



Genome - Wide Linkage Mapping

Introduction

This activity is based on the work of Dr. Christine Seidman et al that was published in *Circulation*, 1998, vol 97, pgs 2043-2048.

Background

The goal in genome mapping is to find the gene in which a disease causing mutation is located. Sometimes it is not possible to find the exact gene, in which case the location of the mutation is determined as closely as possible.

Why genome mapping?

- Genetic testing
- Gene therapy
- Choosing treatments
- Understand molecular mechanisms underlying the disease
- Directions for further research

Steps in genome mapping

- Construct a pedigree
- Determine pattern of inheritance
- Select genomic markers to use in co-segregation studies
- Examine the selected genomic markers in affected vs unaffected individuals
- Compare results, looking for markers which are co-inherited with the mutation
- Examine the region in which the mutation is found. Do any known genes make sense?

Atrial Septal Defect

ASD is a common congenital heart malformation. In this disease the septum, or wall, between the two atria is either incompletely formed or extremely weak. If this septum is extremely weak, an aneurysm can occur resulting in a hole between the two atria. When the two atria are no longer separated, blood can be exchanged. Because pressure is higher in the left side of the heart, atrial septal defect results in atrial blood flowing from the left atrium into the right atrium. This defect is usually treated surgically. Uncorrected ASD can cause right heart volume overload and premature death.

There are many different genes that control heart development. A mutation in any of a number of different genes can lead to atrial septal defect (ASD). In this activity we will map the location of the gene leading to ASD in one particular family.

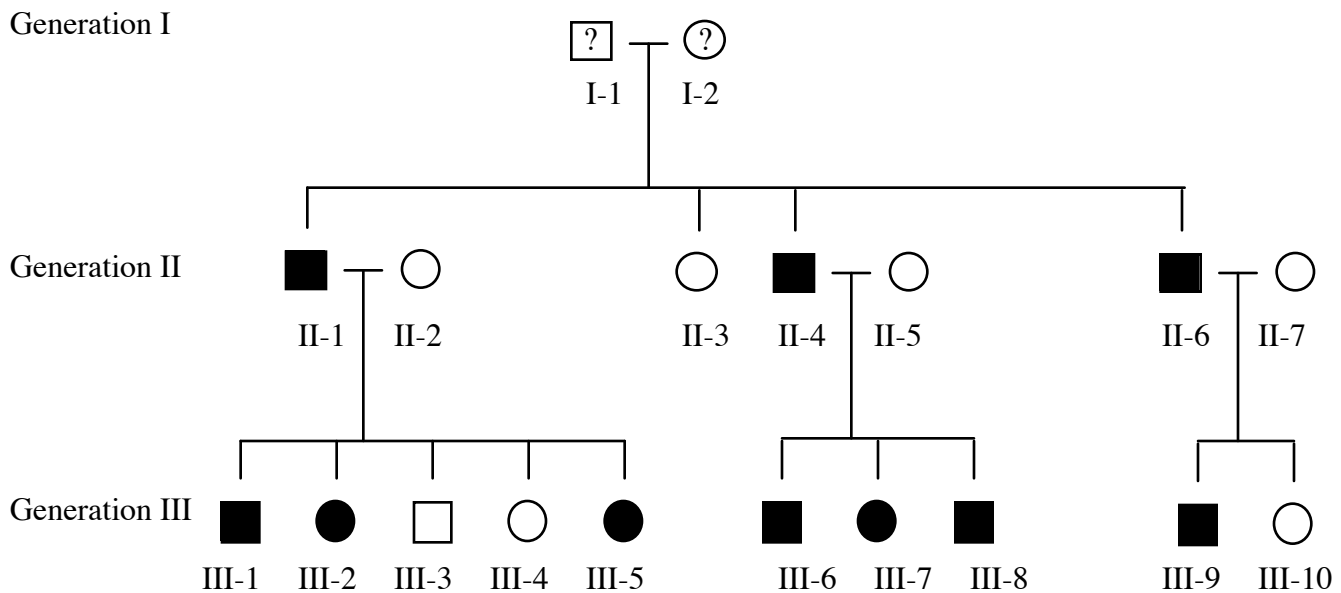
Step 1 - Constructing a Pedigree

The first step in genome mapping is to construct a pedigree of all individuals for as many generations as possible in a family.

This is a crucial step in genome mapping because

- It is the first step in determining the pattern of inheritance (dominant vs recessive)
- It will allow comparison of the DNA between affected and unaffected individuals
- By constructing pedigrees of different families, comparisons between families are possible.

A pedigree is constructed by questioning, examining and running relevant tests on all family members for as many generations as possible. Below is shown the pedigree for one family suffering from atrial septal defects.



In this pedigree, individuals are numbered. The first number in Roman numerals refers to the generation. The second number refers to a specific individual within a generation. This numbering system allows scientist to refer to specific people in the text of their reports. Note that in older generations, if these individuals have already died and the cause of death was uncertain, we may be unable to determine whether they were affected or not. Similarly, it will be impossible to run tests on these family members.



Step 2 - Patterns of Inheritance of Human Disorders

The second step in genome mapping is to determine the mode of inheritance of the disease.

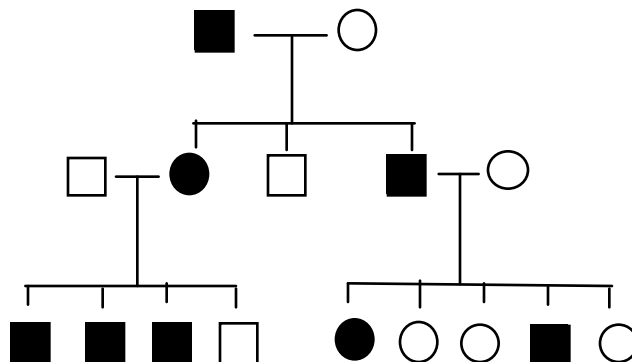
This is an important step in genome mapping because

- It lets us know whether there are one or two copies of the mutation per individual.
- It tells us whether to look on the sex chromosome or elsewhere.

Diseases can show different patterns of inheritance; sex-linked, autosomal dominant, and autosomal recessive (see box to review these patterns). It is important to know what pattern of inheritance you are looking at so that when you look at markers found on the two different chromosomes, you know whether you will be looking for copies of the mutation on one versus two chromosomes. It will also tell you whether to look for the mutation on the X chromosome versus one of the other chromosomes.

Examine the three pedigrees shown below. Each is from a representative family with a different type of inherited heart defect.

Family A



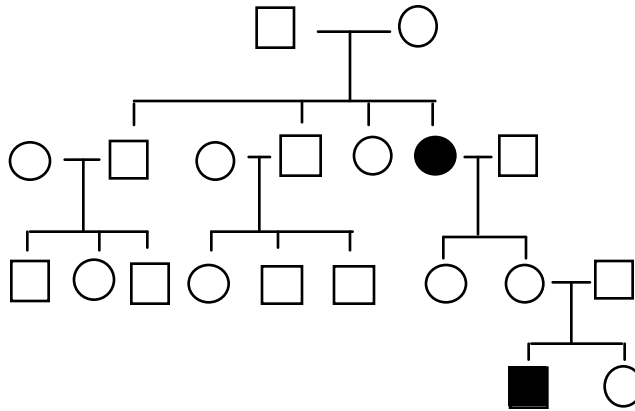
Patterns of Inheritance

Autosomal dominant - The mutation is not located on the X or Y chromosome and is dominant over other alleles at this locus. With diseases that are autosomal dominant, 50% of an affected parent's offspring are affected, and males and females are affected with equal frequency. Since the mutation is dominant, there are no carriers of diseases inherited in this fashion, and the disease will not skip generations.

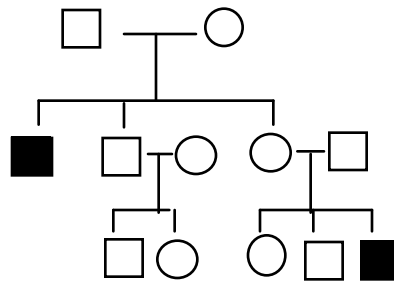
Autosomal recessive - The mutation is not located on the X or Y chromosome and is recessive to other alleles at this locus. For diseases that are autosomal recessive, 25% of the offspring of 2 carriers are affected, while none of the offspring are affected if both parents are not carriers. Males and females are affected with equal frequency. Diseases with this pattern of inheritance will skip one or several generations.

Sex-linked - The mutation is located on the X chromosome. Since males have only one copy of the X chromosome, sex-linked diseases are carried by females and almost always affect males. Diseases with this pattern of inheritance will skip one or several generations.

Family B



Family C



For each of the three families -

1. Identify the percent of family members affected (in this case, count the grandparents).
2. Determine whether males or females or both are affected.
3. Decide whether each of the pedigrees shows an autosomal dominant, an autosomal recessive or a sex-linked pattern of inheritance.
4. Record your results in the table on the next page.

	Percent of Individuals Affected	Sex of individuals affected	Does the disease skip generations?	Pattern of Inheritance
Family A				
Family B				
Family C				
Atrial Septal Defect				

Now return to your family with ASD. Which pattern of inheritance does ASD exhibit? Record your results in the table.

Next you will be looking for the location of the mutation leading to ASD in the human genome. Considering the type of inheritance ASD exhibits, will you look on the X chromosome or one of the other chromosomes?

Again considering the type of inheritance ASD exhibits, will you be looking for one or two copies of the mutation in affected individuals?

Part III – Sequence Tagged Sites as Markers for the Human Genome

The goal in genome mapping is to find the gene in which a mutation is located. A first step towards finding the gene is to determine “markers” that co-segregate with the mutation during meiosis. This co-segregation of markers is a phenomenon known as linkage disequilibrium. The closer a mutation is to a particular marker, the more likely it is to segregate with the mutation.

An early problem in genome mapping was finding a set of markers that covers the entire human genome. To be useful, markers must be evenly spaced throughout the genome and they must be highly variable between individuals.

Sequencing of the human genome as part of the Human Genome Project has led to the discovery of sequence tagged sites (STSs) that are found throughout the human genome and can be reliably used as markers. STSs are repeats of 2, 3, or 4 bases (see box). The number of times each of these sequences is repeated in tandem is highly variable between people. For example, the number of copies of the tetra-nucleotide repeat GATA found on chromosome 5 varies between individuals. There could be 1, 2, 3, or 4 copies of the repeat:

GATA
GATAGATA
GATAGATAGATA
Or GATAGATAGATAGATA

Since each person has 2 copies of chromosome 5, there are 10 possible combinations found in the total human population (1,1 1,2 1,3 1,4 2,2 2,3 2,4 3,3 3,4 and 4,4)..

We can now use a selection of evenly spaced STSs as markers of different parts of the genome. It is these sequences which are currently measured in genome-wide linkage mapping. One list of such STSs can be found at <http://www.chlc.org/ChlcMarkerMaps.html> Several different sets of markers exist, and they are continually being updated.

Sequence Tagged Sites (STSs)

Throughout the human genome are many repeats of 2, 3, or 4 bases. These sequences have been known as "junk DNA" for a long time. The purpose of these sequences remains unclear however it is now doubtful that they are "junk DNA."

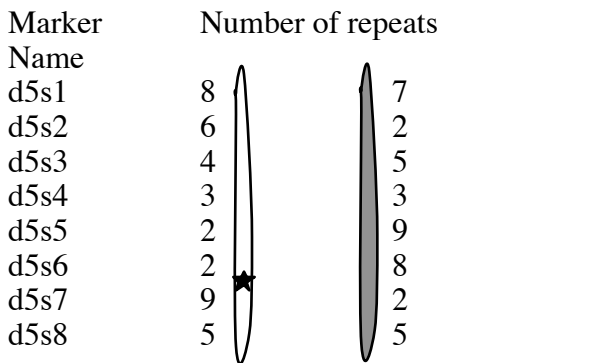
A few of the repeats of 3 bases occur within DNA that codes a protein. The variability in the number of repeats within a gene leads to a string of a particular amino acid, often glutamate. Variable numbers of these repeats can lead to diseases such as Huntington's Disease and Fragile-X syndrome. Also, changes in the number of certain repeats seen in non-coding regions correlate with development of cancer.

In addition to the disease consequences of the repeat sequences, we have put these sequences to use in many ways since they are so highly variable between people. DNA fingerprinting is based on differences in the number of repeat units. These sequences are also used in genome-wide linkage mapping.

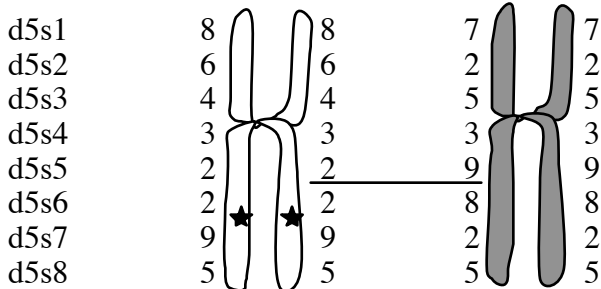
Part IV – Segregation of STS Markers During Meiosis

Genome mapping is based on the principle that the markers most closely linked to the mutation will be less likely to be separated from it during meiosis and crossing over. This is known as linkage disequilibrium. Markers that co-segregate with the mutation are physically near the mutation.

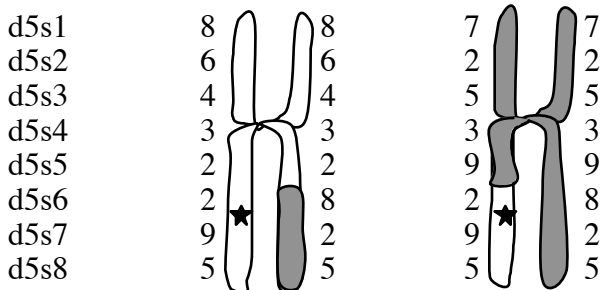
The diagram below represents what happens with one pair of chromosomes during meiosis. The star shows the site of the mutation.



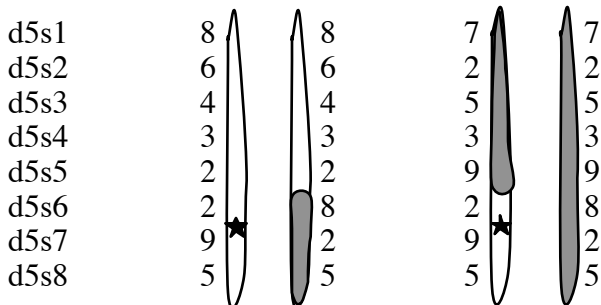
Pictured at the left are the two copies of chromosome 5 found in an individual. They have different numbers of repeats on each chromosome because they have one copy of chr 5 from each of their parents – the two copies of chr 5 are not identical.



Here the chromosomes have replicated. During meiosis I, the homologous chromosomes align and crossing over occurs. In this case, crossing over will occur between the spots marked with the horizontal line.

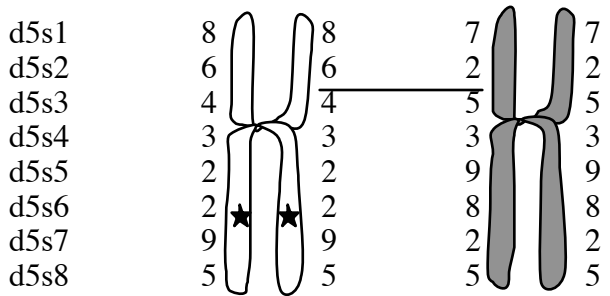


After crossing over, the homologous chromosomes will be separated as meiosis I is completed. Note how the markers have been exchanged between the two inside arms.

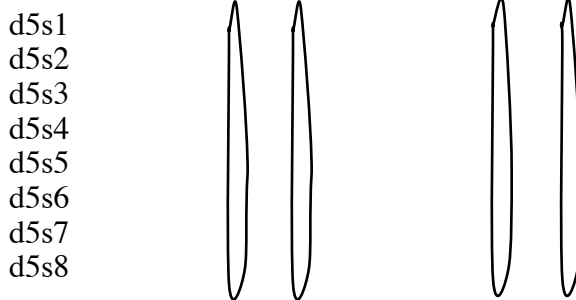
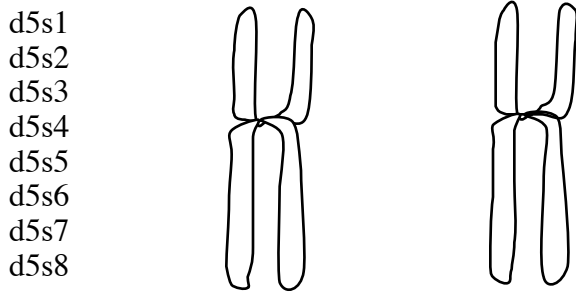


Following meiosis 2 one copy of each chromosome ends up in an egg or sperm. The 4 different possibilities for chromosome 5 in this example are shown at the left.

Now you will do the same thing – except that crossing over will occur in a different spot. Complete the example below by filling in the number of repeats and shading the chromosomes for a cross over event located at the horizontal line in the example below.



Chromosomes have replicated and homologous chromosomes are aligned.



Questions

What marker(s) and repeat numbers are most closely linked to the mutation?

Next, compare the 4 products. Which markers and repeat numbers are inherited with the mutation on both chromosomes that contain the mutation in this example?

in the example on the previous page?

2. Draw the results of your electrophoresis on your gel. Analyze the gel to determine the number of repeats found in each of the two chromosomes in each of your 7 samples. Record your results in the table below.

Lane	Person being tested	Marker being tested	Number of Repeats found
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			

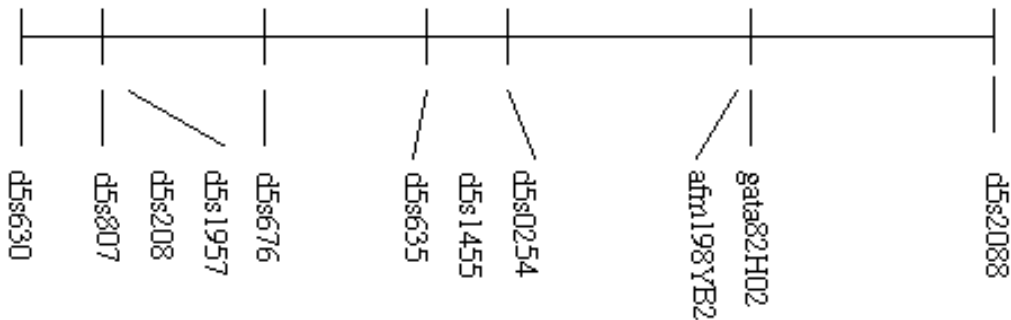
3. Record your data with the other class data (see attached data sheet).

Part VI – Analyze the Data and Determine Linkage

Below is a map of the region of chromosome 5 you are examining. Note that the STS markers you have used are noted on the map. These are also the same markers found on the data table.

Your goal is to find a marker or series of markers that are most closely associated with this mutation. To do this

- Look at the data sheet. Note that each individual has two chromosomes.
- The individuals with ASD are noted in bold. Looking at their chromosomes, which of their two chromosomes is most likely associated with ASD. Here, you are looking for a chromosome common to all people with ASD (remembering that the chromosomes will be highly similar, but not identical due to cross over events that may have occur).
- Next, comparing these common chromosomes, look for a series of STS markers that is the same (or close to the same – remember that crossover can separate markers in some individuals) in all individuals with the disease and is not found in individuals who do not have the disease.
- Narrow down this region to as small an area as possible by looking for the smallest possible number of markers that are co-inherited with the disease.



What series of markers and repeat numbers is inherited together with the mutation?

Mark the region in which the mutation occurs on the map above.

Part VII – Using the information

Why genome mapping?

- Genetic testing
- Gene therapy
- Choosing treatments
- Understand molecular mechanisms underlying the disease
- Directions for further research

Sometimes locating the mutation allows scientists to understand the mechanism of disease and to attempt gene therapies. This is the case for cystic fibrosis. The cystic fibrosis causing mutation is located in a gene for a chloride ion channel found in epithelial cells. Knowing that the mutation is in this gene allowed scientists to:

1. Explain the mechanism by which the mutation leads to disease - the altered shape of the chloride channel leads to increased retention of chloride within cells. This in turn leads to less liquid being secreted from the cells, resulting in thick mucus that clogs the lungs and digestive tract.
2. Test for the presence of the mutation, allowing accurate and early diagnosis of cystic fibrosis and genetic counseling.
3. Begin design of a gene therapy for cystic fibrosis. A copy of the correct chloride channel gene has been cloned into a defective virus, which will temporarily infect human cells. Although gene therapy has not been effective to date, continued progress may lead to good treatments.

Sometimes, however, finding the mutation does not lead to an immediate understanding of the disease process or a cure. An example of this is Huntington's disease. The mutation has been found, however the function of this protein in normal people is unclear. Further, no one understands how the mutation leads to the disease. In this case knowing that the mutation is in this disease allows scientists to:

1. Test for the presence of the mutation, allowing accurate and early diagnosis of Huntington's Disease and genetic counseling.
2. Give researchers an idea where to look further when trying to understand Huntington's Disease.

In the case of ASD, the gene causing ASD in this particular family has not yet been located. However, the gene causing ASD in other families has been found.