Volume 12, Issue 09 September, 2019

Bioinformatics up to Date

(Bioinformatics Infrastructure Facility, Biotechnology Division)
North-East Institute of Science & Technology
Jorhat -785006, Assam

Inside…..

About us

The Bioinformatics Infrastructure Facility (BIF) at Biotechnology division, CSIR NEIST, Jorhat runs under the Biotechnology Information System Network (BTISnet) programme of DBT, Ministry of Science & Technology, and Government of India. The Centre was established on 2nd February, 2008 to promote innovation in Biological research and education through Bioinformatics accomplishment. The main goal is to facilitate and expose students and researchers from different academic institutions of North East India in Bioinformatics. The center conduct training and workshops for enlightening the use of bioinformatics applications in biological research and development. The Centre has access to global information through 24 hour high speed internet facility, and also e-journal facilities with DeLCON, Science Direct etc. To date the Centre has profoundly extended support in R & D work with a great intensity to different biological discipline including medicinal chemistry, computer aided drug design, genomics and proteomic data analysis etc.

Adverse drug reaction detection via a multihop self-attention mechanism

The adverse reactions caused by drugs are potentially life-threatening problems. Detecting ADRs through clinical trials takes a large number of experiments and a long period of time. With the growing amount of unstructured textual data, such as biomedical literature and electronic records, detecting ADRs in the available unstructured data has important implications for ADR research. The relationship of the two entities depends on more complex semantic information.Tongxuan Juang, et al., 2019 proposed a multihop self-attention mechanism (MSAM) model that aims to learn the multi-aspect semantic information for the ADR detection task. First, the contextual information of the sentence is captured by using the bidirectional long short-term memory (Bi-LSTM) model. Then, via applying the multiple steps of an attention mechanism, multiple semantic representations of a sentence are generated. Each attention step obtains a different attention distribution focusing on the different segments of the sentence. The model was evaluated by using two ADR corpora. The architecture of the model consists of four components: (1) The words are represented by word vector embedding and position embedding, respectively. (2) Bi-LSTM are used for extracting the contextual information in the sentence. (3) The multihop self-attention mechanism extracts complex semantic information. (4) The output layer shows the sentence classification. The paper further reports that proposed method significantly improved the learning of the complex seman-

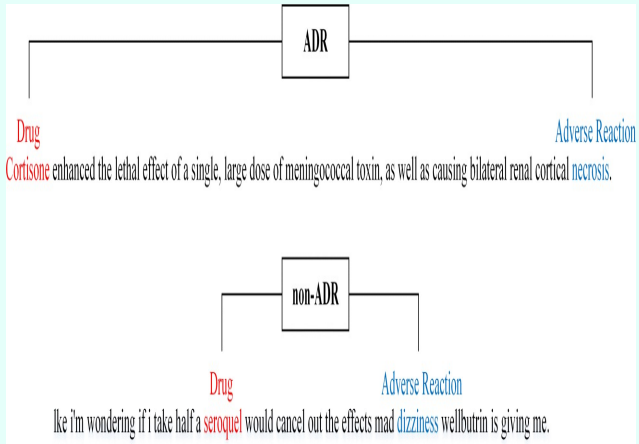


Fig: The examples of annotated sentences in the ADR corpus

Source: Tongxuan Juang, et al. *BMC Bioinformatics* 479(2019)

HPAanalyze: an R package for retrieval and analysis of Human Protein Atlas data

The Human Protein Atlas (HPA) is a comprehensive resource for exploration of the human proteome which contains a vast amount of proteomics and transcriptomics data generated from antibody-based tissue micro-array profiling and RNA deep-sequencing . The Human Protein Atlas (HPA) aims to map human proteins via multiple technologies including imaging, proteomics and transcriptomics. Access of the HPA data is mainly via web-based interface allowing views of individual proteins, which may not be optimal for data analysis of a gene set, or automatic retrieval of original images. HPAanalyze is an R package developed by Anh Nhat Tran, et al. for retrieving and performing exploratory analysis of data from HPA. HPAanalyze is reported to provide functionality for importing data tables and xml files from HPA, exporting and visualizing data, as well as downloading all staining images of interest. The package is free, open source, and available via Bioconductor and GitHub. *HPAanalyze* is designed to fulfill three main tasks: (1) import, subsetting and export downloadable datasets; (2) visualization of downloadable datasets for exploratory analysis; and (3) facilitation of work with individual XML files. The full downloadable datasets include normal tissue, pathology (cancer), subcellular location, RNA gene, and RNA isoform data. Additionally, it also provides an interactive website for non-programmers to explore and visualize data without the use of R.

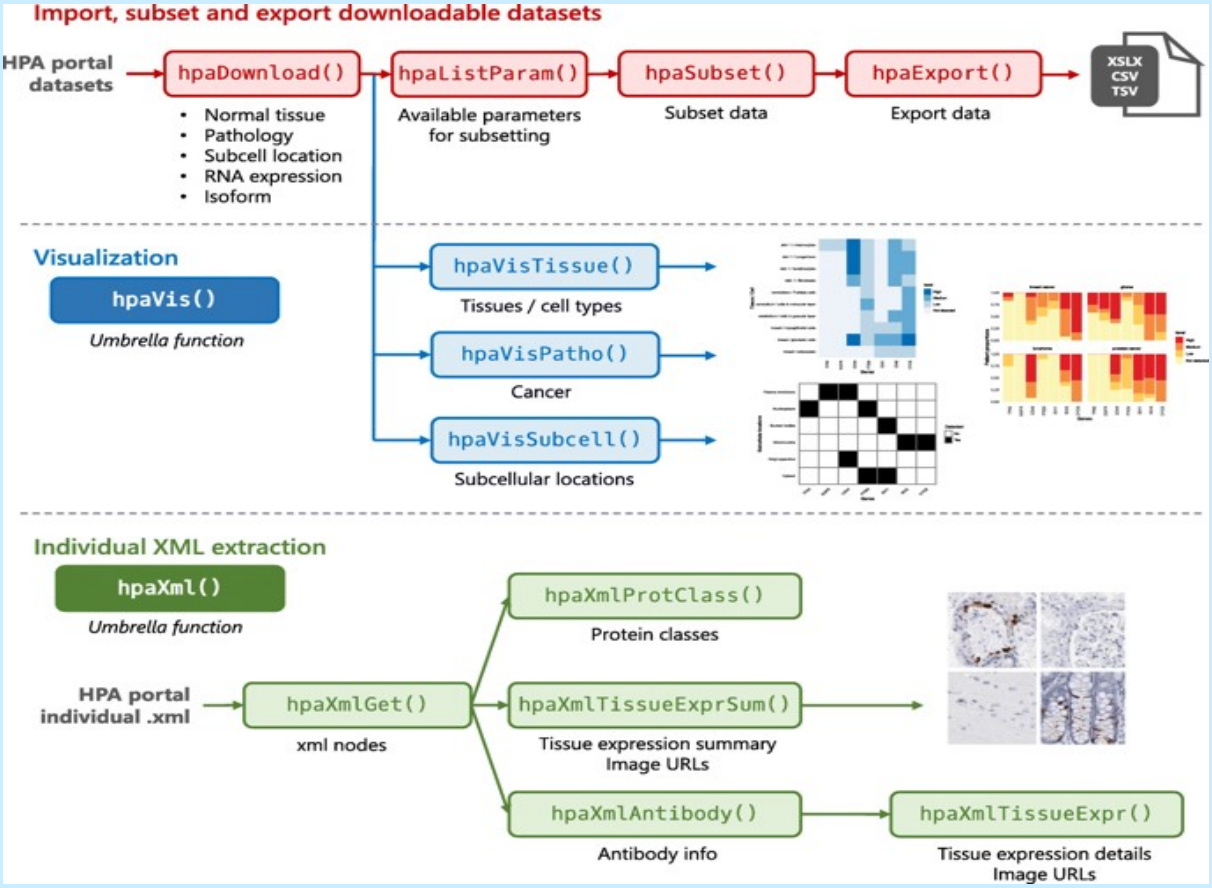


Figure: HPAanalyze Workflow

Source: Anh Nhat Tran, et al., BMC Bioinformatics 463(2019)

Bioinformatics Animation

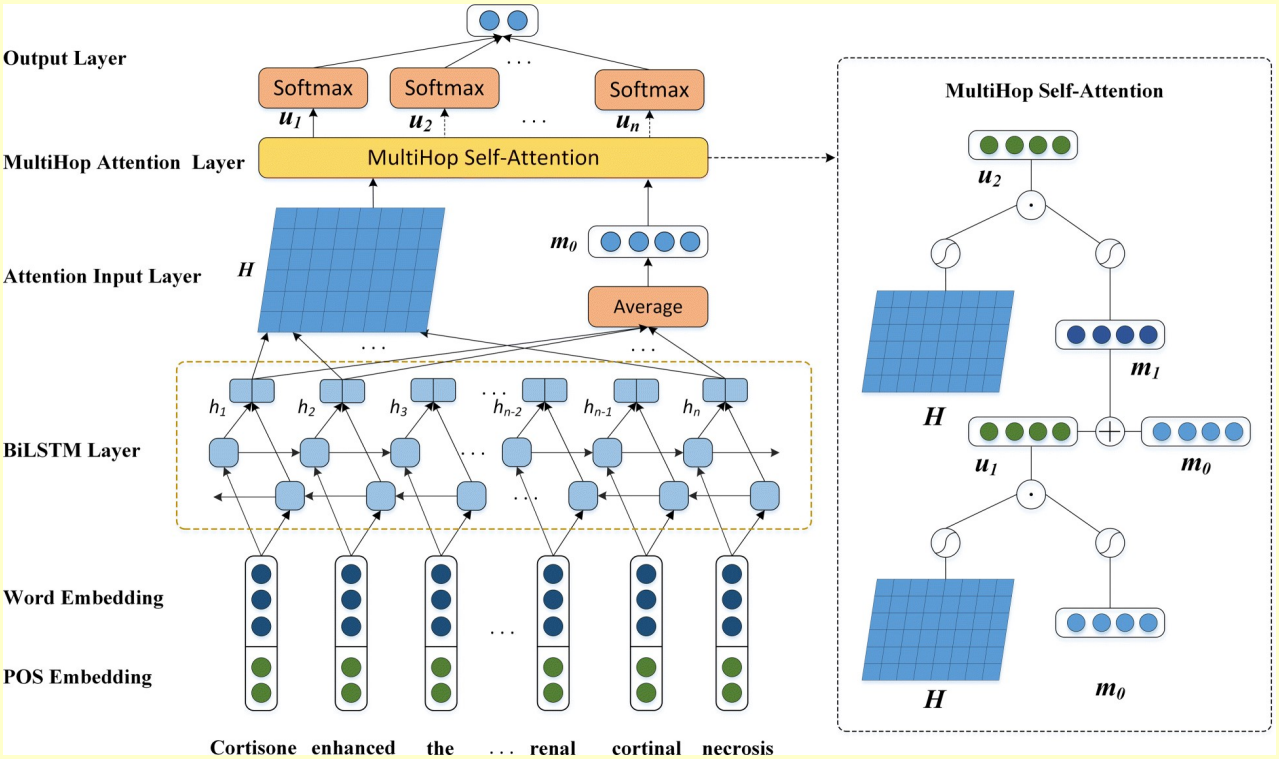
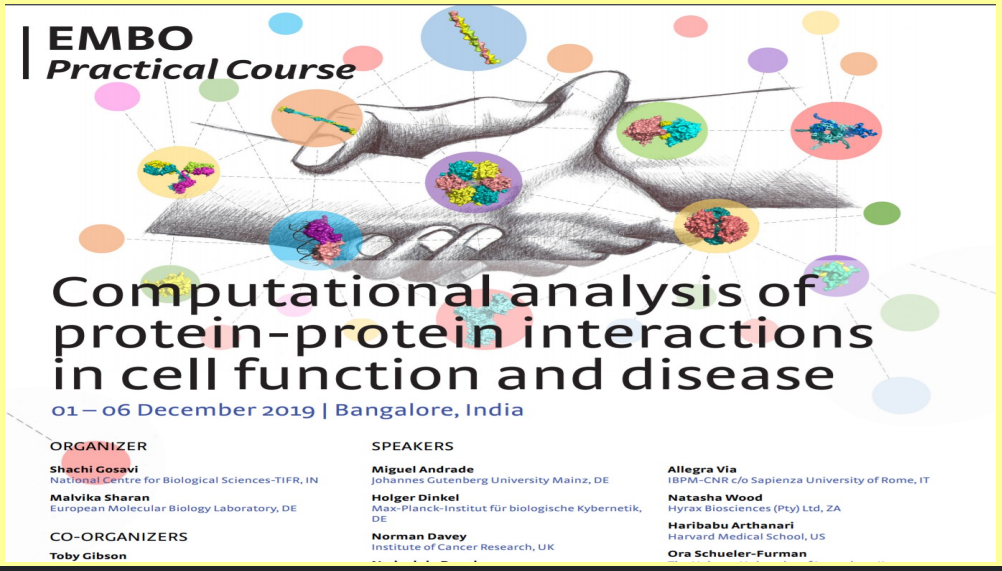


Figure: The sequential overview of Multi step Overview model proposed by Tongxuan Juang, et al., 2019.

Upcoming Events



http://meetings.embo.org/event/19-protein-protein

https://biotexcel.com/event/genomic-medicine-2019-nordic/

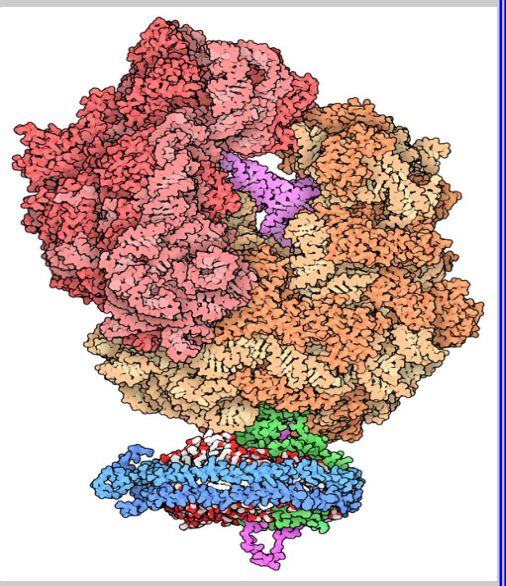
Molecule of the month

Nanodisc and HDL

The revolution in structural biology and new developments in cryoelectron microscopy has led the researchers to determine structures of molecular machines in a larger and complex way. Membrane-bound proteins are difficult to study because they are bound and protected in a lipid environment. Nanodiscs are made of a tiny disc of lipids which is able to carry one copy of a membrane-protein of interest, surrounded by a stabilizing belt of proteins.

When good cholesterol has also been good for science

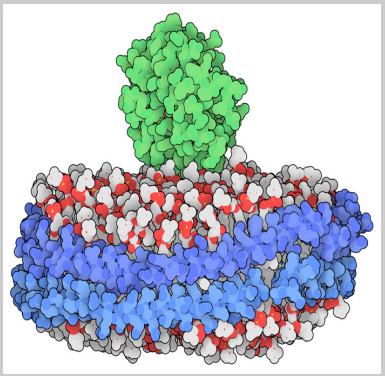
Cholesterol and lipids are transported through blood in small globules, surrounded by proteins called apolipoproteins. HDL particles are composed of large amount of protein which are denser than more lipid-rich varieties of lipoprotein particles. HDL are also known as good cholesterol since they are involved in the transport of excess cholesterol from blood stream to the liver. Nanodiscs were designed after inspired by one form of HDL which has a disc-like shape of high-density lipoprotein or HDL.



Structure of a ribosome (red and orange) synthesizing a new protein chain (magenta, still attached to a tRNA seen at the top), which is being transported through a nanodisc membrane (apolipoproteins in blue, lipids in white and red) by the secretory protein complex SecYE (green).

Nanodiscs are now found with different examples in the PDB archive. However, very of them have coordinates. Tis is because of high mobility of lipids and rather more attention on the proteins by the researchers make the bound-lipids blurred out in the cryoEM map. The researchers used cryoelectron microscopy combined with molecular dynamics simulations to generate a structural model of the protein synthesis machinery complex, including the nanodisc and its constituent lipids. One example is shown in the figure (PDB entry 4v6m). The structure shown here includes a ribosome secreting a newly-synthesized protein chain through the protein export channel SecYE.

Nanodiscs and NMR



Membrane-binding domain of MT1-MMP (green) bound to a nanodisc.

Structural biologists who use NMR also got advantages to study proteins from nanodiscs as they are small and soluble in water. The structure shown here (PDB entry 6clz) is the membrane-binding domain of MT1-MMP (membrane type 1 matrix metalloproteinase), a collagen-cutting enzyme that is needed to allow cells to migrate through existing extracellular matrix when new blood vessels are being built. This protein is also a target to fight cancer, since metastatic tumor cells often use this enzyme to spread. The structure al-

Source: https://pdb101.rcsb.org/motm/237

Kindly send us your feedback to

Dr Yumnam Silla Devi
BIF Center, Biotechnology Group, BSTD
CSIR-North East Institute of Science and Technology, Jorhat, Assam

Dr Ratul Saikia
BIF Center, Biotechnology Group, BSTD
CSIR-North East Institute of Science and Technology, Jorhat, Assam