

Amrita Vishwa Vidyapeetam

Biological Systems – 21BIO211

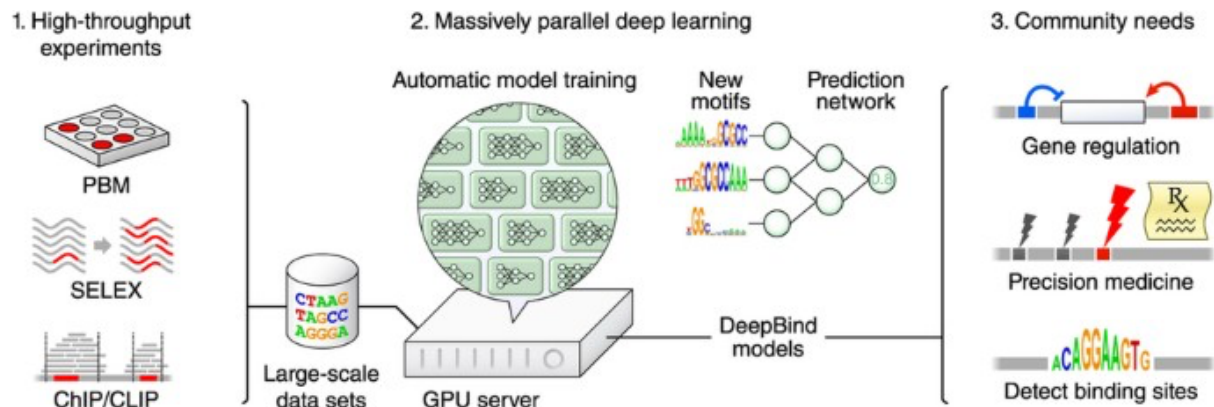
Lab Assignment – 4

NAME:- J Viswaksena

ROLL.NO:- AM.EN.U4AIE21035

1. How to you predict RNA /DNA binding protein using deep learning tool DEEPBIND .

- Access the DeepBind web server.
- Prepare the protein sequence.
- Input the sequence into DeepBind's input box.
- Start the prediction process.
- Await the results, which include a position-specific scoring matrix (PSSM).
- Analyze the PSSM to identify binding preferences and motifs.
- Validate and further investigate using experimental methods.
- Remember to consider other evidence and knowledge for interpretation.



1. The sequence specificities of DNA- and RNA-binding proteins can now be measured by several types of high-throughput assay, including PBM, SELEX, and ChIP- and CLIP-seq techniques.
2. DeepBind captures these binding specificities from raw sequence data by jointly discovering new sequence motifs along with rules for combining them into a predictive binding score. Graphics processing units (GPUs) are used to automatically train high-quality models, with expert tuning allowed but not required.
3. The resulting DeepBind models can then be used to identify binding sites in test sequences and to score the effects of novel mutations.

- ❖ Download fasta file from ncbi website and paste in below field called **Target Sequence in FASTA**.
- ❖ Click for search. To get results in the next page

DeepBind

a tool for predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning

Arabidopsis thaliana
Aspergillus nidulans
Caenorhabditis elegans
Danio rerio
Drosophila melanogaster
Gallus gallus
Homo sapiens

A1CF
ALX3
ALX4
ANKHD1
AR
ARID3A
ARNTL
ARX
ATF1
ATF2

Leishmania major

The selected protein:ALX4

Target Sequence in FASTA: [Example](#)

>RBFOX1
AGGUAAUAAUUUGCAUGAAUAACUUGGAGAGGAUAGC
>MBNL1
AGACAGAGCUUCCAUCAGCGCUAGCAGCAGAGACCAUU
>GATA3
GAGGTTACGCGGCAAGATAA
>CTCF
TACCACTAGGGGGCGCCACC

... or, upload a file:

Result

The table below shows the prediction of protein and sequence binding. The higher the score, the more likely the combination is.

	ALX4(D00292.001)
RBFOX1	-0.24651
MBNL1	-0.04128
GATA3	-0.72431
CTCF	-0.52671

Description

ID	Protein	Species	Experiment	Experiment Details
D00292.001	ALX4	Homo sapiens	SELEX	['CloneType=DBD','Primer=TGTGTC20NGA','Cycle=2','Batch=W']

- **The ID of protein:** A unique id of the form 01234.567 that identifies a model 01234 is unique for each combination of (Protein, Species), and 567 is unique for each combination of (Experiment, Experiment Details, Model ID), In other words, 01234 identifies the protein 'version', whereas 567 identifies the model 'version' for that protein version.
- For more information about the ID of protein,you can visit the website '**DeepBind**'.

DeepBind

a tool for predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning

1.

Arabidopsis thaliana
Aspergillus nidulans
Caenorhabditis elegans
Danio rerio
Pcbp2
Rbm38
Rbm4.3
Sf3b4
Drosophila melanogaster
Gallus gallus
Homo sapiens
Leishmania major
Mus musculus
Naegleria gruberi

2.

The selected protein: Pcbp2

Target Sequence in FASTA:

Example

```
>RBOX1
AGGUAAUAAUUGCAUGAAUAACUUGGAGAGGAUAGC
>MBNL1
AGACAGAGCUUCCAUCAGCGCUAGCAGAGACCAUU
>GATA3
GAGGTTACGCGGCAAGATAA
```

... or, upload a file:

Browse ...

Search

Reset

he search results include target symbol, protein and its predict score. The results can be download.

Result

The table below shows the prediction of protein and sequence binding. The higher the score, the more likely the combination is.

Pcbp2(D00258.001)	
RBOX1	-0.029389
MBNL1	-0.087705
GATA3	-0.530471
CTCF	-0.390984

Description

ID	Protein	Species	Experiment	Experiment Details
D00258.001	Pcbp2	Danio rerio	RNAcompete	['RNAcompeteID=RNCMPT00246']

- **The ID of protein:** A unique id of the form 01234.567 that identifies a model 01234 is unique for each combination of (Protein, Species), and 567 is unique for each combination of (Experiment, Experiment Details, Model ID), In other words, 01234 identifies the protein 'version', whereas 567 identifies the model 'version' for that protein version.
- For more information about the ID of protein, you can visit the website '[DeepBind](#)'.