

Alternative splicing

"The essence of differentiation is the production of different sets of proteins in different types of cells."

Scott F. Gilbert, 2000

Mechanism

Regulated process that occurs during and after **transcription**.

Introns of a gene are **excluded** from the final transcript.

Exons are joined together to form the processed mRNA product.

Consequences

A single gene is **differentially expressed** into multiple proteins, creating **families** of proteins.

Evolution

Allows the human genome to direct the synthesis of **many more proteins** than would be expected from its 20,000 protein-coding genes.

Let's look a little bit more in detail at this mechanism of splicing or alternative splicing. Splicing is a process that occurs during and after transcription, and it is the process by which Introns are being excluded from the final transcript, and so-called Exons are joined together to form the processed mRNA product. Now, splicing most often, most predominantly occurs on mRNA species, but it can also happen on what are called tRNA species. The consequences of splicing are that a single gene can be differentially expressed into multiple proteins, and this can create families of proteins. On an evolutionary time scale, this, we think has allowed the human genome to direct the synthesis of many more proteins than would be expected from its roughly 20,000 protein-coding genes. How does splicing work?

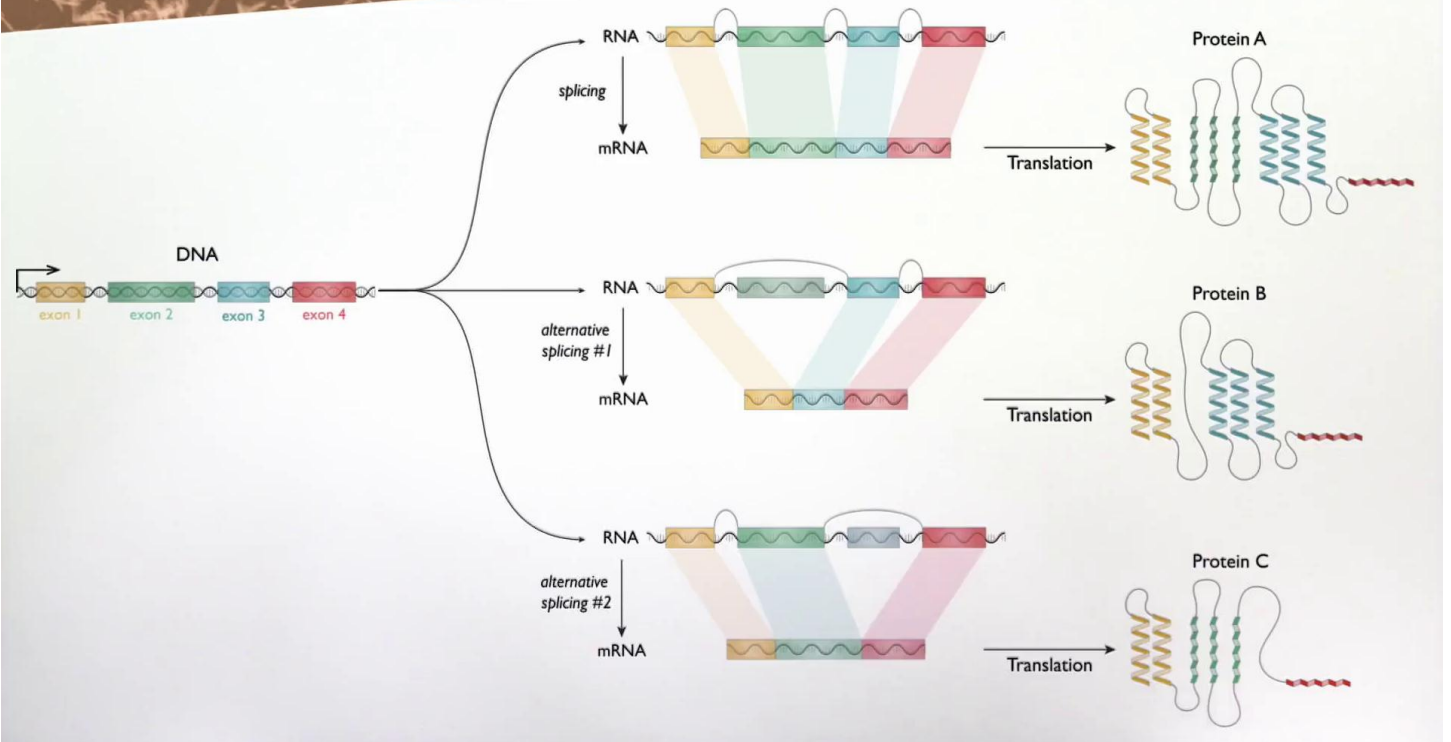
Notes

Summary



0m 05s

Alternative splicing



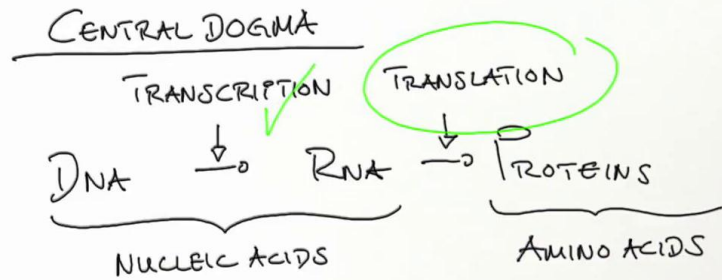
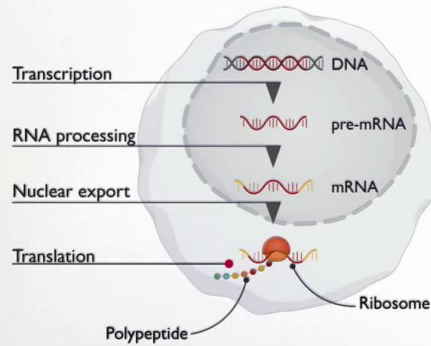
Let's assume that we have a gene that contains four exons, Exon 1 in yellow, 2 in green, 3 in blue, and Exon 4 in red. Let's assume that by splicing, one option is to stitch together all four exons. Then the ultimate mRNA product will be a concatenation of the four different exons, which will then be translated into a protein A with a specific function. But of course, alternative splicing now allows the cell to not only express this mRNA species that has all four exons, but it allows the cell to, for instance, skip Exon 2. What we do here is a process that is called exon skipping. This is one way by which alternative splicing can happen. Then as you see, the final product is now shorter, only composed of Exon 1, 3, and 4. Obviously, this, through the process of translation, gives rise to a protein with a different function. Nothing prevents the cell to not exclude Exon 3, which gives rise to yet another mRNA species, which is composed of Exon 1, 2, and 4. This can then also give rise to a protein C with yet a different function.

Notes

Summary



Translation



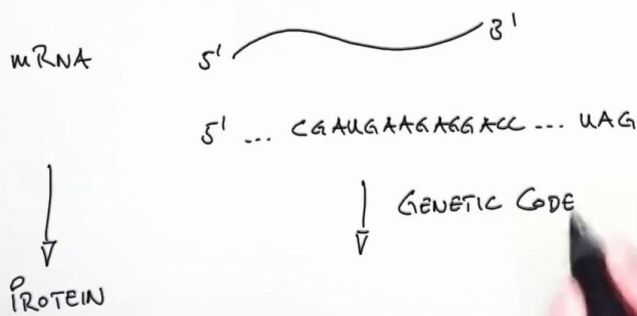
Now, we've covered transcription. Now, let's look at the second process, fundamental process of the central dogma of molecular biology, which is translation. Translation is really the process that translates from the world of nucleic acids to the world of amino acids.

Notes

Summary



Genes encode proteins



In translation, what we want to do is to go from a molecule of mRNA to a molecule of protein. Let's draw this molecule of mRNA single-stranded, five prime to three prime, and we want to go to the protein. Now, let's give this mRNA species a sequence, five prime, CG, AUG, AAG, AGG, ACC, and so on, until a sequence that is called UAG. Now the question is, how does the cell know how to translate this sequence of letters into a sequence of amino acid. The way the cell does this is by using a lookup table which is called the genetic code.

Notes

Summary



2m 55s

The genetic code

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G	Third letter
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G	

redundancy
1 start codon
3 stop codons

What the cell does is to use this lookup table or genetic code and scan the string of letters and look for what are called triplets that are composed of three different letters. We have a first letter here on the left, a second letter on the top, and a third letter. It does this. What the cell knows by using this genetic code is that there is one specific start codon that is called AUG.

Notes

Summary



3m 56s

Genes encode proteins

A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA		AGA		A
	AUG	Met	ACG		AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA		GGA		A
	GUG		GCG		GAG	Glu	GGG		G

mRNA

5' ... C G AUG A A G A G G A C C ... U A G 3'

START

GENETIC CODE

PROTEIN

The cell screens this string of letters and finds the first code that signals start, which would be here, AUG, and this signals start.

Notes

Summary



4m 29s

The genetic code

		Second letter				Third letter
First letter		U	C	A	G	
	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	

redundancy
1 start codon
3 stop codons

Not only does the signal start, but it gives already the information of which amino acid has to be paired with this code, AUG, and in this case, it's the amino acid methionine. Once this is done, the cell switches back to the genetic code and starts to read the second triplet, which in our case would be AAG. The way to read this is to take the first letter here, A, second letter A, and then go to the third letter, which is here, G. Then the cell knows that this triplet signals that it has to transcribe this triplet into the amino acid lysine.

Notes

Summary



Genes encode proteins



A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

And so on. Then the next triplet will be AGG, ACC. This will give rise to the following sequence of amino acids arginine in here and threonine here and so on. It's important to realise that the cell reads this mRNA molecule in so-called triplet, which also get the name of codons. One triplet of nucleic acid gives rise to one amino acid and just the same way the cell knew where to start by having a so-called start codon, it also knows where to stop by having a so-called stop codon. In this case, it will be UAG, which signals the cell to stop transcription. Now, it turns out that there are multiple stop codons. This can also be UAA or UGA, but all of those will signal stop. Now, the stop codon, unlike the start codon, is not translated into an amino acid, but it really is just the end of transcription.

Notes

Summary



The genetic code

		Second letter				Third letter
First letter		U	C	A	G	
	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	

redundancy
1 start codon
3 stop codons

The three-stop codes are depicted in red here. We see UAA, UAG, and UGA, which are not translated into an amino acid, but just signal the cell to stop translation. Then what we also realise immediately here is that there is a certain redundancy. We have three different stop codons. If we look at the amino acid leucine, we can see that it can be coded by as many as six different codons.

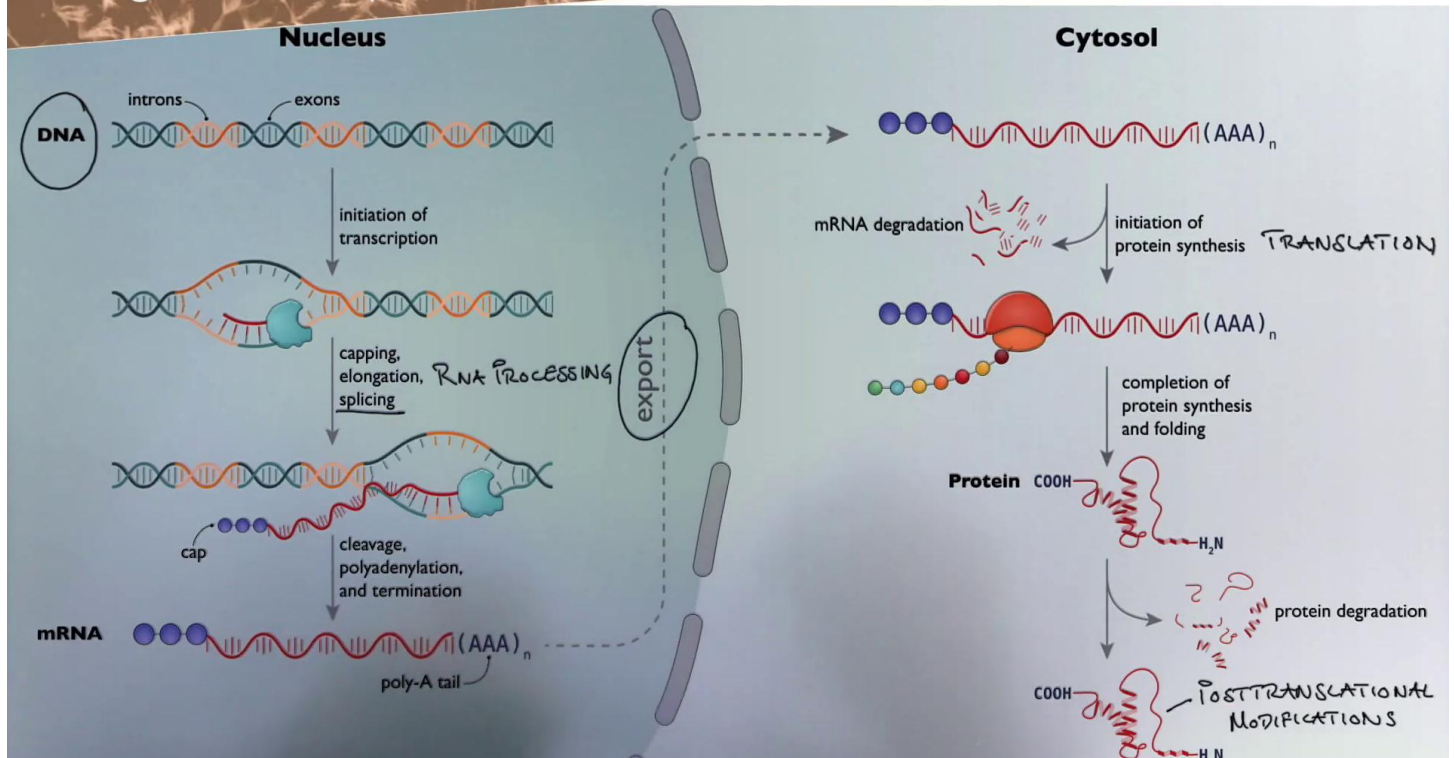
Notes

Summary



7m 04s

From genetics to proteins



If we want to go from genes to proteins, which means from the basic information of genetic information that is encoded on the DNA, then what we have to do is to initiate transcription. We will have a step of RNA processing some of the elements which we have seen today, such as splicing, and then we have the termination site which signals the cell to stop transcription. The final mRNA product will then be exported by an intricate and highly controlled mechanism out of the nucleus into the cytosol where protein synthesis can start. Protein synthesis is nothing else than translation. Once and actually already during translation, protein folding will start and once a protein is completely folded, then it can reach its final destination. But it's important also to keep in mind that there can be what is called post-translational modifications and protein degradation mechanisms that will influence the functionality of a protein and its half-life. As you can see, this is a highly complex and tightly regulated mechanism that brings a genetic information via transcription and via translation into a functional product.

Notes

Summary



7m 34s