Nils Gehlenborg

14 April 2011

Nozzle: Report Guidelines

The main purpose of these guidelines is to ensure that individual pipeline reports and the summary report can be interpreted efficiently by biologists and bioinformaticians, both within and outside the Broad/Harvard TCGA GDAC team. Furthermore, this document is an effort to help pipeline developers decide what to put into their reports and how.

General

Reports should be written for people who are familiar with cancer biology, but most likely unfamiliar with Firehose or the methods used to generate the results that are included in the reports.

Report Title

DOs:

- The report title should be a sentence fragment and begin with a noun describing the “action” of the pipeline (“Correlation”, “Consensus hierarchical clustering”), followed by the data types to which the action is applied (“mRNA expression”, “clinical data”). Of course the correct prepositions, etc. should be used. Examples are shown below.
- The first letter of the first word should be capitalized and all following words should be written in lowercase, unless they are abbreviations (“mRNA”), gene symbols (“TP53”), or similar terms.

DON’Ts:

- Report titles should not include the word “report”, “summary” or similar words. In the future Nozzle may automatically include this if necessary.
- Titles must not include any formatting or links.

Here are some example report titles:

“Correlation between copy number and miR expression”
“Consensus hierarchical clustering of mRNA expression”
“Correlation between clinical data and AGGREGATED_CLUSTER analysis” (This last example is a special case and certainly doesn’t represent an ideal solution and might change in the future.)

Summary Section

Summaries should be concise descriptions of the **results** generated by a pipeline. Input parameters and methods can be mentioned in the summary, but their description should be included in the Methods section.

DOs:

- All variable names and values should be marked up using the appropriate Nozzle commands, e.g. “asValue()” and “asParameter()”. This will allow Nozzle to highlight these parts of the summary, which are typically the parts that change across runs/tumor types.
- If there are multiple results included in a report, they may be included in the summary in form of a list created with “newList()”. This will make them easier to read.

DON’Ts:

- References to Figure, Table, or Citation elements should not be included in the summary, since these will not be accessible in the summary report and are irrelevant at that level. There are currently also some technical limitations that prevent such references from being inserted correctly.
- Please don’t include Figure, Table, or Result elements in the summary.

Here is a (shortened) example for a good summary:

“By testing the association of 18699 genes to 8 clinical features, statistically thresholded by P value <= 0.05 and Q value <= 0.25:
- 31 genes correlated to 'Time to Death'. … “

Table Elements

DOs:

- Variables used as column names should be described in the caption (at least for now, future versions of Nozzle might feature a “table legend” element for this purpose). Common variables such as “p” or “q” don’t have to be described. However, it would be better to name these columns “p-value” and “q-value”.
- Columns that contain only numbers should be of type “numeric”. This will allow Nozzle to automatically round numbers with (too) many decimals to a predefined number (2 or 3) of significant digits. However, this will only be applied to columns for which *is.numeric(table[,x]) == TRUE*.
- Rows should represent genes and other genomic features.
- Columns should be used for samples, groups, variables, and related objects.

DON’Ts:

- Formatted text should be used only where absolutely necessary, e.g. to include links in a table.

- Tables are not supposed to be used to layout text, e.g. by creating sparse matrices that contain text only in a few cells. (If there is a need for more flexible layout of text this can be added to Nozzle.)

Figure Elements

DOs:

- Variable names used as axis labels or elsewhere in the figure should be described in the caption (at least for now, future versions of Nozzle might feature a “table legend” element for this purpose).

Other Considerations

DON’Ts:

- Correlation with batches and other quality control issues should not be marked up as a result (significant or not). In the future, Nozzle reports are going to contain a dedicated section to note such issues, that can then be aggregated into an overall quality control report for all pipelines.

Checklist

- Does the report title follow the guidelines?
- Does the summary focus on results and does not include any references to figures, tables, or citations?
- Are all parameters of the pipeline mentioned in the report marked up with “asParameter()”?
- Are all values of parameters mentioned in the report marked up with “asValue()”?
- Are all links in the report generated with “asLink()”?
- Are all filenames mentioned in the report marked up with “asFilename()”, in particular in the “Input” section.
- Are tables referred to using “asReference(myTable)”?
- Are figures referred to using “asReference(myFigure)”?
- Are citations referred to using “asReference(myCitation)”?
- Do figure and table captions describe the variable names used in column headers and as axis labels?
- Do figures that potentially might get large (wider than 800 pixels) have a link to the “High-Res” image?
- Does the report use nomenclature as agreed:
  - **miR** - Not: microRNA or miRNA
- Is “writeReport()” being called **without** “debug=TRUE”?