• This Power Point presentation is designed to supplement the GeneChip® Microarrays curriculum, Activity #2.
The following will be covered in this presentation:

- A review of basic genetics important to the understanding of GeneChip® microarrays
- The structure of the GeneChip and how it functions
- Focus on the Gene Expression chip, but will also briefly cover the other two types of GeneChip® microarrays:
  - Genotyping and Resequencing
- How GeneChip® micorarrays can be applied to so many areas of science research
- Examples of research / experiments that used GeneChip® microarrays
- Hypothetical ethical issues with the use of GeneChip® microarrays in our society
• Despite the many differences between our DNA, overall, all human beings are actually 99.9% the same

• This is just one of the multitude of genetic information that was taken from the Human Genome Project

• One of the challenges facing us now is determining how all this genetic information will be used.

• How will we work together to prevent it from being used to stigmatize and discriminate?

• How will we use this information to improve the quality of life for human beings?

• We don’t have a choice about whether or not to participate in the Genetic Age. We are in it.
The Genome is Complex

- Only ~1.5% of genome (genetic material) encodes genes
- 20,000 to 30,000 genes
- Millions of differences between any two people

• The Genome is complex – it contains a immense amount of information
• Others have compared the sequencing of the genome to finding the periodic table of the elements.
• In that case we have found most of the elements
• Now we need to understand what they do
• In this context, that is about understand variability – How are we different
• We also need to understand the functions of the genes – what do the genes do
• This is made more complex because of splice variation –
• Genes in the genome consist of introns and exons – The dna is copied into RNA – then the introns are spliced out. The remaining messenger RNA is
Phenotypic Variation

- Single Nucleotide Polymorphism (SNP)
- Polymorphism
- Gene Expression

**Discovery of the genetic basis for the variation in health and appearance among human beings**

- Phenotype - outside appearance or result of the genes in a person
- Polymorphism – differences between any two people’s DNA
- Phenotypic variation is due to many factors within the genome
- We can learn about this variation by looking at SNPs and other polymorphisms, and the gene expression of different people
- This can help us learn about the genetic basis for variation in human health
• Most diseases are a result of the interaction of disease with environmental factors such as exercise, obesity, drugs, infectious diseases, lifestyle choices, and exposure.

• The amount that genetics and the environment play in the disease is variable:
  • Tay-Sachs and Cystic Fibrosis are almost 100% genetic, with little environmental influence
  • Diabetes and Cancer have both a heavy genetic component as well as are affected by environmental factors
  • Heart Disease and HIV/AIDS lean more towards the environment
  • Lead Toxicity is purely due to environmental factors
What is RNA?
The sequence of a strand of mRNA is based on the sequence of a complementary strand of DNA. Template for protein synthesis.

To understand where GeneChip® microarrays fit into this picture, we must first review how the genetic code leads to the building of proteins (protein synthesis):

• The code is held in the DNA, which is found in the nucleus of the cell
• The RNA is basically a movable copy of that code
• The RNA moves outside to the cytoplasm to act as a template guide for the building of the proteins at the ribosome
• The process of building the RNA from the DNA is called Transcription
• The process of building the protein based on the RNA is called Translation
• The term “gene expression” refers to the process that occurs when a gene guides the production
• A gene is “expressed” when the protein is made (the gene is “on”).
• Since the process of making a protein starts with the making of the RNA, one can measure gene expression by looking at the RNA produced by the gene
• An active gene will produce a lot of its RNA product while a low activity gene will produce very little RNA
• Measuring RNA volume to look at gene expression levels can be very powerful and useful when it comes to determining the function of a gene as well as finding drugs to fight the disease
• The example on the slide shows a Normal Cell and the genes that are active in that cell
• The y axis shows the amount of RNA present indicating the “volume” of the gene
• Each # represents a different gene
• Notice that the genes #1-15 and #63-97 are active at variable levels. The taller the line, the more RNA expressed (= the gene is actively expressed)
• The Malignant Cell has difference genes on and in different levels than the normal cell
• Notice the middle genes #17- 31 are very active, but were inactive in the normal cell
• Also, other genes that were on in low levels in the normal cell are much more active in the malignant cells (like #67 and 69)
• So, if someone wanted to work on stopping the malignant cells, they would want to focus their attention to the genes that are active in the malignant line, but not the normal line
• This is where GeneChip® microarrays come in – they help find those genes!
The GeneChip® System consists of arrays, hybridization ovens, fluidics stations, scanners, and software for data analysis.

All of these pieces of equipment are used in a full experiment with a microarray and how the results are analyzed.

Though other parts of the system will be discussed, this power point is focused mainly on how the arrays function and how they can be used (this is why the array is in red).
• The terms "wafers", "chips", and "features" will be used a lot in this presentation. This slide shows how they all relate to each other.
• These GeneChip® microarrays are a fusion of the technology of the semiconductor industry with life sciences – this is why these arrays are called “chips”.
• The chips are manufactured on wafers. These wafers are 5-inch-by-5-inch pieces of glass.
• You can get anywhere from 49 to 900 individual chips out of each wafer, depending on the overall size of the final chip to be made.
• A standard chip is 1.28 cm X 1.28 cm, the size of a thumbnail.
• On each one of those chips there are now over 6.5 million sections (squares) known as features.
• In each feature there are millions of identical DNA probes on them.
• In the diagram above, the probes are represented by the blue spiral connected to the feature.
  • A probe is a 25 base pair (25-mer) length piece of DNA that is attached to the chip.
  • Each feature has millions of identical probes attached to them.
  • The probes are used to "probe" the sample for certain DNA or RNA segments.
• The orange spirals with the stars at the end are RNA from a sample applied to the chip.
• Probes are the essential part of how the GeneChip functions and how you can use these chips to identify if a gene is on or off.
Hybridization is the basis for the function of microarrays

Hybridization: The process of joining two complementary strands of DNA to form a double-stranded molecule. The two strands are known as “complimentary”

This hybridization occurs because of complimentary base pairing rules (A – T, and C – G)

The hybridization of the probes on the features to RNA in sample from a cell being studied is the key to the function of the genechip (see next slide)
• Here is how hybridization fits into an experiment with the GeneChip® microarray
• Shown above is a single feature (grey square) of a chip made of millions of features
• The feature has an “array” (in this case 6) identical probes (purple spirals) on it
• A sample of RNA (orange spirals), each tagged with a molecule known as Biotin (the stars) is added to the feature
• The RNA will hybridize to the probes in the feature only IF the probe and RNA pieces are complimentary
• Since these features are so tiny, there must be a way to visualize this.
• This is performed using a biotin-fluorescent molecule interaction
• The Biotin “tags” on the RNA stuck to the feature are like a glue for a fluorescent stain
• When the fluorescent stain is washed over the feature, it will stick to whereever there is Biotin
• Since the Biotin is stuck to the RNA fragments, the fluorescent molecule essentially only sticks to places where the RNA is found
• Lasers are then used to detect the presence of the fluorescence, thus indicating the RNA target – probe match
Steps of the Expression Assay

- This is another representation of the full experiment (known as an assay) mentioned in the previous slide, though it illustrates the early part of the experiment in much more detail.
- First, the total mature RNA is isolated from the cell being studied. This RNA has already been “processed” (removal of the noncoding introns and splicing together of the coding exon) as well as the addition of a poly-A tail.
- Second, the RNA is turned into a double stranded DNA copy known as a cDNA. This is done through reverse transcription. This is done because RNA itself is not a very stable molecule and the cDNA is a way to store the RNA for a much longer period of time.
- Third, when it comes time to run the array, the cDNA is allowed to go through in vitro transcription back to RNA (now known as cRNA), but this RNA is labeled with Biotin. This is done by having all the uracil bases tagged with the Biotin. So, anytime a Uracil is added to the RNA chain during the transcription, a biotin molecule is also added.
- This labeled cRNA is then randomly fragmented into pieces anywhere from 30 to 400 base pairs in length (there is enough Biotin to make sure each RNA fragment has some biotin found on it).
- The fragmented, Biotin-labeled cRNA is then added to the array.
- Anywhere on the array where a RNA fragment and a probe are complimentary, the RNA sticks to the probes in the feature (remember there are millions of identical probes in each feature).
- The array is then washed to remove any RNA that is not stuck to an array (i.e., no match was made) and then stained with the fluorescent molecule that sticks to Biotin.
- Lastly, the entire array is scanned with a laser and the information is kept in a computer for quantitative analysis of what genes were expressed and at what approximate level.
• This slide shows two lab technicians performing two steps of the assay explained in the previous slide

• The first slide shows a technician adding the tagged RNA sample to the chip just before hybridization

• The second slide is showing a technician adding a tray of chips to a hybridization oven
  • The chips will rotate in the oven at 45°C for 16 hours
  • During this step all complimentary RNA will combine with the probes in the features
• The next six slides will provide a simple graphical explanation of the interaction of molecules in the GeneChip® microarray at a single feature.
• They will help explain how this interaction leads to a fluorescing feature that illustrates a gene is expressed in a cell.
• Most of the molecules in this example are simplified, such as the probes being shown as 7 bases long instead if 25 base pairs.

• This slide shows the dimensions of a typical chip and the connection from chip to feature to probe
• When the tagged (red ball) RNA (in purple) is added to the array, the RNAs “look” for their complimentary match
• This illustrates what happens when there is no match
• The RNA essentially bounces off the probes
• Even one base pair is enough to prevent the match

C does not stick to another C, so no match is made
• If there is a match, the RNA sticks to the probes in the array
• Since there are millions of probes in each array and multiple copies of the RNA (depending on the expression level of the gene that produced of the RNA), many RNA fragments combine to the many probes in the same feature (though only 2 are shown to match up)

• The key is that each probe represents a unique portion of a gene. This sequence is only found within that specific gene.
• When the RNA hooks up to a probe for a gene, it is hooking up to a sequence that is identical to a portion of the gene that produced it.
• Essentially, the connection means that gene was expressed! How is this connection visualized?
• The visualization occurs because the fluorescent molecules are washed over the entire array and stick to a RNA fragment tagged with the Biotin

• When hit with a laser in the scanner, any place the fluorescent molecule sticks, it will glow and be picked up by the computer

*Labeled RNA (purple) binds to the matching probes in the single feature.

*The feature will now fluoresce
• If there is no match between the probe and the RNA fragments, nothing sticks to the feature.
• So, there is no place for the fluorescent molecule to stick to.
• So, there will be no fluorescence at the feature when the laser hits it (the gene is not expressed in that cell line)
• This is an image of an actual expression array after scanning.
• The image to the left is the full array, with every feature represented – notice it looks pretty blurry and it’s hard to make things out.
• The second image on the right is a zoomed-in view of a section of the array.
• It represents about 324 features.
• Each feature is scanned for its’ intensity and it is evident that there is a wide range of intensities in this area.
• The black features represent no intensity (no RNA combined with the probes in the feature).
• The intensity level from lowest to highest by color is:
  • Dark blue -> Blue -> Light Blue -> Green -> Yellow -> Orange -> Red -> White
  • Remember that more intensity means more RNA combined with that feature, which basically means the gene was expressed at a higher level.
In actuality, there are three different GeneChip® microarray formats and each looks at a different aspect of the genome – Gene Expression (mRNA Analysis), DNA Analysis, and Resequencing.

- DNA Analysis looks at the variability a person has for a certain gene or polymorphism, like SNPs (they could have the A/A, A/B, or B/B combination).
- Resequencing is basically used to sequence DNA in a sample as a means for identification.

- These three arrays have a wide range of applications in various parts of research, drug development studies, medical diagnosis and prognosis, and the prescription of drugs (so called “personalized medicine”).
• These microarrays are being used in many different areas of society
• Anytime we can learn something by looking at the DNA involved, these chips are quick, easy, and can provide much more information than any genomic tool in the past
• Because of this a huge market they are being used more and more each year
• Areas of use include:
  • Agricultural Biotechnology (Developing the perfect strains of plants)
  • Environmental Testing (Testing causes for environmental degradation or outbreaks)
  • Food Testing (Testing for contaminants like other foods or pathogens)
  • Livestock Diagnostics / Grading (Designing the best livestock and grading their quality and safety)
  • Identity testing (Who are the parents? Can this person be related?)
  • Individualized medicine (Prescribing the best combination of medicines for a person)
  • Human diagnosis (Diagnosis of diseases)
  • Basic Research (The identifying of key genes/DNA at the start of research)
Plant and Animal Arrays

- Plant & animal research communities collaborate to create novel array
- Organisms planned for 2004:
  - Wheat
  - Rice
  - Poplar
  - Corn
  - Chicken
  - Grape
  - Soy
  - Tomato
  - Cow
  - Citrus
  - Pig
  - Cotton

- As mentioned previously, there are a large number of applications of these arrays beyond human disease research and medicine
- There are a lot of arrays being developed to study many plant and animals important to the agricultural industry.
- These arrays may help answer questions:
  - What is the best conditions to grow these plants / animals?
  - How can we design a “more efficient” plant or animal that grows at the best speed leading to the best taste / color for the consumer market
  - How can we fight diseases that damage our crops and livestock
• The use of microarrays can tell us a lot about specific disease such as cancer.
  • They can help us to
    • (1) Diagnose the specific cancer in a quick and accurate way
    • (2) Classify the specific subtype of cancer to allow for the best treatment
    • (3) Allow the most accurate prognosis of recovery based on the genetics of the tumor
    • (4) Identify the specific treatment for each type of specific genetic profile. The arrays can be used to study the genetics behind how a certain type of cancer reacts to a specific treatment
An example of how the microarray was used to learn about cancer can be illustrated with one of the first major studies into the genetics of cancer using a gene expression array – the classification of leukemia.

For a long time, scientists had a hard time correctly diagnosing which type of leukemia their patients had – ALL (Acute Lymphocytic Leukemia) or AML (Acute Myelogenous Leukemia).

Just looking at the cells was not affective enough and took a very trained eye to see the differences.

This was very important, because ALL and AML react differently to different medicines and treatments and progress differently.

Scientists took RNA samples from both cancers and place them on gene expression arrays.

The results were wonderful – it was clearly evident that the gene expression profile for ALL was different than that of AML.

Different genes were expressed for ALL compared to AML.

The crazy thing about the experiment is that in doing the study, the scientists noticed a third type of expression pattern showing up. This was leukemia, but was clearly not ALL or AML.

Scientists determined that they had found a third type of leukemia that they called MLL (Mixed-Lineage Leukemia).
Here is an example of how GeneChip® microarrays can be used in a comparison analysis study.

This study used a Yeast gene expression array to look at the different gene expression pattern for the yeast cells in two different nutritional conditions – rich medium or minimal medium.

The rich media has all the compounds needed for life by the yeast and much more, while minimal media has only the bare minimum (water, salts, and a few amino acids).

Though the image is black and white, the black represents no expression, and white represents a lot of expression and the various shades of grey are found in between.

In the images, you can see that gene “a” (RPL2A) is highly expressed in both conditions.

Gene “b” (TIP1) and “f” (SUP46) are expressed in both conditions, but slightly less under minimal media.

Gene “d” (VAP1) and “c” (BAP2) is slightly expressed under rich media, but not expressed in minimal media.

Gene “e” (YBR147W) is not expressed under rich media, but is expressed under conditions of the minimal medium.

Those genes expressed under minimal media must be part of a “survival” mode for the yeast, being expressed only when needed.

This shows that you can learn a lot about the function the genes by doing comparison of cells in different situations. Scientists could then focus on those genes that were expressed at different levels.

On a side note, notice how in reality, there are a multiple number of features used to detect the expression of a single gene (it is not really one feature = one gene).
Individual Variations in Drug Metabolism

CYP450 genes metabolize more than 90% of commercially available drugs

Poor Metabolizers
- Two mutant alleles
- No enzyme activity

Intermediate Metabolizers
- One reduced activity allele
- One null allele

Extensive Metabolizers
- At least one normal allele

Ultrarapid Metabolizers
- Multiple functional alleles
- Excess enzymatic activity

- Here is another example of an experiment that used GeneChip® technology
- The experiment used the CYP450 chip that looks at the CYP450 gene, which is found to be involved in the metabolism of 90% of commercially produced drugs
- The chip basically is a genotyping chip that looks at the versions a person has for the gene
- Allele – the different versions of a gene
- Remember that each person will have two alleles for CYP450, since they have a set of DNA – one half from mother and the other half from the father.
- Using the chip in conjunction with people’s individual ability to metabolize drugs, scientists were able to group people into four categories: Poor Metabolizers, Intermediate Metabolizers, Extensive Metabolizers, and Ultrarapid Metabolizers.
- A poor metabolizer is not able to break the drug down to make it useful to the body
  - These people have two mutated versions of CYP450, so the enzyme does not work at all
- A intermediate metabolizer (occurs at the greatest % of the population) can break the drug down better than those that are poor, but still not enough to receive its’ full affect
  - These people have a non-working mutant allele and one reduced activity allele
- Extensive metabolizers can break down the drug better than normal people and are able to get the most affect from the drug
  - These people have at least one normal allele (the second allele can be any version)
- Ultrarapid metabolizers break the drug down so fast that they can not get its’ full affect
  - These people have CYP450 enzymes that works too fast because they have multiple copies of the functional allele
- Using this chip, scientists could theoretically screen a person for their CYP450 genotype before prescribing the drug. This would allow them to prescribe the best amount or look for alternative methods if needed.
- The use of GeneChip® microarrays have a wide range of possibilities that can improve our quality of life
- However, this technology is very powerful, and with power comes moral and ethical concerns that should be addressed by society
- For example, with this technology, there is the possibility of developing a newborn screening chip that could tell parents a lot of things about their child before he/she is even born
- The chip could look for the presence of:
  - Late Onset Diseases (like Alzheimer’s Disease)
  - Treatable Infant Conditions (like Phenylketonuria)
  - Untreatable Infant Conditions (like Down's Syndrome)
  - Genetics that could help sales companies (like preferences to products)
  - General Skills & Traits (like height, athleticism, singing ability, etc)
- A chip like this brings up a lot of questions to society
  - Should parents be able to screen for this things – All of them? Some of them?
  - Will everyone have access to the technology?
  - What are the social implications of the use of this technology?
Questions: Newborn Genetic Screening

- Assuming 100% accuracy, would you test your newborn for:
  - Treatable Infant Condition (e.g. PKU)?
  - Skills and/or traits, (e.g., height, athletic ability, sexual orientation)?
  - Currently untreatable late onset disease (e.g., Alzheimer's)?
  - Untreatable infant condition which will lead to death before the age of 3 years (e.g. Swarey's Syndrome)?
  - Preferences, (e.g., food tastes, smells)?

- Here’s some questions for the audience to finish the presentation.