

MIT OpenCourseWare  
<http://ocw.mit.edu>

7.012 Introduction to Biology, Fall 2004

Please use the following citation format:

Eric Lander, Robert Weinberg, and Claudette Gardel, *7.012 Introduction to Biology, Fall 2004*. (Massachusetts Institute of Technology: MIT OpenCourseWare). <http://ocw.mit.edu> (accessed MM DD, YYYY). License: Creative Commons Attribution-Noncommercial-Share Alike.

Note: Please use the actual date you accessed this material in your citation.

For more information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>

MIT OpenCourseWare  
<http://ocw.mit.edu>

7.012 Introduction to Biology, Fall 2004  
Transcript – Lecture 18

Good morning.

It can't go without at least some acknowledgement and mention.

If you should ever find yourself in life in a situation where you have or are about to give up all hope, you think things are utterly impossible and there's no way, you will remember this week that nothing is impossible.

It is possible to come back three games down in the bottom of the ninth inning, you've got to believe you can do it, and remember to have Dave Roberts pinch run.

Just a general bit of good advice. What an amazing week, just absolutely amazing week. Wow. There are lessons in life to be taken from it. Please do take them. You know, there really are. I mean I'd given up hope by that point, I confess. I wish I could say oh, I knew they were going to pull it out, but I didn't. And, boy, they pulled it out one game at a time. So, all of you think good thoughts this week. This could be a historic week, you know, you were here.

Anyway, onward.

We were talking last time about how to analyze your clone. The notion of cloning random pieces of DNA, identifying your clone within a library, purifying the DNA from a clone, doing some preliminary analysis by maybe cutting with a restriction enzyme, then sequencing it using these techniques that I'd described would allow you to take the clone that was, say, able to rescue the yeast that couldn't grow without arginine and figure out what its DNA sequence was.

You could take the clone that you had obtained by hybridizing with the DNA sequence corresponding to the protein sequence for beta-globin and sequence it and see the beta-globin gene sequence perhaps.

This is very powerful. I want to take a brief moment, we'll come back to it in more detail in a subsequent lecture, but I really described how you would sequence one clone. I just want to make a note, because someone asked about it last time, about how you would sequence an entire genome. Someone asked about this.

Remember before we pulled out our clone, we sequenced it, we got its DNA sequence. What if I wanted to sequence the entirety of a genome? Yeah. Do a lot of this, right, basically if I got a whole genome. Well, somebody asked could I put a primer here and just sequence?

It would take a very long time. And it turns out that it wouldn't work because the separation that you can achieve through gels is a function, the separation between  $N$  and  $N + 1$  in length goes like the logarithm of the ratio. So, it turns out that when  $N$  and  $N + 1$  get to like about a thousand, you can achieve very little physical separation between them. And so, DNA sequencing runs cannot go much past the thousand bases.

So, the problem with sequencing a genome by putting down a primer on an extraordinarily long piece of DNA, a hundred million bases, is you cannot separate the little fragments like that. So, what you do is you break up your genome into lots of pieces. One strategy, break it up into a library of some very big pieces. It turns out you can make pieces at random of a hundred thousand base pairs.

Cloning these in bacterial artificial chromosomes, as we talked about before. Take a library of bacterial artificial chromosomes and then begin sequencing them. And take any given bacterial artificial chromosome and break it up into a whole lot of pieces that are maybe a thousand bases long, and you could sequence all of those. How do you arrange to get just a perfect overlapping set of thousand base pair clones that perfectly tile across the sequence with no redundancy?

You don't. That's the correct answer. That's how you do it, you don't. Instead you just randomly take a bunch of things. And, in fact, typically you might take clones that give you six or eight-fold redundancy. You just sequence a lot of clones and then you ask the computer to reassemble it. And, in fact, all that overlap is very good for being able to stick these pieces together.

Sometimes people do such things as take pieces that might be four thousand bases long and sequence a thousand bases here and a thousand bases here by using a primer that starts there and a primer that starts there. And then you can get DNA sequences from two ends of a clone. And if you had that for zillions of clones your computer program might do an even better job of linking things up. It's one very big crossword puzzle of putting together all of these pieces, a jigsaw puzzle of putting together all these pieces. But, in effect, this is how you sequence a big piece of DNA.

You chop it up into medium-sized pieces of DNA and then tiny pieces of DNA, you sequence them, and you use computational science to reassemble it. Some people, for some genomes, take the whole big genome and immediately go to lots of little pieces. That can work, too. It depends on exactly how complicated your genome is. In the human genome, there are some parts of a human genome that are almost identical that might be like 99.91% identical in two different parts of the genome.

And so, if you do that, you may have trouble telling those pieces apart. So, for really complicated genomes people like sometimes breaking it up into intermediate-sized pieces. But basically the idea of sequencing a big piece of DNA by this process is referred to as shotgun sequencing. Shotgun sequencing, in fact, was developed in about 1980 by Fred Sanger, the same guy who developed the DNA sequencing technique that I told you about using polymerase and dideoxynucleotides.

Sanger very quickly wanted to go from sequencing a single piece to sequencing pieces, and so he developed the shotgun technique there. And it's now been applied in many different forms of intermediate shotguns, whole genome shotguns, et cetera. So, that's in reply to the question someone asked last time about, well, how would you do a whole genome? And, as a matter of fact, this is not theoretical because, in fact, people do whole genomes this way. And we do this at MIT.

Lots of genomes get done here in this fashion. Someone else asked how would you analyze your clone. And, again, I'll just make a brief remark on that in response to the question. So, analyzing some DNA sequence. So, suppose we got some DNA sequence, A-A-T-A, don't bother writing this down. I'm just making up letters here.

How would we make any sense of it? Suppose I give you the ones and zeros from your hard-drive, how would you make any sense out of them? This is about as interesting as the ones as yours from your hard-drive, right? It's got four letters, not two, but this is actually what you get out of any project.

You want to sequence beta-globin? You'll get something like this. You want to sequence the arginine gene? You want to sequence the human genome? You get a very long string of four letters. What do you do with it? Oh, well, you could compare it to a normal copy of the gene. And if I did that I might find a bunch of differences. But how would I even know where the beta-globin gene was within this sequence? This clone contains beta-globin. How would I even find the exons? Yes?

Or whatever? Look at codons. So, let's start looking at the codons. This codon here? Well, or this, maybe it's this codon here. Sorry? Find it. Do you see any start codons here? Oh, there's an ATG there. So, maybe that's the start codon or maybe not. How often do we expect to find an ATG in some reading frame?

You know, it could happen fairly easily. Also, how do we know it's going this way? Maybe we should look for an ATG, we'll put it there, going this way. Sorry? I drew the arrow there? Well, that's because it's where the sequence started out on my page. It doesn't tell me my gene runs that way. Yes? From five prime to three prime.

Ah, but it's a double-stranded piece of DNA. You see, if it's five prime to three prime on this strand, the genome has a, another strand that reads the other way. What did I get? C-A-T-A, right, C-C-T, et cetera. And the gene could be encoded on this strand. This could be the coding strand, that could be the coding strand, and looking for a mere ATG in one of three possible reading frames on one of two possible strands I'll find all sorts of stuff.

So, sorry? Guess. Guess. Guess is good. They don't, don't, won't you, you remember we talked about getting papers accepted. If you were to write up the paper that way the reviewing would probably ding it and say, you know, the guess isn't, isn't good enough. So, that's actually very interesting. How do you actually find the gene sequence?

Well, it turns out to be a non-trivial problem which often gets glossed over in the textbooks. What you might do is if something really were

exonic, if this were any exon, does it have any properties that you can think of? It shouldn't have a stop codon. No stop codon. How often does a stop codon occur at random in a given reading frame? How many stop codons are there? Three out of 64 possible codons.

There's about one in 20 codons in any given reading frame is a stop codon. So, that means if I read for about 20 codons, and I don't encounter a stop, it's beginning to get more likely that that's not random. If I read for, say, 60 codons, 180 bases and I've encountered no stop codon in that reading frame, that chances of that occurring is about either the minus three or so, right? Because if I went through three characteristic lengths, either the minus three, you know, and I don't know, about 5% or something like that.

If I went for thousands of bases without any stop codon, would you be impressed? That's pretty impressive. So, all I have to do is find the few thousands of basis with no stop codon. The problem with that is that in bacteria there are some genes that are a thousand bases long and you, there, you can read them and they have no stop codon.

What's the problem with the human genome? Introns. It turns out that because the coding sequences are broken up into small exons, if I found a thousand bases with no stop codons then it's very likely coding sequence. But a typical human exon is on the order of a 150 to 200 bases.

Very inconvenient because, you know, it's a typical exon encodes 50, 60, 70 codons. So, it turns out that even that is not so easy to do. Well, the answer is it's not a trivial problem. People do all sorts of things to figure out how to decode sequences of genomes. You do run computer filters across there that say, look, there are a bunch of consecutive codons without stop codons.

There tend to be little preferences, like amongst the synonymous choices of stop codons, humans tend to prefer one stop codon, one codon for a specific amino acid over others. So, there are some biases as to which codons get used. And the computer can kind of take a little bit of account of that. Then you can also have made a library of cDNAs and sequence cDNA, the mRNA which will help you a lot and look for where they match up. Then you can take sequences from the human and the mouse.

And it turns out that the sequences in the mouse and the sequences in the human, if you line them up, the exons tend to match up better

than the introns because evolution cares a lot about the exons. But it turns out this is not a trivial problem. And even today, if I give you a random stretch of human DNA, it's not, there is no simple computer program that it's on, that on its own, not even a complicated computer program, but on its own would be able, without auxiliary data, to accurately pick out all the genes.

Even for simple bacteria, we cannot nail perfectly all the genes. Although, the lack of these introns means that the exons tend to be pretty big, it means the coding are pretty big and we can kind of do it. So, I just wanted to point out that, that there's a lot still to be done there. The cell manages, thank you, to read this just fine, but we're not as smart yet as the cell, and so we're not totally able to read out all this stuff. We'll come back to genomics in a, in a further lecture. Yes? Yeah, wouldn't --

What a cool idea. Yes. There are actually some experiments, which maybe if you remind me we can, I can work it into a subsequent lecture, but people have some experiments where they can randomly mutagenize zillions of bacteria and determine which ones will grow and which ones won't. And they can do it all in parallel in a single test-tube. And thereby you can tell which, which nucleotides in the genome matter and which don't. It's a kind of cool procedure. OK. Anyway, I just wanted to sort of tie up that bit here. Now, let's move on to re-sequencing a gene.

So, let's suppose we've managed to sequence, I don't know, the human genome, the entirety of the human genome we have before us, OK? Actually, next week in the journal Nature will appear a paper reporting, in fact, today, yester, no, yesterday, in fact, it was yesterday, yesterday appeared in the journal Nature a paper reporting the finished sequence of the human genome.  
<http://www.nature.com/nature/journal/v431/n7011/pdf/nature03001.pdf>

And so, anybody who wants to go online, in fact, we can get, we'll get copies for the class. Why don't we get you copies of the paper? It's not as long as the last one. But, in fact, I didn't realize it was yesterday. I thought it was next week. Yesterday came out the final report on the finished sequence of the human genome, which a number of us have been laboring on for quite a long time. And it just appeared.

So, we actually have that now. It actually, it's been on the Web for a while, but the paper describing it took a while to write up and it came out yesterday. So, we'll get you a copy of that. But now you've got that whole sequence of the human genome here. I've been, you know, I've been working on this paper with people for so long that, you know, I hadn't actually paid attention to the fact that it just came out. You don't want to know how long it took to write this. The paper, actually, is unusual.

It's the only paper that Nature has ever published where the author list is sufficiently long that we don't have it in the Journal. There's a website that contains the author list. There are, I believe, I don't have the final count, but something in the neighborhood of about 200 authors to the paper. We decided that everybody who'd worked on it should be a co-author of the paper, and we just put it all on a website. So, anyway, I digress. So, suppose we have the beta-globin gene here. So, I've got that in --

I've got the normal form of the beta-globin gene, or I've got one person's form, in any case, in the human genome sequence. Now I want to take a patient with sickle cell anemia and I want to re-sequence their gene. Now, remember what we said, we would, we would make a library from that person, right? So, we'd get that person's blood, we'd purify DNA, we'd cut it, we'd clone it, we'd probe the library with a radioactive probe for the beta-globin gene, we'd pull out the gene and we'd re-sequence it. Suppose we wanted to do that to a hundred patients.

For every patient we'd get blood, we'd make DNA, we'd clone in a plasmid, we'd make a whole plasmid library, plate it out on filter, probe it with a radioactive probe, pull out the clone and sequence it. Now, for any such library you probably need to look through a couple hundred thousand clones to find beta-globin. So, for your DNA and your DNA and your DNA and your DNA, we're going to make libraries of a hundred thousand clones, that's a couple, that's a lot of plates, right?

We're going to put them all on, on nylon filters in these Seal-a-Meal bags with these radioactive probes, and we're going to look for your beta-globin clone, your beta-globin clone, your beta-globin clone, your beta-globin clone, et cetera, et cetera, et cetera. This is really boring. Do you realize how off putting it would be to study sickle cell anemia if we had to do that for each successive patient, make a whole library?

But that was what you had to do in molecular biology because that was how you got the gene.

You build a whole library, you withdraw it from the library. However, if you wanted to do this, could you manage to get the beta-globin sequence from your genome without having to make the whole library? It turns out, and I know it's been covered at least in some of the sections, there's a cool technique to do that. And what is that technique? PCR.

So, it turns out that the next really great advance in molecular biology was the technique of PCR. And what PCR was a way, is, is a way to obtain a piece of DNA corresponding to an already known gene, you have to already know the gene, and what it allows you to do is then obtain that piece of DNA based on knowing at least some of its sequence.

It allows you to amplify just that DNA from a, from any individual. So, as compared to the experiment where I make a library for you and a library for you and a library from you and a library from you, each of which could take a month, PCR would allow us to do it in principle in five minutes.

And, actually, there are machines that would let you do it in five minutes. So, let's discuss how this PCR works. Nobody uses the five minute machines because you usually will then wait an hour or so, but anyway. Suppose I take my DNA sequence here from the human genome.

Five prime to three prime. Five prime to three prime. This sequence here beta-globin. I want to obtain that sequence. The first thing I do is I'm going to heat my DNA sample to maybe 97 degrees Celsius to denature.

Denaturing means, of course, breaking the hydrogen bonds that separate the two strands so that the strands come apart, five prime to three prime, five prime to three prime. Now, what I then do is I take a specific DNA primer matching this stretch just before the beta-globin gene starts.

Or just before where I'm interested in. How do I make a primer that matches just that sequence? I order, well, how do I know what to order? I know the sequence, right?

I've got the sequence already. I just look at it and I say I want that sequence. And then how do I get it? I order it. I type it into the Web and the machine will synthesize me this, this primer. Typically a 20-base stretch will suffice. So, I'll get me a twentymer, a 20 base oligonucleotide complementary to the sequence on this side of the gene. What I'm also going to do is the same thing over here.

I'm going to get a second primer. This is primer number one. This is primer number two. OK? Now, let's see. Five prime. This is five prime, five prime. Now what I'd like to do is add polymerase, I'd like to add dNTPs. So, plus DNA polymerase plus dNTPs.

And what will happen? Polymerase will come along and start copying my DNA, but it will only copy it starting from the primers. Now, this will keep going, of course, but DNA polymerase doesn't go forever, you know, the reactions sort of stops at some point. And so you'll get a strand going off here and a strand going off there.

Now, notice what I've done. I started with an entire human genome, and the number of copies of beta-globin was one per genome. When I'm done with this process, how many copies, how many double-stranded copies of beta-globin do I have? Two. That's still very little, but it's more than I had before.

So, what do I do next? Repeat. So, let's heat up that sample again. We'll denature at 97 degrees, and now we have our initial strand here, we have our strand that came off this primer that runs to here and maybe goes forward, we have this strand here.

We have this strand here. And this was five prime, five prime, five prime, five prime. Now what do we do? We repeat. We'll take our primer, this is primer number one, let's see.

It matches over there. Primer number two over here. Number one over here. Number two. Have I got this right? Yes. Good. Then where does this guy stop? Right at the end where my other primer was. This guy runs along here.

That guy stops right at the end. That guy might go a little further. How many copies of the beta-globin gene do I have now? Four. Two of which, by the way, perfectly sit between my pink primers. What's going to happen if I do this again? How many copies will I get? Eight, six of which will sit perfectly and two might be a little ragged as to where they go.

So, initially, after cycle number zero, that is initial conditions, the number of copies relative to the genome was one. After one cycle it's two. After two cycles it's four. After N cycles it's two to the N copies. Is that clear how the PCR works?

And that on every round you're doubling. And, with the exception of those two white things that go off to the side, they're going back and forth and back and forth between the two primers you chose to put in. What is, ah, when N equals ten, what do you got? A thousand copies. What happens when N equals 20? A million copies.

What is the copy number of beta-globin? Beta, let's suppose beta-globin, for the sake of the argument, sake of argument is about one thousand bases. What fraction of the human genome does beta-globin represent? Yeah, about a millionth for the genome. No, actually, one three millionth, but we'll call it a millionth. So, after I've made a million-fold amplification of beta-globin, beta-globin now represents half of the stuff that's in the tube.

What would happen if I go another ten rounds? How many copies do I have? A billion copies. So, in other words, I started with something that was only present at about one one-millionth of what was in my test tube. If I could make a billion copies of that specific molecule, now it so dominates the mixture that it is a thousand times more abundant than the rest of the genome.

It works. That's the remarkable thing, this works. Any questions about the technique? Now, yes? Well, I need two primers in their sequence. How many copies do I need of each of those primers? Well, I, I obviously need a lot of copies of those primers.

So, primer number one, it's a single sequence, but when I order it from the company, I'm going to order me a boat load, a lot of that primer. So, I'm adding, I better add a billion molecules of that primer because I'm going to make a billion copies starting from such primers. But if I have a billion copies of, of number one and a billion copies of number two and, you know, these days, billions aren't such big, you know, molecules are Avogadro's number and all that. It's not hard to get things.

So, you throw in huge excess, a massive excess of primer number one, a massive excess of primer number two, and you just do this. Now, I mean, what does it cost to make such a massive excess of a

primer? It's about ten cents a base, so it's two bucks, two bucks per primer give or take. You know, so I can get you a better price if you want, but, you know, anyway. It's not a bad price to, to buy primer. So, you can just go out and order a pair of primers.

You can have them tomorrow. And then all you have to do is add the primers to the, so I take DNA. Do I, I, I need DNA from you. It turns out I could draw your blood and purify DNA and all that. But it turns out that if all I wanted to do was amplify one locus, I could actually take a Popsicle stick and ask you to scrape the inside of your cheek. That'll get enough cells off from the inside of your cheek, stick it in a test-tube, and it'll actually have enough DNA there.

It turns out this is a very sensitive and powerful technique, so, but before we get to that notice what we had to do. We had to heat our DNA to 97 degrees and add polymerase. Then we heat again to 97, add polymerase, heat again to 97, add polymerase. Why do I have to keep adding polymerase? Because polymerase gets ruined at 97 degrees so it's denatured. So, the nuisance about PCR is I have to go to my Eppendorf plastic tube, pop open the lid, stick in some DNA polymerase, close it up, stick it back in a heating bath, let it go for a while, take it out, pop it open, add some more polymerase, put it back in the heating bath, pop it out.

And this is actually the way primitive scientists did PCR not so long ago, OK? Wouldn't it be cool if we could engineer a DNA polymerase that didn't denature at 97 degrees? Because then what we could do is just add the polymerase, close up the tube, put it in a machine that goes heat, cool, heat, cool, heat, cool, heat, cool, but you would have, so how do we, what kind of clever biological engineering do we use to modify polymerase so it won't denature at 97 degrees?

Yes? Get it from a bacteria. What kind of a bacteria would you ask for an enzyme that could work in, in basically boiling water? Bacteria that basically live in boiling water. Where would you look for such?

Thermal vents. You'll, geysers, things like that. Life lives everywhere. What you go is you find yourself a bacterium, so you find bacteria that lived in geysers or in thermal vents and you purify their DNA polymerase. The most famous one comes from the organism, the bacterium called *thermos aquaticus*, *aquaticus*.

Which of course means hot water, right? That's what the bacteria is called, *thermos aquaticus*. And, or, and its enzyme is called Taq, Taq.

So, we'll refer to it often, Taq polymerase, meaning from this bacteria *thermos aquaticus*, OK? So, that's Taq. So, it turns out that you can do this now without having to open and close the test-tubes.

Oops, I meant to put that here. How sensitive is PCR? It's very sensitive. You could do, so applications of PCR. Very versatile. First let's just re-sequence a gene.

Gene from yeast or from human. You just need, you know, any DNA sample. Get my gene, get my primers, and as I was indicating with a Popsicle stick, I don't have to have it very pure, although in a laboratory you go to the trouble of making it pure because you want it to be pure and all that. Yes?

Correct. Yeah. So, remember I was making a fuss over the accuracy of replication, right? And I said that on its own a polymerase might have an accuracy to only about ten to the minus five. So, now, what were the two mechanisms for, for repairing DNA, for proofreading DNA?

One was a built-in proofreading activity that the enzyme had. The enzyme would have put in a base, would check the base, and that actually helped by an order of magnitude or two. And some of these polymerases have a proofreading activity. But then we also discussed the mismatch repair activity that would later come along and detect mismatches. You're absolutely right, PCR is not as accurate as cells because it doesn't have that mismatch repair activity.

So, when you take a PCR product, if I were to clone, so if I were to take all the PCR product, say, from my beta-globin gene, so I'm going to take my test-tube, I'm going to add my primers and everything, I'm going to PCR, I'm going to PCR, and then I'm going to get a lot of copies of beta-globin. If I were to take that beta-globin and just directly sequence the DNA in the test-tube. Here's my pieces of beta-globin.

I can now sequence it by adding a primer and doing my fluorescent sequencing and running it on a sequencer and all that. Sorry, going the other way. I'll run a sequencing reaction. I could actually do it, and what I do it on is the whole population of a million or a billion molecules. If any one of them is wrong it's going to be swamped out by others, OK? Because I could do my sequencing reaction on the whole PCR product.

And random mistakes in one molecule or the other will still be a tiny minority of the votes at any given base, right? But suppose I were to take my PCR product, all these amplified molecules here, and suppose I were to clone them individually and I were to sequence each of those individual clones instead of sequencing a mixture of all the products.

I would, in fact, see a higher mutation rate. And you're absolutely right. When people clone PCR products they have to check them afterwards and throw out the ones that are wrong, OK? Absolutely right. Good, good, good. So, you guys are, you know, right on top of the important issues about, about DNA. So, so I can, I can take a gene and I can re-sequence it. I can also do things like take blood and look for the presence of a virus.

So, I could re-sequence beta-globin and study people and see who's got sickle cell anemia and all that. I could take blood and I might want to say do I see the HIV virus present in someone's blood? For example, HIV testing can be done by making PCR primers for the sequence of the HIV virus.

It has a genome. Taking a human's blood sample and PCR-ing it. If you get a positive PCR product, a PCR product that is made by these two primers and if, for example, you checked that it, that it gives you the HIV sequence then you know that that blood sample has, that person has the HIV virus. This is a way to do this. The PCR reaction itself is fast. Typically takes hours.

In fact, can be forced to go much more quickly by machines that rapidly thermocycle. And you can actually PCR in five minutes, although people don't do it very often, but if you put a thin glass capillary and go heat, cold, heat, cold very, very quickly, there's a machine from Idaho Technologies that can do it in five minutes, but it's usually not worth the trouble. And you just put it in and, you know, in a couple hours you'll get an answer there as to whether or not somebody has HIV, for example. So, you can do that to detect relatively low quantities of virus.

How low can you go? Well, it turns out, what's the limit? What's the smallest number of molecules you might be able to detect in a sample? Theoretically. One. You can't fewer than one molecule, right? So, one might be the limit. So, how could I arrange to have a single molecule in a test-tube?

I would like to have a test-tube that has exactly one copy of the beta-globin gene. What, how's the best, what's the best way to get exactly one copy of beta-globin and put it in the test-tube? Sorry? You can't. Why? Just one molecule. I want to get exactly one copy of beta-globin. I could, I could just take total DNA and dilute it so, on average, there's only one copy. Or, actually, is there any way to, I mean can I, I'd just like to buy a package that contains exactly one beta-globin.

Sorry? Bind it to something big. Let's think biologically. Does biology package up a single copy of beta-globin? Sorry? Gametes. How about a sperm? Let's grab a sperm by its tail here, put it in the test-tube. It's one copy of beta-globin. So, you can actually take cell sorters and have it cell sort sperm into individual test-tubes. You now know there's one copy of beta-globin.

Heat it up, it will crack open the sperm, add your primers, you can amplify beta-globin, it's a single copy. That proves its extraordinary sensitivity. You can do it with a single sperm. You can do it with a single egg also, but harder to come by. So, with that level of sensitivity, you could do the following. So, single sperm typing.

Now, single sperm typing is cool but sort of useless. What are you going to do with it, right? But here's another thing you could do. Embryo typing. Suppose someone has a genetic disease in their family, maybe it's Huntington's disease. And suppose that the individual with Huntington's disease wants to have kids. Or the individual, sorry, the individual who is at risk for Huntington's disease or breast cancer or whatever wants to have kids.

What you can do is with an in vitro fertilization clinic you're able to obtain eggs, fertilize eggs in vitro, and grow them up in a Petri plate to 8 or 16 cell stage before re-implanting embryos back in the mother.

Wouldn't it be cool if we could choose to only re-implant an embryo that did not have the genetic disease? How are we going to do that? PCR. How are we, so what do we do?

We take the embryo. We make DNA from the embryo. We do PCR and we say, ah-ha, this embryo did not have the genetic disease. Problem is it has killed the, the cells there, right, it killed the embryo. Any ideas? Pull off one cell. Remove a single cell. It turns out that at this stage the cells are not differentiated. If I remove one cell from an

embryo at that very early stage, the other cells will make a perfectly happy, healthy baby.

That cell is not necessary. This single cell sensitivity is very valuable because I can actually do single cell genotyping on in vitro fertilized embryos and be able to offer parents a chance, the opportunity to re-implant only those embryos that do not have the genetic defect. That's cool. That's really cool. There are other things you might be able to do. If you're treating a patient with cancer, a patient, a cancer patient and you've given chemotherapy you want to know have I managed to eradicate the cancer cells?

And six months later have any of the cancer cells come back? I could look for very low quantities of cancer cells. I can, I can do surveillance for low quantities of cancer cells following chemotherapy.

And, of course, I can also do forensics. I could take a small sample of blood from the scene of a crime or saliva from the back of an envelope that someone has licked, and I could do PCR and look for genetic variations that distinguish people. And, presumably, you see all that stuff on television all the time.

So, that's what PCR is good for. It's good. All right. Last topic, very brief topic, but I do want to mention. This was being able to analyze a gene by directed mutagenesis. And I won't go through the details of all this, but I just want to at least basically describe the concept.

I could take any piece of DNA, say from a drosophila, and I can mutate the DNA in vitro. I can change this base from a G to a C. There's a right, there's a proper protocol and cooking trick for doing that. It involves putting a certain oligo over it and extending, and it doesn't matter exactly how.

I could insert an extra gene into that. I could use a little restriction enzyme to open it up and stuff something in. I could delete something from this. Maybe I'll use a restriction enzyme to cut it open, et cetera. Basically I could, I can fuse genes together. I can do whatever kind of construction of pieces of DNA and modifications of pieces of DNA that I would like to do in vitro.

I can then take that mutated gene, let's say the gene is an enzyme, encodes an enzyme, and the enzyme has an active site. I could change the code for the amino acid right at the active site to see if that amino acid really matters or not. I can do any of those things.

And I can put this back in an organism. Remember that you, I said you could transform DNA back into bacteria? Well, you can also do such things as simply inject DNA into a fertilized egg.

In fact, at the stage where there's a male and a female pronucleus that haven't fused yet right after fertilization. You can take your little pipette and a needle and you can inject some of the DNA you want into the male pronucleus, and then when the male pronucleus and the female pronucleus fuse and the embryo grows it will have your DNA.

You can make mice that carry whatever gene you've modified like this. You can also not, you, you can also not just modify a piece of DNA and add, this is gene addition, you can also do gene subtraction. You can do gene subtraction and, again, I won't worry about the details here, by taking embryonic stem cells.

Much in the news these days, and we may come back to them. And in vitro, working with embryonic stem cells, to transform a piece of DNA that has been arranged to recombine into the gene of interest and know it out. So, if you build, if you build a piece of DNA in vitro and you put it into a whole bunch of embryonic stem cells you can select, by various clever techniques, for those embryonic stem cells that have taken up your gene.

And not just taken it up but slammed it into the normal locus in place of the normal locus. And that way you can knock out a gene. You can do gene knockout. So, the basic point of this now, to summarize these many lectures is we're now at the point where this picture that we saw at the, at the beginning, function, gene, protein, that we understood now first as a methodology, genetics, biochemistry.

And then we understood how genes encode proteins through molecular biology. These tools of recombinant DNA allow us to move in any direction.

You want to find the gene underlying a function, find the gene for Huntington's disease? We could do it. Clone it based solely on its linkage. You want to find the gene encoding a protein? If I know its amino acid sequence, I can find the DNA sequence that corresponds it. If I want to find what a certain protein does, its function, I could get the gene for that protein. I could knock out the gene for that protein and see what its function is.

Suddenly, for the mathematicians amongst the group, this becomes a commutative diagram, which you can chase around in any direction. That is, in a sense, what the 20th century was about, was intellectually these two disciplines merging through molecular biology and then recombinant DNA giving you all the tools that if you're sitting at any place in this triangle you can move this way and that way, from a gene to a protein, from a protein to a gene, from a function to a gene, from a function to a protein.

Much of the rest of the course we'll talk about how you use these tools, but this brings to a close this first chunk of the course about the concepts and the methodologies of molecular biology. Now, if you hang on one more minute, this is my last lecture for a while. I won't be, we're having an exam on, we have a quiz on Monday, and then Bob's taking over again. So, I won't see you for the next week or so. So, two things.

One, I won't see you before the World Series is over so everyone please think good thoughts about the Red Sox. Number two, I will not see you before the election. Vote. It's your choice who you vote for, but vote. Good-bye.