Let me talk about exons.

This is otherwise known as mRNA processing.

And that's in eukaryotes only.

Bacteria don't do this.

So what we talked about in general is you go DNA to RNA to protein.

Well this one gets DNA to RNA; you mess around with the RNA and you go to protein.

So let me show you there's an intermediate step, and people oftentimes get goofed up by this because it isn't at all clear why it should be there.

And I want to say a little bit about that at the end.

Transcription produces what we call a pre mRNA.

And you would have seen that on Gene Explorer.

And it had some strangely colored pieces that I want to now sort of illustrate here.

So your pre mRNA has a 5 prime end and it would extend to the 3 prime end

And let's just say it has ABCDEF…

Right, so those aren't DNA sequences, clearly because B is not a nucleotide.

They're like regions of the gene, so this bunch of bases, rather than writing out a sequence which would be difficult to copy down but just use a familiar string of letters to indicate those various parts.

And what you might have in this RNA is it would be organized into exon so exon one are typically drawn as blocks and intervening sequences intron one

And then you might have another Exon 2.

And then an intron number two and then an exon #3.

Right, and there's a couple things.

How do these boxes know they’re boxes or not is that there are signal sequences at the beginning of each intron.

And so these are start intron signal sequence.

And there's one of them here.

And then there's also gonna be a draw, decided you can symbol a Red Square at the end of each intron and end intron signal sequence.

Right there.

Right and the idea?

Is so there's two of these oops colors, right?

Each intron has a start and an end.

And so, just like every other process we've talked about, there's some machine that reads a sequence.

And that machine needs a start signal and a stop signal.

RNA polymerase looks for promoters and Terminators.

Ribosomes looks for start codons and stop codons.

Splicing machinery looks for start intron sequences and end intron sequences.

They don't have fancy names.

Just start and end.

So it's a particular sequence of RNA nucleotides it says, “Start an intron here and keep splicing it out until you get to this point and then stop splicing.”

And let me show you what the splicing what, what that means in the next drawing.

2 splicing.

It's based on signal sequences and introns are removed, they are depolymerized, and they are recycled.

Break them down to nucleotides and use them over.

Exons are kept and joined.

Right, and the terminology for this is just plain awkward.

It's another one of those things, right?

You might think intron means they stay in here, - no - so that's the thing you want to try to erase from your brain.

The idea is introns are intervening sequences.

They’re stuff in the middle you throw away.

Exons are expressed.

Right, that's just the way the terms got invented, so introns are intervening sequences.

So those sequences in between that we throw away and the exons are expressed.

They are kept, and So what do I mean by that?

The after splicing is done, the exons are joined.

Like this so that the regions AB&C are present.

GH and I are present and M&N are present.

Here's a 5 prime end and a 3 prime end.

And these things here, these are, uh, seamless joints.

That is, they can even break in the middle of a codon.

So let me say just sort of reiterate so sequences DE&F&JK&L are gone, removed, snipped out depolymerized out of nucleotides and recycled right and the other parts are joined seamlessly.

So if you looked at this RNA there would be no way to know that something had been snipped out there.

Like if you look in Gene X, it's just RNA.

There's no way to know that there had been an intron there, it's all been removed.

Right, and again it can break in the middle of a codon.

This is this weird thing, right?

When I when we first when I first learned about it, it was just a fact and it was what they sort of called junk DNA.

Like, why bother transcribing that huge long RNA and throwing it away?

And it turns out in some cases in many human genes you throw away up to 90% of the RNA you go to the trouble to make this humongous mRNA and 90% gets thrown away, 10% gets kept and then goes on for translation.

Why would you do that?

At the time when it was first discovered, it was not at all clear.

It's still not entirely clear.

It turns out, however.

It's beyond the scope of bio 111, but sometimes a single gene can make more than one protein by controlling which exons get kept in, which gets skipped.

Like sometimes I'll skip exon 2.

So you're missing some codons of protein proteins different, it turns out that's an economical way to have one gene make several proteins.

It just happens to be so.

So in things other than bacteria, so humans, plants, animals we have this bizarro thing where the introns are removed and the exons are kept.

And again it's done by its particular signal sequence.

Questions about that odd thing.

We talk about a little bit.

What happens?

There's a few more steps.

So when you look at GeneX, you would have seen the results of this.

If you go back and play with it again, especially for the, for the warm up.

All right, let me say this a little bit more to say about processing, because the other details you might notice if you look at RNAs in GeneX is that there's the final round of processing.

There's a cap at the five prime end and a tail at the three prime end are added.

So you've got your mRNA; your ABC GHI and MN, and there's a cap, a special nucleotide added at the five prime end and then a string of As to the three prime end, that's his tail are added.

And this is now the mature mRNA.

And here it's not again not clear why these exist.

It is likely that the CAP and the tail are added last and to signal “This thing is done being spliced. It's OK to translate.” They're like markers that say processing is done. You can go ahead and translate it because the next steps are #4 where the mature mRNA is exported from the nucleus.

And then #5 the mature mRNA is translated by ribosomes in the cytoplasm.

And so The thing is, if you think about it, the start codon’s probably in in the beginning and the stop codons near the end.

If you tried to translate a not completely spliced RNA, you’d start translating for the introns, which is all these weird junk codons.

Not something stuff that you don't want to translate, so the the Poly a tail and the cap on the end are added when splicing is done to signal “OK splicing’s done; it's OK to ship this out of nucleus, it's good to go. It's good to be translated.”

So there is this odd artifact of that's true of of human and eukaryotic genes; just not in bacteria.

People are coming to understand more about why but mostly it just is.

Alright, let me just I want to make a small summary here.

As I said, all these processes have control sequences, so the process.

Control sequence.

Make a little table here and then is it the same in all organisms?

So transcription; it's controlled by promoters and Terminator.

And these are not the same in all organisms; that is a promoter, sequence in a bacterium would not work for the RNA polymerase in in our cells, and vice versa.

Splicing is controlled by start and stop intron sequences.

These are also not the same in all organisms.

Translation by start and stop codons.

And they are the same.

OK.

That is, start and stop codons the genetic code.

It turns out and the genetic code is universal.

There are some very small exceptions.

The cool thing about genetic code is CCC encodes proline in every living thing on Earth.

Which is why recombinant DNA works that you can take a gene from a human and put it in a bacterium, and it will make the same protein because the codons are exactly the same.

The promoters and a lot of other stuff won't work, but the fundamental coding region part will work.

The other point about this is all these machines are independent of one another.

RNA polymerase knows promoters and Terminators; it doesn't know anything about codons; it doesn't know anything about splicing.

It says, “I see a promoter. I will make an RNA until I see a Terminator. I don't care what's in between.”

It minds its own business.

Splicing things they look in RNA's for starting and intron sequences, and they splice them. They don't know anything about codons or anything else down the line, they just they mind their own business. They splice things.

Likewise the ribosome it finds a start codon, and it rolls along until it makes a stop codon.

It doesn't care whether that protein is good, bad, or indifferent, it just minds its own business and follows the instructions that it's given.

There is a kind of a sequence that as transcription happens first, then splicing, then translation.