DNA Microarrays Primary Knowledge
Instructor Guide

Notes to Instructor:

This Primary Knowledge (PK) unit provides an overview of the DNA (Deoxyribonucleic acid) microarray and GeneChips®. Such devices consist of thousands of DNA or gene probes used to hybridize target molecules consisting of genomic DNA (gDNA) or cDNA (complementary or copy DNA, copied from messenger RNA). The detection of target DNA hybridized to a microarray is used in the medical or research lab to assess the types of mutations found in gDNA or the levels of gene expression indicated by mRNA levels. DNA microarrays are bioMEMS devices that are being used in genetic research, disease diagnostics, and many other fields.

In order to get the most from this learning module, participants should have knowledge of DNA, RNA, DNA transcription, and genomes. It is recommended that participants complete the SCME *DNA Overview Learning Module* and *DNA to Protein Overview Learning Module* prior to completing this PK. It would also be helpful if participants have a basic understanding of bioMEMS devices and their applications. Two SCME Learning Modules that will provide this background are *BioMEMS Overview* and *BioMEMS Applications Overview*.

This unit explains how a DNA Microarray works, how it is used in medical, environmental and food applications, and how it is fabricated. It is designed for high school and post-secondary freshman biology, science, and technology classes.   There are four suggested activities that reinforced the material presented in this PK as well as a final assessment.

This primary knowledge unit is part of the *DNA Microarray Learning Module*.
- Knowledge Probe (KP or pre-assessment)
- **DNA Microarray PK**
- DNA Hybridization Activity
- DNA Microarray Terminology Activity
- DNA Microarray Model Activity*
- The DNA Microarray - An Ethical Dilemma? Activity
- DNA Microarray Assessment

*A DNA Microarray Kit is available to support this learning module. The kit is required for the DNA Microarray Model Activity. The order a kit, please visit the SCME website (http://scme-nm.org).

Description and Estimated Time to Complete

The DNA Microarray Learning Module provides an overview of DNA microarrays – types of arrays, applications, interpretation, design, and fabrication. Activities provide the opportunity for students to gain a better understanding of these concepts as well as to discuss the ethical questions surrounding some of the applications and possibilities of these arrays

This DNA Microarray unit provides an overview of the DNA microarray and the GeneChip[©], devices used to identify specific DNA (Deoxyribonucleic acid) sequences in a sample. A DNA Microarray consists of thousands of DNA or gene probes used to hybridize target molecules consisting of genomic DNA (gDNA) or cDNA (complementary or copy DNA, copied from messenger RNA). The detection of target DNA hybridized to a microarray is used in the medical or research lab to assess the types of mutations found in gDNA or the levels of gene expression, "information about all messenger RNAs that are made in various cell types."[(NCI, 2010)]

Two GeneChips[©] by Affymetrix with projected results. Image courtesy of Affymetrix.

This unit explains what a DNA Microarray does, how it works, how it is used, and how it is fabricated. There are several names used for microarrays, depending on the method of fabrication and its specific application: GeneChips[®], DNA chip, DNA microarray, genome chip, and biochip. In this unit we will stick to the term DNA microarray, however, when quoting from other sources, you might see GeneChip[©] or another term.

To get the most out of this unit, you should have a basic understanding of DNA structure and function, DNA transcription and hybridization (or "base-pairing rules"), and terminology associated with genomes (an organism's genetic material). Some knowledge of bioMEMS would also be beneficial. If you need to review, it is highly recommended that you complete the following SCME Learning Modules prior to this one.
- *BioMEMS Overview*
- *BioMEMS Applications Overview*
- *DNA Overview*
- *DNA to Protein Overview*

These learning modules provide you with a review of DNA structure and function, DNA transcription and hybridization, as well as an overview of bioMEMS devices and their applications. Many of these terms may be new to you; therefore, a glossary is provided at the end of this unit for your reference.

Estimated Time to Complete

Allow approximately 30 minutes to an hour to read through this unit.

Introduction

The Human Genome Project(1990 through 2003) introduced an era in which individualized approaches to medicine are possible through an analysis of a person's DNA, or most specifically their gene expression levels and exact gene sequences (topics we'll discuss in this unit).   One outcome of this project was that the human genome (complete set of DNA) encodes approximately 30,000 genes. A person's specific genes are stored in each of one's cells. In fact, every single cell in a human body contains the exact same genes; however, the "activity" of genes varies from cell to cell. Between different human bodies and between different species the genes are not the same. Having a tool that identifies and compares genes between two or more samples would enable scientists to understand more about how genes affect who we are, who we aren't, why we develop a certain disease and why we don't. Such a tool was one of the outcomes of The Human Genome Project.
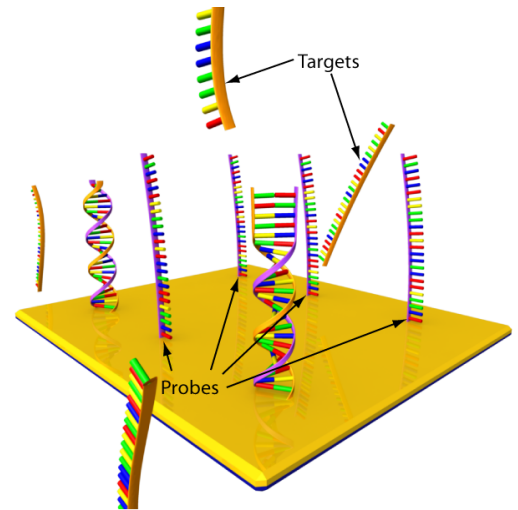
Double-stranded DNA (Deoxyribonucleic acid)

The goal of the Human Genome Project "was to provide researchers with powerful tools to understand the genetic factors in human disease, paving the way for new strategies for their diagnosis, treatment and prevention." [National Institutes of Health, 2009]Based the development of genomic-scale technologies such as DNA microarrays, one can say that the project reached its goal. A DNA microarray is a tool that "uses genome sequence information to analyze the structure and function of tens of thousands of genes at a time." [Bonetta, 2009] Before the invention of DNA microarrays, scientists were limited to the analysis of genes one or two at a time, making the "snap shot" of what was going on with the rest of the genome very limited.

So why is this important?

DNA microarrays are helping researchers learn more about human diseases, what causes them, how to identify them, and how to treat them. We now know more about complex diseases such as diabetes, multiple sclerosis, heart disease, and cancer than we have ever known before. For some diseases, such as multiple sclerosis, researchers have been able to identify specific genes that influence the risk of getting the disease.[Stimson, 2007] They have found that most diseases that are affected by one's genes are influenced by many, many genes and not just one or two. Such discoveries may eventually lead to the development of therapeutics that are needed to prevent a disease or to possibly cure it.

DNA microarrays are not only used in the medical field, but in other industries such as forensics, agriculture, and toxicology. In this unit you learn about the various applications of the DNA microarray and how this device works. You will also begin to study how microarrays are fabricated. A related activity provides the opportunity to simulate a fabrication process in order to gain a better understanding of these devices – how they work and how they are made.

This graphic is a simplification of a DNA Microarray. Fabricated onto a substrate (represented in gold) are single-stranded DNA molecules (probes). These probes are used to identify and recombine (hybridize) to complementary DNA(targets) in the sample being analyzed. We will discuss this process in more detail in this unit.

Objectives

- Describe three applications of the DNA microarray.
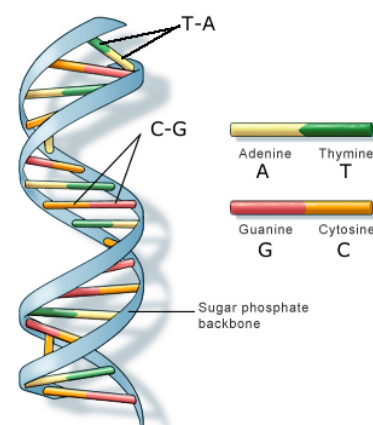- Explain how a DNA microarray works from hybridization to interpretation.

Terminology(Definitions are provided in the Glossary at the end of this unit)

Address
Allele
Capture molecules
cDNA
DNA
DNA microarray
DNA replication
Electrophoresis
Feature
Gene
Gene expression

GeneChip$^{©}$
Genome
Hybridization
Microarray
mRNA
Nucleotides
Oligonucleotide (or oligo)
Polymerase chain reaction (PCR)
Polymorphisms
Reverse transcriptase
Reverse transcription
RNA
RNA polymerase
SNPs (single nucleotide polymorphisms)
Southern Blot
Substrate
Target molecules
Transcription

Let's Review the Basics

Before jumping into how a DNA microarray works, let's have a quick review of Deoxyribonucleic acid (DNA), DNA replication and DNA hybridization. If you get stuck on a term, check the glossary at this end of this unit for its definition.
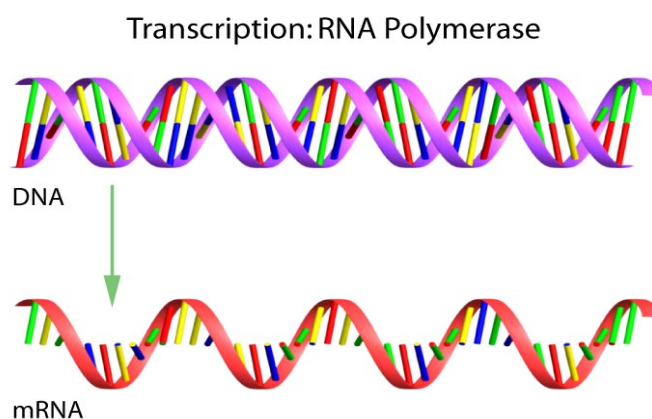
A Review of DNA[(V. Celeste Carter, 2009)]

Deoxyribonucleic acid (DNA) is a long polymeric molecule (found in most cells) that functions in the chromosome as the carrier of genetic information. It is a double stranded helix with a uniform diameter. The steps of the helix are formed by base pairs consisting of four (4) nitrogenous bases (Adenine (A), Thymine (T), Guanine (G), and Cytosine (C)). The ladder or rails are formed by nucleotides which are the nitrogenous bases, each consisting of a five-carbon sugar and at least one phosphate group which join the nitrogenous bases along the length of the DNA molecule. The two strands are anti-parallel and demonstrate complementary base-pairing (i.e., A-T, T-A, G-C, and C-G). *(see graphic right)* The bases are joined in the middle using hydrogen bonds. An A-T and T-A pairing each have two hydrogen bonds while a G-C and C-G pairing each have three hydrogen bonds.

The genetic information in the molecule is stored in the linear sequences of the base pairs. For example, A-T, T-A, G-C, and C-G is one partial sequence whereas T-A, G-C, A-T, and T-A may be another sequence. One DNA molecule may consist of millions of base pairs and thus, millions of linear sequences.

DNA Transcription[(V. Celeste Carter, 2009)]

DNA transcription is the process that creates a messenger Ribonucleic acid molecule or mRNA. A mRNA is needed to create a copy of a DNA molecule. During transcription the DNA sequences are "copied" into the mRNA molecule as shown in the graphic. In a process called "reverse transcription", the mRNA sequences are transferred into a DNA copy or cDNA.

Transcription is defined as DNA-directed RNA synthesis. It requires a DNA template, RNA polymerase (*an enzyme that produces RNA*), and ribonucleotide subunits. Transcription produces RNA complementary to the original DNA. This RNA is called the messenger RNA (**mRNA**).
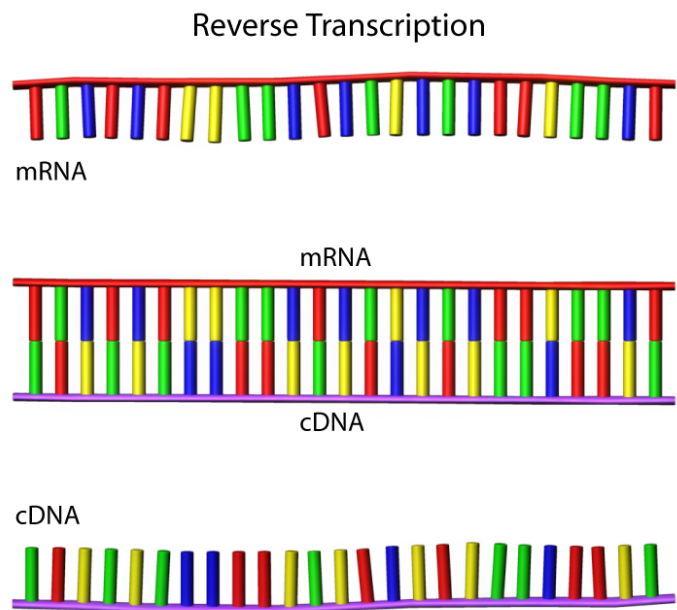
Previously we mentioned a gene's "activity" and that even though each cell contains the same genes, the activity of those genes vary from cell to cell. A gene's activity is described as either "on" or "off". A gene is "on" when it is producing a mRNA (i.e., DNA transcription); otherwise it is "off". DNA microarrays that create "expression profiles" identify the on/off activity of genes within a cell or organism; however, to do this, a DNA copy (cDNA) of the mRNA must be made. This is called "reverse transcription."

There are enzymes on the surfaces of your fingers and in every cell that is designed to break down RNA (not the case with DNA). Scientists will take a special enzyme and take that mRNA in cDNA. It is the cDNA that actually gets used.

Reverse Transcription

DNA microarrays use the cDNA rather than the mRNA because mRNA are unstable and do not last long. Such instability would make a microarray analysis unreliable. cDNA are more stable and thus, more reliable. cDNA are less likely to degrade during the microarray process. In reverse transcription, DNA copies (cDNA) are made from mRNA molecules as shown in the graphic.

Reverse transcription provides the cDNA needed for DNA microarrays to study expression profiles or gene activity in cells.

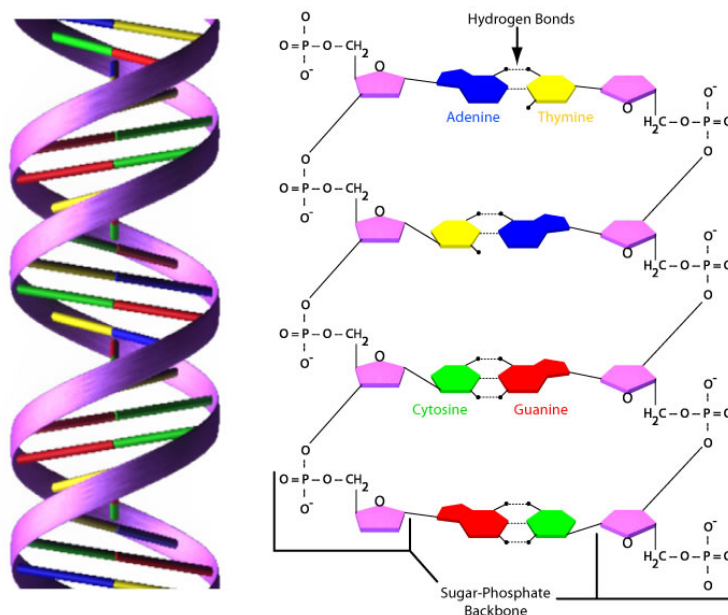### Reverse Transcription

mRNA

mRNA

cDNA
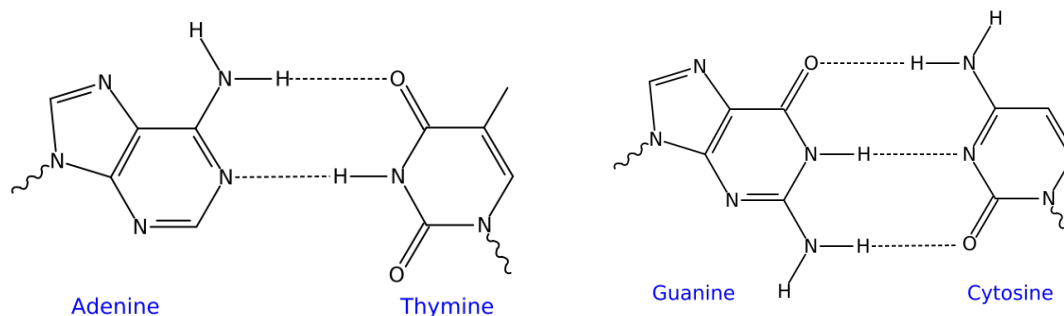
cDNA

DNA Hybridization

DNA hybridization refers to the process in which a double-stranded DNA (dsDNA) helix is denatured (or separated) into two, single stranded DNA (ssDNA) molecules by disrupting the hydrogen bonds that hold the two strands together.*(See the graphics below.)* Since hydrogen bonds are relatively weak, they can be disrupted by simply heating a DNA buffered solution and changing the solution chemistry with extremes of pH and high salt concentrations.

The image below illustrates a dsDNA molecule sequence and the hydrogen bonds that hold the two strands of the DNA together at each base-pairing (A-T, T-A, C-G, and G-C).
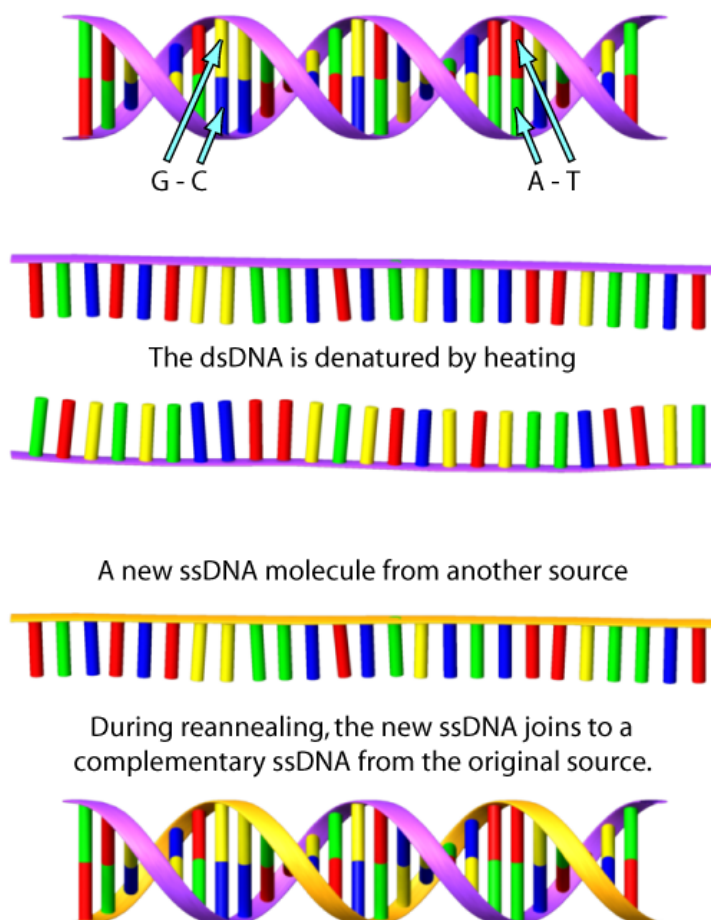
The graphic below shows the non-covalent hydrogen bonds between the pairs as dashed lines.
Left is an A-T base pair with two hydrogen bonds; right is a G-C base pair with three hydrogen bonds.

Adenine   Thymine   Guanine   Cytosine

DNA denaturation is reversible. If the buffer conditions and temperature are slowly changed back to normal, the two strands of ssDNA will again bind to each other, reannealing the two single strands back to their original double-stranded structure.

"DNA hybridization" is when the denatured DNA molecules are cooled down in the presence of ssDNA molecules from another source. These DNA molecules from another source are introduced to the original DNA solution during the reannealing step. If the original ssDNA strands have sequences that are complementary to the introduced ssDNA strands, they can form dsDNA hybrid molecules with one strand from each (an original ssDNA and the source ssDNA). The following graphic illustrates the hybridization process.

G - C   A - T

The dsDNA is denatured by heating

A new ssDNA molecule from another source

During reannealing, the new ssDNA joins to a complementary ssDNA from the original source.

The ability to make DNA hybrids is used in standard techniques in molecular genetics such as the binding of oligonucleotide (a short nucleic acid polymer, typically with 20 – 50 bases) probes in a Southern blot and the annealing of primers in PCR (polymerase chain reaction).

- Southern blot is a technique used to detect a specific DNA fragment in a DNA electrophoresis gel; in other words, to locate a specific base or DNA sequence within an entire genome.
- PCR is used to amplify DNA sequences and to make numerous copies of specific DNA segments quickly and accurately.

In each case, a synthetic single-stranded oligonucleotide is designed to search and find a complementary DNA sequence, then form a DNA hybrid. This process is similar to the "find" command in a word processing program. If the complementary DNA sequence to be found exists within the page of words, the oligonucleotide (*or oligo)* locates the word (or cDNA) and tags it. In DNA hybridizations, the oligo strand that carries a label (or sequence) for detecting the base-pairing of a hybrid molecule is called the probe.
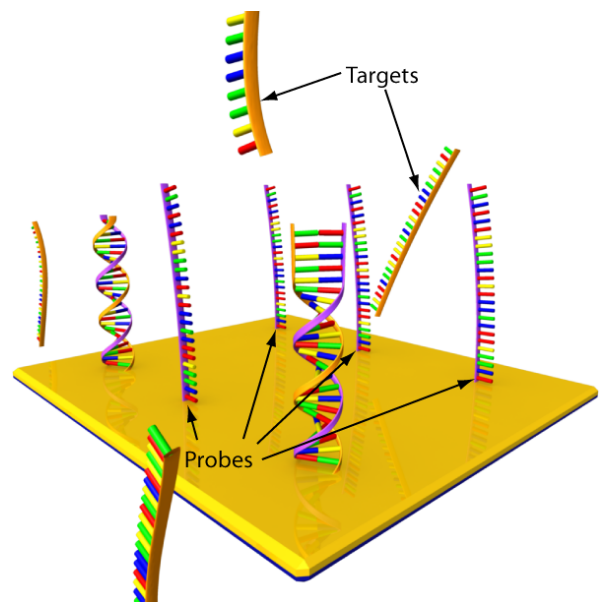
DNA Hybridization Activity

This is a good time to take a break from reading and complete an on-line tutorial on DNA Hybridization. Please complete the DNA Hybridization Activity that is part of this learning module.

What is a DNA Microarray?

DNA microarrays use gene sequencing and DNA transcription and hybridization to analyze and identify thousands of genes simultaneously. Each microarray consists of hundreds or thousands of gene sequences (ssDNA molecules or oligos) mounted on a chip and used as "probes". These probes detect complementary DNA fragments or cDNA copied from messenger RNA (mRNA) in a sample. The cDNA are the target molecules *(as shown in the graphic)*.

The DNA microarray relies on hybridization of DNA fragments to an oligo DNA sequence (a specific A,C, G, T combination). In DNA microarrays synthetic DNA oligos are fabricated and used as the capture molecules or probes. These synthetic DNA oligos are fabricated onto a solid surface (substrate) before the hybridization step. In the graphic to the right, the probes are the oligos. This graphic illustrates six identical probes on a substrate and several possible targets. Three of the probes have identified cDNA (targets) and have reannealed or are in the process of reannealing into a dsDNA. Those cDNA that do not match this specific probe's sequence continue to move through the microarray looking for a matching sequence at another location within the array.

In a DNA microarray, a grid is laid out on a surface such as glass microscope slide or silicon substrate, with an array of different DNA sequences. Each position of the array (called an "address" or "feature" of the grid), is looking for a specific gene sequence found in a particular organism. Therefore, each feature has a unique set of synthetic oligo probes attached to it. Each feature may contain hundreds or thousands of identical probes (oligos), while each array may contain tens of thousands of features. Because an array contains thousands of features, it can simultaneously search and "find" thousands of specific genes.

To ensure the accuracy of any DNA microarray test, several types of controls may be used (more on this later). In many tests, a control sample is used along with the test sample (both containing millions of cDNA). Detection of cDNA captured (or hybridized) on a microarray is most often done by tagging the cDNA in the control and test samples with a green and red fluorescent dye molecule, respectively. After hybridization, a laser scans the microarray and the presence of specific target DNA is detected by the fluorescence of the label at that position on the microarray. We'll discuss the function of the control sample as well as the meaning of the different colors (red, yellow, green, and black)in more detail later on in this unit.
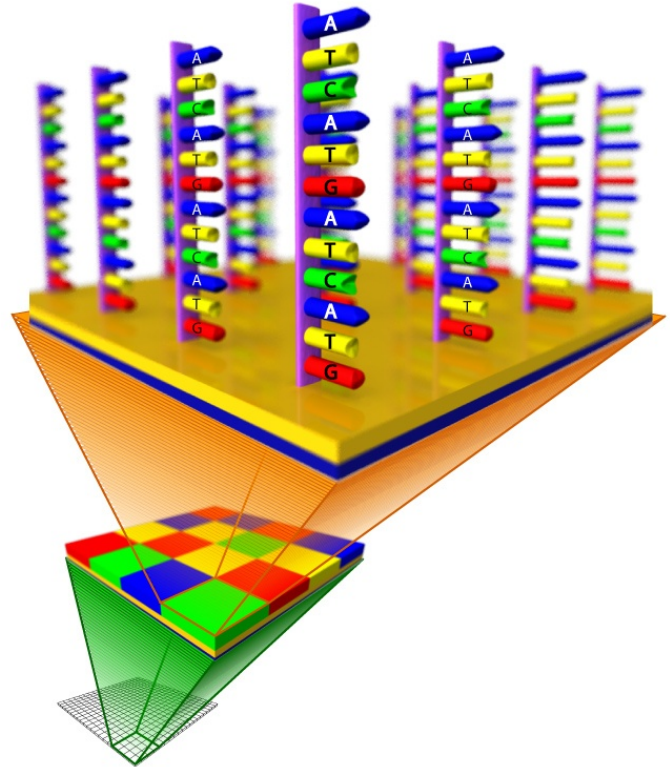
This image to the right is of a DNA microarray as seen through a microscope. Each tiny dot corresponds to an address or feature with oligos corresponding to one of the organism's thousands of genes. The color of the dot indicates the relative activity level of a gene. The red dotsindicate that the targeted DNA is present in the test sample. [Image courtesy of NASA. Image credit: James Smiley.]

How DNA Microarrays Work

This graphic illustrates a DNA microarray prior to being exposed to the control and test sample (or target) genes. The different colored blocks in the graphic (green, blue, yellow, red) represent different features (addresses) in a microarray. The oligos attached at each feature are "ssDNA probes" all having the same sequence (in this feature, G-T-A-C-T-A-G-T-A-C-T-A – bottom to top), of 20 to50 nucleotides, and up to a million probes per feature. Each oligo sequence is complementary to a gene of interest in the control and test samples.

DNA microarray expanded from many features (bottom grid) to a few features (middle grid), to a single feature (top grid) depicting a unique DNA sequence (G-T-A-C-T-A…). The coloration in this graphic is strictly to illustrate different locations (features) of ssDNA sequences in a DNA microarray. DNA, and thus a DNA microarray is actually colorless.

DNA Microarray Controls

Before moving on, let's talk about "controls". There different types of control used in a microarray experiement. Some controls are "positive" and "negative" controls, while others allow for direct comparisons between two different cells – a control cell and a test cell.

Positive and Negative controls
Positive and negative controls exist in all biological assays (tests that analyze biomolecules in specific events or conditions). The function of positive and negative controls is even more important for DNA microarrays because of the complexity of microarray fabrication. The purpose of positive and negative controls is to verify the overall performance of the microarray (Is it accurate and can it be trusted?) and the analytical technique (Were the samples prepared properly and the procedures executed correctly?).

A <u>negative control</u> is an array feature that is designed to have NO binding or hybridization. Its purpose is to avoid getting "false-positives" – positive results that should have been negative. In the design of a DNA microarray, bogus features that have no oligos attached may be interspersed throughout the array. If something does bind to one of these features, then the results of the test should be questioned and the test repeated. Prior to repeating the test, adjustments should be made to the preparation to better ensure accurate results.

A <u>positive control</u> is one that you expect to show a positive result. An example is a feature that contains a gene sequence that is ALWAYS present; therefore, you should see hybridization with both the control and test DNA samples. If this positive control does not show hybrization with both samples, then one would have to assume that there may be "false negatives" elsewhere in the array. Therefore, the assay results should not be trusted. In the case of a "negative" positive control, one should examine the procedure used to prepare the DNA samples to ensure that it was correct. If the procedure was correct, then one should suspect that the microarray is defective. In this case, the test should be discarded.

Direct Comparison Controls
In this type of control, each feature of the array is a comparison of the test sample DNA to the control sample DNA. Such controls are for the direct comparison of each address (or each gene) between two different cells. Below are examples of DNA microarray testing and direct comparison controls:
1. Testing the effects of drugs or toxins on gene expression (which genes are "active" or "inactive") – In this test, cDNA from an untreated cells would be compared to the cDNA from treated cells.
2. Testing the mutation of genes in cancer cells – In this test, the genomic DNA (gDNA) from the patient's noncancerous tissue would be compared to the gDNA from the patient's cancerous tissue or tumor. The results from such a test can help researchers find new cancer genes as well as help doctors decide how aggressive a patient's cancer is and thus, the best treatment to prescribe.

The following cDNA Microarray Experiment is an example of a direct comparison control assay.
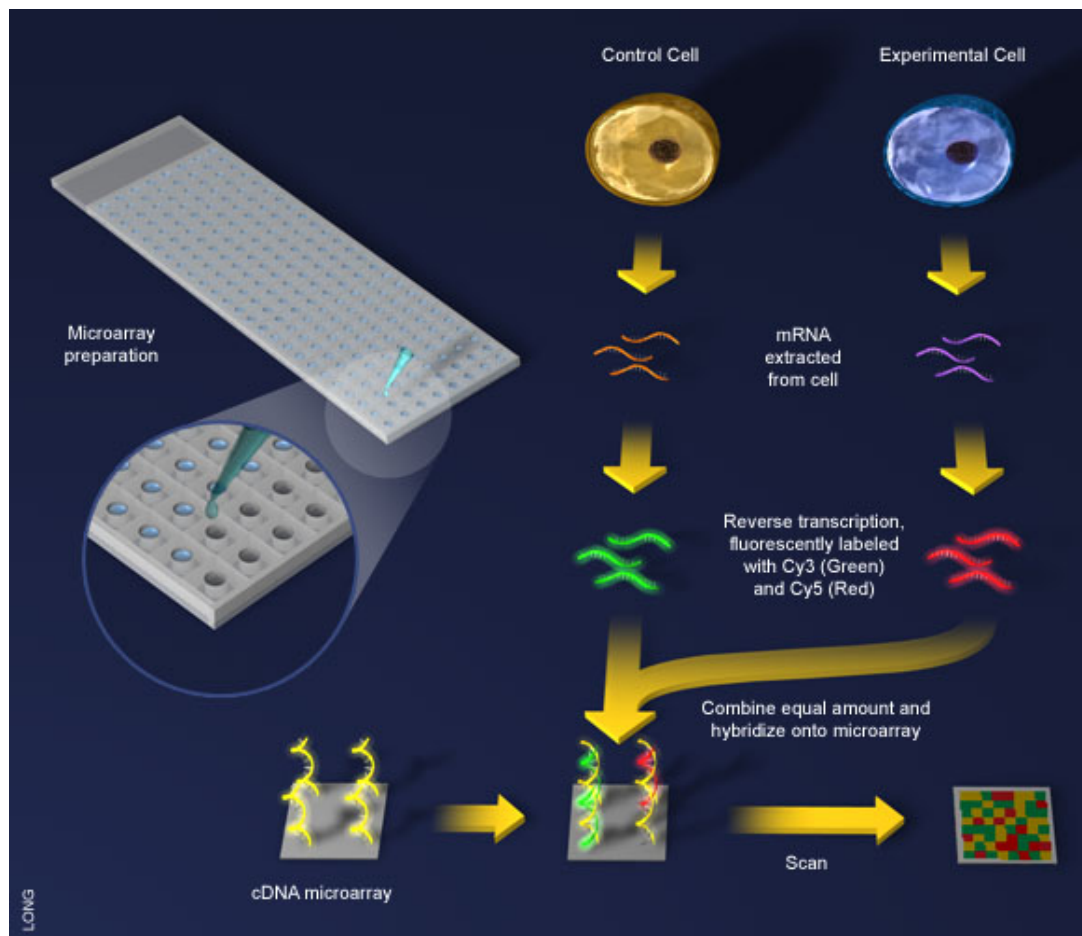
Note to Instructor: A few more notes on controls.

Positive controls also are used to "tune" a microarray. Sometimes the balance of red-labeled to green-labeled DNA is off for a variety of reasons, generally having to do with the procedures for preparing the target DNA and the rate of fading of the red vs. green fluorescent label. This would be detected in the positive control features where the red vs. green intensity should be the same. If they are not, the instrument will add to the intensity of the deficient color throughout the microarray in order to balance the labels out. This is referred to "normalization".

As for the use of ratios of red:green (test:control DNA) - "The control DNA sample rules out other factors that might play a role in a low (or high) signal strength in the binding level of test DNA. An example of another factor that could lead to high signal strength at an address is, if during fabrication of the microarray a more abundant density of capture (probe) oligos is made in that feature, this would result in a higher amount of test DNA bound to that address in the array. Conversely, an address with a lower density of oligo synthesized will result in less test DNA bound there. By comparing a ratio of test DNA to control DNA bound at all of the arrays, it is possible to rule out these types of problems that can arise during fabrication.

Prior to the DNA hybridization reaction between the probes (on the microarray) and targets, two samples are prepared – the test sample and a control sample (the targets). EachssDNA sequence in the control sample is labeled with a green fluorescent marker, while each ssDNA sequence in the test sample is labeled with a red fluorescent label. When the samples are combined and applied to or "washed over" the microarray, DNA targets in the samples find complementary DNA probes within the microarray and hybridize forming dsDNA hybrids. In the graphic below DNA has been extracted from both the control cell (control sample) and the experimental cell (test sample), transcribed to mRNA which is converted to cDNA through reverse transcription. The cDNA are tagged with their respective fluorescent labels (green for control and red for test sample). The two samples are mixed together and each cDNAin the samples hybridize with a complementary synthetic ssDNA at a specific feature in the microarray.
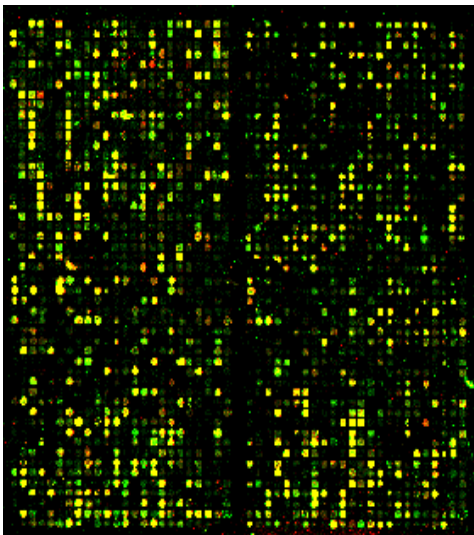

An example of a cDNA Microarray Experiment. RNA is extracted from two different samples and converted into complementary DNA (cDNA), during which the DNA is labeled with florescent compounds. The two samples are then mixed together for comparison and hybridized to the array. Differences in gene expression are revealed by fluorescent patterns on the array.[Courtesy of "The Science Creative Quarterly". scq.ubc.ca. Artist: Jiang Long}

Now let's talk about how to retrieve and interpret the results of the hybridization processes that take place during a test.
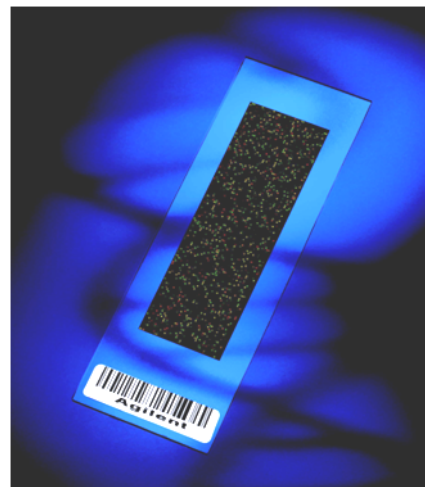
Interpretation of Microarray Results

After hybridization is complete, each feature of a microarray grid is scanned by a green laser and a red laser to detect the presence of both control and test DNA hybridization. As shown below, the scan results in an image of red, green, yellow and black dots or labels. Each color represents the activity or expression level at each address with the test or control sample.

- A green label at an address indicates the presence of DNA mostly from the control sample.
- A red label at an address indicates the presence of DNA mostly from the test sample.
- A yellow appearance at an address indicates an equal amount of target DNA from the control sample and the test sample.
- No color or black at an address indicates that neither the control nor the test samples had DNA complementary to that DNA probe's sequence.

A fluorescing DNA microarray showing the results of DNA hybridization between the probe and target DNAs. . When cDNA is prepared from a test sample (red) and from a control sample (green), and both are hybridized to the microarray, the color of the dot indicates the relative activity level of that gene. A green dot shows activity in the control only, red shows activity in the test tissue only, and yellow shows activity in both. Black is the absence of activity from either sample. [Image courtesy of NASA]
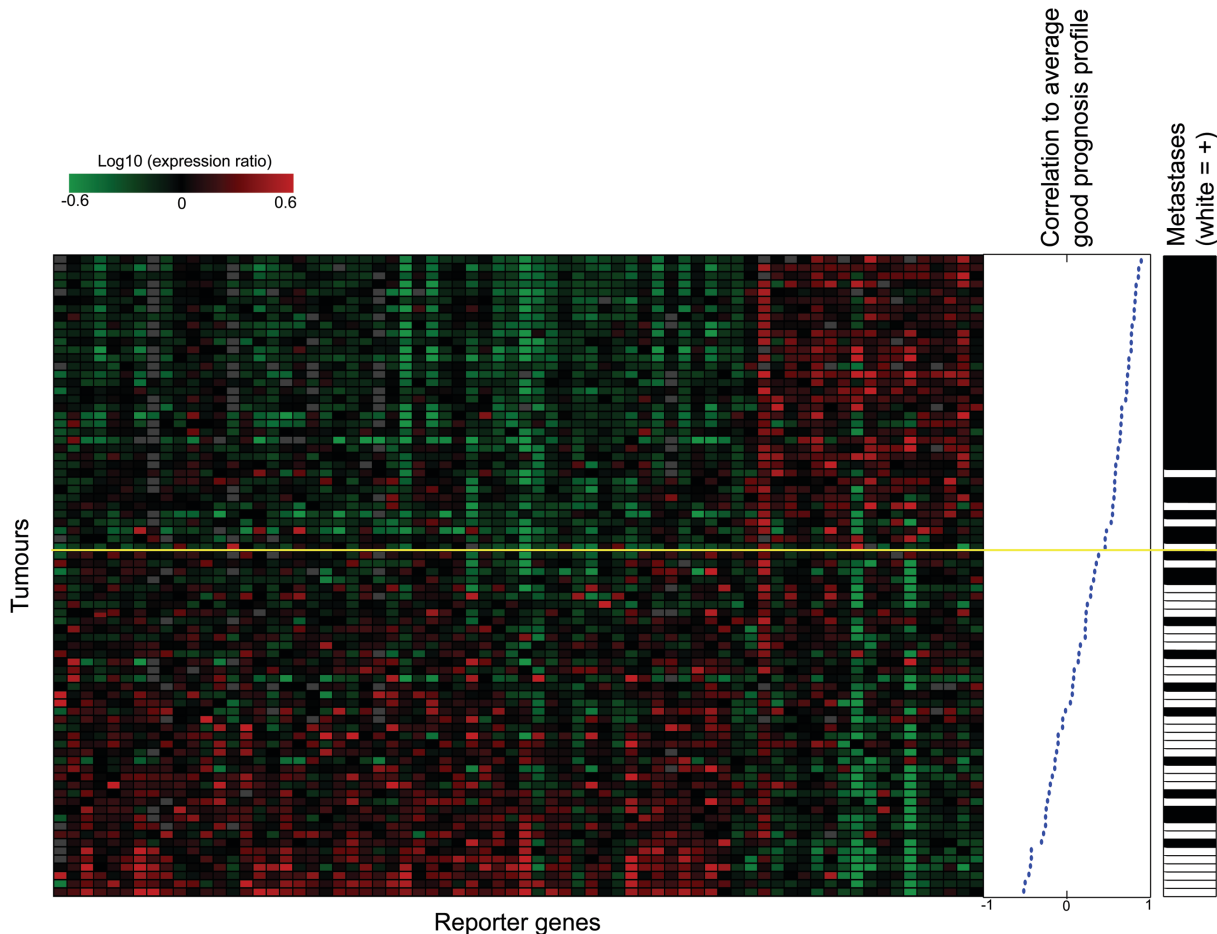
The images below show an Agilent Technologies microarray printed on a1" x 3" glass slide format. The image on the right shows the microarray after hybridization and while being scanned with a laser. One can see the fluorescence of hybridized molecules.[Agilent, 2003] [Images courtesy of Agilent Technologies]

An Example of an DNA Microarray Analysis

In February of 2007, the U.S. Food and Drug Administration (FDA) approved the marketing of the MammaPrint test – a test that uses DNA microarrays to "predict whether existing cancer will metastasize (spread to other parts of the patient's body). The reoccurrence of cancer is partly dependent on the activation and suppression of certain genes located in the tumor. Prognostic tests like the MammaPrint (below) can measure the activity of these genes and thus help physicians understand their patient's odds of the cancer spreading." MammaPrint was developed by Agendia, in Amsterdam, Netherlands. [FDA, 2007]Beloware the results of a MammaPrint developed in 2006.

This image is the gene expression data matrix of 70 prognostic markers genes from tumors of 78 breast cancer patients hybridized using a DNA microarray referred to as the MammaPrint. "Each row represents a tumor and each column a gene…The metastases status for each patient is shown in the right panel. White indicates patients who developed metastases within 5 years after the initial diagnosis, black indicates patients who continued to be metastasis free for at least 5years."[Glas, 2006]Those patients above the threshold (yellow line) have a low risk of the cancer spreading to other parts of the body, while those below the line are considered high risk patients and have a poor prognosis.

Results such as these and similar DNA microarrays tests provide medical professionals with knowledge of one's level of susceptibility to a specific disease, leading to possible prevention or treatment, and probable prognosis.

Applications of DNA Microarrays

DNA microarrays are being use to study the human genome as well as the genome of other species. They are used for many different purposes and applications. For each species and application a specific selection of probes are included in the array. Researchers can purchase a microarray specific to a particular species' genome, such as a mouse microarray or a human microarray. Such microarrays contain a probe for each of the genes found in that organism. Genetic comparisons can now be made between organisms of different species or the same species.

Two GeneChips® by Affymetrix. One for the Human Genome and one for Mouse Genome. Image courtesy of Affymetrix.

Types of Microarrays

There are two basic types of DNA microarrays: direct detection of genes and gene expression.

- Direct detection microarrays detect specific genes or gene mutations within a sample, such as the example above for the *BRCA1* mutation. Direct detection microarrays are being used to identify specific genes that cause a specific disease, and to screen for mutations that are responsible for genetic disorders when there are multiple gene mutations that can possibly cause the disorder. Direct detection microarrays are being used to profile somatic mutations in cancer, for forensic applications, genotyping, identifying DNA-based drugs, and in agriculture, "to guide the genetic selection process of milk-producing animals." (Mertens, 2009)

SNP (single nucleotide polymorphisms) chips – a type of direct detection DNA microarray
The SNP chips in the photograph are bovine assays (or analysis) that "easily and quickly identify regions within the bovine genome that harbor variants that cause the animals to differ in the outward expression of important traits, allowing scientists to predict an animal's genetic merit from its SNP profile."(Mertens, 2009)[Courtesy of Jeremy Taylor, Animal Genomics, University of Missouri]

- Gene expression microarrays detect "expression levels" in a sample - when mRNA copies to cDNA (i.e., which genes are "active" or "inactive"). This information is called a "gene expression profile". Gene expression microarrays detect how cells and organisms change and adapt to specific stimuli such as changes in the environment or one's disease state.

  Let's remember that every cell in a human's body contains the same genes. However, the same genes are not active in every cell. Gene expression microarrays allow us to identify which genes are active and which are inactive in different cell types. They also allow us to determine which

genes are active in different cell states. For example, gene expression microarrays can identify which genes are turned on during cell division or when exposed to an external stimuli, such as a drug. This analysis enables "scientists to understand both how these cells function normally and how they are affected when various genes do not perform properly."[(NHGRI, 2009)]

Gene expression microarrays are being used to study cancer-causing genes and to customize drug therapies based on how our genes react to specific drugs or drug dosages. They are being used to study how cells react to changes in the environment, such as increases in pollutants or toxins.

Expanded Notes for the Instructor

I.   <u>Direct detection of genes</u>. This approach can be used to detect point mutations, as well as deletions of chromosomal DNA. Alternatively, it can be used to detect pathogens (the cause of a disease) through the presence of pathogenicity genes (those genes that cause a disease). Since a large number of pathogens can be detected simultaneously, microarrays are especially useful to quickly screen food and environmental samples for the existence of pathogens. Microarrays are also used for personalized medicine. They are able to profile an individual's unique alleles (alternative forms of a gene) involved in drug metabolism. As a result of this profile, the most effective drug and most effective dosage can be determined.

SNP (single nucleotide polymorphisms) chips
Microarrays developed to locate genetic variations within the same organism or species are mapped for a particular organism (e.g., a microarray to identify the gene variations between the queen bee, worker bee and drone). These microarrays are often referred to as "SNP chips" for "single nucleotide polymorphisms". A SNP is a change of a single nucleotide in the DNA sequence. When a SNP occurs in the coding region of a gene, it is a mutation that can be responsible for a genetic disorder.

SNP chips are especially useful for screening for mutations that are responsible for genetic disorders when there are multiple gene mutations that can possibly cause the disorder. Once a SNP profile is associated with a particular disease, individuals can be tested for that specific SNP profile to determine if he or she is susceptible to or at risk of developing that disease. [(NCBI, 2007)] SNP microarrays are also being used to profile somatic mutations in cancer, for forensic applications, genotyping, and for identifying DNA-based drugs. All of these applications are associated with many genetic polymorphisms working in combination.

SNPs are currently being tested for use in agriculture. For example, the dairy breeding industry is using microarrays to study complex traits in animals, such as profiling the genes of dairy cows and using that information "to guide the genetic selection process of milk-producing animals." [(Mertens, 2009)]

II.   <u>Gene Expression or Changes in Gene Expression Levels</u>. Research that measures mRNA or cDNA from many different genes at the same time can be useful in diagnosis, as well as in identification of target genes for new drugs. The most useful application of this approach has been in the study of cancer, where microarrays have played an important role in the current rapid discovery of new oncogenes (cancer-causing genes) and defective tumor suppressor genes. Such microarrays can

also be used to customize drug therapies based on drug metabolizing profiles of the individual, in other words, gene expression arrays can study how cells react to specific drugs or drug dosages.

Gene expression can also be used to study how cells react and, in some cases adapt, to changes in environment, such as an increase in pollutants or toxins. The information gained by measuring the gene expression levels upon exposure to a chemical "can be used both to provide information about the mechanism or action of the toxicant and to form a sort of "genetic signature" for the identification of toxic products."[(Lettieri, 2006)]

Gene Expression Profiling

A gene expression profile is developed from "information about all messenger RNAs (mRNAs) that are made in various cell types." Gene expression profiling uses a profile "to find and diagnose a disease or condition and to see how well it responds to treatment. Gene expression profiles may be used in personalized medicine." (NCI, 2010) In addition, this process allows us to identify triggers for diseases by studying the activity of genes within our cells.

End of expanded Instructor Notes

As you can see from this discussion, DNA microarrays are being used to classify cancers, to assess one's risk for a specific disease, and to track disease progression and predict prognosis. They are proving to be invaluable for identifying drug therapies and drug response (personalized medicine). In genetics, DNA microarrays allow us to compare the genes between two or more different organism as well as the genes within the same organism. Outside of the medical field, DNA microarrays are being used to ensure that our food and the air we breathe are safe, and to assist farmers and ranchers in food production. "If the work involves the DNA or the genetics of a cell, …microarrays can be used."[(Zaccheo, 2005)]

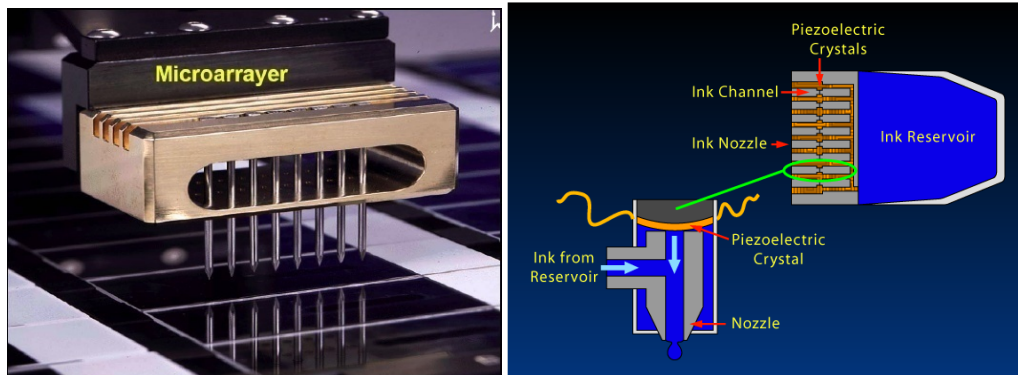**Key features of DNA Microarrays**

The key feature of DNA microarrays is the ability to screen for many DNA sequences in one run. In fact, microarrays have enough features (addresses) to screen ALL of the genes of an organism's genome in one hybridization experiment. Even with thousands upon thousands of different features, the microarray itself is very small - the size of a microscope slide and even smaller. So, although they can store as many as a million features and can screen for the presence of a million different DNA sequences in a single experiment, their size can be on the order of a regular size postage stamp!

The size of the individual feature is even smaller – MUCH smaller. Each feature can be as small as 200 nanometers (200×10^{-9} meters, or 0.2 micrometers) – placing this technology in the "nanotechnology" range. The feature size makes fabrication of DNA microarrays a challenge but not impossible since much of the technology that is needed to fabricate microarrays already exists.

Many of the fabrication methods for microarrays come from semiconductor fabrication technology as well as microtechnology. For example, one type of microarray fabrication technique uses a printer similar to an ink-jet printhead (a micro-size device) to print the addresses onto the microarray slide. However, the "ink" is an oligonucleotide solution. Another method of microarray fabrication uses the photolithography process used in micro-fabrication. Let's take a closer look at both processes.

DNA Microarray Fabrication

The DNA microarrayer (below left) "prints" oligonucleotide probes to specific addresses on a microscope slide using the microtechnology of an ink-jet printer (below right). The inkjet "printhead" consists of an ink reservoir and piezoelectric actuators that allow the fluid ink to flow from the reservoir to the nozzles. Because the diameter of the nozzles is in the micrometer-scale (between 1 to 100 micrometers), capillary action moves the molecules of ink through the tiny nozzles. This same technology is used to achieve the nano-size probes when printing microarrays. Because the size of a probe is in the nano-scale, the printhead nozzles are smaller allowing a single microarray to have as many as a million features printed onto one slide.

Alternatively, the Affymetrix GeneChip® technology uses a fabrication technique called photolithography to synthesize the oligonucleotides on a silica (or "chip") surface. The Affymetrix GeneChip® is a DNA microarray that has oligonucleotide probes synthesized in place on a silicon chip using a photolithography process borrowed from the semiconductor fabrication industry. The oligonucleotides on these arrays tend to be shorter, generally 20 nucleotides long and the array is miniaturized even more than those found in other DNA microarrays. Each feature on a GeneChip® may be as small as 50 nanometers square – almost 2000 times smaller than the width of a strand of hair!

A more in-depth look at this process is included in the The DNA Microarray Model Activity.

[The photo shows a GeneChip® (blue) for the human genome. The case that encloses the GeneChip® is about as wide as the length of a wooden match. Image courtesy of Affymetrix]

The next generation of DNA microarrays will be MEMS devices (microelectromechanical systems) containing mass sensitive microcantilevers and electrochemical detection. These arrays will be 100 times more sensitive, down to attomolar (10^{-18}) sensitivity! They will also have greater precision and better reproducibility, faster and more robust (fewer steps in assay).[Yan, Tang, Zhai, & Ju, 2007]

Summary

The DNA microarray has opened up a whole new frontier for exploration in medical research, drug development, forensics, toxicology, and food production, just to name a few. All of the information derived from DNA microarrays affects us all in one way or another. Through the hybridization of synthetic oligos and target DNA molecules we can identify the presence of specific genes, mutation and pathogens. We are getting closer and closer to knowing what makes us tick, what makes us sick and what can make us well.

Food for Thought

The DNA microarray may be small, but when it comes to career potential, it's huge!
- Discuss at least three applications or potential applications of DNA microarrays.
- How does a DNA microarray identify a target DNA?
- What are some of the careers that one might look into that involve the use of or the fabrication of DNA microarrays?

References

1. National Institutes of Health. (2009, March 11). *Human Genome Project Fact Sheet.* Retrieved from National Human Genome Research Institute. https://bit.ly/2KOBkGJ

2. Agilent. (2003, October 8). *Image Library.* Retrieved from Agilent Technologies: https://www.agilent.com/search/?Ntt=image%20library

3. Bonetta, L. (2009). *The Basics of DNA Microarrays.* Retrieved from BioInteractive - Howard Hughes Medical Institute.

4. Cooper, C. S. (2001, March 20). *Applications of microarray technology in breast cancer research.* Retrieved from Breast Cancer Research: http://breast-cancer-research.com/content/3/3/158

5. FDA. (2007, February 6). *FDA Clears Breast Cancer Specific Molecular Prognostic Test.* Retrieved from FDA-News & Events.

6. Glas, A. M. (2006). *Converting a breast cancer microarray signature into a high-throughput diagnostic test.* Retrieved from BioMed Central: http://www.biomedcentral.com/1471-2164/7/278

7. Lettieri, T. (2006, January). *Environmental Health Perspectives.* "Recent Applications of DNA Microarray Technology to toxicology and ecotoxicology". Retrieved from PubMed Central: https://www.ncbi.nlm.nih.gov/pubmed/16393650

8. Umesh, Singh. (2014 June). *Molecular markers and their applications in cattle genetic research.* ScienceDirect Vol 6, Issue 2, June 2014, Pages 49-58.

9. NCBI. (2001). *Appendix I. Immunologists' Toolbox (Immuno Biology).* Retrieved from National Center for Biotechnology: http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=imm&part=A2395

10. NCBI. (2007, July 27). *Microarrys: Chipping Away at the Mysteries of Science and Medicine.* Retrieved from NCBI (National Center for Biotechnology Information): A Science Primer: https://bit.ly/2GLFyw7

11. NCI. (2010). *Dictionary of Cancer Terms.* Retrieved from National Cancer Institute: http://www.cancer.gov/dictionary/?CdrID=386201

12. NHGRI. (2015, August). *DNA Microarray Technology.* Retrieved from National Human Genome Research Institute: http://www.genome.gov/10000533

13. Stimson, D. (2007, July 29). *After a decades-long search, scientists identify new genetic risks factors for multiple sclerosis.* Retrieved from EurekAlert!: http://www.eurekalert.org/pub_releases/2007-07/nion-aad072607.php

14. V. Celeste Carter, P. D. (2017). *DNA Overview Learning Module.* Retrieved from Southwest Center for Microsystems Education. http://scme-support.org/index.php/category-03/dna-overview

15. Wikipedia. (2018, May 12). *DNA Microarray.* Retrieved from Wikipedia, the Free Encyclopedia: http://en.wikipedia.org/wiki/DNA_microarray

16. Yan, J. W., Tang, J., Zhai, C., & Ju, H. (2007). A Disposable Simultaneous Determination of Tumor Markers. *Clinical Chemistry 53:81* , pp. 1495-1502.

17. Zaccheo, R. K. (2005). *GeneChip Microarrays Curriculum for Life Science Teachers.* Retrieved from Affymetrix. https://bit.ly/2LoYQep

Glossary

Address – A single location on a DNA microarray that can contain up to a million of the same ssDNA sequence. Also called a feature.

Allele – One member of a pair or variation of genes that occupy a specific position on a specific chromosome.

Anti-parallel – Two molecules that are side-by-side by run in opposite directions. In DNA the two strands are anti-parallel because the head of one strand lays against the tail of the other strand.

Base Pair – The pairing of a nitrogenous base using two hydrogen bonds for A-T and T-A pairs and three hydrogen bonds for C-G and G-C pairs.

Capture molecules – Specific DNA binding sequences used as probes on DNA microarrays.

cDNA (Complementary DNA) - Single-stranded DNA that is complementary to messenger RNA or DNA that has been synthesized from messenger RNA by reverse transcriptase.

Deoxyribonucleic acid (DNA) - A long linear polymer formed from nucleotide pairs, and shaped like a double helix; associated with the transmission of genetic information.

DNA microarray - A tool that "uses genome sequence information to analyze the structure and function of tens of thousands of genes at a time." A microchip that holds DNA probes that form half of the DNA double helix (called single-stranded DNA or ssDNA) and can recognize other ssDNA from samples being tested.

DNA replication – The process in which each strand of a double-stranded DNA molecule serves as template for the reproduction of two identical DNA molecules.

Electrophoresis- A method used to separate particles, such as DNA or proteins, in which an electric current is passed from one electrode to another of an opposite charge through a medium, and the separation of the molecules depends on the rate at which the molecules travel towards the electrode based on their electrical charge. *Jonas: Mosby's Dictionary of Complementary and Alternative Medicine. (c) 2005, Elsevier.* – The movement of charged suspended particles through a liquid medium in response to changes in an electric field. *Mosby's Medical Dictionary, 8th edition. © 2009, Elsevier.*

Feature - A single location on a DNA microarray that can contain up to a million of the same ssDNA sequence. Also called an address.

Gene - A length of DNA sequence that contains information that is capable of being translated into a polypeptide product.

Gene Expression – Conversion of the information encoded in a gene first into a messenger RNA and then to a functional gene product, which is often a protein.

GeneChip®- A DNA microarray developed by Affymetrix that uses ssDNA probes to recognize complementary ssDNA from samples being tested.

Genome – An organism's genetic material which is made up of molecules of DNA and found in the nucleus of eukaryotic cells and in the cytoplasm of prokaryotic cells.

Hybridization–A process of combining two complementary single-stranded DNA into a single double-stranded molecule through base pairing.

Microarray– A micro-sized lab-on-a-chip (LOC) with a two-dimensional array *printed* onto a substrate. The array is used to assay (analyze) large amounts of biological materials in a sample solution.

Nucleotides – Molecular subunits consisting of a nitrogenous base (A, T, C, or G), a sugar and at least one phosphate, that when joined together, make up the structural unit of RNA and DNA.

Oligonucleotide – A short fragment of single-stranded DNA typically 5 to 50 nucleotides long.

Polymerase Chain Reaction (PCR) -A laboratory technique that can amplify the amount of DNA from a tiny sample to a large amount within just a few hours. PCR can take one molecule and produce copies of a DNA sequence through replication with a heat-stable DNA polymerase and two oligonucleotide primers that flank the target sequence.

Polymorphisms - Variations of forms. A DNA polymorphism is a mutation or change in DNA sequence that is found in at least 1% of the species.

Reverse transcriptase – An enzyme isolated from certain viruses that can make a DNA sequence complementary to an RNA sequence. The copy DNA sequence is referred to as cDNA.

Reverse transcription– The process of making a copy DNA sequence from an RNA sequence using the enzyme reverse transcriptase.

Ribonucleic acid (RNA) -A polymer consisting of a long, usually single-stranded chain of alternating phosphate and ribose units with the bases bonded to the ribose.

RNA polymerase– The enzyme responsible for locating a gene on a DNA sequence and making an RNA copy from the DNA sequence of the gene.

Southern Blot – The transfer of DNA fragments in an electrophoresis gel to a plastic membrane, followed by hybridization with a radio labeled oligonucleotide probe complementary to specific DNA sequences. Used in DNA fingerprinting.

Single Nucleotide Polymorphisms (SNPs) - Microarrays developed to locate genetic variations (i.e., sequence polymorphisms – the difference between the queen bee, worker bee and drone) within the same organism or species are mapped for a particular organism.

Target molecules–The molecules in a sample solution that are to be assayed or analyzed using a microarray.

Transcription– The process of making RNA sequences from DNA sequences, using RNA polymerase enzyme.

Disclaimer

The information contained herein is considered to be true and accurate; however the Southwest Center for Microsystems Education (SCME) makes no guarantees concerning the authenticity of any statement. SCME accepts no liability for the content of this unit, or for the consequences of any actions taken on the basis of the information provided.