Good morning. So, we are going to see if my voice holds up through this lecture today.

It is a casualty of having been at Foxborough yesterday, and then staying up rather late watching the Red Sox game.

On the whole, both seemed to have come through successfully, but my voice is a bit of a casualty of the events.

So, we'll see. But I'm going to sound a lot scratchier than normal.

So, how many of you stayed up to the end of the game last night?

Good, excellent. I approve.

OK,last time, we spoke about the idea of cloning DNA, to create libraries of molecules.

And again, I think this is just one of the most clever inventions because it's a completely new way to think about purifying molecules.

Rather than purifying molecules, by separating them based on their biochemical properties, it's purifying molecules by diluting them into single components, and then amplifying each back up from its own source. It's really quite a beautiful idea.

And just to go over it, we take, say, human DNA, or we could take drosophila DNA, or we could take yeast DNA, or we could take any other DNA we feel like.

We cut it up in some fashion with a restriction enzyme.

We'll use our favorite restriction enzyme here, echo R1, which cuts a defying side, GAATTC. We take that. We add our insert DNA. These are referred to as inserts because they're going to be inserted into a plasmid. We take a plasmid vector.

The plasmid vector here is a naturally occurring, although sometimes modified, piece of DNA that bacteria have that take an origin of replication that allow it to grow autonomously when put in a bacterial cell, a selectable marker.

The selectable marker, for example, ampicillin resistance, or some other resistance, we add these and then we seal up the pieces of the DNA using the enzyme ligase. Ligase joins and joins producing for us molecules of this sort. We make zillions of them in parallel in one test tube. We then transform them by adding these molecules to bacterial cells that have been appropriately prepared to be transformed, that is, their membranes have been

treated in such a way that they're going to be most likely to suck up pieces of DNA. We then plate them on a plate at a density so that individual bacterial cells are well separated from each other. You try a bunch of different densities so you get one right.

And, you let them grow up. And, every colony here, as we discussed, is the descendant of a single bacterial cell, carrying ideally a single plasmid.

And, that single plasmid, we know it's carrying a single plasmid because we were clever enough to put ampicillin or other selectable marker on this plate. And so, only bacteria that have picked up the plasmid are ampicillin resistant. And there you go.

This is called a library. And, at the end of the day, you may have a library that contains one plate of clones or a library containing hundreds of plates of clones. We're going to see how we last through this. Now, a few people asked me at the end of the last lecture, well, OK, but what about the details.

Is it really going to work like this? How come some of these plasmid molecules don't automatically get closed back up by ligase? Why is it that there's always an insert in the plasmid?

What's the answer to that question? Sorry? There's not an answer because sometimes ligase might close up that molecule.

Now, that would be unfortunate because it would mean that a bunch of the things in your library just had the vector without any insert.

So, and these are details, but over the course of years, recombinant DNA specialists have worked out lots of cute tricks to make better and better libraries. I'll just give you an example of the kinds of things. Remember that in order to ligate DNA, we had a five prime here. We have a phosphate group here, three prime hydroxyl phosphate here, double strand of DNA here. We have a phosphate here. We have a hydroxyl here, phosphate five prime, three prime.

If ligase is going to come along, it turns out that ligase needs the phosphate there in order to seal it up and make a chain.

So, for example, suppose we were to arrange that the plasmid vector didn't have phosphates on its two ends. Then ligase would not be able to re-seal the plasmid vector. That's a cute trick.

This is just cooking, but I'm giving you an idea of the kind of cooking tricks we use in all this. So, ideally, you would like an enzyme that can remove phosphate groups from the end of DNA. How are you going to invent such an

enzyme?

It already exists is the answer to all these questions.

And, bacteria have such an enzyme that can remove phosphate groups.

So, just remove phosphate groups. And of course these enzymes are developed by bacteria because they need them in the course of DNA metabolism. And, what do you think the enzyme is called? Phosphotase, of course. That's what happens, use phosphotase, and you treat that, and it doesn't seal back up. Now, somebody will say to me, well, OK, but now I've got my vector here, and I don't have a phosphate on it, and so this is my vector DNA. And then, I've got my insert DNA, and sorry, my insert DNA here, it has a hydroxyl here and a phosphate here. So, the vector has no phosphate. But, when ligase wants to attach an insert, it's got a phosphate here but not here. What's going to happen? Well, it turns out that ligase will seal up this because it's got a phosphate, but it'll leave this one open. Now, is that a problem?

It turns out, if you just transform it into the bacteria with that hole there on one strand but not both strands, it's still a covalently closed circle on one of its strands. The bacteria will repair it. So, you can take advantage of the bacteria's own DNA repair mechanisms to just throw the molecule in sealed up on one strand and let its repair mechanism; all these tricks we play to our advantage.

Someone else asked after class, what happens if the gene I'm interested in studying has, here's my gene let's say that I'm interested in studying. I take human DNA. I cut it with echo R1. So, I have cut it at all the echo sites.

Well, golly, what happens if my gene happened to have an echo site in it? Then my gene's going to be cut up into two pieces.

Isn't that bad? What do I do about that?

Do I know in advance if my gene has an echo site? Well, no, I don't, because I don't know what my gene is.

I'm making a library of everything in the genome.

So, some genes will have it, and some won't. And, I might not know the gene I'm looking for. So, how do I avoid that? Sorry?

Oh, you've tried another enzyme. You've tried BAM and Hindi, and make a library with different enzymes. That's one way. That works. Another way, just to give you a sense of how fast molecular biologists are with this.

Supposed when we add echo R1 we don't let the reaction go to completion. Suppose we run the reaction under

conditions where it's somewhat inefficient, and instead of managing to cleave every echo site, on average it cleaves, say, one out of every three-echo sites. You can do that.

So, that means you can arrange just by your reaction conditions to on average randomly cleave some but not others.

And, these are called partial digestions. So, it turns out that all of the kinds of things that people were asking me about afterwards, I was very glad people were thinking about would this really work? There are tricks to get around all of it, and there's a whole fat book of protocols about if you want to make a library really, really carefully, how you would do that, how you make sure the vector doesn't re-close, how you make sure that you don't cut every site but random sites, and things like that. And, all of these rely on lots of enzymes and things that bacteria have already invented.

So, I'm just going to put these down as cooking tips.

These are not really necessarily, I don't care whether you know the details or not, rather that there exists a whole 15 years, 20 years worth of ways to make the best possible libraries.

And so, it's quite routine now to be able to make good libraries.

All right, so, having made a library, the challenge is finding your clone. How to find your clone, the clone of interest. So, I need to describe a number of ways that people have for finding a clone of interest. And here, of course, up to this point, the DNA could be zebra DNA, and it could be human DNA and yeast DNA, and it could be something that is an enzyme for arginine, or this, or that.

But now we have to be specific. So, let's suppose we go back to a problem we talked about before about, say, auxotrophy for a nutrient.

So, let's suppose that I have a bacteria, maybe even E coli itself, where I have selected mutants that are auxotrophic for arginine.

So, arginine auxotrophs will grow on rich medium, but on minimal medium they don't grow. But, they would grow if I added arginine to that medium. They don't grow because they have a mutation in a gene. We know it's a gene because we crossed together the mutant and the wild type. We show that we can define this phenotype to be a recessive phenotype.

We can map it in the yeast genome by showing it has linkage to other phenotypes. That's all great. We can do classical genetics, a la Mendel, a la Morgan, a la Sturtevant.

But, how are we going to find the gene? How are we going to, now, use our tools of recombinant DNA to get physically in our hand the piece of DNA that encodes the gene that is defective in the strand?

So, have a mutant bacteria. It can't make arginine. It can't grow in minimal medium. Somewhere in there, you know there's a mutation in the DNA sequence.

How do we find it? What should we do?

This is the whole point of recombinant DNA, to make this abstract notion of, there exists genes, they transmit all this kind of stuff, concrete. How are you going to find it? Any takers? Sorry? Run a gel.

So, I take DNA, cut it up, run a gel.

I have all the DNA from the bacteria schmeered (sic) out.

And somewhere in that schmeer is the gene. So, I take normal DNA from normal bacteria. I take mutant DNA.

One nucleotide is different in the mutant DNA. I run them out, and I assure you, they just look like a schmeer.

It's just a big schmeer of DNA. It's hard to see one nucleotide difference out of the 4 million nucleotides.

The E coli say, how are we going to get that?

This is good. We're thinking practically here.

What else? Sorry? Sorry? Cut it up. I'm assuming she wanted it cut up and run out on the gel. It still will look like a schmeer. Forget the gel. Cut it up. Make a library.

OK, so we're going to make a library. Let's assume now we have a library of different E coli cells containing individual plasmids, containing random bits of E coli. How's that going to help?

Splice it back in. How do I know if I spliced it back in?

Ooh, that's an interesting thought. Suppose I were to make my library using wild type DNA, DNA from the wild type strain.

So, I'm going to make a library containing lots and lots of fragments of normal E coli DNA. This is my library. I'm going to transform it into, what kind of bacteria should I transform it into, wild type or mutant?

Who votes mutant? Who votes wild type?

We'll go with mutant, then. Mutant. We'll put it in mutant. So now, all of these mutant cells, each one is going to suck up a plasmid. We then are going to plate this, and let colonies grow up. One of these colonies contained, so this mutant is arge minus. And, one of these colonies is going to contain the ARG plus gene here. How are we going to know which one?

Sorry? How are we going to know which one has the arge plus gene?

Yes? So, plate it on minimal medium. If I plate it on minimal medium, what will happen to most of my mutant bacteria?

They're not going to grow. But, what's going to happen to the bacteria that happens to be lucky enough to have picked up the plasmid that contains the ARG plus gene? It'll grow. So, whatever grows on minimal medium has been rescued. In fact, we've complemented the defect. Remember, we talked about complementation tests? In a way, it would be the plasmid is complementing the defect. Bingo, that's it. So, we can actually find that gene functionally.

We plate on minimal median, and we look for growth. The only things that will grow have been rescued. So, this is called cloning by complementation because we are complementing the defect in this strand. All right. So, any time I have a functional defect in my bacteria, I can find the gene for that functional defect by simply taking a total library for normal from wild type bacteria, transforming it into a mutant bacteria, and looking for rich bacteria has suddenly been rescued. Then I'll purify that bacterium, and I'll purify out the plasmid. And that plasmid will contain the DNA for the gene. That's pretty cool.

Let's try another one. Suppose, yes? OK, great.

I've got my plate here, and I've said only one of these bacteria will grow. It's the one that happens to have within it the plasmid containing the ARG gene. And, you're fine with that, but you're saying, but how would I get that plasmid back out of the bacteria because the bacteria's got its own chromosome, and I'm making this big deal about how we purified stuff away from all this other DNA.

But, I've thrown this plasmid back into a bacteria that has all its chromosomal DNA. So, who am I kidding?

How are we going to purify out just that plasmid? If I could purify the plasmid, it would be OK right? It turns out I can. Plasmids are little circles of DNA. Chromosomes are big pieces of DNA.

It turns out that the coiling of the plasmid as a little circle gives it different densities and different physical chemical properties to big chunks of DNA which get broken up. And so, there are a bunch of tricks that allow me to get a pretty high purification of a plasmid away from chromosomal DNA based on the different physical properties of a

small circle versus big chromosome. But, good question. Otherwise, how would I get that plasmid out? But it turns out, you can purify plasmids. Good question. OK, so now, let's try another one.

Next cloning expedition: we're going to go to the library, and we want to withdraw a volume from the library.

And, I want now, instead of bacteria that can't make arginine, let's go with human DNA. Let's try human DNA. And, I would like you to now please find the gene that encodes beta-globin.

Beta globin, of course, is one of the two proteins in hemoglobin.

Hemoglobin is a tetramer. It has alpha-globin and beta-globin.

This tetramer is the oxygen carrier in your blood.

It carriers oxygen. Beta-globin happens to be the site of some very important mutations. We know that sickle cell anemia is caused by mutations in beta-globin. We know that diseases like thalassemia are caused by mutations in beta-globin.

And, people knew this before they had recombinant DNA because they could study red blood cells. There's lots of beta-globin in red blood cells. They could see that something was funny about the protein. They could even see that in sickle cell anemia the protein had a different net charge, and it would run differently.

So, they knew something was funny with the beta globin protein.

All I want you to do now is clone beta-globin for me.

Could we do the same thing? Why not?

Bacteria don't make beta-globin. So, what can we do? Well, we could make a library of human DNA. And, we could throw it into the bacteria. So, why don't we just select for a bacteria that makes beta-globin? Could we do that?

I don't know, how? Do you see how? How would we select for that? I mean, there, we could see who grows without arginine. But how are we going to tell which bacteria has picked up beta-globin? I don't know. Yeah? Use mammals. We could take a mouse that did not make beta globin, a mouse that had, say, thalassemia, isolate a naturally occurring mouse with a defect in beta-globin. Then, do injections of plasmids into mouse eggs, grow up the mouse eggs by implanting them back into pseudo-pregnant females, do this for 108 individual plasmids with 108 individual mice, and look for the mouse that is rescued. Intellectually, you're absolutely right, it works. So, that's exactly the cloning by complementation we talked about for bacteria, and you're dead-on right. That would

work. Getting it funded is another matter because it's a hugely expensive experiment to shoot up each egg with this, but it could work. So, we need another solution because we can't rescue the function in mice because it's just not practical to do so.

Of course, if we could do this in mouse cells, maybe we could make it work in cell culture in mice. But, let's suppose we don't have a cell culture phenotype. We just have an organism phenotype.

So, it's not going to work to just do this by complementation.

But, good thinking guys. This is good. So, next trick we might have at our disposal is suppose because beta-globin is so abundant in red blood cells we have purified beta-globin, and we've done amino acid sequencing of the protein. By end degradation, you can work out the sequence of globin. And, you can learn that beta-globin has, here at its amino terminal, val, leu, ser, pro, ala, asp, lys, threonine dot, dot, dot, dot, dot off to the carboxy terminal, OK?

If I knew that this was the amino acid sequence of the beginning, just the beginning of beta-globin, couldn't I figure out what that initial portion of the DNA sequence must be?

Wouldn't this give me a clue? If I knew a little bit of the protein sequence, wouldn't this give me a clue about the nucleotide sequence that must be there in the human genome to encode this protein? So, a biochemist has purified the protein. Biochemists have studied the protein well enough to know some of its amino acid sequence. Can I infer the DNA sequence from the amino acid sequence, or at least a little snippet of it?

Sorry? Multiple possibilities, but an infinite number? No. Why do you encode valine? Well, GT something; something could be actually A, T, C, or G. What about luecine. Well, it's either a T and a C, or is T in the first place? There's always a T there.

There you go, and it can be either of those. There's a T, C, anything, or an A, G, and a T, or a C.

Here, we have C, C anything. Here we have a G, C anything. We have a G, A, T, or a C. For leucine it's an A, an A, either an A or a G. Here, it's an A, a C, an anything. Here, it's an A, an A, a T, or a C, here a G, a T, anything, an A, an A, A or a G. You're right. There are multiple possibilities. But, it's not an infinite number, right? There are certain possible DNA sequences that might be encoded here. If I just work it out, it's either two choices here. There are four choices here. There's two choices here.

There's four choices here. There's two choices here, two choices, etc. If I just look at, let's take a segment of this. Let's try one, two, three, these six amino acids. Four choices here, how many possible DNA sequences could encode these six amino acids in this order? Four times four times two times two times four times two, what is that?

256, let's see, two, two, to the two, to the four, to the five, to the six, to the seven, eight, 512. I think it's about 512 possibilities.

So, 512 possible nucleotide sequences could work here.

Well, 512's not infinite. There's 18 bases of sequence, 512 possible 18 base long nucleotide sequences.

Just suppose that you knew which one it was. Now, you have to suspend your disbelief for a second. I'm not going to tell you how you might know, but suppose you knew which of the 512 it was.

OK, could we use that little fact of knowing a stretch from about 18 bases of the sequence to find the clone?

How could we find that clone in our library that has that 18 bases of sequence? Google. [LAUGHTER] And, of course, you are totally right because as we'll come back to, that is the way you would do it today if it's the human genome because the entire sequence of the human genome's on the web.

But, you might have an organism where it's not on the web.

But, we'll come back because, of course, the human genome project changes everything as to how you would approach this.

Google is how you would do it today. But, in the absence of Google or the absence of the entire sequence of the human genome, but I'm glad you raise it because it's absolutely right, how could I find the clone that has that specific 18 base pair sequence?

Who has my 18 base sequence. Well, here's a trick.

I could chemically synthesize an oligonucleotide that matches my sequence: an 18 base pair long ologonucleotide encoding my sequence.

What I'd like to do is use this ologonucleotide as a chemical probe to wash over my library. And, by washing it over my library, I'd like to see where it sticks. Now, that's kind of interesting.

What do I mean by that? What I'd really like to do would be to kind of crack open all the cells of my library, and then the DNA would be sitting there. And, I'd like to take my ologonucleotide probe for a little snippet of the gene and wash it over the library. And then, by the amazing powers of Crick and Watson base pairing, it should stick to the right place.

Could it do that? Turns out DNA, given time to wash around, will stick to its own complement. So that's the idea.

How in the world do I do this in practice? So, here's what you do in practice.

In practice, let us grow our bacteria. Let's plate the bacteria on an agar plate on which we have put a membrane a nitrocellulose filter or some other kind of filter.

Just imagine it being a piece of filter paper. And, I'm going to plate my bacteria on the filter paper that's here.

I'll let them grow up because there's nutrients here.

The nutrients diffuse through the filter paper. And then, I have a piece of filter paper that I can pick up with my tweezers, and on that filter paper are bacterial colonies growing.

So, this is a filter. Then, what I'm going to do is I'm going to take this filter with these glistening bacterial colonies, and I'm going to stick it in the autoclave.

And, I'm going to heat it up in the presence of wet heat, and the bacterial cells will crack open. And, under these conditions, the DNA will tend to stick to the filter because I've picked the filter that the DNA tends to stick to. And, I'm going to wash this filter in a certain way that all the usual junk, some of the proteins and cell surface junk washes off. And, the DNA from each bacterial colony will stick. So now, I have the DNA from each colony sticking to that spot. Then, what I'm going to do is I'm going to take my filter and I'm going to add my ologoprobe.

This thing is now called a probe. I'm going to add the probe to the filter, and I'm going to put this in a, I need some sort of a hybridization device in which the probe and the ologonucleotide and a little water can swish around. And here, we use a technical device called a baggy, or some other kind of, basically, a Ziploc bag or you can heat seal it or something like a freeze meal. In fact that's actually what's used in the lab is Freeze-a-Meal. You get these Freeze-a-Meal bags, you toss your filter in, you squirt a little bit of your probe in, and you put it in the Freeze-a-Meal bag, and then you put it in a water bath. And, it switches back and forth.

And, the probe just goes washing all over the place.

And, wherever the probe finds its corresponding cognate sequence by Crick and Watson, it'll stick. And there you go.

That clone contains your sequence. Now, we have a few problems here, don't we? What are some of the problems with this? Yeah?

Sorry, what if it sticks what? So, the probe, I thought this filter likes DNA. So, why won't the probe just stick nonspecifically everywhere? We treat it in some way so that after we've got the DNA adhering to it it's now not going to stick everywhere. Good, next problem. Well, even before that, yes? No, we'll take the whole library.

We've gotten the library scattered out on this filter.

Good, so hang on to that one for a second. First off, do we even know where that clone is? How did we know where the piece of DNA stuck? I mean, I drew it as red. But, how do we know where that red spot is? Yeah? Oh yeah, you see the problem is if I just wash it over there, unless you have, you know, Superman vision, you're not going to know where that probe is. So, you're proposing, the first thing we better do is radioactively label the probe.

So, let's put a radioactive label on the probe, OK?

Radio label, and it turns out you can radio label probes by using these enzymes that can add a radioactive phosphate group, etc. So, now, when it's radioactive, we put it here.

And now we have a radioactive signal here. How are we going to find our radioactive signal? We put it up against x-ray films.

We take our filter. We dry it off.

We slap it onto a piece of x-ray film. We let it expose overnight.

We develop the x-ray film. And, we'll see a black dot.

We'd better actually have taken some care to take a little radioactive pen and make a couple of fiducial marks around the corners.

Otherwise, we're not going to know where our black dot corresponds to.

But, assume we've made a couple of dots and we know how to line up our x-ray film to our filter. Now, we go back to our filter.

We say, uh-huh, there is a black dot corresponding to the location of the radioactive probe right there.

That was, as you said, where the colony used to be that we wished we still had [LAUGHTER] because we cooked it in the autoclave, which is too bad. So, what should we do about that?

Yep? So, if I did it one colony at a time, I would know exactly which one it came from. But, it could take a long time.

Sorry? So, plate it first onto a plate of agar.

Take a filter, and press the filter up against the plate and make a copy of it. Replicaplate (sic) that. It turns out, that'll work. There are two different approaches and both of you were right. One approach is to replicaplate it.

Plate it first on a normal plate, and lay a piece of filter on top of it, and a little bacteria will stick in the same patterns.

Peel it off, and you now have it. Alternatively, now in the presence of robotics, you can use a robot to take these colonies into microtiter plates, and you can screen the individual wells by stamping them onto a filter, things like that.

And frankly, that's how we do it now. If you want to screen the human genome, at least set up a library with a few tens of thousands or hundreds of thousands such things. And, we can read off from a grid which one it was, and we go back to our master microtiter plates where we have. But, either way, we need to have a living copy of the library. But, that's how you do it.

So now, we're in business. We have a living copy of the library. We make a filter containing that.

We cook the filter in the autoclave. We add a radioactive probe.

Wherever it sticks, it matches by the wonders of Crick-Watson base pairing. We're in business. Yes? So now, there was this issue.

I mean, how do I know that that sequence doesn't appear multiple times in the human genome? That's one issue. So, I'm going to have to pull out each of the positive hits I get and check it out.

I'm going to have to analyze the clone because just knowing that it hybridized to that might not tell me it's the beta-globin gene, but at least it's probably a good start, right? I've narrowed it down.

But, yes? Wait a second, right. We said there were 512 possibilities, and I said, bear with me, let's suppose we knew which one it was and we used it. Well, how are we going to know which one it is? Well, we could do the experiment 512 times, and one of them would work. That's lousy.

We could go and make 512 ologotes and simultaneously throw them in the same seal-a-meal bag. That actually works.

How do you make 512 ologotes? How do you make an ologote, by the way? To make an ologonucleotide, there's very fancy chemistry that's been developed, which someone won a Nobel Prize.

Nowadays, of course, if you need an ologote made, how do you do it?

Go to the catalog, that's right. In fact, you can go on the web, type in the sequence you want, and there's a machine that will make it. You can have it tomorrow. So, it turns out, that's how you make ologonucleotides today. There are good machines for it.

And, it turns out that if you wanted to, so what you do is you type into the computer the following. You type in, please make me an ologote that starts, put a C in the first position, a C in the second position. And, what are you going to put in the third position? Just tell the computer to put in a random mix of all four. Then, a G in this position, a C in that position, and then a random mix of all four.

Then, put in a G and an A, and then put in a 50/50 mix of T and A. In fact, in one synthesis, by telling the computer to just add a mixture at certain steps, it'll simultaneously synthesize a mixture of all 512 possibilities for you. So actually, a single synthesis will suffice to get a mixture of 512. You take your mixture of 512, wash it over the filter, etc. Now, your point still stands. How do we know that there's not something else in the genome that has this, etc.? But at least we can find all the specific positives associated with this, and we can analyze them further as we'll talk about next time more about how you actually analyze them.

And, of course, whether 18 is the right number of bases, or you might prefer to have a longer probe or shorter probes, or two probes, these are all the cooking tips molecular biologists worry about. But, given a sequence of an amino acid sequence, you can infer, although with redundancy, a nucleotide sequence. Given a nucleotide sequence, you can make an ologonucleotide probe. Given a nucleotide probe, you can wash it over the filter. You can find the colonies that have it, and therefore you could clone by hybridization.

So, we'll call this one cloning by hybridization, or cloning by sequence. OK, now, there are other ways to do it, or by sequence here. Of course, as someone correctly noted, if the entire sequence of the human genome has been already sequenced as it has right now, if you knew the amino acid sequence, you could do this hybridization not using filters and radioactive probes, but just doing it in silico. You can do it in the computer, and that will work as well. So now, let's do the next one. Last cloning expedition: I'd like to clone the gene for Huntington's disease or cystic fibrosis or something like that. Cloning a disease gene, such as Huntington's disease, is a dominantly inherited disorder passed to some of the offspring, causes a brain degeneration that onsets typically in the fifth decade of life.

Let's clone that gene. Can we do it by method number one, cloning by complementation? No, because we don't have a bacteria that has Huntington's disease. We don't have mice that have Huntington's disease. And, we can't certainly shoot up people and try to rescue the phenotype and all that.

That's not going to work. Number two, how about doing it by number two? Let's just get the protein for Huntington's disease, get its amino acid sequence, and then find its nucleotide sequence. Pretty good. What's the

protein for Huntington's disease? Huntase. No, it's actually called Huntington it turns out. But, at the time that people went off trying to find the gene for Huntington's disease, I'm afraid they didn't know. They had no idea what the gene was that caused Huntington's disease. That was the point. They wanted to use molecular biology to find the gene when they didn't even know the protein. So, we can't use our method number two. So, how are we going to find it? The disease does lead to degeneration of nervous cells. Study nerve cells. So, we could take brain biopsies from patients who have died of Huntington's disease, and people did that. But, nerve cells that die, a lot of stuff goes on. All sorts of proteins go wrong, and it's stuff. The problem with studying tissue from people who have a disease is that it's diseased tissue.

And, just because you see something wrong doesn't mean it's a cause rather than the effect of the disease. That's why we really want to find the gene and find its mutation because we know then that's the primary cause. But, how are we going to do that?

We don't know its sequence. We can't rescue it by complementation.

As a pure geneticist, what can we do?

Yeah, we know the sequence of the human genome. So, we just sequence the entirety of the genome of somebody with Huntington's disease and compare it to normal. That actually may become a reasonable way to do things, but the first sequence of the human genome costs a couple of billion dollars. Doing it again would be cheaper. We'd spend about $30 million or so, but it's pricey. Also, there would be a lot of genetic variation, just random, meaningless polymorphism between individuals. The human genome differs between any two people by about one letter or 1, 00. So, we would see about 3 million differences between the person with Huntington's and the wild type reference sequence on Google. We wouldn't know which one causes it. Suppose you have a family tree. How could we use it?

Compare the children and the parents.

That's all right. What does a geneticist do with a family tree? What did Sturtevant teach us: genetic mapping.

Suppose we were to study a family tree of individuals with Huntington's disease. And suppose on the chromosome where the Huntington's disease gene lives, we were to look at genetic markers.

Could we do genetic linkage analysis?

Genetic linkage analysis that would allow us to know that there was a marker here, some kind of a marker, a DNA marker, a DNA variation that was co-inherited with that showed linkage with Huntington's disease?

We could do that just by finding that across a family, there tended to be very little genetic recombination between this marker and Huntington's disease. Now, how would we know to look here?

You wouldn't. We'd try markers all over the genome.

Next chromosome, next chromosome; if we tried genetic variations all over the human genome, we would eventually find that some genetic markers in the human genome tended to be co-inherited along with Huntington's disease. It turns out that that's enough.

This will tell us approximately where this unknown gene must live.

Here's a portion of the chromosome where the unknown Huntington's disease gene lives. Here's a genetic variant, and here's a genetic variant, a marker, that shows correlation.

Maybe there's only 1% recombination here, and 1% recombination here.

And, that's the powerful thing about Sturtevant's idea.

It works in fruit flies. It works in humans. If I have any genetic variation and it's 99% correlated, or only recombines 1% of the time, it tells me that this unknown gene must be nearby.

So, I could use this genetic marker as a DNA probe to wash over a library to get a big piece of DNA from this region.

I can take this piece of DNA and use it as a probe, a radioactive probe, to get an overlapping piece of DNA.

I can use the end of this DNA as a probe to wash over a library and get the next piece of DNA. And, I can do the same thing here.

Once I have any piece of DNA that's even vaguely in the neighborhood, I can use it as a probe to wash over a library and get a piece of DNA, use it to get the next piece, the next piece, the next piece, in a process that was called chromosomal walking.

That gives me a series of clones that I know must cover the region for this unknown gene. I then begin to analyze them and I say, let's look at some more genetic markers, a genetic marker a little closer and a little closer and a little closer.

Which ones show perfect correlation with Huntington's disease?

And, that narrows me down to a small number of clones that must contain the gene, even though I had no idea in

advance what that gene was.

This is called cloning by position. And, that's a very powerful technique of genetics because you don't need to know in advance what's wrong with a diseased gene. You first figure out where it is, and then you get the clones to figure out what it is. So, this actually works. Now, the process of getting the next piece, and the next clone, and the next clone, is unbelievably boring and tedious. And, for Huntington's disease, this process took nine years. Of course, now, how would you do it?

Go to the web because with all of this process of the human genome, you've got all these clones laid out already. And so, the work that used to take years now is, once you have a genetic marker that's close to Huntington's you can just look up all the clones in the neighborhood and actually all the sequences in the neighborhood.

So, this process has gone from nine years to, if you have do this again, you could get that region for Huntington's disease in a couple weeks. Now the question is, how do you analyze that region?

How do you know what's in that region? How do you know what the genes are that are in that region? And that's what we'll talk about next time.