# MYOP A GENE PREDICTOR FRAMEWORK

Kashiwabara, A.Y, Paschoal, A.R, Vicente, F.F.R., Durham, A.M.

Computer Science Department, University of São Paulo, Brazil

With the advent of massive sequence processing the amount of genomic data available far exceeds the capacity of researchers to investigate the data by purely biological means. Biological investigation not only is time consuming, but also expensive, making computational analysis a necessity to manage the large amount of data available. Genomics tasks involving the intensive use of bioinformatics include, but are not limited to: sequencing, assembling genes and genomes, predicting protein and ncRNA genes, assigning function to genomic sequences, predicting RNA secondary structure, characterizing families of genomic sequences (proteins, ncRNAs), and finding low complexity regions. Most genomics analyses involve the integration of different individual techniques and programs into a single analysis path. However, the straightforward solution of implementing a specific program to solve the immediate necessities of a project generally needs to solutions that are very hard to re-use or modify. The solution to this problem is the development of *Analysis Platforms* and *Frameworks,* systems that propose a general model for the solution of problems for a given scope, and that provide the general model for the solutions. In our presentation we will describe *MYOP*, a Framework for describing, training and running gene predictors. MYOP can be used to implement the models proposed by many of the most successful gene predictors available today like TIGRScan, Glimmer, GlimmerHMM, GenScan, GeneMark and PHAT. In a preliminary study we have used MYOP to implement 96 variations of the architectures of Glimmer, TIGLScan and GeneMark, and discovered novel and more efficient prediction architecture. We are also considering extending MYOP to include secondary structural signals for alternative splicing and ncRNA prediction.

**Financial support:** CAPES and CNPq
**Keywords:** RNA and gene prediction, software framework