

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

7.014 Introductory Biology, Spring 2005

Please use the following citation format:

Penny Chisholm, Graham Walker, Julia Khodor, and Michelle Mischke, *7.014 Introductory Biology, Spring 2005*. (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>

7.014 Introductory Biology, Spring 2005
Transcript – Lecture 11

So just trying to remind you that the replication fork looks something like this where 5 prime to 3 prime and 5 prime to 3 prime. This is what's known as the leading strand because DNA, the synthesis of the new strand can go --

Which is going 5 prime to 3 prime is going in the same direction as the movement of the replication fork. The other strand, which is known as the lagging strand, the DNA synthesis is actually going backwards to the movement of the replication fork, which means it has to go and then start up here and go again.

And it's continually jumping. And I told you that the little RNA primer is used to start each strand. And then the DNA polymerase is able to elongate that. And then at the end these little nicks in here, the RNA has to be removed, fill in the gap and then it's sealed up by the enzyme DNA ligase, which we'll talk about when we talk about recombinant DNA.

Someone asked, I had mentioned why this strategy of using RNA was beneficial, and that has to do with the fact that the fidelity, which is going to be the next thing I'm going to focus on of DNA replication as not, you can get a much higher accuracy if you have the end of a primer already there and then carry out the chemistry in there. No enzyme has ever achieved the accuracy that you see in DNA replication if it's starting a strand.

So RNA polymerase, which constantly starts strands to make RNA copies, as we'll talk about, is not as accurate as DNA replication. And by putting a little bit of RNA, because the cell has to start a new strand. Before it gets here there's no strand at all on this lagging strand so it needs to make this little RNA primer. It needs to make a little primer.

And by making it out of RNA then it can tell what doesn't belong there. It doesn't matter if it's not quite as accurate as the rest of DNA replications because it's going to take it out anyway and fill it in using the DNA polymerase. And if you think about that maybe you can see one of the reasons that the cell has chosen or nature has chosen through evolution to use little RNAs to begin the strands. OK. Well, in any case, the fidelity of DNA replication is really pretty amazing.

Incidentally, just speaking of DNA, many of you wrote some very thoughtful things about Vernon Ingram's visit. I didn't give him a whole lot of warning and he had to go and change his schedule and move meetings around in order to

come to talk to you. And it was very nice of you. Many of you wrote some very thoughtful things, which I'm going to pass onto him. I want him to know that many of you appreciated his visit. I also saw a lot of you reacted to his advice about crowded labs.

That has been my experience, too. And one thing about the scientific process is it's not just one person. You're in with a group of people, just as Vernon described, and that group of people becomes the creative engine that drives all the science within that lab. And so you're not only picking your project, you're looking for a group of people to work with. And, as Vernon said, if the lab is really doing hot stuff they tend to attract a lot of people. So a crowded lab can sometimes be a really good indicator.

No absolutes, and there's an exception to everything, but that was a good piece of advice he gave you if you're looking for UROPs sometimes. OK. So, anyway, DNA fidelity. Remember I said we've gone from, our bodies have somewhere from like 10 to 20 billion miles of DNA in them if we could take all the human DNA and stretch it out? But that fidelity is done at an error rate of about one mistake to every ten to the minus tenth nucleotides replicated.

Which I said if you were typing all the time it would be like sort of making one mistake every 38 years. So it's an astonishing degree of fidelity. Something that's beyond anything within our experience. And there are three principles that go. One is polymerase is really good at the base pair recognition telling that an A is paired with a T or a G is paired with a C.

And discriminating against everything else there's a phenomenon known as proofreading, and I'll tell you how that works. And then there's a third system called mismatch repair. And all three of these contribute to this very, very low-frequency of errors, one mistake for approximately every ten to the tenth nucleotides replicated.

So the first thing is I've pointed out to you several times that if you draw the hydrogen bonds between an A and a T base pair, the two hydrogen bonds or the three hydrogen bonds between a G and C base pair, that the shapes of this pair and that pair are virtually identical. You can pick them up and lay it right down on top.

Now, if you actually look at it you'll see you could draw some base pairs between, for example, a G and a T. In fact, you can draw two hydrogen bonds, which is the same as between an A and a T. But the one thing I hope you can see, just from the shapes even without being able to see the individual atoms, is that a GT base pair doesn't have the same shape as the correct base pairs. So when I showed you that little movie the other day where this is the template nucleotide, this is the incoming nucleotide and there's this alpha helix that's swinging up.

What's happening in there is that the enzyme is checking the way that the incoming nucleotide is the correct shape to go with the base pair. And you can sort of see it's flipping it right into a very narrow little slot in the enzyme. So it's not only asking for sort of hydrogen bonds, it's asking for the exact shape.

If you just did it by thermodynamic grounds you'd make about one mistake in a hundred because that's about the discrimination between the correct base pairs and some of these other ones. This works so well. You get more like one mistake in ten to the fourth or ten to the fifth. We're still quite a distance away from the ten to the tenth, but this is one of the things. It's looking for the correct shape of the base pair.

Now, the second thing that helps with fidelity is a phenomena known as proofreading --

-- exonuclease. Things called a nuclease.

That means it can degrade DNA. And the exo works at an end.

And, furthermore, the directionality of this proofreading was something that puzzled people initially because it's going 3 prime to 5 prime.

And when people started to purify DNA polymerases or complexes of DNA polymerases involved in replication there seemed to be a puzzle because the polymerase, as I've told you, goes 5 prime to 3 prime, but the same enzyme complex had an exonuclease that went in the opposite direction.

So this seemed very peculiar at first in the sense if you were trying to polymerase DNA in this way why in that same enzyme would you have something that wanted to degrade DNA in the other way? And the answer turned out that this was known as a proofreading exonuclease, as I've put up here. And here's the principle of how it works. Suppose you were replicating the DNA and there was a G. And if you put a C in there it very quickly goes on and continues the replication.

If it puts in a T, let's say, this is not a very good base pair. It wouldn't have the right shape. So when the enzyme came up looking for that 3 prime hydroxyl, which would be right at the end of that T, things are not in the right place. And so the polymerase activity slows down. And as that primer terminus, if it sits there for a little bit, it's able to just peel off the DNA, flip up, and there's this function that does just what you'd do if you were typing and you made a mistake.

You'd just hit the delete key and take off the last nucleotide that you did. And I have a little movie showing you that. This is a crystal structure. This is the DNA template. And the polymerase catalytic activity site is right here. And in this little movie it's just added an incorrect base pair and the polymerase is sort of stalled.

And the actual nuclease function is physically separate on the protein structure. But what you'll see in the movie is that if the polymerase cannot go very well eventually this thing will come up and it will chop off one nucleotide, come back and try it again. Let's see. I think if we do this, oopsy-daisey. Let me see if I can get this to work here.

Nope, it's not working. OK. Well, anyway, I'm going to skip it for right now. I don't want to waste time. But, in any case, the end would go up here and it would take off one nucleotide. So there at least are two of the ways that the polymerase is able to work with such fidelity. It selects for the correct base pair shape.

And then after it's done in addition it sort of looks back, just as if you were a very slow typist, and every time you typed a letter you looked back and said did I make a mistake? And if you made a mistake then you'd delete and then just try again. And that gets the cell another maybe two orders of magnitude of accuracy. So we're up to about one mistake in ten to the seventh base pairs replicated. The third system, which is called mismatched repair, turns out to be very important for a whole variety of reasons.

And before I tell you about it, I want to first introduce the idea of DNA repair in general. One of the things that's wonderful about DNA --

-- as you've learned, is it's got the information in two copies. It's in a complimentary form but it's like having the photograph and the negative. And if your kid sister pokes a hole with a pair of scissors through the picture of your boyfriend or your girlfriend, you're not really in trouble as long as you've got the negative because you can get the information back again.

And that same principle applies in DNA repair. So if you have some kind of lesion in DNA, and this might have come from going outside in the sunlight, your DNA absorbs in the UV and it undergoes photoreactions, they tend, for the most part, to just effect one of the two strands of DNA.

Or if you smoke, which I hope none of you do, there are many chemicals in smoke that will react with DNA, and they'll modify one strand. And so what the cell has is a system that has many kinds of repair systems, but it has a special type of repair system known as nucleotide excision repair.

And you could think of this as a protein machine that constantly scans the DNA looking for little distortions. And if it finds it then what it needs to do is it needs to make cuts, remove the DNA and make a little gap. And now you can see what it can do now. Once it's got a little gap the information over here is a complimentary form.

So if a DNA polymerase were to come along it could fill in that gap and seal it up and then you'd be back to ordinary DNA, the lesion would be gone. And I made a silly little PowerPoint thing here to show it. So if you were to, say, damage the guanine with something, say one of the carcinogens you find in cigarette smoke, you could think of this protein machine as being a sort of pair of scissors that have a conditionality in them.

As this protein machine scans along the DNA the scissors aren't activated until it recognizes there's a distortion here, at which point then it senses that there's some bump in the DNA. And it's very clever the way it does it because the nuclease activities, the things that are going to cut the DNA are actually some distance away, a few nucleotides away from the lesion.

So even if this is distorting the DNA, the scissors are able to work out here and out here. It makes two cuts. That was a huge surprise. Nobody expected that when they started to do the biochemistry. And then in principle once you cut it now you can remove this little nucleotide and then a DNA polymerase can just come in, and following those A pairs with T, G pairs with C, copy it along and then would seal it up to get to the end. And I've actually shown you a picture of what happens if a human is missing that system.

When I was showing you how profound an effect you could get from just losing one single gene or a mutation affecting one single gene, this disease called xeroderma pigmentosum. They're a variety of different groups. And the one on the left is an example. That's someone who is missing one of the genes that encodes one of the proteins involved in nucleotide excision repair.

And this is really, really important for fixing up the damage we get all the time in sunlight. So if you miss that repair system and you got out in the sun then you get all kinds of lesions and people are very susceptible to skin cancer. And I told you fortunately now you don't find people with this disease looking like that because at least in developed countries we recognize it. They're kept out of the sun. And these were the kids who I said are called "children of the moon" because they, for example, go to summer camps where they do everything at night so they won't get exposed to sunlight.

But that's what happens to us if we miss that excision repair. And, again, what makes that possible is that the information is there twice in a double-stranded DNA. I also showed you a little movie early on when I was showing you, I'm going to actually run this in QuickTime because it works a little more smoothly, I think.

So I showed you this when we were talking about DNA because I wanted you to sort of get that sense of what it was like to kind of fly down the groove of a DNA. But what I didn't emphasize was this protein that was bound to the DNA. That's a protein that's a DNA repair protein. And it's one of these things that looks for

lesions in the DNA. And as we fly along the major groove this little green thing is actually the lesion that that protein is looking for.

And it sort of puts fingers down into the groove and it's able to sense that. And you can sort of see how this protein is bound to DNA. This is a lesion that we get all the time from oxidative damage. And remember I said oxygen is bad for DNA? So our bodies have to have systems that are able to do that. So DNA repair is very important for life.

We'll just finish flying down the major groove one more time here. OK. I'm going to go back to PowerPoint.

OK. So mismatched repair is a form of repair that's got that same idea.

Let's think about it if we had a replication fork here, and let's say there was a G here and the T got misincorporated, but in this case it wasn't removed by the proofreading which happens about one in ten to the seventh times. Now if that strand is fixed up, excuse me, is continued then you'd end up with a GT base pair.

And the next time you copied it this strand would give rise to a GC but this one would give rise to an AT. And then you'd have a mutation that now would have changed. And if it affected an important gene that could be bad for you. So the cell has what's known as a mismatch repair --

-- that works in exactly the same logic as here. That it basically comes along.

It scans the DNA. It finds the bump because this is not a proper base pair. And then it fills it in and you're back to ordinary DNA with a GC base pair. There's one little wrinkle. For this system to work it has to do one other thing that's different from that kind of DNA repair.

Can anybody see what it is? Why don't you talk to the person next to you and see if you can figure it out. This system must be doing something else in order for this to work. OK, you can ask somebody. What do you think?

What if I removed the gene? Would that work?

What would happen if I took the gene instead? Say I made the little gap over on this strand instead, cut it here?

Yeah. So which one is the one that's right, the old strand or the new strand?

The old strand, yeah. See, this is the old and this is the new. And the term that's usually used, it's known as the daughter strand, the new strand. So the other thing this system has to do is it not only has to be able to detect that there's an

incorrect little base pair in there, but it also has to know which is the parental strand, the template strand, and which is the daughter strand, the newly synthesized strand.

And this system makes the assumption that the strand that's old is the one that's correct and the mistake is on the new one. You guys see that? OK. So that gets another two or three orders of magnitude in accuracy and that's what brings it up.

Now, the people who made this, who formulated this model for mismatch repair, complete with the feature that it needed to recognize the old and new strand, that's a bit of a trick, if you think about it because it's DNA on both sides. And there are several different ways used in nature, so I'm not going to go into it, but there's at least a couple of different ways of doing that trick. You could sort of see if you were the replication fork and you talked to that you could certainly, just from the geometry of that, if you wanted, you could probably keep track of who's old and new.

E. coli has a very cute trick, but it's not universal so I won't go into it, but the people who did the seminal stuff, I had to just quickly show you a couple of pictures. When I showed you that picture of the DNA 50th, the guy sitting in the front row was Miroslav Radman who was one of the two people. He's a European scientist originally from Croatia. And he collaborated with someone you've heard about before, Matt Meselson, who was up at Harvard. And it was with the Meselson-Stahl experiment that showed the semi-conservative mechanism of DNA repair.

This was a little reception. And Matt was talking to Alex Rich who's in the MIT Biology Department. And I was amused because remember how Vernon told you how Francis Crick would run up and down the stairs in the Cambridge lab and he was talking all the time? And I've heard Vernon say you could never really tell whether an idea came from Watson or Crick because they'd just talk, talk, talk all the time. So this was at sort of nice reception at the DNA 50th.

And within a couple of minutes, I looked over and there were Miroslav Radman and Matt Meselson talk, talk, talk. They were in the corner drawing pictures on a board. I also showed you actually a picture of one of the genes that's involved in recognizing this mismatch, because there's a protein that recognizes that mismatch and it's given the name of Mute S. And when I was showing you some proteins it had one that had a lot of alpha helices.

This is actually a picture of Mute S. It's a dimer. That's why some of it's green and some of it's blue. And this is DNA viewed end on and it's recognizing a GT mismatch in DNA in that picture. Now, this may sound very esoteric, you know, and obviously important for life and an important part of sort of understanding how life works if you're interesting in studying molecular biology.

It may not seem to have very much connection to your real life. But, in fact, in this case mismatch repair does because it affects the frequency with which, if you lose it, then when you replicate your DNA you're going to make more mistakes. And I need to just give you a very quick introduction to cancer so you can see why this is important.

Cancer comes from the fact that remember a human cell or a multi-cell like us that has many kinds of different cells starts out from one cell. And I talked about first you get the embryonic stem cells that can become anything. And the cells become successively more and more and more specialized as they go along. So ultimately a cell that's in your retina or in, say, the lining of your colon needs to know that's where it belongs.

And it also needs to know that it cannot just keep replicating. So if this is actually showing a little picture of the lining of your intestine. And there's a single layer of cells right along the inside edge of your intestines. This is the cells through which all the nutrient exchange happens and everything else when your body extracts nutrients as food stuff passes through your intestine.

And so what happens with cancer is a cell that's normally a part of your body has to obey a whole set of rules. And what you can think of when someone starts to develop cancer is that what started out as an ordinary cell undergoes some kind of successive changes in its DNA that gradually causes it to forget the rules that make it be part of an organized body system. So if we take a look here at all these different cells.

But let's imagine just one of the gets a change that makes it forget to stop, or it should know to stop replicating when it touches its neighbors, but if a cell were to lose that control what would happen? Well, it would then begin to proliferate. And then what happens in cancer is the cell will, now there are more of them, and one cell will acquire an additional mutation that will lead to a further loss of growth control.

You can see now the cells are starting to become sort of funny shapes. And then one of the cells in here will undergo yet another change. And right at this point, up until now, the cancer has, even though the cells are dividing and have lost some of their growth control they're still staying in the same place. So that would be sort of, you know, like a wart or something like that, or what you would hear as a benign tumor.

You can go in surgically and take it away. But then the other thing that can happen is cells can forget where they're supposed to be in the body. And when that happens they say the cells metastasize and become metastatic or a malignant tumor. And what that means is the cell is beginning to, it's acquired yet another change that's made it forget which part of the body it's supposed to

be in. And they've signified it here as being a change in this cell that then leads to, you can see here right now it's starting to invade into the whole intestine.

Or if one of those cells comes off loose in your bloodstream it can land somewhere else in your body and then start to grow there. And that's what happens when somebody has metastatic cancer. You cannot really cure it because now there are cancer cells all over the body. And that usually is a very difficult situation to get any kind of cure on. So to put this in perspective, you needed to have a number of changes to go from an ordinary cell to a metastatic cancer cell.

So each one of these changes there was some kind of change in the DNA. Either there was a mutation or maybe a chromosome was lost or something like this so that you need a series of successive genetic alterations. So there was a very key insight that a number of people had after we understood the mechanism of mismatch repair.

Because some people realized that if a human cell had lost mismatch repair then the frequency of each one of these changes would go up. It wouldn't affect what the change was. It wouldn't actually have anything to do, if you lost mismatch repair it wouldn't affect directly the ability of this cell to stop dividing when it touches its neighbors.

But it would increase the chances that a mutation somewhere would have that effect. And if every one of these steps goes now a hundred or a thousand times faster, you can see that if somebody loses mismatch repair in a cell then the chances of that cell coming into a cancer are very high. So there was a kind of human cancer, it's a susceptibility to colon cancer called hereditary nonpolyposis colon cancer.

You don't need to remember the name. It's often abbreviated HNPCC for people who cannot remember the name. But it was a kind of susceptibility to cancer that ran in families. So it was thought to be genetically determined in some way.

And one of the interesting things was a number of the people who had this disease would show a kind of instability of the genome if they looked in the tumors. They just looked at the DNA. It seemed to be undergoing changes at a much faster rate. And the insight that came out was that the people who had this disease had, for example, a mutation affecting what we can think of as a human homolog of mute S.

And we'll talk about genetics of humans in a small number of weeks, but I think most of you know that for most genes, except for the genes associated with the sex chromosomes, you get one copy of a gene from mom and another copy of a gene from dad. So under most circumstances we would have two good copies of this gene encoding a human homolog of mute S.

What does that human homolog of mute S do? The same thing as the bacteria. It recognizes a mismatch in DNA and fixes it up. So it turned out that what the people with this disease have is they have one of the genes. The gene they got from mom or the gene they got from dad is broken. So they're still OK. They have one copy of mismatch repair in every cell.

But if a cell ever had lost that copy of the good version now that cell and all of its descendants would mutate at something like a hundred or a thousand times the normal probability. And so they would progress down this pathway.

And so the polyposis means that if they look in the colons of people who have this disease they find lots and lots of little growths or polyps that are on their way to progressing down this disease. Even in these people it takes quite a while. And so once they knew that they were able to go in and through colonoscopies find these cancers and remove them.

And most of you will not have that disease, but this is now a kind of cancer that's pretty much preventable as long as it gets detected. It can take in a normal person as long as 20 years or something for an initial cell that underwent this initial change to go all the way down to becoming metastatic. So when you get older, and this certainly applies to most of your parents or in this age group, you should have ask them if they've had a colonoscopy.

It's not the world's most fun procedure because, you know, they stick a probe and look inside your intestine, but it isn't that bad. And what they do is if they see one of these little polyps they can catch it before it's progressed far enough to be metastatic. And then there's no problem. I had my first one done about, I don't know, three or four years ago and they found one. And they took it out and I'm fine. But if it had been left there and allowed to progress then some years down the line I would have gotten colon cancer.

And I'm going to have to go back and get checked again in another year or two. But it is something that you should check with your parents because everybody should have a colonoscopy. My hope is by the time you guys reach an age when this comes they'll probably have some kind of little blood test or something where you won't have to go through this indignity. But right at the moment it's something everyone should do, I think. I just wanted to make one other comment about basic research because there's another thing here.

Actually, my lab was the first lab to clone the mute S gene. We cloned it, we sequenced it, and we looked in the databases. And at that time in the late eighties there was nothing else that looked like it. I thought it would be like, there were some sort of similar mutants, and here's what it looked like. This is a culture of *E. coli*. And there are about ten to the ninth cells per mil. And we plated about ten to the ninth or ten to the eighth on a plate with a drug on it.

And you can see they almost all died, but there were maybe three or four that survived. And then their descendents were able to grow up and form a colony. This is how we recognized something was defective and what we now know as mismatch repair. If you took this mutant of *E. coli* and plated it out, you'd see you got a lot more drug-resistant colonies. That's the difference that I was describing, the importance of mismatch repair. If you don't have mismatch repair you can see, you get a lot more mistakes that show up as mutants.

So I was studying that. And we cloned the mute S and mute L genes which are another gene that's involved in this. Didn't see anything in the database, but there were very similar mutants in *Streptococcus pneumoniae* that people had isolated. Remember *Streptococcus pneumoniae* in the transformation experiments? So I thought, well, maybe these are the same genes on an evolutionary basis. So I phoned some labs, and I found one that was sequencing what turned out to be a homolog of mute S.

We tried to publish our papers in a medium fancy journal because I thought this was a pretty cool result that two bacteria that were evolutionarily very diverged had this conserve mechanism for mismatch repair, but the reviewer said, you know, this is a pretty specialized topic, it's not of general interest, it should go in, the phrase they use is "a more specialized journal". So it was published in the *Journal of Bacteriology* which is a really wonderful journal, but it basically deals with bacteria.

And about a week after that paper came out my phone rang and it was a guy from Emory. And he said, "I work on mouse. We were sequencing a gene, it doesn't matter what, and we sequenced in the wrong direction. And we seem to have something called mute S. Do you know anything about mute S?" And a couple of days after that I got a phone call from somebody at NIH. And they said the same thing, "We were trying to sequence this gene in humans. We kind of sequenced in the wrong direction and found mute S." So within a week of the paper coming out I knew there were mouse and human homologs.

And that led from these sorts of studies, which my first graduate student worked on, to the identification of the human homologs. And then not me but others made the connection between mismatch repair and cancer. But this is the way a lot of things happen with basic research. This doesn't look like anything that's very important.

And it sure doesn't look like it's going to lead to an insight into cancer, but this is very much the way it goes. I've had this happen twice with another set of genes in my life that turned out to be important for cancer as well. And, as I said, what happens, if you lose mismatch repair, then all these alterations happen much more quickly and the cells can become cancerous. I've included a couple of

outtakes because I actually made this slide with my son's pillowcase on our dining room counter.

And our cats, who you saw at some point earlier in the year, thought this was the weirdest thing they had ever seen, when I brought these plates home. So, OK, anyway. All right. So one other thing to tell you about DNA replication before I move on, and that is --

-- the initiation of DNA replication. In *E. coli* there's one great big piece of DNA. And it's all one giant circular chromosome. And if you realize what I've told you about DNA replication, I've talked to you only about once you have a replication fork established how you keep it going. But, as you might guess, a really important point of biological control is the initiation of DNA replication.

And so the way cells do that is they have a special sequence in their DNA. It's written just with Gs and Cs and As and Ts, but it's a word sort of written in a different language than the kind of genetic code we're going to be talking about in the next couple of lectures. And what it means is *start replication here!*

And so in *E. coli* these terms are called origin DNA replication. And, for example, in *E. coli* it's a stretch of DNA that's about 250 base pairs long. And it's got a sequence that lets proteins bind and they kind of are able to make a little bubble like this.

And it's at the edges of this little bubble where it's able to start a replication fork. And one of the secrets to control of cell division is that cells are able then to control whether the protein that sees the origin is there or not. And it won't start a new round of replication unless everything is right. Then it can make the things that initiate a new round. And after that it finishes. Our eukaryotic cells with a lot more DNA use the same thing.

The same idea, but there tend to be multiple origins. And you get a little bubble and another little one down here. And once you get the replication forks established then these kind of merge. And then eventually we end up with the two strands of DNA. But I just mention that in passing because it's an example of how even though the DNA is nothing but Gs and Cs and As and Ts, you can kind of write words in there that mean different things.

Some of them on the genetic code tell you what the order of amino acids in the cell are, but everything else has to be encoded in the DNA, too. And here's a really nice example of how that works. Now, we're going to switch at this point from worrying about how DNA is replicated to how information is stored and interpreted.

And there's a figure that most of you have probably seen, DNA goes to RNA goes to protein. This is the usual direction of information flow. The information

for making proteins is encoded in the DNA, as we'll talk about in more detail, and an RNA copy of some piece of that, one gene's worth usually, gets made in RNA.

And then that information in the RNA is used to direct the sequences of amino acids that appear in a protein. And this is a four letter alphabet, if you want, A, G, T and C.

This is a four letter alphabet, A, G, U and C, where the uracil and the thiamine have the same base pairing capacity. And this is a 20 letter alphabet.

All those 20 amino acids that you were looking at, at the chart over at the back of the exam.

So from the point of view of information storage and information flow there are some interesting things that had to come up in order for the information to flow in that way. But before I do that I want to just get you to think about DNA as an information storage device. This is MIT. I'm almost sure in this room there are some people that are experts in high density information storage.

And even if you're not most of us have now a lot of experience with it. Your computer can do gigabytes of information. Your iPod probably has a 40 megabyte hard drive in it or something like that. So you have some experience with high density information storage. So here's the question. How much DNA would it take to encode everybody who's alive on earth today, 6 billion and a bit people?

And let's argue that all we need is a single cell's worth of DNA because everybody started out a single fertilized egg and went on. Yeah? OK. Enough DNA to fill one human being. Anybody else got any sense? All right. This is, I think, the most amazing demo. I did this when I was teaching for the first time.

The amount of DNA it would take to encode everybody who's alive on earth, one cell of everybody who's alive on earth today is this little thing in here, which you probably cannot see even, but I took a picture of it. There are about six times ten to the minus twelfth grams of DNA in a human cell. And if you multiple that out by 6 billion people it comes out to 36 milligrams of DNA. And I weighed out 40 something milligrams of DNA.

So there's actually more DNA there than you need to encode everybody who's alive on earth today. And I don't know how this hits you, but I've been working on DNA my entire life. And every time I do this, you know, I think I understand this molecule, but I don't really think I do at some more fundamental level. It's absolutely amazing how much information is stored in that molecule.

So the one point I will, actually, I think it's close enough. Why don't we just call it a day, and I'll pick this stuff --