



## Molecular Genetics

Michio Hirano, MD

December 10, 2005

### Chapter 1: Introduction

**Michio Hirano:** In the first talk, I'll talk about classical and molecular genetics and give you some background in it. Dr DiMauro will talk to you more about mitochondrial genetics, which is very different. So hopefully these two talks will be complementary.

I'll start by giving you some background about Mendelian genetics, which most of you are familiar with, and I'll talk about some work that was done by Morgan-Hughes in the early part of the 1900s here at Columbia. And then I'll talk about molecular genetics and how that's applied to identify the causes of monogenic disorders—diseases due to mutations in a single gene—and how this has had an impact on clinical neurology; and then finally I'll talk a little bit about complex traits—that is, traits that are due to either multiple genes or a combination of genetic factors and environmental factors.

### Chapter 2: Medelian Genetics

Okay, so genetics really began with this fellow, Gregor Mendel, in the 1800s, who was an Augustinian monk in Brno, which is now in the Czech Republic. He studied garden peas in the monastery, and here's a picture of the monastery and this is the little garden where he worked. And he started out trying to become a schoolteacher and he failed his examination to become a science teacher; so he went back to the university, learned how to do research, and applied that to the



garden peas. Over a period of about eight years he studied these peas. His work culminated in this publication, which was published in 1866 in the *Society for the Study of the Natural Sciences*. In this study he reported that there are hereditary traits that exist in alternative forms. So, for example, in his peas he saw that one hereditary trait would produce a smooth pea, and another hereditary trait would produce a wrinkled pea. So he crossed these peas and he came up with this simple algebra for heredity, which I'm sure all of you are familiar with. When he crossed the smooth peas with wrinkled peas he produced hybrid offspring—they were all smooth. And from that he deduced that the smooth trait was dominant over the wrinkled trait. He then took those hybrid peas, they're called *F1* for filial 1, and he crossed them with each other and he produced *F2* (filial 2) generation peas, and he found that three-quarters of these peas were smooth and one-quarter were wrinkled—so the wrinkled trait reappeared.

How could that occur? Well, as I said you have these hereditary traits, *S* for smooth and *W* for wrinkled, and when you cross smooth peas with wrinkled peas, you end up with these *F1* peas that are hybrids of smooth and wrinkled; but because the smooth trait is dominant, you see them as smooth peas. But when you cross these *F1* hybrid peas with each other, you get some peas that are going to be smooth-smooth, will carry smooth and wrinkled traits, but again they'll be smooth because the smooth trait is dominant; and then a quarter of the peas will have the wrinkled-wrinkled traits and will appear wrinkled.

From these experiments he deduced two laws of genetics. The first is called the law of segregation, that there are pair units of heredity—now they're called alleles—that segregate from each other into gametes; and then the law of independent assortment—that is, these hereditary traits assert independently of each other. So the traits for the smooth and wrinkled pea is independently inherited from the color of the peas, like green or yellow peas.

Here are some examples of Mendelian inheritance. And I'd like the audience to participate at this point and try to tell me what kind of inheritance this might be. All of the affected people in this family are marked in red, so there are 3 generations, and there are affected people in each generation. Autosomal dominant,



right, because you'll see this person who's affected has 4 children and half of them are affected. And again this male person has 4 children, again 2 are affected. So there's a dominant trait that's being passed. When there's male-to-male transmission, it's almost always dominant inheritance, so if you see male-to-male transmission that's a clue. Okay, so there's the answer, autosomal dominant.

Here's another example. This is a family in which 2 out of 8 siblings are affected. And you'll see that the parents here are consanguineous—the double line means that they're related to each other because they have common grandparents. So what is the inheritance pattern here? Autosomal recessive, right. And obviously in consanguineous families, autosomal recessive traits are expressed because they carry the same mutations. You could see here I've marked the people who carry the mutations with little red dots. So you can see that this grandmother here of the parents had the trait and passed it on to her children, who then passed it on to the parents of the affected people. And you'll see that half of these people in this generation carry the trait and one-quarter have the disease.

### Chapter 3: Non-Mendelian Inheritance

Now there are non-Mendelian forms of inheritance, and here's an example. This is not autosomal dominant or recessive, so what kind of inheritance pattern might this be? There are 4 generations affected. Someone said mitochondrial—absolutely. This is an example of maternal inheritance. You'll see that both males and females are affected, but only females transmit the disease. Here's an example of a man who has the disease but doesn't transmit it to his children. So this is maternal inheritance, or mitochondrial inheritance. And Dr DiMauro will talk in more depth about that in the next hour.

And how about this one? Right, X-linked, and specifically X-linked recessive, because there are only males affected in this family. You'll see that females who presumably carry this and transmit it to the males—the females here do not have the disease. So only males have it because the mutation is on the X chromosome, and women have 2 X chromosomes obviously, and therefore they have a wild-type allele and a mutant allele, and the wild-type allele protects them. Here



you can see that the women here carry the mutation and pass it on to their children. And here's an instance of a woman who passes it on to both an affected son, and the mutation's passed on, to an unaffected daughter.

Gregor Mendel's work was largely forgotten for several decades, but it was rediscovered in 1900 by three different biologists, who reproduced experiments that confirmed Mendel's law of segregation. It was quite remarkable that his very important work was ignored for many decades, but fortunately it was rediscovered. And that set the stage for this gentleman, Thomas Hunt Morgan, who worked down at the main campus of Columbia University in Schermerhorn Hall. That building still exists. And he had what's known as the Fly Room in that building, and he studied *Drosophila*. He made many important contributions. He and his students were absolutely phenomenal in what they studied and what they found. They discovered the chromosomal theory of inheritance: that genes are located in chromosomes, that each gene is located in a particular chromosome, and they discovered sex-linked inheritance that I showed you earlier.

An important concept that they came up with is that genes are located in chromosomes and that they're linked together on those chromosomes. But there are chromosomal recombinations that occur during the formation of the egg and the sperm that's called meiosis. During meiosis you get recombination of chromosomes in the cells, and that leads to separation of linked genes. So genes that are located physically far apart on the chromosome are more likely to be separated, and those that are close together are more tightly linked. This is a very important concept that you need to understand so you can understand how linkage studies are done. Is that clear? Okay.

In the 1940s, subsequently Hunt Morgan's pupils, Beadle and Tatum, came up with the idea that each gene encodes one enzyme. Actually it's become a little bit complicated than that—genes encode not just enzymes but other products, such as ribosomal RNAs, transfer RNAs, but in general genes encode proteins or enzymes or structural RNA molecules. In 1953, obviously the important discovery of Watson and Crick was reported in *Nature*. They described the double helix structure of DNA, and this really marked the beginning of molecular genet-



ics. By understanding the structure of DNA, we now can understand a lot more about how it's inherited, how mutations occur, and so on.

#### **Chapter 4: Key Concepts and Terms of Molecular Genetics**

So I need to define a few terms for you. First is the concept of a gene. A gene is a chromosomal segment that encodes a single functional polypeptide or sometimes enzymes or an RNA molecule. A gene is a part of the chromosome that encodes a functional product. And genes are encoded in DNA—as you know, deoxyribonucleic acid—and RNA is a related molecule, it's ribonucleic acid and codons. A triplet of nucleotides in DNA encodes a single amino acid, or terminal signal, a stop code to end the polypeptide after it's produced. For example, the triplet of nucleotides in DNA, called CAG, represents the amino glutamine.

The central dogma of genetics is that DNA, which contains all of this genetic information, is transcribed into messenger RNA, which in turn is translated into protein products. Sometimes the DNA is just transcribed into RNA, which has a function on its own, and it's not translated into proteins. But by and large, most of the DNA genes are transcribed and then translated into proteins.

Another couple of terms I need to define for you. An allele is one of several alternative forms of DNA at a chromosomal locus. As I described for you before in the peas, you can have an allele that will produce a green pea, and a different allele will produce a yellow pea, or a wrinkled pea, or a smooth pea. All of these traits are encoded in DNA and each trait would be considered encoded in an allele. And a clone is a large number of molecules or cells that are identical, that are derived from a single precursor.

The human genome is composed of 23 pairs of chromosomes. Twenty-two are autosomes, and 1 pair of 6 chromosomes—in women, obviously XX, and in men, XY. And of course there's the mitochondrial genome which we shouldn't overlook, although it's small. The human genome is comprised of about 3 billion base pairs of DNA. By contrast, the mitochondrial genome is about 16 and a half thousand base pairs in length; small, but equally important.



Here's a picture of human chromosomes. This is from a woman who has 2 X chromosomes, and there are the chromosomes numbered 1 to 22 for the autosomes. Why are they numbered this way? The answer is on the screen. Size, right. Chromosome 1 is the largest. Actually, chromosome 21 is a little smaller than 22—they got these backwards. And chromosomes are divided into short and long arm. Short arm is designated P, long arm is designated Q. Anybody know why the short arm is P? Petite, yes, it's the French name, *petite*, so it's abbreviated P. And why Q? Because Q comes after P in the alphabet.

So there have been a series of breakthroughs that I've mentioned historically. And this is another one. This is the two publications from 2001, four years ago, that describe the initial draft sequences of the human genome. One was produced by the Human Genome Project, and the other was produced by Craig Venter from Solara. But the sequences are virtually identical. There was an update in 2004, and we know now that there are about 3 billion base pairs of DNA in the human genome. There are only a few gaps that haven't been sequenced yet. About 99% of the coding regions of the human genome have been sequenced. About 22,000 genes have been identified, so more will be identified; so the exact number of genes in humans is not known, but it's somewhere between 22,000, and maybe as high as 30,000—but it's not a lot, there's a limited number of genes obviously.

So we've learned that any two unrelated people have virtually the same genetic information; 99.9% of DNA between any two people will be identical; only 0.1% will vary, and that will account for the variations in skin color, hair, height, so on. But that 0.1% also confers important other genetic variance besides physical characteristics: they may also carry increased risk or decreased risk of developing disease, and also response to drugs can vary according to one's genetic background.

## Chapter 5: Techniques of Molecular Genetics

A lot of molecular genetics is driven by technology. Two very important tech-



niques are PCR, polymerase chain reaction—this is a powerful technique to amplify segments of DNA—and DNA sequencing, which is usually performed using this Sanger method, which I'll explain just very briefly. The idea behind PCR is you take DNA from a person, or from virtually any organism, and you can heat it and separate the double strand of DNA, and then you have specific primers for specific segments of the DNA that anneal to the template DNA. Then you use an enzyme called polymerase, in most cases Taq polymerase, which then uses this primer to initiate replication of that piece of DNA. And you can repeat this cycle—heating, annealing and extension, polymerization of the DNA—25 to 35 times, and so you exponentially expand that piece of DNA many-fold. So it's a very powerful technique to amplify bits of DNA, and it's been applied in many, many ways.

Then related is DNA sequencing using that same type of polymerase, where you use the primer and you use either radioactive labeling or fluorescent labeling of the DNA molecules. And instead of just allowing the DNA polymerase to extend indefinitely across a template DNA, you add dideoxy nucleotides that stop the polymerase reaction. Then you can separate out these strands of DNA of different sizes, and then you can read the sequences of DNA based on the labeling, either radioactively or fluorescently. So that's how we do DNA sequencing.

## Chapter 6: Monogenic Disorders

I'm going to talk a little bit about how we apply some of these techniques, first starting with monogenic disorders, and then I'll briefly talk about complex disorders at the end. Monogenic disorders are caused by mutations in a single gene. And there are multiple examples of that, almost 2,000 at this point. How do we find these causative genes and mutations that are associated with human diseases? Well, first we have to start with identification of families with the hereditary condition, and identify whether they have an autosomal dominant, recessive, or mitochondrial inheritance pattern. We then identify the chromosomal locus causing the disease by performing marker studies and linkage analysis. And then after we identify the chromosomal locus we try to narrow it down using fine mapping techniques, and then eventually we sequence genes within this chro-



mosomal segment and look for mutations.

So how does that work? I have the family, and you have to understand that in each meiosis there's a 50/50 chance that if you use a marker allele that it will randomly segregate with the disease. So if there are 10 meioses in a family, the chances are that if you take a marker for a particular chromosomal segment, the chances of that marker segregating with the disease are about  $\frac{1}{2}^{10}$  or roughly 1 in 1,000. So if you apply 400 markers across the genome, eventually you will find a marker that looks like it will segregate with the disease. Here's an example of that, graphically anyway. Here's an autosomal dominant disease—here the father has the disease, and all 10 children, unfortunately, inherited the condition. Now using colored chromosomes to make the concept easier for you to follow, here you can see that the father has a yellow chromosome and a red chromosome. It turns out that every person who's infected inherited the red chromosome from the father. The mother has 2 different-colored green chromosomes and they are randomly inherited in the subsequent generation.

What are the chances that every person who's affected inherited the red chromosome? Well, each person has a 50/50 chance of inheriting the red chromosome, and they also have a 50/50 chance of inheriting the disease. So this person has a 50/50 chance of having the red chromosome, so it's one-half; and if all of them have the red chromosome, it's  $\frac{1}{2}^{10}$ , or 1 in 1,000 chance that all 10 people who are affected with the disease happen to inherit the red chromosome by chance. That's a very slim chance. So that's the idea, how we do linkage. Instead of coloring the chromosomes, we have to use markers. We have to look at specific sequences—sequence tag sites in the genome in order to identify which segments of the chromosomes are inherited within a family.

There are various types of these markers. Initially, there were restriction fragment length polymorphisms—these are sites where restriction enzymes cut. That's been replaced largely by simple sequence repeats. These are short segments of repeated DNA sequences; like CA repeats, they vary in length. And more recently than that, we've started using single nucleotide polymorphisms, single base variations in the human genome to map out these disease loci.



I'm going to show you an example of how we did this kind of linkage analysis in families with this very rare disease, mitochondrial neurogastrointestinal encephalomyopathy, or MNGIE. This is an autosomal recessive disease, and although it's a nuclear DNA defect, there's instability of mitochondrial DNA. So a factor involved in mutating the stability of mitochondrial DNA, it's mutated and results in instability of mitochondrial DNA. The disease is characterized by droopy eyelids (ptosis), ophthalmoparesis (inability to move the eyes), severe gastrointestinal problems leading to cachexia, peripheral neuropathy, leukoencephalopathy—so this is a multisystem disease.

Here's a family with this condition, and we applied these microsatellite markers in this family. Here we see that the father and mother, who are carriers of this disease mutation, have 4 out of 8 children who are affected. You'll notice that every single person who is affected has inherited the lower band from both parents. The parents have band 1, 2, and 3. They both share band 3. But you'll see that every person who's affected has inherited 2 copies of band number 3. So that suggests that the disease locus could be very close to this marker.

How do we express that mathematically? We do something called linkage analysis which assesses whether 2 loci are physically close to each other. We express that mathematically by the logarithm of the odds [LOD], and that's a logarithm that the disease is linked or not linked to a locus. And an LOD score of 3 is required to consider linkage significant to a disease. Here's how it's expressed mathematically. This is the logarithm that the marker and the disease are linked versus the odds that they're randomly segregated. So in the example that I showed you, where there were 10 people who were affected with the disease, if the red chromosome was inherited by every single person, or a particular marker was inherited by every single person, if the marker and the disease are tightly linked the odds that they are linked is 1; but if they're not linked and it just randomly happened, it would be 1 in 1,000, so the logarithm of that would be log base 10 of 1,000—would be a LOD score of 3. And indeed that's a little complicated. To do this kind of study, we use multiple markers, generally around 400 across the chromosome, and then we perform these linkage analysis studies



that determine the lod scores. Then we try to fine map the chromosomal locus through recombination events.

As I told you in the beginning, Thomas Hunt Morgan came up with the idea that genes that are very close together will be tightly linked in a chromosome, whereas the genes that are very far apart on a chromosome will not be linked because there are recombination events through meiosis. And here's an example of recombination. Here you can see that the person is affected, has the red chromosome, and the children who are affected inherited either the whole red chromosome or just the P arm, the short arm, of the red chromosome. This tells us that the disease is probably located on the short arm of this chromosome because there were recombination events here. Does that make sense? So you can, through recombination events, narrow down the locus and hopefully you'll find some individuals in a subsequent generation that will have inherited only part of that chromosome.

In the example that I started to tell you about before in MNGIE disease, we had four families and we used microsatellite markers to map the chromosomal locus. Here in the boxes I show you the markers that were shared by the affected people in four different families. In the boxes these are the regions that are shared with affected people, and there are recombinations between the boxed alleles and the other alleles, so this narrowed down the chromosomal locus. In this case we were able to narrow it down to the tip of the long arm of chromosome 22. We express the size of chromosomal loci in genetic units called centimorgans, named after, of course, Thomas Hunt Morgan. One centimorgan is defined as a 1% recombination frequency, and corresponds to about a million base pairs of DNA. Once we narrow it down, the chromosomal locus for disease, to about one centimorgan—actually we could probably do it with larger regions now because of the Human Genome Project; in the old days we did positional cloning that was to find a gene based on just knowing its chromosomal location. In the old days we had to sequence the region and find genes, but now because of the Human Genome Project, we can just go to the Web sites—there are three major Web sites—and identify the genes in that region, and identify the ones that look like best candidates for the disease—and that's the positional candidate approach.



In that example of MNGIE, we mapped the disease locus here to the tip of the long arm of chromosome 22, and we were fortunate that this was the first human chromosome to be sequenced. We went to one of the Web sites and looked up the genes that were located in this region of chromosome 22 here, and we found one that was encoded endothelial cell growth factor or thymidine phosphorylase, and we sequenced it and we found mutations. Here are the various mutations that we found. So that's how we find the causes of monogenic disorders.

### Chapter 7: Types of DNA Mutations

Various types of mutations can occur in DNA: you can have point mutations, deletions, duplications and dynamic mutations. Here I'm using words to represent the codons—the codons are triplets, in this case here I'm using three-letter words. So if you change one letter in this sequence of letters you'll change this word from *let* to *get*, so that changes the meaning of the sentence from “The law may lie and let men die,” to “The law may lie and get men die.” This is an example of what we call a missense mutation, a point mutation that alters the word—or in the case of DNA, the DNA alters the amino acid.

You can have deletions of DNA, and you can have deletions that are in frame. So here there are nine bases deleted, so the sentence changes from “The law may lie and let men die,” and here you delete nine letters, and so the sentence becomes, “The law let men die.” Here you can see that the words before and after the deletion are unchanged, so this is an in-frame deletion. But here's an example of an out-of-frame mutation, a friendship mutation. Here there's a 7-based deletion, so the three words before are the same that are normal, but the words afterwards change completely because there's a frame shift, so the reading frame of these letters is altered. Obviously these types of mutations that cause frame shifts are more severe.

To give you an example of how that occurs in people, about two-thirds of patients with mutations in dystrophin will have deletions. They can either be in-frame, like this, and cause the milder phenotype called Becker's muscular dys-



trophy, or you can have a frame shift mutation and it will cause the more severe disease called Duchenne's muscular dystrophy. Although they're very similar mutations, in these patients the consequences are different just because of the way it alters the reading frame of the DNA.

I forgot to include a slide on dynamic mutations, and by that I mean generally repeated segments of DNA, like CHE repeats, that become expanded so they're dynamic and the degree of expansion affects the severity of the disease. Obviously the number of diseases associated with genetic mutations has been rising, and this is what the graph looked like in 2001. But actually the rate of new gene discovery in diseases has declined. In the last four years, only about 400 new genes have been described, so this curve is beginning to slope downwards because most of the major common monogenic disorders have been identified. There are many rarer diseases that are yet to be defined, but most of the common genetic disorders that are due to one-gene mutations have been defined. In 2001 there are about 1,400 diseases, clinical diseases, associated with molecular genetic defects. But the number of disease genes is only about 1,100 at that time. So there were more diseases than genes.

Why is that the case? It's because sometimes mutations in a single gene can cause different diseases, and this is an extreme example. Mutations in this gene, which encodes a nuclear protein envelope called lamin A/C, can cause seven different diseases. It can cause muscular dystrophy—it causes Emery-Dreifuss syndrome which is a muscular dystrophy with cardiac involvement; it can cause just heart disease (cardiomyopathy); it can cause lipodystrophy; it can even cause premature aging (progeria, Hutchinson-Gilford progeria syndrome). So it's quite remarkable that a single gene can be mutated in different ways and cause various diseases.

I talked about monogenic disorders that are caused by mutations in a single gene. Polygenic disorders are diseases or conditions that are due to the effect of multiple genes. Complex disorders include polygenic disorders, so they're disorders that can be caused by multiple factors—either multiple genes or genes plus environmental factors. And many of the traits in humans, like diabetes, hyperten-



sion, can be considered complex disorders because they require the interactions of genetic risk factors with environmental factors.

## **Chapter 8: Impact of Molecular Genetics on Neurology**

So how has molecular genetics impacted neurology? Of course it's to find the causes of many neurological diseases that are inherited. Now we can provide definitive DNA testing for many of these monogenic disorders, and we can provide more accurate genetic counseling and offer some information about prognosis. Unfortunately therapies, particularly gene therapy, have lagged behind—and it's a lot of the more rational classification of many diseases. For example, the spinocerebellar ataxias can now be classified according to their mutations, and many of them are due to dynamic mutations—that is, expansions of simple repeats like CAG repeats. The limb-girdle muscular dystrophies are another example. That was a wastebasket diagnosis: when a patient didn't have Duchenne or facioscapulohumeral or ocular pharyngeal muscular dystrophy, and they just had proximal muscle weakness, they were called limb-girdle muscular dystrophy; but now we know that there are many genetically distinct forms of limb-girdle muscular dystrophy. And Leigh syndrome is another example—that's a syndrome that's due to multiple genetic causes. There's an update on limb-girdle muscular dystrophy. Now there are at least 18 genetically distinct forms of limb-girdle muscular dystrophy, whereas before it was considered just one entity, but now we know that there are various dominant and recessive forms, due to different mutations in muscle proteins.

## **Chapter 9: Complex Disorders**

So going back to the concept of complex traits. It can be very difficult to sort out what are the genetic factors responsible for complex disorders, and sometimes there are monogenic disorders that can cause these diseases, like Alzheimer's disease; and mutations in 3 genes have been identified in the APP precursor protein—presenilin-1 and presenilin-2 have been associated in families with early onset Alzheimer's disease. Late onset Alzheimer's disease has been associated with certain alleles of the apolipoprotein.



The mutations in familial Alzheimer's disease have been very informative. They have shown us that mutations in amyloid precursor protein can cause Alzheimer's disease and presenilin-1 and -2, which are important in processing of amyloid protein in cleaving the amyloid precursor protein into A-beta protein, polypeptide. So obviously this suggests that the amyloid protein and its processing are very important in the pathogenesis of Alzheimer's disease, and it obviously helps explain why these amyloid plaques occur. But obviously that's not the full story because most patients don't have familial Alzheimer's disease. We know now that variants of this apolipoprotein-E can confer increased or decreased risk of Alzheimer's disease; for example the apolipoprotein-E2 is associated with resistance to Alzheimer's disease, but apol-E4 allele is associated with susceptibility to Alzheimer's disease; people who are heterozygous for this E4 allele have a threefold risk of Alzheimer's disease, and those who are homozygous have a fourteenfold risk of developing Alzheimer's disease. Here's an example of a genetic predisposing factor for developing Alzheimer's disease—obviously a complex disorder which can be influenced not only by genetic factors but environmental factors as well.

About these complex traits, how can we identify these genetic variants that contribute to these complex traits? The apol-E4 allele was found somewhat fortuitously. It's a very difficult problem. There are 3 billion base pairs of DNA—how do you sift through all of that information to find the variants in the DNA, predisposed to these complex diseases? You can't sequence the entire genome—it's too expensive, it's too laborious. You can look at candidate genes, sequence candidate genes; and look for variants, polymorphisms in the DNA; but that's a hit-or-miss proposition.

There is a strategy using single nucleotide polymorphisms. These are specific bases in the human genome that vary quite a bit from person to person, and there are about 10 million SNPs, or single nucleotide polymorphisms, that have been identified. And you can use those SNPs—10 million is certainly better than 3 billion bits of DNA—and arrange them in haplotypes: haplotype 1 would have an A and a T at these particular sites, and haplotype 2 would be an AG, and a GG would define haplotype 3. So you can use haplotypes—that is, groups of



SNPs—to follow the segregation of chromosomes through populations. The human genome doesn't recombine at equal rates throughout the chromosomes. There are hot spots for recombination. If you can identify which segments of DNA tend to segregate together, then you have to look at only a few SNPs without those haplotype groups. And so instead of using the 10 million SNPs to characterize inheritance patterns, you can reduce the number down to a more manageable number, between 300,000 and a million SNPs. This is the HapMap Project. The HapMap Project is trying to define these haplotype blocks that are defined by a limited number of SNPs that will allow us better and faster characterization of inheritance chromosomes across large groups of people.

To give you one more concrete example of a complex disease, now genetics are beginning to define some of the risk factors for multiple sclerosis. Multiple sclerosis obviously is not a purely genetic disease—it's multifactorial presumably, and it's influenced by the environment, it occurs in temperate climates—but also we know that there are some genetic factors because there's a higher incidence of multiple sclerosis among identical twins, so there are clearly genetic factors involved.

In this study that was just published earlier this year, they studied 730 families with multiple sclerosis. They used more than 4,500 genetic markers on each of these and they found linkage to chromosome 6—the short arm of chromosome 6—which is the locus of the histocompatibility complex. Then it was published in September, and the following month, in October, 2 studies came out in *Nature of Genetics*. The first one fine mapped that chromosomal locus on chromosome 6—and they mapped the locus using another 1,100, almost 1,200, families from northern climate countries with multiple sclerosis—and they were able to refine the locus to a specific HLA class 2 region in this chromosome 6. Back-to-back with this article was another interesting study by David Rice at Harvard, in which they used a very interesting technique called admixture analysis. It's known that African Americans are less likely to get multiple sclerosis than Caucasians, but we know that in African Americans about 20% of the ancestry can be traced to European lineages. You can try to use these markers and define which regions of the chromosomes in the African Americans are associated with the risk of



developing multiple sclerosis, and presumably that came from the European background. And so they took this admixture analysis strategy and they were able to define a new chromosomal locus on chromosome 1 that's different from the previously described MHC loci that are associated with multiple sclerosis. This is a very new way of trying to identify genetic factors for complex diseases.

## Chapter 10: Clinical Applications of Molecular Genetics

I'll just end with a couple of points about molecular genetics and raise a couple of questions. Will molecular genetics provide insights into the pathogenesis of common complex diseases such as stroke or Alzheimer's disease? Will molecular genetics influence the way neurologists prescribe drugs—that is, pharmacogenomics will be able to predict how patients will respond to a drug based on their genetic background? And ultimately, will gene therapy be effective for some of these hereditary neurological diseases?

The answers are yes to the first one. Clearly, molecular genetics has provided insights into common disease like Alzheimer's disease, as I described earlier. Will it allow us to custom-tailor drugs in the future? Almost certainly it will. And that's just beginning to occur. Will it allow gene therapy to be effective? We don't know, but many people are working on gene therapy, and the answer is probably gene therapy will be on the horizon for these often devastating neurological conditions.

The human genome is, as I said, 3 billion base pairs in length. It's a very tedious project to read through all of this, but now that it's been sequenced, there's a lot of information out there, and we're beginning to unravel a lot of this; and it has a lot of implications, and it's beginning to overwhelm many people, but it's obviously very important because the information is out there. Molecular genetics nowadays is more like doing crossword puzzles than doing fundamental basic science, because the information is out there—the 3 billion base pairs of DNA are available on the Internet—and now we have to be clever in how we analyze that information, and how we use that information is going to be even more important in the future.

So I would like to end and ask if there are any questions.