MITOCW | 7.013-S05-L11

OK. And here we are in the molecular biology section. And the goal of this section, as Professor Jacks started to tell you during the Genetics module and Professor Baker told you at the beginning of last lecture is try to link together in molecular terms the question of genotype and the question of phenotype. And we presented to you this notion that goes by the ponderous name of the central dogma that the link between genotype and phenotype is related to DNA as a genetic material that then proceeds to transmit its information to a final outcome, which is very often a protein, through an RNA intermediate. And the point of these molecular biology lectures is to tell you about the molecular biology, and then at the end to try to bring together this genotype and phenotype in molecular terms. Now, last lecture you talked about DNA replication, DNA as the genetic material required to be replicated faithfully and accurately so it can transmit its information to the next generation. Professor Baker I know stressed the requirement for accurate replication, but she did not do one part of this lecture that I want to spend a couple of minutes now discussing with you. And that is the question of DNA repair. So there are two types of DNA repair. Actually, three types that I want to talk to you about. And the first pertains to the accuracy of the DNA polymerase that replicates the DNA. So DNA polymerase -- -- three, or the DNA polymerase that replicates the DNA makes mistakes. It puts in the wrong nucleotide. It puts in the wrong base. And it does so about one in ten to the fifth bases. OK? So one in a hundred thousand bases is wrong. Now, if you think about the fact that there are more than ten to the ninth bases in a human genome, every cell cycle that translates to ten thousand or so mistakes, that's a lot of changes in the DNA. That's not a very faithful kind of DNA replication. So this has been selected against evolutionarily. And there is a mechanism that's called proofreading -- That allows this high error rate to be corrected. And it's actually very cleaver. So this DNA polymerase has what is called an exonuclease activity. Exo meaning out. Nuclease meaning to break down nucleic acids. And this exonuclease proceeds from the 3 prime to the 5 prime direction, the 3 prime nucleotide being the one that was added last as you should now know. And so what DNA polymerase does as it is replicating is it kind of feels whether or not the double helix has reformed in a smooth way. And if it feels that there is a bubble there, a bubble where the two bases, actually look at me rather than the diagram. I think it's easier. I can do it better with my hands. Where you've got a nice smooth helix, if there is a mismatched nucleotide, the bases do not pair, there will be a bubble. OK? There will be a bubble in the helix or a space in the helix. The two strands will not be joined together. And the DNA polymerase can sense this and it can go back and it excises the wrong nucleotide and puts in the correct one. OK? This is called proofreading. And it's extremely necessary and it's actually very good. And what it does is to decrease the error rate to one in ten to the ninth bases. OK? So you get four orders of magnitude improvement in the accuracy of DNA replication. Now, there is another set of things that can go wrong. And these actually fall under the heading of mutagens. Mutagens, as Professor Jacks mentioned to you, being agents which change the base sequence of the DNA once the DNA is there. And these can either be chemical or these can be ionizing radiation. And in those cases also the helix gets changed because the wrong base gets put in. No, not because the wrong base gets put in. But because there is a chemical reaction which might modify a base, which might, for example, covalently link two bases. thymine for example. If two thymines are sitting next to one another in the helix, ultraviolet light is very good at cross linking those. And you now have something called a thymine dimer. And that is very bad because that is not a normal base sequence. And when replication time comes along that DNA helix is abnormal and the replication machinery doesn't know what to do about it, and that can lead to all sorts of problems and to mutations. So there are mechanisms that can get rid of abnormal bases. So mutagens can chemically, actually, maybe not say chemically. Let me just say change bases. They change base structure either to something that looks like another normal new base or to something that looks abnormal. And there are two mechanisms to get rid of this. One is called excision repair and the other is called mismatch repair. I have them written in the reverse order than is on this diagram from your book. In mismatch repair there is one nucleotide that looks normal, but it's different. It doesn't match the, usually it looks normal. It doesn't match the one opposite to it. And in that case the repair machinery can go in and remove the abnormal or the mismatched nucleotide, and there's another enzyme that will go and correct it. In excision repair, one very often, excision repair occurs when, for example, two nucleotides have become covalently linked to one another, and the one strand of the helix is just a mess. And there is an enzyme, or enzyme complex that will go in and actually excise a little chunk of the helix. And then another enzyme will come in and fill in the gap so that you get the helix repaired. Now, the challenge in this, and you may be asking yourselves this, is how does this repair machinery know which the correct strand was? In the case of proofreading it's very interesting because initially after replication the newly synthesized DNA strand is not modified. It's just a normal nucleotide polymer. However, the template strand, the template strands, the parental strands over time become chemically modified. The bases actually get, especially adenine gets some methyl groups added to it. And this is different than the newly synthesized with doesn't have these methyl groups. And so the polymerase knows which strand is the old strand and the correct one and which is the new strand and the incorrect one. In the case of excision and mismatch repair, that's sometimes not clear. Where you've got these thymine dimmers, these Ts that are joined together then that's clearly the wrong, that's clearly wrong. OK? The enzymatic machinery can take that out and copy the other strand. Sometimes, though, if you just have a chemical conversion of one base to another, the repair machinery does not know which strand is the correct and which isn't. And that's when you'll get mutations fixed in the DNA because at replication you really may get the changing, you may get the incorrect, you may not get the correct base repairs. And then that incorrect base will be passed on through the next generation. OK. So this is a very rapid zip through DNA repair that I wanted you to be able to think about. I want to move onto the next step in the transmission of information from gene to final product today. And I want to talk to you about the generation of RNA. And so let us begin with a quiz. And I have for you a new incentive to pay attention, a new prize that you can use to think about the conversion of potential to kinetic energy, and also you can use to amuse yourself when you're downloading very poor, when you're downloading things from the Internet and have nothing better to do. I can usually get this right across the room. There you go. You can also use it to think about the nature of amphibians, they're nice flying frogs. OK. So let us pose the question here, what is RNA? And you've had some of this on a problem set, but you really need to know what I'm talking about. This is a ribonucleotide. How do I know that this is a ribonucleotide? Think about it. You can put your hands up, but I want everyone to think about it. OK. And you need to identify the precise chemical group, please, that tells me. I saw you two first, so yes. What does the lower right mean? Give me a name. It's the? There's a number there. The? Ah, we have a discrepancy of opinion here. Someone says it's a 3 prime hydroxyl on the ribose and someone says it's the 2 prime hydroxyl on the ribose. Let's take a vote. Who thinks that this is identified as a ribose because of this 2 prime hydroxyl? Thank you. And who believes it's the 3 prime hydroxyl that identified riboses? OK. We have a smaller but firm contingent. In fact, it's the 2 prime hydroxyl that identifies this is ribose. You remember, and you really need to remember that this three prime hydroxyl is the reactive group that allows the sugar phosphate backbone to polymerize. This 2 prime hydroxyl is a reactive group. It identifies this as ribose rather than deoxyribose, and it also is an additional reactive group. And the fact that it is a reactive group makes RNA rather labile. OK? So let's write a couple of important things here. So this is RNA as the nucleic acid polymer. You should really know this. Ribose has both a 3 prime hydroxyl and this 2 prime hydroxyl. And this is a reactive group. And because of this RNA is a much less stable polymer than DNA. Here's another one. What type of polynucleotide is this and how do you know? Yes. You. OK. It's RNA. And it's RNA we know because of these uracil groups. OK? So uracil is an alternate base to thymine that's found only in RNA. Here are the Us. It tells you it's RNA. OK? So you need to know those facts about RNA. Good. So let me pose a question to you. In this litany that you've had several times now where the flow of information moves from DNA to RNA to protein, why is the RNA there? This is a rhetorical question. I'm going to try to answer it for you. Why is the RNA there? Why is there an RNA intermediate? You could imagine that the DNA double helix could open up and that nucleic acid could be directly translated or could be directly converted or the code could be changed to form a protein without any RNA intermediate. But, in fact, universally throughout biology, throughout our world anyway, throughout our earth, RNA is there as an intermediate. Why? Well, I think the answer actually lies in evolution. RNA is probably the most ancient of the information polymers. That is widely believed now. So RNA is ancient. It was the first, strongly believed now that it was the first information carrying polymer. RNAs themselves were catalytic. They became able to replicate. And they also probably became able to be translated into protein before DNA was invented. OK? So DNA's chemical structure is different and it's a derivative of ribonucleic acid, and undoubtedly came second. There was an advantage of having DNA because it's so much more stable, and it made the hereditary material much more stable and much more faithfully transmitted from generation to generation. So RNA was ancient. And the relationship between RNA and protein is probably a very old one, and we'll talk about this relationship next lecture. And I believe that that relationship has persisted, and then DNA was kind of an add-on. And the DNA to RNA to protein does not necessarily reflect the only way or the best way to do things. Evolution is a capitalization of various changes. And RNA to DNA, DNA to RNA to protein is how things work now. But this, I think, is a consequence of the evolutionary past. Now, however, in our modern world RNA serves two main purposes. One of the things it does is to allow one to use just a subset of the genes to make proteins. So, as you've been told several times, you and I have about 30, 00 genes in our genomes. Not all of those genes, and we will discuss this in great depth as the course goes on. Not all of those genes are used at any one time. We use just a subset of the genes. And having them converted into an RNA intermediate is one of the ways that you can allow just a subset of the genes to be used. So I'm going to write here subset. Subset of gene usage. OK? Because you can turn just some of those genes, or you can convert some of those genes into RNA, the information in some of those genes into RNA. And the other thing it lets you do is to amplify the signal from each gene. So there are two copies of each gene in a diploid cell. When it comes to RNA there can be up to 10,000 copies of RNA per cell of a particular RNA. OK? So you can get an amplification of the signal -- -- from each gene. RNA copy number per cell ranges from about one copy to about 10, 00, that's rare, copies per cell. All right. So here we are. Why RNA? We've dealt with that. So I want to talk to you about two things. I want to talk to you about synthesizing the RNA, and then I'm going to talk to you about modifying the RNA a bit. And the first thing I want to cover is something called transcription. Which is also known as RNA synthesis. And you all should have this handout. So I'm not going to draw it but I will write some salient features on the board for you. And we're not quite ready to use that. I'm going to leave this up here, but I'm going to work on the board for a little bit. The basic idea behind transcription, RNA synthesis, is that one copies a DNA template into a complementary RNA strand, complementary RNA. And one does this, as I've alluded to, only from the genes. And this is an interesting point because although you have 30, 00 genes in your genome, in fact, those 30,000 genes only take up about 5% of the total amount of DNA in each of your cells. So 5% of your total DNA of your genome comprises the genes, the information carrying entities in your DNA. And the rest is other stuff. So the 95% is not genes. It consists of various repeats, repetitive DNA that can be there at just a few copies per genome or at 10,000 copies per genome. They can be real little, 10 base pairs, six base pair repeats, or they can be a few kilo bases repeated many times. Oris, Origins of Replication that you talked about last time are not genes. Those are there, too. Centromeres, the middles of chromosomes. Telomeres, the ends of chromosomes. All of these things are not genes, and they comprise the bulk of your DNA. Now, this isn't true in all organisms. OK? Some organisms have got very little of this repetitive extra DNA. We happen to have a great deal of it. OK. So let's pursue this a bit more. And let's think a bit more about these genes. And in particular let's think about the kinds of RNAs that those genes make. So I'm going to talk about gene classes or classes. And this is with respect to the RNA and the functional RNA that comes from those sets of genes. And I want to distinguish two major classes of genes. The first are the protein encoding genes. And protein encoding genes move through a type of RNA that is called messenger RNA, abbreviated mRNA. Messenger RNAs comprise about 1% of the total amount of RNA in a cell. And they can range in size from let's say 100 base pairs to 10, 00 base pairs. OK? So there's a very wide size range. No, not base pairs. Yell at me. Why not base pairs?

Why was I wrong saying base pairs? Tell me about RNA. Raise your hand. This is worth a frog. I caught myself, but if you can catch me, too. Yes. You. Good. OK. Generally RNA is single, woops. RNA is single-stranded. It does not form, it can form a double helix, OK? It's not as stable as the DNA double helix, and many RNAs, probably most RNAs have some double-strandedness to them, but that is an intromolecular double strand in this. There are some RNAs that form intermolecular double strands, but in generally I'm going to assume that RNAs are single-stranded. So we talk about 100 bases rather than 100 base pairs. OK? Second class of genes are the ones that do not code for protein, and in this case the RNA is the final product. And this litany of DNA to RNA to protein doesn't hold. You just stop at the RNA. And the RNA is the functional thing. So here RNA is the final product. And we can break these into a bunch of different classes. Ribosomal RNAs, abbreviated rRNA are a very abundant class of RNA that comprise about, I've moved over here, let me move here, 98% of total RNA. And there are a few thousand bases in length that say 2, 00 to 4,000 bases in length. OK? So this is 98% ribosomal RNA. This is fascinating. I'll tell you next time. This is the RNA that comprises a very large proportion of the ribosome that is the factory that makes the proteins. OK? And so I will tell you more about these next time. Some other ones, tRNA, the T for transfer RNA. tRNA comprise about 1% of all RNA and are about 100 base pairs, 100 bases long. OK? And then an interesting one that MIT has had a huge role in discovering and studying, these things called micro RNAs, abbreviated miRNAs, which are, they're at relatively low abundance. Less than 1% of total RNAs. And these are small. In their mature form they're about 22 bases in length. OK. So now, and I believe I cannot do anything with these boards. Ah, I can do something with this one, but that one is stuck. All right. So I'm going to do something with this one. And then I'm afraid it's going to disappear, but it's not going to matter because you have the handout in front of you. So now I'm ready to move on with you to the basic idea of transcription. And I'm going to write some facts on the board, and we're going to look at these cartoons that I drew for you together because I think your book is kind of difficult. So I decided to draw some cartoons to help you with the basic idea. Transcription or RNA synthesis takes place in the nucleus. Anyone else need a handout? Why don't you come on down. Actually, one of the TAs, could you be an emissary and just hand out to those people with raised hands? Thanks. Transcription takes place in the nucleus. And the idea is really analogous to DNA replication with a difference. The analogy is the synthesis of a complementary strand of nucleic acid on a template strand. So this is an enormously important principle that you need to have. Super important that you get the principle. The basic idea involves synthesis of a complementary strand of nucleic acid from a template strand. The template, actually, let me start even earlier than that. We start with a gene that generally comprises double-stranded DNA. There are exceptions to almost everything that I will tell you, or that Professor Jacks will tell you. You should understand that there are exceptions. Some organisms, particularly viruses have genomes that are RNA that can be single-stranded or double-stranded RNA. Some have genomes that are single-stranded DNA. But in general most genomes are double-stranded DNA. And the deal is this. The double-stranded DNA separates its strands, and one of the strands, and this is the difference between DNA replication and transcription, one of the strands becomes the template strand. And this template is copied to form a complementary strand. And it's copied by an enzyme called RNA polymerase. So RNA polymerase synthesizes the complementary strand to the template strand. complementary strand. And it does so, of course, as RNA, because we're talking about RNA synthesis and this is RNA polymerase. It does not, unlike DNA polymerization, require a primer. So this does not require a primer. OK. You should know, and it should be getting deep within your neural circuitry that polymerization occurs by adding nucleotides to the 3 prime end of the growing polymer. Yes. If that didn't make, you know, if you didn't say "yeah" to that, go back and think about it, go back and look at problem sets and you'll get more practice in this. But you really need to know that the growing chain adds onto the 3 prime end. OK. So after the polymer, after the RNA polymer is made the RNA is released from the template strand. As its being transcribed it forms this complementary strand. And, as you know, complementary strands can base pair. After it's made it is released from the template. And it usually then goes into the cytoplasm where it does its thing. So if you look at the diagram I gave you, that's what's up here, here's your double-stranded DNA, your gene. The strands separate. One strand is transcribed into RNA. The RNA is release. Obviously, your double-stranded template, or what was your double-stranded template will reform its double strand. So perhaps that's not so obvious, but the double-stranded, originally double-stranded template will reform its double strands, thus released RNA, then goes into the cytoplasm where it is translated into protein, or where the RNA is the final product. So let's look at that in a bit more detail. I've got here a template strand shown in red. This is, again, the second picture in the handout in front of you. And I've got three features added here. I have got a precise start site of transcription. I've indicated elongation where the polymer is elongating. And I have a precise termination site where transcription ends. OK? Now, let me see what I have here. I have a movie here. Watch the movie. I'll show it to you once, and then you can go and watch it at your leisure. This is meant to be RNA polymerase. There's the helix opening up locally. Here are ribonucleotide triphosphates coming in, and RNA polymerase is catalyzing their synthesis. OK? So the template strand is the bottom and here is the RNA being released. There's RNA polymerase moving along the helix. And the depiction is that the helix is opening locally and then closing again behind the RNA polymerase. At transcription termination, the helix, the gene helix zips up again and the transcript is released. So this is a very much simplified story. But is the basic principle of transcription. And you should know it. And in particular I have put onto this second diagram, and because you have him in front of you I'm not going to write this on the board, I'm going to use this as something to tell you, I have put the directionality of the strands of the double-helix on this diagram. This should be something you can deal with. 5 prime to 3 prime on one strand. The other strand is anti-parallel. RNA, any nucleic acid is synthesized by adding onto the 3 prime end. And that newly synthesizing nucleic acid polymer is anti-parallel to the template. This is also something that you should be familiar with. And you will have, will have, have not yet, will have practice on doing this kind of polymerization, but it should be something you really, really should be familiar with, this anti-parallel requirement. So, in fact, you can tell the direction of transcription because of the directionality of the template strand. OK. So this is very important for you to go and think about after class the directionality of the template and of the newly synthesized polymer. These are some diagrams from your book, and you can go and look at them. I'm not going to dwell on them. They indicate the different between, or the steps in transcription initiation, elongation and termination. And I've put them up there just to tell you there are these diagrams in your book and you can go and take a look at them and read the accompanying text. OK. So I see three problems with transcription that are very interesting problems. One is how to find the genes. I'll write them on the board and then we'll go through them. 5% of the genome is genes. That's most of it that is not genes. How does the transcription enzyme, how does the RNA polymerase know which is a gene and which is not a gene? How does it know, even if it finds the gene, which strand is the template strand and which is not the template strand? I could have drawn your previous diagram where the top strand was the template, and that what would have happened would be that the RNA synthesis went in the other direction. So which strand is the template? And that also gives you the direction, of course, of transcription. And the third one I'm going to write, and then I'll tell you about this in a moment. I'm going to write how to unwrap chromatin. OK. So in each of your cells, you have to look at me for this. In each of your cells there is this length of DNA. One meter. This is a little over, but one meter of DNA. How big is the average cell in diameter? Give it to me in micrometers. Worth a frog. On average. Well, that's actually a really big cell. It's about ten times less than that. But whoever that was, who was it? No way. These are very bad to throw. Very bad to throw. You can have it because you caught it. See me afterwards. I'll give you one. OK. [LAUGHTER] OK. So how do you pack a meter of DNA into a cell that is about ten microns in diameter? OK. So, OK, Jamie, you want to hazard an answer here? Your hand was up. OK. Good. You can wind it up. OK. The other thing you have to do, of course, is to make it really thin. It has to be a lot thinner than my piece of rope. But once you've made it really thin you can then wind it up. OK. It's logical and this is how it's done. And you can wind it up and then it will fit into your ten micron cell. Now, in actual fact, there's a whole process to do that. And I'm going to go through them as we go through these problems here. So here is problem one exemplified. I've got red little dots for each of the genes. How does RNA polymerase find these genes in this vast amount of DNA that is not genes? Here's the other one. Which strand is the template? Oh. And here is a nice problem that in the interest of time I am not going to do here in class with you, but I want you guys to go and do this as an exercise. I will tell you that the answer is not on your handout on the Web. I took it off. Sneaky, ha? So that you can go and think about this. I want you to go and understand that the products of synthesis from either strand of a DNA double-stranded helix are not the same. OK? And I'm going to zip through this because I want to move on here. OK. And I want to move to problem three which is this thing I called chromatin. DNA is wound up around proteins. These are called histones, and we'll have more to say about them later in the course. And wound up and wound up and wound up. And there is a very set number and type of proteins that the DNA is wound around. And once the DNA has been wound around once, those DNA protein complexes are wound up some more, and then wound up some more. And eventually you get them wound up and wrapped up so much you get the characteristic rather large chromosomes which are very much packed DNA. Now, this is a great way to fit DNA into a cell. However, this wrapping up of the chromatin into, the wrapping up of the DNA into this chromatin structure inhibits transcription. And in order to allow transcription to proceed, you have to remove these proteins from the DNA and allow it to unwind locally. And that takes a whole series of enzymatic steps, again that we'll explore more later in the course. But the problem I throw out at you now is hw do you unwrap the chromatin where transcription is needed? And the answer to all of these things lies in a specific, no, stop. Stop. Down. Up. OK. The answer to all of these questions lies in a specific DNA sequence or a series of specific DNA sequences that are collectively called -- -- the promoter. Here's another one. What is a promoter? And I need to make the distinction now between transcribed DNA of a gene and untranscribed DNA of a gene. The promoter is part of a gene but it is not transcribed. It usually depends on the gene and the type of gene. It usually lies 5 prime to the transcriptional start site. And it is a DNA sequence that says this is a gene, and it also says transcription should proceed in this direction. OK. And the way it does these things, I'm going to each of the answers to each of the problems now, is that it binds proteins that specifically recognize the sequence of the promoter. So you talked about the DNA replication origin and proteins that specifically recognize the nucleotide sequence of the origin. This is analogous. There are proteins that recognize promoter sequences which are similar but not identical from gene to gene. So it binds proteins. And these are called transcription factors. And these transcription factors bind in a DNA sequence specific way. OK. It also binds RNA polymerase which I'm going to abbreviate RNA polymerase, RNA pol. OK? And the answers to the three questions are that firstly the protein-DNA interaction is sequence specific, sequence specific, and so this allows you to actually find the genes. Secondly, and this is cool and I'll show you a picture of this in a moment, the proteins interact with a promoter DNA differently on different strands of the helix so they bind asymmetrically. They may bind more to one strand than to the other strand. And this gives directionality to the transcription so the protein binding is asymmetric or strand specific. Not for all of these proteins but for a significant number. And that helps you decide which strand you're going to use as the template. And thirdly these proteins have got associated with them activities that will unwrap the chromatin, that will unwrap the DNA from its protein complexes and allow it to be accessible to the transcription machinery. OK. Let's zip so I can show you this. You can look at this on your slides. Here are some pictures from your book. Really important in this, don't move. Really important in this is a protein called TF2D which recognizes a sequence called, that goes T-A-T-A-A-A. This is called the TATA binding protein. And it's really important. And it's the one thing that, the major thing that gives asymmetry to this transcription, set of transcription factors on the promoter. Once TF2D has bound to the promoter, other proteins come along, including these various other things called BFHG and so on. And here's RNA polymerase. And you can see this complex positioned asymmetrically on the DNA. And this complex you should know the name of, I'm going to put it in its own box here, is the initiation complex. And I want to show you a crystallographic rendition of the TATA binding protein called TBP, also sometimes called TF2D. But TATA binding protein here shown in purple. And if you look here, you're looking head on at the double helix. OK? Here's the helix. You're looking down the helix. And you can see that this protein is positioned on just one side of the helix, so that gives you asymmetry. Here's another transcription factor bound to DNA. This is a protein called GAL-4. It binds as a dimer. And you can see that GAL4 is the blue. And here it is contacting just one side, one strand of this red double helix. And I'm going to stop there and finish off the last little bit on Monday.