

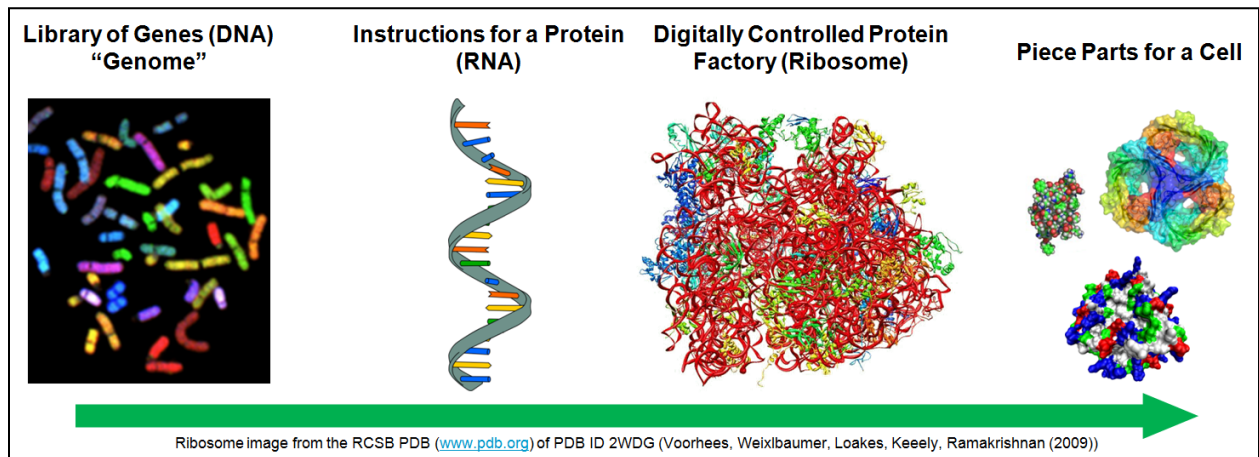
Agilent 101: An Introduction to Microarrays and Genomics

By Jim Hollenhorst

We are in the midst of a revolutionary era in which the mysteries of life are finally being explained. This will lead to breakthroughs in health care, agriculture, alternative energy sources and many other things that will improve the quality of our lives and those of our children.

The life science industry represents Agilent's fastest-growing market opportunity, and microarrays are one of our most prominent technology platforms. But in order to explain microarrays, I first need to talk about the machinery in our cells.

I use the term "machinery" on purpose. Consider a slightly out-of-date, numerically controlled machine tool. You would feed the machine a block of material and a paper tape. The paper tape contained a program – numerical instructions to move both the tool and the block of material. The result was a precisely machined part. To build a complicated product, such as an automobile, you'd need many different piece parts, but each could be made by using the appropriate program from a library of paper tapes.




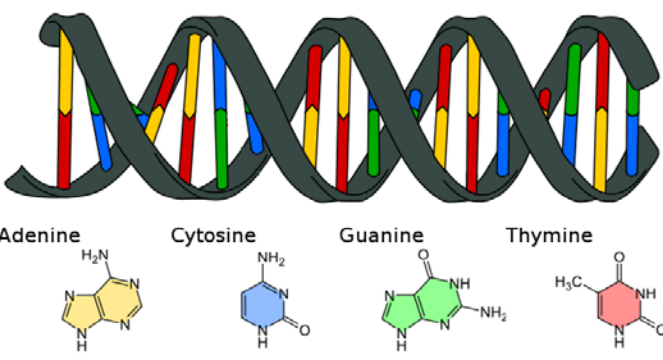
Our cells work in the same way. In the center of each cell is a bag of molecules called the nucleus, which holds all the instructions for making and controlling each structure and activity of the cell. These instructions are encoded in molecules of DNA that are analogous to paper tape. Taken as a whole, we call this DNA our **genome**. The instruction set for a single piece part is called a **gene**.

As I will explain, Agilent microarrays are made using DNA, so we call them **DNA microarrays**, or gene arrays. The measurement and study of this DNA is called **genomics**.

The piece parts of the cell are a class of molecules called proteins. Each cell contains thousands of tiny factories that make proteins. These factories, which work like our numerically controlled machine tool, are called ribosomes. The instructions come from the DNA library in the nucleus.

When a particular gene is needed, the cell doesn't check it out of the library. Instead, it makes a copy. The copy is slightly different from the original; it is a specialized molecule called **messenger RNA** or **mRNA** for short.

The figure below shows how information is encoded on paper tape versus in DNA or RNA. The paper tape stores letters of the alphabet and punctuation, encoded like Morse code as the presence or absence of holes in the paper. The DNA molecule is a chain of almost identical units arranged like a spiral staircase. Each step in the staircase can have one of four different chemical structures. These steps are called **bases** or **nucleotides**.

	
<p>8 holes per symbol (character) Each hole has 2 states (present or absent)</p> <p>$2^8 = 256$ symbols Letters, numbers, and punctuation</p>	<p>3 bases per symbol (codon) Each base has 4 states (A, C, G, T)</p> <p>$4^3 = 64$ symbols Commands to start, stop, or to add a specific amino acid</p>

Each base can be the chemical adenine, guanine, cytosine, or thymine, conveniently represented as **A**, **G**, **C**, or **T**. Taken together, three adjacent bases represent an instruction to add an **amino acid**, or to tell the ribosome to start or stop making a protein. Amino acids are the building blocks of proteins and are assembled by the ribosome according to the instruction specified by the messenger RNA.

The code to specify a human being is a spiral staircase of DNA with about 3 billion steps, less than a gigabyte of digital information. Darwin once talked about life's "endless forms most beautiful." He didn't know that every creature reproduces and thrives by following instructions from the same simple molecules of DNA and RNA.

DNA molecules have two interlocking strands, comprising a so-called **double helix**. The strands are held together by weak bonds between the bases. Importantly, **A** bonds only with **T**, while **G** bonds only with **C**.

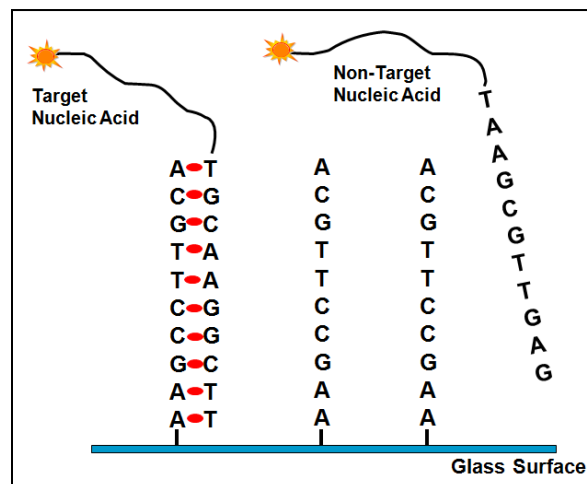
DNA can exist as single strands, but these strands will only stick strongly and form a double helix if the two strands are complementary – if the **A**'s in each strand line up with the **T**'s in the other, and the **G**'s line up with the **C**'s. This complementary bonding gives DNA a number of remarkable properties, including the ability to make RNA copies of genes and to reproduce the entire molecule. It is also the basis for Agilent's genomic measurement products.

From DNA to microarrays

A DNA microarray is simple in concept, but the technology behind it took years to develop and perfect. We take a piece of glass and deposit hundreds of thousands of spots of single-stranded DNA. Each spot contains molecules with a unique sequence of bases, i.e., a different genetic code.

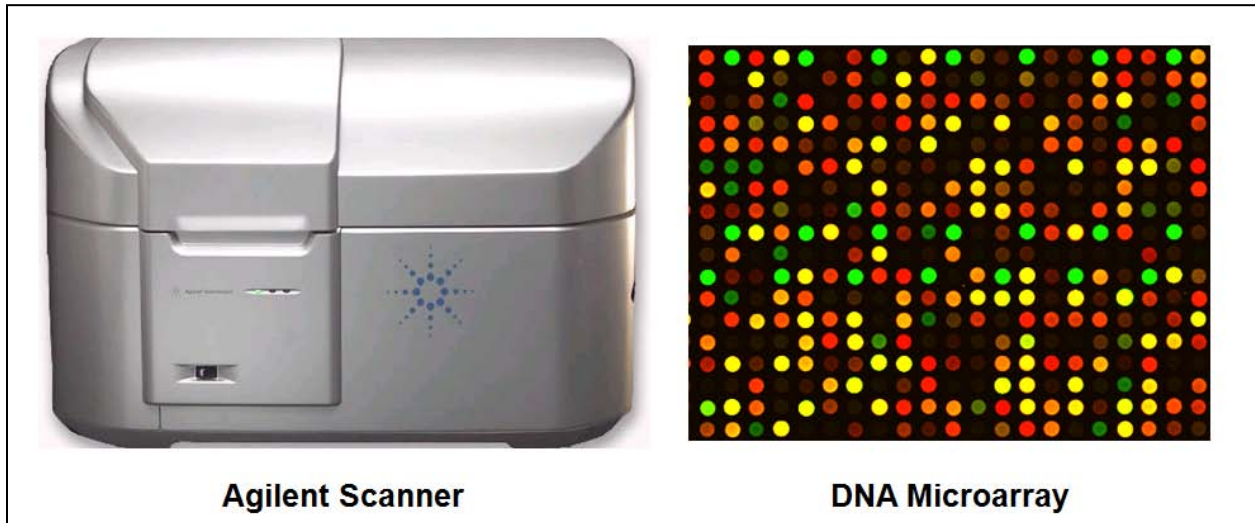
Since the molecules are single-stranded, they are itching to get together with a second strand with the complementary sequence. If you bathe the array with a complex mixture of single-stranded DNA, each spot will glom on to any complementary strands that happen to come by, but it will ignore the other molecules. When you wash away the mixture, any molecule that did not find its soul mate will also be washed away.

By configuring the microarray with the *complement* of the genetic code you're looking for, you can separate and isolate it from all of the other DNA.



So that is really cool, but how do you measure the amount of DNA that stuck to the spot? This is where it gets really colorful. Before you wash your sample across the array, you attach a fluorescent molecule to each DNA strand using a protocol that Agilent designed. Just like those paints that glow in the dark under black light, these fluorescent tags will light up when illuminated by the right kind of laser.

We sell an instrument called a **microarray scanner** that scans a laser across the microarray and produces an image of each spot. The brightness of the spot tells us how much DNA was in the sample. If the spot is black, we know nothing stuck to it. If it is bright, we know there was a lot of the complementary sequence in the sample. We can even use two different-colored fluorescent tags and do two experiments at the same time on one microarray.



How do we put all those different sequences on one chip? Well, Agilent came out of Hewlett-Packard, right? So we use ink-jet printing, of course.

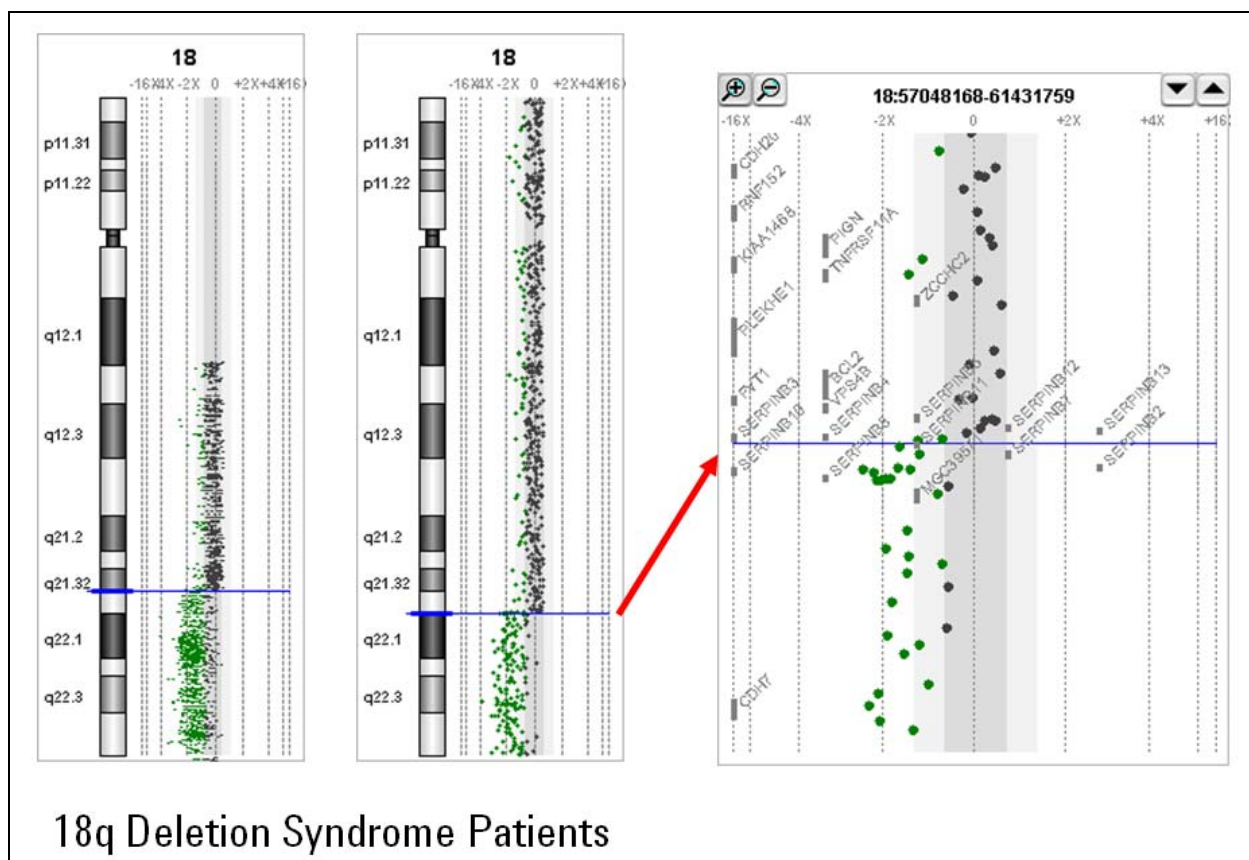
Think of **A**, **T**, **G**, and **C** as different colors of ink. We build up the molecules one base at a time, printing first the **A**'s, then the **T**'s, **G**'s and **C**'s, so each spot gets printed with the correct base. Then we go through the process over again.

With each four passes of the ink-jet head, we add one more base to the molecule. We repeat this between 60 and 200 times, depending on how long we want the final strand to be. Of course, we don't use the kind of ink-jet heads that you can buy at the computer store; our printers are specially designed to make DNA molecules.

Microarray applications

We now know that there's wide variability in the genomes of individuals, in the form of missing or duplicated genes. The extent of this was underappreciated before the advent of microarrays. Agilent pioneered the use of microarrays to measure these so-called **copy number variations (CNVs)**.

Let's imagine that you are a medical researcher. You are studying a rare congenital disease called "18q deletion syndrome." Human DNA is stored on 23 pairs of chromosomes. In this disease, a portion of chromosome 18 that should contain hundreds of genes is missing. This leads to mental retardation, limb abnormalities and many other symptoms. But there are large variations in the symptoms seen in different children. You want to find out why.



Fortunately, you're using Agilent's microarrays. It turns out that the position of the break in chromosome 18 is variable. With Agilent's CNV application, you can measure – with great precision – where the break occurred and how much of the chromosome is missing. You can now correlate the location of the break with the variable symptoms in your patients.

Other scientists are using Agilent microarrays to study cancer. Since the chromosomes in cancer cells are abnormal – with many pieces missing or duplicated – CNV measurements play a key role in understanding and diagnosing the pathology of these cells.

Remarkably, the same technology that can measure DNA can also measure RNA, since RNA also binds to single-stranded DNA. This opens up a wealth of additional studies. I will mention just one, called **gene expression** analysis.

Remember that to make a protein, you need a gene in the form of messenger RNA. If we measure the amount of messenger RNA in a sample, we can find out which proteins the cell is trying to make, and in what quantity. With microarrays, scientists have learned that certain disease states, environmental conditions or developmental stages lead to wide variation in the amount of messenger RNA for each gene within the cell.

In addition to being used for research, our gene expression arrays are also being used in medical practice. A partner of Agilent has developed the first of a new class of diagnostics to be approved by the FDA. This gene expression measurement can predict the severity of breast cancer and enable doctors to optimize the treatment of the disease.

Molecules gone mad

Another important innovation of the past few decades is the **polymerase chain reaction**. PCR is a method that enables you to take a tiny amount of DNA and make as many copies as you like.

You start with a piece of double-stranded DNA – which is the normal form – and heat it up. The heat causes the two strands to separate. Then you cool it down. If you have the right enzymes and nucleotides in the beaker, each strand will actually regrow the missing complementary strand of DNA. There are now two pieces of double-stranded DNA identical to the one you started with.

If you cycle it again, the original molecule separates into four halves, each of which will regrow to yield four double strands. As you can see, with each stage of thermal cycling you get twice as many molecules. Thirty cycles gets you a billion for each one you started with. Forty cycles gets you a trillion.

It's amazing what you can do thanks to Kary Mullis, who dreamed up this idea while driving from San Francisco to Mendocino. He won the Nobel Prize for that little trip up Route 128.

It's a tad more complicated than I've just described, but that's essentially the mechanism. PCR is now a workhorse tool of the biotech industry. One of its many applications is to amplify the DNA in a nasal swab, to get enough molecules to determine what bacteria you may be infected with.

Agilent now sells enzymes, reagents and thermal cycling tools for PCR. These became Agilent products when we acquired **Stratagene** in 2007. With our acquisition of **Velocity11**, also in 2007, we make it possible for customers to automate their PCR and microarray experiments for higher throughput and better reproducibility. Our automation solutions are flexible enough to do all kinds of complicated manipulations to enable measurements in molecular biology.

As you can see, our tools enable researchers to do some pretty amazing things to help scientists increase our understanding of living things. Agilent will continue to increase its life science capabilities through R&D as well as through acquisitions. In this rapidly growing field there are enormous opportunities to change the world for the better. As an Agilent employee, I'm excited to be a part of this grand endeavor.



Jim Hollenhorst is senior director of technology for Agilent. He has a B.S. in physics from the University of Minnesota, as well as an M.S. and Ph.D. in physics from Stanford University.

March 2011