

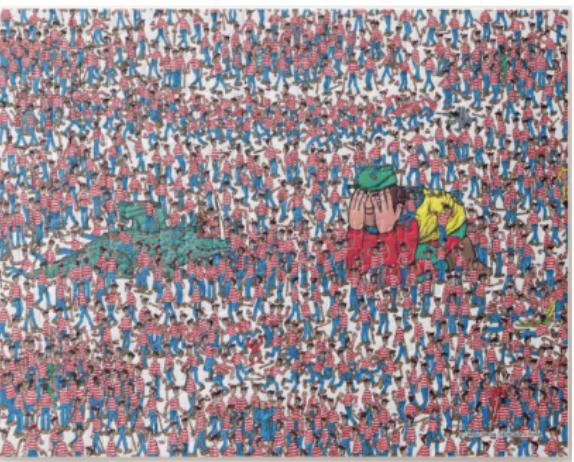
*Remember, everything on the lecture slides and the accompanying notes  
is potentially examinable!*

# FRST302: Forest Genetics

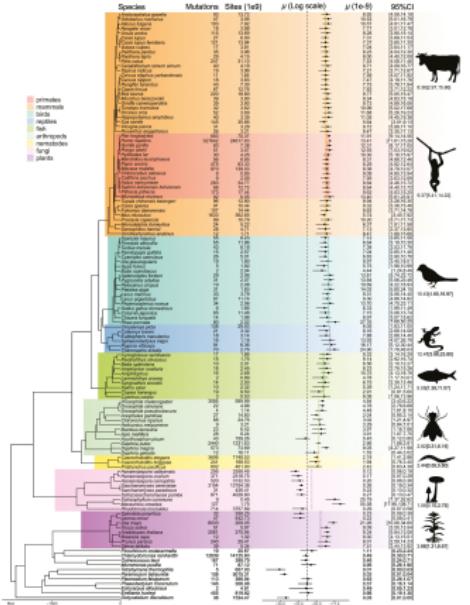
---

Lecture 1.5: Gene Expression

- Introduction to different sequencing methods
- An introduction to genomics in conifers
- The difficulty of repetitive DNA for genomic analysis
- The pros and cons of different sequencing methods



# Why do mutations have the potential to influence phenotypes?



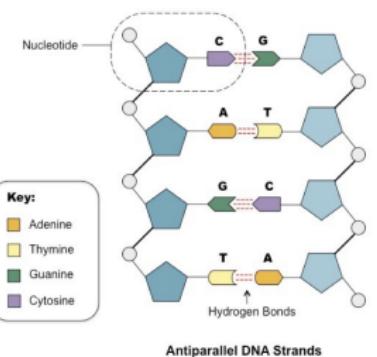
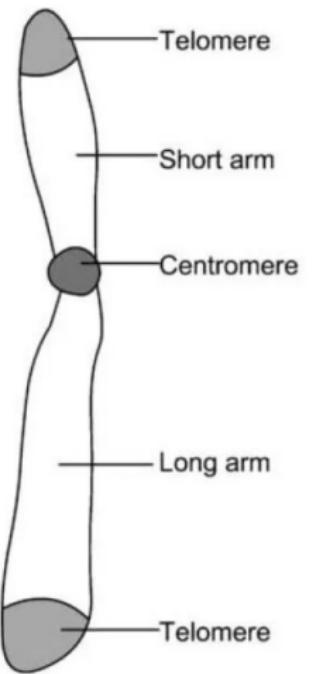
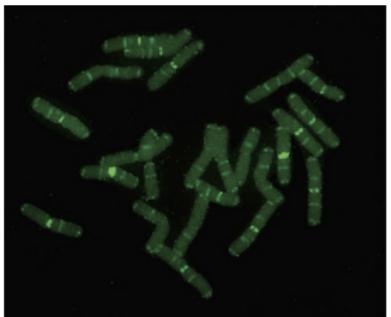
The genome is the “blueprint of life”, so alterations to it can influence phenotypes



---

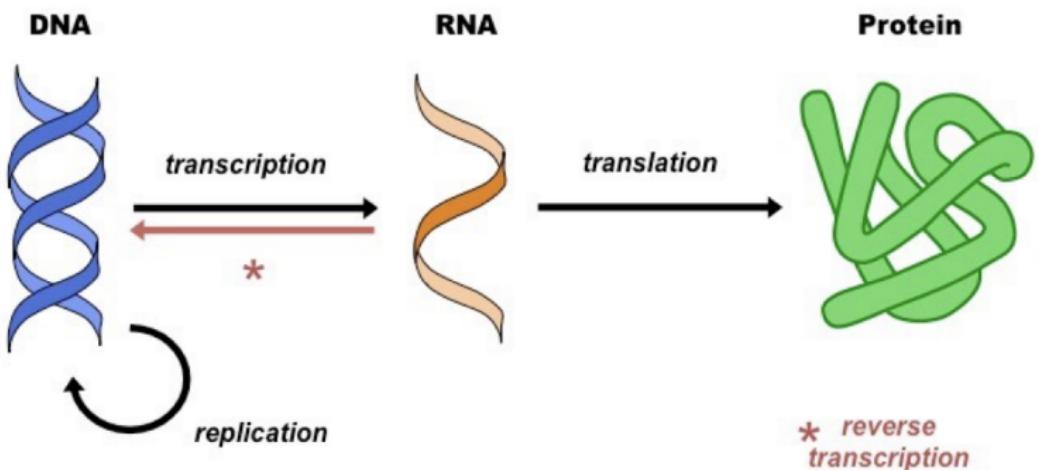
What is a phenotype? Any measurable characteristic of an individual. But remember that an individual's phenotype represents a combination of its environment and its genotype

# What's in a Genome?



But what does the DNA actually do?

*"DNA makes RNA, and RNA makes protein"*

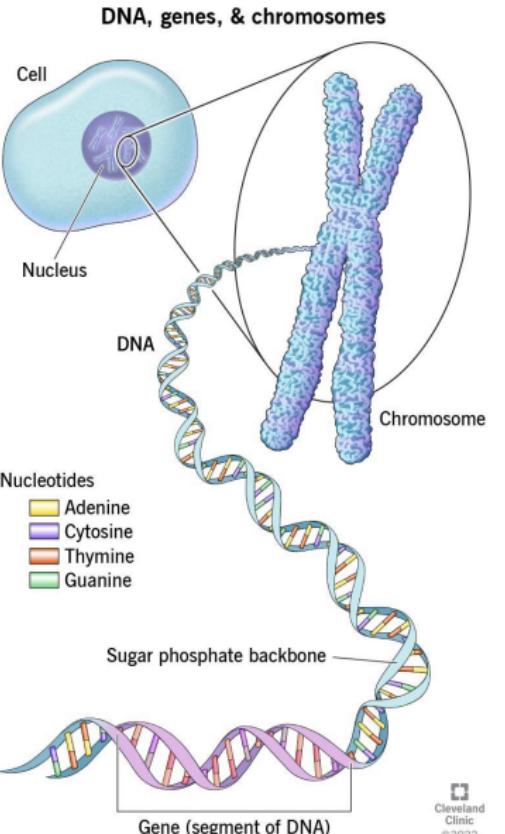


It means that there is a one-way flow of information  
(*But exceptions abound!*)

A **gene** is a section of DNA that controls a certain trait by encoding **proteins**

**Proteins/enzymes** do much of the work in the cell and the body

- **Enzymes** are proteins that catalyze chemical reactions
- Some proteins give cells their shape and structure and others carry out processes like energy conversion and photosynthesis.

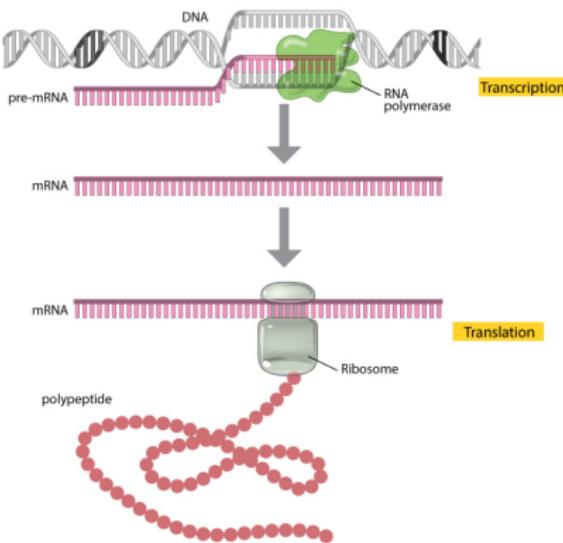


# Genes - Transcription and Translation

Genes are expressed through the processes of **transcription** and **translation**

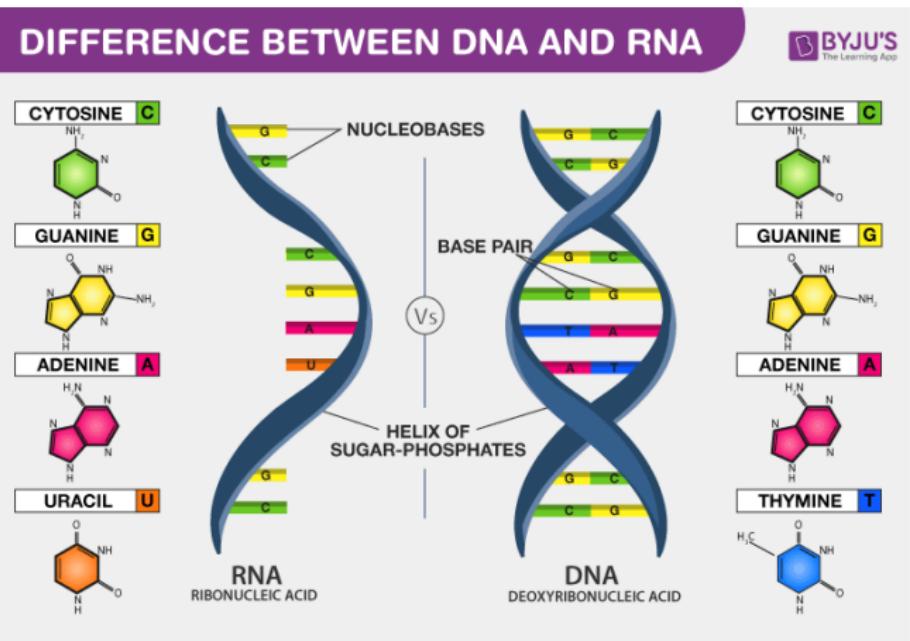
In **transcription**, DNA is used to produce a single stranded m(essenger) RNA molecule (note the pre-mRNA)

In translation, the specific set of nucleotides in the mRNA are used to build polypeptides (i.e. protein)



Ribonucleic acid (RNA) is present in all cells

Structurally similar to DNA, but usually single stranded



The genomes of many viruses are encoded by double-stranded RNA (e.g. rotavirus)

## Types of RNA

RNA plays a large number of roles in species' lifecycles,  
characterising these is an area of very active research

mRNAs – messenger RNAs transcribed from genes

tRNAs – transfer RNAs for transferring amino acids

rRNAs – ribosome RNA involved in protein synthesizing (most abundant)

ncRNAs – non-coding RNA

lncRNAs – long non-coding RNAs (>100 base) seem to be involved in more or less everything!

miRNAs – microRNAs (21-23 bases) involved in mRNA silencing (gene regulation)

piRNAs - Various roles including suppression of transposable elements

**All RNAs are encoded by DNA, but not all DNA encodes RNA!**

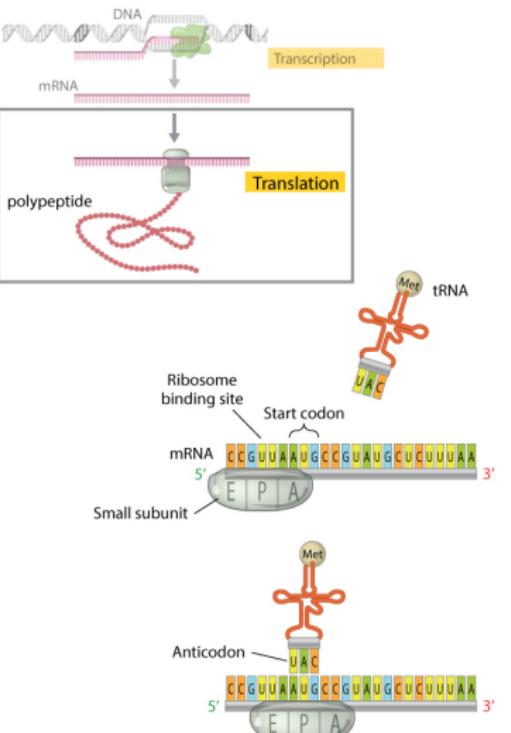
# Genes - Translation

In translation, ribosomes are recruited to synthesize the polypeptide

The sequence of the mRNA serves as the template

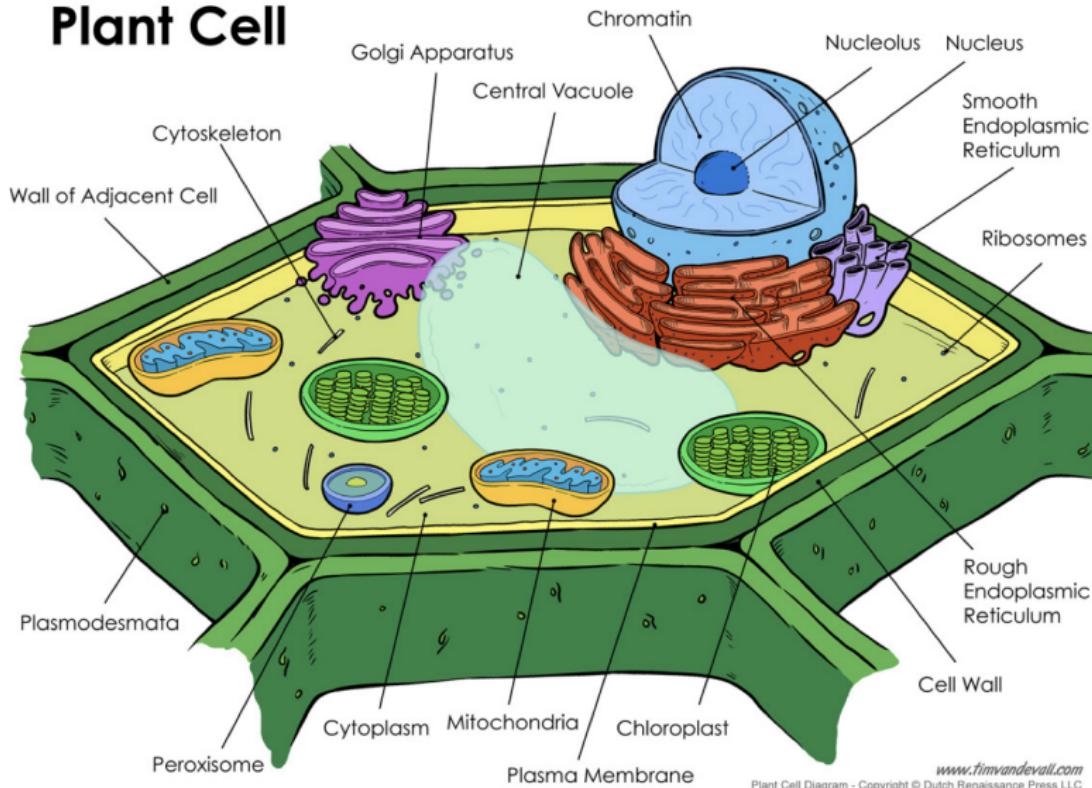
t(transfer)RNA matching both amino acids and the genetic code of the mRNA are recruited by the ribosome

Watch this video in your spare time if you want to see a visualisation:  
<https://youtu.be/gG7uCskU0rA>



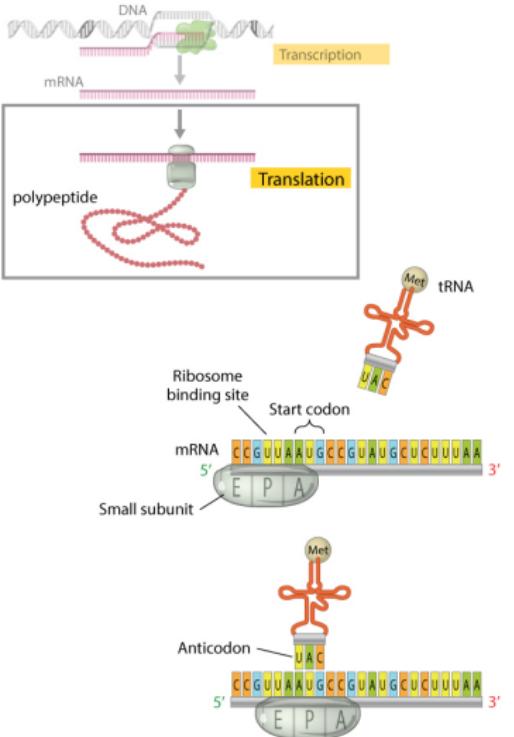
See <https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393/> for a more complete primer

## Plant Cell



[www.timvandevall.com](http://www.timvandevall.com)  
Plant Cell Diagram - Copyright © Dutch Renaissance Press LLC

# Genes - Translation and the Genetic Code

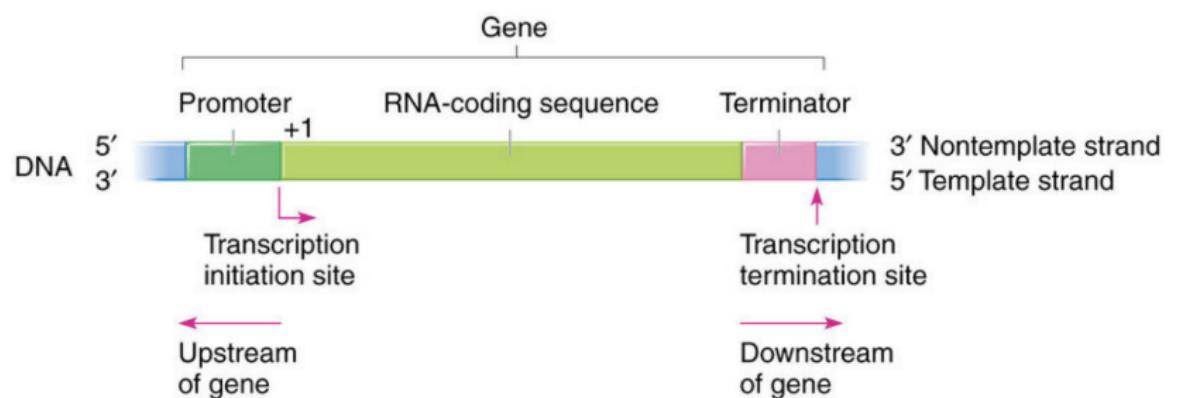


Second nucleotide				
U	C	A	G	
U	UUU <b>Phe</b> UUC UUA <b>Leu</b> UUG	UCU UCC UCA UCG	UAU <b>Tyr</b> UAC UAA <b>STOP</b> UAG <b>STOP</b>	UGU <b>Cys</b> UGC UGA <b>STOP</b> UGG <b>Trp</b>
C	CUU CUC <b>Leu</b> CUA CUG	CCU CCC CCA <b>Pro</b> CCG	CAU <b>His</b> CAC CAA <b>Gln</b> CAG	CGU CGC CGA <b>Arg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

- Each possible set of three RNA nucleotides (i.e. codon) corresponds to a particular amino acid
- Note the synonymous codons!
- Codons are written from 5' to 3'
- AUG (i.e. Methionine) is the start codon

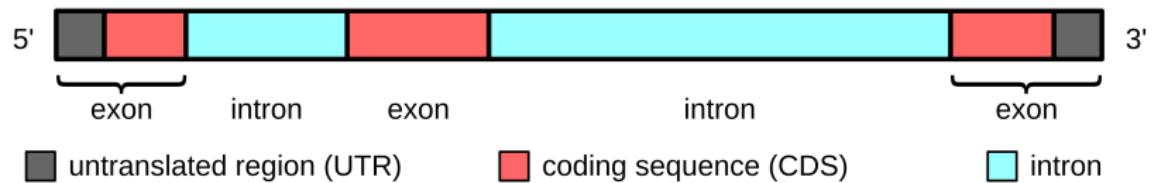
See <https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393/> for a more complete primer

# Gene Structure



© 2010 Pearson Education, Inc.

## Gene Structure - Exons/Introns



**Exon** - DNA/RNA sequence that encodes the amino acid sequence

**Intron** - Non-coding DNA/RNA interspersed between exons in most genes

**UTR** - Untranslated regions of the mRNA that allow for binding of the ribosome (5') and for translation termination (3')

# Douglas-fir Genes

Total genes	51,419
Average gene length (bp)	17,967.11
Median gene length (bp)	1,962
Multiexonics	41,595
Monoexonics	9,824
Longest intron (kb)	778,429
Average number of exons per multiexonic gene	4.73



Think about all the different tissue types, cell types and developmental steps that characterize a Douglas-fir throughout its life

**How does a set of 51,419 proteins do all of those things?**

Questions?

Let's take a short break

# The Man in Me

Bob Dylan

The musical score for "The Man in Me" by Bob Dylan consists of two systems of music. The top system starts at measure 72 in A-flat major (A-flat, D-flat, E-flat) and includes staves for Stimme (vocals), Orgel (organ), Akustische Gitarre (acoustic guitar), Elektrische Gitarre (electric guitar), Klavier (piano), Elektrischer Bass (electric bass), and Schlagzeug (drums). The bottom system begins at measure 73 in E-flat major (E-flat, B-flat, E-flat) and includes staves for Stimme, Orgel, Git (guitar), E-Git (electric guitar), Pfp (percussion), E. B. (double bass), and Schlgz. (drums). The vocal parts feature lyrics such as "La la la la" and "The". The score uses a mix of standard notation and rhythmic patterns.

If genes are the orchestra, gene expression is the music

# Components of Gene Regulation

Chromatin accessibility

Transcription

Promoters, enhancers,  
silencers

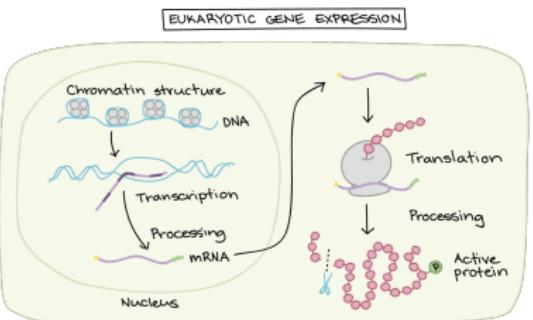
RNA processing

RNA stability

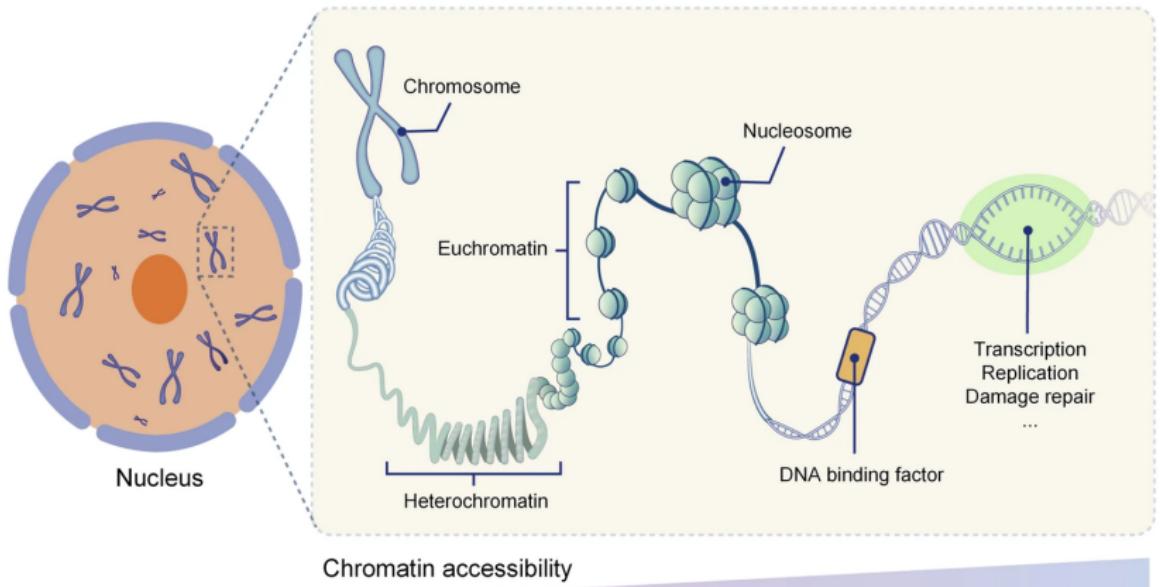
The longer it stays, the greater  
chance of translating the  
protein

Translation

Protein activity



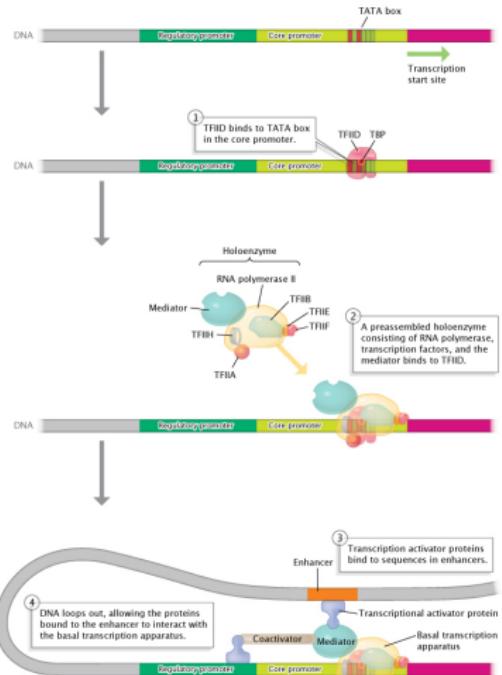
# Chromatin Accessibility



In order for transcription to occur, a gene needs to be located in accessible chromatin

The dynamics of regions becoming more or less accessible is characteristic of eukaryotic organisms

# Gene Regulation - Promoters and Enhancers



## Promoters

100-1000bp long, located just upstream of transcription start site

TATA box usually located 5-bp upstream of transcription start site

Contains specific sequences that bind proteins to initiate transcription

## Enhancers

Potentially located far away from transcription start site

Contains specific sequences that bind proteins to amplify or make transcription more likely

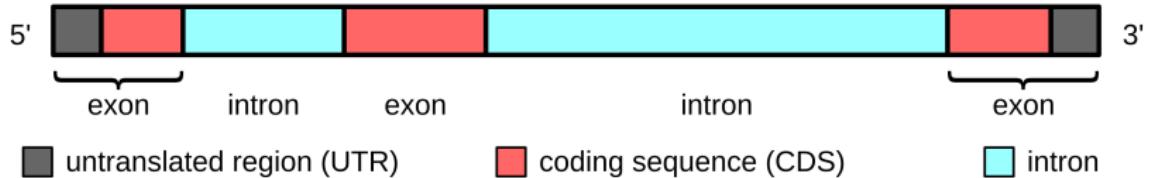
## Silencers & Insulators

Other binding sites that can influence or prevent transcription at particular sites in the genome

Contains specific sequences that bind proteins and influence transcription

See a more detailed primer here: <https://www.nature.com/scitable/topicpage/what-is-a-gene-colinearity-and-transcription-430>

## Gene Structure - Exons/Introns



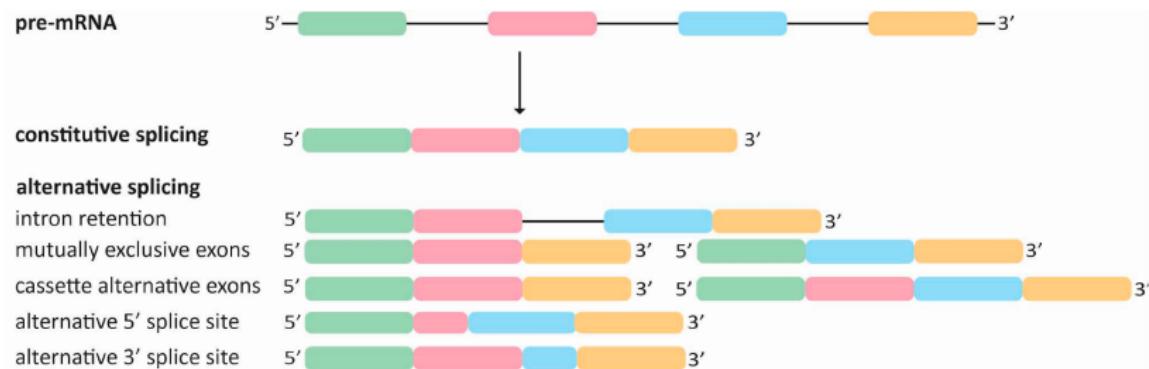
**Exon** - DNA/RNA sequence that encodes the amino acid sequence

**Intron** - Non-coding DNA/RNA interspersed between exons in most genes

**UTR** - Untranslated regions of the mRNA that allow for binding of the ribosome (5') and for translation termination (3')

## Gene Structure - Alternative Splicing

A multi-exonic gene is capable of making many different isoforms (or splice variants) that potentially encode different protein products





Every cell contains the same genes in a multicellular organism, how do cells know whether to make a pollen or seed cone, a shoot or a root?

The **timing, location and extent** of gene expression are controlled

The set of genes **expressed** in a cell determines the properties and the functions of the cell

In eukaryotes, gene regulation can occur at many steps

Questions?

Let's take a short break

## How Do Mutations Influence Phenotypes?

Well, you could mutate an exon potentially influencing the structure of an important protein, or you could influence the binding affinity of a promoter, an enhancer, a silencer



Changes in DNA can potentially influence what proteins do as well as how and when they are regulated

Functionally characterising the effects of individual mutations is extremely time-intensive and difficult for an organisms like Douglas-fir

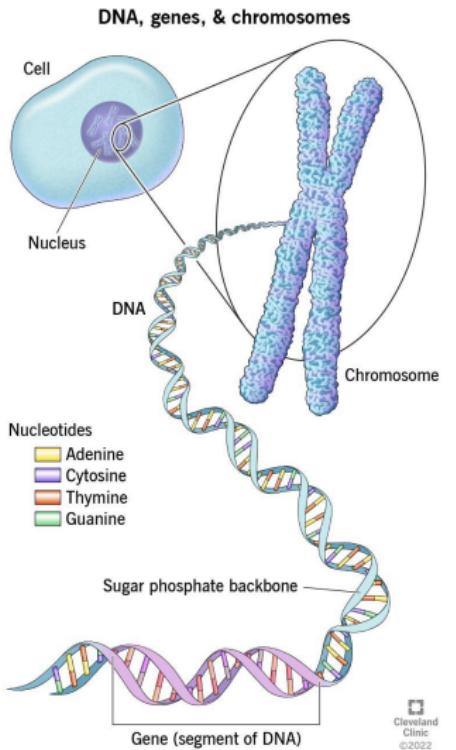
Logan is right, but genetic variation comes in many forms and these different forms may influence genes and their expression in different ways

How can we locate genomic regions that could potentially influence phenotypes?

# Identifying Genes in DNA Sequences

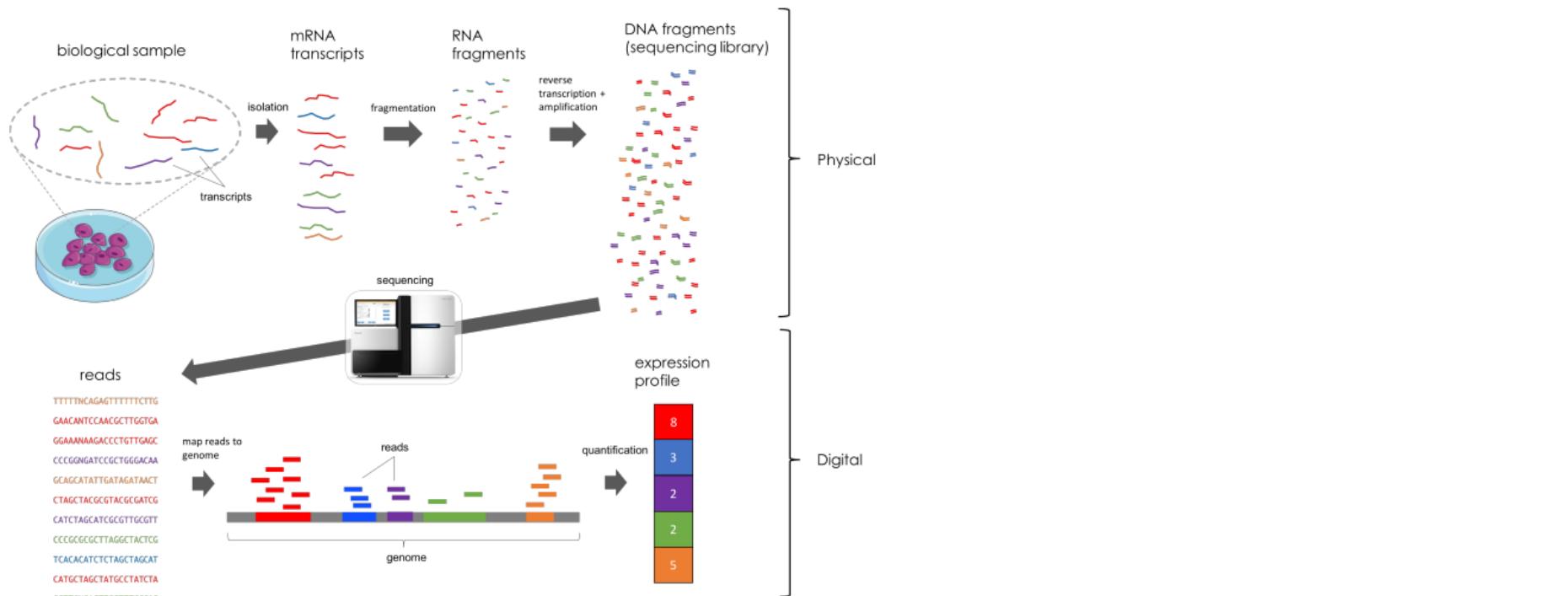
## Two Fundamental Methods

- Search for start/stop codons remember the genetic code!
- Sequence RNA and map that to the genome



We refer to the identification of genes in a genome as "annotation"

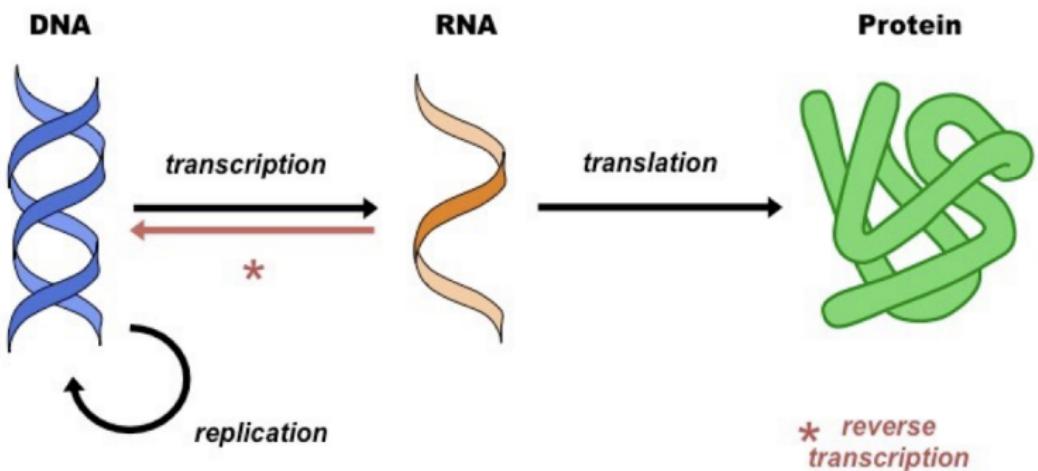
# RNA Sequencing - Aligning to Reference Genome



Note the use of reverse transcription to convert RNA to DNA!

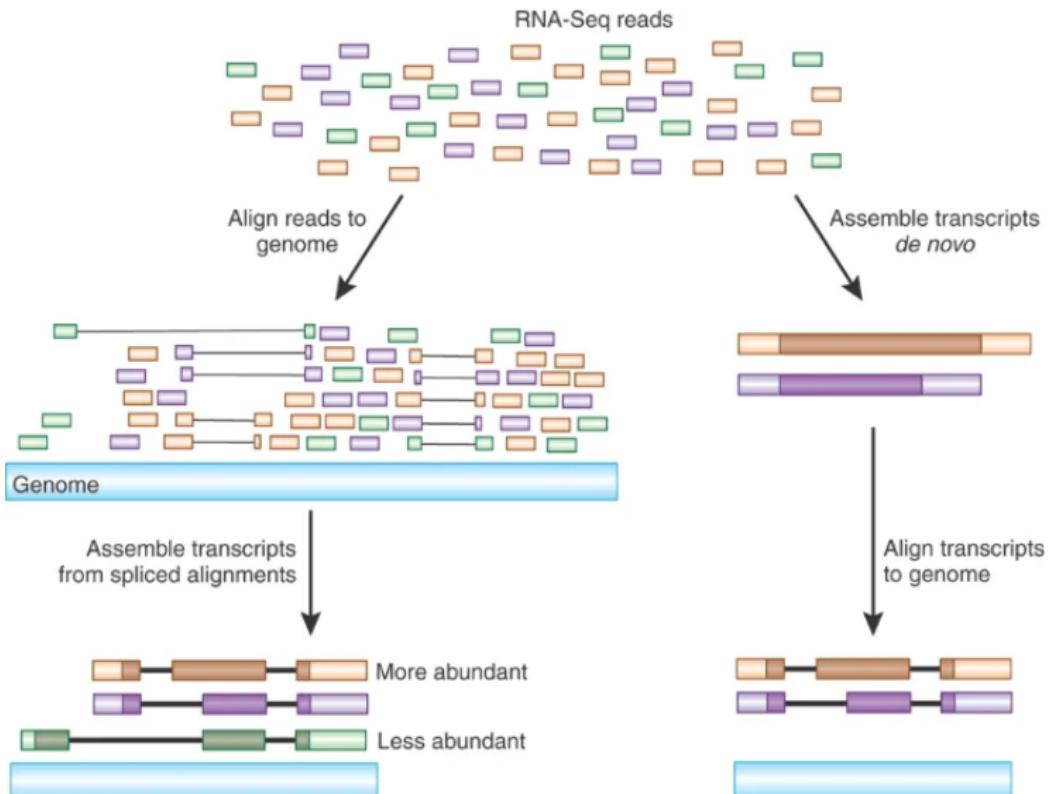
Image from: [https://mbernste.github.io/posts/rna\\_seq\\_basics/](https://mbernste.github.io/posts/rna_seq_basics/)

*"DNA makes RNA, and RNA makes protein"*



It means that there is a one-way flow of information  
(*But exceptions abound!*)

# RNA Sequencing



# Douglas-fir Genes

The following data comes from a recent annotation of the Douglas-fir genome

Total genes	51,419
Average gene length (bp)	17,967.11
Median gene length (bp)	1,962
Multiexonics	41,595
Monoexonics	9,824
Longest intron (kb)	778,429
Average number of exons per multiexonic gene	4.73



51,419 genes with an average length of 17,967 bp that gives us an estimate that 924 Mbp of the Douglas-fir genome codes for protein

That's roughly 5% of the 16 Gbp genome - **what's the rest of it doing?**

# Identifying Gene Regulatory Elements

If particular regions have functional roles (e.g. enhancers, promoters etc.) we expect that mutating them would be harmful most of the time

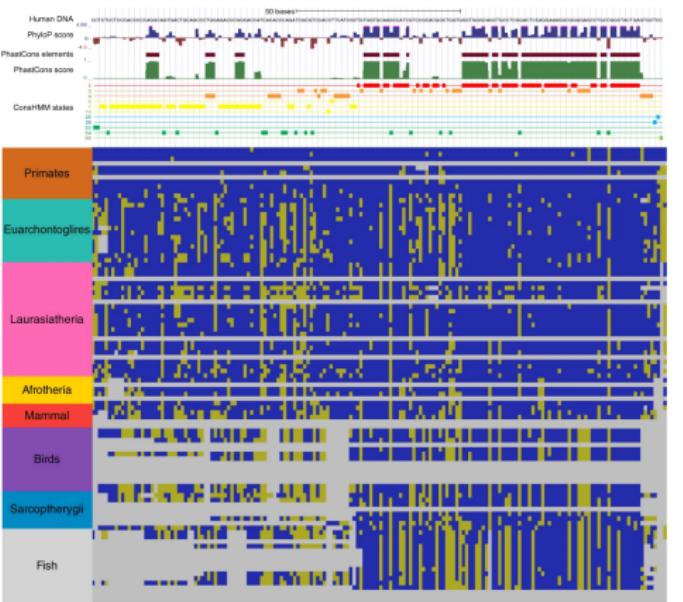
For that reason, we expect that **functional** regions of non-coding DNA to be conserved through evolutionary time

**a**

Human	T	T	T	C	C	T	G	A	C	T	T
Chimp	T	T	T	C	C	T	G	A	C	T	T
Bushbaby	T	C	T	G	C	T	T	C	C	T	T
Rat	-	C	T	T	C	T	G	A	T	T	-
Alpaca	-	-	C	C	T	T	G	C	A	T	-
Megabat	T	C	-	C	C	T	G	A	T	T	T
Parrot	-	-	-	-	-	-	-	-	-	-	-

- Aligning and matching the human sequence
- Aligning but not matching the human sequence
- Not aligning to the human sequence

**b**



Conserved non-coding elements identified in this way overlap with regions of accessible chromatin

# Noncoding ("Junk") DNA

Noncoding DNA sequences are components of an organism's DNA that do not encode proteins

Noncoding DNA makes up the vast majority of the total DNA in Douglas-fir, the precise fraction varies a lot among species

Beyond genes and gene regulatory elements, tree genomes are filled with stuff:

- Some noncoding DNA is transcribed into non-coding RNAs
- Transposable elements (see last lecture)
- Pseudogenes - the remnants of old genes left in the DNA sequence



- Describe gene structure
- The various roles of RNA
- Describe gene expression
- Identifying functional regions in a genome