**Genomic Interpreter: A Hierarchical Genomic Deep Neural Network with 1D Shifted Window Transformer**

**Group members:**

21BD1A05AY**-** M.Shivani

21BD1A662L- G.Nagamythreya

21BD1A662Q- K.Sowmya

21BD1A662Y- K.Srivarsha

21BD1A6631- K.Karthik

21BD1A6633- M.Rishitha

**Genomic interpreter:**  
 A genomic interpreter is a professional who specializes in analyzing and interpreting genomic information, particularly genetic data obtained through techniques such as DNA sequencing.

· We introduce Genomic Interpreter, an attention-based model for genomic assay (techniques used to analyze genetic materials) prediction

· Transformer: 1D Swin (shift window)

**Swin Transformer:** is a type of deep learning architecture built upon the foundation of the Transformer architecture, originally introduced for natural language processing tasks. Swin Transformers were designed to address the limitations of traditional Transformers when applied to vision tasks, such as image classification and object detection. Swin Transformers have gained popularity for their ability to handle images of varying sizes and their efficiency in capturing long-range dependencies in images.

Swin Transformers introduces a concept called "shifted windows" to reduce computational complexity. Instead of processing the entire image simultaneously, they split it into non-overlapping patches or windows. The window-based processing helps reduce traditional Transformers' quadratic complexity associated with self-attention mechanisms.

**Genes** are segments of DNA (deoxyribonucleic acid) that contain the instructions for building and maintaining an organism.

**DNA Structure:** Genes are made up of DNA, which is a long molecule composed of four nucleotide bases: adenine (A), thymine (T), cytosine (C), and guanine (G). The sequence of these bases within a gene determines the specific genetic code or information it carries. DNA stores genetic info permanently.

**RNA:** RNA has one oxygen mole more than rna. RNA has multiple functions, including:

* Messenger RNA (mRNA) carries the genetic code from DNA to the ribosomes for protein synthesis (translation).
* Transfer RNA (tRNA) brings amino acids to the ribosome during translation.
* Ribosomal RNA (rRNA) is a structural component of ribosomes, where protein synthesis occurs.
* Other non-coding RNAs (ncRNAs) play roles in gene regulation, splicing, and various cellular processes.

**mRNA:** It serves as a temporary carrier of genetic information, allowing the cell to translate the genetic code into functional proteins that perform a wide range of cellular functions.

**Genomics:** is a branch of molecular biology and genetics that focuses on the study of an organism's complete set of genes, collectively known as its genome. It considers the entire genome.

Genetics: considers one single gene.

Gene expression is the process by which the information encoded in a gene is turned into a function.

**Uses:**

Genomic interpreters often work in genetic counseling, where they assist individuals and families in understanding their genetic test results. They provide information about the implications of specific genetic variations and the risk of inherited conditions.

They provide guidance to healthcare providers and clinicians in making treatment decisions based on genetic information. This can involve recommending specific therapies, medications, or monitoring strategies tailored to an individual's genetic profile.

Genomic interpreters may also work in research settings, helping researchers analyze and interpret genomic data to identify genetic factors associated with diseases, conduct population studies

**Non-coding RNAs:** they do not contribute to protein synthesis but have various other functional roles.

**Transcription:** Copying some part of DNA to RNA. It is the first step in gene expression. This resulting RNA molecule is called mRNA.

**Translation:** The next step in gene expression is translation, where the genetic information in the mRNA is decoded to produce a specific sequence of amino acids, which forms a protein.

**Swin transformer:**

· Divides input image into non-overlapping blocks each of which is passed as the input

· Then the positional encoding is done (positional encoding is a mechanism used in the Transformer architecture to inject information about the positions of elements in a sequence). (swin is used to reduce computational complexity. This mechanism limits the attention computation to nearby patches, rather than considering all patches in the sequence.)

· It is a self-attention-based transformer. In genomics, we use a 1D-shifted window-based deep learning transformer, which is a hierarchical structure.

· Swin Transformers include feedforward neural networks within each block or layer.

· Swin transformers can work with large data with less computation. Hence, they are better than traditional vision transformers.

PROCESS:

<https://youtu.be/ORWdELQ1h9M?si=imldle0-F2p46zpE>

· Patch partition: The image is divided into patches

· Linear Embedding: pixels are converted into numerical data

· In this transformer, attention is computed within each window

Encoded dataset: [basenji/data/human.hg38.genome at master · calico/basenji](https://github.com/calico/basenji/blob/master/data/human.hg38.genome)

Reference:

bioinformatics process: <https://youtu.be/4CkRXGWmAbU?si=h-2tyZMUeDhzccMd>

Basics: <https://www.genome.gov/about-genomics/fact-sheets/Genetics-vs-Genomics>