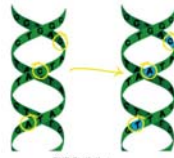
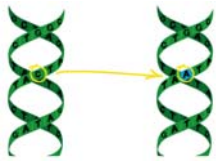




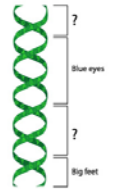
How SNPs Help Researchers Find The Genetic Causes of Disease



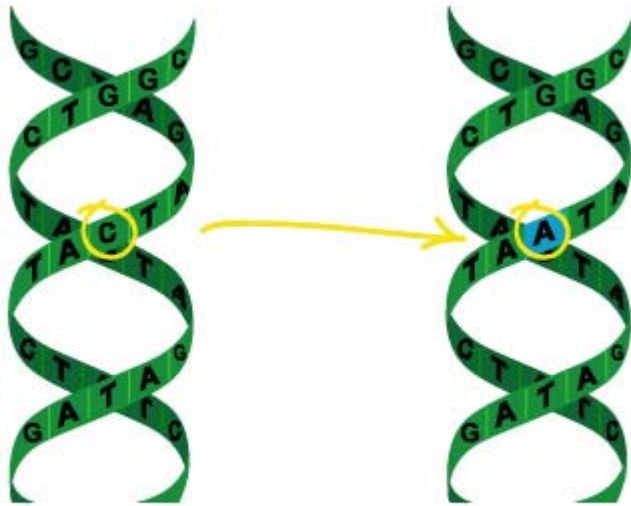
SNP Statistics
-Estimated SNPs in human genome: 10 million
-Number that have been seen twice: about two million
-Seen more than twice: about XX thousand



SNPs (red squares) tell scientists where to begin looking for the autism mutation. By sequencing the DNA between those SNPs, scientist can find the actual mutation (yellow star).



DNA is inherited in blocks that contain multiple traits



Single Nucleotide Polymorphisms (SNPs) are single base differences in otherwise identical DNA sequences.

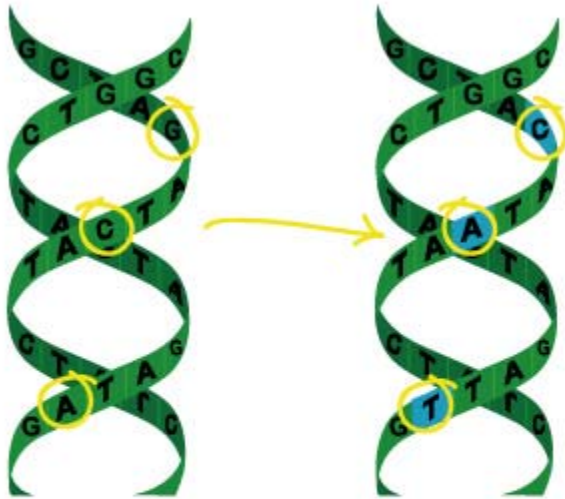
SNP Essentials

One of the findings of the Human Genome Project is that the DNA of any two people, all 3.1 billion molecules of it, is more than 99.9 percent identical, but that 0.1 percent accounts for all the genetic differences between people. In literal terms, that means that one person might have blue eyes rather than green, or a susceptibility to lung cancer, or perfect pitch, because the sequence of their DNA -- a long chain of adenine (A), guanine (G), cytosine (C) and thymine (T) molecules -- differs from another person's. Rather than having an A-T pair of molecules at a certain spot on the DNA chain, a person might have a G-C pair. On the other hand, that difference might not have any effect at all on a person's health or appearance.

These differences in DNA sequence are called single nucleotide polymorphisms, or SNPs.

The same SNP story

SNPs do not occur randomly. There isn't an equal chance that any one of the 3.1 billion base pairs in your genome will be different from someone else's. SNPs are mutations that occurred once in history and then were passed on to future generations. So if your ancestor developed a SNP 5,000 years ago, you, along with many of your other very distant relatives would inherit that SNP, but those not descended from that ancestor would lack it. Perhaps in 15 percent of the population the 1,253,334,078th base pair along the genome at the very end of Chromosome 16 is a T-A, not a C-G like it is in the other 85 percent of the population. And most SNPs that we care about are like this, they are common, to a greater or lesser degree, throughout large parts of the population. This makes sense, since very few attributes, like eye color or a disease, occur only in one person.



SNP Statistics

- Estimated SNPs in human genome: 10 million
- Number that have been seen twice: about two million

How many SNPs are there?

SNP research, like the rest of genome research, is definitely a work in progress. No one knows how many SNPs there are, but some people estimate that there could be as many as 10 million. When scientists are sequencing DNA in drug or disease research and they see a discrepancy in the sequence between people, they will record that in a public SNP database. Right now there are about two million entries like that in public databases. There are far fewer well-annotated SNPs; those are SNPs that have been seen at least twice by researchers.



How do SNPs help disease researchers?

Finding DNA mutations in genes that cause or contribute to a disease is one of the most challenging tasks for a researcher, because the mutation could be anywhere in the 3.1 billion A, C, T and G molecules that make up our genome. It's like looking for a needle in a haystack, and scientist often don't even know where to begin looking. SNP analysis tells them what section of the genetic haystack to start looking in, and this allows them to find the disease-causing gene much more quickly.



Over 3.1 Billion Molecules

To understand how invaluable SNPs are in tracking down mutations that cause disease, you have to appreciate the immense size of genome.

Consider this: if each of the DNA molecules in our genome was about the size of a ping pong ball, the long unraveled chain of molecules would circle the earth 3 times, or just over 75,000 miles. The real difficulty is that less than 2 percent of that -- about 1500 miles, or a little less than the distance from Los Angeles to Chicago -- is DNA that we know codes for proteins. These protein-coding areas are what has traditionally been referred to as "genes." But those 1500 miles worth of "genes" isn't all in a row. Genes are scattered throughout the genome, and in between

them is the so-called "junk" DNA. Since scientists estimate that genes are on average about 600 base pairs long, a gene on our global ping pong scale would be 24 meters (80 feet) long. Given a genome that wraps around the world three times, 24 meters is miniscule. If you were walking – or swimming – the entire trip, you'd be likely to encounter a gene an average of once every 2.5 miles (4 kilometers).



Genetic Postal Codes

Because the genome is so immense, it is practically impossible to find a specific gene or disease-causing mutation without having a rough idea of where to begin looking. Searching for disease genes without SNPs would be like searching for an address without a postal code. The address could be anywhere in the US and you would have no clue where to start. But with a postal code, you could narrow your geographic area and then methodically search a local map to find the street. SNP analysis does the same thing, reduces the possibilities so that researchers can better focus their search and find the disease-causing mutation they are looking for within the vastness of the human genome.

How to Find Genes Associated With Disease

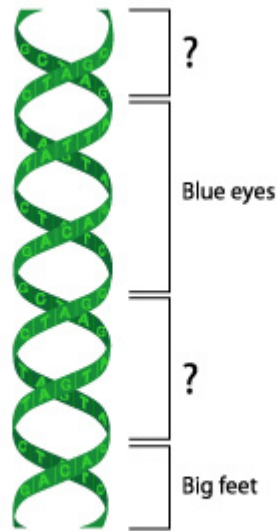
Researchers make the assumption that if 1000 people share the same disease they should also share the genetic mutations that contribute to that disease. If researchers can pinpoint the genetic differences that all these people share – genetic mutations that healthy people don't have – they can understand how these mutations contribute to a disease. By understanding the cause, they can hopefully find a treatment.



A comparison of two genomes with SNP positions shown in blue. Most SNPs are different, but some are common (yellow circle).

Comparing Genomes

In an ideal world, researchers would just sequence the genomes of all 1,000 people – effectively lay them side-by-side and compare the sequence of the As, Cs, Gs and Ts in each person's DNA. That would show them the mutations that people with the disease share and scientists would start their research there. Unfortunately, with current technology, sequencing the 3.1 billion bases in a single human genome is too expensive and time consuming to be practical for disease research – after all, it took the Human Genome Project 10 years to sequence a single human genome. SNPs offer a more practical way to find the genetic differences that cause disease.

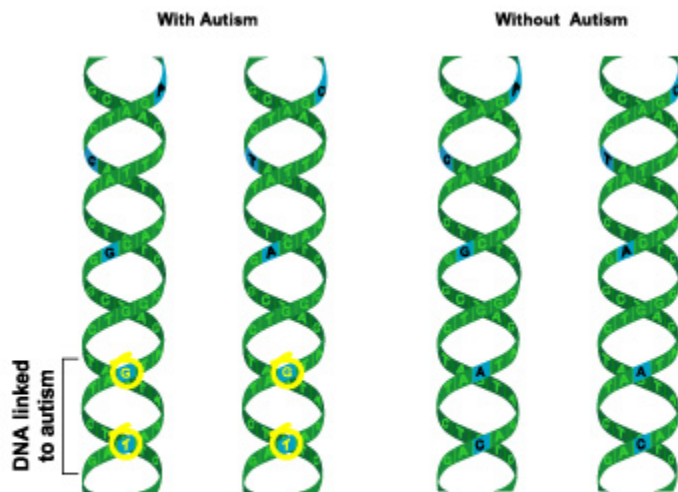


DNA is inherited in blocks that contain multiple traits

DNA Moves in Blocks

To understand how SNPs help scientist locate disease genes, you first need to understand how genes are inherited. When you inherit a trait or disease, you don't just inherit the DNA for that trait. Instead you get a long chunk of DNA that may affect many characteristics. So maybe the piece of DNA from your dad that gave you his big blue eyes, also gave you his big feet. In this hypothetical example, big-blue-eyes DNA and big-shoe-size DNA make up a block of DNA that is always inherited together. You inherited this genetic chunk from your father, he from his father, and so on, all the way back to the original ancestor who first developed this particular trait. So, even in a large mixed population, anyone with this specific chunk of

DNA would be genetically related to each other, because they share a common ancestor – the first big-blue-eyed big foot.

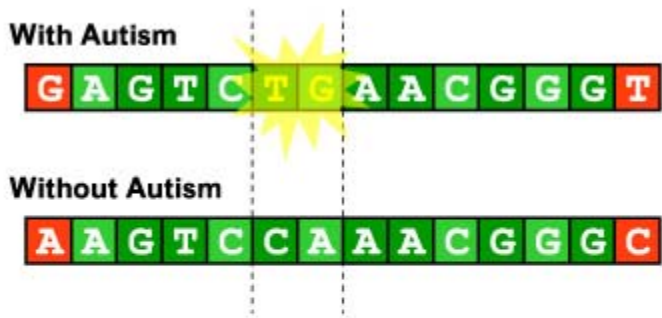


SNP genotypes that are common to people with autism (circled), but are absent from people without, define a block of DNA linked to the disease.

Tracking DNA with SNPs

The fact that we inherit our DNA in these consistent, predictable blocks is key to understanding how SNPs are used to track down a disease-gene. Once a disease-causing mutation occurs in this block of DNA – either by chance or by environmental factors – that mutation is passed on to descendents who inherit that block of DNA generations later. The various SNPs that occur within the block of DNA will also be passed on. So when researchers see a SNP shared by a lot of people who have a disease like autism, (but not shared in a group of people that don't,) they think "These people share a similar block of inherited DNA and there may be a disease causing mutation in that block." In this way, SNPs from an ancestor

who might have lived 5,000 years ago, can serve as a marker for a disease gene you could have inherited today.



SNPs (red squares) tell scientists where to begin looking for the autism mutation. By sequencing the DNA between those SNPs, scientist can find the actual mutation (yellow star).

Finding the Disease Mutation

Scientists' next step is to look for mutations in the DNA surrounding the SNPs that the patients have in common. The Affymetrix 10K Mapping array basically screens the entire human genome for 10,000 SNPs that scientists have discovered. On average, those SNPs are about 20,000 bases apart (an A, C, G or T molecule is called a "base"). In the example we've been using, scientists have found two SNPs, a G and a T, that are shared by people with a disorder like autism. That means out of the whole genome, scientists only have to look at the block of DNA containing those two SNPs to find the autism mutation. The next step would be to find out the exact order of the As, Cs, Gs and Ts on that

block of DNA, which is called "sequencing." Researchers would sequence then that block of DNA from everyone in the study and then do a base-by-base comparison to try and find other mutations that people have in common, mutations that might be contributing to the autism.

Does that mean the marker SNP is responsible for the disease?

It's possible, but it would be quite a stroke of luck. The SNP is a mutation and could be part of the problem, but scientists think that most SNPs have no effect at all. For a researcher, a SNP's primary function is to serve as a marker, or a sort of sign post along the genome that says to the researcher: "Out of the 3.1 billion base pairs in the human genome that could have mutations that cause this disease, you might start looking here, around this SNP which everyone with the disease shares." SNPs are not the only types of mutations either. Deletions and duplications of DNA can also cause disease, but by analyzing SNPs, scientist have a way of finding any kind of mutation linked to disease.

So is any single base mutation a SNP?

By definition, any single base pair that is different from the reference sequence drafted by the Human Genome Project is a SNP. But if, say, only five people in the world share the same SNP, it's not going to be much good to researchers that are trying to find genes associated with diseases. If you have a list of 10,000 SNPs and you want to see if a group of 100 people with colon cancer share any of them, you probably won't get many matches if the SNPs you have only appeared in a handful of people in the population. You want the popular SNPs, the ones that show up a lot. Remember, the SNP's purpose is to point you to a block of DNA that people in the disease group share, it may not have anything to do with the disease.

Why are some SNPs rare?

If a SNP mutation has happened recently, not much time has elapsed to allow it to be transmitted and inherited by a large number of people. This kind of SNP is a rare SNP. On the other hand, if we are looking at a SNP mutation that happened 25,000 years ago, there's a much greater chance for that SNP to have been inherited by a lot more people. Scientists say that these types of SNPs are "common".