

ERIC LANDER: So, now what I'd like to do is turn to variations on the theme. One of the best ways to understand what's going on with DNA goes to RNA goes to protein is to consider how it works in different organisms. And the organisms we'll consider are eukaryotes, like you; prokaryotes, like a bacterium; and viruses. And each does the same basic copying of nucleic acid to nucleic acid, transcription, and translation.

But there are some pretty fascinating variations on the theme. So let's turn to those variations. Let's start with DNA replication. For eukaryotes, your genome are long, double-stranded DNA molecules, and they're linear. Long, linear, double stranded-- ds for double. I'll write it out-- double stranded DNA. You know this.

And you've got a lot of it. The human, you actually have 23 pairs of chromosomes. And the total length of your DNA is about 3 times 10^9 base pairs. So about 3 billion bases or so, typical chromosome on the order of about 150 million bases or so.

The mouse, pretty similar. Mouse, 20 pairs of chromosomes. And it's something like 2.7 times 10^9 bases. The dog's about 2.5, the elephant's about 3.3, essentially all mammals are about 3 times 10^9 .

Tomatoes, 12 chromosomes. They're also in the neighborhood of 3 times 10^9 to the ninth. You don't have more DNA than a tomato, for example. Yeast, much smaller genome. Yeast is like-- it's got 16 chromosomes. Yeast is in the neighborhood of about 13 million bases instead of billions of bases.

Fruit flies, four chromosomes, including a pretty measly one. And it's in the order of 200 million bases, et cetera. Many different sizes for eukaryotes. But there's one significant issue that all eukaryotes have, which is this chromosome here, this linear chromosome, how do we replicate it? Well, we told you. We open up a bubble. We start making little primers. Primase makes primers, you extend it, it's all just fine. Except in one place, the end.

Primase makes a primer, continues to the end, no problem. What happens out

there at the end? Suppose primase sits down here. What happens at the very end of the chromosome? If my primer wasn't exactly at the end of the chromosome, what happens?

STUDENT: A little bit of it doesn't get replicated.

ERIC LANDER: A little bit doesn't get replicated. Does that matter? Just a little bit--

STUDENT: [INAUDIBLE]

ERIC LANDER: --per cell division. Every cell division, you lose a little bit of information off the end of your chromosome. That's not so good. The word for end is tel. And so the ends of chromosomes are called telomeres. And if you have a linear chromosome, you have a problem with replicating your telomeres because it's going to get a little short each cell division.

So what do you think the cell does? Special mechanism, special enzyme that comes along and adds-- there's a repeat sequence that occurs. There's a repetitive sequence that occurs out here. T2AGGG in humans, different things in other organisms. And there's a specific enzyme that comes along and adds back telomeres so you don't get in trouble failing to replicate enough stuff at the end. By chance, the enzyme happens to be called--?

STUDENT: Telomerase.

ERIC LANDER: Telomerase. And some folks got a Nobel Prize for this last year, for understanding how telomerase works. What cells in your body are in need of replicating and replicating and replicating? Well, not necessarily in your body. What cells in some people are replicating and replicating and replicating and replicating? Cancer cells.

Cancer cells probably need their telomerase, right? So one way to possibly treat cancer might be to inhibit telomerase. So you see, all the things I'm telling you about, these are useful fun facts to know and tell about how the cell works.

They're also the heart of a lot of approaches in medicine. Because if you could specifically inhibit telomerase, you might specifically create a liability for rapidly

dividing cells. And so telomerase, very interesting.

Anyway, so a linear chromosome has that problem. And if you understood our mechanisms for replication, you'll understand there why the linear chromosome needs some special mechanism.

Now, here, prokaryotes are much easier. Because most prokaryotes have circular chromosomes, double-stranded circular-- ds circular-- DNA. This is much easier because there are no ends. They don't have to worry about the telomerase problem. They start somewhere, they start replicating around, it all works fine.

E. coli, for example, 4 million bases of DNA. The smallest are microbacteria. They're on the order of about a million bases of DNA. And they work just like we talked about.

But now, viruses, they are weird. Viruses -- Turns out some viruses have some double-stranded linear DNA, and some with multiple chromosomes, even. Some viruses, though, have double-stranded circular DNA. So it turns out viruses can do either of those.

It turns out they can do more than that. Some viruses have single-stranded circular DNA. That is to say, when they're traveling around in their capsid, in the protein coat, what's in the protein coat that you get infected with is a single-strand of DNA. Well, how does it replicate?

When it gets in, it becomes double stranded. The first thing it has to do is polymerases end up making this guy double stranded. But it travels around in its single-stranded form.

Why? Because it decided to. The great thing about viruses is they are small. They've had a chance to experiment with a zillion different things.

But some viruses don't have any DNA at all when they travel around. Instead of having DNA, they alternatively could have RNA. Remember I said Crick already figured out DNA and RNA are essentially equivalent. They're both nucleic acid. You

can go from one to the other.

Some viruses decided to bring along single-stranded RNA. So when the virus attaches to the cell, it injects an RNA. The RNA is in the cell. But how is the virus going to do anything? How's it going to replicate itself?

How do you replicate RNA? Well, the same stuff I told you about replicating DNA-- namely, for replicating DNA you use a DNA polymerase. It's a DNA-directed DNA polymerase. It uses DNA as a template. Any reason not to have an RNA-directed RNA polymerase? No. You can have one of those.

So the way this works is this gets replicated into a strand of RNA. It makes double-stranded RNA by an RNA-directed RNA polymerase. That's a kind of weird enzyme. It takes RNA as its template.

And it uses RNA as its template, and it makes another copy. It makes a strand of RNA to make it double stranded. And then it goes back and makes another strand of RNA.

But you don't have that enzyme. Where does that enzyme come from?

STUDENT: The virus.

ERIC LANDER: The virus? Did it bring it with it?

STUDENT: Another cell that it infects?

ERIC LANDER: Another cell that infects?

STUDENT: The RNA encodes--

STUDENT: The RNA does--

ERIC LANDER: Whoa! Wouldn't it be cool if the RNA was a messenger RNA and it encoded a protein, and the protein it encoded was the RNA-directed RNA polymerase? Bingo. That is, in fact, what happens with a certain class of what are called plus-strand viruses.

This is a messenger. It's a messenger RNA. And it's actually encodes the instructions to the cell, please make me an RNA-directed RNA polymerase. That's way cool.

It also turns out that some viruses are what are called minus-strand viruses. They don't bring a messenger RNA. But instead, like you said, they bring their own polymerase with them. The polymerase comes--

So here, these bring the instructions for a polymerase. These actually bring the polymerase itself. The polymerase then copies this strand, which is the messenger RNA. And it makes more RNA-directed RNA polymerases.

So both of your two solutions-- the virus brings a polymerase with itself, or the virus brings the instructions for the polymerase. Both of those actually happen. Pretty much a good rule with viruses is anything that can happen does happen. This is pretty much Murphy's rule for viruses there.

Turns out viruses can do one other thing. It turns out that viruses can take that RNA strand-- RNA-- and, although I won't go into all the details, copy that RNA strand into a DNA strand and then copy that DNA strand to a second DNA strand, to make double-stranded DNA. So some viruses that bring RNA with them copy themselves not into more RNA, but back into DNA. So instead of an RNA-directed RNA polymerase or a DNA-directed DNA polymerase, what is this? It's an RNA-directed DNA polymerase.

So this is an RNA-directed DNA polymerase. In effect, what is this thing doing? It's doing the exact opposite of transcription. What's transcription? Reading DNA into RNA. What's this guy doing? Reading RNA into DNA. It's the reverse of transcription.

What is the enzyme called? Reverse transcriptase. It's called reverse transcriptase. And then what happens is quite insidious, is that if this is your own chromosomal DNA, that piece of double-stranded DNA from the virus can be inserted into your own human chromosome.

The virus can then make more copies of itself by transcription of that. This is a truly insidious virus because it doesn't just infect your cells and grow. It infects your cells, turns into double-stranded DNA, and installs itself. And how do you get that DNA from the virus out of your chromosome?

You don't. You can't get it out. It's stuck there. These things, because they work in this fashion of back from RNA into DNA, have the name retroviruses. That's what these are, retroviruses.

And can anyone name a particular retrovirus?

STUDENT: HIV.

ERIC LANDER: HIV. And that's how HIV works. It turns out that David Baltimore won a Nobel Prize for the discovery of reverse transcriptase. And again, the reason I tell you about all these mechanisms is A, they're cool, and they're about biology. And B, they're medically important.

Why is reverse transcriptase not just cool but medically important? Because you could inhibit it. If you wanted to fight the HIV, if you wanted to fight the AIDS virus, you could come up with chemicals that inhibit reverse transcriptase. And of course, the cocktails that are given to patients who have been infected by HIV that now keep them alive very, very nicely include reverse transcriptase inhibitors. It's a very important thing.

If you understand the biology of retroviruses, you understand the targets you can use for drug development. And they have saved billions of lives. So again, I say this is not entirely unimportant stuff. This is kind of cool stuff, to understand how this works.

All right. Next up, transcription. Let's turn to transcription. Transcription is a little bit easier. Transcription varies. Here, eukaryotes, prokaryotes, and viruses. Let's see, let's start now with prokaryotes.

Prokaryotes, pretty simple. I have a chromosome, I have my promoter, I make my messenger RNA, I'm done. Just like I taught you before. Transcription looks just like I told you.

But for eukaryotes, it's a little weirder. Transcription, I have my RNA. I have my promoter. I make my RNA. And my RNA, it turns out, gets processed in all sorts of interesting ways.

The three ways in which eukaryotic RNA are processed-- first, there is a modification put onto the five-prime end. It is called-- well, it's basically a backwards G. It's a G triphosphate that is put on backwards, so it goes GPPP, right there.

And this is called a cap. And it's important for message recognition and stability. Eukaryotes put a funny little chemical modification there.

Eukaryotes also do something where somewhere near the end, they cut the message, and they stick on a bunch of A's as a tail at the end. And this is also important for message recognition and message stability and all that. And this is called the poly(A) tail.

Most of the names are quite reasonable. The tail of lots of A's is called the poly(A) tail. So eukaryotic messages have a cap at the front, the poly(A) tail at the back.

But the truly weird thing that they have is, if this is my eukaryotic message here, some chunks of the message are cut out and discarded entirely. They are spliced out. You might make a longer message, and whole chunks are spliced out. This is called splicing.

Splicing throws out sequences. And it could start with a long mRNA and make it a short mRNA. Now, Phil Sharp, who is on the faculty here at MIT, won a Nobel Prize some years ago for his discovery, together with someone else, of the phenomenon of splicing. So Phil is really cool. You should talk to Phil.

This splicing involves leaving some things in and excising other things. The things that are -- the things that go out are called introns. The things that stay in and are

not excised are called exons. The nomenclature is a little nuts. If it stays in, it's an exon. If it goes out, it's an intron.

As I told you, the phenomenon was discovered by Phil Sharpe at MIT. The nomenclature was proposed by Wally Gilbert at Harvard, [LAUGHTER] who's a good friend. I'm teasing Wally. I'm teasing Wally, but Wally is responsible for this nomenclature that confuses generations of students.

Why is this called an intron? Not because it stays in, but because it's an intervening sequence. It's an Intervening sequence when it is called an intron. And once something was called an intron, the other thing became an exon.

But I've got to say, for all purposes, to me, in means in an ex means out. But it's exactly backward. I've now said that a few times. That may help you remember that it goes the other way. Introns are intervening sequences, okay?

So there are some pretty impressive splicing events that go on. A typical gene, might start off 30 kilobases, 30,000 bases. And it might get spliced down to 3 kilobases, 3,000 bases.

But for example, the Factor VIII gene that encodes the factor that hemophiliacs lack, it starts off with 200 kilobases, 200,000 bases. And all but 10,000 are cut out. So it starts off 200,000, you throw out 190,000, you get down to 10.

The winner, the Duchenne muscular dystrophy gene, starts out at 2 million bases. And it gets cut down to 16,000 bases. You make 2 million bases of RNA, and you throw out almost the entirety of it and retain only 16,000.

What a waste. Why do this? Why break up your genes in patches of exons separated by big intervening spaces and then make a big RNA and splice it together? Why do something that dumb? Yeah?

STUDENT: [INAUDIBLE]?

ERIC LANDER: Sorry?

STUDENT: Are they recycled?

ERIC LANDER: The nucleotides, you mean, in the RNA? The nucleotides are recycled. But remember, you spent the trinucleotides there. So that was an energy expenditure. Well, it turns out the energy expenditure, big deal, who cares?

But it turns out that this is actually very interesting evolutionarily. I'll just tell you for a second. Maybe you'll forget it.

In a given organism, it might be more efficient to not have introns. Well, actually, there's one use for it. If I had introns, the cell could do alternative splicing. It could take the same message and splice it different ways in different cells.

Your liver might splice the message one way to make one protein. Your muscle might splice the message a different way to make another protein. So you could actually make multiple, different mature messages. That's cool, and that's used, and most genes actually have alternative splice forms. They can be spliced up in different ways.

It's also cool evolutionarily. Because it turns out that if your genes are broken up into patches like that, when random breaks happen in your genome, and this bit of chromosome attaches to that bit of chromosome, as happens sometimes-- you get hit by some little radiation, it breaks something, it puts it together-- the gene you could have a functioning gene. Because since your cell knows how to take a long message and splice it together, the fact that there was a break and a reunion, if it happens in one of those intervening sequences, would give you a new gene, and a new functional gene.

And so, in fact, some people think that this is one of the tricks evolution uses to create diversity is by breaking up its information like that into little patches of files that can then be recombined with each other in different ways. So those are the ideas. In any case, it happens a great deal.

Now, viruses. Viruses it turns out, will behave with regard to this like the organism in which they live. Prokaryotic viruses will behave like prokes. Eukaryotic viruses will

behave like eukes.

Now, let's turn to translation. How does translation work between--? Well, now here, eukaryotes, for the most part, are well behaved. A eukaryote makes an mRNA, it goes to the ribosome, the mRNA makes your protein, just like I taught you.

But now, prokaryotes are a little weird. Prokaryotes, you know, their messages are fine. None of this funny cap business, none of these poly(A) tails, none of this splicing. But here's the weirdness that prokes do. This mRNA, one mRNA might encode multiple different proteins. Multiple proteins. So this is one protein. This, multiple proteins.

So what happens is this RNA, the ribosome gets on here at a certain site. Now remember I told you the ribosome finds the first AUG? Well, it's a little more complicated. There's a particular ribosome binding site that ribosomes like to go to.

It turns out that this prokaryotic message has several of those ribosome binding sites. Ribosomes sit down. They then scan for the first AUG that they see after that point and start making a protein. So I might have several different proteins all being made by one messenger RNA, one RNA making multiple proteins.

Why would I do that? Wouldn't it be simpler, make your head hurt less, to have one gene making one protein instead of this one mRNA encoding multiple proteins?

STUDENT: They're related to one another.

ERIC LANDER: They're related. Tell me about that.

STUDENT: Well, like their functions, if they had similar functions.

ERIC LANDER: If they had similar functions, it might be very efficient to have regulatory controls that made one message. And then I get all the proteins made together, rather than having separate regulatory controls for each message. Suppose I'm a prokaryote, and I'm dividing very rapidly, and I care about having a small genome. It might be very efficient to do it this way.

So what related genes might you put together on the same message?

STUDENT: Pathway.

STUDENT: Same pathway. Maybe if you're in charge of the committee, you would say, I would like to arrange the genes encoding the multiple steps of a biochemical pathway. And that's often what happens. That's exactly what happens.

So you can get co-regulation of genes, of genes in the same pathway. And such a thing, which is called polycistronic message. Polycistronic is an unhelpful name for it, but it's what they call it sometimes, or an operon, a regulatory thing that makes many such things.

Now, why do bacteria do it? To minimize DNA. They just want to use as little space as possible. So once it's invented a regulatory control that turns on this message when I want to make arginine, well, just stick all the genes on the same message. It's cheaper, simpler, less DNA needed.

Now, who really has trouble with the amount of DNA they can have? Do you have trouble with-- Is replicating your DNA a rate limiting step in your replication as an organism? No. What is the rate limiting step in your organismal replication, your having offspring?

STUDENT: [INAUDIBLE].

ERIC LANDER: Sorry?

STUDENT: When it comes to an age that--

ERIC LANDER: It's graduating from MIT, getting a job, things like that, right? And DNA replication is not rate limiting, actually, right? Getting a degree, all that, that's rate limiting.

But for bacteria, who are reproducing not every 20 years but every 20 minutes, DNA replication is a rate limiting step. It's important to their replication. And they want compact genomes.

You, you don't care about a compact genome. It's simply not a big metabolic cost to you. Viruses-- think about viruses. Their genomes are tiny. They're much tinier even than prokaryotes.

Typical viruses might have a few thousand or tens of thousands of bases. Bacteria were millions of bases. Viruses sometimes might only have 10,000 letters of nucleic acid. They really have to use their information very compactly.

So you know what viruses do sometimes? Not all, but some? Some viruses make a messenger RNA. They have a ribosome binding site right here. And they start translating the message-- let's give it a sequence. ACUUGAGCAA, and we'll put an AUG in front of that. They can start translating here. Whoops, I'm just going to fix that.

But you know what? There's also an AUG over here. They could start using that one. And you know what else? If they found another AUG off frame, some of them use that too.

Remember, the first AUG in normal messages get used to set the phase of the codons. But in theory, I could be reading those codons in a different reading frame, shifted over by one or shifted over by two. Some viruses are so clever that they encode three different proteins smack on top of each other by reading the same nucleic acid sequence. It shifted three different ways.

It makes your head hurt to think, how could I make a functional protein? I've got to manage to get three separate functional proteins built out of a single nucleic acid sequence. One, it's easy. Just tell me the amino acids.

Now this becomes an interesting puzzle. Can I make a sequence where read one out of frame, it also makes a protein I want. And another one where-- And some viruses have evolved to do that, showing you how much they care about economizing their DNA.

All right. So what have we got? We have DNA replicates. We've seen how it replicates.

Its mostly the same by nucleic acid copying, but all these variations-- linears, circulars, single strand, double strand, RNA, reverse transcriptase, transcription, translation. Look over these examples. And what they should do is help solidify how these things work and how they're used in slightly different fashions.

Next time.