

## Gene regulatory network inference on transcript isoform level

### *Central dogma of molecular biology*

The central dogma of molecular biology is a framework that describes the flow of genetic information within a biological system from DNA to protein. The central dogma consists of three main processes:

1. **Replication:** This is the process by which DNA molecules make copies of themselves. During cell division, each new cell needs a complete set of genetic instructions, so DNA must replicate to ensure that the genetic information is passed on to the next generation.
2. **Transcription:** In this step, a specific segment of DNA is used as a template to synthesize a complementary RNA molecule. This process occurs in the cell nucleus and results in the formation of messenger RNA (mRNA).
3. **Translation:** The mRNA carries the genetic code from the DNA to the ribosomes, the cellular machinery responsible for protein synthesis. During translation, the information in the mRNA is used to assemble a chain of amino acids, forming a polypeptide chain. This chain then folds into a functional protein.

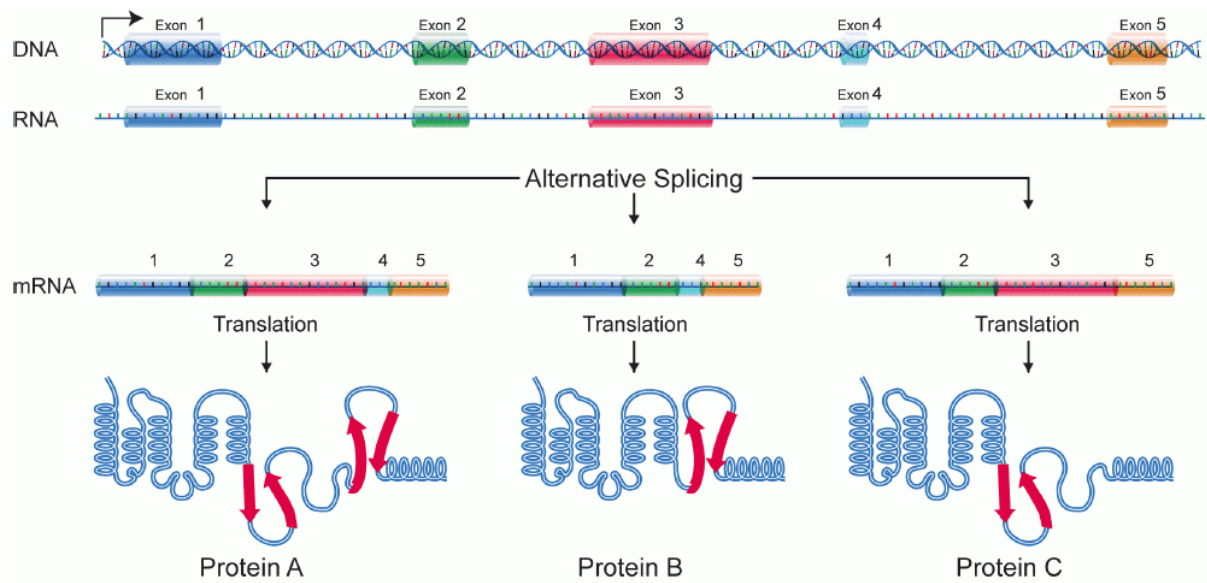
In summary, the central dogma describes the unidirectional flow of genetic information from DNA to RNA to protein. It is important to note that while information flows from DNA to RNA in transcription and from RNA to protein in translation, the reverse processes (reverse transcription and the synthesis of RNA from RNA) are also observed in certain biological contexts, such as in the replication of retroviruses. However, these processes are exceptions to the general flow described by the central dogma.

### *Alternative splicing*

After transcription, the pre-mRNA undergoes the process of splicing where the non-coding elements are removed from the mature mRNA which is translated by the ribosome. A pre-mRNA can give rise to several mature versions of mRNA, called isoforms. There are multiple ways in which exons can be combined to generate the final transcript, see Figure 1 for an example.

### *Gene regulatory networks*

Gene regulatory networks attempt to model the regulatory relationship between genes. The genes are depicted as nodes and the regulatory as edges. GRNs can be directed or undirected, the edges can be weighted or unweighted. An example for such a regulatory relationship would be transcription factors (TFs) whose expression can lead to an increase in the expression of a gene, called activation; or the reduction of the expression of certain genes, called repression. There are a plethora of tools and algorithms attempting to reconstruct GRNs from gene expression data (Genie3, GRNBoost, WCGNA, ...), but not many take into account the presence or absence of splice isoforms.



*Figure 1: Schematics of alternative splicing*

### *Project outline*

The goal of this project will be to develop a method to construct gene regulatory networks from transcriptomics data.