Analysis of Cancer Genes Duplicability

ID - 174

Rambaldi Davide[1], Francesca Ciccarelli[1]

[1]Bionformatics Group, IFOM-IEO Campus, Milano

Motivation

This research investigates the relationship between duplicability of cancer genes and their tumorigenic potential. A major question in molecular evolution concerns the role of gene duplication in the organization of genomes: to which extent are duplicated genes retained into the genome? How does this retention correlate with the gene phenotype and/or with the genome features? Genes coding for complex subunits and for highly connected proteins duplicate at a lower rate when compared to random sets, and their duplicates are generally not retained. Similarly, an inverse correlation has been observed between alternative splicing and gene duplicability. All these observations seem to suggest that a correlation does exist between the tendency of a gene to duplicate and its function and regulation. We want verify if there is a relation between the cancer genes duplicability and their tumorigenic potential.

Methods

In order to define which is the duplicability of genes related to tumor initiation and development, we rely on the Cancer Gene Census (below reported as reference set). This is a dataset of genes collected at the Sanger Center that are implicated in oncogenesis and show two independent experimental evidence of mutation in primary tumours.

To detect the duplications specifically associated to cancer genes, we aligned the protein sequences corresponding to the reference set over the human genome using an in-house developed bioinformatics pipeline. In respect to a protein-protein comparison, this pipeline has the advantage of extracting not only the putative protein-coding paralogs, but also pseudogenes that are not transcribed.

For each cancer protein in the reference set, a duplicate is defined as a genomic match other than the best hit over a given threshold. This threshold was chosen as the best compromise between sensitivity and specificity. We applied the same pipeline for the identification of duplicates also to a random datasets of human genes, and to a collection of genes with the same functional distribution.

Results

According to our pipeline, the cancer genes duplicate at a rate significantly lower than random sets of human genes.

We are now trying to combine this lower duplicability signal with other characteristics to predict the oncogenic potential of human genes that have not been yet related to cancer.

Email: davide.rambaldi@ifom-ieo-campus.it