

CSI Results Visualisation

A results visualisation tool has been developed in order to help you explore some common features output from the CSI model. Roughly, the tool allows you to visualise the resulting marginal and MAP networks, applying thresholds as appropriate. One can also view the model predictions while comparing them to the input data.

The main display is broken into three main sections; the top right of the page displays the network, the top left shows a table of genes and various attributes associated with them, at the bottom of the page a series of plots show a selected gene and its parents. The menu-bar at the top of the page allows some other operations to be performed.

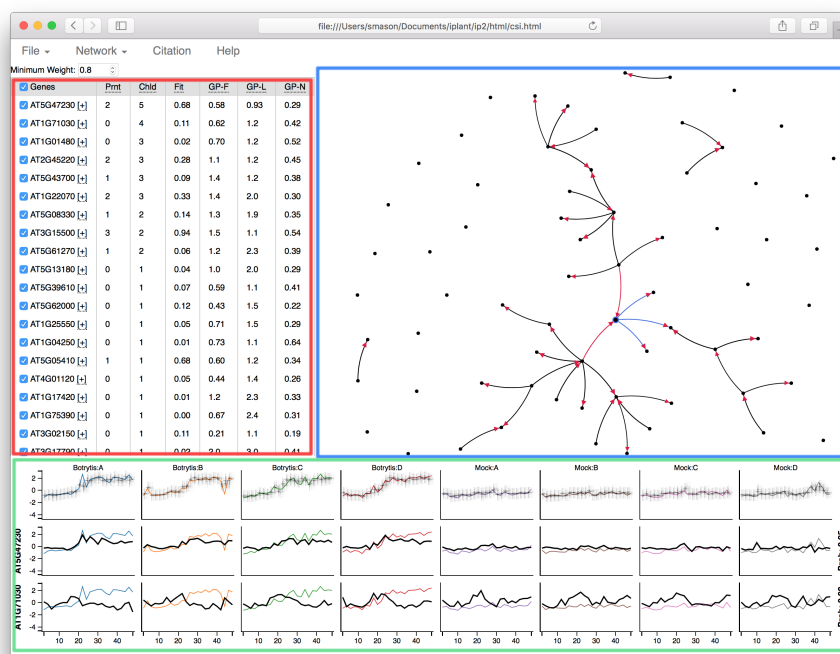


Figure 1: CSI results visualisation with main areas of the display highlighted

Network Graph

The top right of the page (highlighted blue in the figure) displays a directed graph showing all the selected genes with arrows pointing from “parental” genes to their targets. The plot can be dragged after clicking with the mouse and

zoomed by “scrolling”. Hovering over a node displays its name, while clicking on it allows it to be rearranged within the graph.

The actual nodes and edges displayed within the graph is a representation of the graph with the current options applied, with a number of controls in other sections controlling what’s displayed.

Table of Genes

The table in the top right of the page (highlighted red in the figure) displays a list of genes and various attributes associated with them. The checkbox shows which genes are selected to be displayed within the network graph and can be toggled individually or across all genes by clicking on different checkboxes. The other columns show results from the model that are helpful in extracting information from its fit.

All columns can be sorted by clicking on their title, clicking on the same column a second time will reset the sort. Sorting by any column and then inspecting the top and bottom genes can be informative, either of biologically relevant information or of the model misinterpreting the input data and further preprocessing.

Number of Parents and Children

The columns titled **Prnt** shows the number of parents (*indegree*) inferred for this gene, and **Chld** the number of genes who consider this gene to be their parent (*outdegree*). These values are effected by the current network type as well as weight threshold.

Hovering over one of these values highlights the parents or children as appropriate in the plot, while clicking shows/hides the those items in the plot.

Fit of Best Parental Set

The **Fit** column shows the weight associated with the best parental set and is useful to find genes where the model is able to accurately predict expression. This column can be sorted by clicking on the title, and examining genes with the best fits in the expression plots can be useful.

Inferred Gaussian-Process Parameters

The columns **GP-F**, **GP-L** and **GP-N** show the hyper-parameters associated with the model fits for a given gene. Clicking on the titles to sort them and then examining the outliers (i.e. those genes with the largest and smallest values) can help indicate genes that the model isn’t fitting.

The actual values will be dependent on the data you have analysed and it is difficult to make suggestions on the values.

Expression Plots

The bottom of the page (indicated by green in the figure) is dedicated to plots of the raw data and model predictions. There is a column for each replicate/condition included in the input data. The first row displays the data for the selected gene along with estimates from the models. Only the best models are included, as selected when the data was converted from HDF5 to JSON—when run from within iPlant this a weight cutoff of 0.01, i.e. approximately those that would be visible. The estimates from the model predict expression at a given time point given the gene and its parents' expression at the previous time. The display highlights the “better” models by making the line width proportional to the model's weight.

Below the gene's expression are plots of the parents, labelled on the left with the parent's name and on the right with the parent's marginal-likelihood. The coloured lines are those of the selected gene and the black lines the parents', with the line's width proportional to the marginal-likelihood.

Manual Result Extraction

In order to open CsiVis the HDF5 results from CSI need to be transformed into a JSON file. This process happens automatically within the iPlant App, but can be triggered manually by running a command like:

```
python csi-postprocess.py output.h5 -v > output.json
```

This command executes the Python script `csi-postprocess.py`, telling it to take the complete results from the HDF5 formatted file `output.h5`, extract the *better* models and predictions and write them out to `output.json`. The `-v` option causes output to be *verbose*, that is it display some progress messages.

A few options exist to control this extraction process, and these can be seen by running:

```
python csi-postprocess.py --help
```

A more complete set of results can be viewed by running:

```
python csi-postprocess.py -v output.h5 \  
    --weight 1e-6 --predict 1e-3 > output.json
```

will cause models with even lower probabilities to be written to the JSON file, as well as more predictions. The trade off is larger JSON file size which can cause issues when loading CsiVis in a web browser.