Can we predict protein essentiality based on their pyshico-chemical features ?

Mauricio Lopes Casagrande, Ney Lemke, Marcio Luis Acencio

Abstract

The way genes organize and behave during the life of a living being hints that there may be a set of genes essential to life and reproduction, acknowledged as essential genes. These genes show high evolutionary conservation rate when compared to genomic media and stay relatively preserved through time, making it possible to find homology even on distant species. The process of discovering essential genes in vitro requires gene expression modulation techniques such as single-gene knockout, RNA interference, conditional knockout or CRISPR, and has always been an extensive, laborious effort. Nowadays it is possible to shorten this effort by computational, ortogenetic or phylogenetic approaches, but any method able to shorten this effort even more would be welcome. This work intends to present an in silico method to aid prediction of gene essentiality based only on physico-chemical features of the proteins synthesised by a gene. We applied machine learning techniques to verify if there really is some kind of relation between these features and gene essentiality and to evaluate the predicting capability of two algorithms using two distinct organisms. The knowledge about the set of essential genes of a certain organism could help improving our comprehension about the mechanisms related to the genetic base needed to sustain life, signaling new effective antimicrobial drug targets, increasing our knowledge about synthetic life or even guiding genetic therapy.