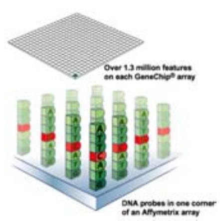
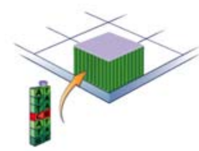
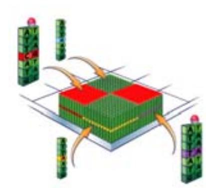
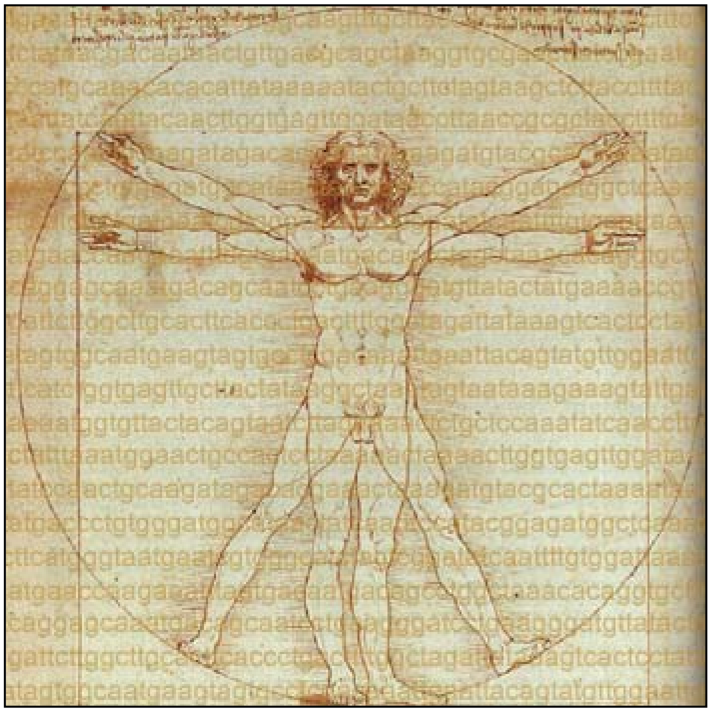
How Affymetrix GeneChip® DNA Microarrays Work







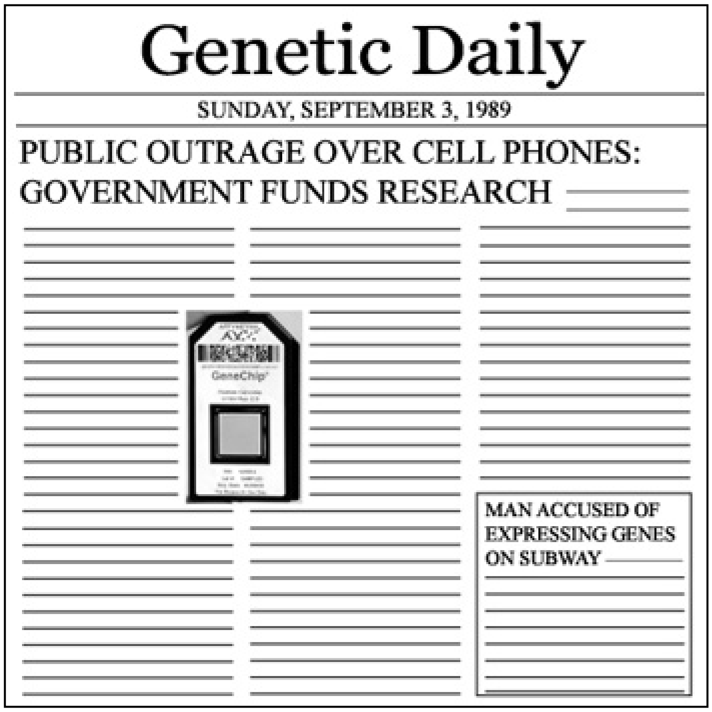


Genotyping DNA

The Human Genome Project documented our genetic sequence and discovered everyone's DNA to be 99.9% identical. However, even very small differences in DNA sequence can have very big effect on health and disease.

The same gene may be working properly in one person, but a small mutation could cause it not to work at all in another.

To find those disease-causing mutations, researchers use microarrays to genotype a patients' DNA and determine the exact sequence – A, T, C, or G – for thousands and thousands of single nucleotide polymorphisms (SNPs) distributed throughout the genome.



Understanding Disease

For simplicity sake, let's use a hypothetical disease to illustrate a common way that SNP genotyping arrays could be used. Suppose that the public outcry over rude cell phone behavior – loud, inane, and inappropriate conversations in public places – reaches such a pitch that the government decides to fund investigative research looking for a treatment. Scientists hope to find some sort of genetic basis contributing to the behavioral phone disorder and to then develop a drug that can treat it.

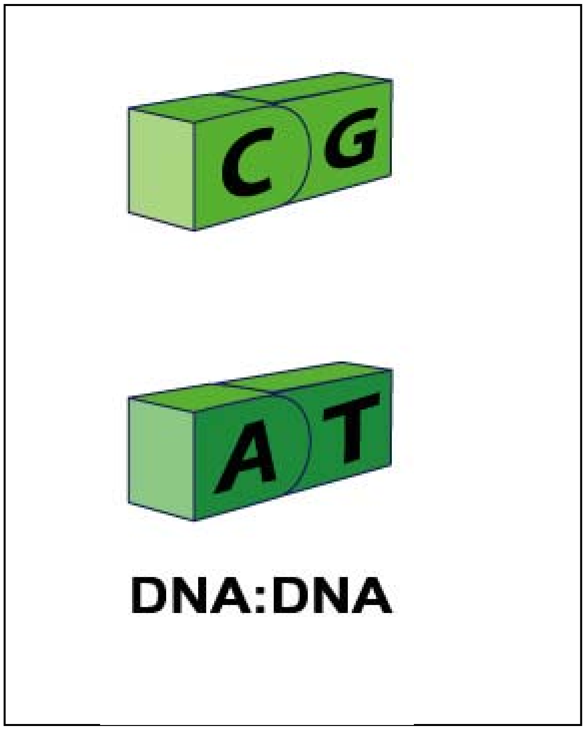
Genes and Disease

To find a treatment, researchers first have to find a cause, a mutation or malfunction in one or more genes. The cause might be a genetic change that inhibits the normal functioning of the "polite" and "common sense" regions of the brain. But no previous research has been done on rude cell phone behavior, so scientists have very little idea of what genes may be malfunctioning to cause the rude behavior. The mutation they are looking for could literally be any one of 3.1 billion base pairs in the human genome. How do they know where to start?

Finding a disease gene

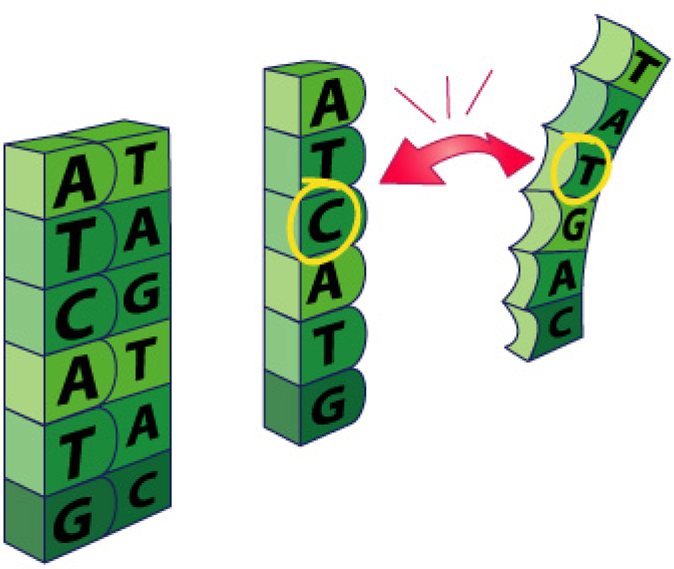
Before the advent of microarrays, scientists would have searched previous academic research for genetic links to other rude behavior, like talking in movies or standing in the express checkout lane with 25 items. They would have tried to form a hypothesis based on any previous research and tried to find a link to cell phone talkers.

Since the advent of the microarray, researchers can simply go to the genome and look directly at the DNA and see what that tells them. Researchers use microarrays to scan the whole genome and look for genetic similarities among a group of people who share the disease. Using microarrays to genotype 10,000 or even 100,000 SNPs, researchers can pinpoint the gene, or group of genes, that contribute to disease. For instance, if 500 people with the same disease all share a half dozen SNPs in common, but a group of 500 healthy people don't share those SNPs, researchers will start looking for mutations behind the disease around those SNPs. The SNPs don't give you the exact mutation, but they do tell you where in the genome to look for the disease-causing mutation. This approach is only possible through the use of high-density microarrays.



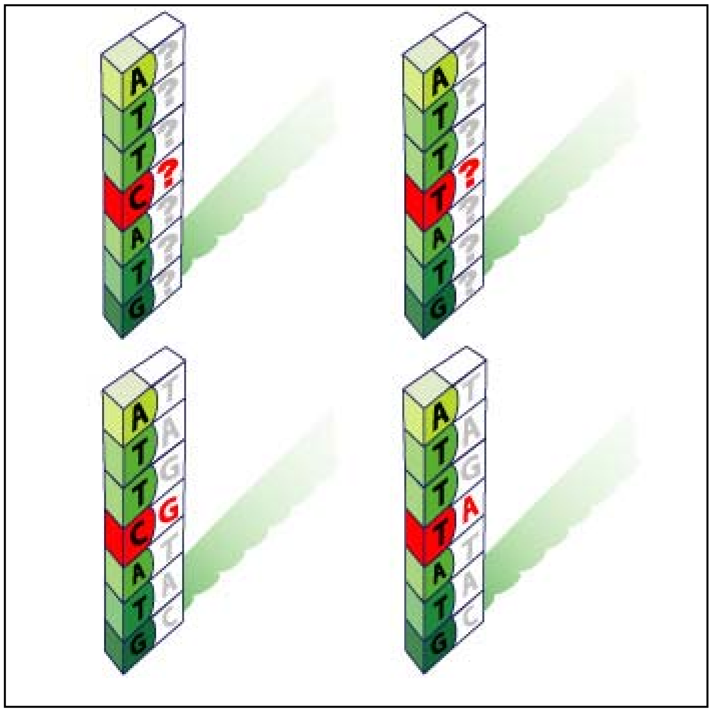
How Affymetrix genotyping microarrays work

All GeneChip microarrays take advantage of the natural chemical attraction between DNA molecules. There are only four molecules, or "bases", in every DNA chain: adenine (A), guanine (G), thymine (T) and cytosine (C). These four molecules partner: C partners with G and T partners with A. Pairing is a natural state for DNA and if you pulled the double helix apart, it would inevitably move back together, like two long chains of magnets that are attracted to each other.



A good match sticks, a bad match doesn't

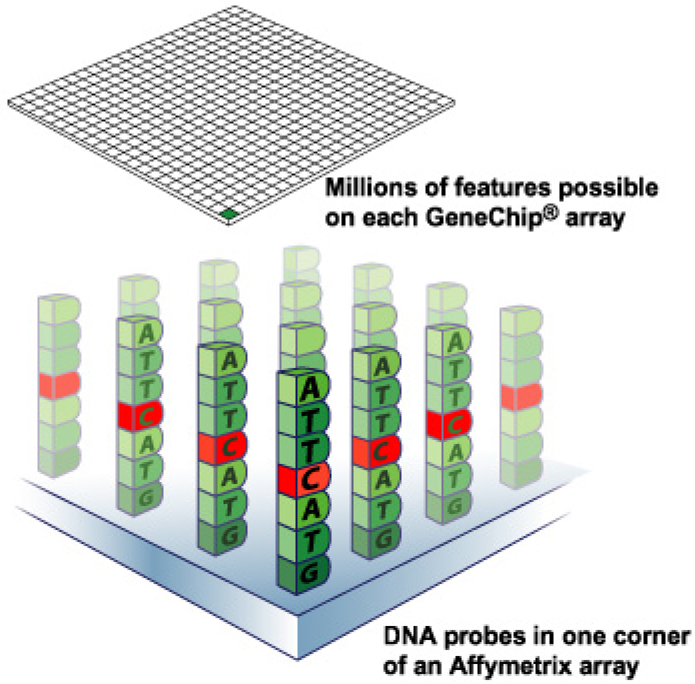
When a single strand of DNA (**ATCATG**) matches a complementary strand of RNA (**TAGTAC**), the two strands are "complementary" and will stick to each other. However, if the bases aren't complementary, they won't fit together. An A won't pair with another A or with a C or a G. Even a single base that doesn't match its partner (like a **C** and a **T** *as shown in the adjacent diagram*) could keep one single strand from sticking to another single strand.



The Basic Principle
Affymetrix microarrays use this base pairing attraction – a property known as hybridization – to help researchers identify each SNP genotype. The sequencing work of the Human Genome Project told us the DNA sequence around any SNP we'd like to genotype. Microarrays use that information to determine SNP genotype. Because C always pairs with G, and T always pairs with A, if you know that one side of the chain is **ATTCATG** (the SNP is in the middle), you don't even need to see the other side — you know it's **TAAGTAC**. The SNP genotype would be C/G. Likewise, if you know that sequence of one strand is **ATTTATG**, you automatically know the sequence of the other side is **TAAATAC**. And this DNA would have a T/A SNP genotype.

So, if you're not sure what the SNP genotype is, all you have to do is look at both strands. If a person's DNA matches **ATTCATG** then you know they are a C/G genotype. If a person's DNA matches the **ATTTATG** strand then you know they are the T/A genotype.

But what's amazing about Affymetrix DNA analysis SNP arrays is that they don't just do it one SNP at a time. The arrays tell you whether a person has a single SNP, they tell you whether they have 100,000 SNPs or more. To understand how, let's look at how an actual Affymetrix DNA analysis array works.



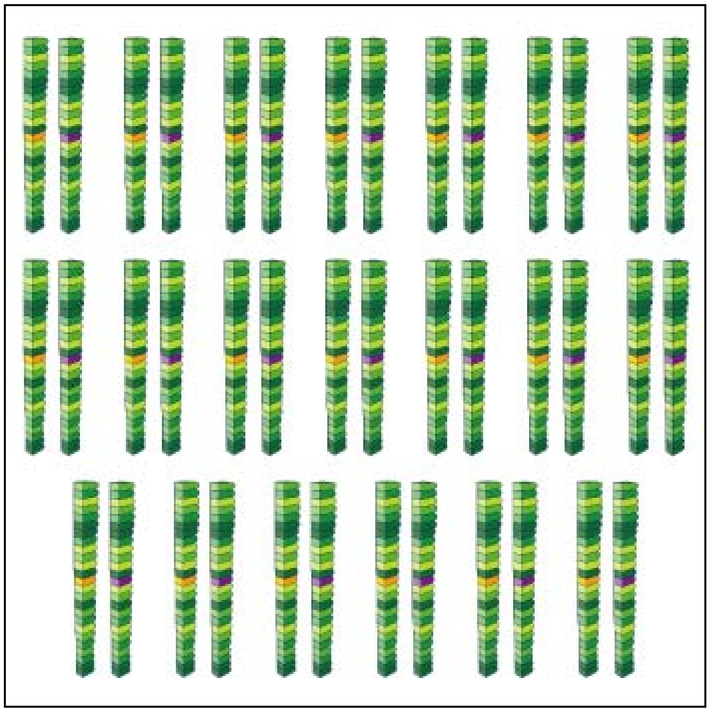
Features and Probes:

The surface of the Affymetrix array is like a giant checkerboard. The 10K array, which is a piece of glass about the size of a thumbnail, has over 400,000 squares. These squares are called features. The features on Affymetrix arrays are incredibly small – about 8 microns across. By way of comparison, a human hair is about 50 microns wide.

Each of these squares holds one unique type of DNA strand – called a probe. So, in our example, one square would hold the **ATTCATG** probe we built, and another square would hold the **ATTTATG** probe. But, while each square holds one *type* of probe, there isn't just one probe in each square, but millions of identical copies of the same probe.

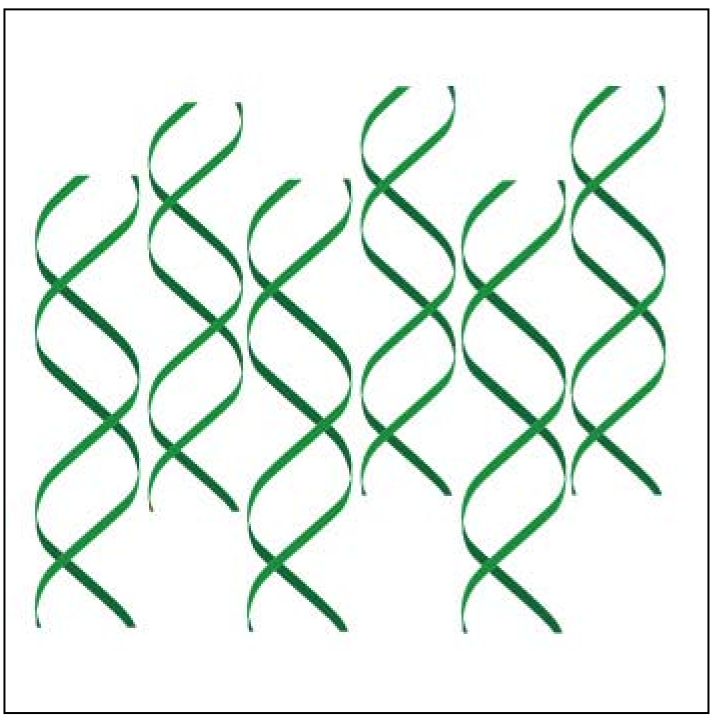
Affymetrix builds these probes one layer at a time, using the same type of manufacturing technology that is used to build computer semiconductors. The molecules are built one layer at a time, one stacked on top of another, like bricks. Multiple probes are synthesized in parallel.

For our example, we are just going to look at one corner, or a single feature, of an imaginary array. Normally, each probe is 25 bases long, but for purposes of illustration here, let's abbreviate this standard probe length and say the probe is only a 7 base sequence, **ATTCATG**. The SNP is in the middle position and is highlighted in red. Somewhere else on the array, there is a probe representing the other possible SNP genotype, **ATTTATG**.



More information

Quick math will tell you that if we are looking for 10,000 SNPs and there are over 400,000 features on an array, that there are a lot more features than SNPs. That's because Affymetrix genotyping arrays essentially ask the same question – which SNP is at this particular location – in 40 different ways to make certain that the answer they get is correct. Each time we ask the question, we use a different probe, meaning that there are 40 different probe sequences used to decide the genotype of every SNP on the array. For a microarray that genotypes 10,000 or even 100,000 SNPs, you can quickly see how that adds up to millions and millions of probes.

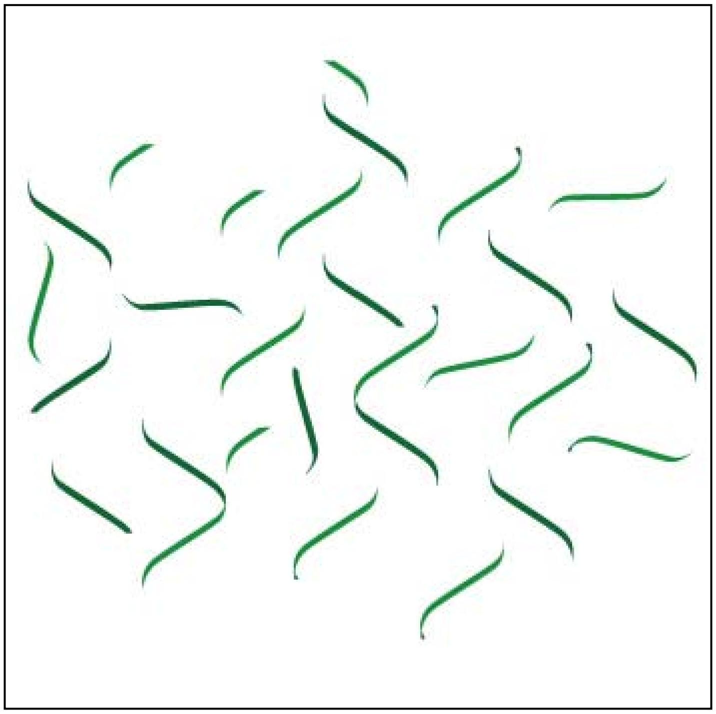


Extract DNA

Now that we have a probe designed to genotype a SNP in a DNA sample, we have to extract the DNA from our subject. DNA can be taken from any biological sample such as blood or saliva. Unlike gene expression analysis, where different cells in the body have different amounts of RNA, a persons DNA sequence is the same no matter what cell you extract it from.

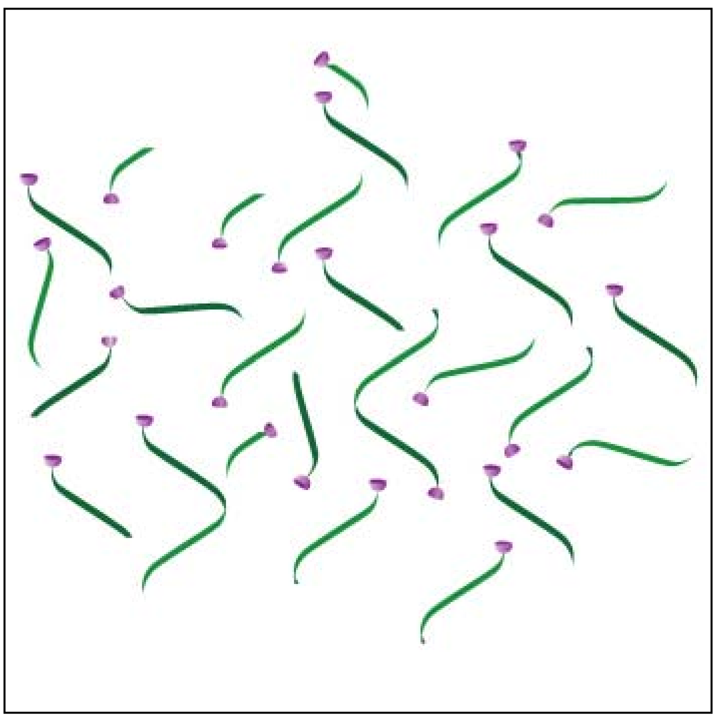
Getting the DNA ready to be genotyped

Before you can analyze the SNPs from the DNA you extracted, you need to make millions of copies of each piece of DNA containing a SNP. Copying the sequence containing the SNP allows the DNA to be more easily detected on the array and the SNP to be more easily genotyped. Previously, this required performing a separate experiment for each SNP, so to genotype 100,000 SNPs, you would first need to perform 100,000 experiments to get the DNA ready. Affymetrix microarrays use a newly developed method that allows you to perform only one experiment and have the DNA ready to go for SNP analysis. With one experiment you can analyze 10,000 SNPs, or, you could scale this up, performing just two experiments to genotype over 100,000 SNPs.



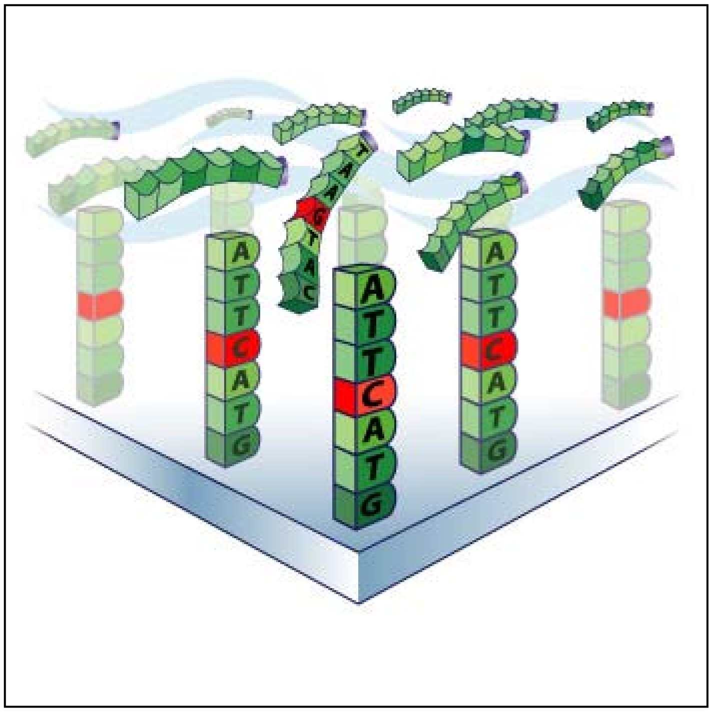
Chop up the DNA

Once we've made copies of all the SNPs that will be analyzed on the microarray, we use another chemical process to fragment the DNA chains up into millions of short pieces.



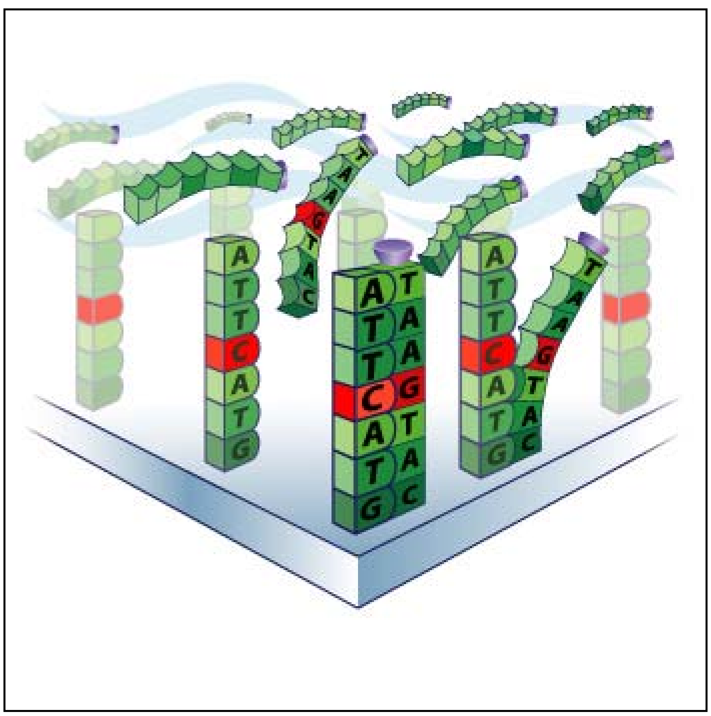
Making Sticky DNA

After the DNA is chopped up, a chemical called biotin (purple cups) are attached to each strand. These biotin molecules will act as a molecular glue for fluorescent molecules that will later be washed over the array. When researchers eventually shine a laser on the array, the fluorescent molecules will glow, showing where the sample RNA has stuck to the DNA probes on the array.



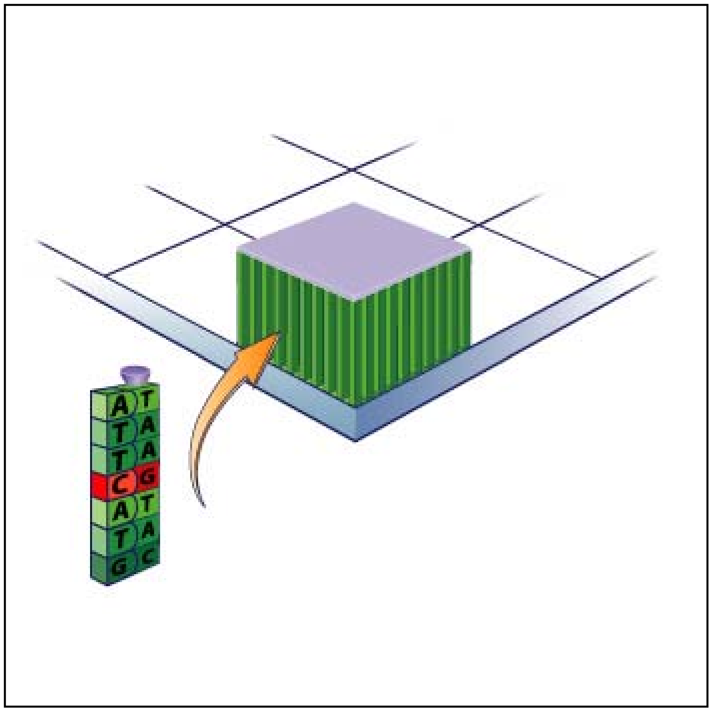
Wash Sample Over the Array

All of the prepared DNA sample is washed over the array for 14 to 16 hours. The number of molecules involved in this wash is staggering. There are millions of copies of each DNA probe in every square on the chip, and there are also millions upon millions of pieces of tagged DNA from the sample. It's like the world's largest molecular singles bar. All of the DNA strands from expressed genes are swimming around, looking for their perfect complement—molecular true love on a microarray. Most of them will not find a match.



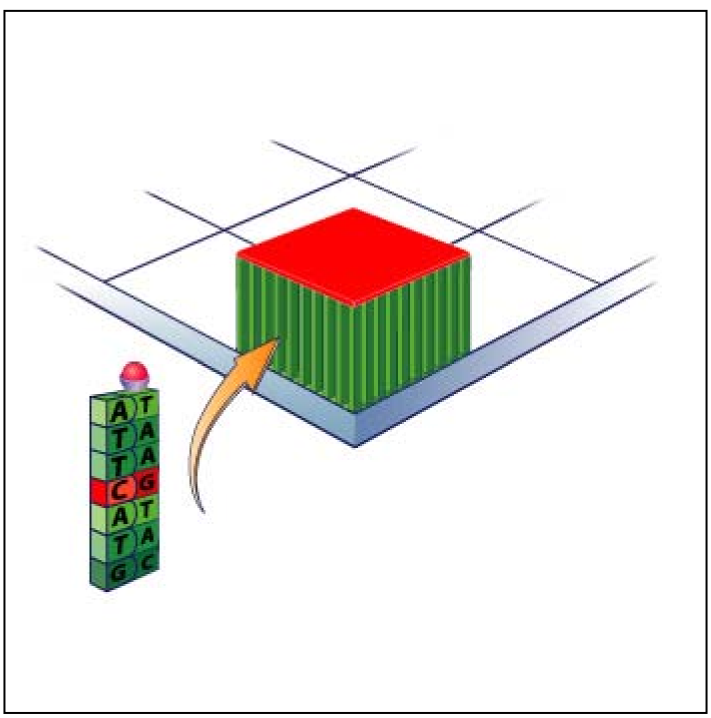
A Committed Relationship

But somewhere, in the tagged sample of DNA washing over the array, a match will be made. If the sequence of bases in the sample DNA matches that of a DNA probe, then there will be a perfect match and the sample will stick to the probe.



Determining a Match

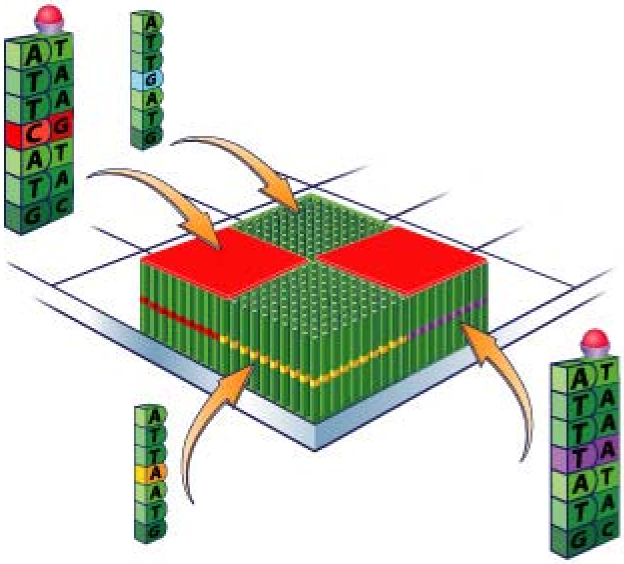
Let's assume that we have a match and that the SNP genotype in the sample DNA matched the probes built on the array. We then rinse the array, so that any DNA that didn't pair is washed away. Because the hybridized RNA is tagged with molecular glue (biotin), it’s as if each hybridized square on the array has been coated with sticky glue.



A Genotype is Determined

Researchers wash the fluorescent stain over the array, and the glow in the dark molecules (red ball) stick to the biotin glue (purple cup). It’s like glitter painting in elementary school -- after pouring sparkle glitter all over the paper, you shake it off and the glitter only sticks to the places where there is glue. With microarrays, the fluorescent molecules are “shaken” away and the stain only sticks to those places on the array where DNA has bound. After all of this, researchers shine a laser light on the array, causing the stain to fluoresce or “glow”.

In this example, the DNA in the sample matched the **ATTCATG** probe we built on the array, so we now know that the person has a **C/G** genotype for that SNP.



Measuring Multiple SNP Genotypes at Once

So far we've been looking at just one possible genotype – C/G -- for the SNP. But, it's also possible a person could have the A/T genotype, or even both. That's because humans inherit genes from their mother and father. So you may have inherited the C/G genotype from your mom, but also inherited the A/T genotype from your dad. This condition is referred to as heterozygous. It means that you have a mix of genotypes. If you had inherited the same SNP genotype from both parents, than you would be homozygous for that SNP. So, in this example, it's possible to be homozygous for the C/G genotype, homozygous for the A/T genotype, or be heterozygous containing both the C/G and A/T.

Diseases usually develop when a person is homozygous for a SNP genotype. When that happens, it means that you inherited the same faulty chunk of DNA from both your mom and your dad, hence the SNP genotypes are the same.