cyTRON and cyTRON/JS: two Cytoscape-based applications for the inference of cancer evolution models

Lucrezia Patruno*[1], Edoardo Galimberti*[2,3], Daniele Ramazzotti[4], Giulio Caravagna[5], Luca De Sano[1], Marco Antoniotti[1,6], and Alex Graudenzi[1,7]

[1] *Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy*
[2] *Department of Computer Science, University of Turin, Turin, Italy*
[3] *ISI Foundation, Turin, Italy*
[4] *Department of Pathology, Stanford University, Stanford, CA, United States*
[5] *The Institute of Cancer Research - ICR, London, UK*
[6] *Milan Center for Neuroscience, Milan, Italy*
[7] *Institute of Molecular Bioimaging and Physiology - IBFM of the National Research Council - CNR, Segrate, Milan, Italy*

Abstract

The increasing availability of sequencing data of cancer samples is fueling the development of algorithmic strategies to investigate tumor heterogeneity and infer reliable models of cancer evolution. We here build up on previous works on cancer progression inference from genomic alteration data, to deliver two distinct `Cytoscape`-based applications, which allow to produce, visualize and manipulate cancer evolution models, also by interacting with public genomic and proteomics databases. In particular, we here introduce `cyTRON`, a standalone `Cytoscape` app, and `cyTRON/JS`, a web application which employs the functionalities of `Cytoscape/JS`.

cyTRON was developed in Java; the code is available at https://github.com/BIMIB-DISCo/cyTRON and on the Cytoscape App Store http://apps.cytoscape.org/apps/cytron. cyTRON/JS was developed in JavaScript and R; the source code of the tool is available at https://github.com/BIMIB-DISCo/cyTRON-js and the tool is accessible from https://bimib.disco.unimib.it/cytronjs/welcome.

*Equal contributors.

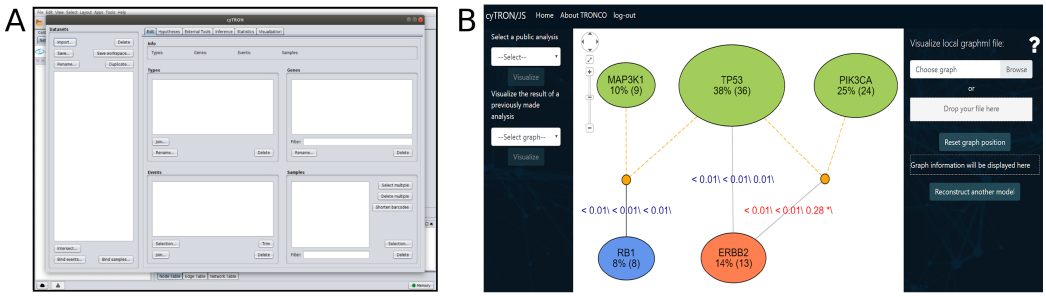

Figure 1: We show in the figure a view of `cyTRON` workspace (A) and an example of output model by `cyTRON/JS` (B).

1 Scientific Background

Cancer is a complex disease, whose development is caused by the accumulation of alterations in the genome. Some alterations may confer a selective advantage to cancer cells, and this may result in the expansion of cancer clones. In order to understand how cancer evolves, it is of great importance to understand how such *driver* alterations accumulate over time [Nowell, 1976, Burrell et al., 2013]. This goal can be pursued by reconstructing cancer evolution models, which are graphs that encode the evolutionary history of drivers and their temporal relationships. The reconstruction of such models is a complex task mainly because of two reasons: first, much of the data publicly available from initiatives such as `TCGA` [`https://portal.gdc.cancer.gov/`] comes from cross-sectional samples, and hence they lack of temporal information. The second main reason can be found in the heterogeneity of tumors [Caravagna et al., 2016, Ramazzotti et al., 2018].

2 Materials and Methods

In order to learn meaningful evolution models, we developed a pipeline, `PICNIC` [Caravagna et al., 2016], which includes the following steps: *i*) identification of homogeneous sample subgroups (e.g., tumor subtypes), *ii*) identification of drivers (i.e., the nodes of the output model), *iii*) identification of mutually exclusive patterns, *iv*) inference of cancer evolution models via distinct algorithms (e.g., [Loohuis et al., 2014, Ramazzotti et al., 2015, Ramazzotti et al., 2019]). This pipeline was implemented within the widely used `TRONCO R` suite for TRanslational ONCOlogy [De Sano et al., 2016], [Antoniotti et al., 2015], which was recently employed, for instance, to analyze the largest kidney cancer cohort currently available [Turajlic et al., 2018].

However, TRONCO presents two practical limitations: first, it requires at least some basic programming skills due to its underlying R infrastructure; second, TRONCO is not integrated with publicly available genomic databases, hence providing a non-interactive visualization of the output graphs.

Therefore, to improve the practicality, effectiveness, interactivity and diffusion of our framework, we integrated it within Cytoscape, an user-friendly open-source platform for the visualization and manipulation of complex networks [Shannon et al., 2003]. We here present cyTRON, a stand-alone Cytoscape app, and cyTRON/JS a web application which employs the functionalities of Cytoscape/JS, both of which allow to produce, visualize and manipulate cancer evolution models, also by interacting with public genomic databases. Figure 1B shows an example of the output in cyTRON/JS, which exploits Cytoscape/JS to provide an interactive visualization of the evolution model.

cyTRON and cyTRON/JS were designed for two main purposes:

- Providing an *interactive* and *user-friendly* visualization of TRONCO models: while TRONCO R-based graph display is static, cyTRON and cyTRON/JS provide interactive views, which allow to directly retrieve information about genes involved in the study, by accessing widely-used public genome databases.

- Making TRONCO *accessible* to users unfamiliar with R programming: cyTRON and cyTRON/JS provide interfaces that enable the usage of TRONCO respectively from Cytoscape and a Web browser, thus removing the need for users to execute any code in order to complete a whole analysis.

The architecture of both tools can be conceptually defined as follows:

- The *front-end* side is composed of an interface that can be used to:

    1. Select input data for the TRONCO analysis [De Sano et al., 2016]: the input files should be either MAF, GISTIC or user-defined Boolean matrices that contain information about the mutations observed in each sample.
    2. Set the parameters for the inference in order to access most TRONCO capabilities. Users can specify which driver mutations to include in the analysis, which algorithm among those implemented in TRONCO to use for the reconstruction and the algorithm's corresponding parameters.

3. Visualize cancer evolution models and dynamically interact with the result: for instance, by clicking on the genes of the output graph, it is possible to retrieve the information available on public genomic databases. `cyTRON` gives access to gene information through databases such as `Ensembl`[1] and `Entrez`[2], that are accessible through the Cytoscape interface.
   In `cyTRON/JS` the data displayed for each node are retrieved from the Entrez database `Gene` [3] using `E-Utils`, an API provided by the National Center for Biotechnology Information.

- The *back-end* side includes the communication channel with R. For `cyTRON`, a Java bridge with R is built by means of `rJava`. Instead, for `cyTRON/JS` it is based on `js-call-r`, a `Node.js` package which collects the data and parameters set by the user, encodes them in `JSON` and sends them to R. Then, R commands are transparently executed in order to perform any specific step of the analysis by TRONCO.

Figure 1 shows a view of `cyTRON` workspace (left) and an example of output model by `cyTRON/JS` (right). In order to choose between the two tools, users should take into consideration the data and the type of analysis they need to carry out. In particular, since `cyTRON/JS` is a web application, it is readily accessible from any device and computations are carried out on the back-end side. This feature is useful in case a user needs to carry out a computational-expensive analysis. However, `cyTRON` is more complete with respect to all the functionalities implemented in TRONCO: for example, it implements also the option of testing hypothesis on mutations through the algorithm Capri [Ramazzotti et al., 2015].

3 Conclusion and future work

`TRONCO` is an `R` package that implements state-of-the-art algorithms for the inference of cancer evolution models with the ultimate goal of understanding the evolutionary trajectories driving tumor evolution. In such a multidisciplinary domain, where computer scientists actively cooperate with biologists, being capable of visually understanding the data is crucial to both parties. In order to effectively allow the use of `TRONCO`, we here presented `cyTRON` and `cyTRON/JS`, two `Cytoscape`-based applications which translate many of

[1] `http://www.ensembl.org/index.html`
[2] `https://www.ncbi.nlm.nih.gov/search/`
[3] `https://www.ncbi.nlm.nih.gov/gene/`

TRONCO functionalities into Cytoscape. Our effort aims at designing user-friendly and accessible tools to support the user in the task of exploring cancer genomic data.

As the TRONCO functionalities are constantly updated and improved, these two new tools need to be kept up to date. Thus, future work will be focused on integrating TRONCO's new algorithms for analyzing single cell datasets [Ramazzotti et al., 2019]. In addition to this, cyTRON/JS needs to be extended with the hypothesis testing functionality, in order to enable users to carry out more complex analysis.

Acknowledgments

cyTRON and cyTRON/JS were developed within the Google summer of Code program, in collaboration with the NRNB organization. The authors declare no conflicts of interest.

References

[Antoniotti et al., 2015] Antoniotti, M., Caravagna, G., De Sano, L., Graudenzi, A., Mauri, G., Mishra, B., and Ramazzotti, D. (2015). Design of the tronco bioconductor package for translational oncology. *Submitted. Available on bioRxiv. org, ID*, 27524.

[Burrell et al., 2013] Burrell, R. A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467):338.

[Caravagna et al., 2016] Caravagna, G., Graudenzi, A., Ramazzotti, D., Sanz-Pamplona, R., De Sano, L., Mauri, G., Moreno, V., Antoniotti, M., and Mishra, B. (2016). Algorithmic methods to infer the evolutionary trajectories in cancer progression. *Proceedings of the National Academy of Sciences*, 113(28):E4025–E4034.

[De Sano et al., 2016] De Sano, L., Caravagna, G., Ramazzotti, D., Graudenzi, A., Mauri, G., Mishra, B., and Antoniotti, M. (2016). Tronco: an r package for the inference of cancer progression models from heterogeneous genomic data. *Bioinformatics*, 32(12):1911–1913.

[Loohuis et al., 2014] Loohuis, L. O., Caravagna, G., Graudenzi, A., Ramazzotti, D., Mauri, G., Antoniotti, M., and Mishra, B. (2014). Inferring tree

causal models of cancer progression with probability raising. *PloS one*, 9(10):e108358.

[Nowell, 1976] Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28.

[Ramazzotti et al., 2015] Ramazzotti, D., Caravagna, G., Olde Loohuis, L., Graudenzi, A., Korsunsky, I., Mauri, G., Antoniotti, M., and Mishra, B. (2015). Capri: efficient inference of cancer progression models from cross-sectional data. *Bioinformatics*, 31(18):3016–3026.

[Ramazzotti et al., 2018] Ramazzotti, D., Graudenzi, A., Caravagna, G., and Antoniotti, M. (2018). Modeling cumulative biological phenomena with suppes-bayes causal networks. *Evolutionary Bioinformatics*, 14:1176934318785167.

[Ramazzotti et al., 2019] Ramazzotti, D., Graudenzi, A., De Sano, L., Antoniotti, M., and Caravagna, G. (2019). Learning mutational graphs of individual tumour evolution from single-cell and multi-region sequencing data. *BMC bioinformatics*, 20(1):210.

[Shannon et al., 2003] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504.

[Turajlic et al., 2018] Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., OBrien, T., Lopez, J. I., Watkins, T. B., Nicol, D., et al. (2018). Deterministic evolutionary trajectories influence primary tumor growth: Tracerx renal. *Cell*, 173(3):595–610.