

Professor Jacks is out of town so I am going to tell you about Recombinant DNA 3, then he's going to come back and tell you about Cell Biology, and then you will have finished the foundations part of the course. And we'll move onto things that build on the foundation, the Formation Module and the part of the Systems Module, which I'll be teaching you for the next few weeks, but today is Recombinant DNA 3. And, as you've been hearing for the last couple of lectures, this is one of the How-To Modules that we've put in the course. How to make use of the information that you have been learning in Molecular Biology and in Biochemistry and in Genetics to use these disciplines or these pieces of information to do something useful. And recombinant DNA is really an extraordinary set of technologies that just keeps getting more and more extraordinary. And the way one can manipulate biological systems now is really very exciting. And it continues to be exciting. When I was a beginning graduate student we were able to clone the first pieces of DNA. And now we can really do a lot more than just clone DNA. So I want to tell you about some of the things that are really essential to understand about this technology, and then take you through some of the forefronts of where recombinant DNA technology is now. We're going to cover three things in this lecture. DNA sequencing, using genetic polymorphisms for various genotyping analyses, and then I'm going to try to touch on, and we'll have to see how we do here, making animals that are so-called transgenic. So transgenic technology. And I'm going to use PowerPoint pretty much for most of the lecture, so you have most of the relevant stuff in front of you. I'm going to frame this in terms of a human disease, familial hypercholesterolemia. So you may remember way back when in biochemistry we talked about cholesterol. Anyone remember what class of macromolecules cholesterol belongs to? Lipids. Thank you. Lipids. OK. I'm not even going to give a frog for that. And we have this sense of cholesterol being a really bad kind of molecule but, in fact, cholesterol is an essential lipid. It's extremely important. Without cholesterol you'd die and you need it for many things. Not only for building membranes in your cells but also, if you think way back, you may remember me telling you that cholesterol was part of or had a chemical structure that was very similar to the steroid hormone family. And steroid hormones, and we'll discuss this more in the future, are very important molecules that tell one part of the body what to do, that regulate what different parts of the body are doing. So cholesterol is part of this whole signaling system. And really it's not actually understood all of what cholesterol does, but it's very important. However, too much of it is not good. And it's probably not good because, but it's not actually clear. I'll tell you what happens if you have too much cholesterol, but actually why it happens is not that clear. So let me talk about this slide up here, and then we'll talk about what too much cholesterol does for you. So familial hypercholesterolemia is an inherited disease, and it's caused by mutations in a gene called the LDL receptor, that encodes for something called the LDL receptor. Now, LDL stands for low density lipoprotein. And you had this in a previous lecture because I'd been mentioned these to you. Low density lipoproteins. And these bind to various lipids, including cholesterol, and are taken up into the cell. And some of them are OK, you probably need some LDLs, but too much LDL is bad. And if you have too much LDL receptor, the thing that actually binds to the LDLs, you get too much LDL taken up into the cell. So this LDL receptor, you'll talk more about this in cell biology, this LDL receptor, and you've already had some of this, the LDL receptor is a protein that binds to these LDLs, takes them into the cell, and then your cell gets full of LDLs. OK? And as a consequence of this, your cholesterol levels go way up. Now, you can be heterozygote or homozygote for familial hypercholesterolemia, for the LDL receptor gene. OK? For the familial hypercholesterolemia gene. Try to say that one quickly. All right. So if you're heterozygote, you have an increased risk of heart disease. In particular for this thing called atherosclerosis I'll talk more about in a moment. If you are homozygote, so you have two copies of a mutated LDL receptor gene, you get severe heart symptoms and you die early. OK? What is atherosclerosis? Atherosclerosis is a disease that occurs because you get these buildups of stuff in the blood vessels. And the stuff is fat and it's proteins, and it basically makes a big lump that eventually occludes or blocks the blood vessel. And so atherosclerosis is bad because impedes blood flow. And if you impede blood flow, eventually your heart will seize up and you will have a heart attack, and that can have, obviously, very severe consequences. So atherosclerosis occurs because you have high levels of LDL. And it's really, the actual etiology of atherosclerosis is not really clear. Part it may be that there's just too much fat around and that starts actually getting deposited out of solution, but it's much more complicated than that. And there seems to be a very complicated chain of events by which you get these atherosclerosis plaques sitting on the lining of blood vessels and impeding blood flow. OK. So there is a lot of interest medically in atherosclerosis, particularly in countries such as ours where food is plentiful and people tend to have too much. And obesity is a problem anyway because that is part of the set of risk factors for atherosclerosis. So here are the risk factors. High levels of LDL, high blood pressure, diabetes, cigarette smoke and so on. And familial hypercholesterolemia is contributory to high levels of LDL and atherosclerosis. OK. So one of the things I want to do is to keep thinking about this disorder and walk you through how you figure out who's got FH. OK. What you can do is to get blood cells from people at-risk, and you can actually examine the LDL receptor gene in the blood cells of people who are at-risk for familial hypercholesterolemia. And what I tell you about is how you can actually sequence the gene, the FH gene, see if you can find the mutation and see whether or not you can then identify people who are at-risk for the disorder. So the first thing I want to tell you about today is DNA sequencing. DNA sequencing. What is DNA sequencing? Does someone care to give me a definition or think about what I might mean by DNA sequencing? In particular, what part of the DNA are we sequencing? Thank you, Jamie. You want to say it louder? The bases. Yes. So in DNA sequencing, and maybe I even wrote this, what is this, what you want to do is to determine the base sequence of the DNA. OK? You want to determine the sequence of AGCT along a DNA fragment. This technique is powerful beyond almost anything else. It's an extraordinary technique. The ability to sequence DNA is extraordinary. And it's extraordinary because you can get out of it information that is absolutely essential for understanding life. What you can get from DNA sequencing is an understanding of the coding capacity of a gene. So, just like you did in your exam, we gave you a string of DNA and you conceptually translated it into the protein.

Well, you can do that in real life by looking through the genome, the human genome and finding stretches of DNA and conceptually turning them into RNA and into protein and saying, OK, is this a gene? Does it code for something? And what does it code for? So you can figure out the coding capacity of a gene. Part of that is actually identifying is a gene a gene? So we've sequenced the entire human genome. And I've told you previously that only about 5% of the genome is actually genes and the rest is other stuff. So one of the things you want to do with DNA sequencing is to identify genes. And that's actually very difficult to do it turns out. But that's one of the things you can do with DNA sequencing. I'll talk more about identifying genes that are associated with disease, that are causative of disease. And particularly alleles that are associated with disease such as in the case of familial hypercholesterolemia. One can figure out evolutionary relationships between organisms. So you've probably heard for years about how similar we are to chimpanzees or how similar we are to dogs or to dolphins or whatever. But, actually, we didn't really know. Now we can sequence a human genome, we can sequence a chimp genome, a dog genome, a dolphin genome, and we can actually look and see how similar we are. And we can try to figure out, in evolutionary time, what's changed between the dolphin and ourselves and what makes a dolphin a dolphin and ourselves ourselves. It's a very tough question, but DNA sequencing is essential for trying to answer that kind of question. And then one can ask about the genome is other ways. Can one find the promoters of all the different genes? Remember promoters that make genes be transcribed? The centromeres, the middle of chromosomes. Various other elements in the genome that are essential for its function. So I'm going to spend quite some time talking about DNA sequencing and tell you that DNA sequencing, most of the DNA sequencing we do uses a trick. And it's a terrific trick. It really is. So this DNA sequencing, I'll write it because I don't think I have this on one of your PowerPoints. The method of DNA sequencing I'm going to tell you about was devised by a scientist called Fred Sanger. So I'll tell you about it. It's called dideoxy, it's also called chain termination, and it's also called Sanger sequencing. Professor Sanger is a British scientist who received two Nobel Prizes. The first was for figuring out how proteins, how to sequence proteins, and the second was for figuring out how to sequence DNA. When I was a student, I heard Professor Sanger talk. And he gave a lecture which was really memorable. It was packed, a packed auditorium. And he spoke the entire time like this. I don't think he looked up once. He gave the entire lecture like this, and he was barely audible. But at the end of the lecture he got a standing ovation from everybody because really what he's done, figuring out how to sequence proteins and how to sequence DNA was really an extraordinary accomplishment. So that's the method I'll tell you about. And it uses a cool trick. So you know now that the sugar in DNA has a 3 prime hydroxyl group, and that hydroxyl group is the group unto which the phosphate gets added. Right? And without that hydroxyl group you could not add on the next nucleotide, right? It's a question. Think about it. OK? I don't mean it to be rhetorical. I want you to really be thinking, OK, about this, because otherwise you won't understand the method. So here's the 3 prime hydroxyl on regular deoxyribose. OK? In the Sanger or dideoxy method one uses in the reaction mix, and I'll go through this with you in a moment, a sugar or nucleotide that's a dideoxy nucleotide. In other words, on both the 2 prime and the 3 prime of the sugar, of the ribose there is no hydroxyl group. There are just those hydrogens. Now, a dideoxy nucleotide such as this one can get incorporated into DNA just fine because this phosphate, the triphosphate here can react with a regular nucleotide that's got a 3 prime hydroxyl. However, once it's been incorporated you cannot elongate the chain anymore because there is no reactive hydroxyl group. OK. So based on this principle let me explain. I've got one of your handouts here. OK. So here we go. Revision, your template, your primer, here's your template strand, always goes 3 prime to 5 prime. Here's your 5 prime to 3 prime primer. If you add nucleotides, deoxynucleotide triphosphates and DNA polymerase, you will polymerize the whole fragment. If you add, however, to the mix of dNTPs and DNA polymerase a low-level of dideoxy nucleotide triphosphates, every time you add on a nucleotide the polymerase can either use a regular nucleotide triphosphate, in which case the chain can elongate subsequently, or it can use a dideoxy nucleotide triphosphate. If it uses one of the dideoxy NTPs the chain will terminate. It cannot be elongated any further. So you get something like this. And the trick here is really this low-level of ddNTPs. OK? So if you have your template and your primer and you do a reaction with your dNTPs at a reasonable level and you spike the reaction with a low-level of dideoxy NTPs, you get a whole bunch of different length chains polymerized. Because there is some probability, at every position, that you're either going to get a ddNTP incorporated, in which case the chain terminates, or you're going to get a regular nucleotide incorporated in which case the chain can continue for a bit. OK? So that is paramount to dideoxy sequencing. So let's continue now by looking at a specific polymer and following through exactly what happens. So here I've given you a template and a primer. And we're going to do the same reaction that we just did conceptually. We're going to do it again conceptually except with letters. We're going to mix together. And we're going to do, and I see a mistake up here already, but that's OK. You'll bear with me. What I've done here is to put in some dideoxy ATP. And I meant to say here I've got dATP at high levels. And I've got all the other nucleotides here, too, at high levels. OK? That's my error and I will correct it. You should correct it now in your handout. So where it says dATP high, that should actually say dNTPs high, not just dATP. OK? All right. So let's look and see what happens to this reaction. And I've noted here that this dideoxy ATP can be radioactive or florescent. Or actually it doesn't have to work that way but let's just leave it that way for now. OK. That actually is not necessarily true. So let's just focus on the ddATP plus the high dNTPs, and let's see what happens. OK. So one thing that can happen is that, here's your primer in red and here's the polymerized DNA in blue, you get a bit of DNA polymerase. Now here's an A. See? It goes GAGTAA. And I've given you a reaction where the first two As use regular dATP. And so the chain will continue after that. All right? So here we go, GAGTA. And then the next A that's put in is a dideoxy A. And that's the end of that polymerization reaction, and the fragments you're going to get out of it is this little red and blue composite there. You can do the same thing where you say actually in some molecules you get polymerization past the second A, and you keep going until you get to the next A. And at that point, by chance, you get a dideoxy ATP added to some molecules. That is the end of polymerization for those molecules. The chain terminates. For some molecules. however. you'll put in a regular dATP and the chain will continue. But it will terminate. excuse me. at

the next A that's put in because you put a dideoxy A in. So in different molecules you're going to land up with a spectrum of elongated products of different length. All right? And what's crucial here is that the length of the molecules that chain terminate, because they incorporated dideoxy nucleotide, correspond to the position of that particular nucleotide along the chain. So you're only going to get a molecule chain terminating with A when there was a T on the template strand. OK? And so you can map the positions of the T on the template or the A on the elongated strand by the length of the elongated products that come out of this reaction. I'm going to assume you're with me here. OK. So the point is the polymerized fragments terminate where dideoxy A incorporates. Now, you've got to do four reactions to determine the sequence of something. OK. And I've noted here. And the length of the terminated fragment indicates the position of A. You may need to go and work with this a bit. OK? It's a very clever method but it may not be something that's immediately apparent, so go and work with it if you need to. So the length of the terminated fragments indicates the positions of A in the elongated strand, or if you want in T of the template strand. In order to get the positions of all the different nucleotides along that DNA fragment you have to do four separate reactions. One that includes dideoxy ATP, one that includes dideoxy CTP, one dideoxy GTP and one dideoxy TTP. And you do those separately so that you can monitor the positions of each of those four nucleotides by the position of chain terminating as you're going along. OK. So assuming that you guys are with me here at this point, are you? No. That's an honest answer. Raise your hands if you're with me. OK. If you're not with me, don't worry about it. You have to go work with it. It's not intuitive. It's very clever. I mean there's a reason this guy got the Nobel Prize for this. OK? It's a really clever method. OK. So the deal is this. So now what you get out of this is a whole mix of fragments of different lengths that have terminated at positions of particular nucleotides, depending on how you've spiked the reaction. And you've got to separate them from one another somehow to figure out what those positions are. And you can do this in a couple of ways. You can use gel electrophoresis, which was discussed with you previously, where you separate the DNA on the basis of size where the DNA migrates in a gel in an electric field and long fragments stay near the top of the gel and short fragments go to the bottom of the gel because they migrate quickly. And what you can do on a gel, and you've somehow labeled, don't worry about this right now, but somehow you're able to detect each of the fragments that has come out of your mix. OK? So remember you're doing the sequencing reaction on millions and millions or billions of molecules. And so you've got this kind of stochastic mix of molecules of different lengths. And you want to separate this mix of molecules of different lengths. OK. So what you can end up with, once you've separated all these different molecules, is in your dideoxy A reaction mix a series of one, two, three, four, five different sized fragments. In your ddG mix, you got out of that also a series of five different sized fragments. And notice that they're different in size from the ones in the ddA lane, the ones in the ddC lane and the ones in the ddT lane. And the reason they're different in size is because their size indicates the position of where a particular nucleotide is in the DNA fragment or particular bases in the DNA fragment. And then the trick is you could look at this gel and you could read off the sequence. So the shortest fragments that you're going to get are the ones that are nearest the beginning of that molecule you made, nearest the 5 prime end. So the bottom one is G, here's the band in the ddG lane. Then up above it there is this band indicating a fragment in the ddA lane. Above it there's one in the G lane again. Above it there's one in the T lane. So the sequence goes G-A-G-T, and then you can keep reading A-A-C-G-G-T-A-T-G-C-A. OK? Literally like that on a gel. OK? So you can do that on a gel. It's really fantastic. And this is what old sequencing gels look like. And, actually, I used to run them. I used to spend hours and hours running these gels. They're very, very thin. They're about a millimeter thick acrylamide so that you can resolve the fragments that are one nucleotide different in size. Think about that. OK? Each of these fragments, indicated by a band, is one nucleotide different in size. Otherwise, you couldn't get the one nucleotide resolution. So you do that by running very, very thin gels so that you can resolve the fragments well, and then you read off the bottom. OK? I've thrown out all my old sequencing gels. And the reason that I have is that there is new technology where you don't use this kind of display anymore. This is a display where your fragments were labeled with radioactivity and you exposed them to x-ray film and you read the sequence after exposure. Nowadays this is done by machine. And the dideoxy nucleotides are labeled fluorescently. OK? So they're not labeled with radioactivity. They're literally labeled with labels that fluoresce with different colors when you put UV light on them. And you do your dideoxy reaction and you run a gel. Again, it's a gel. It's actually a very thin tube of a gel mostly, but you run your gel. And, again, it's the same idea. You resolve fragments at single base resolution, single nucleotide resolution, and they keep, the gel keeps running and running. And single fragments actually run off the bottom of the gel. And as they're passing down the gel they are detected by a laser. A laser excites the fluorochrome. And the detector, there is a detector which will detect whether or not it's yellow, orange, blue or green. OK? And that will tell you which base is being, has been incorporated at that position. So you get things that come out. It's kind of small but you can go back and look, where instead of getting a gel with those bands that I showed you, you get these peaks and valleys that are different colors. And that's what current DNA sequencing readout looks like. And, in fact, there are machines. What did I do? Lots of primers. Well, it depends. Many copies of the same primer, right. Yes. Dr. Gardel is pointing out that there are many copies of the same primer in a reaction mix. Certainly there are. There are billions of molecules in the reaction mix, and so there are billions of primers. OK, so you have to have a primer for each molecule. OK. And each band, you should realize, is not a single molecule. It's a composite of many, many molecules, many thousands of molecules that have all chain terminated at the same position. So what I want to point out here is that this is what today's readout looks like. And, in fact, nowadays you just get a printout from the company or from the machine that tells you a DNA sequence. And it's this improvement in technology, but that basically uses this chain termination method, that has allowed one to sequence, rapidly enough to sequence the human genome and to sequence multiple human genomes in multiple animals. OK. So let's see. Actually, I have a movie. I guess we can take the time to watch this movie. Let's see if it will work. All right. So primer template. Four reactions, each with lots of molecules, each with their primer. DNA polymerase, dNTPs, dATP, dGTP, dCTP, dTTP. dCTP. excuse me. OK. They're your four reactions. OK. I think is a less dorky movie than some. OK. So

here we go. Here's your primer and your template, and here's polymerization. And, ah, there we go, chain termination, dideoxy nucleotide incorporation, and you cannot get elongation. The poor G is thwarted in its desire to elongate. OK? So you land up with this mix, just like I showed you, and you land up with a set of four reactions, each with molecules of different lengths in them. And here's your gel, and you load them on your gel, and they migrate through your electric field. And there you have your things, you have your fragments. This is a piece of x-ray film you put on top. There are your little bands, your radioactive bands, and here we go. GT, you can read it. OK. Enough. Enough. OK. You can go and look at this yourself. This is an old gel apparatus that one used to do DNA sequencing on. This was the first generation of machine that you could do the fluorescent sequencing on. This is a room full of sequencing machines of the kind that was used to sequence the human genome. In fact, many rooms of machines going all day and all night sequencing and sequencing and sequencing. We have a lot of nucleotides. And it takes a long time to sequence. Although, in retrospect it's not such a long time. And now all the sequencing machines that sequence the human genome are sitting around looking for other work because they all exist. And so that is why we are sequencing things like dolphins and dogs and multiple strains of dogs, multiple breeds, excuse me, of dogs because we have all these sequencing machines sitting around. OK. Honestly, I think that's true, not that it's not useful. All right. So I'm going to move on here. This is Professor Jack's joke that I decided to use also. OK. This is something about DNA sequencing and the implications of being able to use DNA sequencing for genotyping. So I'm going to use that. You can go and read that on your thing. I'm going to move on right to talking about familial hypercholesterolemia and the notion of a disease allele. So here's part of the normal FH gene, the LDL receptor gene, and here it is. And there is a T here in red. And here is the mutant gene sequence and there is an A. So if you're wild type you have a T at this position that's arrowed and if you're a mutant you have an A. And if you do your conceptual protein translation here you get your amino acid, part of the amino acid chain. Obviously it's not at the beginning. And obviously this is DNA and this is protein, so we've removed the RNA here, the RNA step. And you can see here is the amino acid of your wild type, the sequence of your wild type gene. And in your LDL receptor mutant there is a stop codon at this position that terminates the LDL receptor. And so the receptor gene is mutant and does not function as it should. OK. All right. So let me move onto the next thing I want to talk about, which is this question of polymorphisms. What is a polymorphism? Anyone. All right. I'll tell you what a polymorphism is. A polymorphism is defined as some kind of variation in DNA sequence. And it's defined as a variation in DNA sequence at a particular position. So our DNA, all of us have very similar DNA. If we were to sequence me and we were to sequence you and we were to sequence you, we would find that our DNA was greater than 99% identical. If we lined up our three times ten to the ninth base pairs in a very long line, we would find it was very similar. There was about 1% difference in sequence between each of us. And most of that, some of that corresponds to disease gene alleles. We all are supposed to carry about a thousand bad genes, or a thousand genes that if homozygous would give us something bad, and sometimes do. And some of those correspond to changes in differences in DNA sequence that are not directly in genes. All of these differences between different individuals are called polymorphisms, DNA sequence variation. And you can use these to help figure out whether or not someone has a particular disease allele, and also you can use it to figure out where the DNA from a sample comes from me or from you or from Dr. Gardel. OK? And I'll talk about this, using polymorphisms to map genotype. I'm going to talk about a particular kind of polymorphism, and these are called SNPs which is pronounced  $\text{snip}$ . This stands for single nucleotide polymorphisms. So I've said again that human genomes are 99% identical, but there are throughout the genome changes, differences between regions. Single nucleotide polymorphisms are variations in one region. Here's a sample sequence I made up. Here's a G in one individual and an A in another individual. And if you take the population, you find very often that there just is a choice of two, sometimes more, but often just a choice of two nucleotides in one position. Most of the genomes are identical, but you find these little regions where in many individuals of a population there are these variations. In fact, these variations have to be present in more than 1% of the population for this thing to be called a SNP. This is a definition that humans have given but it's a useful definition as a genetic tool. So if there is a polymorphism present in about 1% of the population, whereby I might have an A here, excuse me, and Dr. Gardel has a G at that position, that would be a SNP, and we would be polymorphic for that SNP. In fact, my two chromosomes, OK, that are homologous chromosomes might on one copy carry an A and on the other copy carry a G. Now, these different bases are present at different frequencies. So, for example, it might be very common to have a G at this position in the sequence and it might be very rare to have an A at that position. All right? And that's useful because you can use the frequency of these different nucleotides, these different bases to help you use the SNP to genotype. And I want to point out that usually SNPs occur outside coding regions because 95%, actually more than that, 99% of the genome is not coding per se. 95% is not genes, but then if you remove all the introns and promoters and so on, 99% does not code for any protein. OK. So usually these SNPs are present outside coding regions. So here's to explore this a bit more. You can find lots of these SNPs. There are about three million SNPs in the human genome, and a very large percentage of those SNPs has been identified by DNA sequencing. So you can get the idea. You have to sequence DNA from lots and lots of individuals to identify these SNPs, but people have done it. And we know now more than a million SNPs in the human genome that are located all over different chromosomes, and we know where they're located on different chromosomes. And so you can use these SNPs to make kind of a map, I'll tell you in a moment. So here are some possible genotypes. I've given you a choice of two for each of these. OK? So, for example, for this red SNP here you can be AA, AC or CC on the two homologous chromosomes. All right. So let's keep going with this thread. So because you have these SNPs all over your genome and you know where they are, you can use them to make a map of your entire genome. That doesn't depend on the genes. It just depends on the sequence. And knowing these SNPs is a lot easier to work with than having to sequence the entire genome of somebody every time you want some information. So you can use these SNPs to identify each person. So I have a SNP map of all these hundreds of thousands of SNPs. or up to a million. The usual maps presently used are about 300,000 SNPs per genome. I

have a map of 300,000 SNPs where there are different, actually, I don't, but I could, where there are different alleles at different frequencies, different bases present at different frequencies at specific positions. And we could pick any one of you and make a SNP map for you. And it would look really different from mine, not because the SNPs themselves are different, they'd be the same SNPs, but the actual bases and the combination of bases between all these different SNPs would be different between different individuals. And this SNP-type map is the basis for DNA fingerprinting that is used in forensics and to figure out disease alleles. I'll talk more about this in a second. I want to point out that there are other kinds of polymorphisms that are used in genotyping, restriction fragment length polymorphisms and things called simple repeat polymorphisms. And you can look in your book for these restriction fragment length polymorphisms, but let's talk more about SNPs. So SNP genotyping, here's a whole list, but the ones I'm going to focus on are disease gene mapping and forensics. Also, you use SNP genotyping for paternity suits. OK? So if someone comes and, you know, if someone says it's my kid and the other one says it's my kid, you can figure out very easily whose it is by looking at these various SNPs and figuring out what pattern of SNPs is present in the offspring. OK. So let me actually consider, let me not deal with genotyping for disease alleles at this point. Let me talk about forensics a bit because it's kind of interesting. So how do you do this? Let's look through this slide. You have it as a handout. Here are SNPs. And I've just given you two chromosomes each with two SNPs. OK? And different people will have different bases at these particular SNPs, or they'll have different combinations of these bases. So here's the spot of blood at the crime scene. OK? Our red blood cells do not have nuclei so you cannot get DNA from those, but there are enough white blood cells that do have nuclei so you can. And, actually, you know from PCR now that you need very little to amplify something up by PCR. One cell is sufficient, right? It's pushing the technology, but you can really use one cell. So there are plenty of cells in a spot of blood at a crime scene to isolate the DNA and to PCR amplify the regions surrounding the SNP. So you're not just dealing with these two nucleotides or the choice of these two nucleotides at the SNP. You've got a little piece of DNA that's usually maybe 20 or so bases that includes this choice of single nucleotide polymorphism. So you amplify the SNP region, OK, a region that's constant, that includes the nucleotide polymorphism, and you determine the sequence at the different single nucleotide polymorphism regions. So you might get someone who, at the red position you can be A or C, at the green you can be G or G, purple GT, and yellow you can be A or C. And here the example is C and C. So here are the four suspects, numbers one to four. OK. And here are their genotypes. OK. And here is the spot of blood at the crime scene that actually has this genotype. OK. So let me go back here. This is the genotype in the blood at the crime scene. OK. So the red sequence on one chromosome is an A, on the other is a C, so you have AC. On the other, the green sequence you have GG, purple you have GT, and yellow CC. So you're looking to see whether or not any of the suspect genotypes map up with a spot of blood, right? So we're assuming that a spot of blood, you know, comes from one of the suspects that was attacked by the person who was the victim. OK. So you have a victim with scratch. Someone has a spot of blood. And you see whether or not, or you can use semen samples, you can see whether or not the DNA in the human tissue that is believed to come from the attacker is matching of any of the suspects' genotypes. So there are a lot of assumptions there, right? You have to have tissue at the crime scene that you believe to come from the attacker. And then, once you have that, you can determine its genotype and compare it to the genotypes of the suspects. And you find, for example, here that, let's see, yeah, so I believe the suspect number three has the same genotype as the DNA that was in the spot of blood at the crime scene. And that would be some evidence that this suspect number three was the person who did it. Now, in actual fact, you do this not just for four SNPs, you do it for thousands of SNPs. You don't usually do this for 300, 00 SNPs because that's expensive and it's a lot of work. And forensics doesn't put that much money into this. However, the more SNPs you use for genotyping the more sure you are of the suspect's identity. OK? Because it's really a matter of frequency of whether or not you're going to get the same combination of these different SNP bases in different potential suspects. So the greater the spectrum of SNPs you look at, the more sure you are of the suspect's identity. Now, in some cases this has been very, very useful. And there are a number of people on Death Row who have been exonerated by going back to DNA recovered from the crime scene sometimes years ago, doing SNP mapping and showing that they really couldn't have done it because the genotypes did not match up. Usually these were rape cases and the semen genotype just did not match up with the semen genotype of the person on Death Row. So this is very valuable technology. OK. It was used in the O.J. Simpson trial, but not as well as it could have been which lead to equivocation there. OK. So time is fleeting. I'm going to mention a technology to you in the last couple of minutes, and then we'll come back to it as we go on through later parts of the course. So I've talked today about DNA sequencing. I've talked about using polymorphisms to genotype people either, well, for disease alleles I focused on who-done-its. Something else that I want to throw out at you at this point is the notion of transgenic technology. And I'm going to tell you what transgenic organisms are as part of completing the Recombinant DNA Module. And then we'll come back in future modules and talk more about how you make these things. But I want to have this as part of your compendium now. A transgenic animal or transgenic organism is an organism where you have manipulated its genome in some way, where you've either inserted extra DNA into its genome or you've removed DNA from its genome or you've done something to its genome such that it was not the organism that you started off with. Genetically modified organisms. The food that you eat that is genetically modified has had its genome tampered with. This type of transgenic technology is very, very useful, not only for creating genetically modified foods, but it's very, very useful for creating disease models of animals. And I'll tell you now that there is a mouse model of human familial hypercholesterolemia that has been created by making a specific mutation, that T to A mutation in the mouse LDL receptor gene. Another thing that is extremely useful about transgenic animals is that you can get them to make specific proteins. So, for example, there are goats that have had inserted into their genomes genes that encode for particular medications, for particular drugs. And you can get these drugs out of the milk of the goats usually or out of the serum of the goats because they are constitutively producing them

because you've put various genes into their genome. So I'm going to leave it there and we'll talk about how to make transgenics in a future lecture.