



GeneChip[®] Microarrays

Activity #4 – Reading and Analyzing GeneChip microarray Results from Real Life Scenarios

Introduction:

In this activity, you will get the opportunity to use what you learned about GeneChip microarrays from Activity #2. Students will be organized into groups of five or six and each group will be assigned a different scenario that uses one of three different types of DNA chips, Gene Expression, Resequencing, or Genotyping, in the research. Each scenario will have its' own set of results that the group must analyze, interpret, and then present to the class what they determined in a short, five minute presentation. The presentation should include relevant experimental background, the experiment itself, the results, the analysis, and possible future research.

The purpose of this activity is to challenge you to analyze and interpret data in a group setting and work out a real life research problem. All scenarios and results are simplified, but are related to real research and medical studies occurring in recent studies.

Goals of Activity:

The goals of this activity are:

- #1 – To apply knowledge of the function of DNA chips to the analyze results of DNA chips in simulated scenarios
- #2 – To work as a group to analyze and interpret data
- #3 – To work as a group to present the results and their implications to the class
- #4 – To communicate within a small group and to an entire class

Procedure:

- (1) Teacher will organize students into up to 6 groups
- (2) Teacher will assign each group a different scenario with a background, results, and data to analyze
- (3) Students will spend 15 to 20 minutes reading the background to the scenario, looking at the results, interpreting the results, and discussing the implications of the results to the research and to further studies
- (4) Student groups will spend 15 to 20 minutes working on a class presentation focusing on experimental results, data analysis and future research implications. This presentation should be well organized, clear in its interpretation, and should include overheads / visuals to help the class understand the results
- (4) Each group will get 4 to 5 minutes for their presentation (not every person is required to talk for each group, but more than one student needs to speak)
- (5) Students will be graded on their work while in the groups and on the presentation using a rubric as a scoring guide.

The Scenarios:

- A – Food Contamination Analysis using the Fish DNA GeneChip microarray
- B – Drug Metabolism Analysis using the CYP GeneChip microarray
- C – HIV Genotyping using a HIV GeneChip microarray
- D – Gene Expression in *Vitis vinifera* (grape) using a Grape Genome Array
- E – Breast Cancer progression research using Human Genome Array
- F – E. coli form strain identification using Genome (Resequencing) Array

Scenario A – Fish CSI?

Background:

A recent development in the food industry is the substitution of very expensive meats with a “fake” or less expensive version while still selling it at a relatively high price. Imagine you are a group of scientists hired by the FDA to go to some of the most expensive sushi restaurants in the area and randomly test their more expensive sushi for “imposter” fish. You decide to use the Fish DNA GeneChip microarray* to do the testing. This microarray contains probes representing specific gene segments of 15 different fish species. The features in the array are organized in the following way:

| | | | |
|----|-----|-----|-----|
| AC | AB | JSM | AM |
| AS | BT | EE | EH |
| GC | JE | ME | RT |
| ST | SkT | SpT | BFT |

The abbreviations are as follows:

| | | |
|---------------------------------|----------------------|------------------------|
| AC = Arctic char | AB = Atlantic bonito | AM = Atlantic mackerel |
| AS = Atlantic salmon | BT = Brook trout | EE = European eel |
| EH = European hake | GC = Greenland cod | JE = Japanese eel |
| ME = Mozambican eel | RT = Rainbow trout | ST = Sea trout |
| SkT = Skipjack tuna | SpT = Spotted tunny | BFT = Blue Finned Tuna |
| JSM = Japanese Spanish mackerel | | |

The array is simple to read. DNA is isolated from the sample (food) and then reverse transcribed into tagged RNA. The RNA is added to the array, allowed to hybridize, fluorescently tagged, and analyzed. If the DNA came from one of the species above, the RNA would hybridize to the probes in that specific feature and would fluoresce. In short, if a specific feature fluoresces, then DNA from that fish species is present. (Note, on an actual array, multiple features would be used for each species. For simplicity, in this example each species is represented by a single feature).

*Note: This scenario uses a “Fish DNA GeneChip microarray” which is based on an actual chip known as the FoodExpert-ID GeneChip microarray from the company [bioMérieux](#).

Your group has recently collected samples of sushi from a first restaurant. You received some inside information that the restaurant may be trying to increase profits by substituting cheaper fish for some of their more popular sushi. You have collected four suspicious samples that the restaurant claimed to be the following: sake (Atlantic salmon), unagi (Japanese eel), maguro (blue finned tuna), and saba (Atlantic mackerel). For each sample you collected enough for 20 tests.

Results:

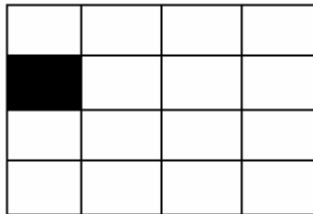
The results are below. For simplicity's sake, use the following key to help you interpret the amount of DNA in each sample take:

Black feature = high DNA levels
Light Grey = low DNA levels

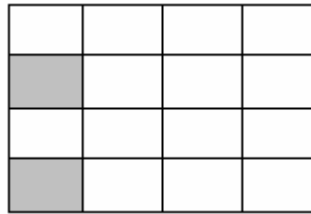
Dark Grey feature = medium DNA amounts
White feature = no DNA present

Use the key on the page before to interpret what each feature represents.

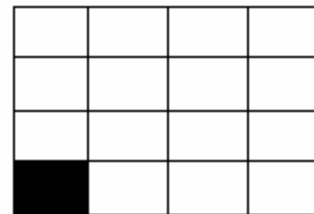
Sample #1 – samples sake (Atlantic salmon)



4 test results

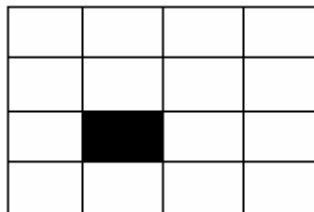


13 test results



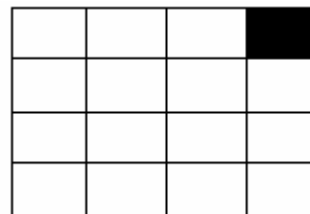
3 test results

Sample #2 – samples of unagi (Japanese eel)



all 20 test results

Sample #3 – saba (Atlantic mackerel)



all 20 test results

Sample #4 – maguro (Blue finned tuna)

| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |

4 test results

| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |

4 test results

| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |

12 test results

Directions:

Analyze each sample carefully and work together as a group to come up with a hypothesis on what is happening in each situation. For each situation, be sure to think about why the restaurant may or may not have switched the fish type. What do they have to gain or lose? When finished, be prepared to report back to the class what you have found. Prepare visuals to help with your explanation and make sure you are very clear.

Scenario B – Drug Metabolism

Background:

You are a medical research group studying drug metabolism in humans. Understanding the variability among humans in terms of how they metabolize drugs has a wide range of applications for the medical community, including helping with questions like: How much of a drug should be prescribed? When should a person take the drug? Should the person take the drug at all? What alternatives should be looked at, etc.? The gene CYP450 is known to be involved in the metabolism of 90% of commercially prescribed drugs. You decide to do a research study into the CYP450 gene activity in group of 250 patients using the CYP GeneChip microarray.

For your study, you are aware of 6 important alleles (versions) of the CYP450 gene. For simplicity's sake we shall call them allele #1 – 6. Each version of the gene has a different level of activity due to point mutations, base insertions or deletions. Past research has shown that the 6 alleles can be placed into three categories based on their activity (i.e., how much they influence drug metabolism) – Normal Activity, Reduced Activity, and Inactive. Here is the category that each allele is known to lead to:

| | |
|-----------------------------|-----------------------------|
| Allele 1 – Reduced Activity | Allele 2 – Normal Activity |
| Allele 3 – Normal Activity | Allele 4 – Reduced Activity |
| Allele 5 – Inactive | Allele 6 – Inactive |

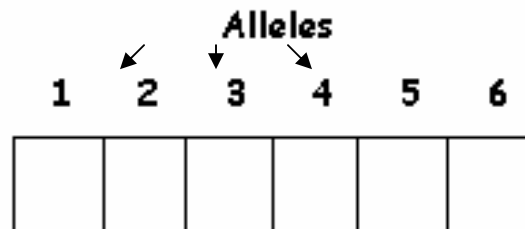
Remember, in most cases, everyone has two alleles for each gene and could be homozygous (two of the same alleles) or heterozygous (two different alleles). Thus, a person could have the genotypes such as 1/3, 2/5, 4/4, 3/6, or 5/5. In this case there are many possible genotypes – 21 in total. Your group decides to look into the different genotypes that exist predominantly in the population to help with the prescription of drugs. In other words, if people who do not metabolize drugs well have the genotype 5/6, then the doctor could adjust the prescription accordingly.

You want to find out what types of “metabolizers” have which type of genotypes. Recent research has shown there to be 4 categories of metabolizers (or four phenotypes): Poor Metabolizers (PM), Intermediate Metabolizers (IM), Full Metabolizers (FM), and Ultra Metabolizers (UM). UMs break down the drug so fast that the drug does not have sufficient amount of time to work for the person. PMs are unable to use the drug efficiently. IMs show a low level of metabolism, but more than PMs. FMs show a normal, full range of metabolism. This is the most satisfactory phenotype where the person is able to get the most from the drug.

To do your study, your group uses the CYP450 GeneChip microarray * to study the possible genotypes leading to the four phenotypes.

* Note: this hypothetical chip is based on a real chip used to study the CYP450 gene. This array, sold by Roche is called the Amplichip™ (www.roche.com)

You study 250 patients. First, you administer Dextromethorphan (a cough suppressant) and take urine samples at specific times to determine how well they metabolize the drug. If they are a poor metabolizer you should find high levels of Dextromethorphan in their urine indicating that very little of the medicine has been metabolized. If they are ultra metabolizers there should be very little of the medication in their urine because most of it has been metabolized while in the body. This will allow you to determine which of the four phenotypic categories they belong to. Next, you take blood sample from which you isolate DNA. You prepare the sample and hybridize it to the chip in the normal manner. Reading of the chip signatures will be simplified to showing one feature for each allele (although typically multiple features are used for each allele). Here is a blank, sample data set:



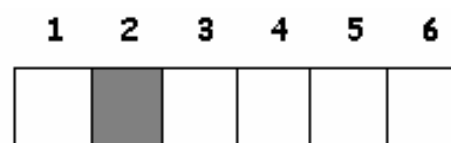
If a feature is shaded in, that means the particular allele is found in the DNA from that person. The one catch is that there is the possibility of a person containing more than two alleles in their genome. This mutation is known as gene duplication and is usually an error in replication during meiosis from the mother or father. Gene duplication usually leads to over expression of the gene due to the multiple copies. The codes for the different shades are as follows:

- Black - multiple copies of the allele (more than 2)
- Dark Grey – two copies of the allele
- Light Grey – one copy of the allele
- White – no copies of the allele

So, here are a few examples. The one on the left shows a person that has the genotype 4/5 and the one on the right shows a person that is 2/2 (homozygous for allele 2):



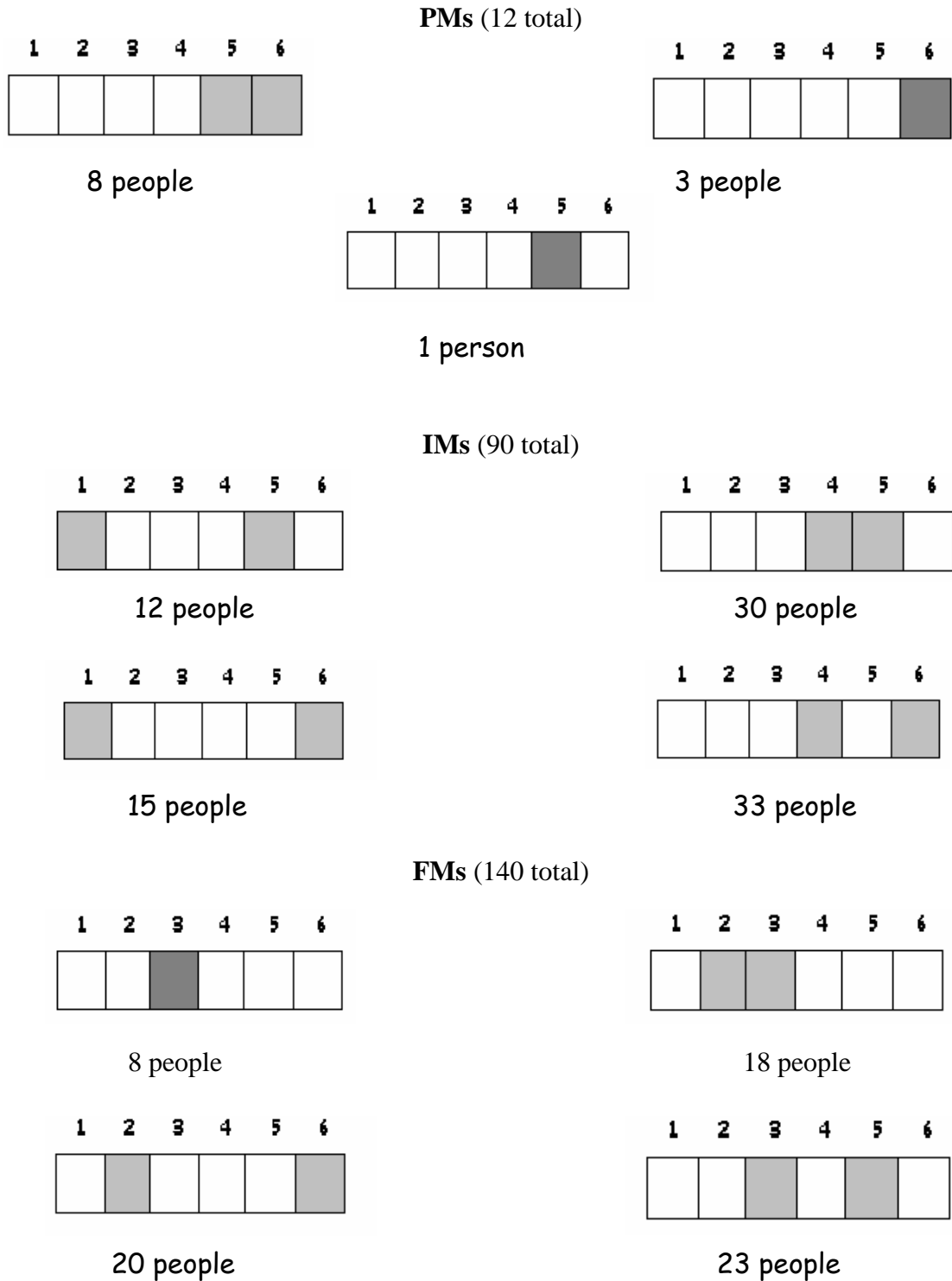
4/5 genotype

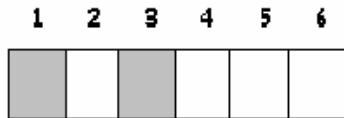


2/2 genotype

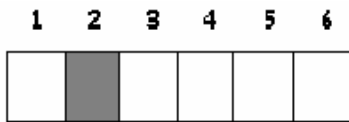
Results:

Here are the results. They are grouped by metabolizer level, showing which alleles were contained by the people in that group.

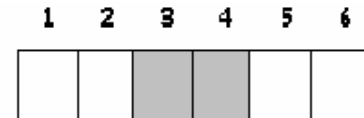




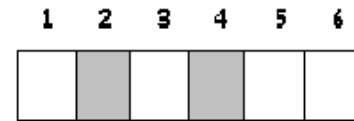
28 people



7 people

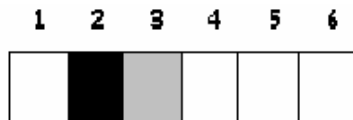


22 people



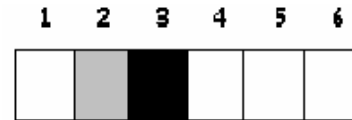
16 people

UMs (8 total)



4 people

3 people



1 person

Directions:

Your job is to analyze the results and attempt to determine all the possible genotypes and which phenotype they connect to. Be careful and make sure to look at all possibilities. Then, form a hypothesis about the causes of differing metabolism levels. That is, try to categorize what type of genotypes (allele combinations) lead to which of the four types of metabolizers. For example, having two reduced activity alleles, such as allele 1 and 4 or 1/4, may lead to a specific ability to metabolize, such as intermediate metabolism (this may or may not be true – you need to study the data!)

Scenario C – HIV Genotyping of Drug Resistant Strains

Background:

HIV is a virus that kills hundreds of people throughout the world each day. One of the main problems with the fight against HIV is that the virus is constantly mutating into newer, and sometimes stronger, forms that make detection and treatment very difficult. The high natural mutation rate of HIV also results in increased resistance to anti-HIV drugs used to help the patient fight the disease. Just as antibiotic use leads to resistant strains of bacteria, anti-HIV drugs such as protease and reverse transcriptase inhibitors can be overcome by new, mutated, strains of the virus that form during the infection. The high mutation rate of HIV makes this drug-resistant strain formation occur at a much higher rate than most viruses and bacteria.

Your team is a research group that has decided to study the HIV mutations occurring during the use of two different common drug cocktails prescribed to HIV infected patients. The two drug cocktail treatments you decide to research are commonly prescribed combinations of different reverse transcriptase inhibitors. The two treatments are:

Combination #1 – zidovudine (AZT) + efavirenz (sustiva) + didanosine (videx)

Combination #2 – lamivudine (epivir) + efavirenz (sustiva) + didanosine (videx)

Looking at the two treatments, you can see that the variable factor in the experiment is whether AZT or epivir is taken. In the experiment you will compare the mutations that occur during each form of treatment as well as look into which specific types of mutations lead to drug resistance. All other variables are controlled as much as possible. Variables that were controlled included the infection stage of the patient, the age of the patient, and whether they were a male or female. Also, their symptoms were all matched as much as possible. You have enrolled 20 patients in the study. None of the patients have taken any of these drugs – individually or as a cocktail – before the study began.

The study has two parts to it. In part one, you did a phenotypic study to determine whether drug resistance of the HIV changes as a result of taking the drug cocktail. This requires isolation of the HIV from the patients' blood, identifying the strain, and doing an in vitro tissue culture study on the resistance to anti-reverse transcriptase drugs.

In part two, you need to perform a genotyping study that looks at the change in the HIV genome after taking the drug cocktail (i.e., the mutation(s)). You decide to do this genotyping study by using an HIV Genome GeneChip microarray*. The full microarray looks at 347 exons of the HIV genome from two genes that code for the protease enzyme and the reverse transcriptase enzymes (both vital for replication of HIV).

*Note: This array is based on the real GeneChip microarray known as the HIV PRT Genome Microarray made by Affymetrix.

This microarray basically resequences each exon, allowing you to look for most mutations such as point mutations (single mutations of one base pair in the DNA). To narrow it down, you decide to look at segments of the two most commonly mutated exons within the reverse transcriptase gene – exon 74 and 184. Notice that you will not look at the entire exon, but rather a 20 base pair segment of each exon known as “hotspots” for mutations, where most mutations in the exon are commonly found.

This second part of the experiment requires you to have isolated HIV from the patient prior to the start of drug treatment. This provides you with an original “baseline” sequence of the HIV at these two codons. Then, the patients take the drug cocktail for approximately one year, at which time you isolate the HIV in the patients blood stream and repeat the resequencing, looking for mutations in the two exons.

Results:

In the study, patients #1-10 received Combination #1: AZT + sustiva + videx and patients #11-20 received Combination #2: epivir + sustiva + videx

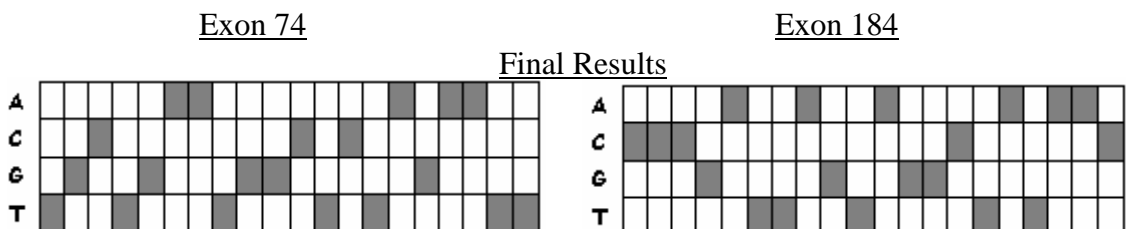
Baseline Study

*Phenotype study: Each person was tested to determine their HIV strain and that strain’s resistance to reverse transcriptase drugs. Their results are below. The scale of the resistance is from 1 to 20 where 1 is “low resistance”, 10 is “mid level resistance”, and 20 is “high resistance”.

| <u>Resistance level</u> | <u>Patients</u> |
|-------------------------|----------------------|
| 1 | 1 -5, 8, 12-16, & 19 |
| 2 | 6-7, 11, 17-18, & 20 |
| 3 | 9 & 10 |

*Note: no patients showed a resistance level beyond 3.

*Genotype study: The baseline HIV sequence study found that none of the patients showed mutations in either exon 74 or 184 at the start. They all had the same sequence at these two exons. (This does not mean they have the exact same strains, but that all the strains they were infected with have no differences at these two exons). Remember, the researchers did not look at the entire exon, but rather the area of the exon known to have the most mutations (called “hotspots”). Here are the sequences of the two (non-mutated) segments you are studying:



*Phenotype study: At the end of the study, you isolated HIV from each person's blood stream and once again tested each for its resistance to reverse transcriptase. The final results for each patient are below.

| <u>Patient</u> | <u>Final HIV Resistance Level</u> | <u>Patient</u> | <u>Final HIV Resistance Level</u> |
|----------------|-----------------------------------|----------------|-----------------------------------|
| #1 | 2 | #11 | 4 |
| #2 | 1 | #12 | 1 |
| #3 | 2 | #13 | 10 |
| #4 | 1 | #14 | 10 |
| #5 | 1 | #15 | 2 |
| #6 | 2 | #16 | 1 |
| #7 | 4 | #17 | 20 |
| #8 | 1 | #18 | 10 |
| #9 | 15 | #19 | 1 |
| #10 | 3 | #20 | 2 |

*Genotype study: After a year of taking the drug cocktail medication, you isolated HIV from the patients and resequenced exons 74 and 184. Here are the results for each patient:

Exon 74

Exon 184

Patient #1



No Mutation

Patient #2

No Mutation

No Mutation

Patient #3



No Mutation

Patient #4

No Mutation
Exon 74

No Mutation
Exon 184

Patient #5

No Mutation

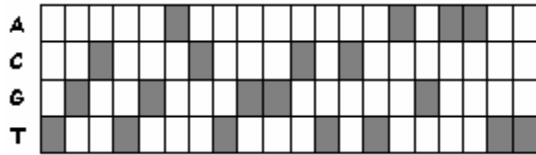
No Mutation

Patient #6

No Mutation

No Mutation

Patient #7



No Mutation

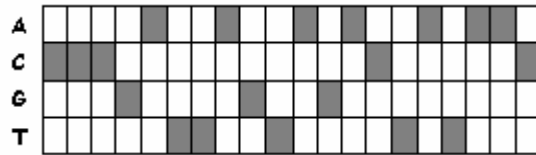
Patient #8

No Mutation

No Mutation

Patient #9

No Mutation

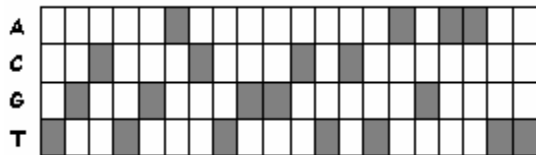


Patient #10

No Mutation

No Mutation

Patient #11



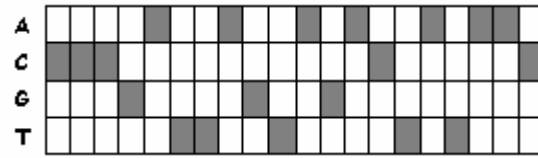
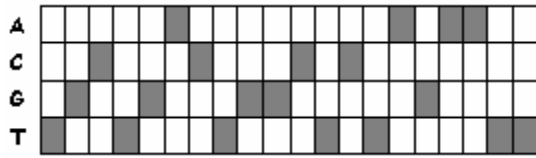
No Mutation

Patient #12

No Mutation
Exon 74

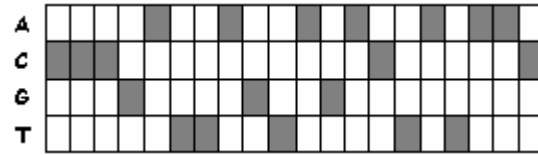
No Mutation
Exon 184

Patient #13



Patient #14

No Mutation



Patient #15



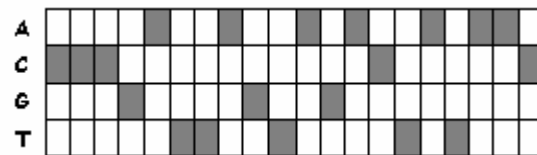
No Mutation

Patient #16

No Mutation

No Mutation

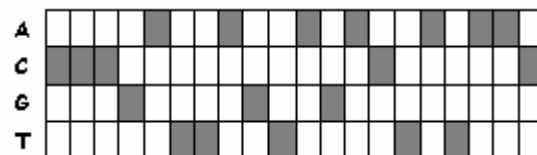
Patient #17



Patient #18

No Mutation

Exon 74



Exon 184

Patient #19

No Mutation

No Mutation

Patient #20

No Mutation

No Mutation

*Notice – for simplicities sake, only those exons that show mutation(s) are shown

Directions:

Now that you have the data, you need to make sense of it. What is happening here? Which drug combinations lead to the proliferation of which resistant strains? Which drugs lead to more mutations? What type of mutations occurred? Which drugs lead to which mutations and, therefore lead to greater resistance to the drugs? How much of a change in resistance occurred?

There is a lot to take a look at! It is suggested that you build some sort of table to help organize and analyze the data. Maybe include columns for changes in resistance, type of mutation, etc. Good luck!

Scenario D – A Grape Gene Expression Study

Background:

The food industry is one of the largest industries in the world. It is an industry that is constantly moving forward and looking for ways to improve. Recent technology has led to large advances in the areas of agriculture. Some of the latest technology looks at the genomes of plants and animals with the hopes of understanding gene function. It is hoped that by understanding the genetics of food plants and animals that farmers will be able to increase their productivity (i.e. grow more food in less time with less money).

One of the newest users of this technology are the grape growers. Grapes are a large market, primarily in the area of fruit grapes and grapes used in wine making. Learning more about grape genetics could have a big impact on both of these markets. For example, if you could determine the best conditions for growing the different types of grapes you would know exactly where to plant each grape plant rather than making your best guess as you have done in the past.

You are part of a research group assigned to study gene expression in wine grapes at different conditions. You are looking for the perfect way to manipulate the environmental conditions of the grape plant in order to grow the grape with the precise acid and sugar content needed for producing wine. Your group decides to analyze as gene activity throughout the grape's lifecycle. You turn to the use of the newest GeneChip microarray – the *Vitis vinifera* Genome Array. The *Vitis vinifera* Genome Array covers the entire genome of the grape and can be used to tell which gene expression pattern produces the “perfect” grape.

Vitis vinifera needs hot dry summers and cool winters. Your group decides to test the best conditions during the winter since less information is known about these conditions. The grapes were all grown in the same conditions during the summer, but when winter came, the conditions of the grapes were varied. Your team tested each group of grapes and you extracted total mRNA. After the end of the growing year, the grapes were harvested and final tests were made to determine if they had the precise chemistry needed to produce the best wine.

Prior to starting your experiment, you did some research and decided to focus on 6 genes known to be involved in the biochemical pathways responsible for the acid and sugar content of the grape. For simplicity's sake, we will just number the genes from 1 to 6.

The conditions studied were: soil pH, temperature, and water amount. You can manipulate the soil pH by adding lime and a weak acid, vary the amount of water each plant receives by adjusting the irrigation, and control the temperature by growing the plants in a separate enclosed green house. This is just a preliminary study, so your group tests each condition in to high, medium and low. This should help you narrow it down to the perfect conditions for further studies. Your objective is to help growers determine how to best manipulate the pH and water amount during the changes in temperature.

Results:

Grape growers rate the grapes on a 1 to 5 scale depending on how well the sugar and acid content of the grapes fit with what is needed. Here is the scale”

| <u>Rating</u> | <u>Description</u> |
|---------------|---|
| 5 | the “perfect” grape / has all qualities needed for wine production |
| 4 | - “great” grape with perfect acid content but slightly off sugar content (or vice versa) / may need to add in some 5 grapes |
| 3 | “good” grape with slightly off sugar and acid content / will need add in some 5 grapes before production |
| 2 | “poor” grapes with sugar and / or acid contents off what is needed / can only be used to produce low quality wine |
| 1 | - the “terrible” grape / unusable on any levels |

You grow the grapes in three different temperatures (8 °C, 12 °C, and 16 °C). For each temperature group, you try combinations of low, medium, and high soil pH, and water amount. The low and high amounts are not extremes, just above or below the medium.

Here are the levels for each test:

| | | |
|----------------------------|---------------|----------------------|
| Low Temperature = 8 °C | Low pH = 5 | Low Water = 4 dm |
| Medium Temperature = 12 °C | Medium pH = 7 | Medium Water = 14 dm |
| High Temperature = 16 °C | High pH = 9 | High Water = 24 dm |

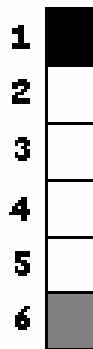
Now, since there are three different levels for each test and if you are trying each combination, then there are 27 different possibilities. Each different test will be given a letter from A to Z and @ to help with organization of the data. Once again, you will group them by the temperature, so there are 9 combinations for each test:

| <u>Low Temperature (8 °C)</u> | <u>Medium Temperature (12 °C)</u> | <u>High Temperature (16 °C)</u> |
|------------------------------------|------------------------------------|------------------------------------|
| A (low pH, low H ₂ O) | J (low pH, low H ₂ O) | S (low pH, low H ₂ O) |
| B (low pH, med H ₂ O) | K (low pH, med H ₂ O) | T (low pH, med H ₂ O) |
| C (low pH, high H ₂ O) | L (low pH, high H ₂ O) | U (low pH, high H ₂ O) |
| D (med pH, low H ₂ O) | M (med pH, low H ₂ O) | V (med pH, low H ₂ O) |
| E (med pH, med H ₂ O) | N (med pH, med H ₂ O) | W (med pH, med H ₂ O) |
| F (med pH, high H ₂ O) | O (med pH, high H ₂ O) | X (med pH, high H ₂ O) |
| G (high pH, low H ₂ O) | P (high pH, low H ₂ O) | Y (high pH, low H ₂ O) |
| H (high pH, med H ₂ O) | Q (high pH, med H ₂ O) | Z (high pH, med H ₂ O) |
| I (high pH, high H ₂ O) | R (high pH, high H ₂ O) | @ (high pH, high H ₂ O) |

At the end of the study you were able to obtain the quality rating of each grape at each condition. Here are the results:

| <u>Grape</u> | <u>Rating</u> | <u>Grape</u> | <u>Rating</u> |
|--------------|---------------|--------------|---------------|
| A | 4 | O | 4 |
| B | 3 | P | 3 |
| C | 2 | Q | 4 |
| D | 5 | R | 3 |
| E | 4 | S | 2 |
| F | 3 | T | 3 |
| G | 4 | U | 2 |
| H | 3 | V | 3 |
| I | 2 | W | 4 |
| J | 3 | X | 3 |
| K | 4 | Y | 4 |
| L | 3 | Z | 5 |
| M | 4 | @ | 4 |
| N | 5 | | |

Since none of these conditions gave the poorest quality grape, you were a little worried that you didn't a full comparison. At the last minute, your group was given some very poor quality (1) grapes that you decide to analyze for further information. Here is what the Gene Expression array output looked like:



Directions:

Make some sense of this data! How did the changes in conditions change the gene expression? Which gene expression profile leads to the best quality (5) grapes? Which genes were on when? And how high were the expression levels of each gene? What gene expression profile leads to the other qualities? How can a grape grower use this information to grow the best quality grapes? These are just some of the questions to look into. There are many other aspects to look at. Good luck!

(Hint: You may want to develop some table that allows you to easily compare the expression patterns with each condition and each quality rating)

Scenario E – Monitoring the Gene Expression of Cancer

Background:

One of the first uses of GeneChip microarrays was to monitor the genes active in cancerous tumors. These arrays were perfect for the job because they allowed the researcher to monitor the activity of multiple genes at a time. Since most cancers are due to mutations in multiple genes, this technology is a perfect fit.

Your group is a research group assigned to look at the activity of specific genes during the progression of cancer cells from normal cell to a tumor cell (a cell dividing out of control). Understanding which genes are responsible could lead to multiple therapies and treatments in the future. The cancer you are studying is a type of lung cancer known as small cell lung cancer (SCLC). Currently, this cancer has only an 8-12% five year survival rate once diagnosed; therefore, knowledge of this cancer is crucial. The cancer has four basic stages to it. The following is the basic description for each stage:

Stage 1 – Localized (cancer is small or growing into the airways)

Stage 2 – Early Spreading (cancer has grown larger and has spread to lymph nodes closest to lungs)

Stage 3 – Multiple Spreading (cancer has now spread to nodes furthest from the lung on other side of chest)

Stage 4 – Other Body Parts (cancer has spread to another lobe of the lung and to another part of the body such as liver or bones)

Your group decides to use a Human Genome microarray that looks at the expression of approximately 21,000 human genes. You isolate mRNA from each patient's normal, non-cancerous cells. Then, you isolate mRNA from tumor cells from patients as they progress from stage 1 to 4 and apply each to the array and scan for results.

The main idea is to look at which genes change their expression as the cells change from a normal, non-cancerous cell through the cancer stages. Those genes that become active as the cell changes might be a target for drugs to suppress the gene. On the other hand, those that turn off as the cell progresses might be ones that could be a target to turn on the expression.

To simplify the readouts, you decide to focus on an area of 10 genes that your research shows to be implicated in small cell lung cancer development. You take 8 total mRNA samples from each patient throughout the study at seven stages – Early Stage 1 (ES1), Middle Stage 1 (MS1), Early Stage 2 (ES2), Middle Stage 2 (MS2), Early Stage 3 (ES3), Middle Stage 3 (MS3), and Stage 4 (S4) as well as a sample from a normal, non-

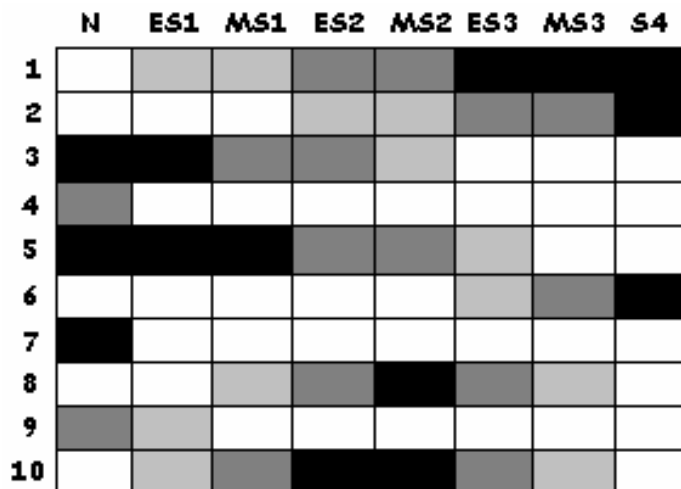
cancerous cell (N). Here list below is of the ten genes. Each one is given a specific number:

| <u>Gene #</u> | <u>Gene Symbol</u> | <u>Gene#</u> | <u>Gene Symbol</u> |
|---------------|--------------------|--------------|--------------------|
| 1 | SCLC1 | 6 | TP53 |
| 2 | SCLC2 | 7 | KRAS2 |
| 3 | EGFR | 8 | BRAF |
| 4 | NULL | 9 | RB1 |
| 5 | MGP | 10 | IGHG3 |

Results:

After collecting samples from each patient as they progress through lung cancer, you noticed that, except for few abnormalities, all of the patients had the same results. This means that their gene expression patterns are all very similar. The resulting microarray output is the common expression pattern that over 90% of the patients in the study showed. Even those that didn't have this exact pattern had mostly similar results with only a few genes acting differently.

The features numbered from 1 to 10 are the genes and those going from "N" to "S4" are the different stages. The features have three possible colors in them. The white feature means the gene is non-active, completely shut off. The light grey means the gene is expressed at a low level while the dark grey indicates the gene is expressed at a higher level. Finally, the black feature means the gene is extremely active at that stage.



Directions:

Your job is to analyze the results above and determine what is happening. Here are a few questions to think about when you analyze them. How did the genes change in activity as the cell progressed from normal cell to stage four cancer cells? How does the normal cell compare to the cancerous ones? If you were a drug company, which genes would you target to try to turn on or off? Why?

Scenario F – An *E. coli* Outbreak

Background:

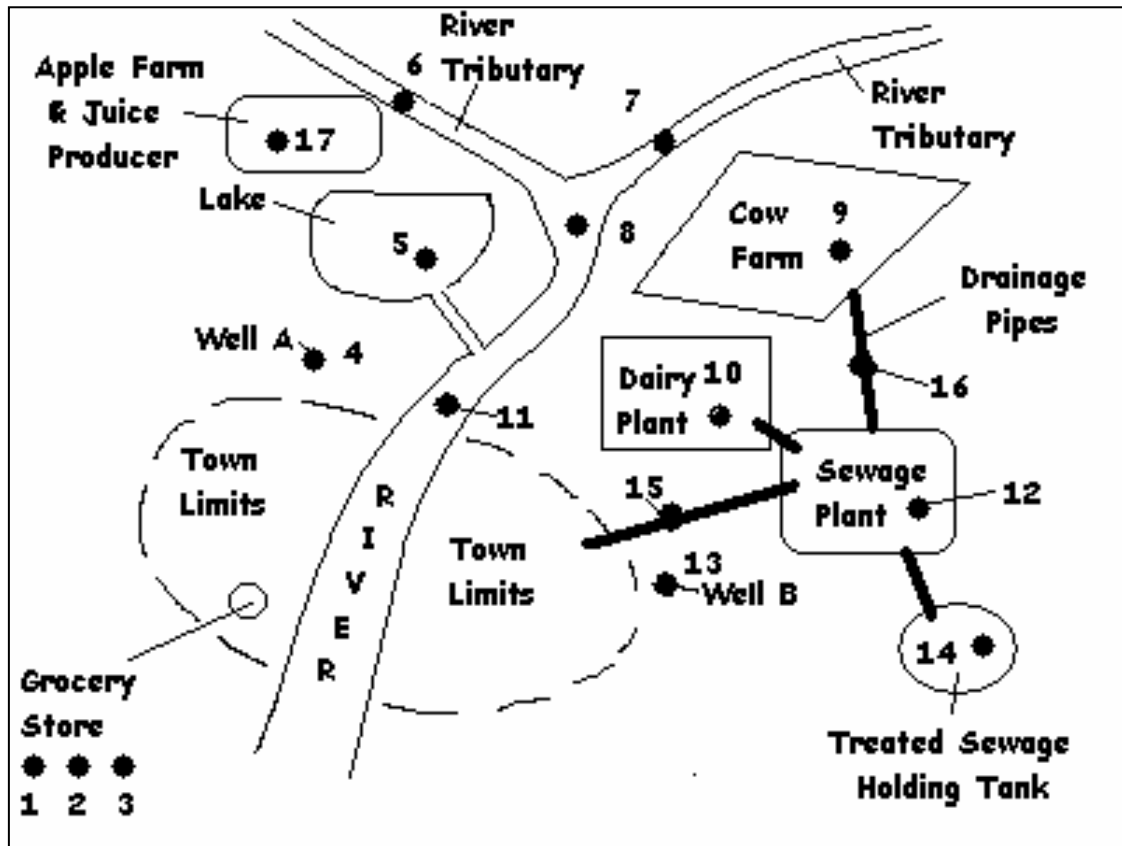
In a small rural town, a group of people – mostly young adults and children – have come down with a case of *E. coli* poisoning. Their hypothesis is that the culprit could be the dangerous and sometimes deadly *E. coli* strain, O157:H7 which has contaminated the food or water. No one is really sure, but preliminary tests do show it to be some type of *E. coli* that is making people sick. No one has died yet, but over twenty people are very ill. Your group of scientists has been assigned to go to the town and test all possible sources of the outbreak and answer these questions. What is the strain of *E. coli* that is causing the sickness? Where is it coming from? How is it getting there? And how are the people being infected?

You have decided to use GeneChip microarray technology and quickly get in communication with the manufacturing company. This company has the ability to custom order the array for you. You order a resequencing array that has the ability to resequence DNA in a sample that comes from specific *E. coli* strains. There are hundreds of strains of *E. coli* besides O157:H7, but you decide to resequence only the 6 most common strains that cause food and water borne illnesses, including the O157:H7 strain.

When you arrive in town, you are given a map of the area which identifies possible areas of contamination as well as possible sources of contamination. You decide to take a total of 17 samples from various areas around the town. Here are some facts about the town that could be helpful:

- there is a nearby river that flows right through the town
- there are two wells that the town draws most of its public water from
- there was a huge storm and flooding a few months ago & the areas north of the town (near the farms and lakes) were particularly affected
- there is a huge farm northeast of the town that has a large herd of cows
- there is a milk processing plant also northeast of the town
- there is an apple farm that produces its' own juice northwest of the town
- there is a lake that people commonly go to swim during the summer months just north of the town
- there is a sewage treatment plant that takes sewage from the town and cow farm and treats the raw sewage before moving it to an outdoor percolation pond just east of the town
- people commonly swim in both the lake and the river during the summer months (it now coming to the end of summer)
- the pipes in and outside of the town are getting very old
- the cow farm and milk plant provide most of the meat and milk consumed by the people of the town
- the town has one main grocery store where the people get most of their food (and all of the meat, milk, and apple juice come from the local producers)

Here is a map that has all the above areas and the spots of the seventeen samples you decide to test (marked with a spot and number).



● = sample point

Here what each sample was taken from:

- #1 – locally processed Apple Juice from store
- #2 – locally processed Milk from store
- #3 – locally processed meat from store
- #4 – water from well A
- #5 – water from lake (common swimming and boating area)
- #6 – water from upper river tributary
- #7 – water from upper river tributary
- #8 – water from river after tributaries meet
- #9 – cow feces from cow farm
- #10 – cow feces from dairy plant
- #11 – water from river after lake tributary (common swim area for town people)
- #12 – raw sewage from treatment plant
- #13 – water from well B
- #14 - water from sewage holding tank (post treatment)
- #15 – water from drainage pipe taking sewage water from town to treatment plant
- #16 – water from drainage pipe taking sewage from cow farm to treatment plant
- #17 – samples from soil taken at the farm

The GeneChip microarray you are using sequences any *E. coli* DNA found in the samples. It is a resequencing array that outputs a unique section of each strain. Remember, only if DNA from *E. coli* strains that you are sequencing for is found in the samples taken will you view results. The probes are specific to sequences in these strains. A blank result means there is no *E. coli* from the 6 strains you chose in the sample at all.

The *E. coli* genome is 1,725,750 bases in length. You have decided to look for a 20 base pair segment that is highly variable among *E. coli* strains and unique for each of the six strains at bases 1,725,600 to 1,725,620. Here are the 20 base pair sequences for each strain along with a bit of information on each.

JM 101

- harmless to humans
- commonly found contaminating dairy products
- unique sequence: TCCTGCAAACCTATTACAAC

RR1

- harmless to humans
- commonly found in water from lakes and rivers and other recreational areas
- unique sequence: GACTGCCACAGATTACAAC

O157

- dangerous and potentially deadly to humans
- causes intestinal strong problems and leads to very high fevers
- normally found contaminating raw meat, unprocessed dairy and apple juice
- unique sequence: TCCTGCAGGCAGATTACAAC

0111

- non-dangerous but does cause some sickness in humans
- causes intestinal problems and slight fevers
- commonly found in intestinal tract of cows
- unique sequence: TCCTGCAGGCAGATTAGTTC

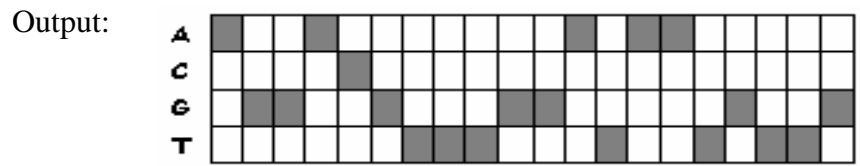
Q358

- harmless to humans
- found in the gut of humans, but aides in normal digestion of our food
- unique sequence: TCCTGCACCCAGAGGACAAC

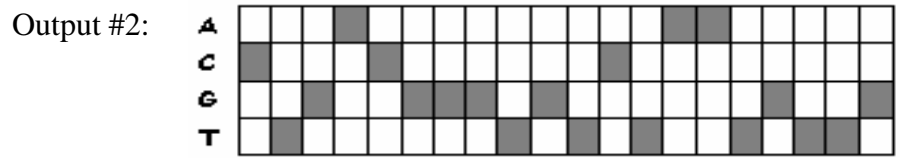
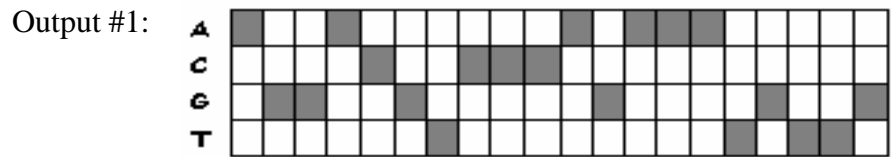
Y1088

- non-dangerous to humans but does lead to some severe illness (intestinal cramps, fever)
- commonly found in moist soil, but known to also be water borne
- important part of natural soil ecosystem
- unique sequence: TCCTGCAGGGTCTTTACAAC

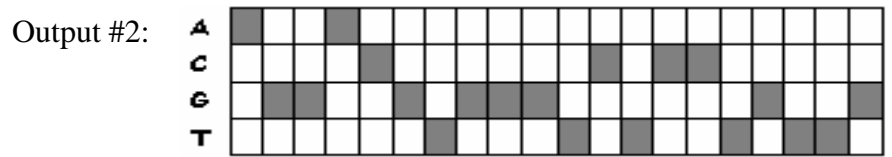
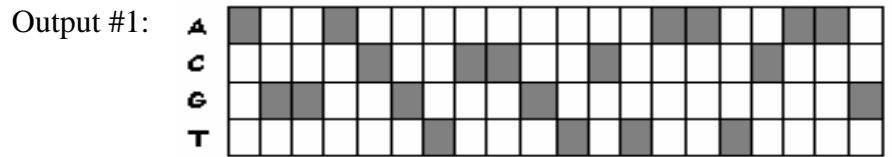
Sample #10 – Some *E. coli* found



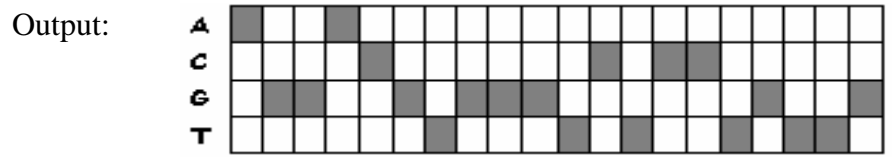
Sample #11 – Some *E. coli* found



Sample #12 – Some *E. coli* found



Sample #13 – Some *E. coli* found



Sample #14 – No *E. coli* found (blank output)

