

Moonlight: a tool for biological interpretation and driver genes discovery

Colaprico* A(1,2,3)[†], Olsen C(1,2)*, Cava C(4), Terkelsen T(5), Bailey MH(6,7),
Silva TC(8) Olsen AV(5), Cantini L(9,10,11), Bertoli G(4), Zinoviyev A(9,10,11),
Barillot E(9,10,11), Castiglioni I(4), Noushmehr H(12), Ding L(6,7), Papaleo E#(5),
Bontempi G#(1,2)

(1) *Interuniversity Institute of Bioinformatics in Brussels (IB)2, Brussels, Belgium*

(2) *Machine Learning Group, ULB, Brussels, Belgium*

(3) *Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, FL 33136, USA*

(4) *Institute of Molecular Bioimaging and Physiology of the National Research Council (IBFM-CNR), Milan, Italy*

(5) *Computational Biology Laboratory, and Center for Autophagy, Recycling and Disease, Danish Cancer Society Research, Center, Strandboulevarden 49, 2100, Copenhagen, Denmark*

(6) *Department of Medicine, Washington University in St. Louis, St. Louis, MO 63110, USA*

(7) *McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108, USA*

(8) *Department of Genetics (CISBi/NAP), Department of Surgery and Anatomy, University of Sao Paulo, Monte Alegre, Ribeirã o Preto, SP, Brazil*

(9) *Institut Curie, 26 rue d'Ulm, F-75248 Paris France*

(10) *INSERM, U900, Paris, F-75248 France*

(11) *Mines ParisTech, Fontainebleau, F-77300 France*

(12) *Department of Neurosurgery, Brain Tumor Center, Henry Ford Health System, Detroit, MI, USA*



[†] Email: antonio.colaprico@gmail.com

Motivation

Cancer is an extremely complex disease and the cancer-related alterations are highly heterogeneous. It is crucial to identify the key driver genes and their roles in cancer mechanisms. Driver genes are often elusive and their discovery is even complicated by the fact that the same gene can play diverse roles in different cancer types or stages. Key biological processes, such as proliferation and apoptosis, have been linked to cancer progression and they can be used to classify the cancer genes. Furthermore, high-throughput genomic initiatives have been providing a precious source of biological signatures associated to cancer. We propose a new framework, called Moonlight, to conduct an integrative bioinformatics analysis of (epi-)genomic and transcriptomic experiments to (i) classify and identify genes that act as tumor suppressor genes (TSGs) and oncogenes (OCGs) (ii) characterize intra- and inter-tumoral heterogeneity in cancer tissue (iii) evaluate and validate the prognostic value of dual role genes in cancer cell lines and in patient tumor xenograft

models by means of in-silico validation and in-vitro/in-vivo experiments. Our new and original approach relies on functional enrichment analyses, prediction of gene regulatory networks and upstream regulator analyses to score the importance of known biological processes with respect to the cancer type. The methodology not only allows us to identify OCGs and TSGs, but it also predicts genes with dual roles, i.e. TSG in one cancer and OCG in another but also to elucidate the underlying biological processes. The three objectives mentioned above will allow a better comprehension of global tumoral heterogeneity and to evaluate its impact on prognosis, resistance to therapy, as well as to guide therapeutic decisions.

Methods

The Moonlight protocol aims: To identify biological processes enriched by a gene set of differentially expressed genes (DEGs) between two conditions by means of Functional Enrichment Analysis (FEA); To Identify the upstream regulators of those biological processes that are enriched significantly in the comparison by means of URA; To identify 'driver genes' with dual role that acts as TSG or OCG in different cancer contexts (i.e., different cancer types, subtypes or stages).

Results

In summary, we have presented a method to determine which role cancer driver genes play in cancer development. Moonlight distinguishes itself from state-of-the-art approaches in multiple ways. It allows the understanding of cancer development from a different angle, gene expression, rather than the classic determination based on mutations. Moonlight also assigns relevance scores to the underlying biological processes, which is a unique feature. Lastly, Moonlight facilitates the identification of the specific role cancer genes play in a variety of contexts, such as specific cancer types or in even in different stages of cancer development. Recently two works within The Cancer Genome Atlas Research Network such as: (i) Ding et al., Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics, Cell, 2018 and (ii) Malta et al., Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation, Cell 2018 showed the capabilities of Moonlight functions such as DRA and URA in a combination using both Moonlight z-score and GSEA NES-score. These examples can help to elucidate further mechanism underlying cancer complexity. We are interested as future work to investigate causes of these effects captured by using Moonlight as tool of gene programs altered at gene level. We would like to hypotheses that more possible alterations in a gene, more is a key gene for a disease? Epi-genetic changes such as microRNA or long-noncoding RNA and DNA methylation differences regulator of these genes can be therapeutic tool? In particular for the the association of a miRNA with a TR could determine or not the degradation of the TR mRNA. Consequently the expression of TR could be down-regulated, in the case of degradation, or at least unchanged, when mRNA degradation does not occur allowing mRNA stability and accumulation.

Availability:

<https://bioconductor.org/packages/MoonlightR> & <https://github.com/ibsquare/MoonlightR>
[https://www.cell.com/cell/fulltext/S0092-8674\(18\)30313-1](https://www.cell.com/cell/fulltext/S0092-8674(18)30313-1)
[https://www.cell.com/cell/fulltext/S0092-8674\(18\)30358-1](https://www.cell.com/cell/fulltext/S0092-8674(18)30358-1)