check your understanding of the reading.

5. How many genes account for the wide variety of coat types in dogs?

6. In two or three sentences, describe how scientists identified these genes.

7. Why do you think it is important to analyze the DNA of many dogs when doing this research?

8. Humans have SNPs too. In general, how might GWAS studies with dogs benefit humans?
PART 2: Applying GWAS to Dog Fur Color

Let’s explore how a GWAS works using a simple example that compares two groups of dogs: dogs with black fur and dogs with white fur. Table 1 shows the dogs’ SNP alleles at 17 specific locations in the genome. These specific locations in the genome are called loci (singular: locus). The SNP alleles at each locus are represented by two nucleotides, one from each parental chromosome.

Table 1. SNP alleles at 17 different loci in dogs with black fur (first four rows) and dogs with white fur (last four rows).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CC</td>
<td>AT</td>
<td>CC</td>
<td>G</td>
<td>G</td>
<td>AA</td>
<td>TC</td>
<td>TT</td>
<td>CC</td>
<td>GG</td>
<td>AA</td>
<td>TT</td>
<td>GT</td>
<td>AG</td>
<td>AA</td>
<td>CC</td>
<td>GG</td>
</tr>
<tr>
<td>2</td>
<td>CC</td>
<td>AT</td>
<td>AC</td>
<td>GG</td>
<td>GG</td>
<td>TT</td>
<td>TT</td>
<td>CC</td>
<td>GG</td>
<td>AG</td>
<td>TT</td>
<td>GG</td>
<td>AG</td>
<td>AG</td>
<td>CC</td>
<td>GG</td>
<td>AT</td>
</tr>
<tr>
<td>3</td>
<td>CC</td>
<td>AA</td>
<td>AC</td>
<td>CG</td>
<td>GG</td>
<td>TT</td>
<td>TT</td>
<td>CT</td>
<td>GG</td>
<td>AA</td>
<td>TT</td>
<td>GT</td>
<td>AG</td>
<td>AG</td>
<td>CC</td>
<td>TT</td>
<td>AT</td>
</tr>
<tr>
<td>4</td>
<td>CC</td>
<td>AA</td>
<td>AC</td>
<td>CG</td>
<td>AG</td>
<td>TT</td>
<td>TT</td>
<td>CT</td>
<td>GG</td>
<td>GG</td>
<td>TT</td>
<td>GG</td>
<td>AG</td>
<td>AG</td>
<td>CC</td>
<td>GT</td>
<td>AT</td>
</tr>
<tr>
<td>5</td>
<td>CC</td>
<td>AT</td>
<td>CC</td>
<td>CG</td>
<td>AA</td>
<td>TT</td>
<td>AA</td>
<td>CT</td>
<td>GG</td>
<td>AA</td>
<td>TT</td>
<td>GT</td>
<td>AA</td>
<td>AA</td>
<td>CC</td>
<td>GT</td>
<td>AT</td>
</tr>
<tr>
<td>6</td>
<td>CC</td>
<td>AT</td>
<td>CC</td>
<td>CG</td>
<td>AA</td>
<td>TT</td>
<td>AA</td>
<td>TC</td>
<td>GG</td>
<td>AA</td>
<td>TT</td>
<td>GG</td>
<td>AG</td>
<td>AA</td>
<td>CC</td>
<td>GG</td>
<td>TT</td>
</tr>
<tr>
<td>7</td>
<td>CC</td>
<td>AT</td>
<td>CC</td>
<td>CG</td>
<td>AA</td>
<td>TT</td>
<td>AA</td>
<td>CC</td>
<td>GG</td>
<td>AA</td>
<td>TT</td>
<td>GG</td>
<td>AA</td>
<td>AA</td>
<td>CC</td>
<td>GT</td>
<td>AT</td>
</tr>
<tr>
<td>8</td>
<td>CC</td>
<td>AA</td>
<td>CC</td>
<td>GG</td>
<td>AG</td>
<td>TT</td>
<td>AA</td>
<td>TC</td>
<td>GG</td>
<td>GG</td>
<td>TT</td>
<td>GG</td>
<td>AA</td>
<td>AA</td>
<td>CC</td>
<td>TT</td>
<td>AT</td>
</tr>
</tbody>
</table>

If a SNP is found much more frequently in dogs with white fur than in dogs with black fur, the SNP is associated with the white fur color.

9. Give two possible reasons for why a SNP would be associated with a trait like fur color.

To determine whether any of the SNPs in Table 1 are associated with fur color, you can compare the SNPs of the dogs with black fur to those of the dogs with white fur. A SNP is completely associated with fur color if all dogs with white fur share the same alleles at that position, and all dogs with black fur share different alleles at that position. A SNP that is completely associated with a trait is likely located within or close to a gene responsible for that trait.

10. Which SNP in Table 1 do you think is completely associated with fur color? Explain the reasoning for your choice.

A SNP is completely unassociated with fur color if its alleles occur with equal frequency in dogs with black fur and dogs with white fur. A SNP that is completely unassociated with a trait is unlikely to be located within or near the gene responsible for that trait.

11. Which SNPs in Table 1 do you think are completely unassociated with fur color? Explain the reasoning for your choices. (Hint: There are five in total.)
The other SNPs in Table 1 have varying strengths of association with fur color. You’ll learn more about how to evaluate the strength of an association in the next part of this activity. For the question below, make your best guess based on what you’ve learned so far.

12. Which SNP in Table 1 do you think has the next strongest association with fur color, after the completely associated SNP you identified in Question 10? Explain the reasoning for your choice.

**PART 3: Identifying Associations for Dog Coat Length**

During a GWAS, scientists must evaluate the strength of associations between traits of interest and millions of SNPs to figure out which SNPs are located closest to the genes of interest. You’ll now learn more about how to evaluate the strengths of associations by doing a card activity that uses genetic data from real dogs.

First, you’ll investigate which SNPs are associated with dog coat length. Your instructor will give you 12 “SNP Cards” for dog coat length (examples shown in Figure 4). Each card represents DNA sequence data from one dog. The sequences show seven SNPs at different loci on chromosome 32.

Take a few moments to look at the cards. Sort the cards into two groups, dogs with long coats and dogs with short coats, then compare sequences of the two groups.
13. Which SNP on the cards do you think has the strongest association with dog coat length? Explain the reasoning for your choice.

To see if the data support your choices, you’ll now learn a way to evaluate which SNPs have the strongest association with a trait. Follow these steps for each of the seven SNPs shown in the tables below:

- Count the number of times each allele for the SNP appears on the cards. For example, if a dog has the alleles TT for that SNP, count the T allele twice. If it has the alleles TC, count the T allele once and the C allele once.
- Record the numbers of each allele, across all dogs in a group, in the appropriate boxes in the SNP’s table.
- Calculate the difference in the numbers of each allele between the two groups of dogs. Record that number in the last column of the SNP’s table.
- Add the differences for all the alleles, then record the total number of differences in the last box of the SNP’s table. A larger total difference indicates a stronger association between that SNP and the trait (in this case, coat length).

14. Complete the SNP tables below using the method described above. The first two tables have been completed for you as examples.

**chr32 7420804**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Short Coat</th>
<th>Long Coat</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td><strong>Total number of differences</strong></td>
<td><strong>0</strong></td>
<td></td>
</tr>
</tbody>
</table>

**chr32 7472206**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Short Coat</th>
<th>Long Coat</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td><strong>Total number of differences</strong></td>
<td><strong>6</strong></td>
<td></td>
</tr>
</tbody>
</table>

**chr32 7473337**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Short Coat</th>
<th>Long Coat</th>
<th>Difference</th>
</tr>
</thead>
</table>

|        | **Total number of differences** |                  |

**chr32 7479580**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Short Coat</th>
<th>Long Coat</th>
<th>Difference</th>
</tr>
</thead>
</table>

|        | **Total number of differences** |                  |

**chr32 7482867**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Short Coat</th>
<th>Long Coat</th>
<th>Difference</th>
</tr>
</thead>
</table>

|        | **Total number of differences** |                  |

**chr32 7490570**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Short Coat</th>
<th>Long Coat</th>
<th>Difference</th>
</tr>
</thead>
</table>

|        | **Total number of differences** |                  |
15. Using your results from the tables above, graph the **total number of differences** in alleles for each of the seven SNPs.

16. Which of these SNPs has the *strongest* association with coat length?

17. Which of these SNPs has the *weakest* association with coat length?

18. Based on these data, would you revise your answer to Question 13? Why or why not?

19. Based on these seven SNPs, where would you search if you wanted to find a gene involved in dog coat length?

20. Based on what you’ve learned in Part 3, would you revise your answer to Question 12 from Part 2? Why or why not?

**PART 4: Identifying Associations for Dog Coat Texture**

Now, you’ll investigate which SNPs are associated with dog coat texture. Your instructor will give you 10 “SNP Cards” for dog coat texture. Again, each card represents DNA sequence data from one dog. The sequences show SNPs at six loci on chromosome 27. As marked on the cards, five of the dogs have curly coats, and five of the dogs have straight coats.

21. Fill out the following tables using the same method as before.
22. Using your results from the tables above, graph the **total number of differences** in alleles for each of the six loci.
23. Which of these SNPs has the strongest association with coat texture?

24. Based on these six SNPs, where would you search if you wanted to find a gene involved in dog coat texture?

PART 5: Which Genes Determine Dog Coat Traits?

Now you know the basic idea of how to identify SNPs associated with certain traits. Scientists look in the region of the associated SNPs to find genes that may be responsible for the traits. So which genes are responsible for the dog coat traits that you explored?

Using GWAS, scientists analyzed the DNA from over 1,000 dogs from 80 recognized breeds with different coat types — including long and short coats, curly and wire (stiff, rough hair) coats, and coats with furnishings (tufts of hair over the eyes and around the mouth). Their study also included several gray wolves, which are ancestors to modern domesticated dogs. Gray wolves have short, straight coats without furnishings.

The scientists identified the three SNPs that had the strongest associations with different coat phenotypes. These SNPs occur within the genes *FGF5*, *RSPO2*, and *KRT71*, which are the three genes that you read about in the news release at the end of Part 1. Each of these three genes has multiple alleles. They each have an ancestral allele, which is the version of the gene found in gray wolves. They also have more recent alleles, which are versions of the genes found in some dog breeds. The more recent alleles differ from the ancestral alleles by single-nucleotide changes in their DNA.

Figure 5 below shows how various combinations of ancestral and more recent alleles account for the seven major coat phenotypes of purebred dogs.
25. Which dog breed in Figure 5 has the ancestral allele for all three genes, similar to gray wolves?

26. Which dog breeds in Figure 5 have a more recent allele for FGF5 but an ancestral allele for KRT71?

You can use Figure 5 to figure out which alleles are associated with certain coat types. For example, if you look at the alleles of the FGF5 gene in different dog breeds, you can see that the ancestral allele (−) is associated with short coats, and the more recent allele (+) is associated with long coats.

27. Which coat type is the ancestral allele of the KRT71 gene associated with?

Identifying which genes affect dog coat types is important for dog breeders, but there are benefits beyond dog breeding. It turns out these genes produce proteins that regulate a variety of processes in all mammals, not just coat variations in dogs. For example, the FGF5 gene (full name fibroblast growth factor 5) plays a role in embryonic development, cell growth, morphogenesis, tissue repair, and tumor growth.

28. How might understanding the functions of genes in dogs help us better understand human health?

29. The methods described in this activity can be used to study the genes of many different organisms. Pick an organism other than dogs or humans that you are interested in. Describe a specific problem or question that you could investigate by doing a GWAS with the organism you picked.

Figure 5. Dog coat phenotypes and alleles for three different genes. Coat phenotypes are listed in the left-hand column of the table, and the genes associated with each phenotype are listed along the top.

A minus sign (−) indicates that the dog has the ancestral allele for the gene. A plus sign (+) indicates that the dog has a more recent allele.
EXTENSION: Chi-Square Test of Independence Analysis

In Parts 3 and 4 of this activity, you identified SNPs that appear to be associated with coat length and coat texture in dogs. You determined the associations based on how frequently different SNP alleles appeared in different groups of dogs. However, it’s possible that allele frequencies you saw happened just by chance, meaning the SNPs aren’t actually associated with the genes. So how can you determine whether the associations you found are statistically significant — that is, unlikely to be due to chance alone?

To determine if the association between two factors — for example, a particular SNP and coat length — is statistically significant, you can do a chi-square test of independence analysis, which has the following steps:

1. Calculate expected values. You calculate the results (e.g., numbers of SNP alleles in different groups of dogs) that you would expect due to chance.
2. Calculate the chi-square value. You calculate a quantity called the chi-square value using your expected values from Step 1 and the actual data.
3. Find the P value. You use the chi-square value from Step 2 to determine whether the association is statistically significant.

Go through each of the following sections to learn more about these steps.

Step 1: Calculate Expected Values

In this step, you will calculate the results you would expect due to chance if there was no association between the two factors you are looking at. The hypothesis that the two factors are not associated is called the null hypothesis. In this case, the null hypothesis will be that a particular SNP is not associated with coat length, meaning that any differences in the distribution of the SNP’s alleles between dogs with long and short coats occurred just by chance. Your goal is to see if you can reject this null hypothesis, which would mean that the SNP is associated with coat length.

First, you will need to calculate the expected values for each of the SNP’s alleles — that is, the number of times that allele would appear in dogs with long and short coats under the null hypothesis. Under the null hypothesis, you would not expect to see any difference in the distribution of the SNP’s alleles between dogs with long and short coats. As a result, you’d expect the allele to appear equally frequently in both groups of dogs. So, the expected value for the allele in each group would be the total number of times the allele occurs across both groups, divided by 2.

Example

The SNP at the locus chr32 7492364, which is one of the SNPs you explored in Part 3, has two alleles: C and G. The C allele occurs 3 times in dogs with a short coat and 1 time in dogs with a long coat, so:
- total number of C alleles = 3 + 1 = 4
- expected number of C alleles = 4/2 = 2

The G allele occurs 9 times in dogs with a short coat and 11 times in dogs with a long coat, so:
- total number of G alleles = 9 + 11 = 20
- expected number of G alleles = 20/2 = 10

So, the expected value of the C allele is 2 (meaning we’d expect it to appear 2 times in each group of dogs), and the expected value of the G allele is 10.
1. Calculate the expected numbers of each allele for the SNP that had the strongest association with coat length (your answer to Question 16 in Part 3). Show your work.

Step 2: Calculate the Chi-Square Value

You will now calculate a quantity called the chi-square value ($\chi^2$), which depends on the expected values you just calculated and on the observed number of alleles for the SNP. The equation for the chi-square value is:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

- $\chi^2$ is the chi-square value.
- $O$ and $E$ are observed and expected values — in this case, the observed and expected numbers of alleles for the SNP. The larger the difference between $O$ and $E$, the larger the chi-square value.
- $\sum$ is a symbol that means summation, or adding together. In this case, we will calculate the fraction on the right using values for each allele, then add all the fractions together.

Example

Let’s calculate the chi-square value for the SNP at the locus chr32 7492364, which was described in Step 1. We’ll record the observed and expected values for each allele in the following table.

<table>
<thead>
<tr>
<th>Allele</th>
<th># Observed in Short Coats</th>
<th># Observed in Long Coats</th>
<th># Expected in Each Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>9</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

The chi-square value for this SNP is calculated as follows:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

- for C in short coats: $\frac{(3 - 2)^2}{2} = 0.5$
- for G in short coats: $\frac{(9 - 10)^2}{10} = 0.1$
- for C in long coats: $\frac{(1 - 2)^2}{2} = 0.5$
- for G in long coats: $\frac{(11 - 10)^2}{10} = 0.1$

$$\chi^2 = 0.5 + 0.1 + 0.5 + 0.1 = 1.2$$

2. Complete the table below for the SNP that had the strongest association with coat length. (You calculated the expected numbers of alleles for this SNP in Step 1 above.)

<table>
<thead>
<tr>
<th>Allele</th>
<th># Observed in Short Coats</th>
<th># Observed in Long Coats</th>
<th># Expected in Each Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Calculate the chi-square value for this SNP. Show your work.
Step 3: Find the P Value

You will now use the chi-square value you calculated in Step 2 to determine whether the association between your two factors (the SNP and coat length) is statistically significant. You will do this by finding the P value, which is the probability that the results you observed occurred by chance — that is, the probability that the null hypothesis is true and your two factors are not associated. A P value of 0.05 indicates that there is only a 5% chance that your factors are not associated. We will consider P values of 0.05 or lower as statistically significant. In other words, if the P value is 0.05 or lower, it’s very likely that your two factors are associated.

First, calculate a quantity called the degrees of freedom (df), which depends on the number of categories for your two factors:

\[
df = (\text{# categories for factor } 1 - 1) \times (\text{# categories for factor } 2 - 1)
\]

In this case, there are 2 categories for the SNP (the two different alleles) and 2 categories for coat length (short and long). So:

\[
df = (\text{# of SNP categories } - 1) \times (\text{# of coat categories } - 1) \\
= (2 - 1) \times (2 - 1) \\
= 1
\]

Next, use the df you found to determine the P value for your chi-square value. This involves looking at a chi-square distribution table, like Table 1. As shown in Table 1, when you have a df of 1, a chi-square value larger than 3.84 gives a statistically significant P value (0.05 or lower).

**Table 1.** A chi-square distribution table. Degrees of freedom (df) are shown on the left and P values are shown on the top. The other numbers in the table are the corresponding chi-square values ($X^2$).

<table>
<thead>
<tr>
<th>df</th>
<th>0.995</th>
<th>0.975</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.002</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000393</td>
<td>0.000982</td>
<td>1.642</td>
<td>2.706</td>
<td>3.841</td>
<td>5.024</td>
<td>5.412</td>
<td>6.635</td>
<td>7.879</td>
<td>9.550</td>
<td>10.828</td>
</tr>
</tbody>
</table>

**Example**

Let’s find the P value for the SNP at the locus chr32 7492364. As shown in Step 2, this SNP had a chi-square value of 1.2. As shown above, the df is 1.

Go to the row in Table 1 that corresponds to a df of 1, then compare the chi-square values in that row. We see that a chi-square value of 1.2 would have a P value between 0.975 (which corresponds to $X^2 = 0.000982$) and 0.2 (which corresponds to $X^2 = 1.642$). Since a P value in this range would be larger than 0.05, the probability of our results occurring by chance is higher than the 5% cutoff for statistical significance.

As a result, the association between this SNP and coat length is not statistically significant. That means we cannot reject the null hypothesis, so it’s possible that our results occurred by chance and that the SNP is not actually associated with coat length.
4. What is the $P$ value for the SNP that had the strongest association with coat length? (You calculated the chi-square value for this SNP in Step 2 above.)

5. What does this $P$ value tell you?

6. Which SNP had the strongest association with coat texture in Part 4?

7. Is this association statistically significant? What is the evidence for your choice?