

Data Analysis of High Content Screens

Outline

- cellHTS2
- biomaRt
- ranks
- visualisation
- hierarchical clustering
- checking original images
- (Birmingham paper comparison)

cellHTS2

An R library for analysis of cell based RNA interference screens, part of BioConductor project.

We have modified a bit to add more functionality (non-robust scoring methods, multiparametric analysis), and to integrate better to KNIME.

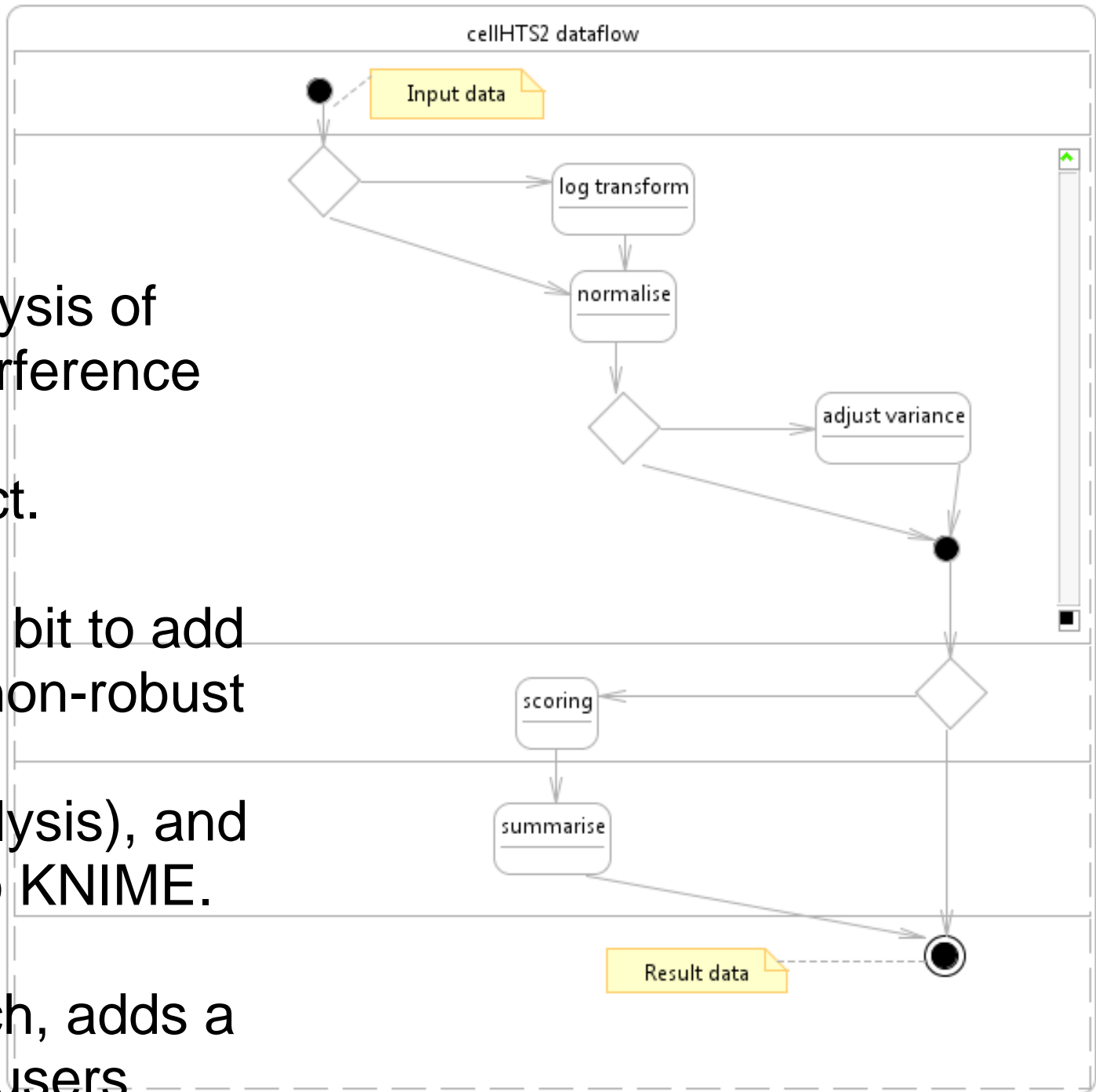
It is really feature rich, adds a lot of options to the users.

cellHTS2

An R library for analysis of cell based RNA interference screens, part of the BioConductor project.

We have modified a bit to add more functionality (non-robust scoring methods, multiparametric analysis), and to integrate better to KNIME.

It is really feature rich, adds a lot of options to the users.



biomaRt

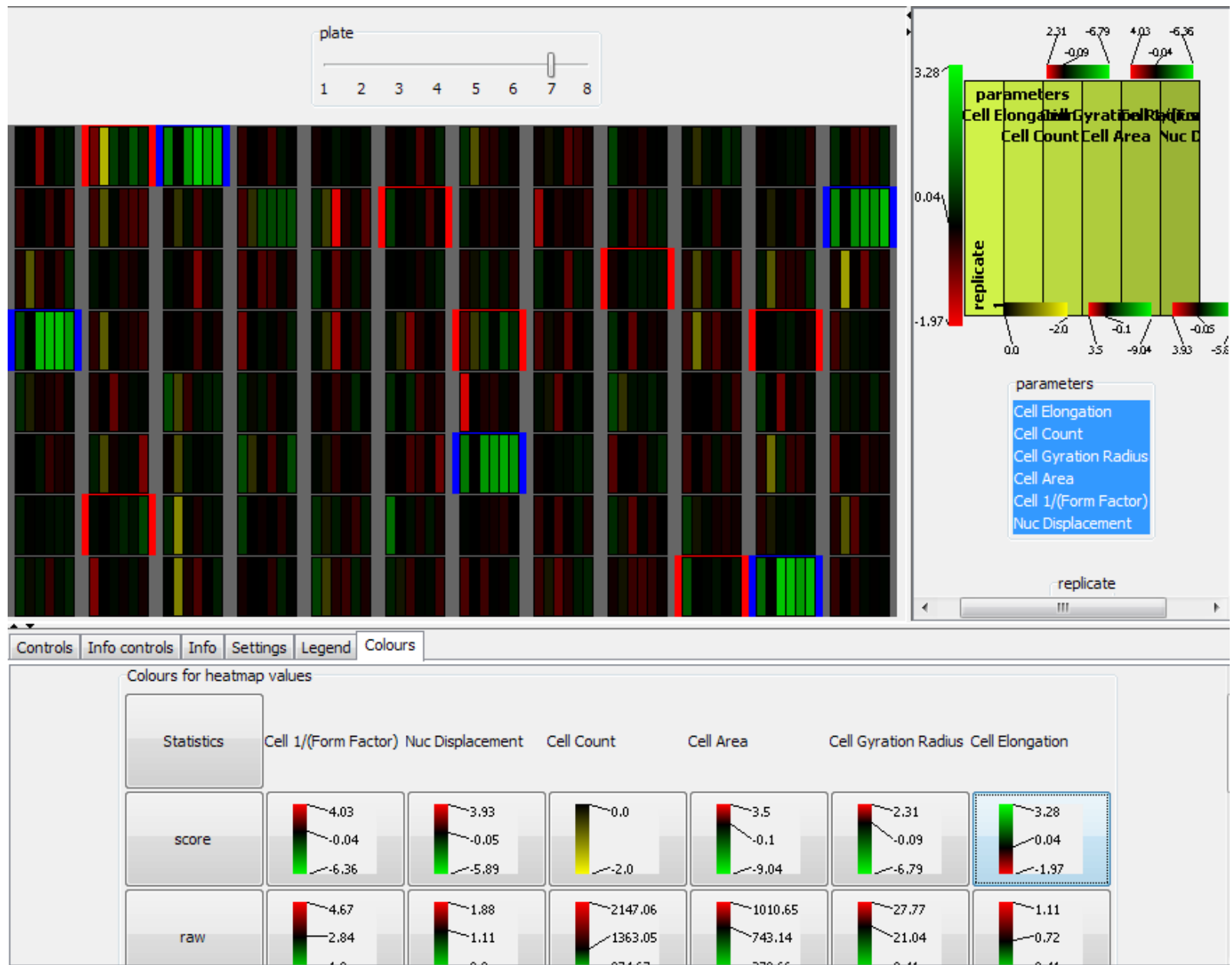
Also an R library from the BioConductor project.

Its aim is to call web services and add information to the screen data.

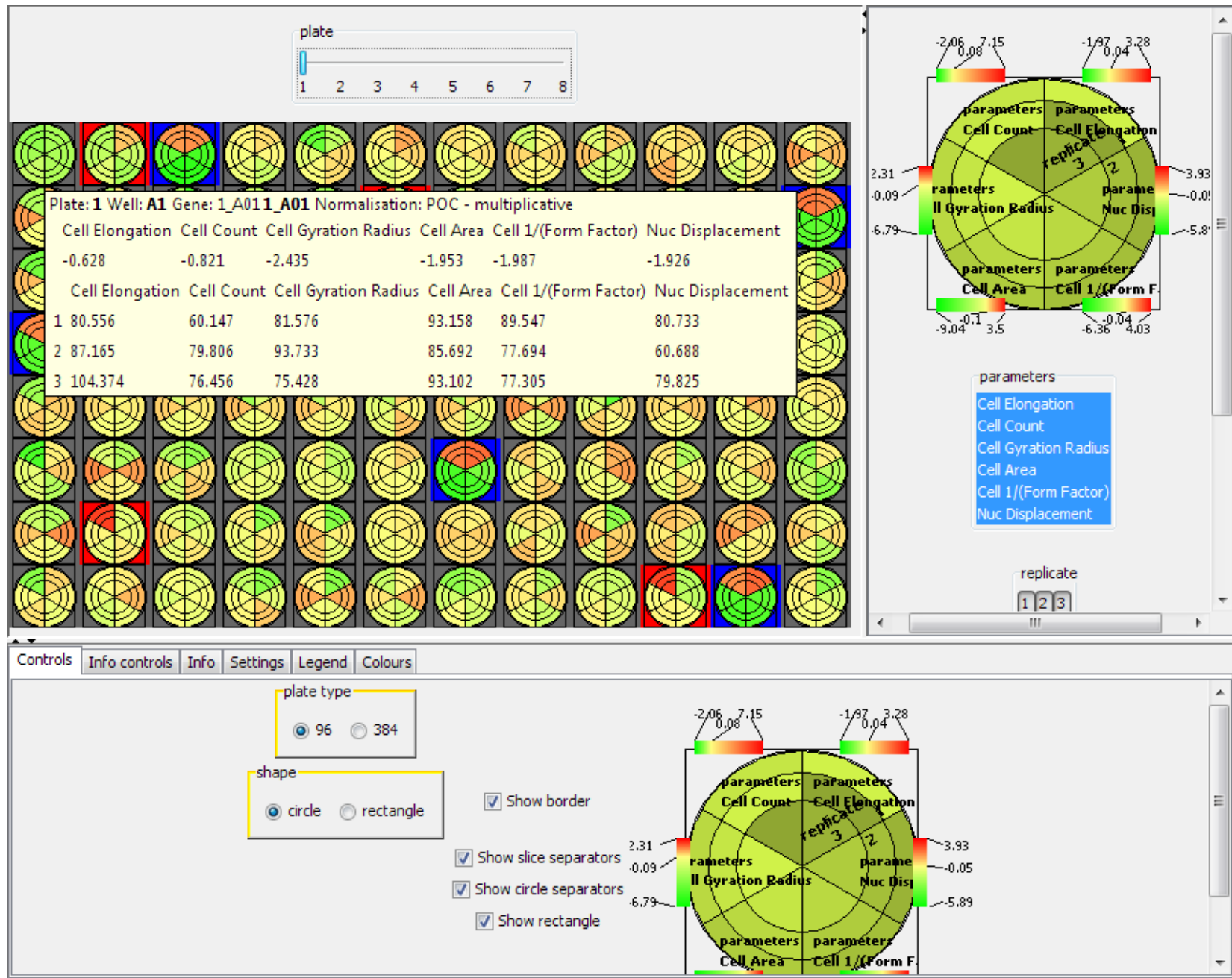
Ranks

Adding ranks can be easy with **Sorter** and **Java Snippet** nodes.

But it becomes tricky when you want to handle differently the controls and your samples, so we created the **Rank** node.



Visualisation - Adjust colours for different parameters



Visualisation - details on demand

Hierarchical Clustering

We added some nodes to the basic hierarchical clustering:

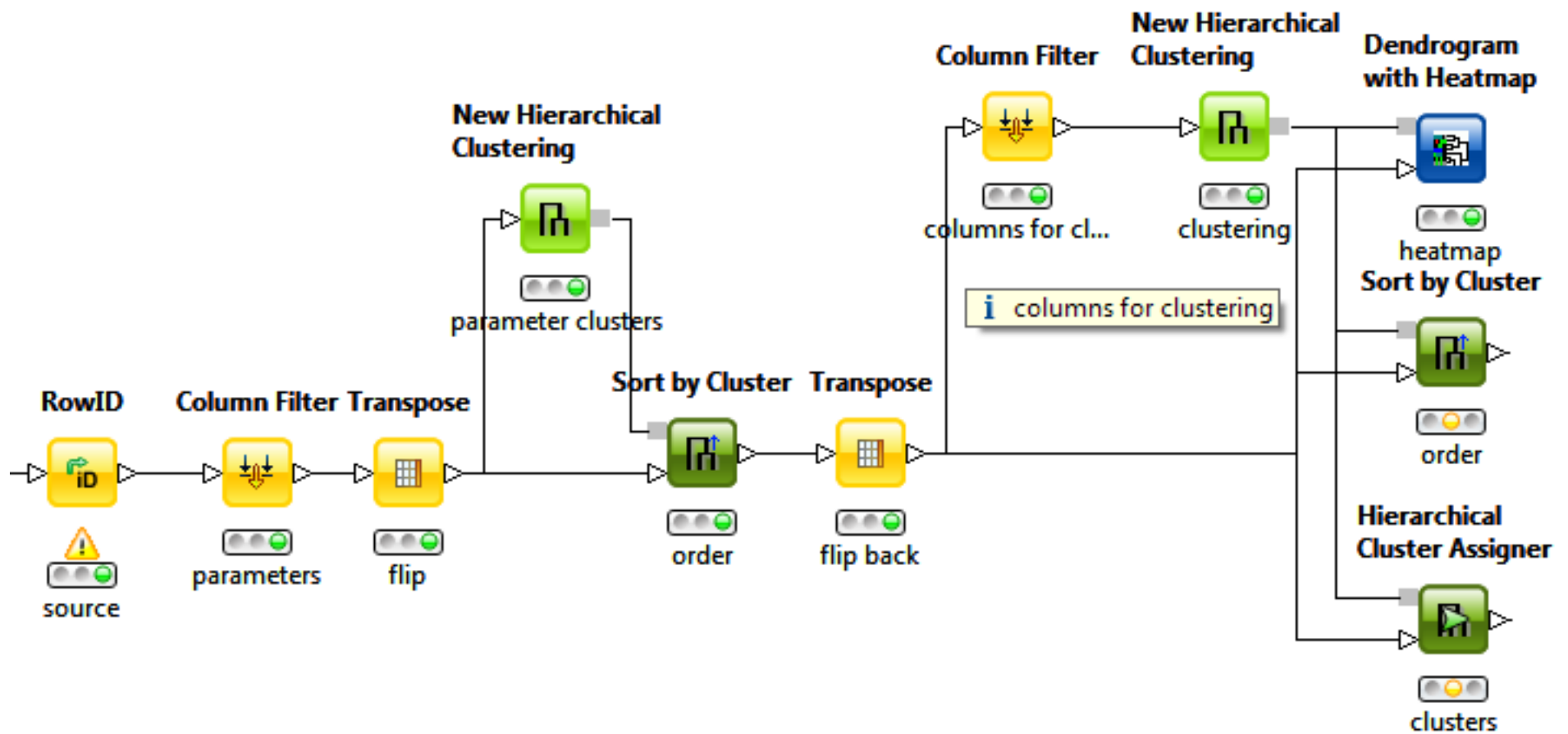
- Leaf Ordering
- Reverse Order
- Sort by Cluster
- Dendrogram with Heatmap

Leaf Ordering makes the clustering more robust, deterministic.

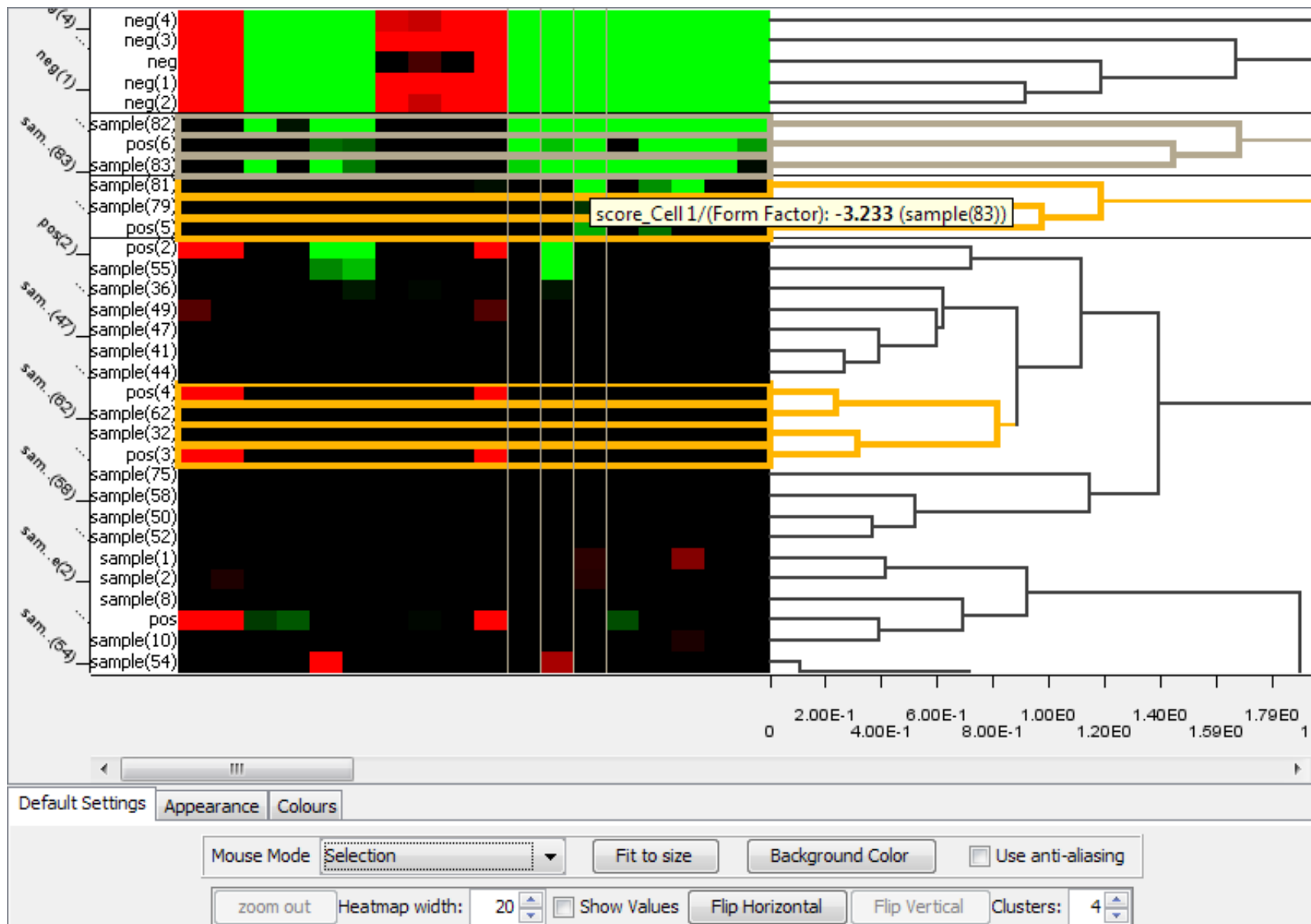
Reverse order is just for esthetic changes.

Sort by Cluster sorts the rows according to the hierarchical clustering.

Dendrogram with Heatmap gives a nice view of the clustering.



Clustering by parameters and then by values



Hierarchically clustered data with heatmap

Analysis of HTS - Statistical methods (1)

data triage

Supported by HiTS	Not supported by HiTS
plate visualisation scatter plots replicate correlation plots heatmaps	

normalisation

Supported by HiTS	Not supported by HiTS
percent of control percent of samples robust z score B score local regression	

Amanda Birmingham, et al.:

Statistical methods for analysis of high-throughput RNA interference screens, *Nature Methods*, 2009

Analysis of HTS - Statistical methods (2)

quality metrics

Supported by HiTS	Not supported by HiTS
Z' factor correlations Q-Q plot	SSMD ROC

hit identification

Supported by HiTS	Not supported by HiTS
mean +/- sd median +/- mad quartile based selection simple rank based	SSMD t-test redundant siRNA activity rank product Bayesian model

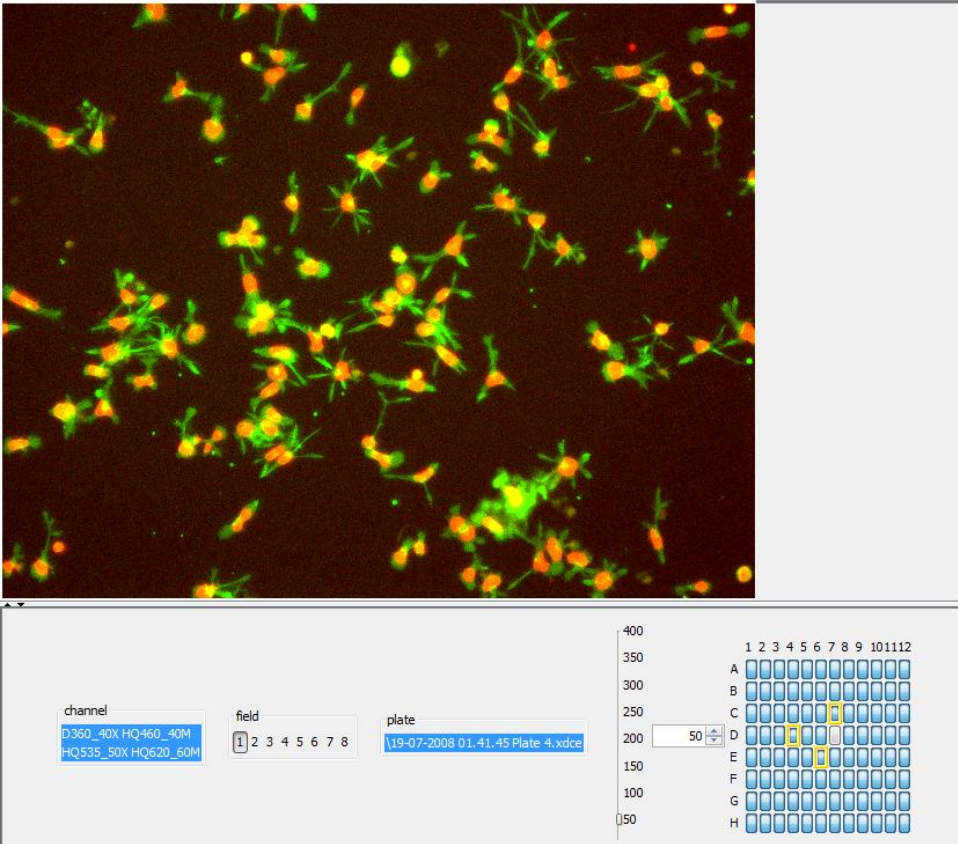
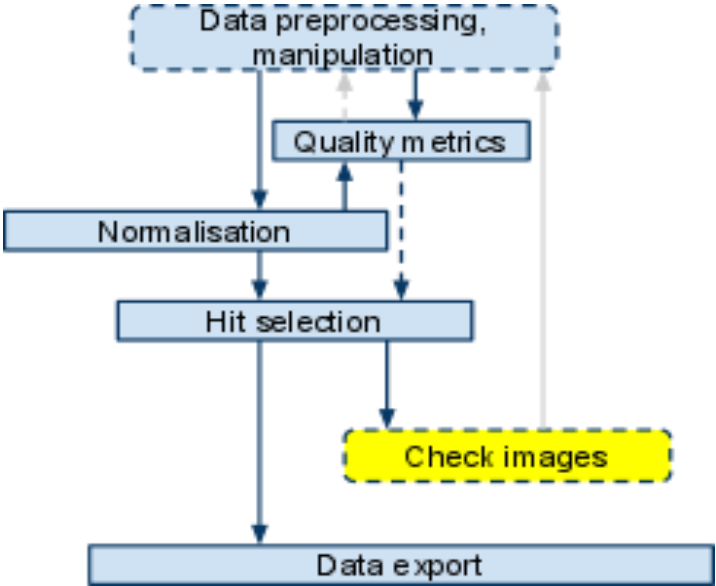
Amanda Birmingham, et al.:

Statistical methods for analysis of high-throughput RNA interference screens, *Nature Methods*, 2009

Check original images

Checking some of the initial images might also help deciding about hits, so we have implemented a simple image viewer for Bio-Format supported image formats (although tested only with IN Cell Analyzer 1000 images).

HCDC offers more device specific readers.



Acknowledgements

Supported by a Marie Curie Fellowship in Trinity College Dublin.

Thanks for

- Aideen Long,
- Dara J. Dunican,
- Emliy Bennet,
- Antje Hoff,
- Michael Freeley, and all other users.

evopro Kft - workshop

Questions?

<http://code.google.com/p/hits>

Thanks for your attention