

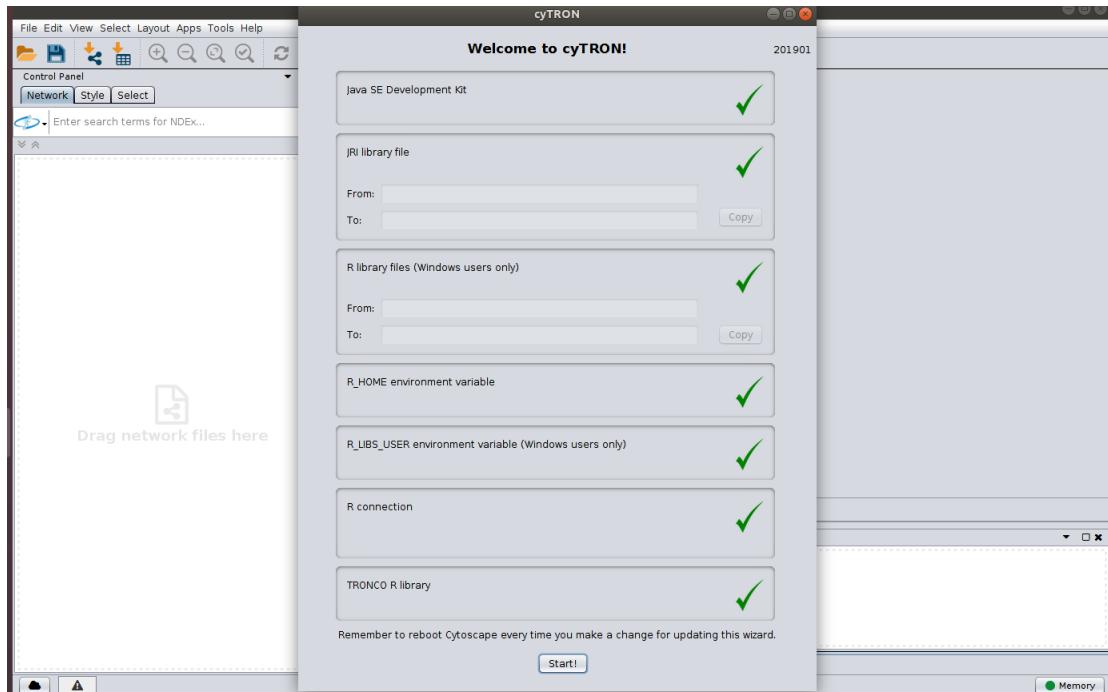
cyTRON: Case Study

We now provide a detailed tutorial for cyTRON replicating most of the analyses from [Caravagna, Giulio, et al. "Algorithmic methods to infer the evolutionary trajectories in cancer progression." *Proceedings of the National Academy of Sciences* 113.28 (2016): E4025-E4034].

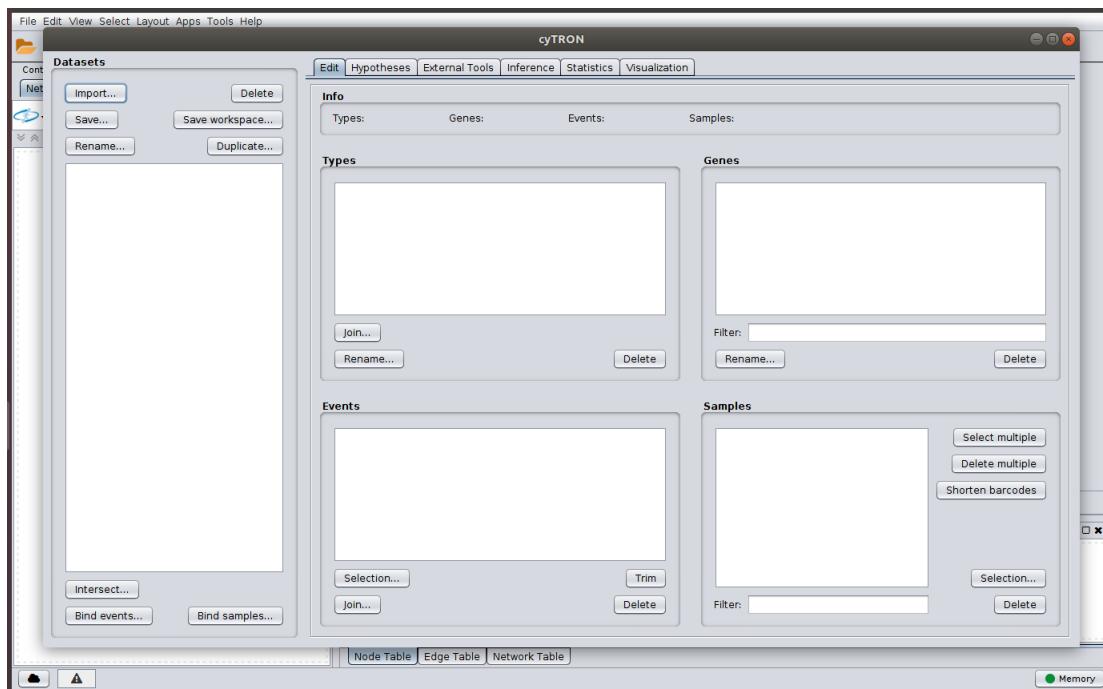
The data used in this tutorial are available for download on Github at: <https://github.com/BIMIB-DISCo/datasets/raw/master/TCGA-COADREAD-TRONCO.zip>.

cyTRON provides all the functionalities available from the TRONCO package. We also refer to the documentation and vignette of the package for more details: <https://github.com/BIMIB-DISCo/TRONCO>.

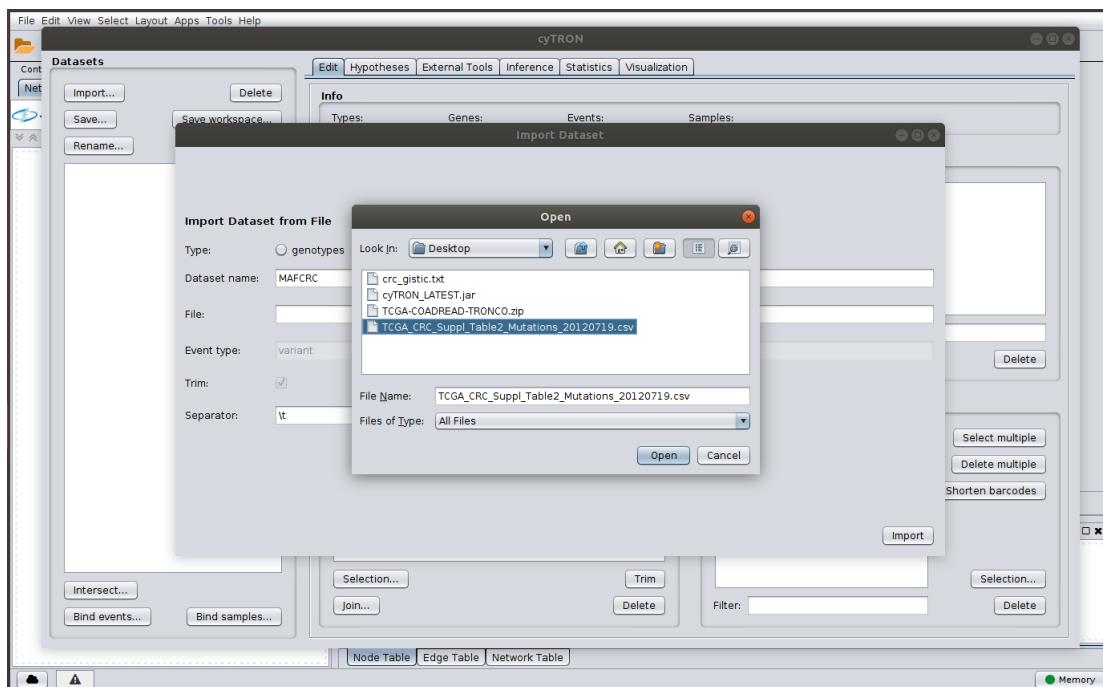
Running cyTRON: we first run Cytoscape and execute the cyTRON app within it. At the first run, cyTRON will display a welcome screen assuring that all the setup of the tool went well. In case of problems, these would be indicated at this point.

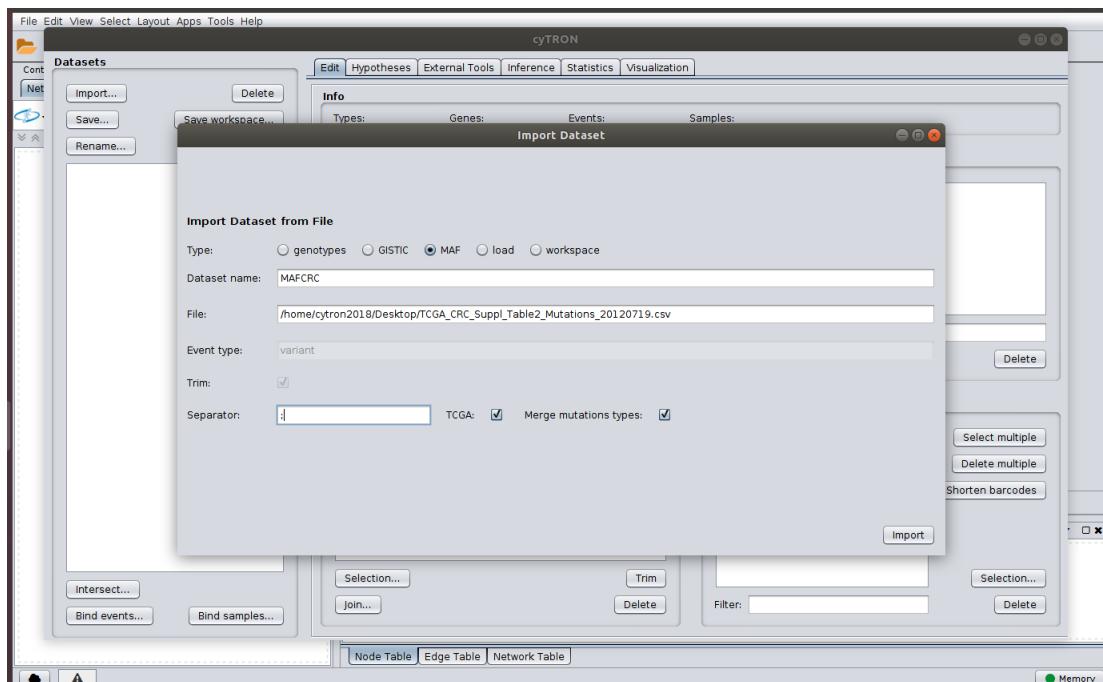


cyTRON workspace: we now leave the initial screen by clicking on the "Start" button and enter cyTRON workspace. The workspace has a fixed component on the left, where it is possible to load, edit and save datasets. The main component, instead, has multiple tabs, each of them reflecting specific functionalities available in the TRONCO R package.

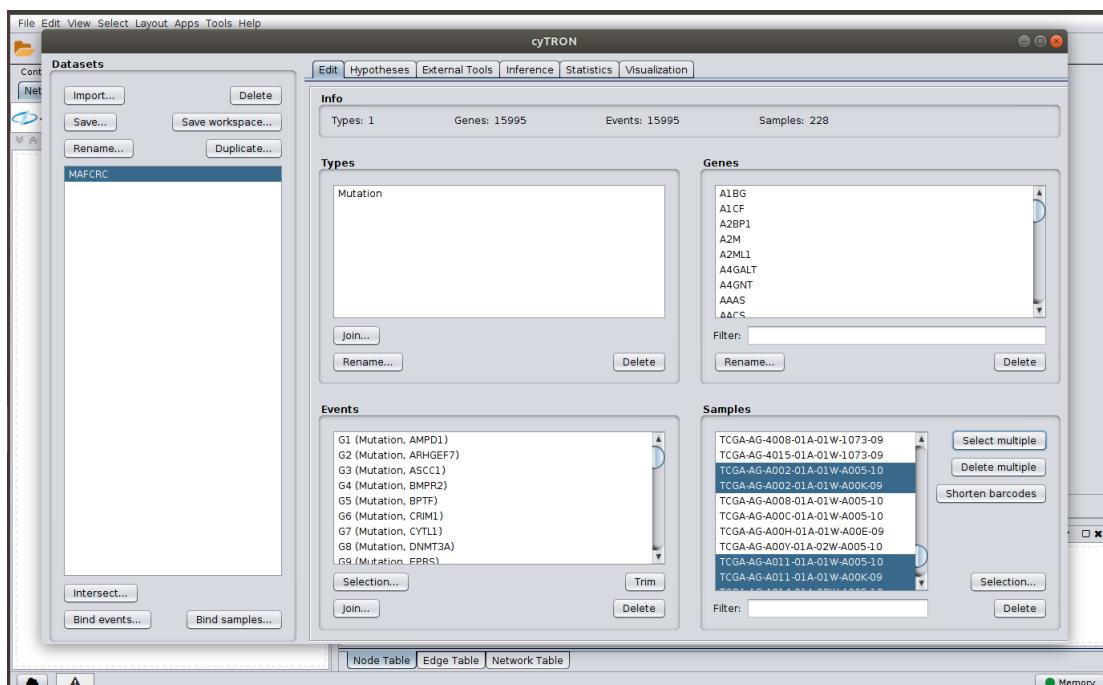


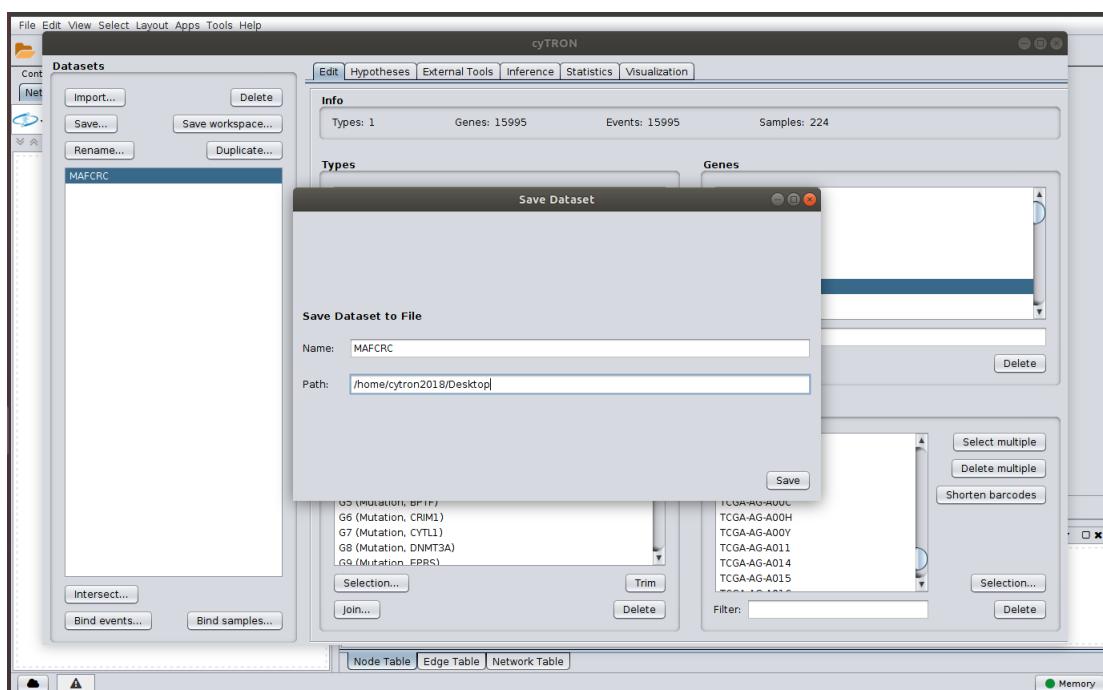
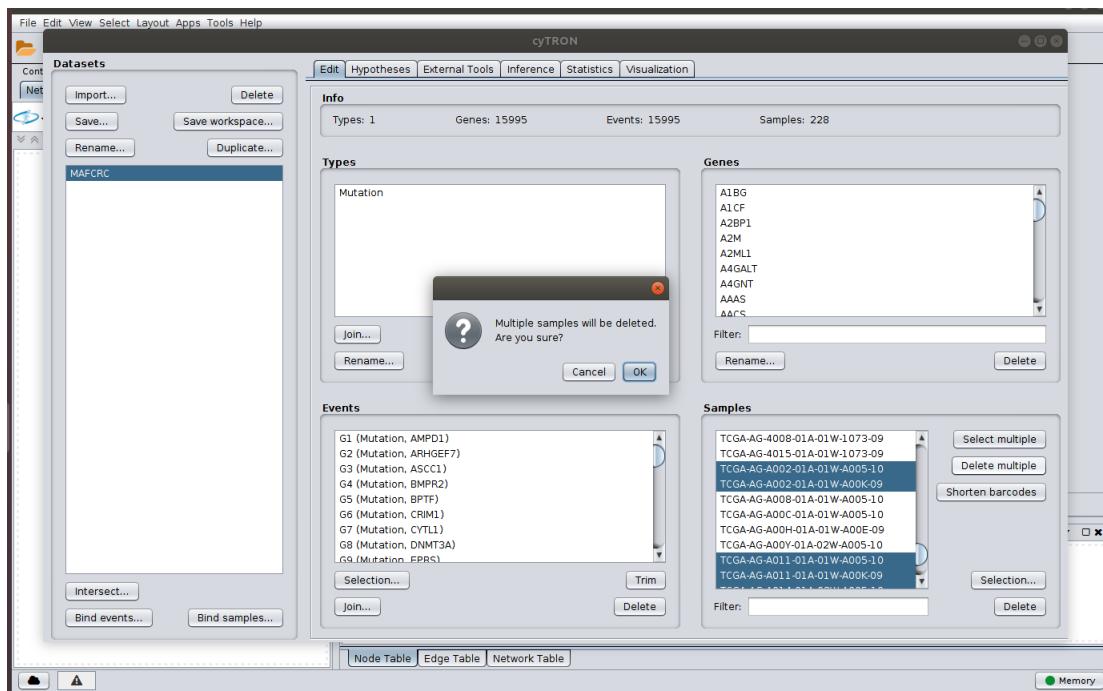
Importing a MAF file: we import point mutation data for a cohort of colorectal tumors by clicking the button “Import” on the top/left and selecting the file TCGA_CRC_Suppl_Table2_Mutations_20120719.csv. For more information about MAF format, see https://docs.gdc.cancer.gov/Data/File_Formats/MAF_Format/.



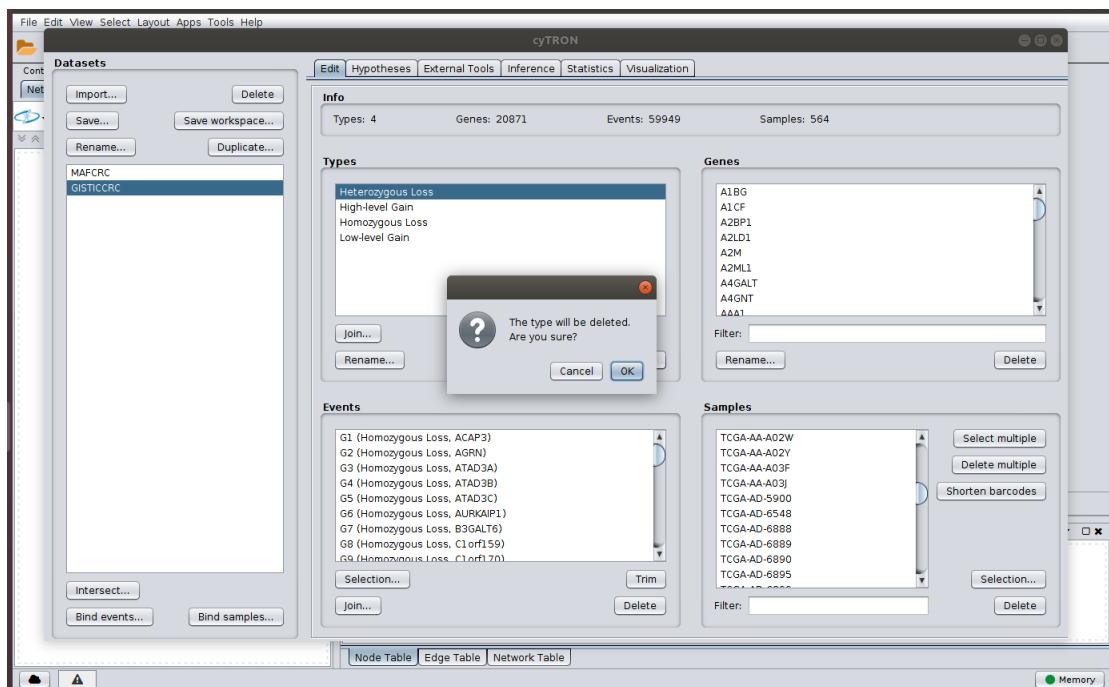
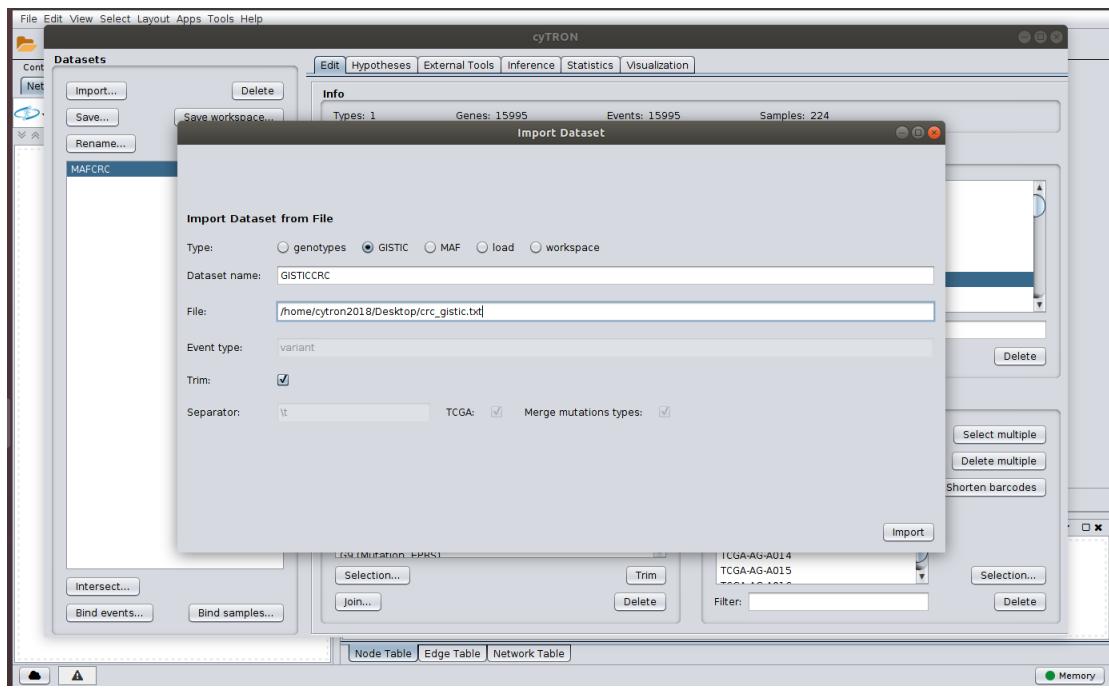


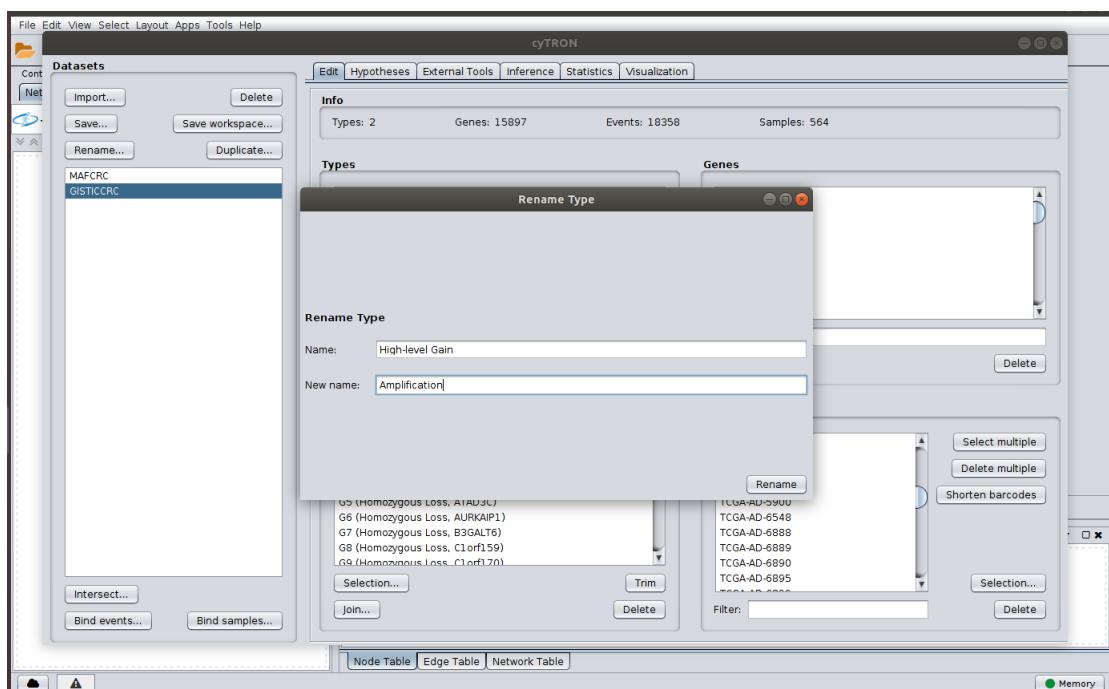
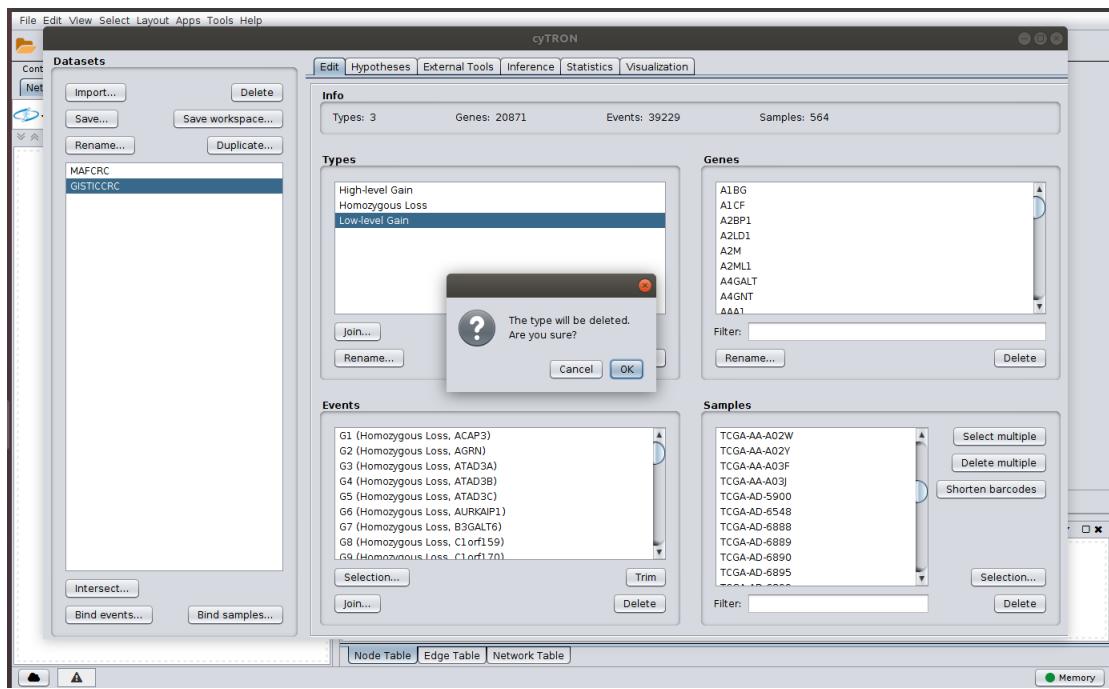
Deleting multiple samples and saving the results: we then select multiple samples within the dataset by clicking on the “Select multiple” button on the bottom/right of the Edit tab and delete multiple sample following TCGA guidelines. Finally, we save the results to file in order to be able to load them in the future. This action will create the file MAFCRC.rds.

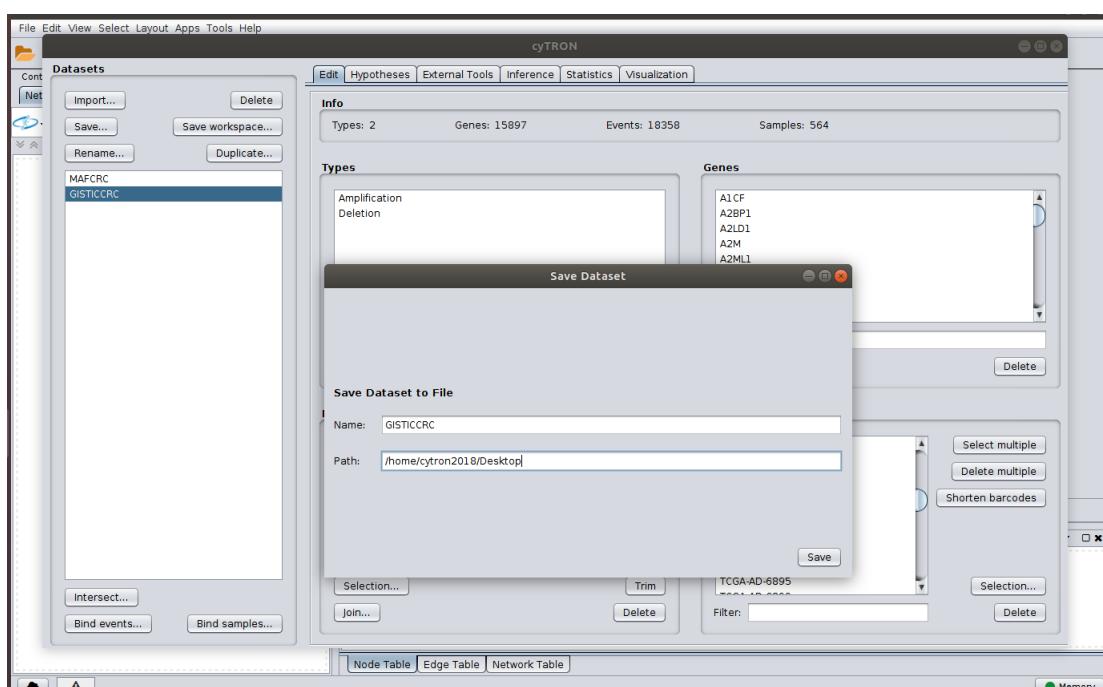
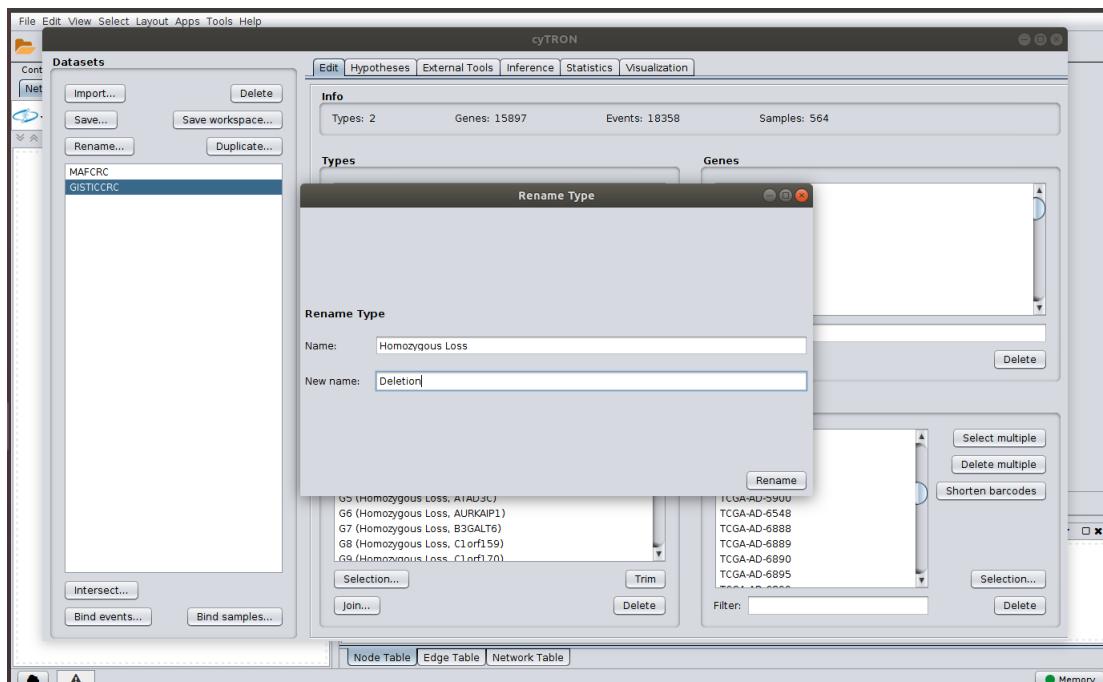




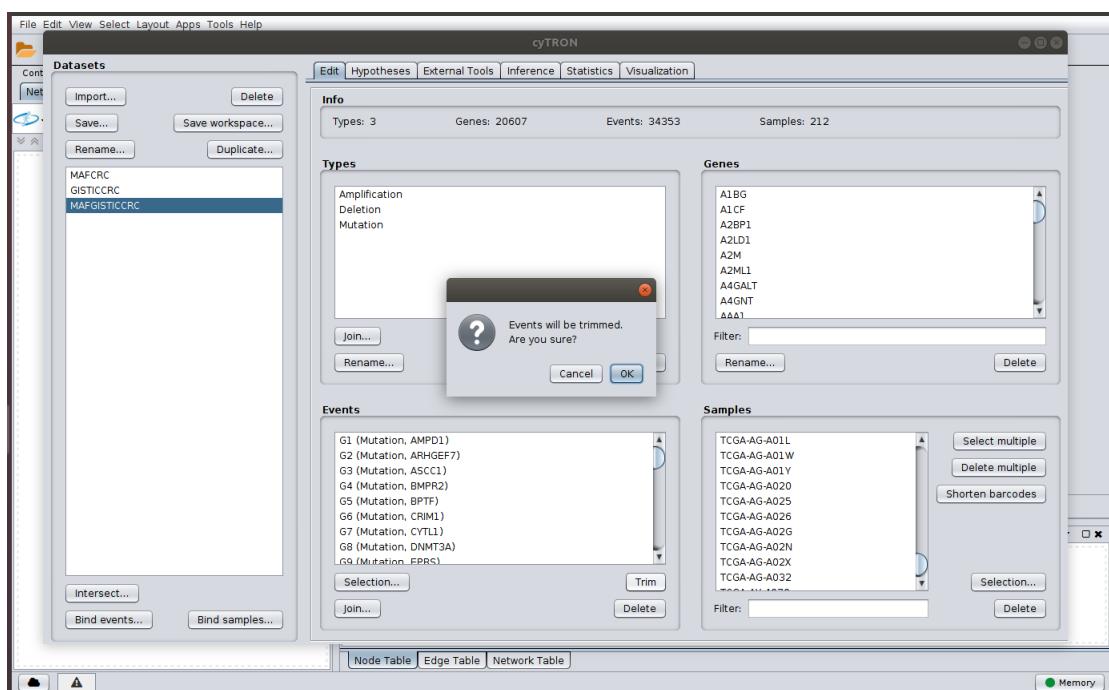
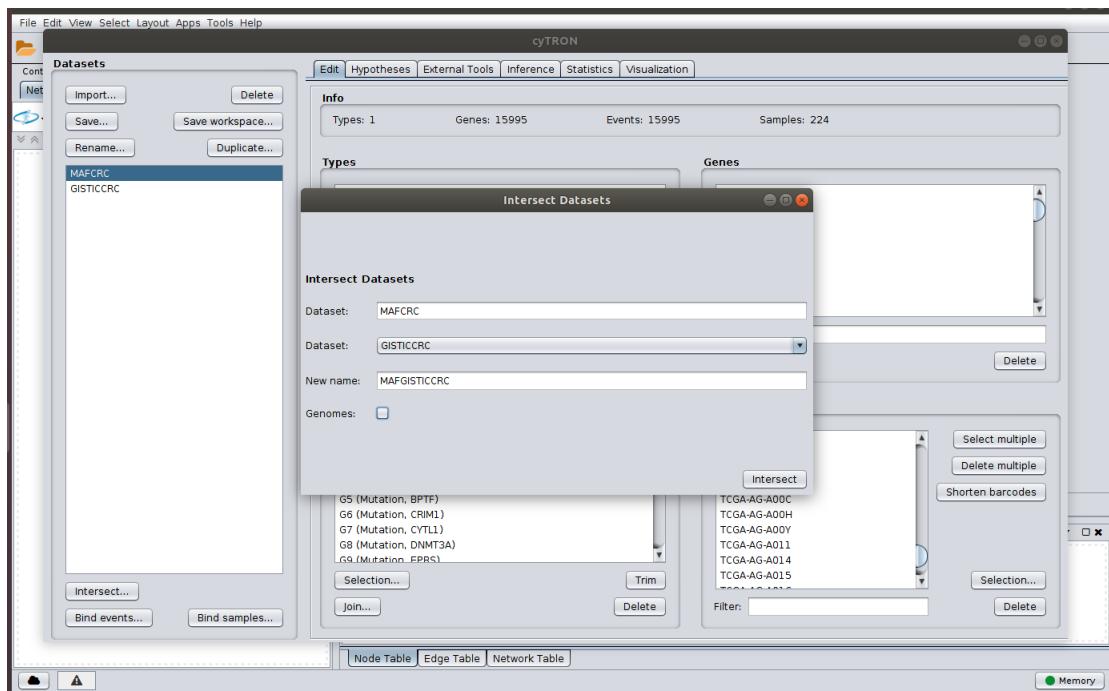
Importing and saving a GISTIC file: We now import copy number data from a gistic file. Gistic values provide 5 level for copy number: -2 and -1 being deletions, 0 being normal, and +1, +2 being gains at different intensities. See ftp://ftp.broadinstitute.org/pub/GISTIC2.0/GISTICDocumentation_standalone.htm for full documentation. We first import the Gistic file named crc_gistic.txt; then, we edit the data by removing +1 and -1 values using the section Types in the workspace and rename +2 and -2 events with the general name "Amplification" and "Deletion". We finally save the results to the file GISTICCCRC.rds.

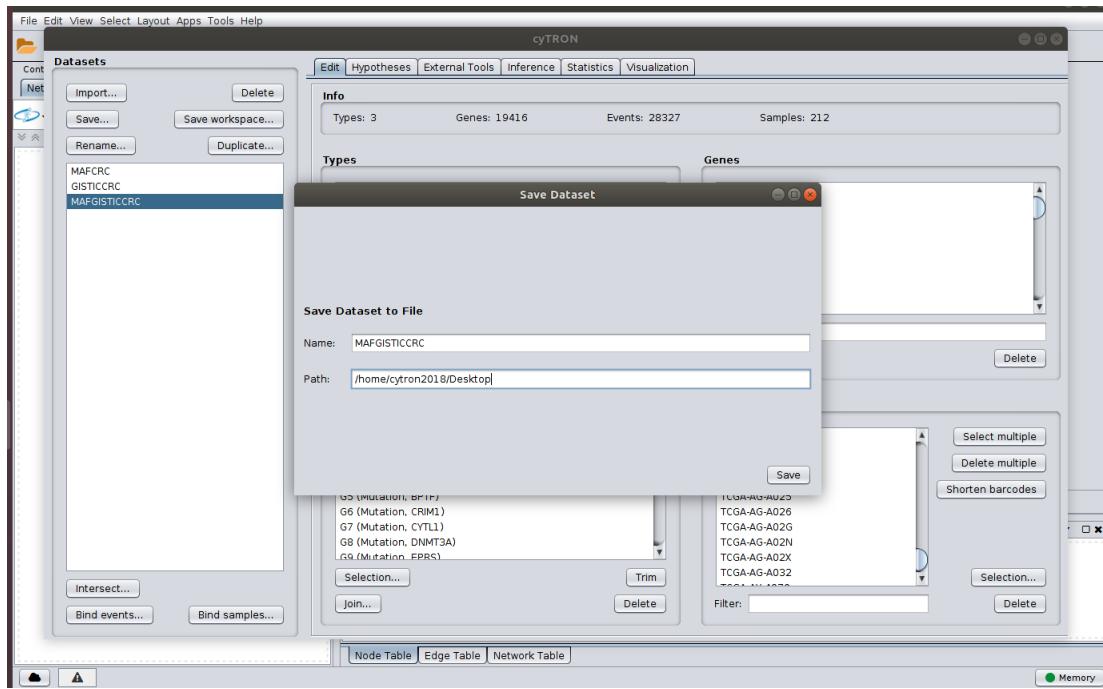




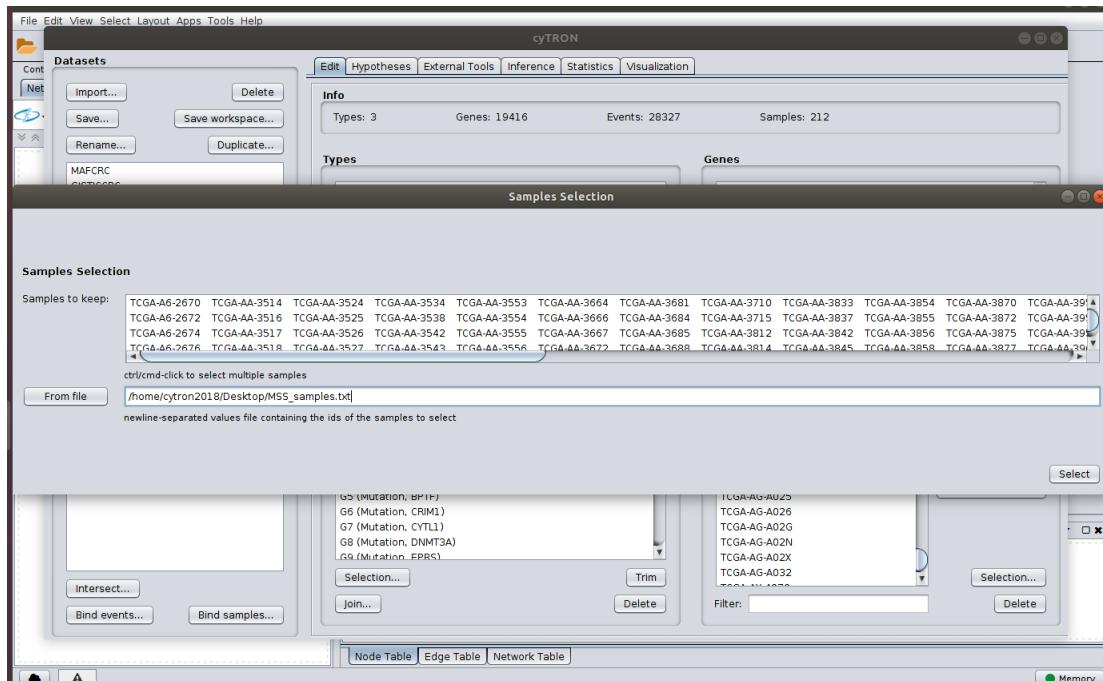


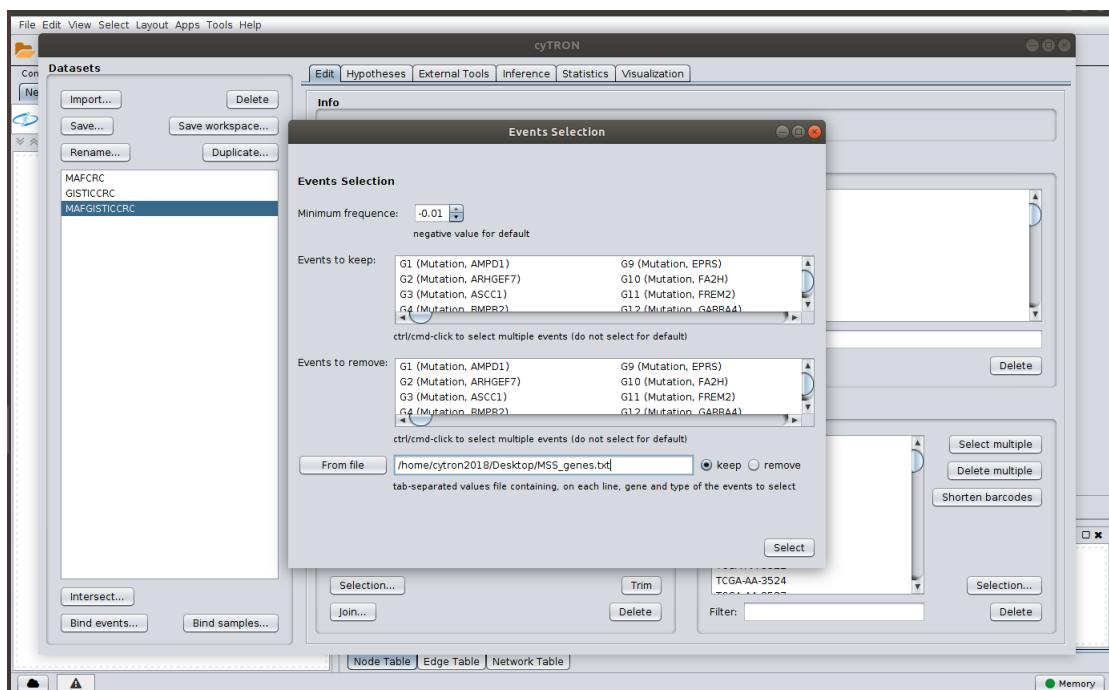
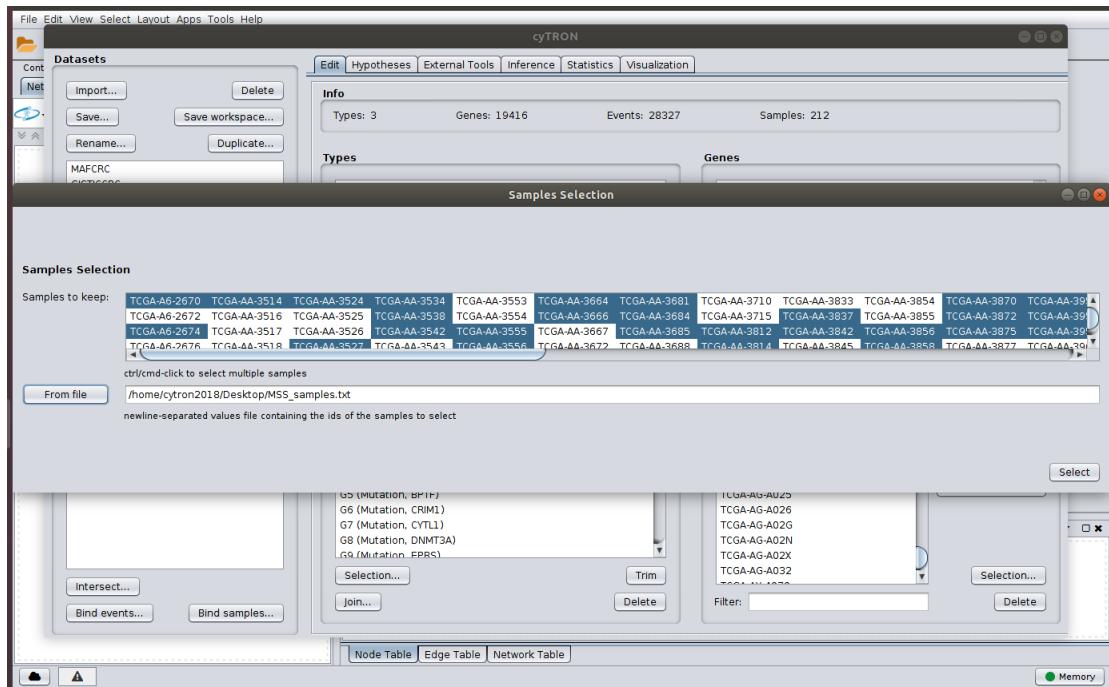
Merging datasets: next, we merge the data resulting from the previous steps by clicking the button “Intersect” on the bottom/left of the workspace. Then we remove genes without any alteration (button “Trim”) and finally save these results to the file MAFGISTICCRC.rds.

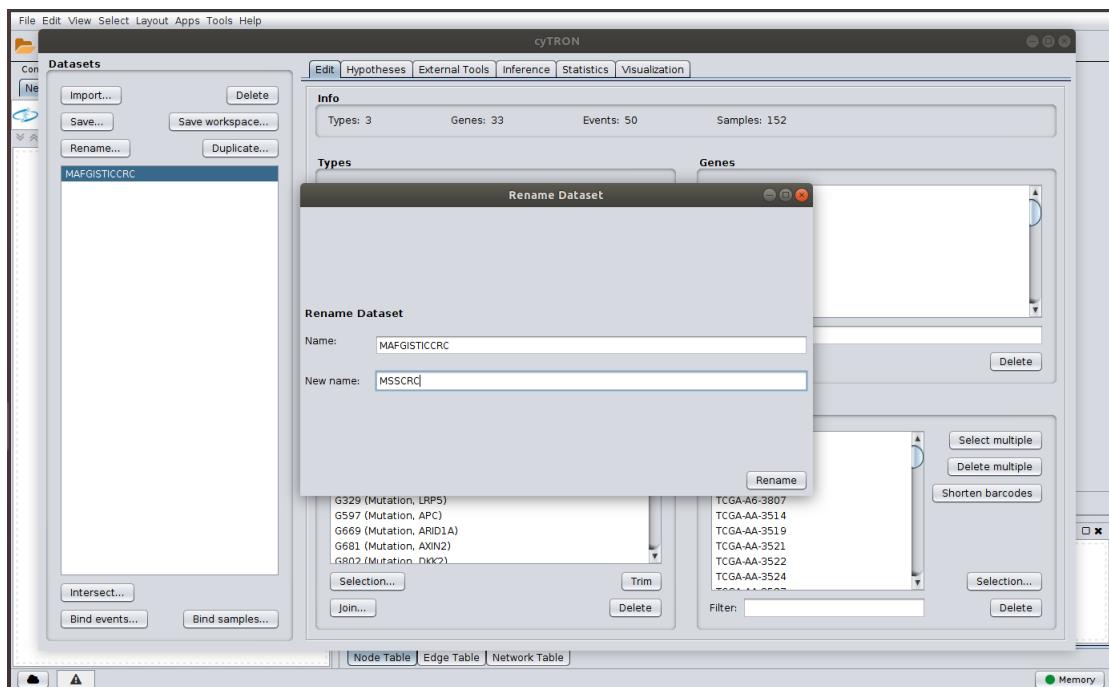
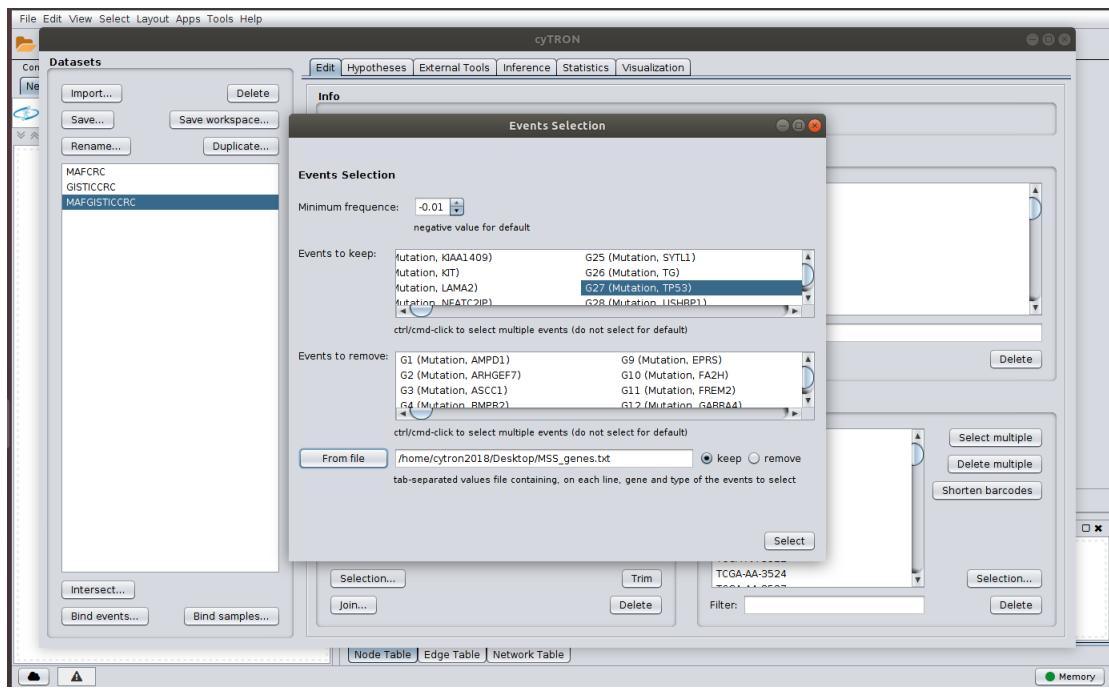


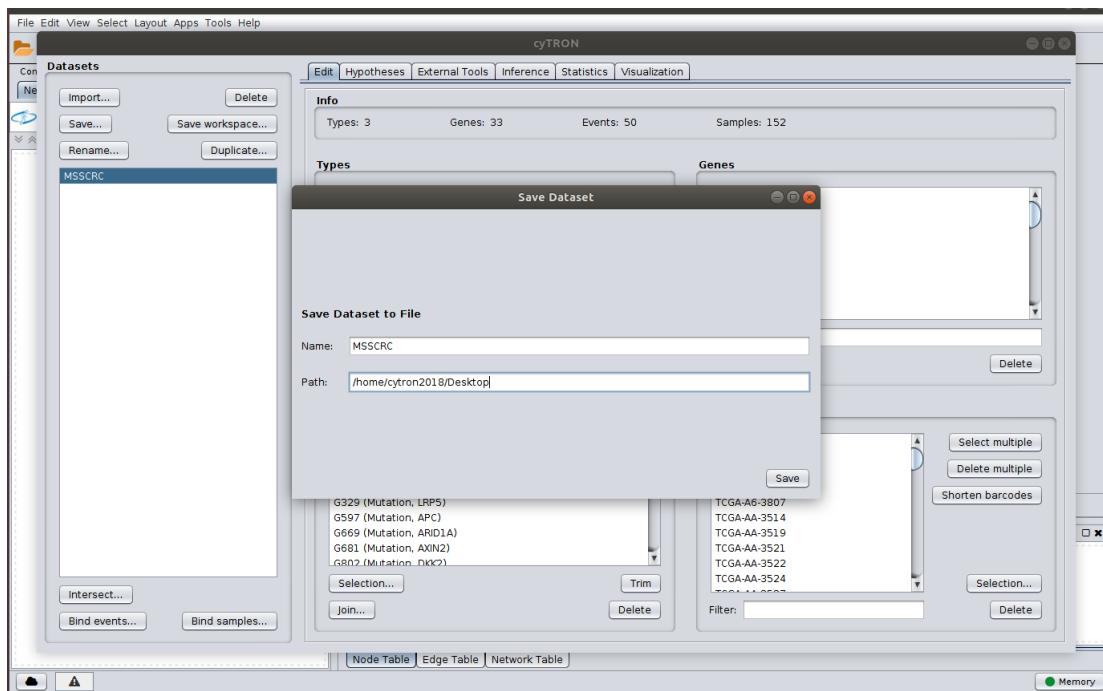


Selecting samples and alterations: in this step of our analysis, we select a subset of samples and alterations of interest within the dataset. We do this by selecting the samples from the file MSS_samples.txt and the alterations from the file MSS_genes.txt. We finally rename the obtained dataset and save it to the file MSSCRC.rds.

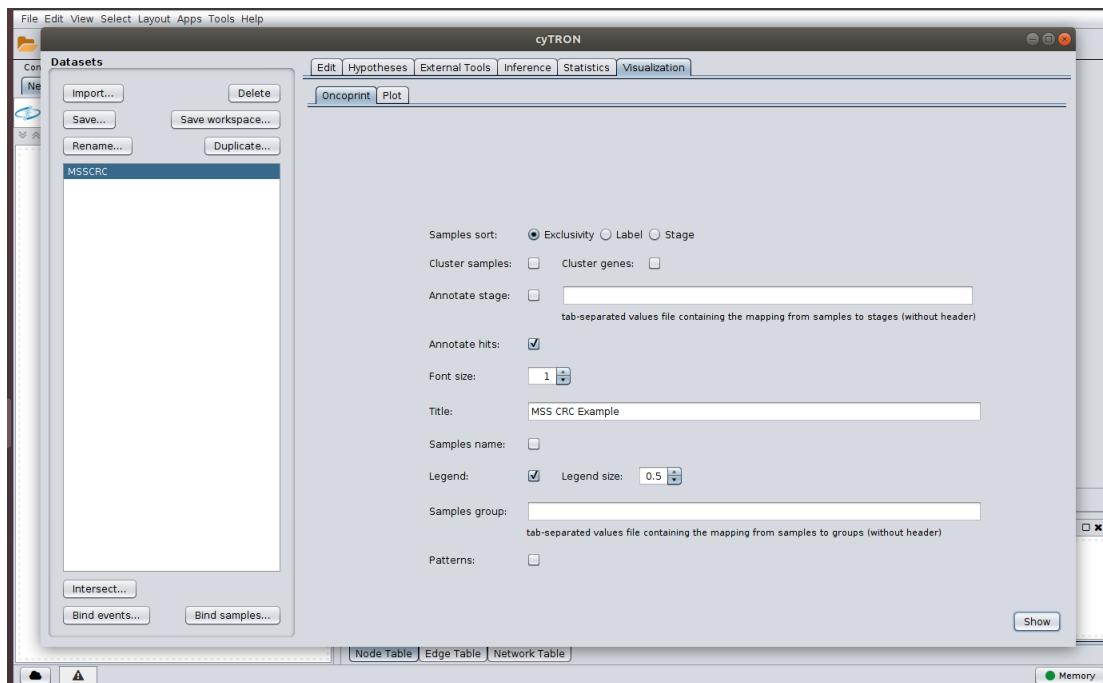


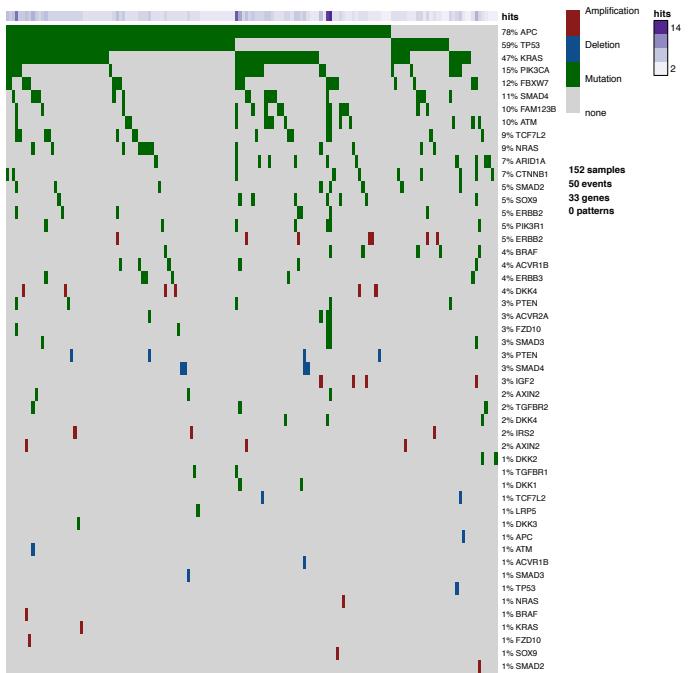




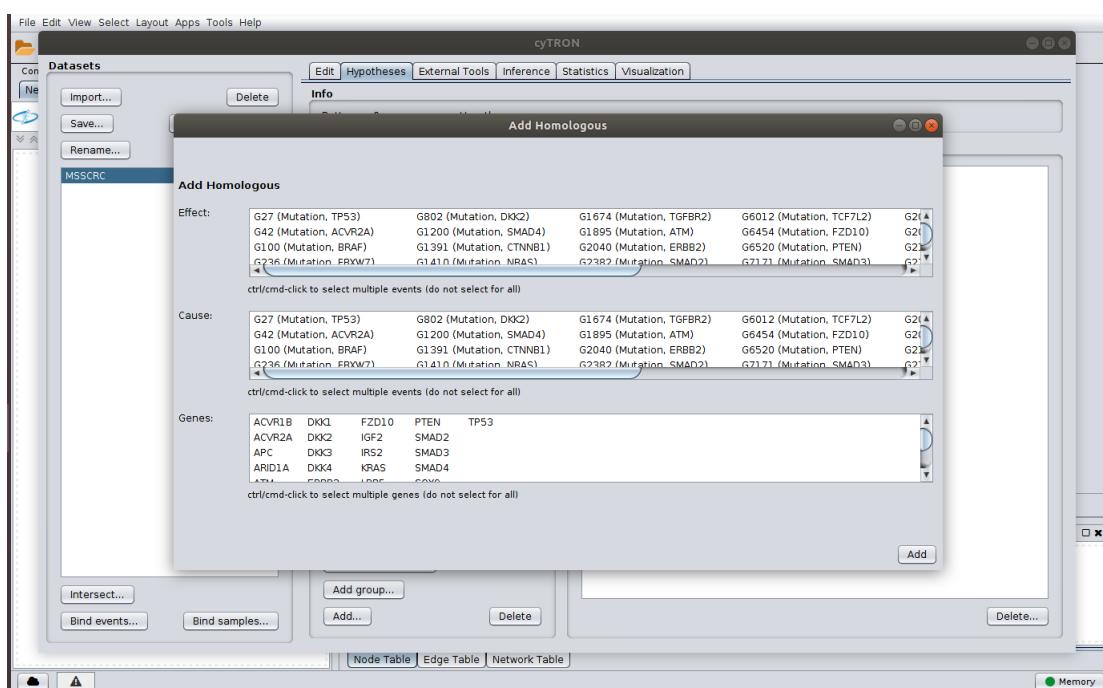


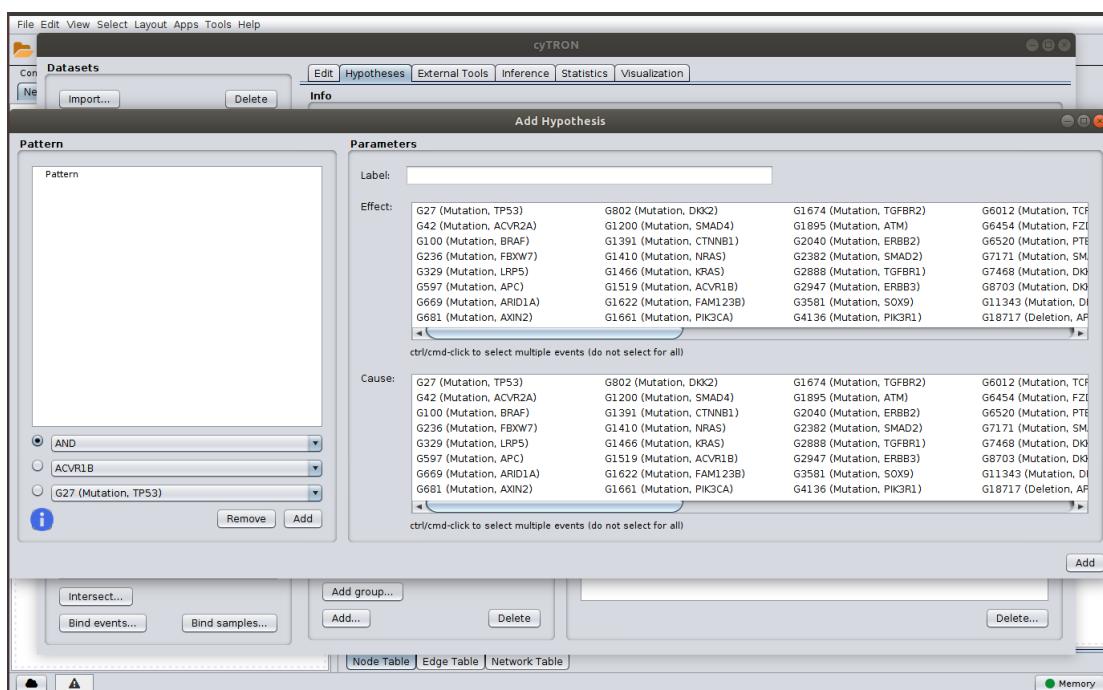
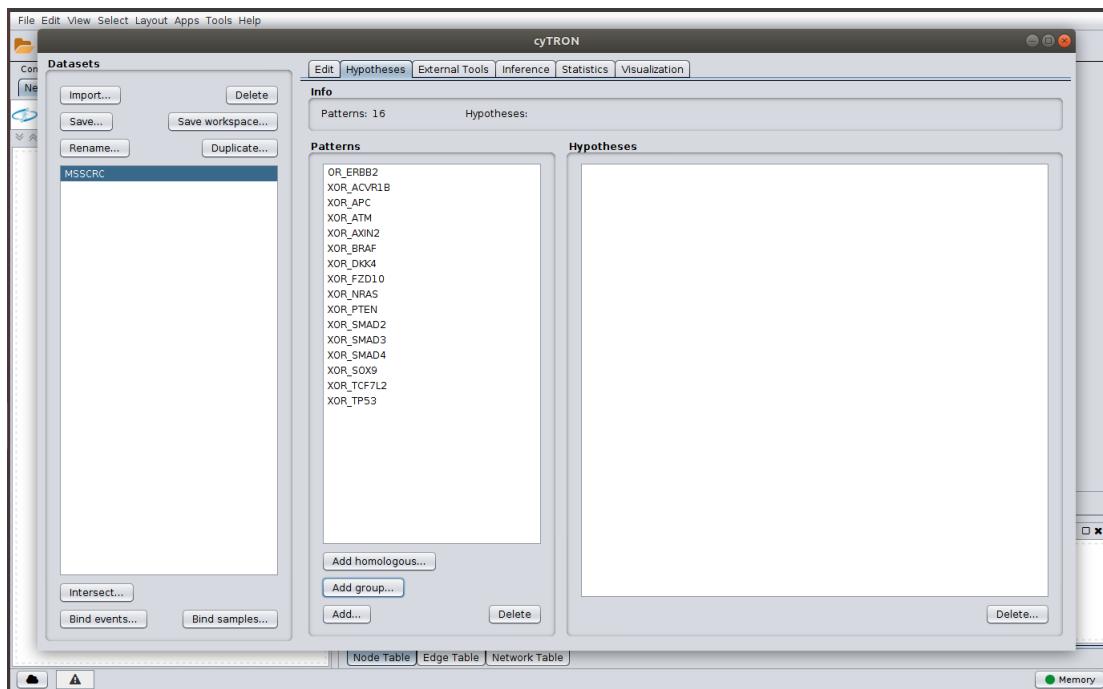
Plotting Oncoprints: at this point we move to the “Visualization” tab and plot an oncoprint of the MSSCRC dataset. Note that this operation requires the correct configuration of the Rscript command at the setup of cyTRON (see setup guide).

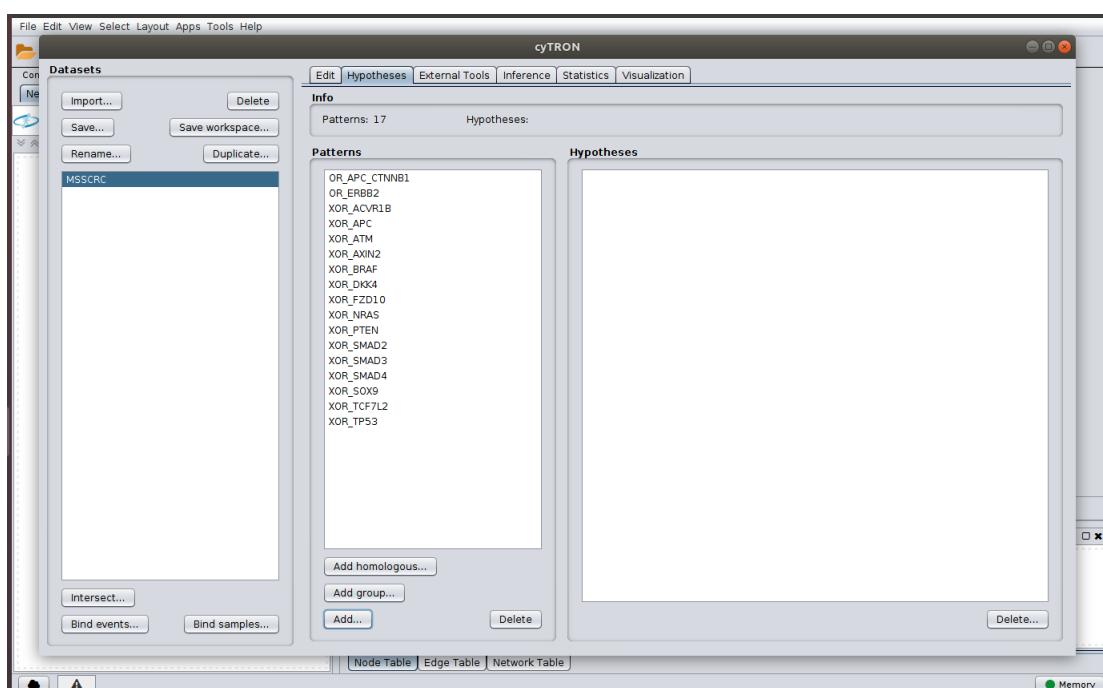
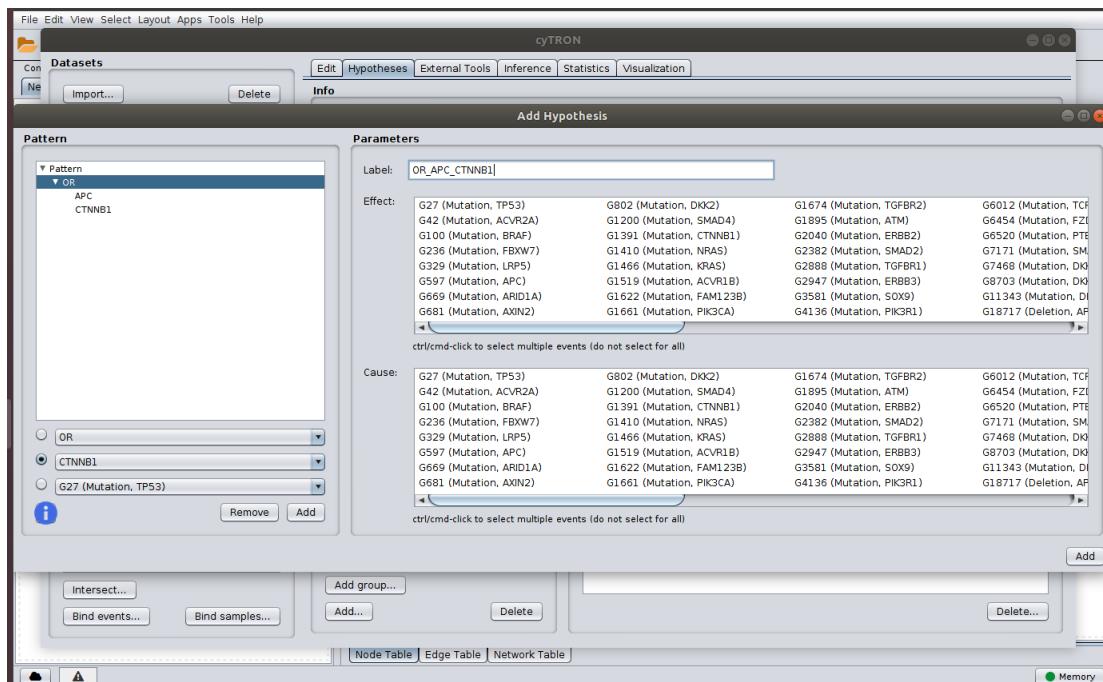




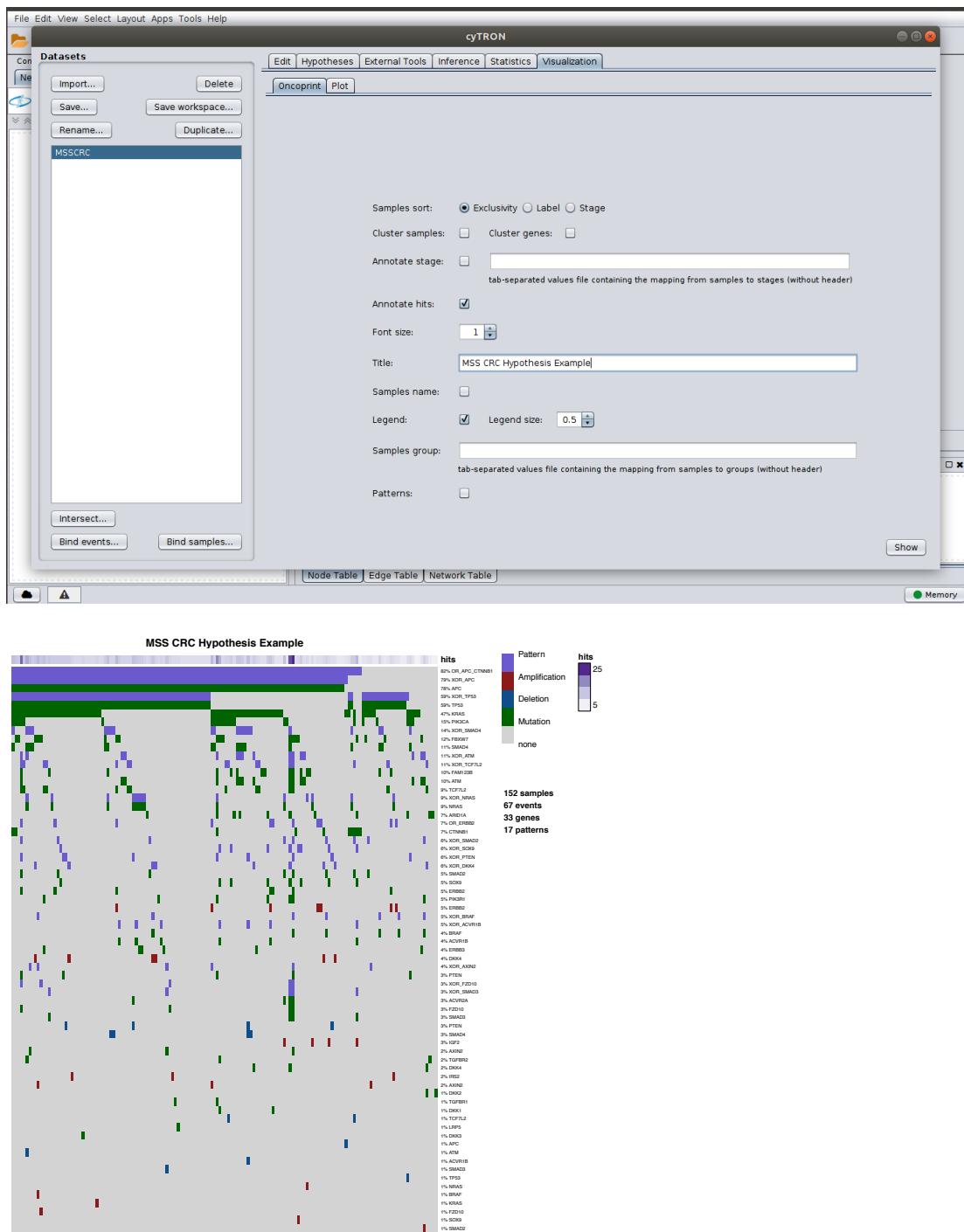
Adding hypotheses: we now proceed by adding hypotheses we want to test to the current dataset. We refer to [Caravagna, Giulio, et al. "Algorithmic methods to infer the evolutionary trajectories in cancer progression." Proceedings of the National Academy of Sciences 113.28 (2016): E4025-E4034] for details. Here, we first add homologous events and then a hypothesis involving APC and CTNNB1 alterations.



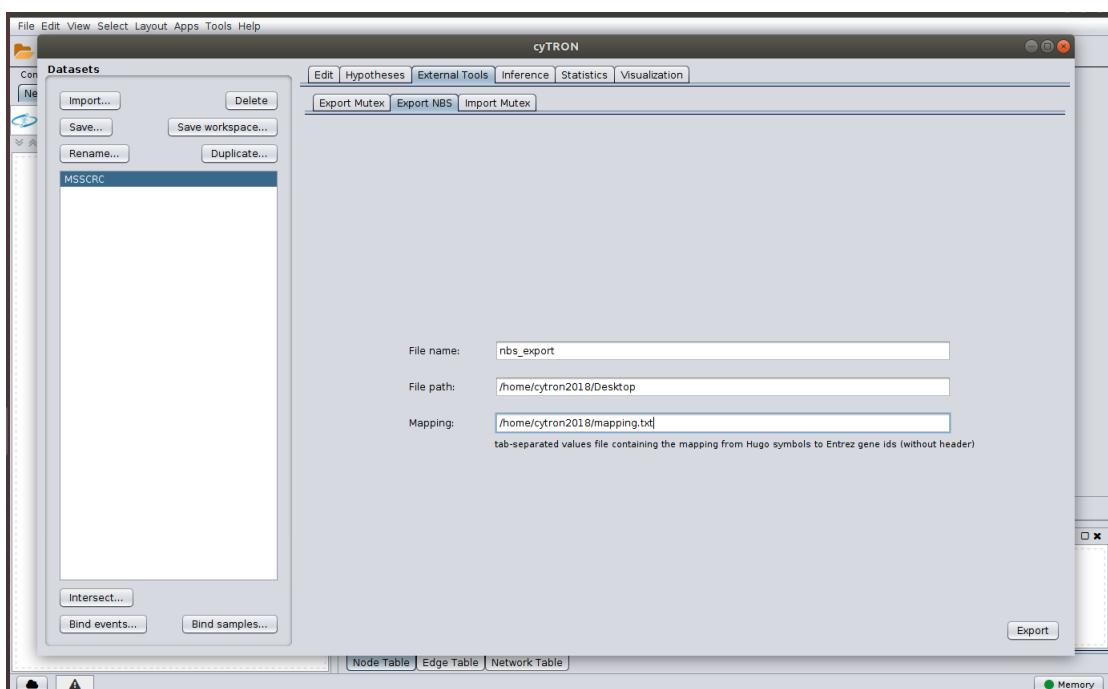
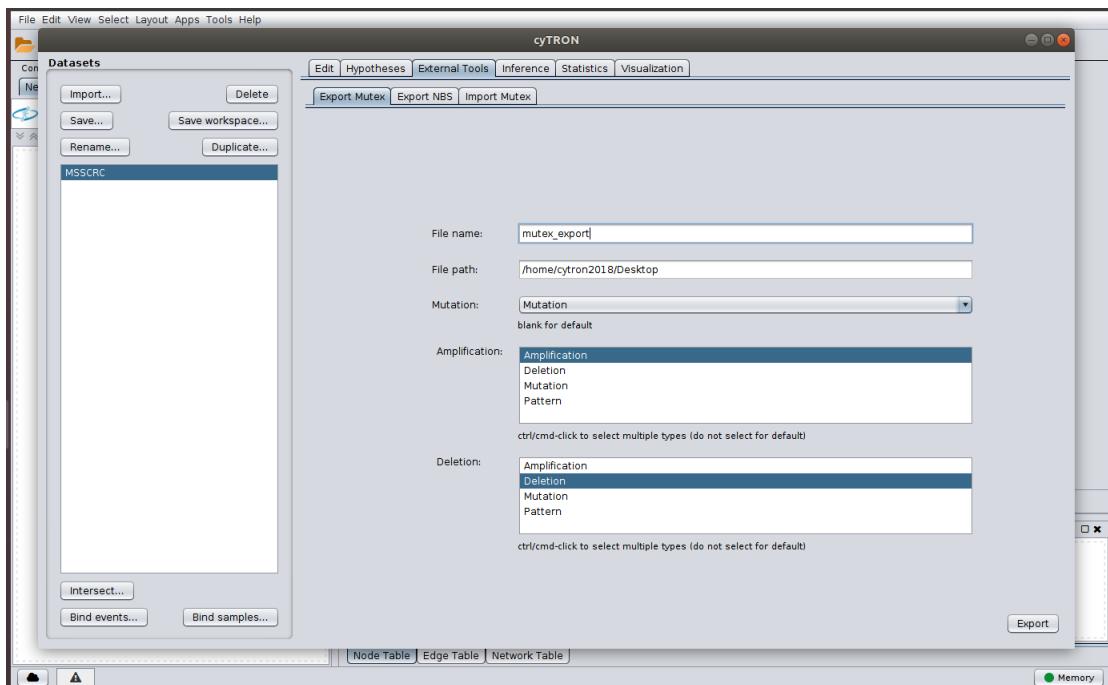


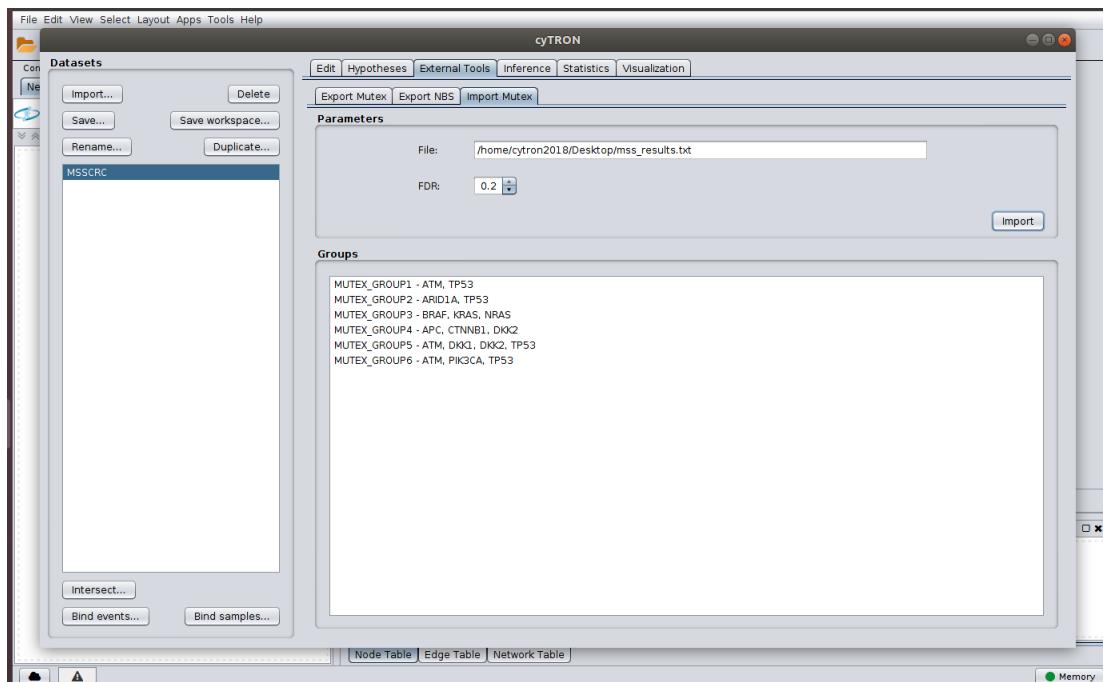


More on Plotting Oncoprints: we plot a second oncrint of this latest dataset, once again going to the tab "Visualization".

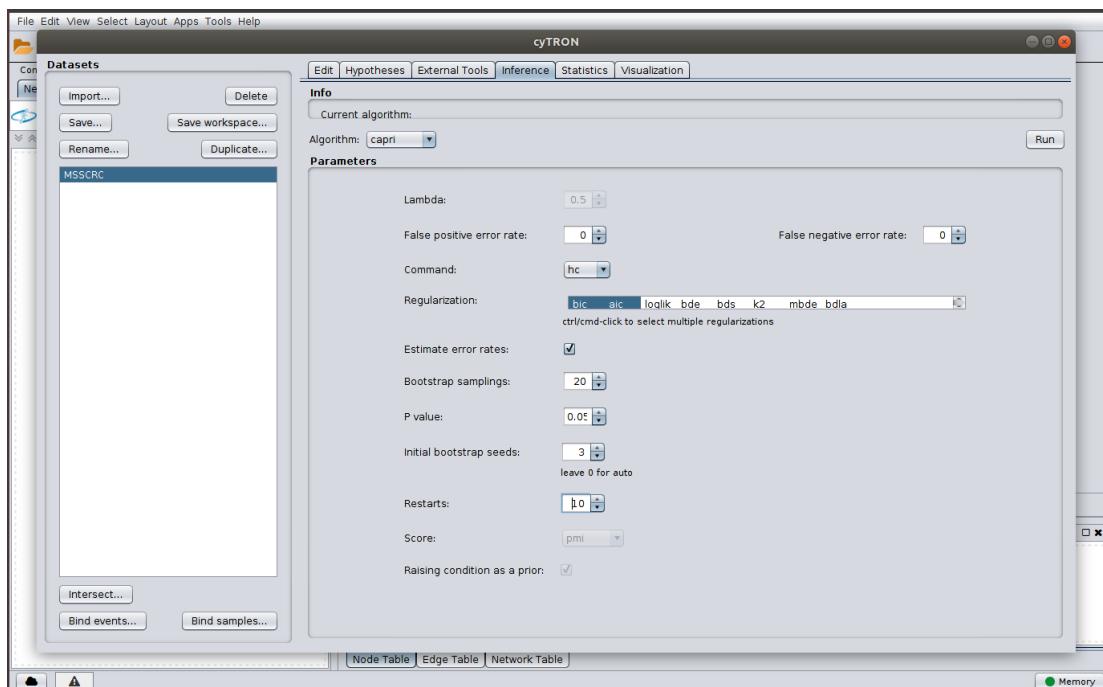


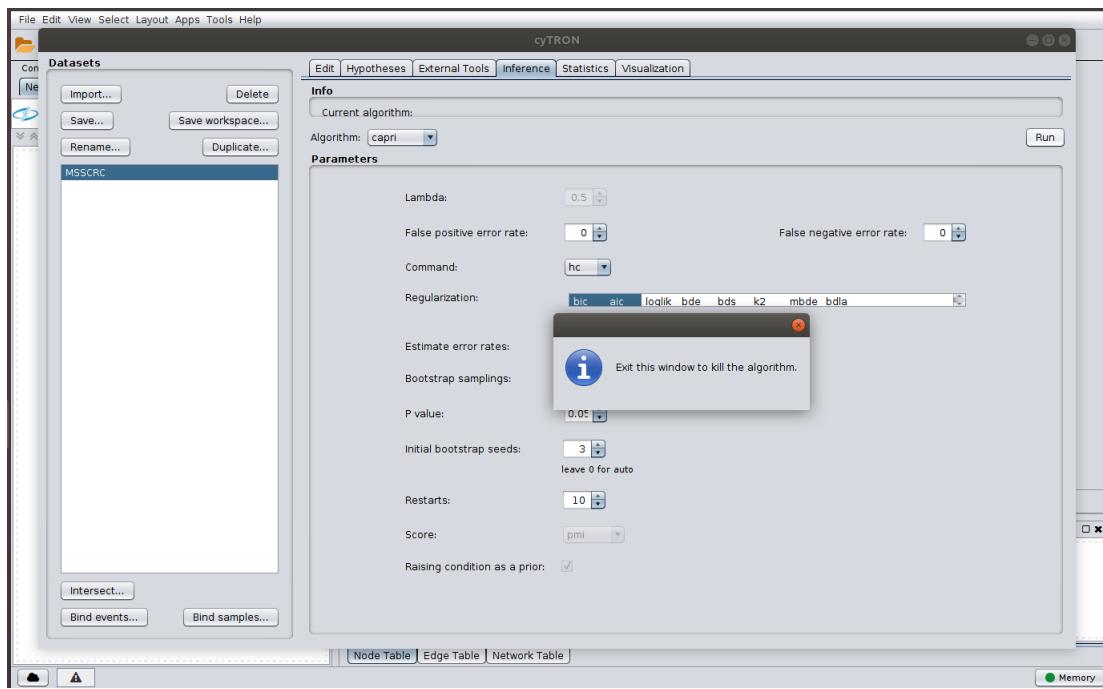
External tools: cyTRON also allows to interact with external tools (namely Mutex and NBS, see [Caravagna, Giulio, et al. "Algorithmic methods to infer the evolutionary trajectories in cancer progression." Proceedings of the National Academy of Sciences 113.28 (2016): E4025-E4034] for details) to build hypotheses or cluster the data. In the following three figures we show examples of such interactions.



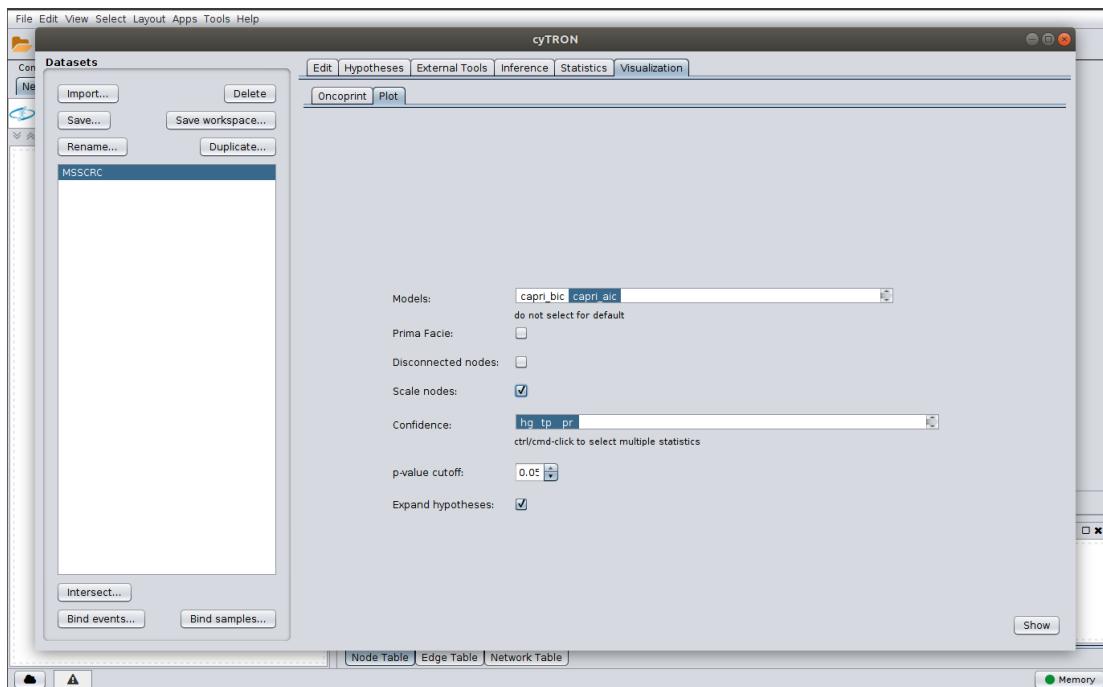


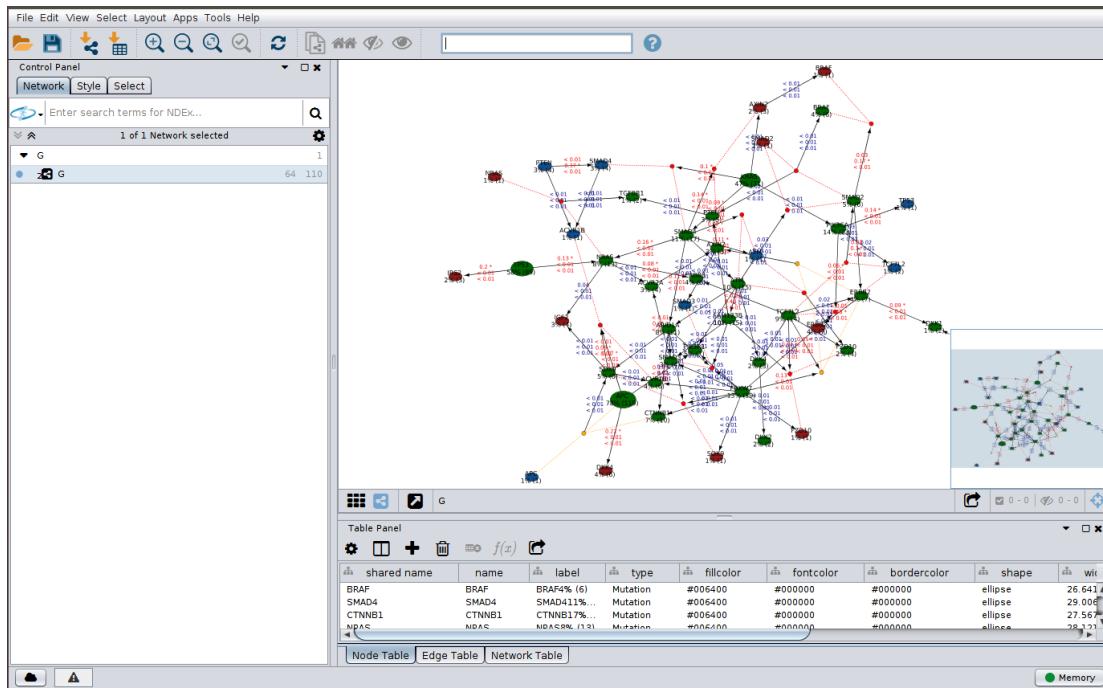
Inference with CAPRI algorithm: now that we have concluded the data preprocessing and visualization, we finally perform the inference by CAPRI algorithm by moving to the “Inference” tab and selecting capri and the desired parameters.



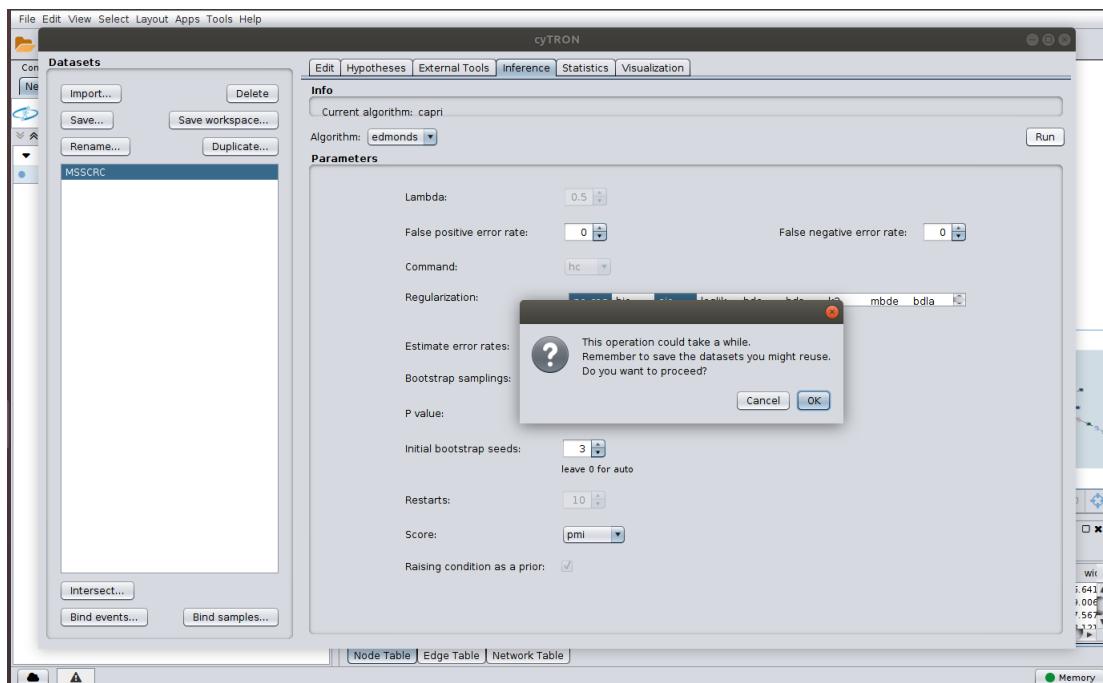


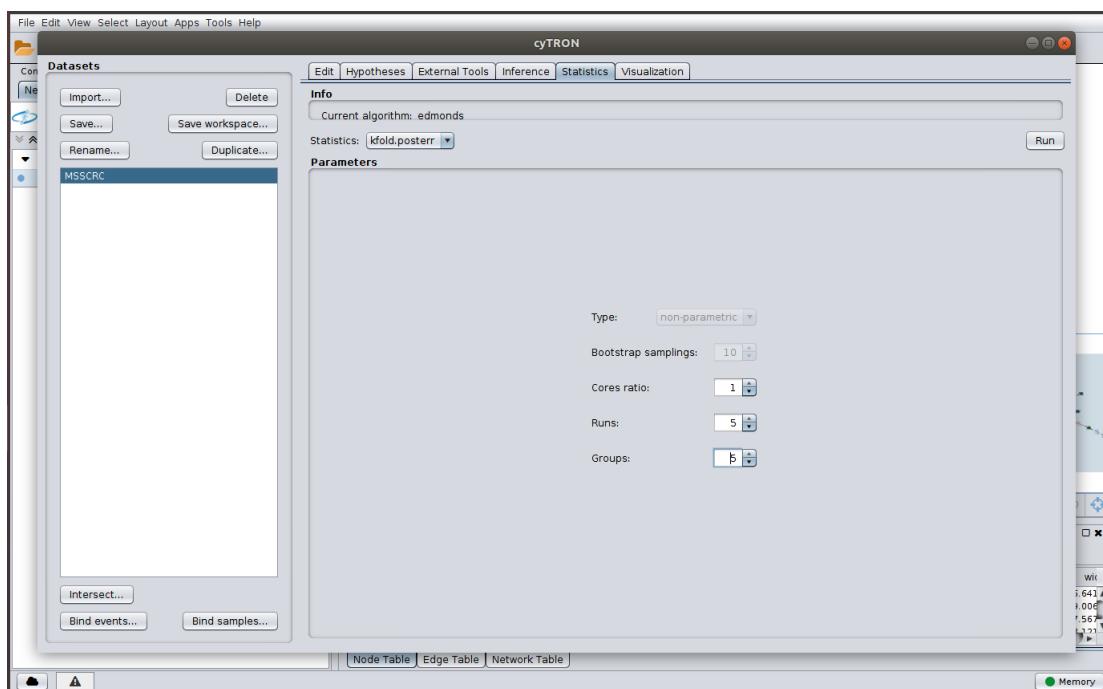
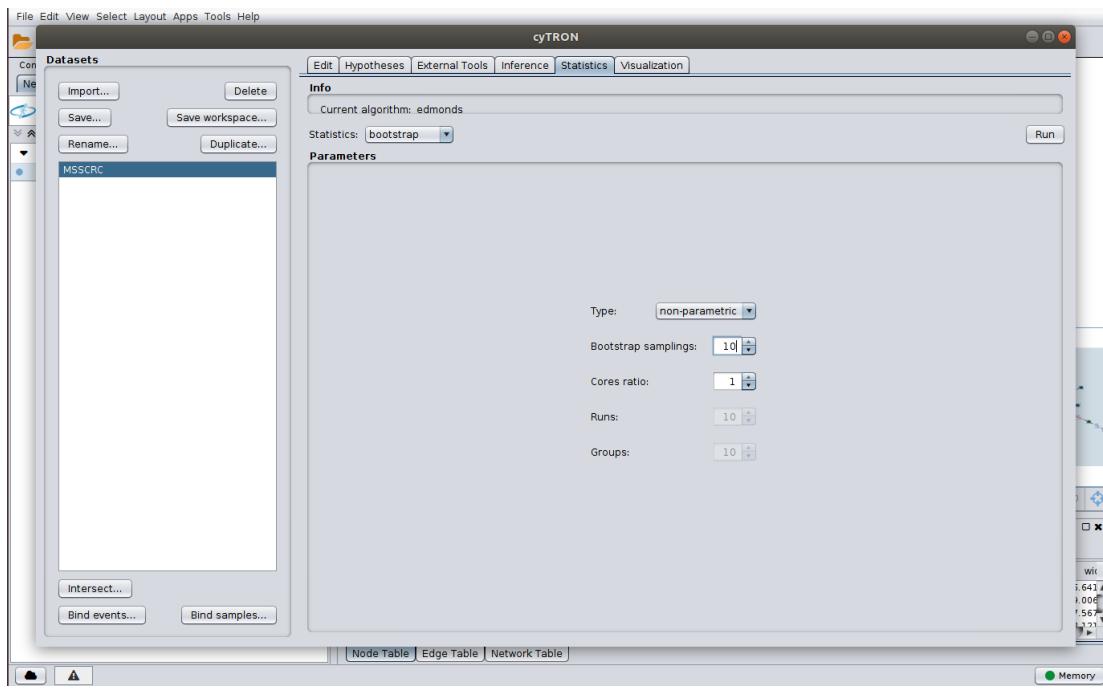
Visualization in Cytoscape: once the inference is concluded we return to the “Visualization” tab and plot the resulting models in order to visualize them in Cytoscape. This results in a Cytoscape fully customizable and interactive graph.





Inference with Edmonds algorithm and statistical evaluation: we repeat the inference but this time we select Edmonds algorithm. This time, we also evaluate the statistical robustness of the resulting models by going to the “Statistics” tab and performing bootstrap and cross-validation assessments. Once again we refer to [Caravagna, Giulio, et al. "Algorithmic methods to infer the evolutionary trajectories in cancer progression." Proceedings of the National Academy of Sciences 113.28 (2016): E4025-E4034] for more details.





Visualization of models and statistics: we conclude this guide by visualizing in Cytoscape the computed Edmonds model and its statistical evaluation.

