BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,
Tamil Nadu, India.

Programme : M.Sc., Biomedical Science

Course Title : Bioinformatics

Course Code : BM35S1BI

Unit-IV

TOPIC: GENSCAN

Dr. P. JEGANATHAN
Guest Lecturer
Department of Biomedical Science

GENSCAN

INTRODUCTION

- ***GenScan*** is a bioinformatics software. it's mainsail function is to acquire a DNA sequence and find the open reading frames that accord to genes. It is a type of Gene finders along with "GLIMMER' and Various other Gene Prediction Programs.

WHAT IS MEANT BY GENE PREDICTION OR GENE FINDING ?

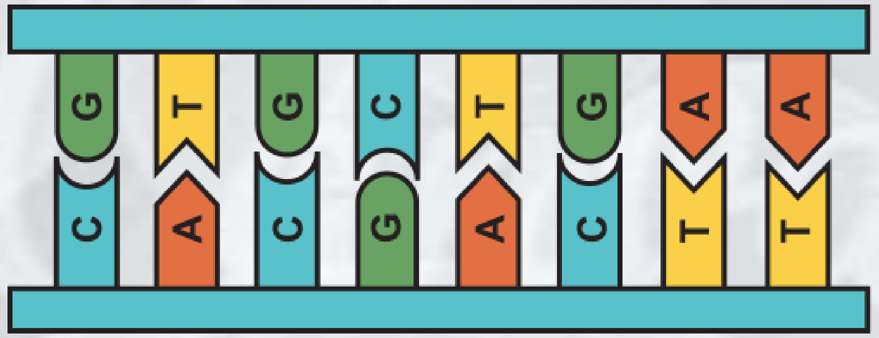
- In computational biology, gene prediction or finding refers to the process of identifying the regions of genomic DNA that encode genes. This includes Protein- coding genes as well as RNA genes, but may also include prediction of other functional elements such as regulatory regions.

IMPORTANCE OF GENE PREDICTION

- Helps to annotate large, contiguous sequences
- Aids in the identification of fundamental and essential elements of genome such as functional genes, intron, exon, splicing sites, regulatory sites, gene encoding known proteins, motifs, EST, ACR, etc.
- Distinguish between coding and non-coding regions of a genome
- Predict complete exon – intron structures of protein coding regions
- Describe individual genes in terms of their function
- It has vast application in structural genomics ,functional genomics , metabolomics, transcriptomics, proteomics, genome studies and other genetic related studies including genetics disorders detection, treatment and prevention.

WHY DO WE USE GENSCAN?

- The GenScan program is designed to predict complete gene structures, including exons, introns, promoter and polyadenylation signals, in genomic sequences.



GenScan

- In Bioinformatics ***GenScan*** is a program to identify complete gene structures in genomic DNA. It is a GHMM-based program that can be used to predict the location of genes and their exon-intron boundaries in genomic sequences from a variety of organisms. The ***GenScan*** Web server can be found at MIT.

CHRISTOPHER BURGE, PHD

GenScan was developed by Professor *Christopher Burge* in the research group of *Samuel Karlin* at **Stanford University**

- This program is not only used to detect genes in a sequenced set of DNA, it can also be used to determine a specific sequence using measures of the C+G content. As a matter of fact there are mainly four profiles for a particular species in a parameter file and each profile is equivalent to C+G%

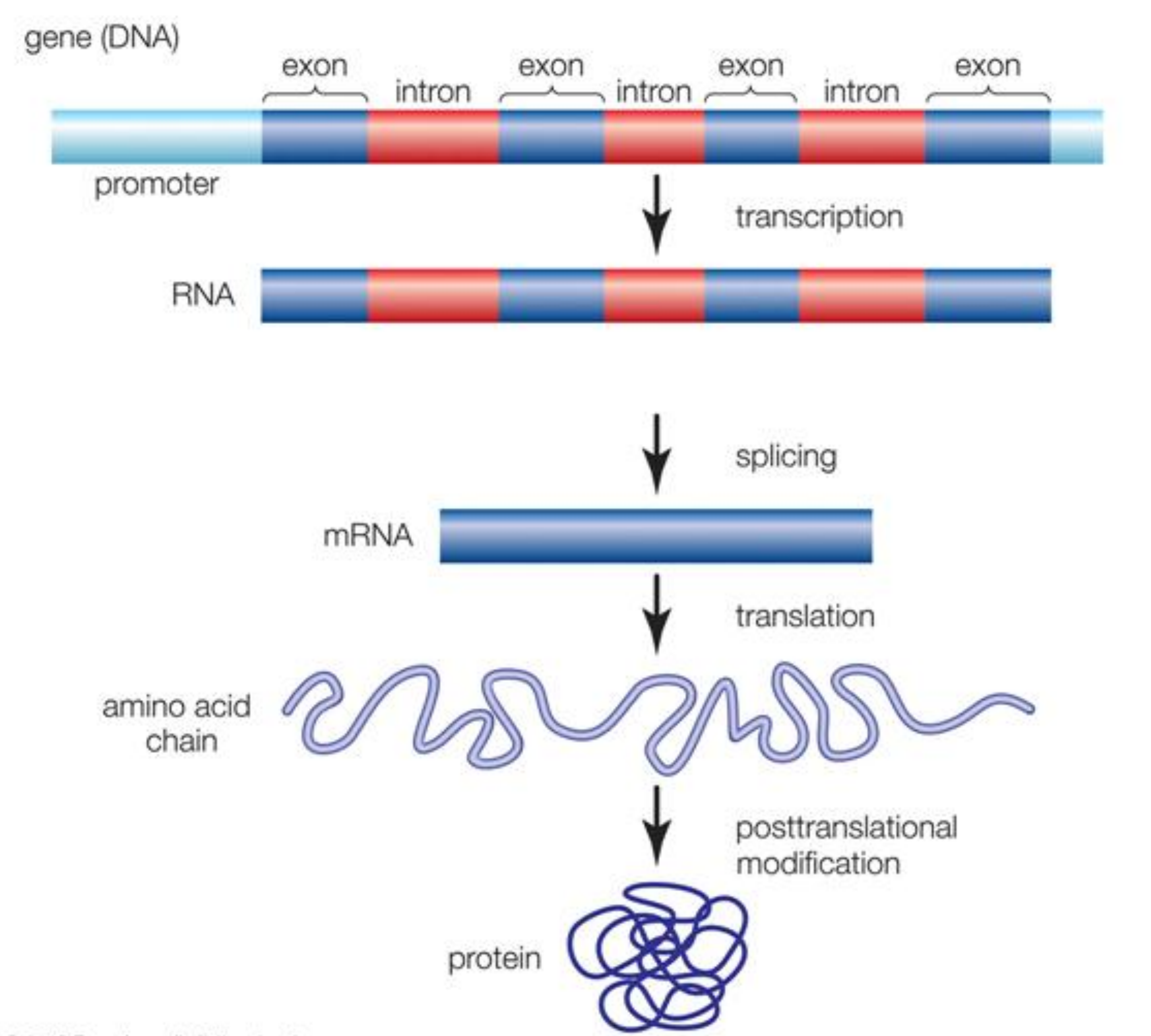
- There are mainly two types of structure predictions; one is for proteins and the other is for genes.
- For the understanding of gene prediction and analysis terms one should know what exactly genes actually are.
- Genes are biological unit of heredity; they are located at distinct position (locus) on a specific chromosome. Basically genes are of two types:

1.) CODING GENES OR EXONS:

- A sequence of DNA that carries instructions to perform specific function or translated into protein. For example, there are globin genes that imparts instructions for the manufacturing of hemoglobin (which carry oxygen throughout the body) protein. As humans contains 50,000 different genes that work collectively in intricate way to perform different functions.

2.) NON-CODING GENES OR INTRONS:

- Usually do not contain any instructions for a function. Non-coding genes never translated into proteins.



STRUCTURE OF GENE

- In a count to regions that overtly code for a protein mostly genes are comprised of regulatory regions. Given below are the regions that all genes have :

Promoter region: It is a regulatory region of gene, which controls the initiation of RNA transcription.

Enhancers: They enhance the transcription.

Introns: are non-coding regions and these regions are transcribed but do not translate into proteins.

Exons: are coding regions and these regions are translated into proteins

GENSCAN : GENE STRUCTURE PREDICTION

Now for the complete structure presiction of gene by using computational advances is to find out the location and function of gene. The main problem is to separate and define the exon-intron boundaries of a gene.

Two approaches followed by gene prediction :

- Statistical patterns identification
- Sequence similarity comparison

1.) STATISTICAL PATTERNS IDENTIFICATION:

- This approach of gene prediction uses all-purpose knowledge about gene structure i.e. statistics and rules. Knowledge of gene structure as discussed earlier includes promoter region (where transcription initiates), Start and end sequences of intron and exon etc.

2.) SEQUENCE SIMILARITY COMPARISON:

- This approach of gene prediction uses similarity based techniques. As similarity is based on evolution, either our sequence is homologous or not. It takes advantage on yhe fact that if the sequence is similar it will have same function. But the structure of gene cannot be predicted accurately based on sequence information alone.

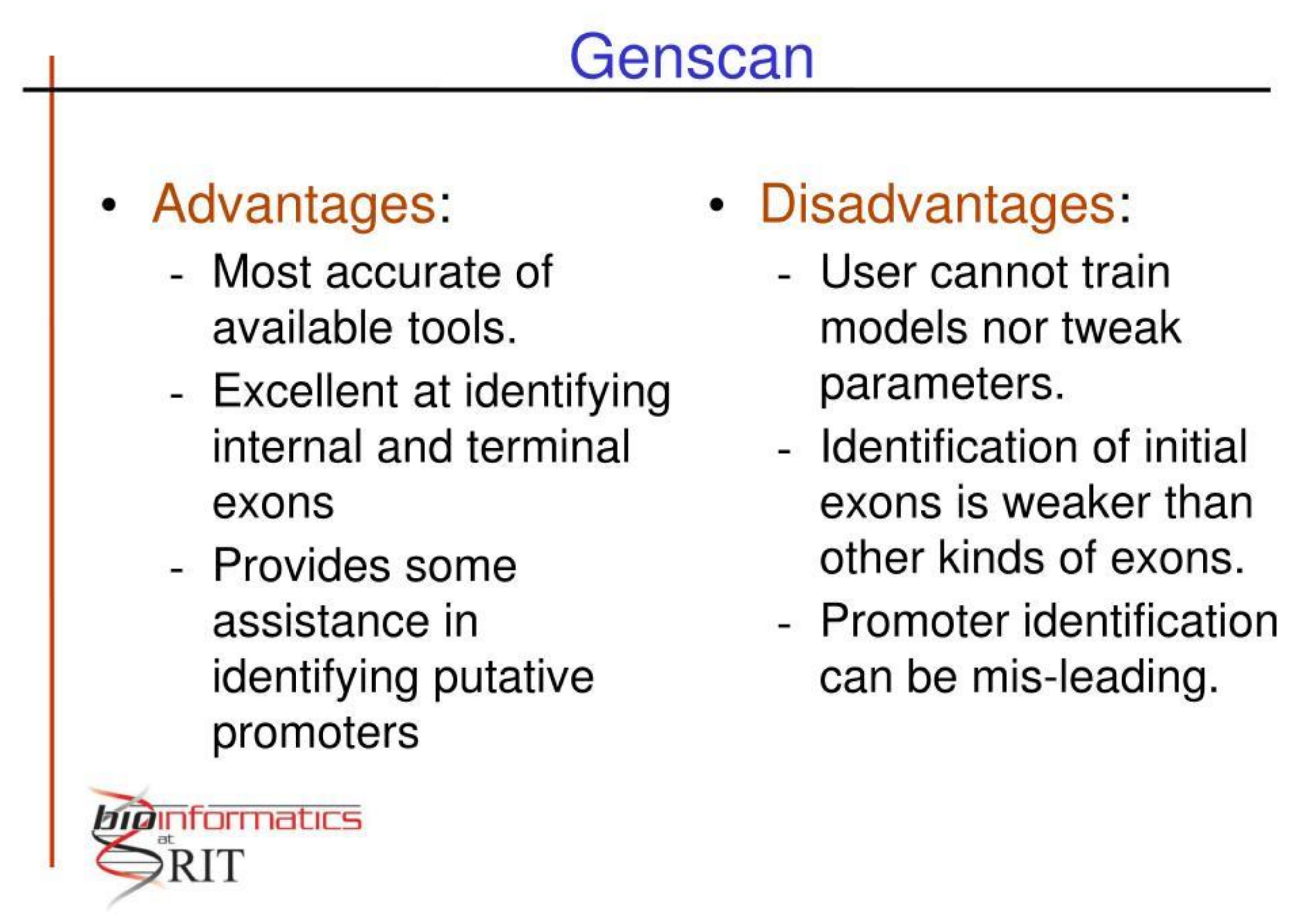
GENSCAN : ANALYSIS

- For the large scale analysis of gene, the typical strategy is to completely inactivate each gene or over express it. In each case resulting phenotype may not be informative. The loss of many proteins is lethal and this tells us the protein is essential but it does not tell us what actually protein does.
- After the prediction of gene structure we can investigate its function, expression level, diseases, mutations, etc. By using this information we can cure different diseases.

WHY TO USE GENSCAN?

- It can help identify disease severity.
- It can help children and adults to live a long, healthy life.
- It can help people concentrate on precautionary measures against numerous, serious debilities.
- It can predict gene structure to investigate its function, expression level, disease, mutation.
- It can prevent the disease by delay of occurence of disease.

○ ● ○



Genscan

- Advantages:
  - Most accurate of available tools.
  - Excellent at identifying internal and terminal exons
  - Provides some assistance in identifying putative promoters
- Disadvantages:
  - User cannot train models nor tweak parameters.
  - Identification of initial exons is weaker than other kinds of exons.
  - Promoter identification can be mis-leading.

bioinformatics
at
RIT

BIOINFORMATICS SOFTWARE DEVELOPED TO SOLVE SAME PROBLEMS

GENSCAN (*Burge* 1997)
FGENESH (*Solvyev* 1997)
HMMgene (*Krogh* 1997)
GENIE (*Kulp* 1996)
GENMARK (*Borodovsky* & *McInich* 1993)
VEIL (*Henderson,Salzberg* & *Fasman* 1997)

REFERENCES :-

- HTTPS://THEBIOLOGYNOTES.COM/GENE-PREDICTION-IMPORTANCE-AND-METHODS/

- HTTPS://BIOTECHARTICLES.COM/BIOINFORMATICS-ARTICLE/GENSCAN-BIOINFORMATICS-SOFTWARE-FOR-STRUCTURE-PREDICTION-AND-ANALYSIS-OF-GENE-284.HTML

- HTTPS://GENOMEBIOLOGY.BIOMEDCENTRAL.COM/ARTICLES/10.1186/GB-2000-1-1-REPORTS222

THANK YOU !!!