BITS :: Call for Abstracts 2019 - Oral communication

Type	Oral communication
Session	Machine Learning in Bioinformatics
Title	A deep-learning method for predicting internal mitochondrial localization of proteins
All Authors	Castrense Savojardo (1), Nicolò Bruciaferri (1), Giacomo Tartari (1,2), Pier Luigi Martelli (1), Rita Casadio (1,2)

Affiliation

(1) Biocomputing Group, Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Bologna, Italy.
(2) Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), Italian National Research Council (CNR), Bari, Italy.

Motivation

The prediction of protein subcellular localization is a key step of the big effort towards protein functional annotation. Many computational methods exist to identify high-level protein subcellular compartments such as nucleus, cytoplasm or organelles. However, many organelles have their own internal compartmentalization. In mitochondria, the outer membrane separates the interior of the organelle from the cytoplasm, while the inner membrane encloses the mitochondrial matrix. The two membranes are separated by the intermembrane space. Such an internal compartmentalization suggests that proteins residing in the different compartments are specialized to fulfill different functions. Hence, knowing the precise location of a protein inside mitochondria is crucial for its accurate functional characterization.
The prediction of internal mitochondrial localization from sequence has been hampered in the past years by a substantial lack of experimental evidence in public sequence databases. Recently, thanks to the increasing amount of sequence data and experimental annotation, methods for predicting internal mitochondrial localization have appeared. Here, we explore the adoption of deep-learning approaches for improving prediction performance in this task.

Methods

Our method is based on artificial neural networks, in particular it adopts a One-Dimensional Convolutional Neural Network (OD-CNN) architecture to extract relevant patterns from primary sequence features and discriminate four different sub-mitochondrial compartments: outer membrane, inner membrane, intermembrane space and matrix. The architecture of our OD-CNN as well as the optimal input configuration were calibrated adopting a rigorous cross-validation procedure on a newly generated dataset comprising 424 highly-curated protein sequences extracted from UniprotKB/SwissProt and endowed with experimental evidence for sub-mitochondrial localization in the four considered compartments. Moreover, the performanceof our approach was evaluated on a number of different settings, including direct comparison with previously-developed approaches and proteome-wide analysis.

Results

In cross-validation experiments, our approach reported very good performances, reaching Matthews Correlation Coefficients (MCCs) of 0.46, 0.47, 0.53 and 0.65 in the discrimination of outer membrane, inner membrane, intermembrane space and matrix proteins, respectively. Interestingly, our approach is robust to class imbalance, reporting a good MCC score in discriminating proteins in the compartment characterized by the lowest number of sequences i.e. the intermembrane space.
When compared with the only method available in literature which is able to discriminate the four considered compartments, our method significantly outperformed its competitor in all MCC scores, showing performances that are much more stable across the four different classes.
Finally, we demonstrated the utility of our approach for proteome-scale analysis, including human data, with very high matching of predictions and available experimental information.

Info

-

Figure

-

Availability	http://busca.biocomp.unibo.it/deepmito

Corresponding Author	
Name, Surname	Castrense, Savojardo
Email	castrense.savojardo2@unibo.it
Submitted on	02.04.2019