Evolution of protein interaction network properties of cancer genes

D'Antonio M, Ciccarelli FD

Motivation

Our lab has recently shown that cancer genes (i.e. genes whose mutations actively contribute to cancerogenesis) tend to be singletons and to code for highly connected and interconnected proteins of the human protein-protein interaction network (PIN). Surprisingly, in the entire human PIN, proteins encoded by singleton genes are in general less connected than those encoded by duplicable genes. Our analysis therefore suggested that cancer genes have evolved differently than other human singletons and that they display peculiar genetic and network properties. We intend to understand the reasons of this peculiarity and the potential contribution to cancer development. An important question that arises in this context is whether the evolutionary appearance of a gene influences its network properties.

Methods

We defined the "age" of a gene as the deepest taxonomic branch of the tree of life where an ortholog could be detected, and we used eggNOG in order to gather the orthology relationships. We gathered PIN information from several publicly available databases and integrated these data in order to have a non-redundant set of protein-protein interactions. We used two datasets of cancer genes: the Cancer Gene Census (CGC), which is literature-based and manually curated, and the candidate (CAN) genes from high-throughput mutational screenings in breast, colorectal, pancreatic cancer and glioblasoma.

Results

When comparing cancer genes to the Rest Of Human Genes (ROHGs), we observed statistically different evolutionary tendencies. For both datasets of cancer genes, there is a significant burst at the Metazoan level and depletion at Mammal and Primate levels. Out of more than 700 cancer genes, only four are primatespecific. Furthermore, when comparing the evolutionary pattern of oncogenes and tumor suppressors, we found that tumor suppressors behave in a different way, with a burst at the last common ancestor and then no differences against the ROHGs. We analyzed the network properties of cancer genes and their orthologs in four model organisms (Homo sapiens, Drosophila melanogaster,

IFOM-IEO CAMPUS, via Adamello 16, 20139, Milan, Italy

Saccharomyces cerevisiae and Escherichia coli) in order to understand whether the property of the higher connectivity is specific for human cancer genes or it is conserved also in their orthologs and, thus, it is a result of the evolution of these genes. With few exceptions, mainly due to the dimensions of the dataset, there is a strong tendency for the orthologs of cancer genes to be more connected. We then analyzed the network properties of all the orthologs of human genes, in order to understand whether the higher connectivity is related only to cancer genes or it is a common property of the genes which are conserved in evolution: we demonstrated that the orthologs of cancer genes are more highly connected than the genes which do not have human orthologs in all the three species. We further investigated this finding, in order to understand whether a relationship exists between the age of a gene and its network properties. We divided the genes into age classes and compared connectivity and centrality between the different classes: older genes are more connected and more central than younger genes. This tendency is consistent in all the organisms with one exception in D. melanogaster. We have demonstrated that cancer genes have evolved in a different way, compared to the rest of human genes. They follow a different evolutionary pattern, although not all cancer genes behave in the same way: oncogenes have a burst at the metazoans level and are depleted in the youngest groups, while tumor suppressors have a burst at the last common ancestor level and, after that, they do not display appreciable differences when compared to the ROHGs. The higher connectivity of cancer genes seems to be due to their evolutionary pattern: the higher connectivity is not a peculiarity of cancer genes, but older genes are more connected than younger genes and, since cancer genes are depleted in the youngest genes (mammal- and primate-specific genes), they display a higher connectivity. The relationship between connectivity and gene age is very strong and it is seen also in other model organisms, which are evolutionarily distant from human, such as D. melanogaster, S. cerevisiae and E. coli. This relationship may be explained by the preferential attachment theory, since new nodes in a network preferentially connect to the already highly connected nodes, thus the ancient nodes are more likely to have a high number of interactions.

Contact e-mail
matteo.dantonio@ifom-ieo-campus.it