https://participant.turningtechnologies.eu/en/join
https://go.epfl.ch/TurningPointPoll

Session ID: **bio411a**

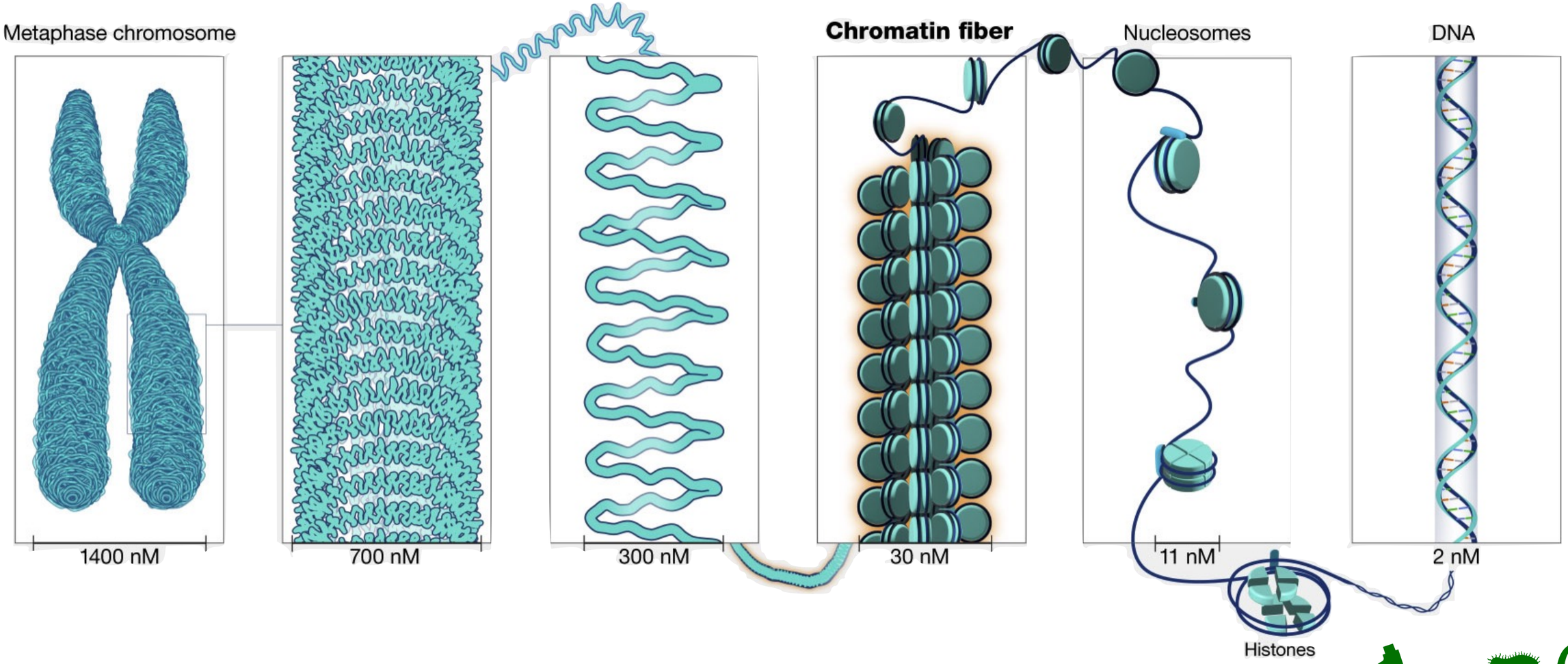
Decoding & Understanding Genomes

What is a Genome?

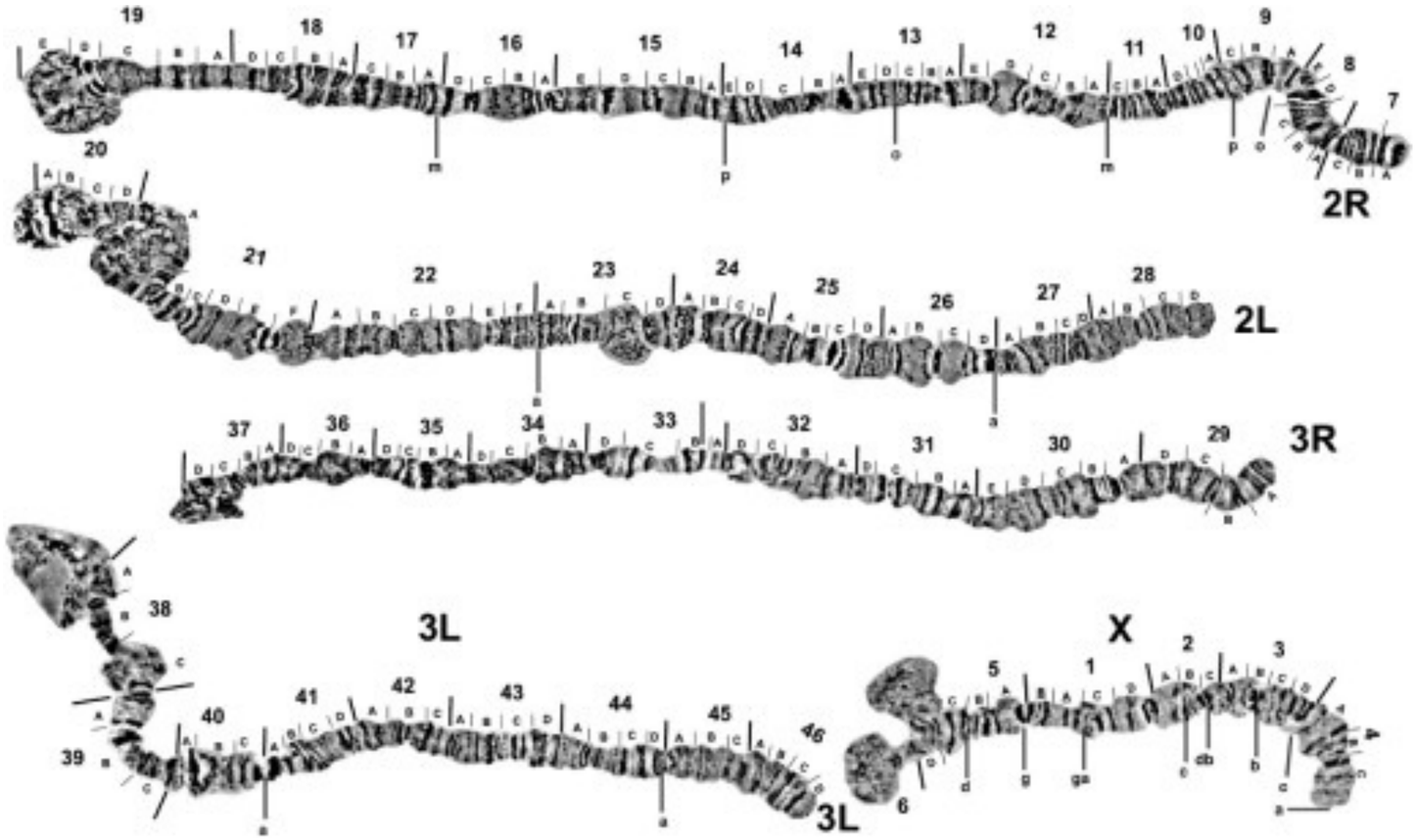
- A **complete set** of an organisms DNA is called its genome
- Human **diploid** genome is 6.37/6.27 (female/male) Gigabase pairs (Gbp)
- Weighs **6.41 picograms** and end-to-end would be **205 cm** long
- Full copy in the majority of cells in the body

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

How is DNA packed into the cell nucleus?



Aside - Polytene Chromosomes



Occur when many chromosomes align with each other without cell division

Allowed the visualisation of chromosome bands leading to the idea that genes resided on chromosomes.

Found in pea plants (aka Mendel) and Drosophila (aka Morgan).

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Which is correct?

A. A DNA $\longrightarrow$ RNA $\longrightarrow$ Protein

B. B DNA $\rightleftharpoons$ RNA $\longrightarrow$ Protein

C. C DNA $\rightleftharpoons$ RNA $\rightleftharpoons$ Protein

B. RNA can make DNA

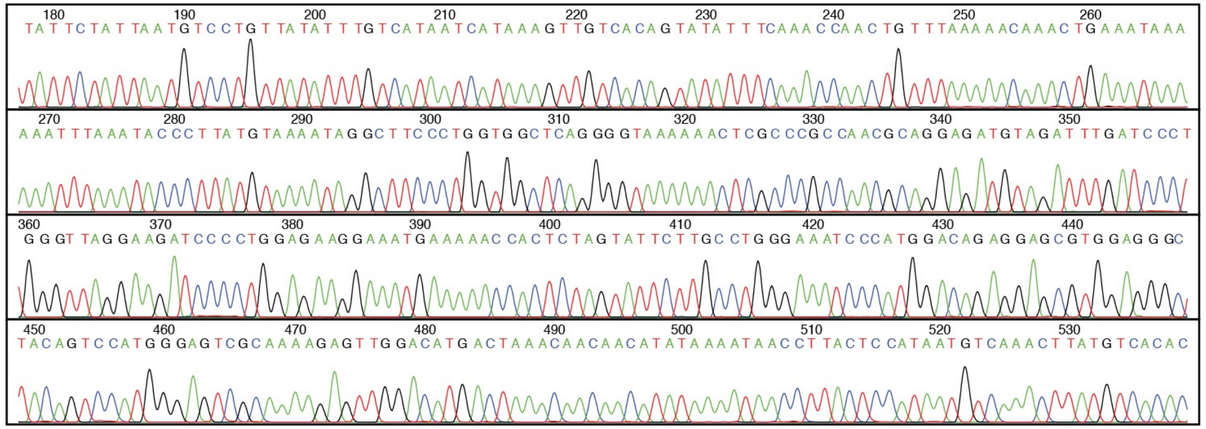
A. DNA ⟶ RNA ⟶ Protein

B. DNA ⇌ RNA ⟶ Protein

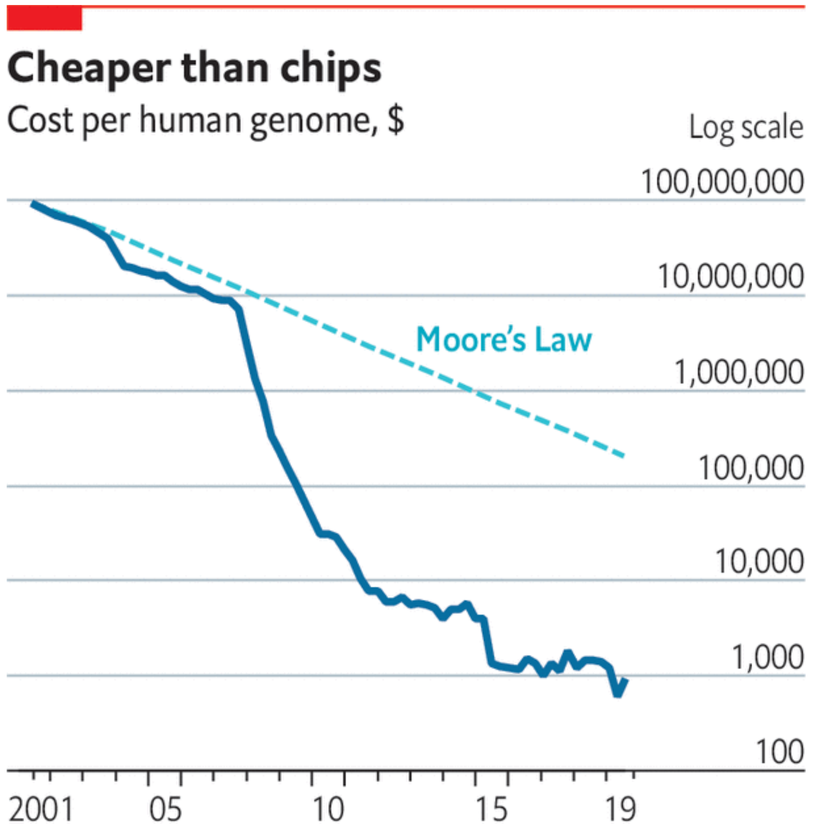
C. DNA ⇌ RNA ⇌ Protein

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

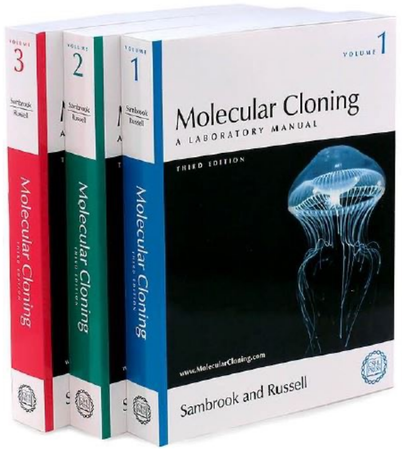
DNA Sequencing

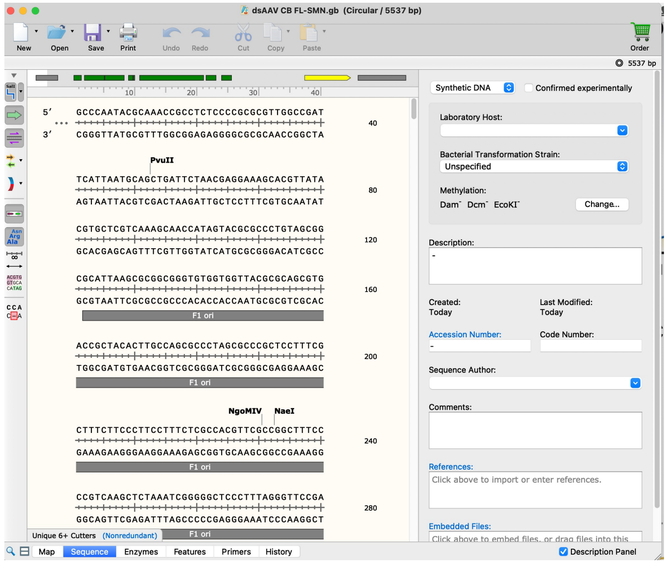


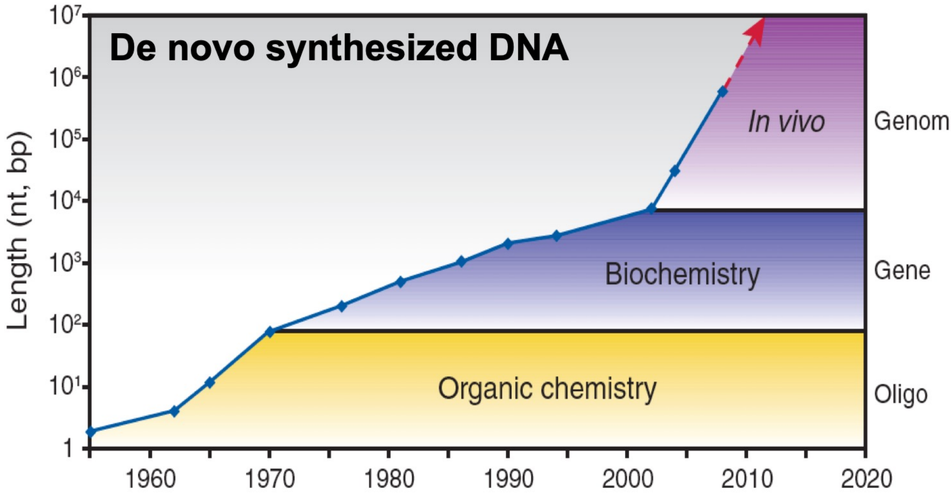
DNA sequence data from an automated sequencing machine



DNA Synthesis is also now cheap





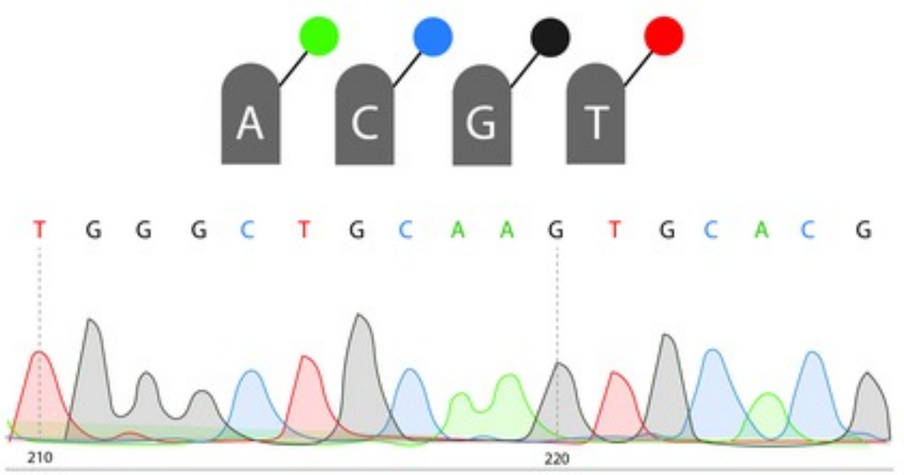


GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Sequencing technologies

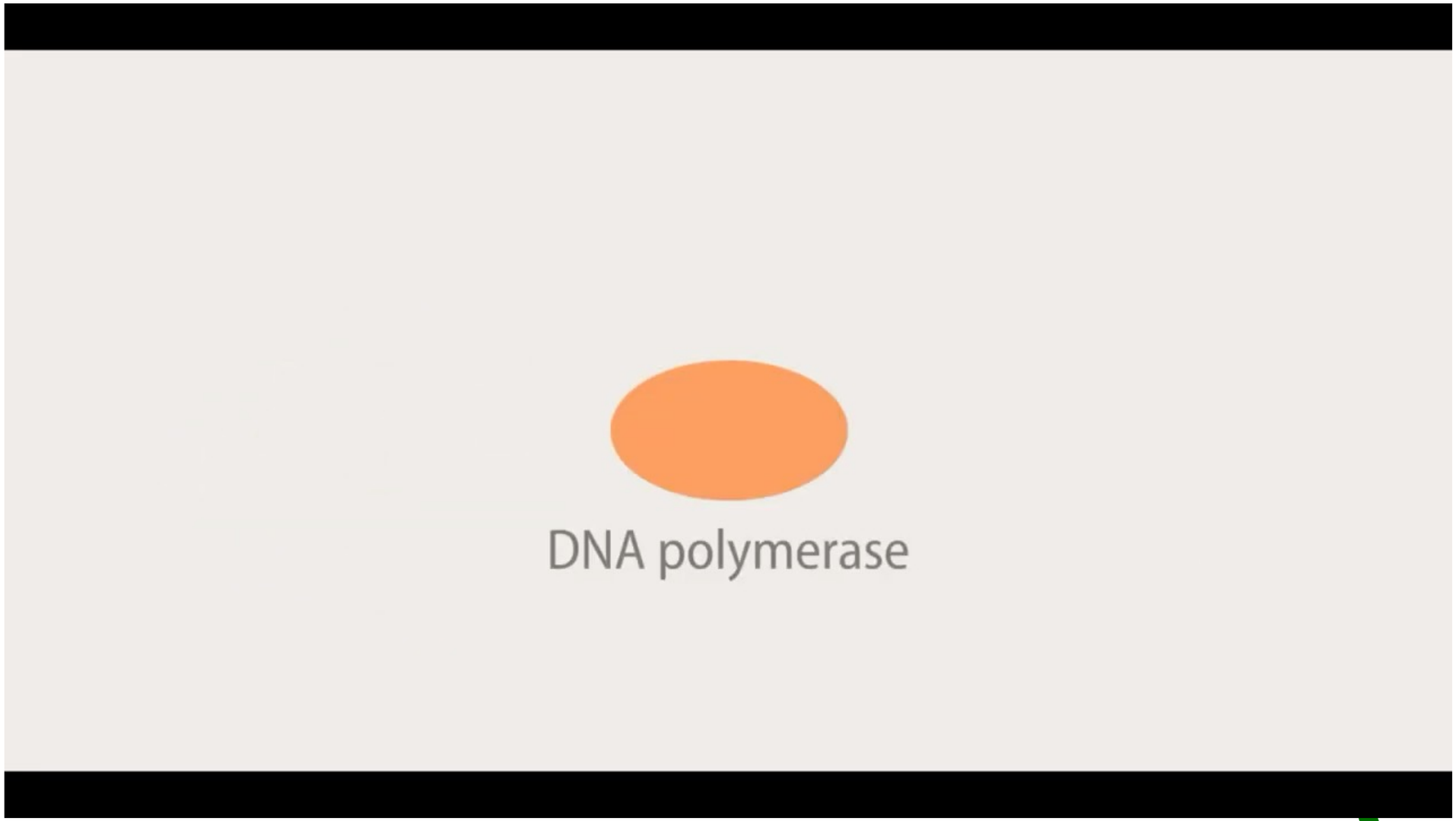
Sanger Sequencing

- Sometimes called chain-termination sequencing or dideoxy sequencing.
- Based upon in vitro DNA replication
- Utilises random incorporation of modified, fluorescently tagged bases onto the growing DNA strand.
- The 4 standard bases are tagged with a different fluorophore so they can be distinguished from one another.



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Sanger Sequencing



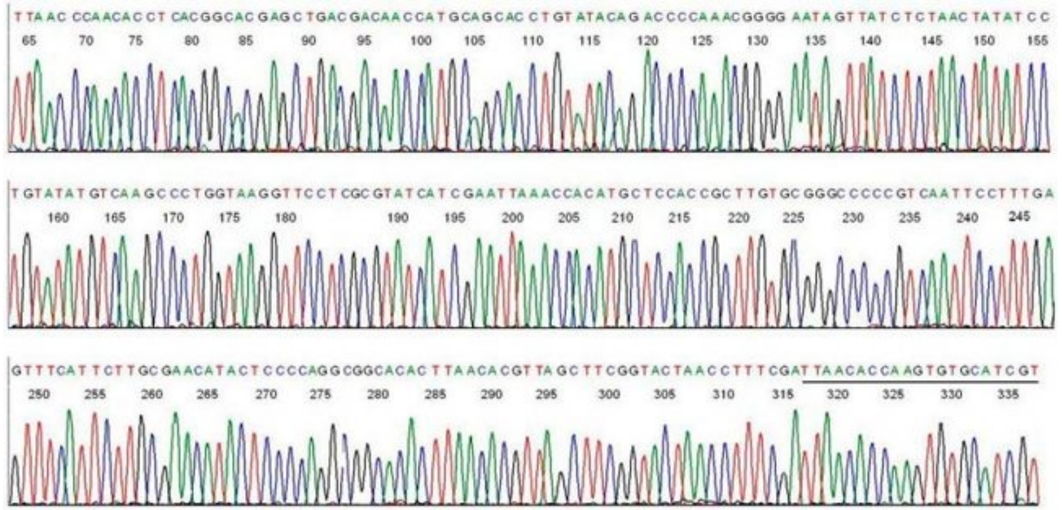
Adapted from - https://www.youtube.com/watch?v=gDdoGcGS014

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

DNA Sequencing technologies

Sanger Sequencing

- Used for routine sequencing
- Reads are 500 -1000bp
- Efficient and reliable
- Low throughput



TOPICS in LIFE SCIENCES

NGS DNA Sequencing technologies

Next Generation Sequencing technologies

- High-throughput, multi parallel sequencing
- Up to 600 billion bases in one reaction
- Like Sanger sequencing, is based upon in vitro DNA replication
- Does not require known primers

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

NGS DNA Sequencing technologies

Next Generation Sequencing technologies

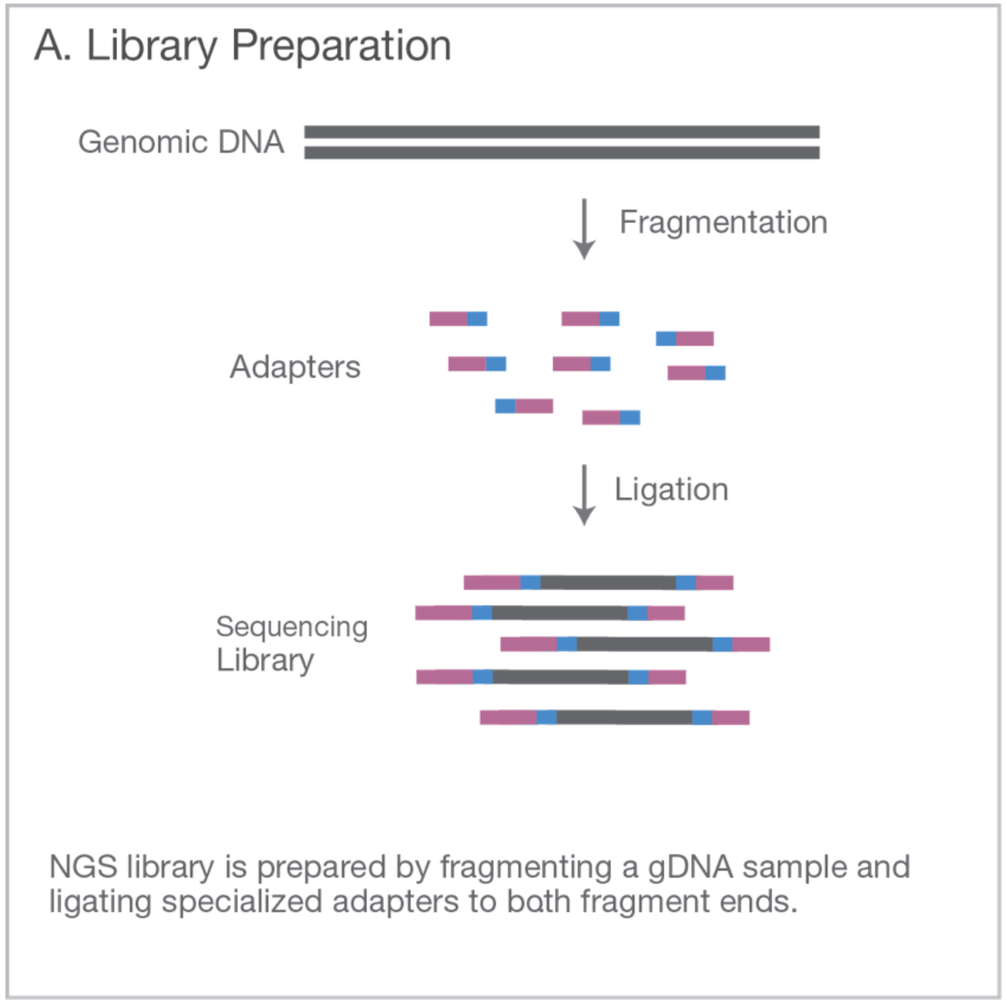
- Different sequencing chemistries (sequencing by synthesis, sequencing by ligation, pyrosequecing and ion semiconductor sequencing)
- All require samples to be prepared into libraries
- All require specialised machines with solid surfaces to create ‘clusters’ of DNA
- Machines output raw data and increasingly this is directly used for bioinformatic analysis

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

NGS DNA Sequencing

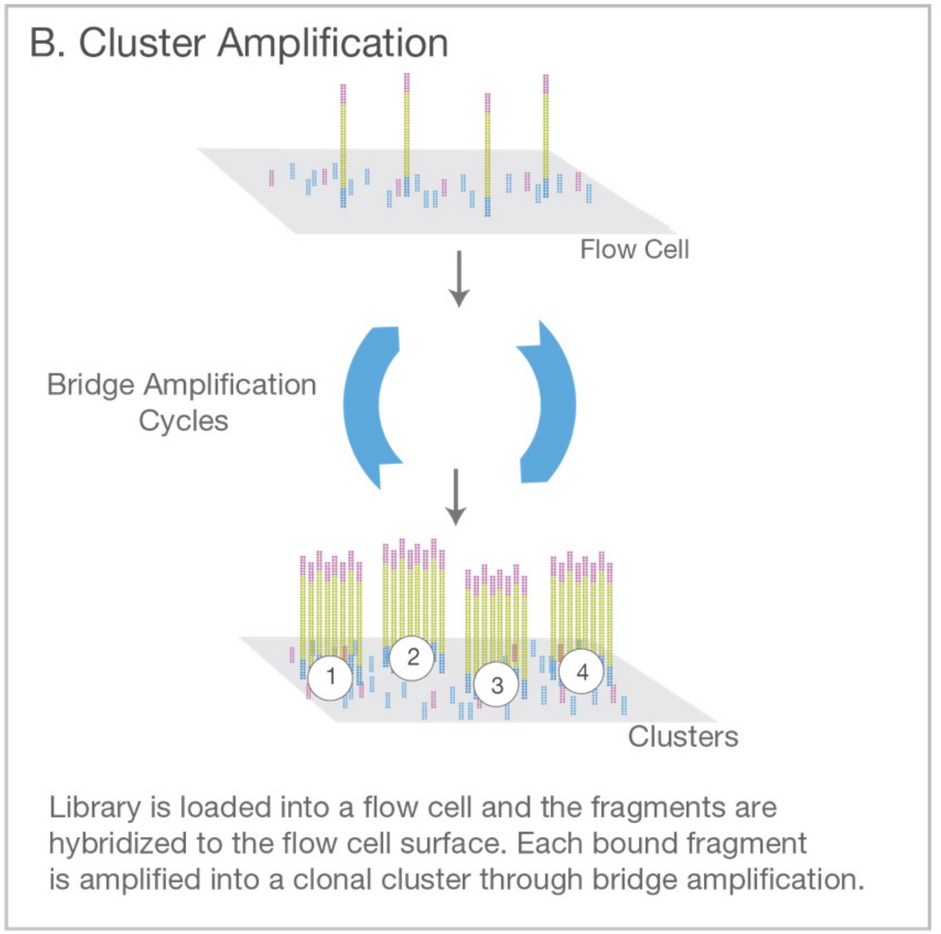
sequencing by synthesis

- Genomic DNA is fragmented
- 5′and 3′adapters are ligated
- Adapter-ligated fragments are PCR amplified.



NGS DNA Sequencing

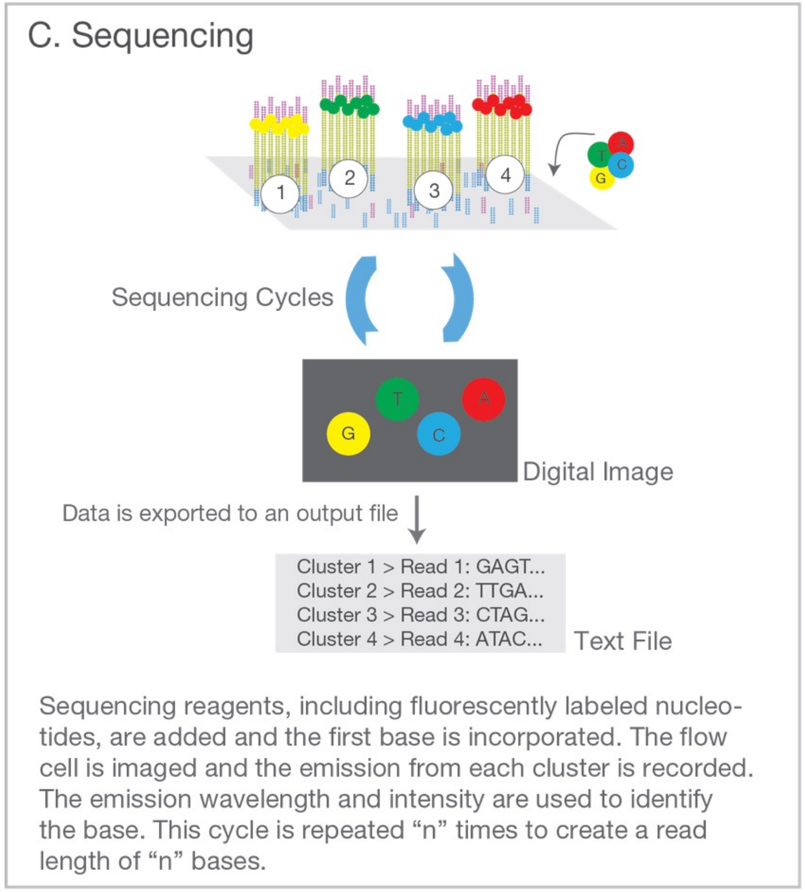
- Library is loaded into a flowcell
- Fragments are captured on a lawn of surface-bound oligos complementary to the library adapters
- Each fragment is then amplified into distinct, clonal clusters.



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

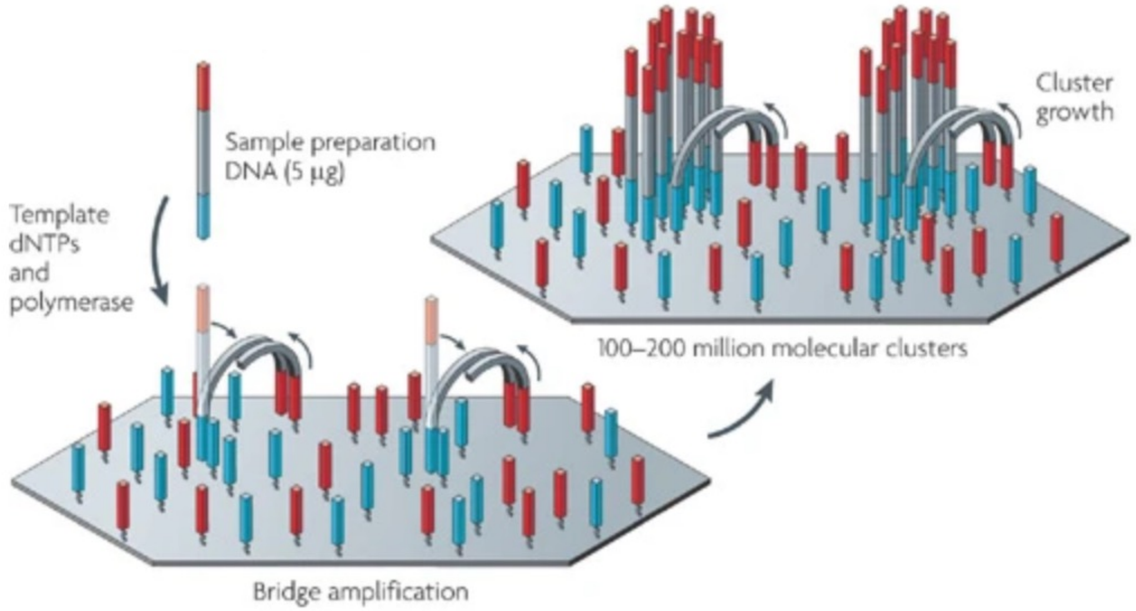
NGS DNA Sequencing

- All four terminator–bound dNTPs are present during each sequencing cycle
- Unlike Sanger sequencing, these modified bases can be converted back to a ‘regular’ bases and do not halt the reaction i.e. no normal bases required



NGS Sequencing

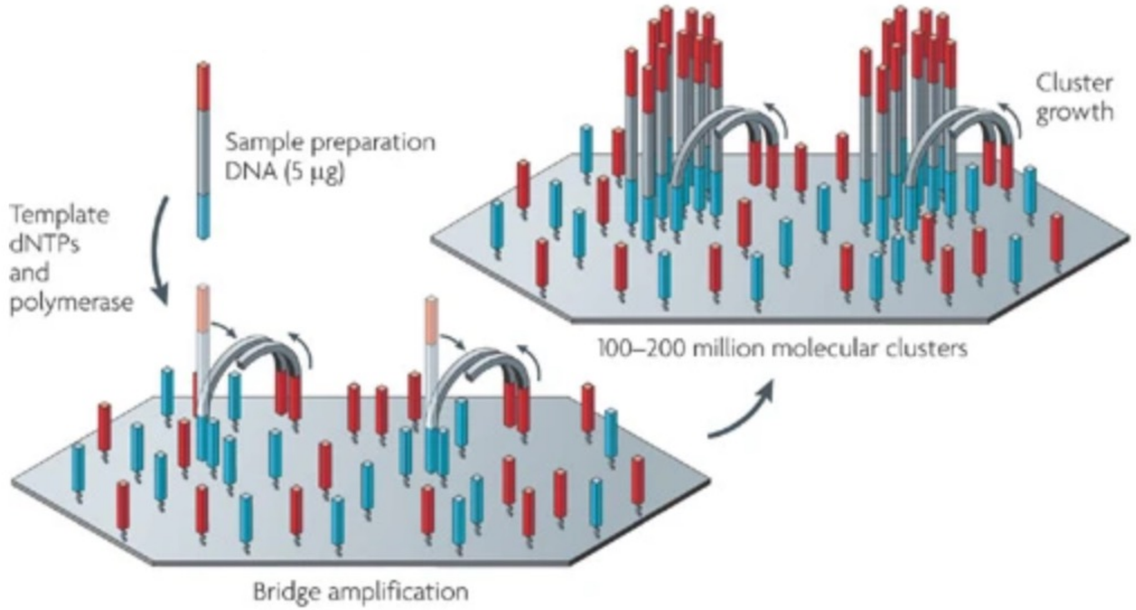
- Used for high throughput analysis
- highly accurate sequencing that greatly reduces context–specific sequencing errors, even within repetitive sequence regions
- Expensive and requires dedicated machines
- Not suitable for routine sequencing
- Enables personal genomics



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

NGS Sequencing

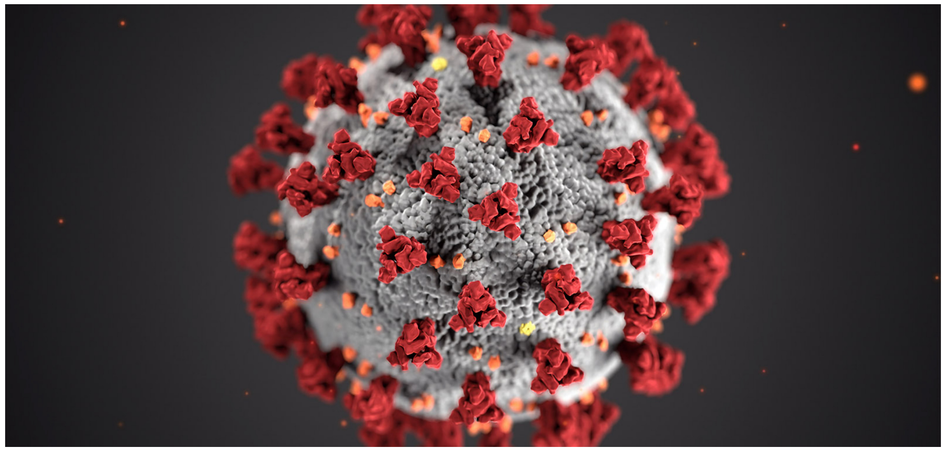
- Used for high throughput analysis
- highly accurate sequencing that greatly reduces context–specific sequencing errors, even within repetitive sequence regions
- ~~Expensive~~ and requires dedicated machines
- Not suitable for routine sequencing
- Enables personal genomics



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Novel Applications

- Identification of potential pathogens
  e.g. new viruses



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

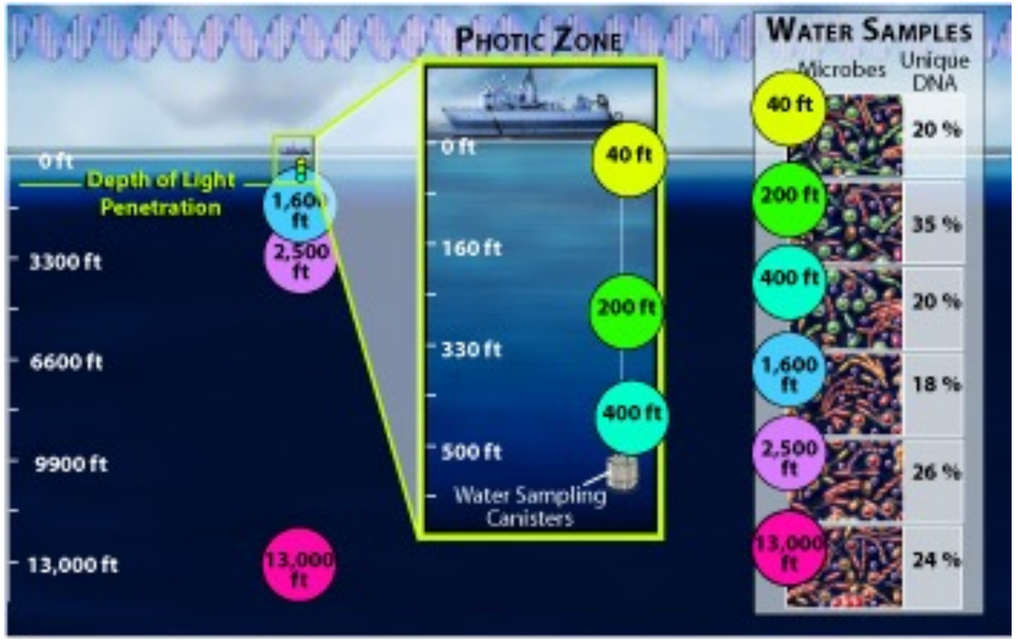
Novel Applications

- Identification of potential pathogens e.g. new viruses
- Disease monitoring for public health

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Novel Applications

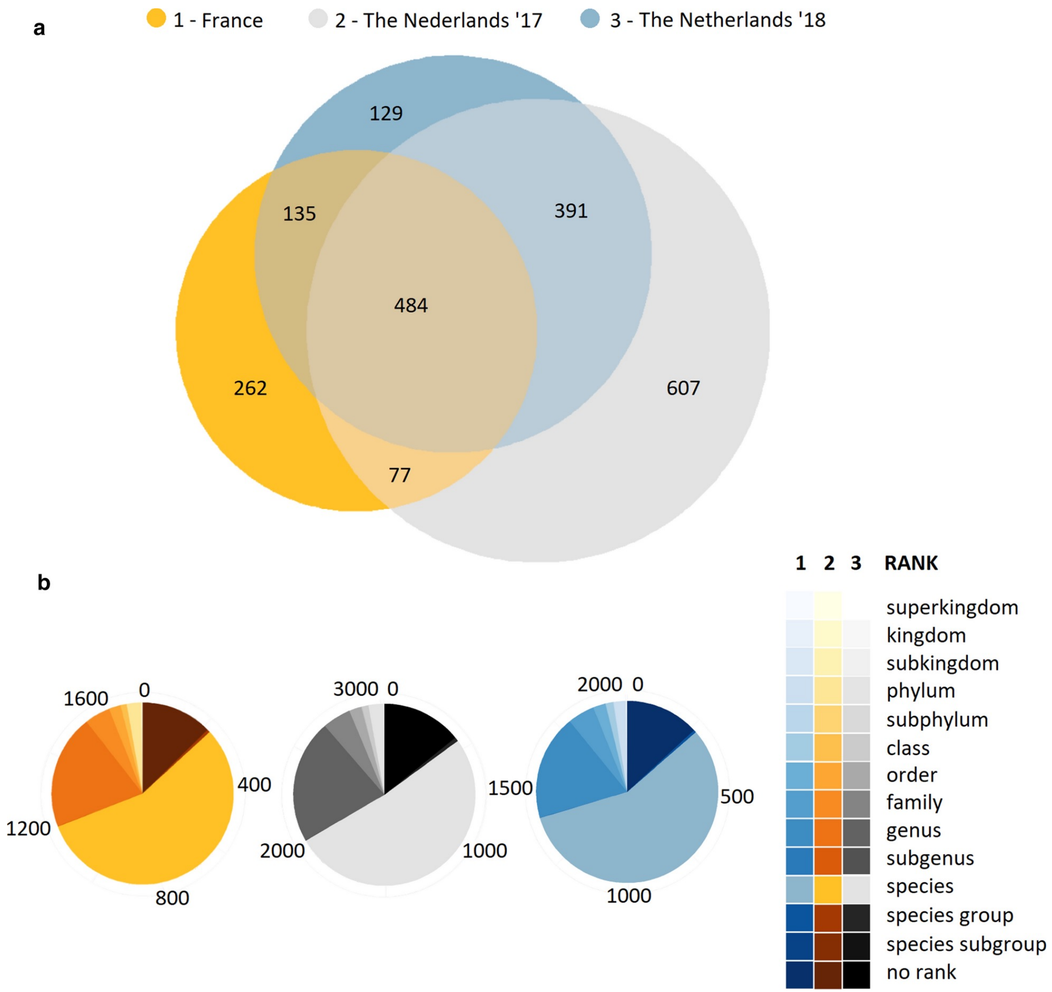
- Identification of potential pathogens e.g. new viruses
- Disease monitoring for public health
- Identification of new species and novel molecules



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Novel Applications

- Identification of potential pathogens e.g. new viruses
- Disease monitoring for public health
- Identification of new species and novel molecules
- Monitoring environmental change



Liem, M et al (2021). https://doi.org/10.1186/s13104-021-05457-3

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Reconsidering DNA technologies

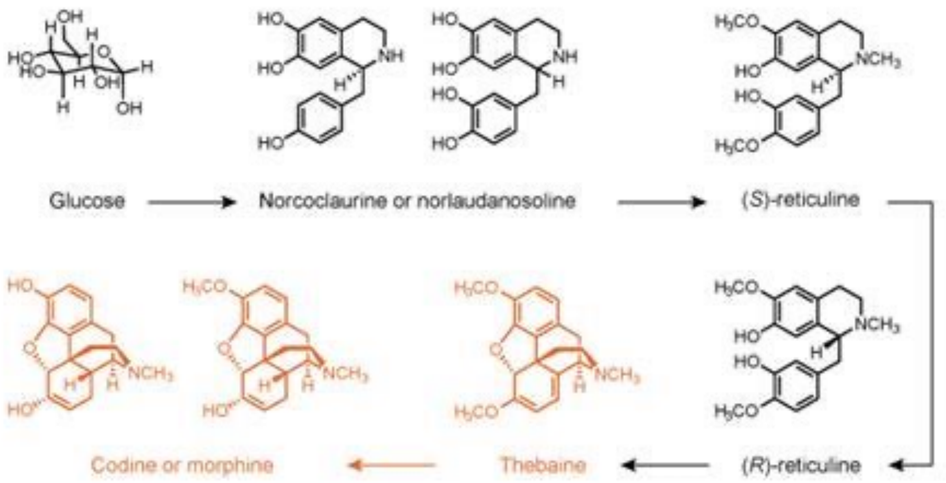
NGS sequencing is now commercially cheap and fast

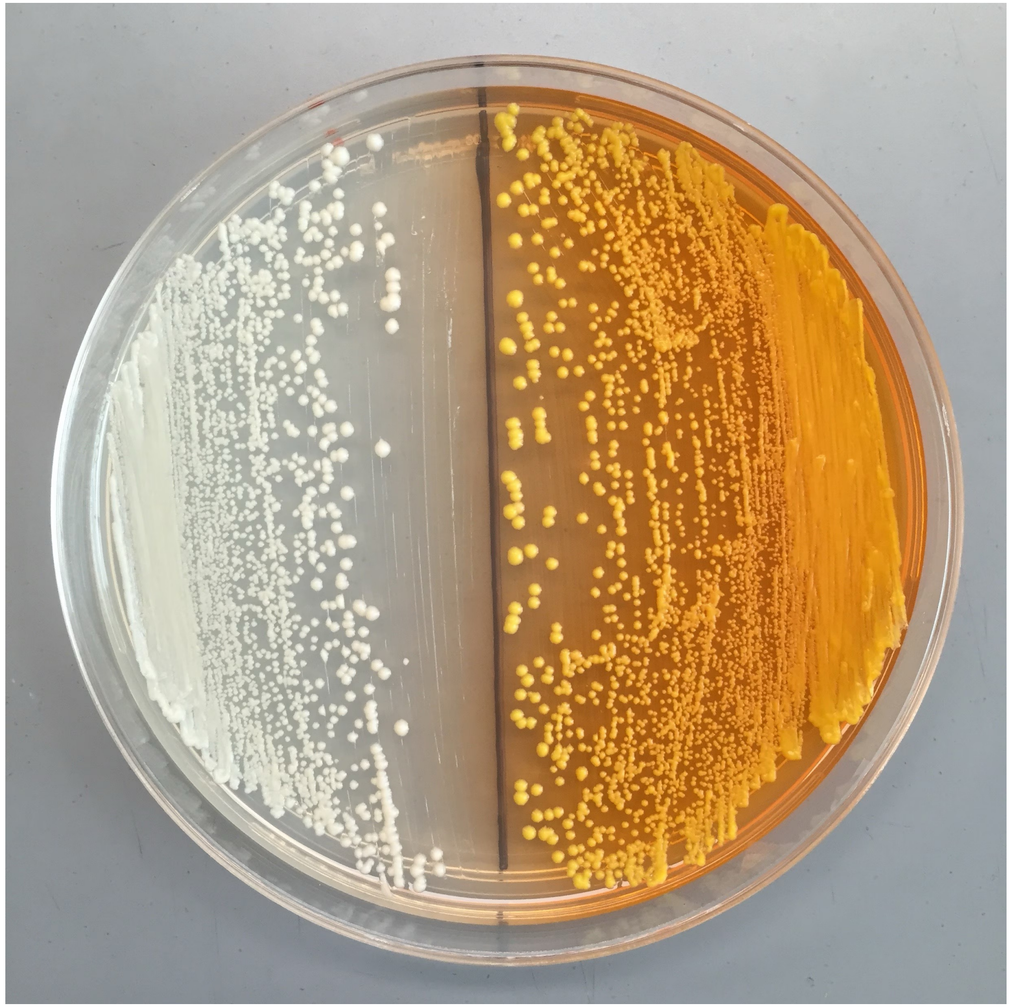
DNA synthesis is now commercially cheap and fast

Your time will be spent **designing molecules** not making them

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Reconsidering DNA technologies





DeLoache *et al. Nat Chem Biol* **11**, 465–471 (2015).
https://doi.org/10.1038/nchembio.1816

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Reconsidering DNA technologies

NGS sequencing is now commercially cheap and fast

DNA synthesis is now commercially cheap and fast

Your time will be spent **designing molecules** not making them

These technologies enable fields such as **synthetic biology** offering the ability to **control matter**

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

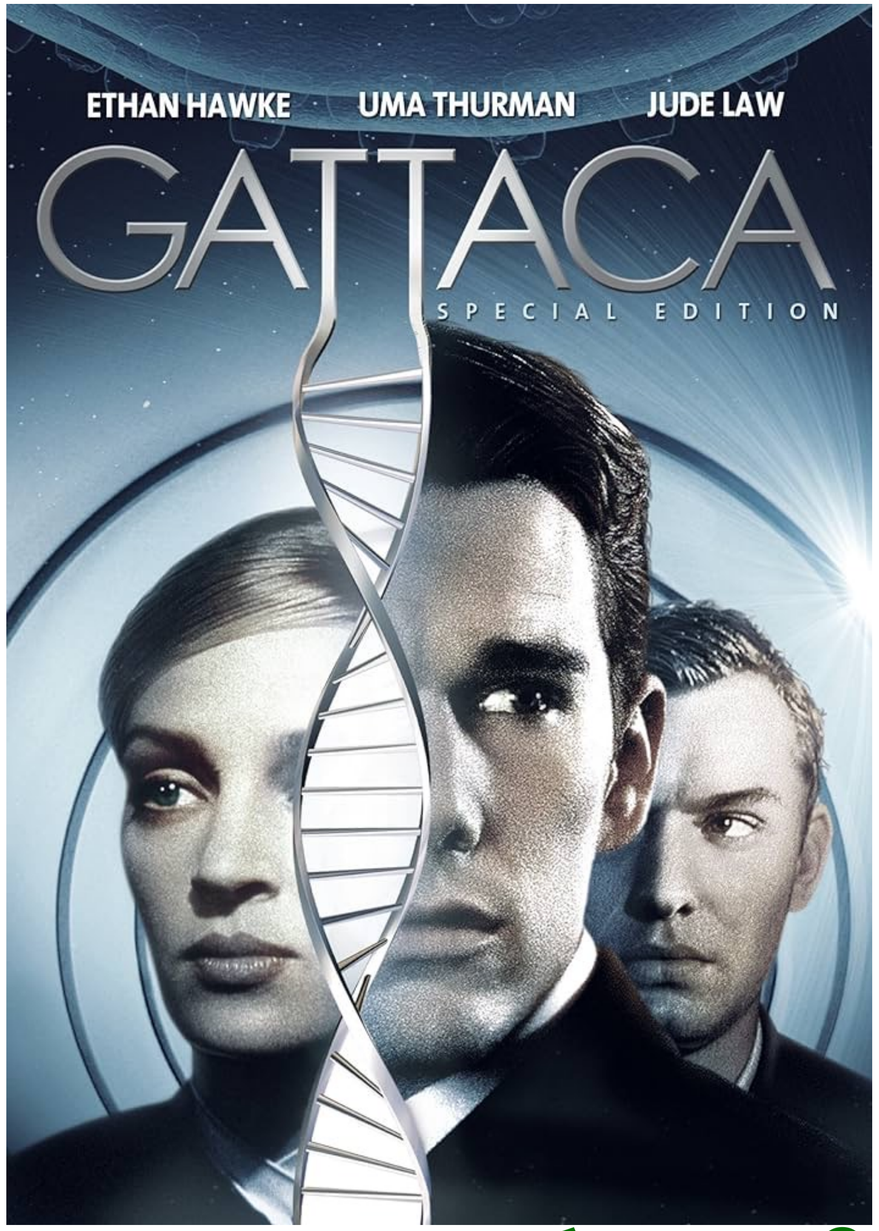
Would you have your genome sequenced for free?

A. Yes

B. No

Personal Genomics

- A new age of eugenics?
- Gattaca (1997) Written and Directed by Andrew Niccol



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

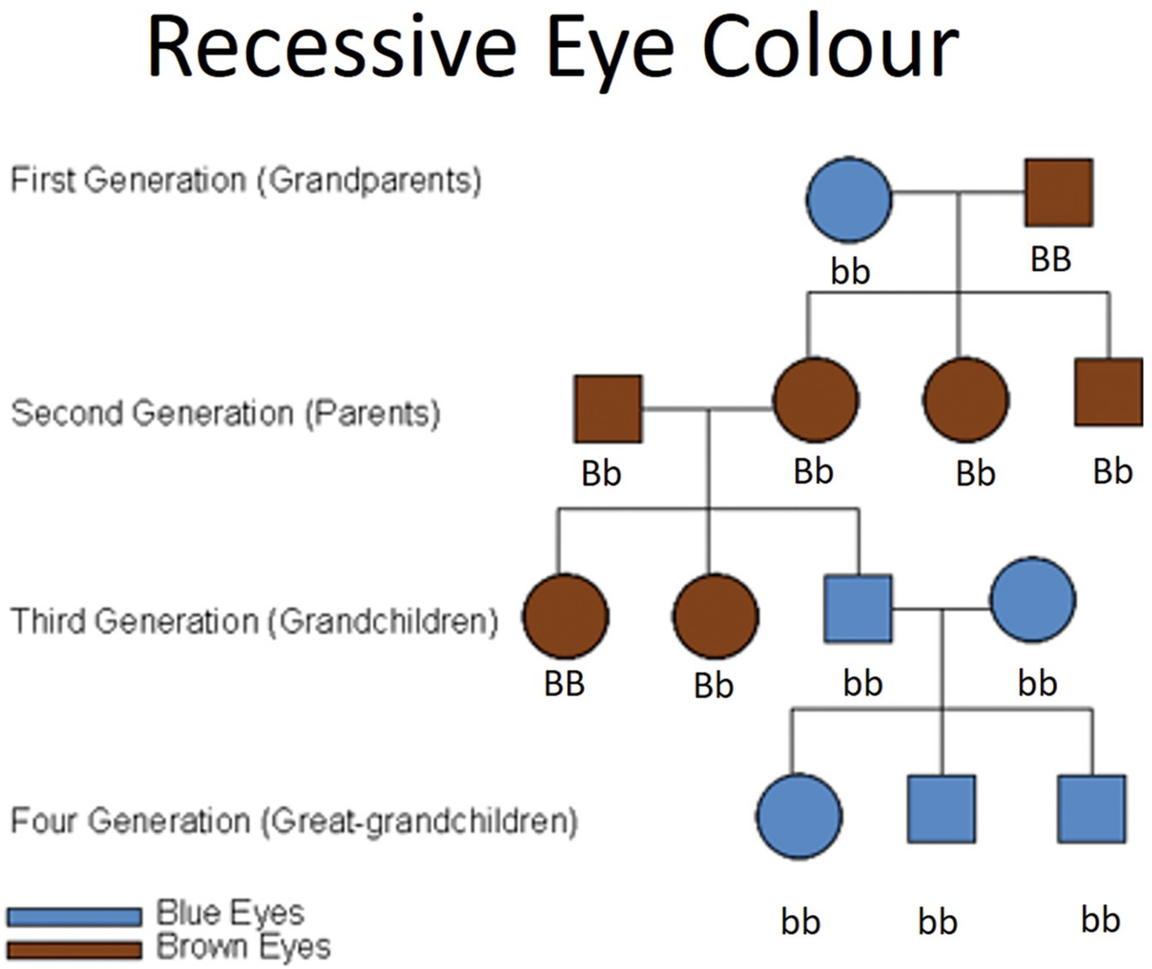
What does my genome tell about me?

Some genetic risk factors with **strong effects** can be identified (e.g. BRCA alleles and breast cancer).

But most ‘gene’ associations are **weak and effects are hard to predict**.

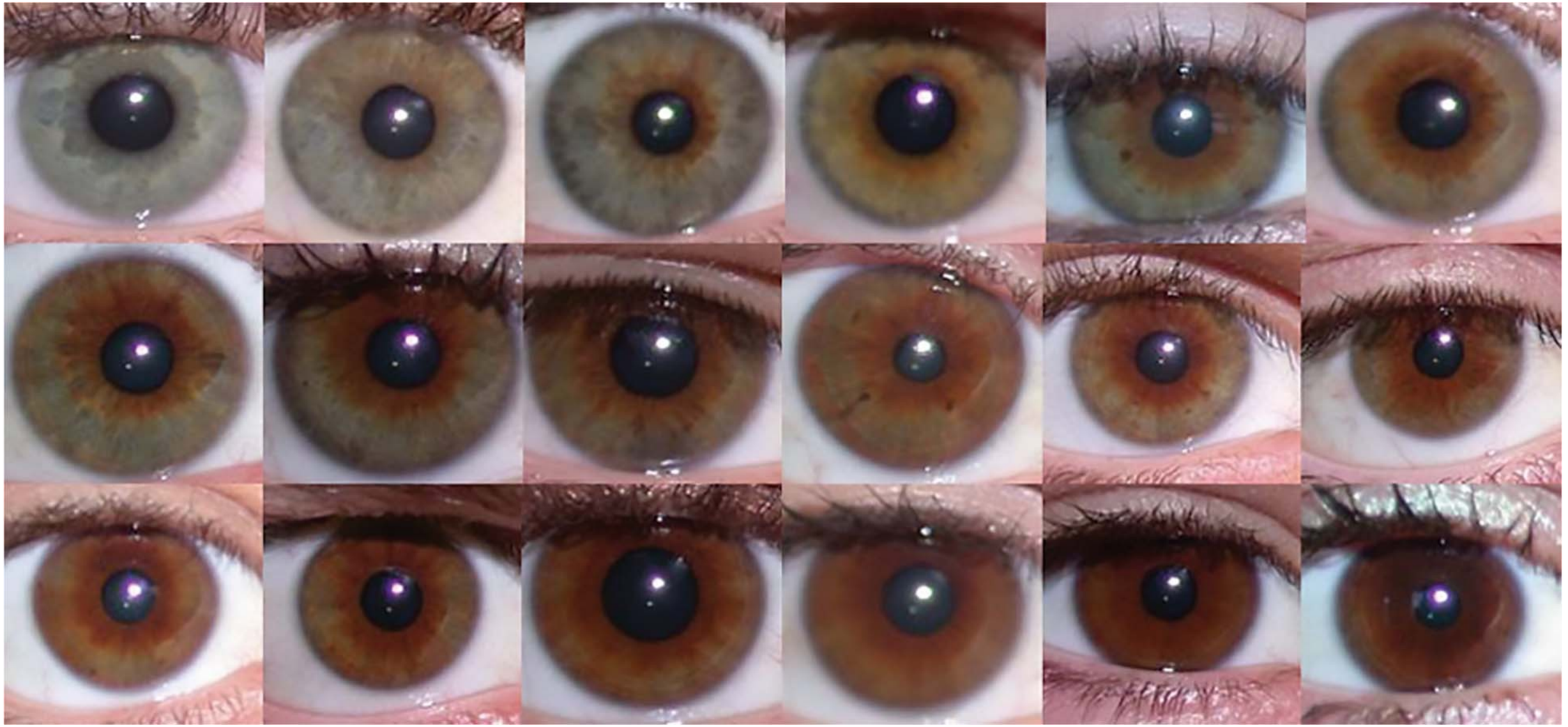
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Human Eye Colour



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Human Eye Colour



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

What does my genome tell about me?

Some genetic risk factors with **strong effects** can be identified (e.g. BRCA alleles and breast cancer).

But most ‘gene’ associations are **weak and effects are hard to predict**.

Gene to phenotype associations are **complex and penetrance** can vary.

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Personal Genome risks

It is possible to **establish your identity** from your genome alone by comparison to related individuals with known identity.

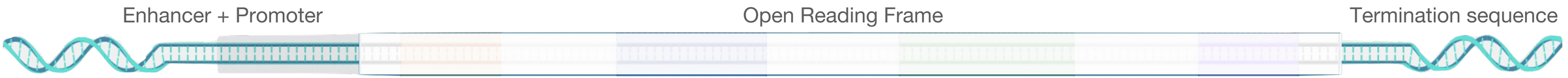
‘Free’ or ‘low-cost’ genome sequencing firms **may sell your genome** or leak it – e.g. 23andMe

Identity plus genome information opens the possibility for discrimination e.g. for health insurance

Do you want to know if you are at risk for incurable future diseases?

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Controlling gene expression



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Transcription Factors

Transcription factors are proteins involved in **transcribing DNA into RNA**.

Many have **DNA-binding domains** that allow them to bind to specific sequences of DNA in gene promoters or enhancers. Transcription factors can also interact with each other or additional protein co-factors to change their DNA binding properties.

~1500 estimated transcription factors in the human genome

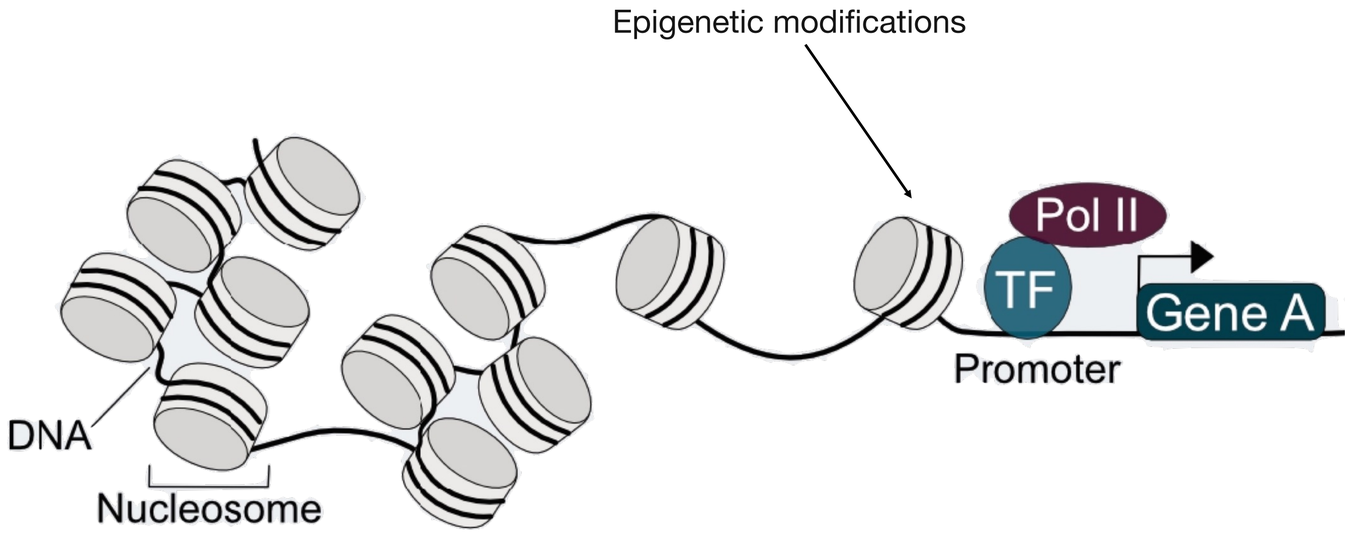
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Transcription Factors

Salzer & Kumar Plos One 2010

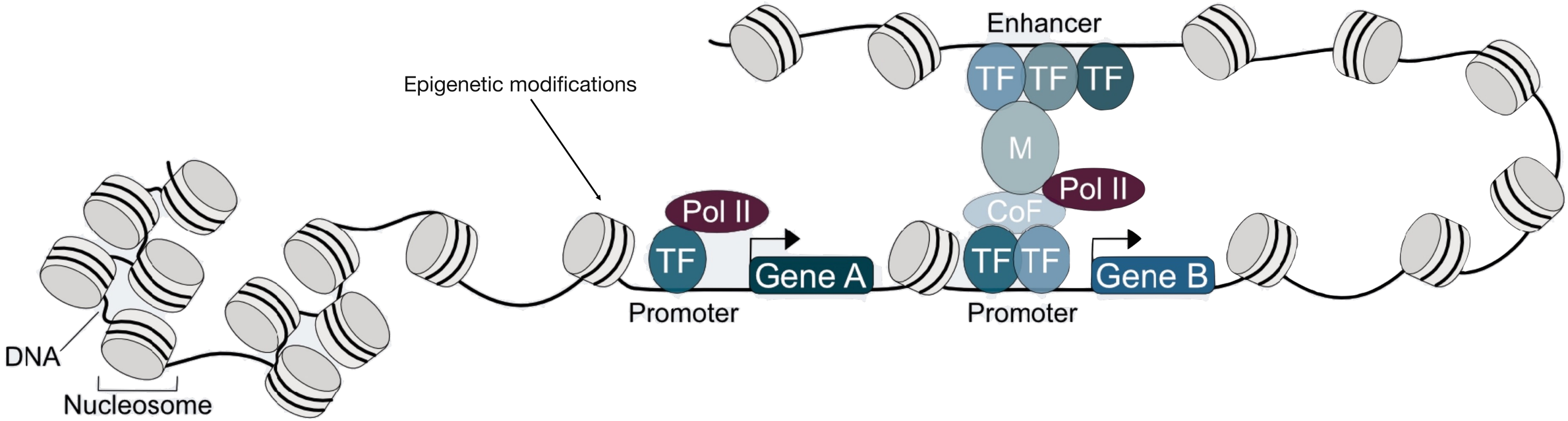
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Controlling gene expression



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Controlling gene expression



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Transcription Factors

Transcription factors are proteins involved in **transcribing DNA into RNA**.

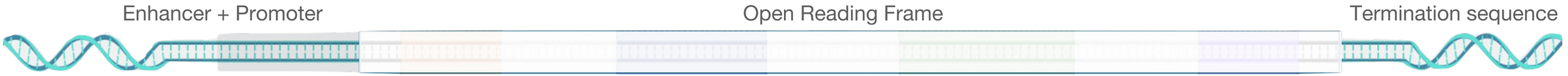
Many have **DNA-binding domains** that allow them to bind to specific sequences of DNA in gene promoters or enhancers. Transcription factors can also interact with each other or additional protein co-factors to change their DNA binding properties.

~1500 estimated transcription factors in the human genome

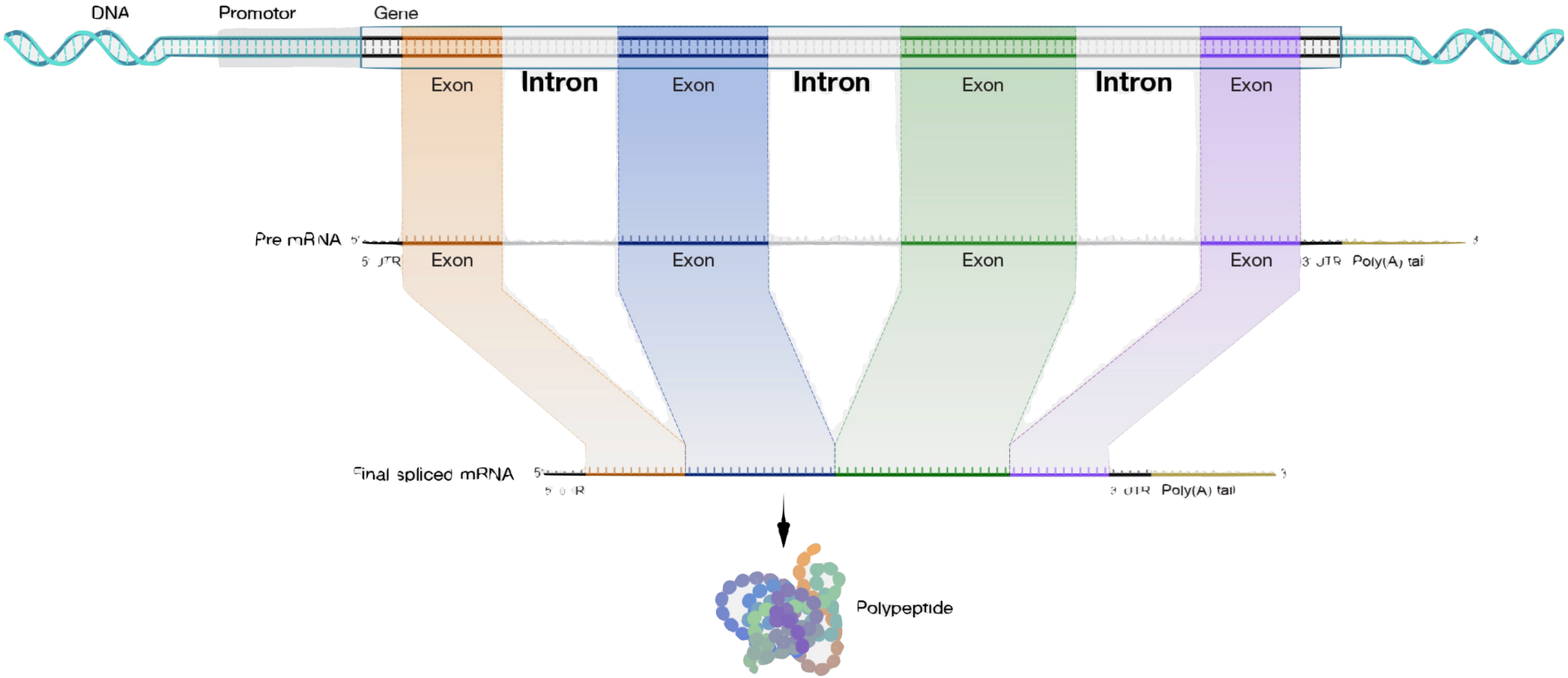
Predicting functional enhancer sequences is difficult.

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Controlling gene expression

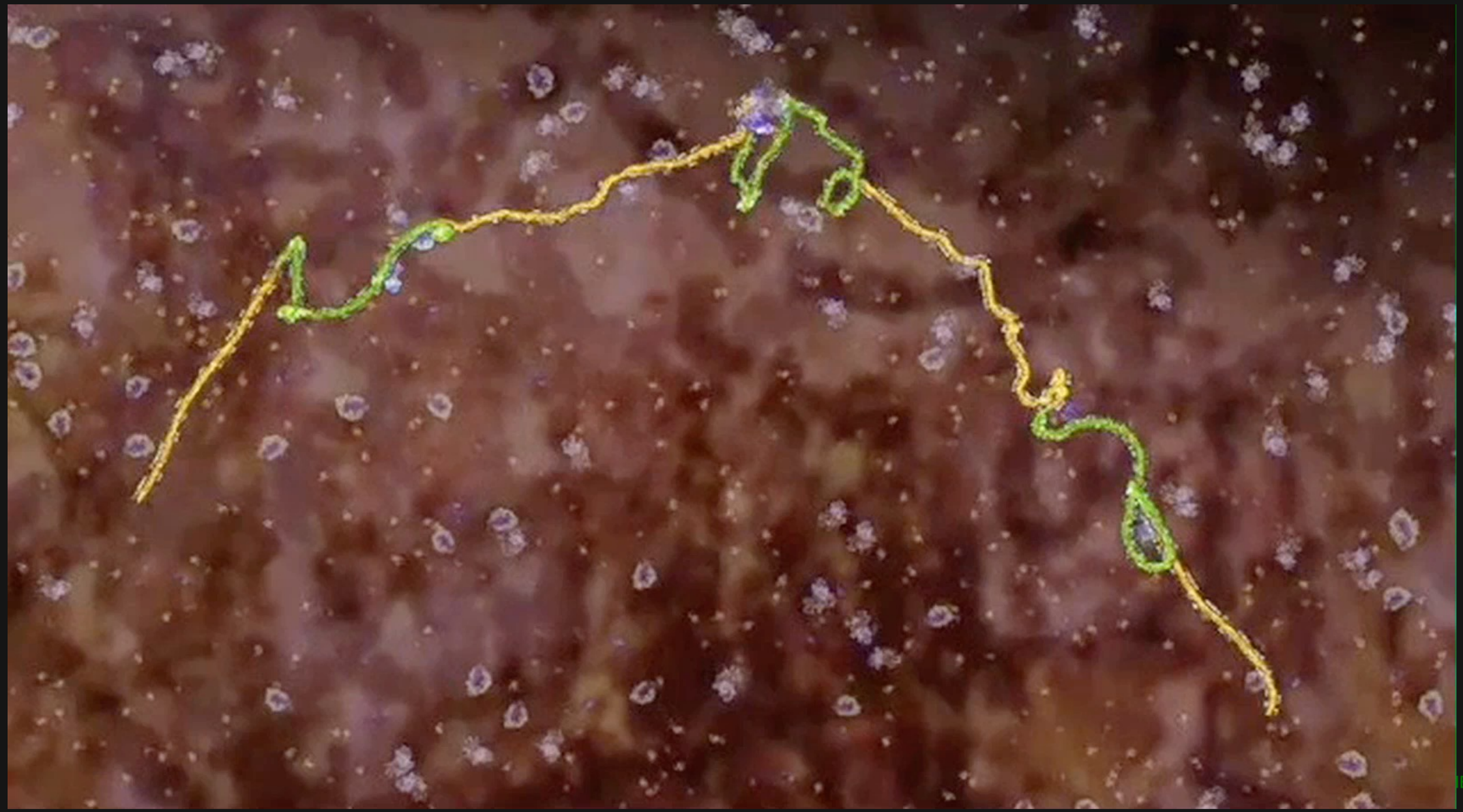


Protein Coding Genes



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

mRNA Splicing



Pre-mRNA Splicing

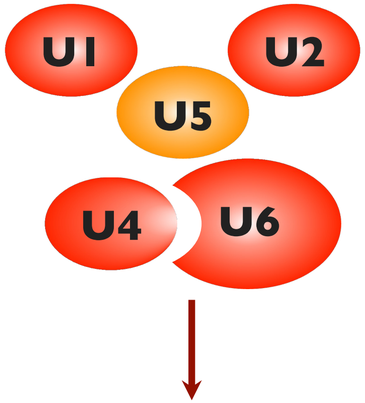
RNA Splicing: catalysed by a large ribonucleoprotein (RNA & Protein) complex called the **spliceosome**

The spliceosome is a highly dynamic assembled by sequential binding and release of the small nuclear RNAs (snRNAs) and protein factors that can influence splicing.

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

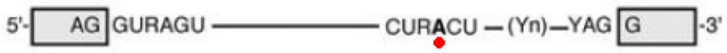
Pre-mRNA Splicing

U2 (Major)

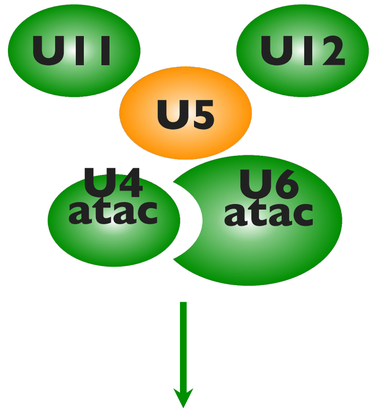


>99% of introns

U2 type



U12 (Minor)

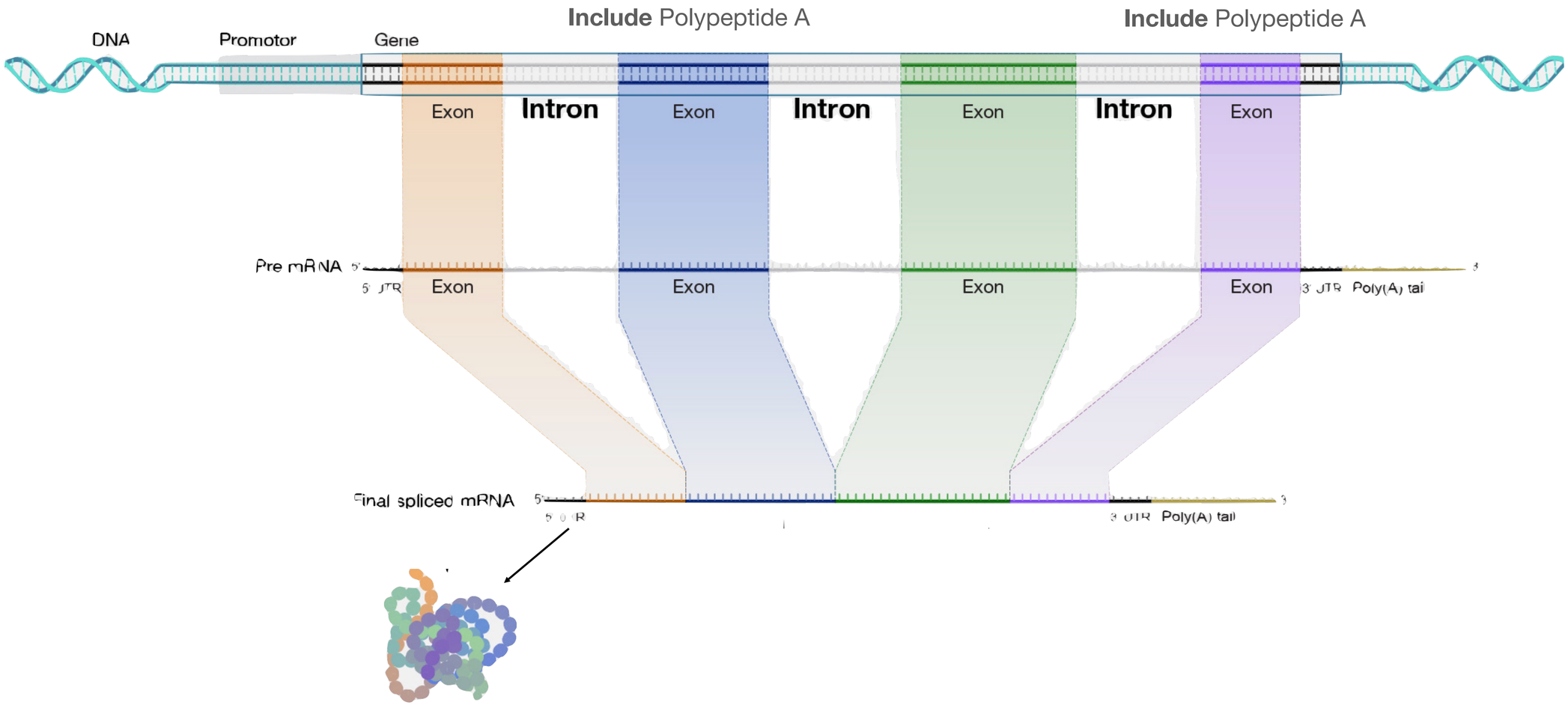


<1% of introns

U12 type

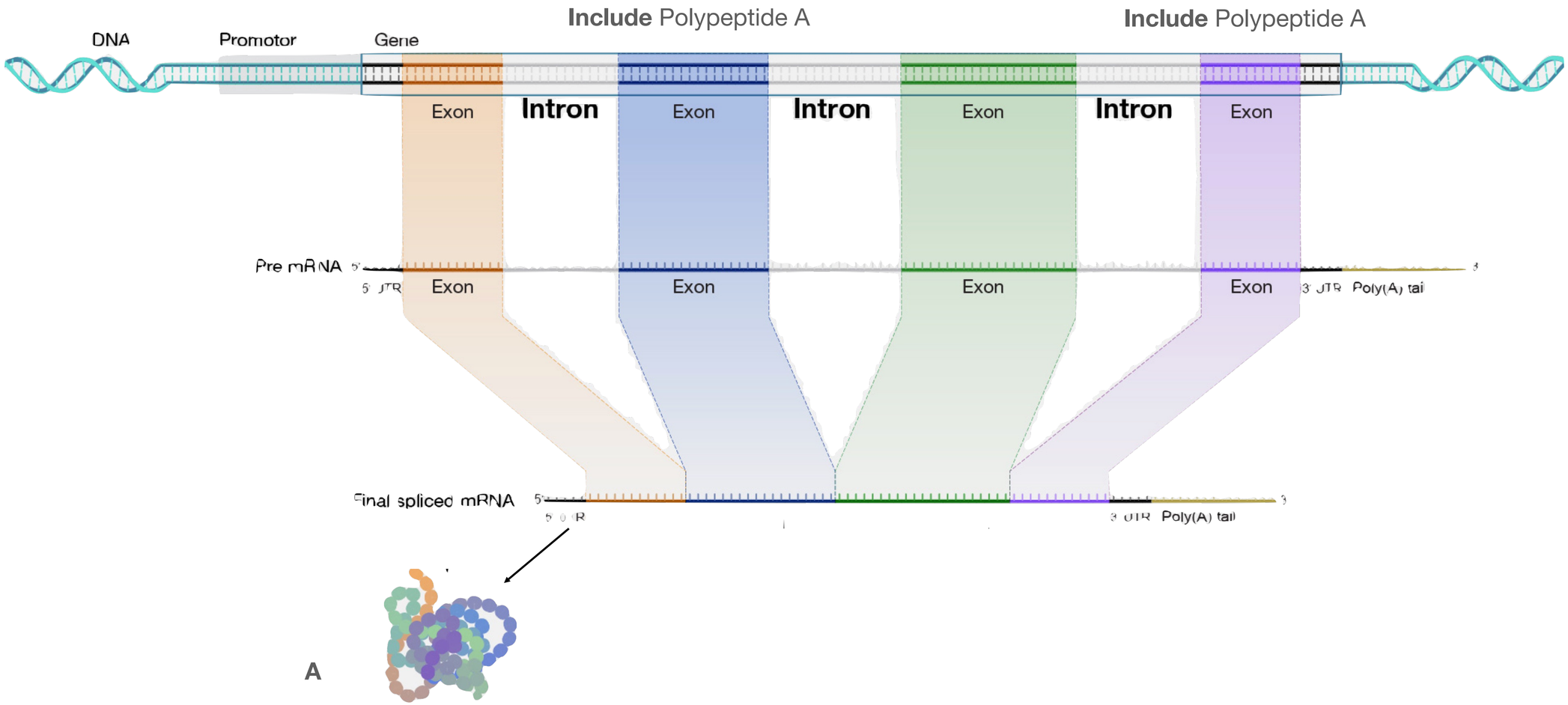
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

mRNA Splicing



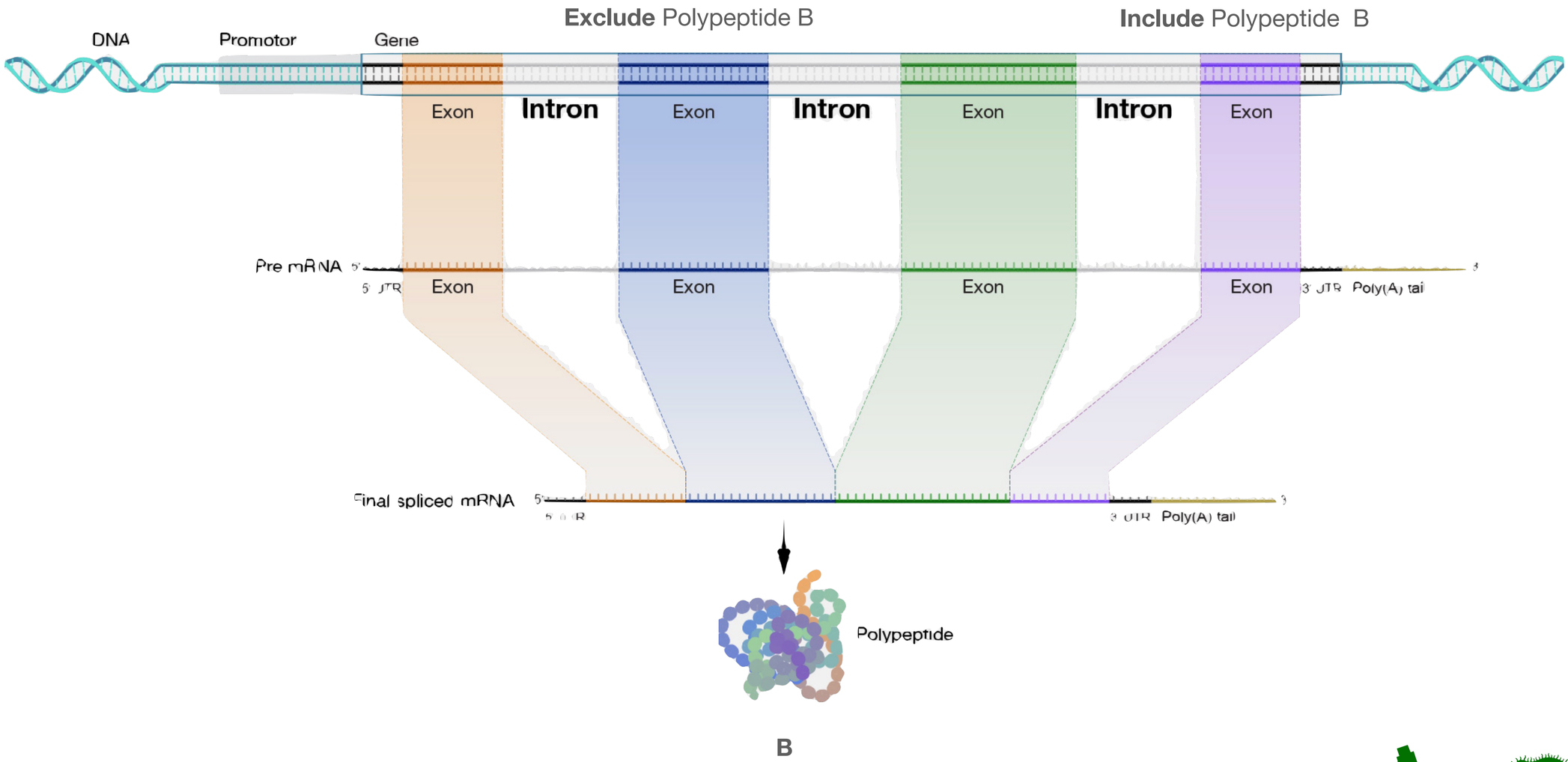
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

mRNA Splicing



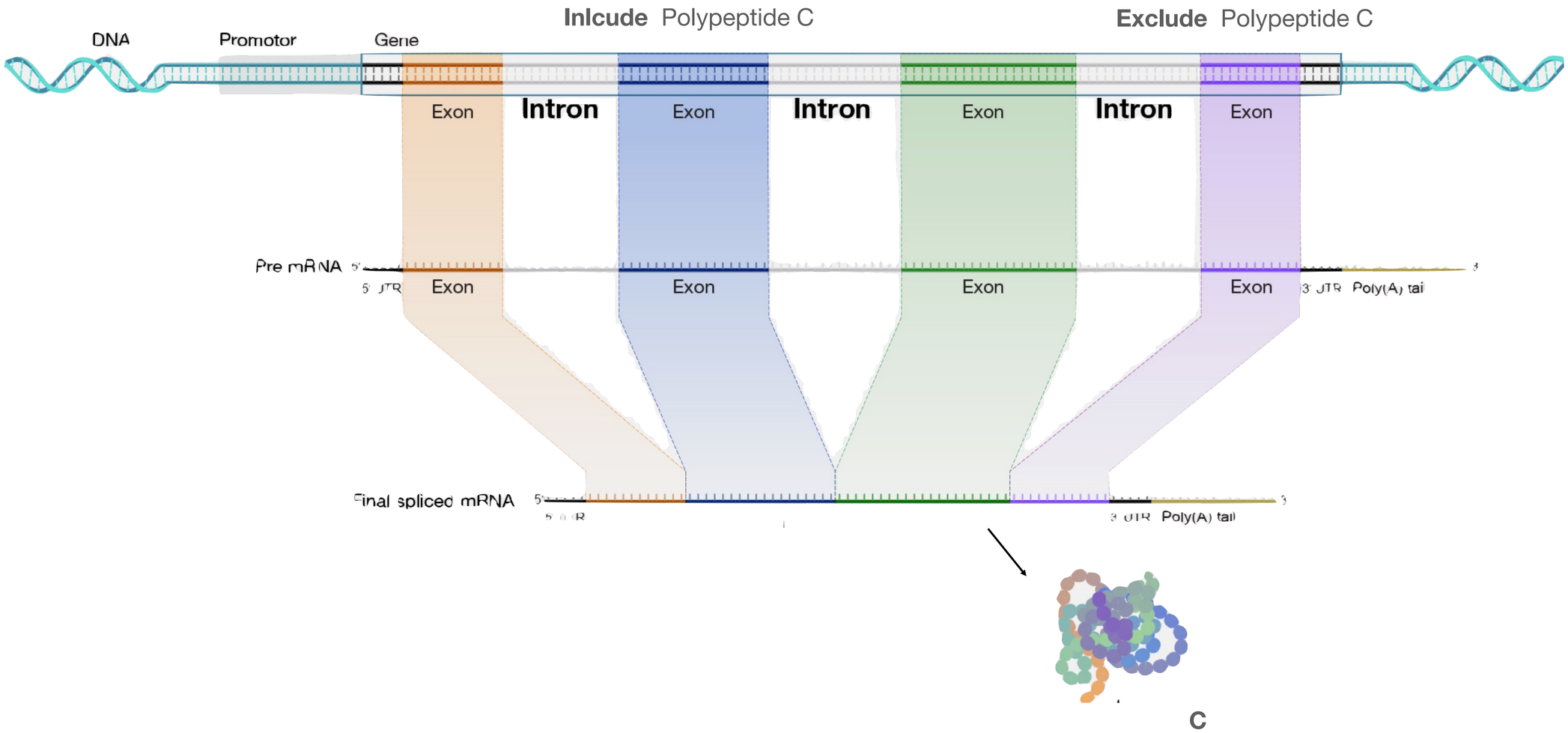
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

mRNA Splicing



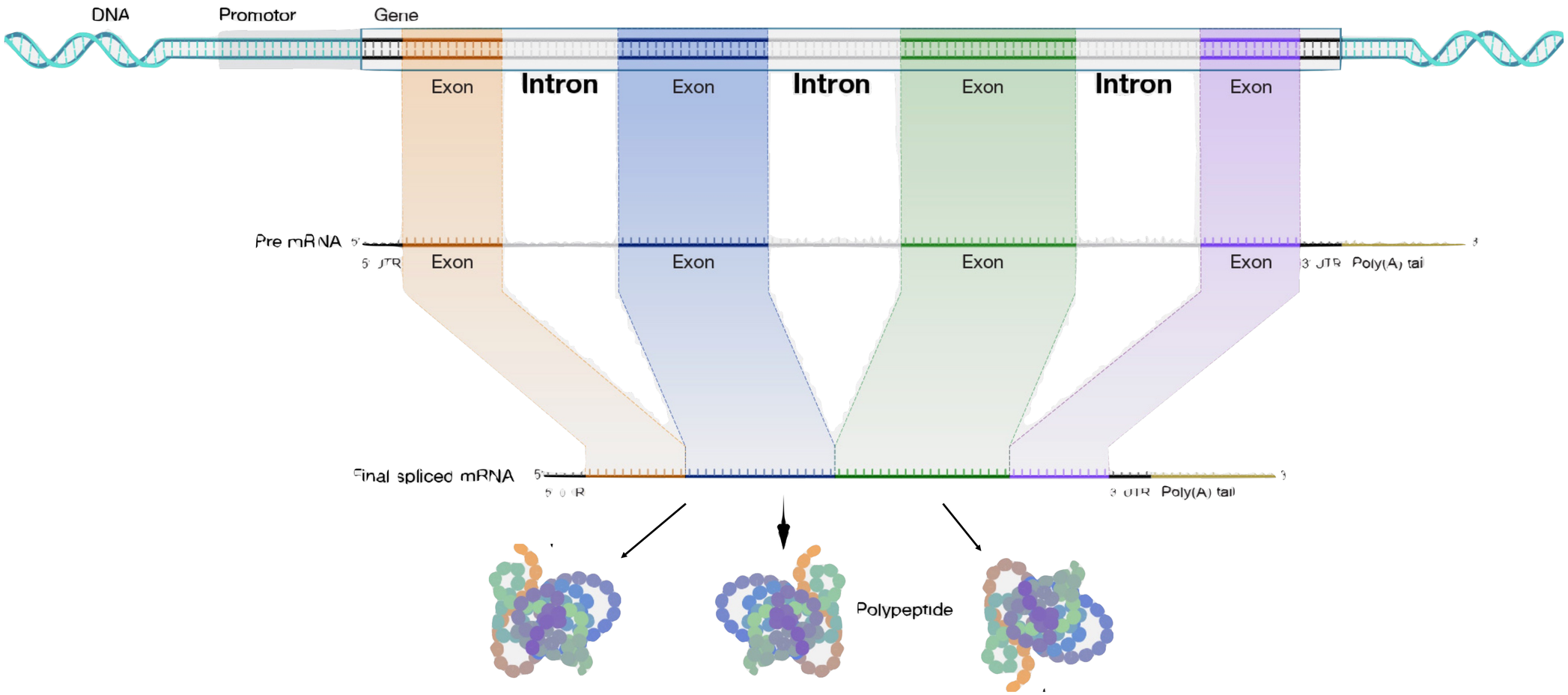
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

mRNA Splicing



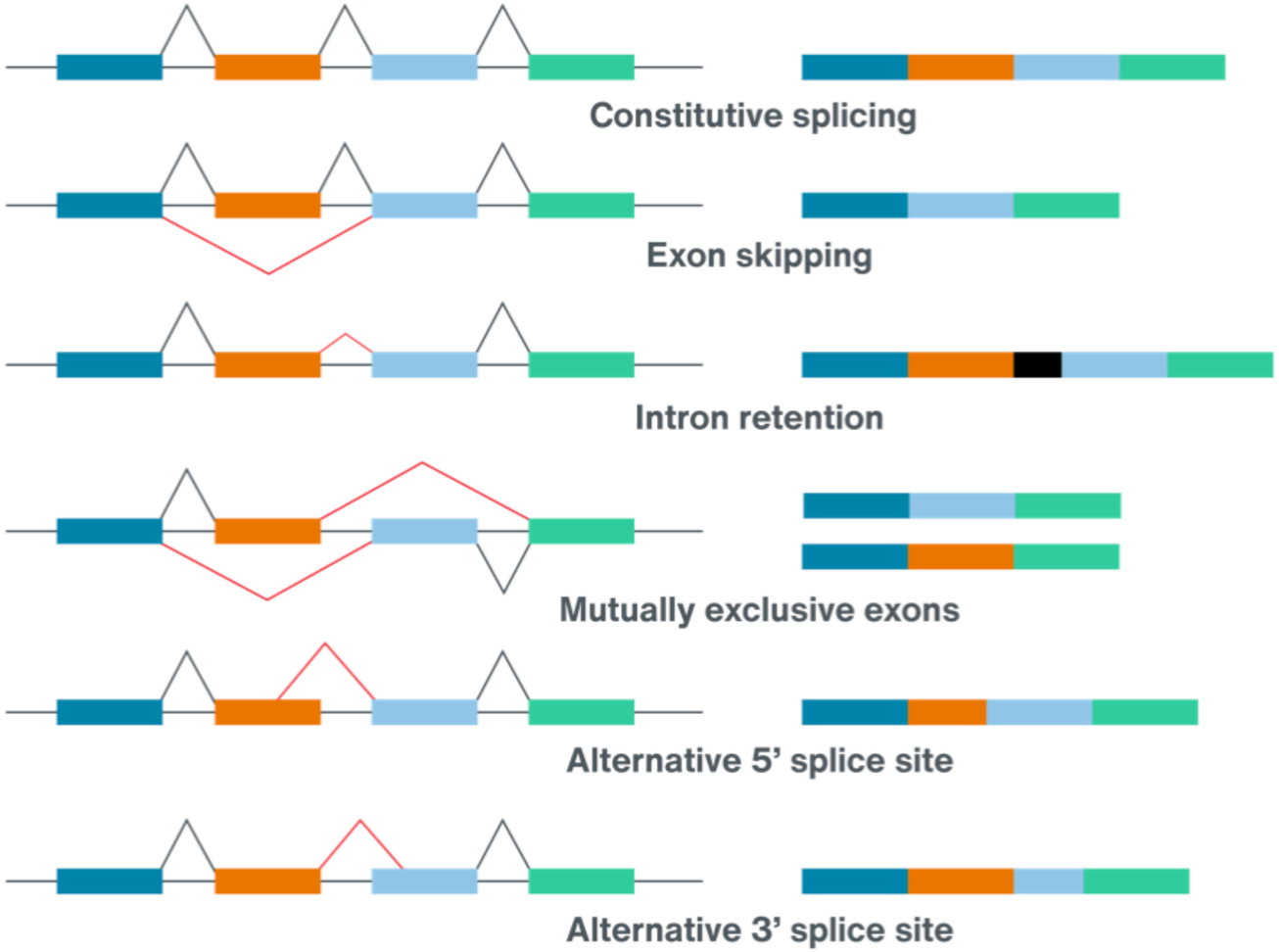
GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Protein Coding



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Pre-mRNA Splicing



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Pre-mRNA Splicing

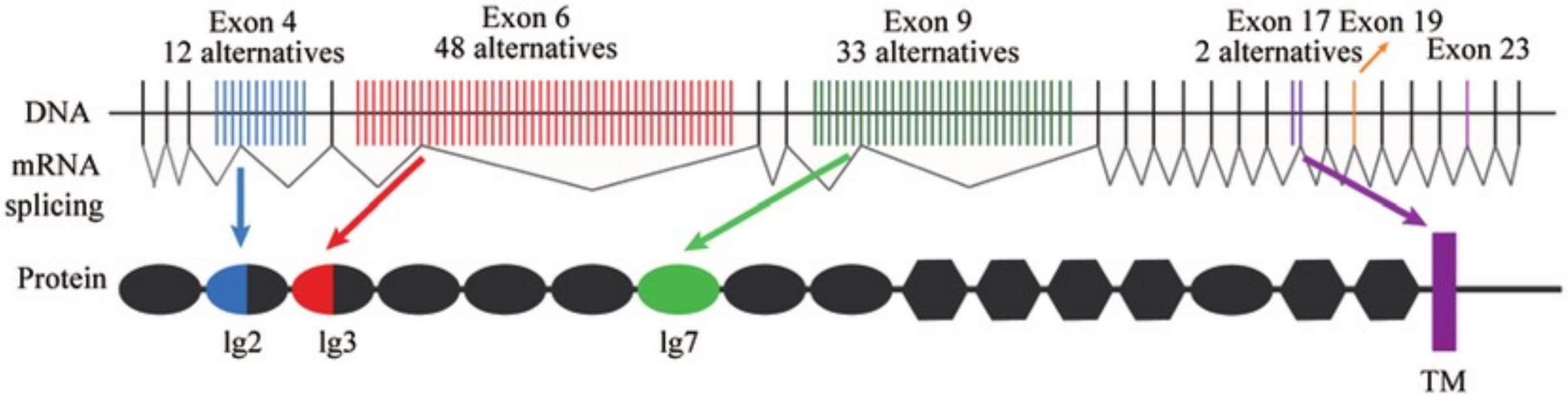
RNA Splicing: catalysed by a large ribonucleoprotein (RNA & Protein) complex called the **spliceosome**

The spliceosome is a highly dynamic assembled by sequential binding and release of the small nuclear RNAs (**snRNAs**) and protein factors that can influence splicing.

Alternative splicing factors can produce a **large diversity** of protein products from one gene

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

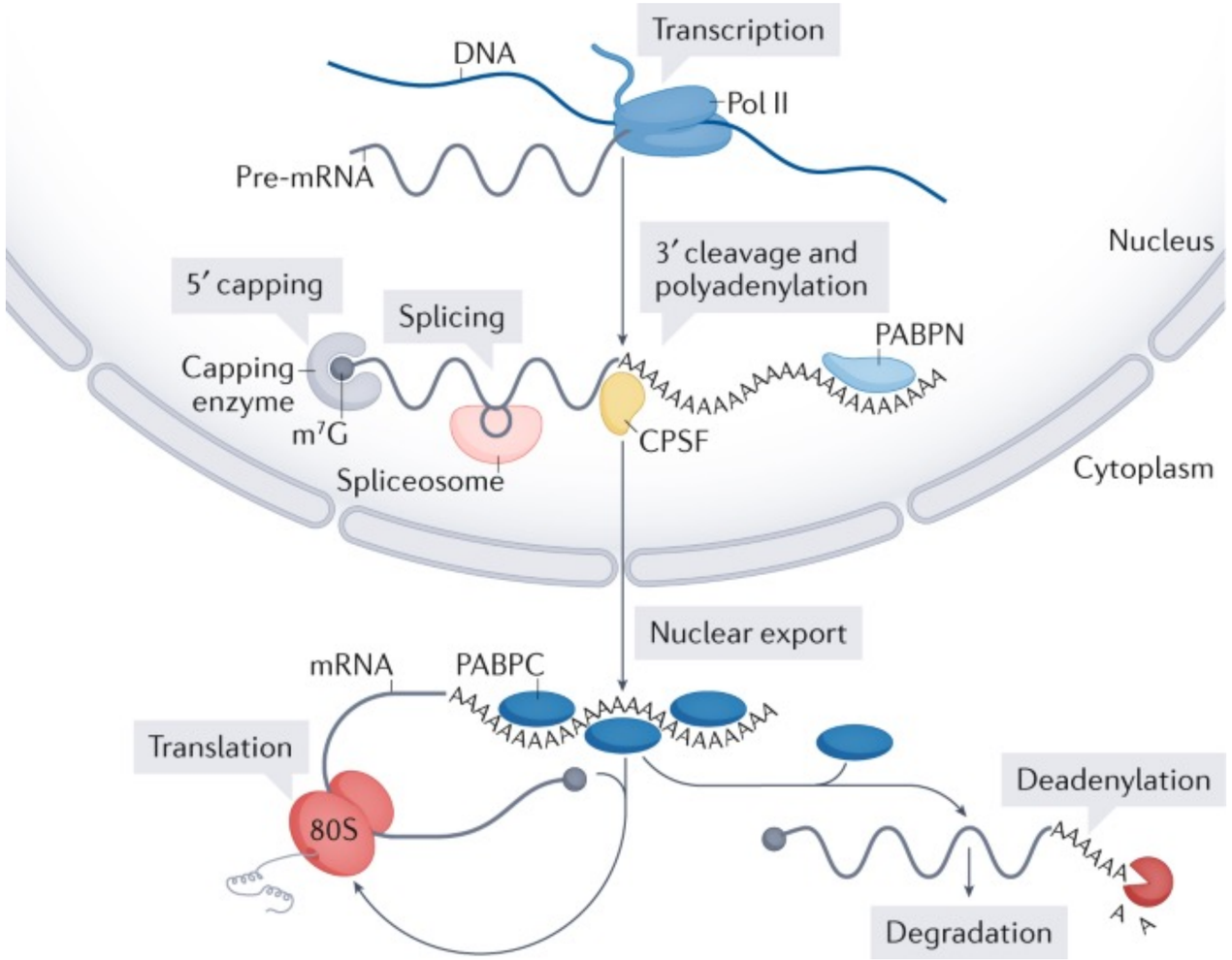
Pre-mRNA Splicing



Drosophila Dscam: **38,016** different peptides from one gene

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Post Transcription regulation of RNA



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

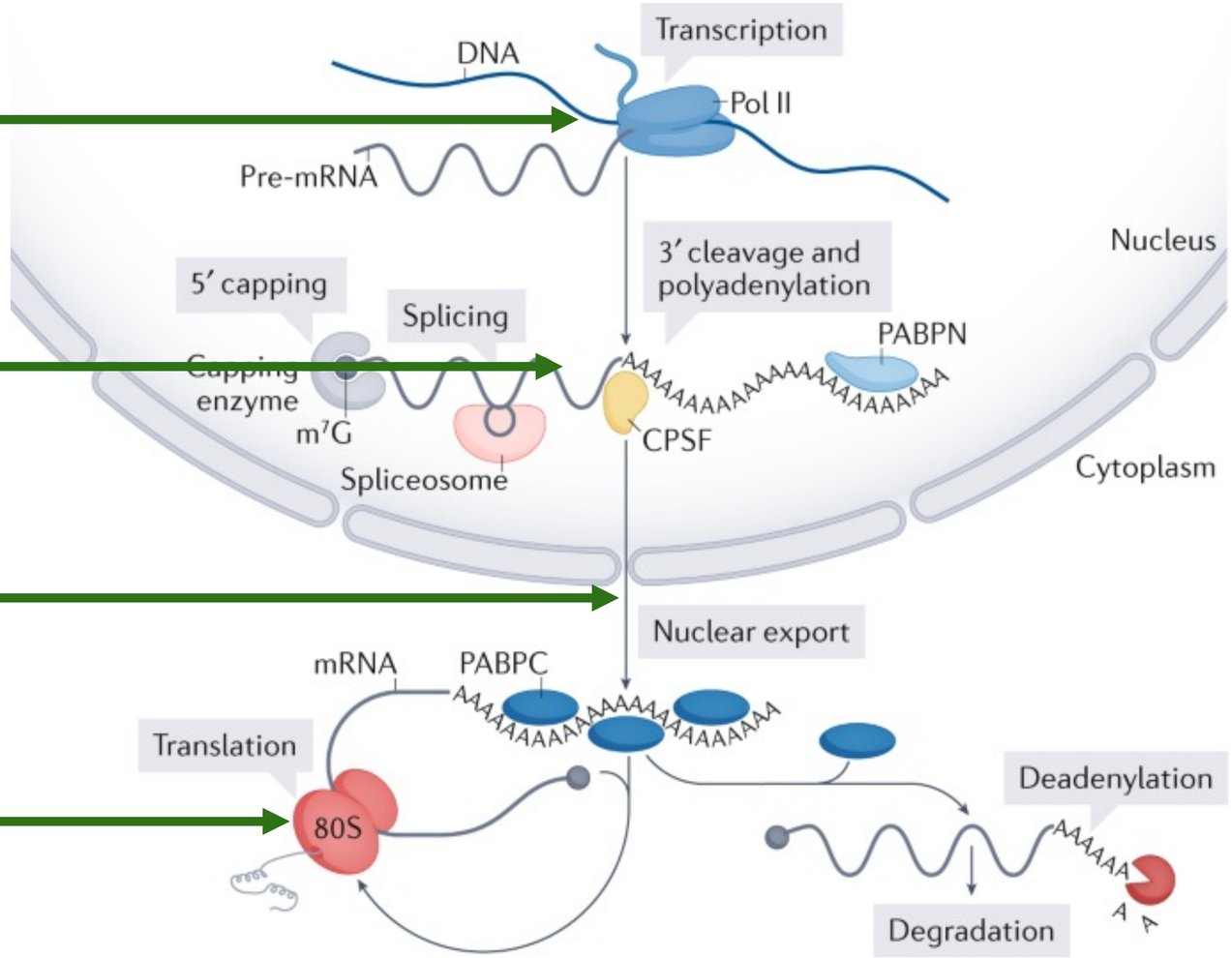
Post Transcription regulation of mRNA

Transcription control

RNA processing control

RNA transport control

Translation control

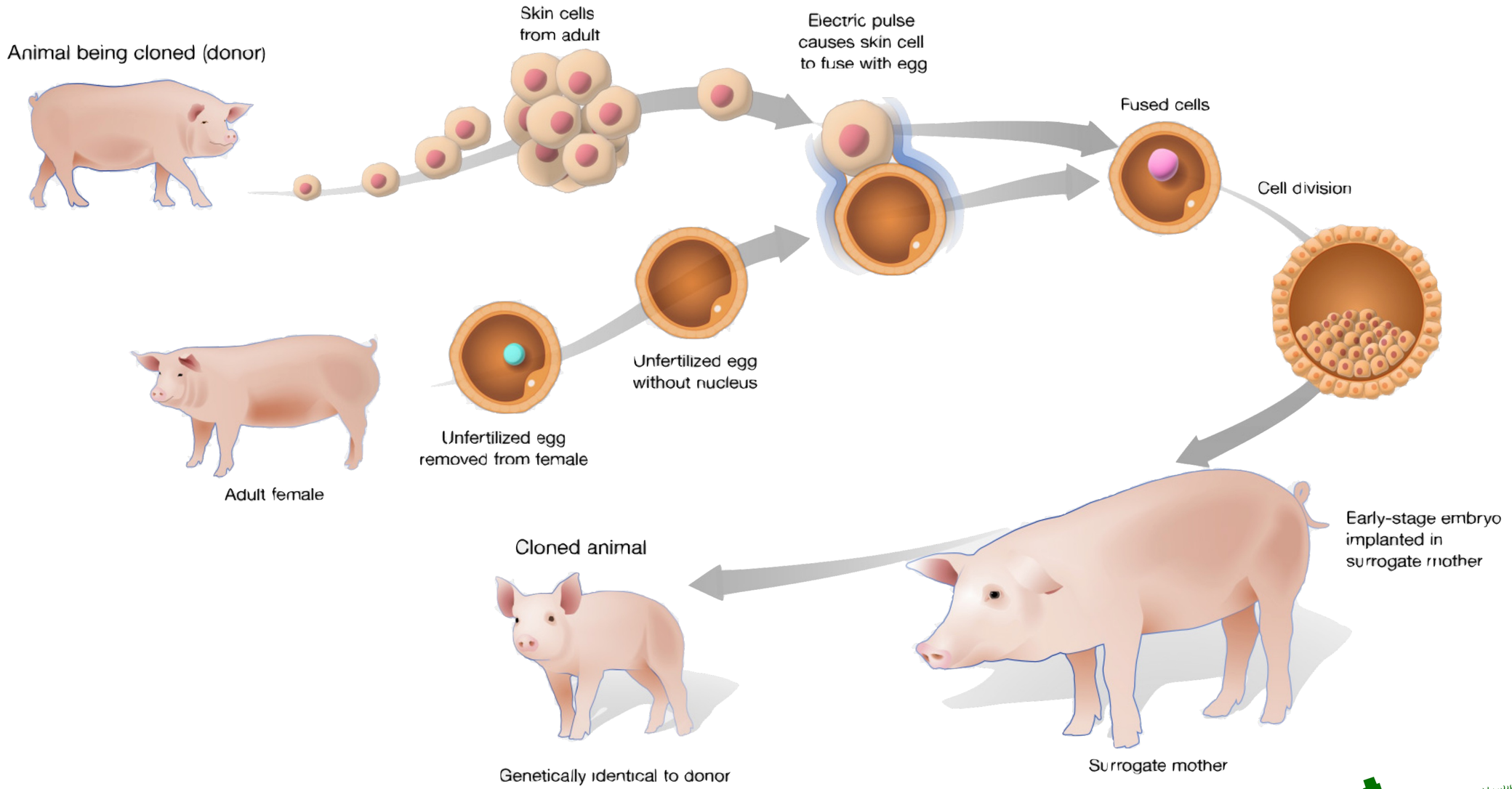


GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Cloning

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Cloning



GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING

Cloning

Inefficient: **Multiple surrogates** and **many cloned embryos** to achieve a single successfully cloned animal.

Can be useful for **valuable agricultural animals** and **rare species**. Used extensively for pets (~40K CHF to clone your dog)

Seems **safe** – clones and their progeny have normal lifespan and no seemingly no adverse health issues

Proposals to **clone extinct animals** (e.g. Wolly Mammoth) rarely talk about the number of surrogates required, in addition to other ethical issues.

Outcome not guaranteed

GENOME TO FUNCTION
LIFE SCIENCES ENGINEERING