QPCR*in*R

v.0.9.7

Powerful and interactive
RTqPCR data analysis
with Linear Mixed Models

User's Guide

BENAKI
PHYTOPATHOLOGICAL
INSTITUTE

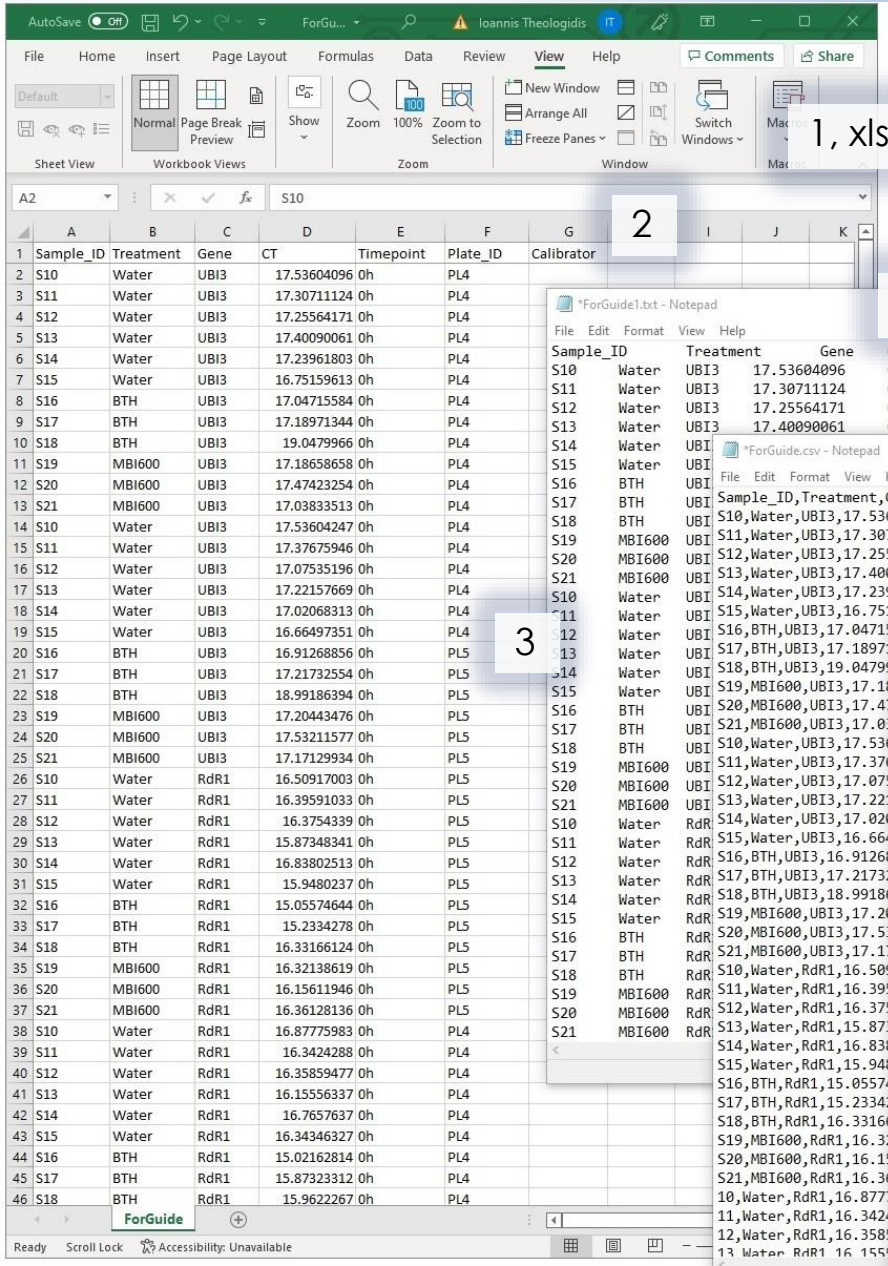
R Programming

Table of Contents

Quick Start

- Load data
  - Choose data file, xlsx, csv or txt (Comma, Tab or Semicolon separated).
  - Define column/variable classes (**Notice**: All but CT variable columns should be "factor". CT as "numeric" – "num"). Do not forget to click [*Submit Class Change*] if you make changes.
- Specify Model
  - Drag CT values into *Response* bucket.
  - Drag Gene and Treatment(s) into *Fixed Effects* bucket.
  - (Optionally) drag variables such as SampleID, PlateID, BlockID, etc., into *Random Effects* bucket.
  - Click [*Run Model!*].
- Assign Comparisons
  - Select Reference Gene(s) from list.
  - Select the Column that contains the experimental condition that will be considered as baseline (control) and then select the Control Treatment itself.
  - Click [*Confirm Contrasts*].
- Get Result Table and Plot
  - Inspect, explore Results and click [*Download Result Table*] to save as a csv file.
  - Select axes for Plot and click [*Create Plot*].
  - Download Plot by clicking the camera [📷] button on the top right-hand side of the plot.

Before Loading Data



- Data file preparation

1. rQPCR accepts *.xlsx*, *.csv* or *.txt* files with column data that are separated by either "tab", "comma" or "semicolon" characters. Their general structure is depicted here.
2. The first row should be the descriptive name of each variable.
3. Data structure must be in the, so-called, long format.

Example: Assume an experimental assay that involves testing three different concentrations of an antibiotic. All of them should be in the same column, under the general header, *Antibiotic*", or "*Treatment A*", or "[the name of the Antibiotic]", etc. The control of this treatment (i.e., "water" or "no Antibiotic" or "control", etc., should be included under the same header, as well.

If this experiment was held at, say, three different conditions (i.e., temperatures, time-points), then another column should be populated with all the levels of this condition, under the general header("*Treatment B*", "*Temperature*", "*Condition*", etc.).

Before Loading Data

1. **Important!** Ensure that the column that contains the gene names is titled as "Gene(s)" or "gene(s)" (or even "GENE(S)"). If not, QPCR*in*R will ask the user to indicate the "gene" column and will rename it accordingly.

2. For Inter-Run Calibration, "calibrator" samples are required (see Model Specification). These samples must be "tagged" in a separate column with a specific name (i.e., "Calib"). Everything else in this column can be empty; any header name is allowed.

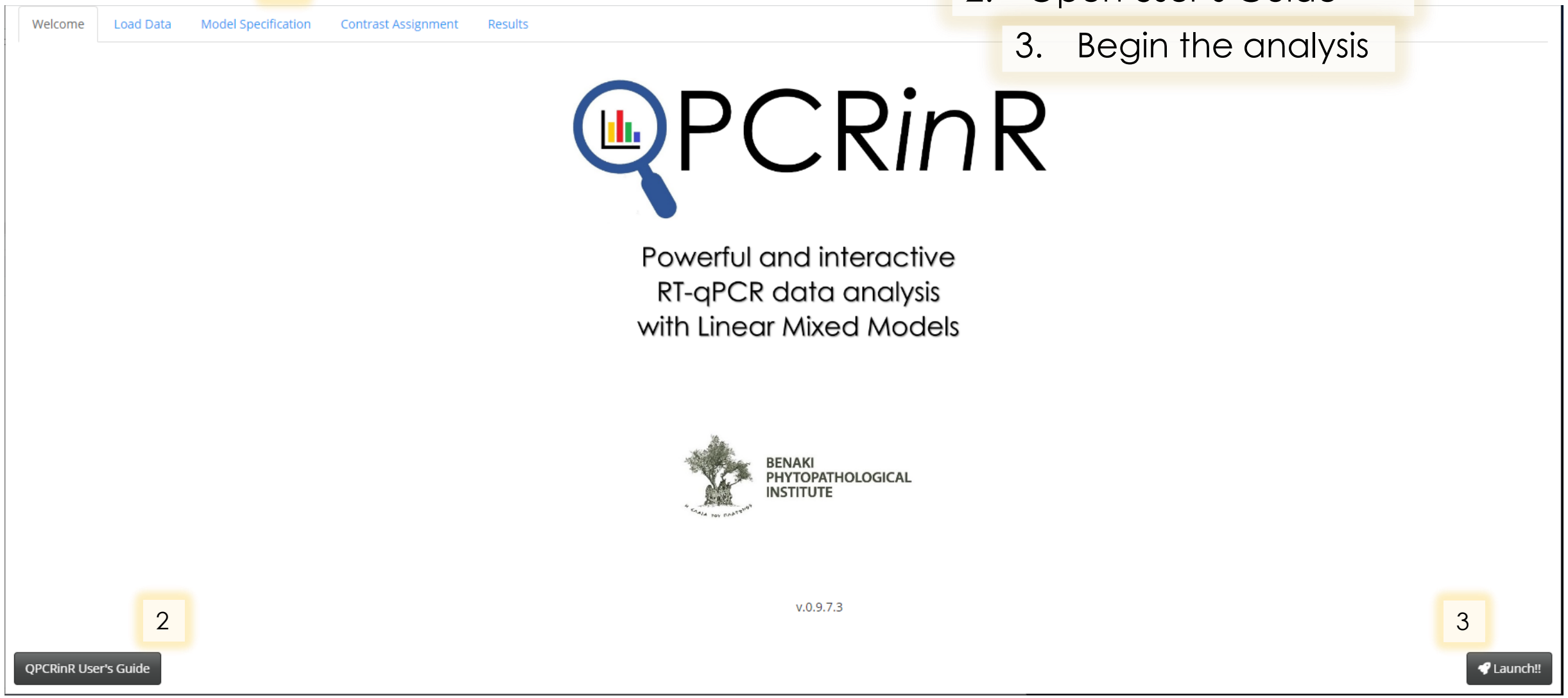
- **Technical and Biological Replication**

  Technical replication is important in RT-qPCR experiments. QPCR*in*R requires that technical replicates in the data set are essentially the same except the C_T value. In other words, two technical replicates' rows contain identical inputs except their C_T values.

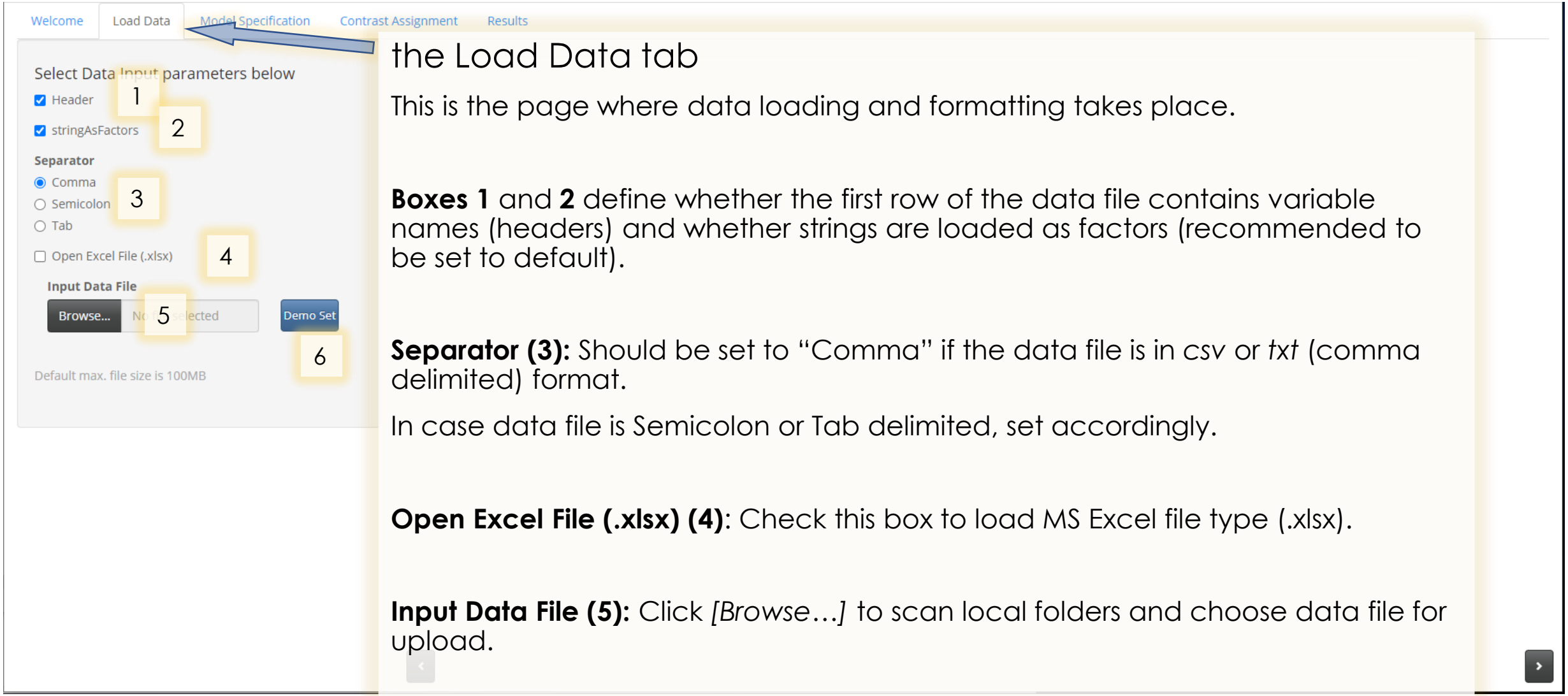
  Accordingly, biological replicates share one more difference: All but C_T values *and* sample_IDs (or the equivalent column name) are equal among biological replicate rows in the data set.

Welcome screen

1. Navigation tabs
2. Open User's Guide
3. Begin the analysis



Load Data



the Load Data tab

This is the page where data loading and formatting takes place.

Boxes 1 and **2** define whether the first row of the data file contains variable names (headers) and whether strings are loaded as factors (recommended to be set to default).

Separator (3): Should be set to "Comma" if the data file is in *csv* or *txt* (comma delimited) format.

In case data file is Semicolon or Tab delimited, set accordingly.

Open Excel File (.xlsx) (4): Check this box to load MS Excel file type (.xlsx).

Input Data File (5): Click *[Browse…]* to scan local folders and choose data file for upload.

Load Demo Dataset (6): Click this box to load a dataset for practice.

Load Data

As soon as data uploading is completed, additional option menus and panels appear:

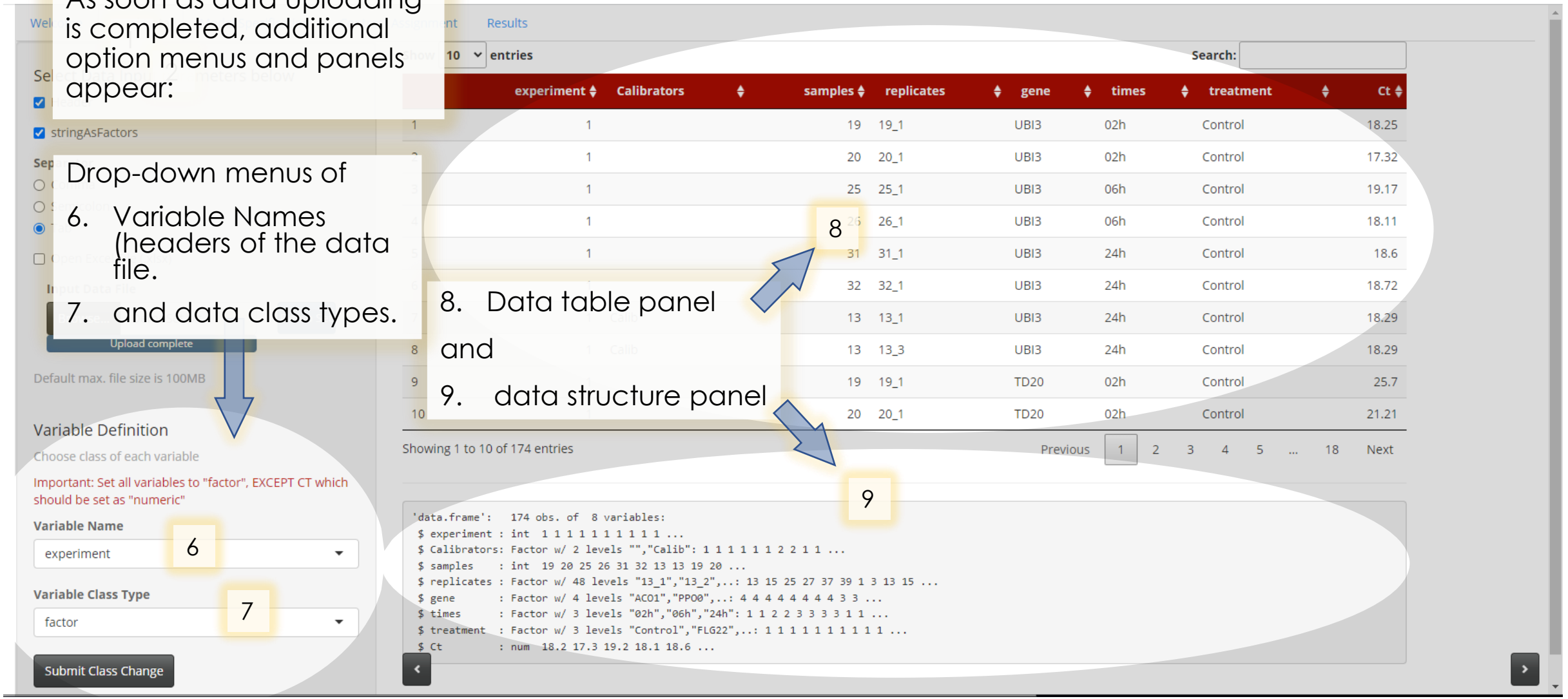
Drop-down menus of

6. Variable Names (headers of the data file.
7. and data class types.

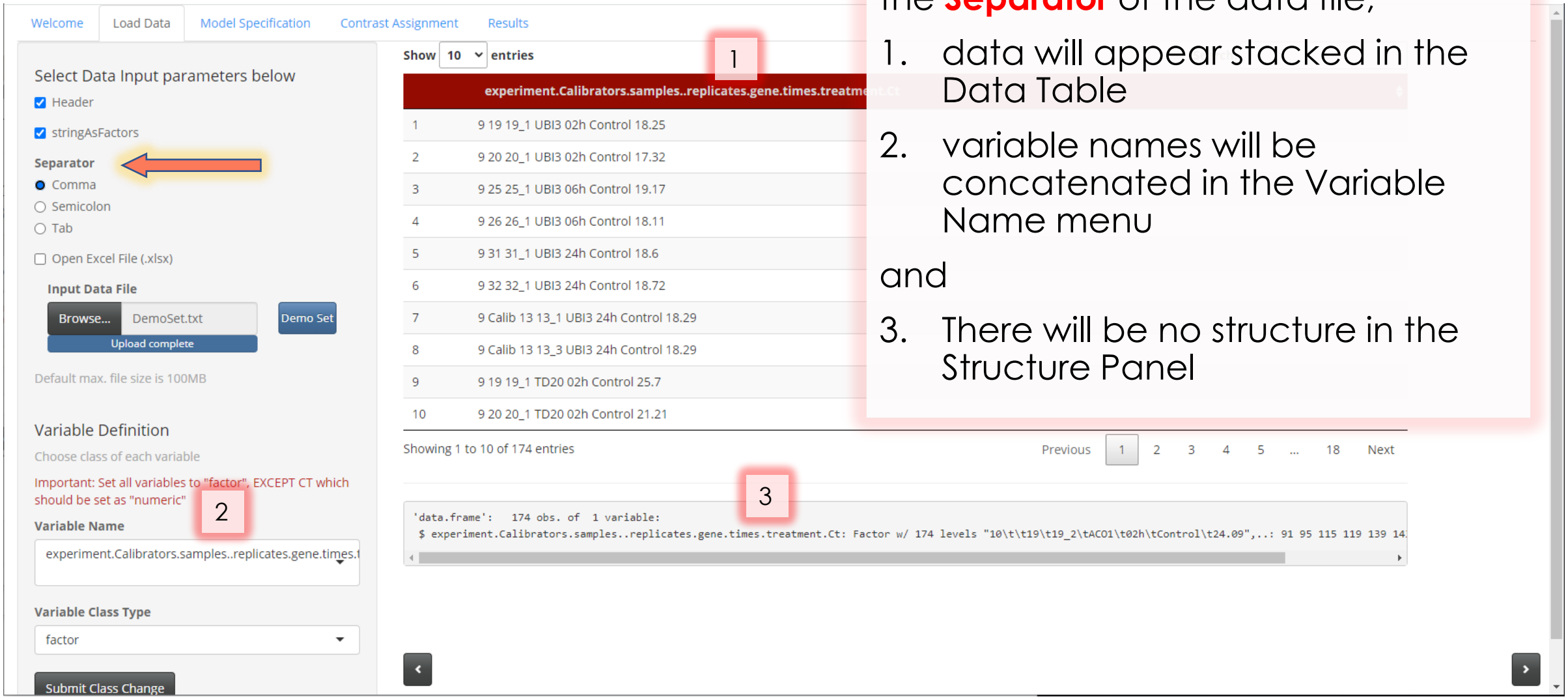
8. Data table panel

and

9. data structure panel



Load Data



In case of inappropriate selection of the **Separator** of the data file,

1. data will appear stacked in the Data Table
2. variable names will be concatenated in the Variable Name menu

and

3. There will be no structure in the Structure Panel

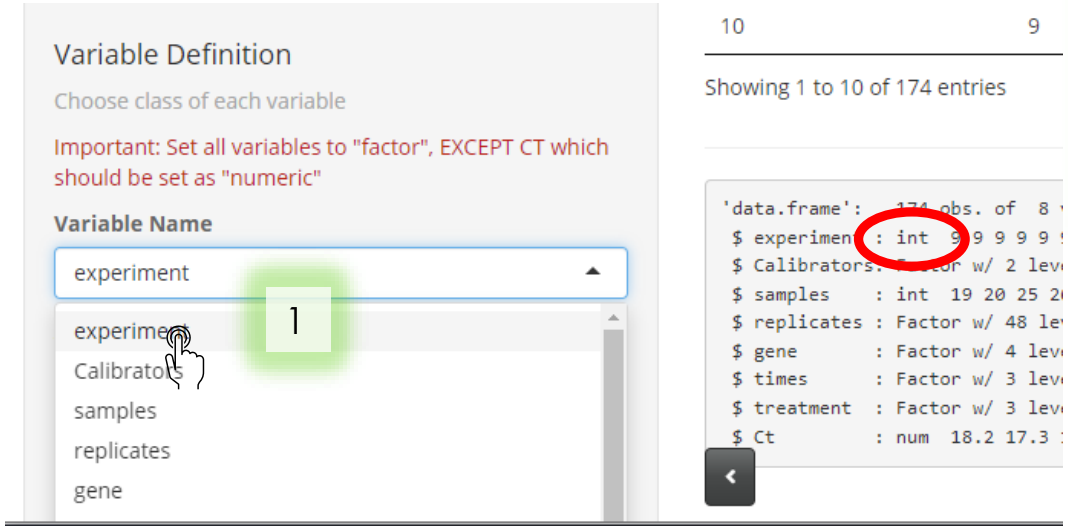
Load Data

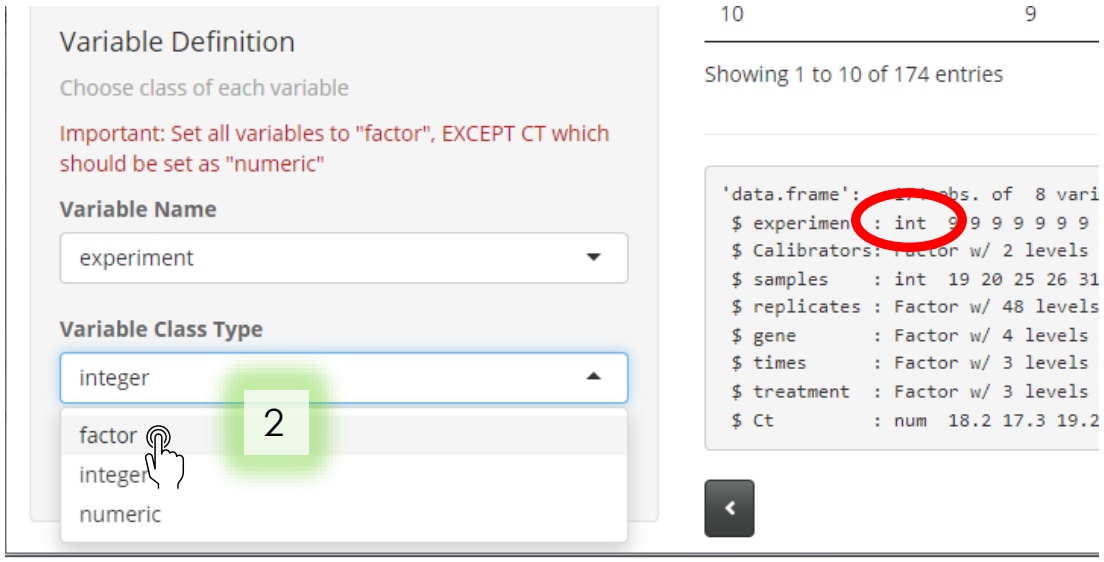
Change variable class type

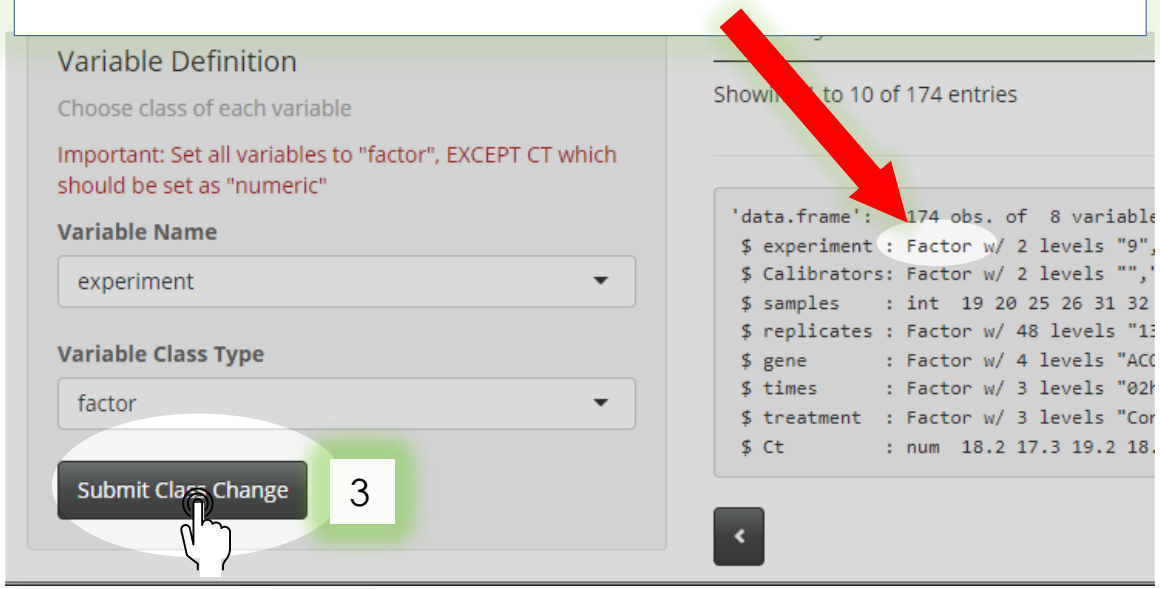
QPCR*in*R performs factorial analysis. Users must inspect and, if necessary, change the class of the data variables to "factor", except CT values.
In the Variable Definition menu:

1. Select Variable Name
2. Select Class type
3. Click [*Submit Class Change*]

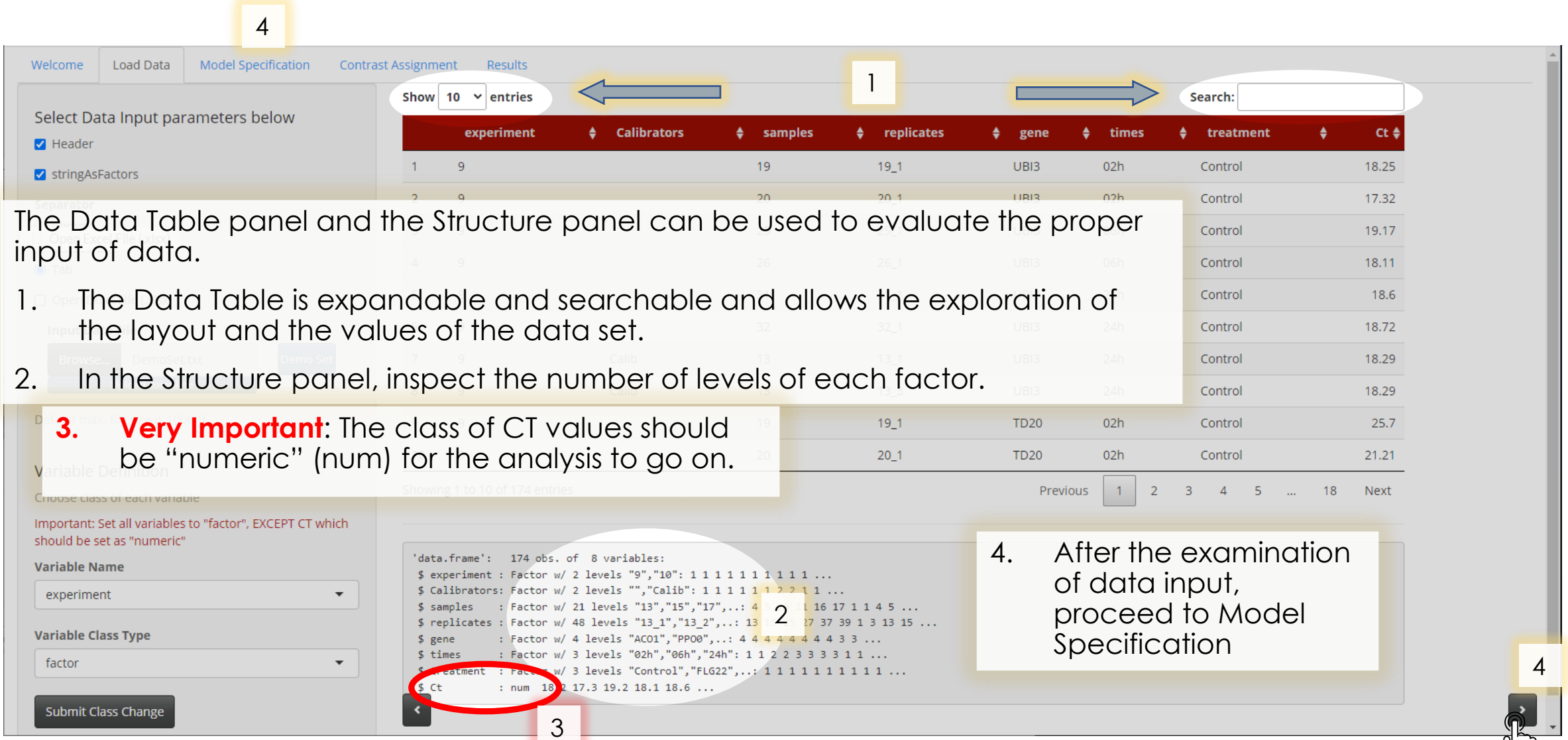
Important: Check the structure panel to confirm the class change







Load Data



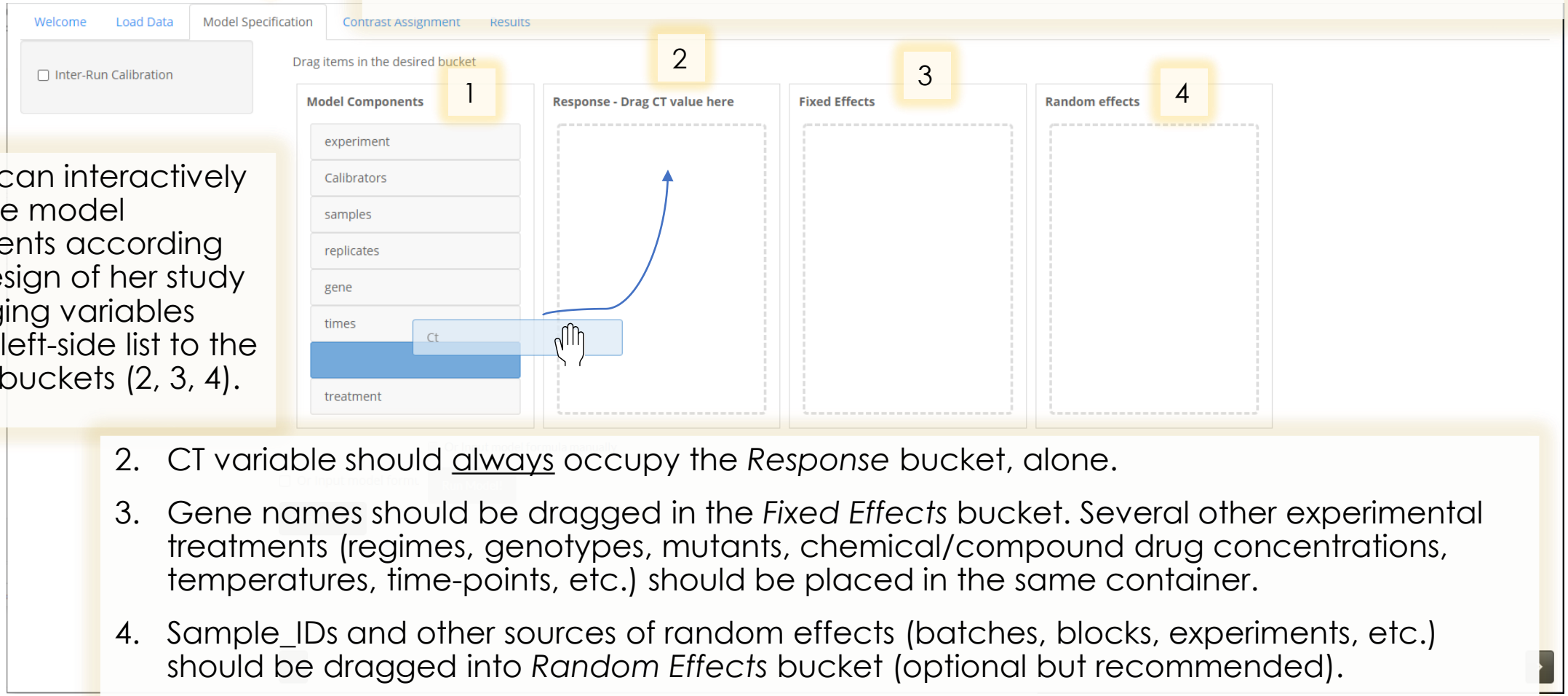
The Data Table panel and the Structure panel can be used to evaluate the proper input of data.

1. The Data Table is expandable and searchable and allows the exploration of the layout and the values of the data set.
2. In the Structure panel, inspect the number of levels of each factor.
3. **Very Important**: The class of CT values should be "numeric" (num) for the analysis to go on.
4. After the examination of data input, proceed to Model Specification

Model Specification

the Model Specification tab

QPCR*in*R implements Linear (Mixed) Models for the analysis of qPCR experiments. This framework offers possibilities for incorporating complex designs and controlling for several variance components (See main article and references therein for details).

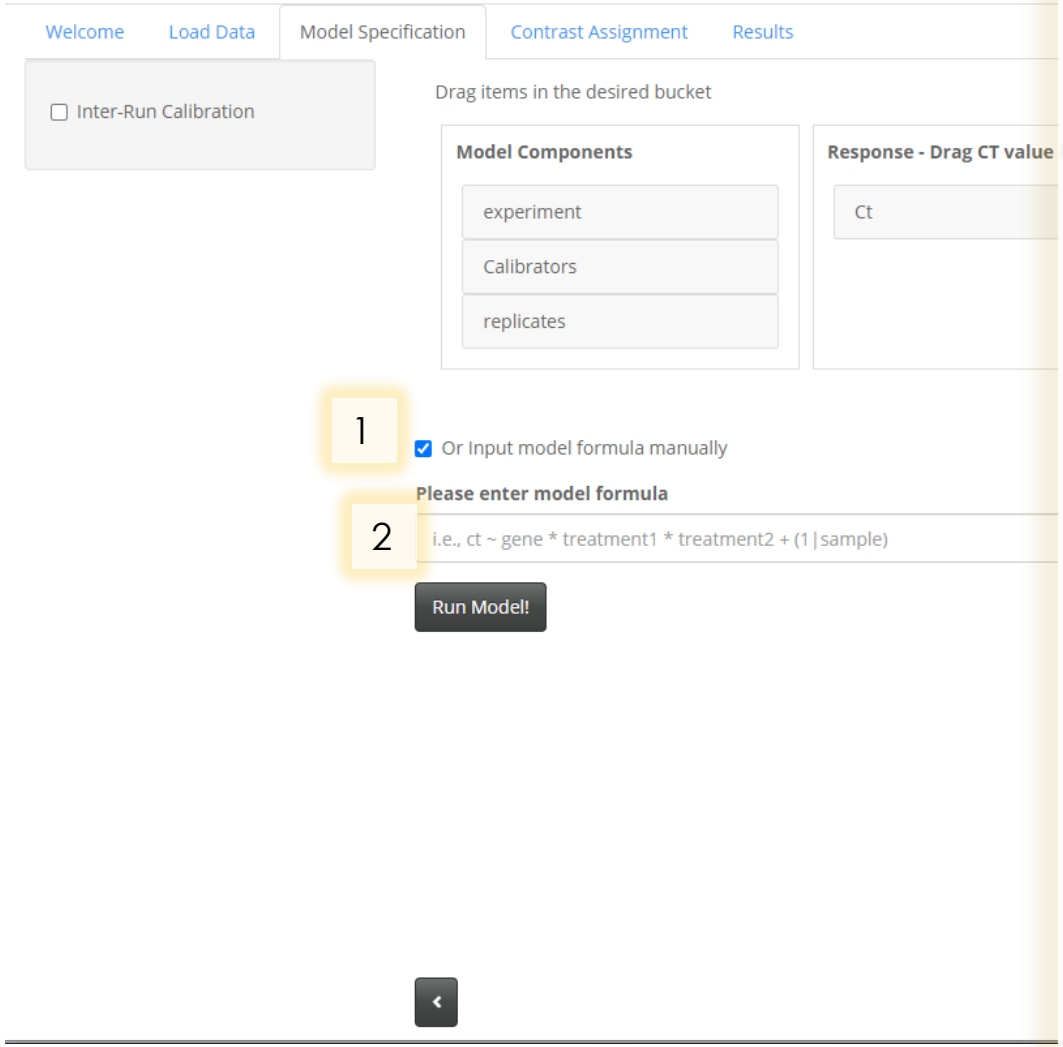


1. The user can interactively define the model components according to the design of her study by dragging variables from the left-side list to the relevant buckets (2, 3, 4).

2. CT variable should <u>always</u> occupy the *Response* bucket, alone.
3. Gene names should be dragged in the *Fixed Effects* bucket. Several other experimental treatments (regimes, genotypes, mutants, chemical/compound drug concentrations, temperatures, time-points, etc.) should be placed in the same container.
4. Sample_IDs and other sources of random effects (batches, blocks, experiments, etc.) should be dragged into *Random Effects* bucket (optional but recommended).

Model Specification

Manual Formula Input



For more complex model specification, the option for manual model input is available.

The formula should be expressed in accordance with R programming, as an *lm* or *lmer* function input.

1. Check box (1).
2. Type the desired model formula in box (2):
   - *I.* *Response*: the [Ct] variable followed by '~'.
   - *II.* *Fixed Effects*: each [variable] followed by '+' or '*' when modelling interactions with the next variable.
   - *III.* *Random Effects*, optional but recommended: Input variables after the string: "+(1 | [variable1])". For interactions and nesting, the general formula would be: "+(1 | [variable1]:[variable2])"

Example: Ct ~ gene*treatment*times+(1 | samples)+(1 | samples:gene)

Response — Fixed Effects — Random Effects

Please, refer to R language manuals for more information about formula expressions.

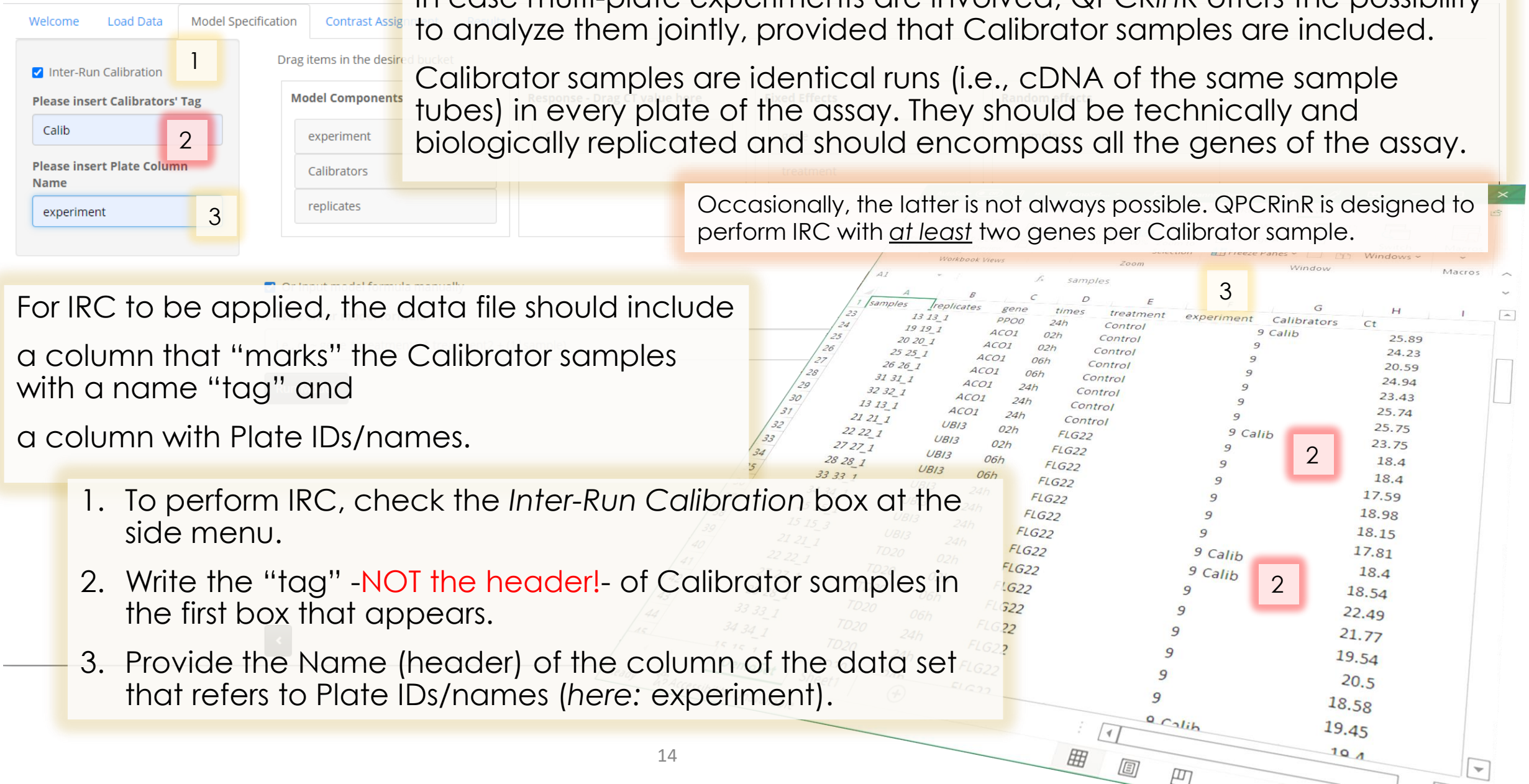
Model Specification

Inter-Run Calibration (IRC)

In case multi-plate experiments are involved, QPCR*in*R offers the possibility to analyze them jointly, provided that Calibrator samples are included.

Calibrator samples are identical runs (i.e., cDNA of the same sample tubes) in every plate of the assay. They should be technically and biologically replicated and should encompass all the genes of the assay.

Occasionally, the latter is not always possible. QPCRinR is designed to perform IRC with <u>*at least*</u> two genes per Calibrator sample.



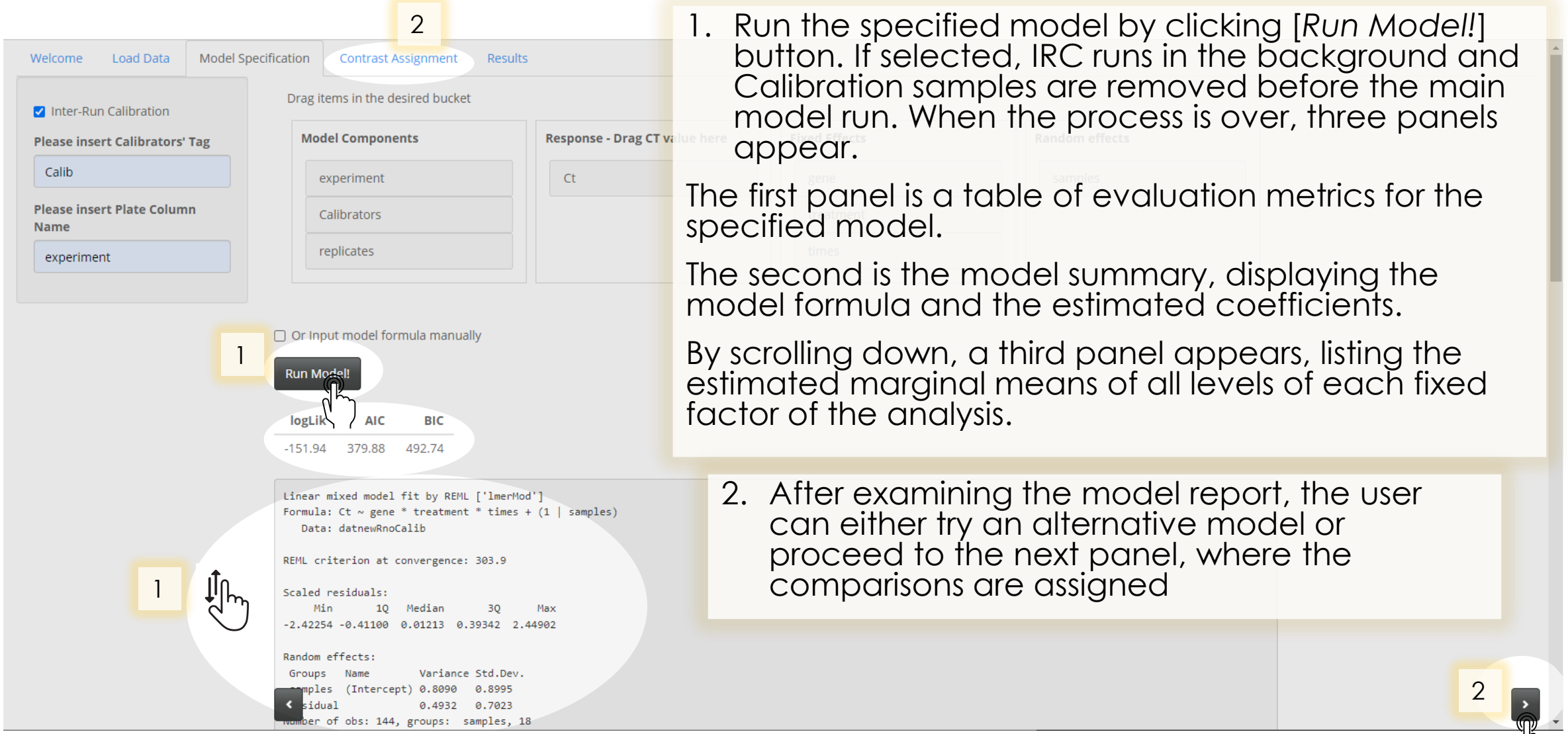
For IRC to be applied, the data file should include

a column that "marks" the Calibrator samples with a name "tag" and

a column with Plate IDs/names.

1. To perform IRC, check the *Inter-Run Calibration* box at the side menu.
2. Write the "tag" -NOT the header!- of Calibrator samples in the first box that appears.
3. Provide the Name (header) of the column of the data set that refers to Plate IDs/names (*here*: experiment).

Model Specification



1. Run the specified model by clicking [*Run Model!*] button. If selected, IRC runs in the background and Calibration samples are removed before the main model run. When the process is over, three panels appear.

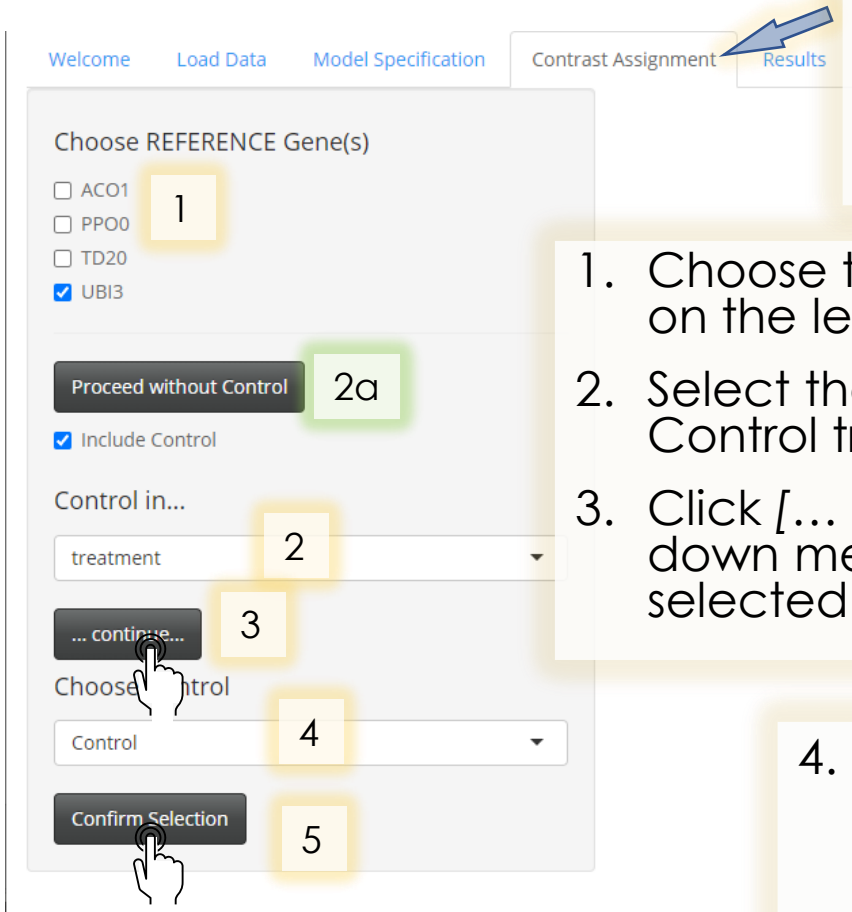
The first panel is a table of evaluation metrics for the specified model.

The second is the model summary, displaying the model formula and the estimated coefficients.

By scrolling down, a third panel appears, listing the estimated marginal means of all levels of each fixed factor of the analysis.

2. After examining the model report, the user can either try an alternative model or proceed to the next panel, where the comparisons are assigned

Contrast Assignment



the Contrast Assignment tab

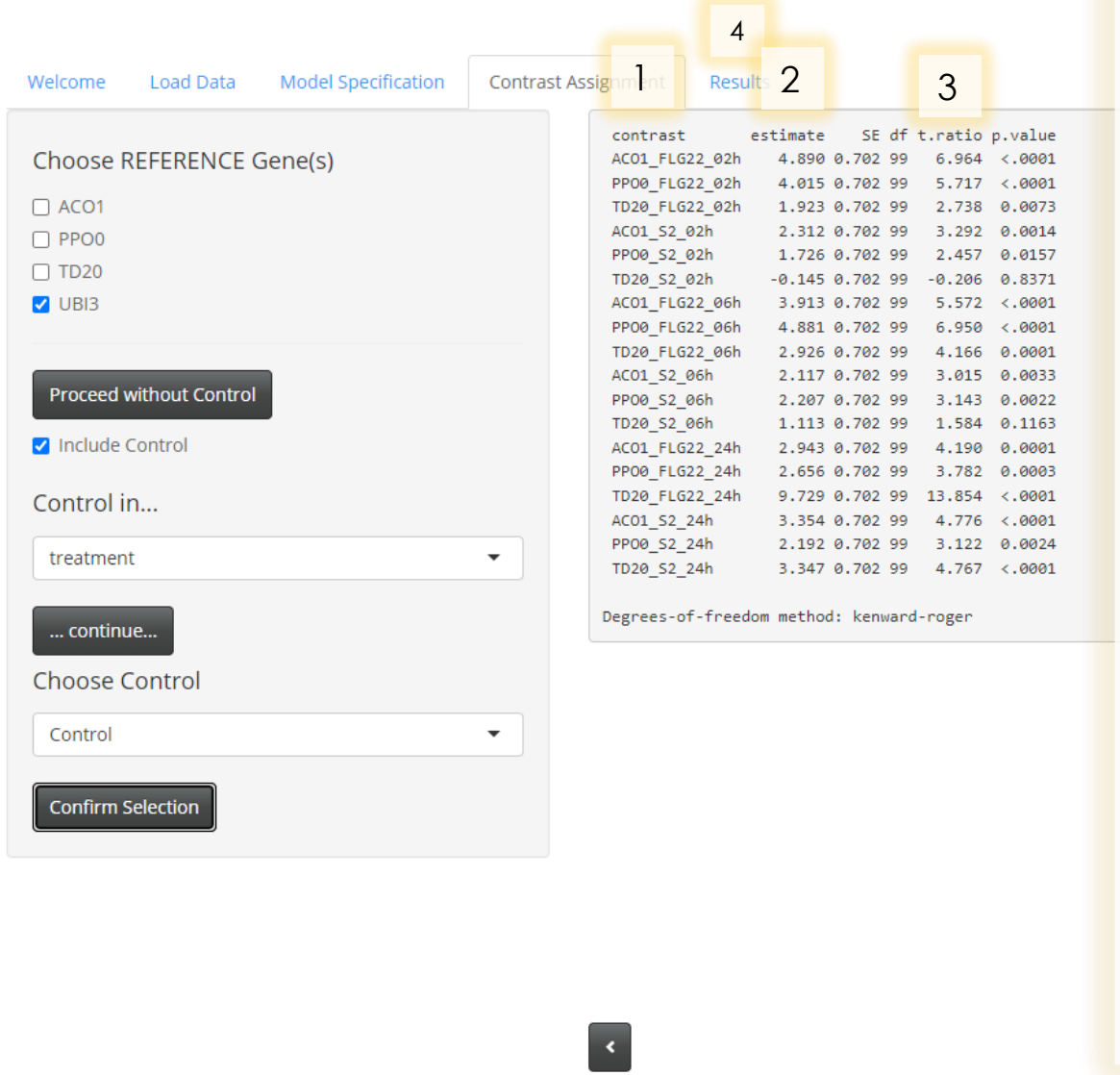
In this tab, the reference gene(s) and the baseline (control) treatment are defined for relative comparisons.

1. Choose the reference gene(s) from the list on the left
2. Select the variable that contains the Control treatment
3. Click *[… continue…]* to activate a drop-down menu of all levels of treatment selected at step 2

2a. Alternatively, the analysis can proceed without control assignment by unchecking the [*Include Control*] box and clicking *[Proceed without Control*].

4. Select the level that corresponds to Control (it doesn't necessarily have to be named 'control'; it may be any level of any of the *Fixed Factors* that will be considered as the baseline).
5. Click [*Confirm Selection*]

Contrast Assignment

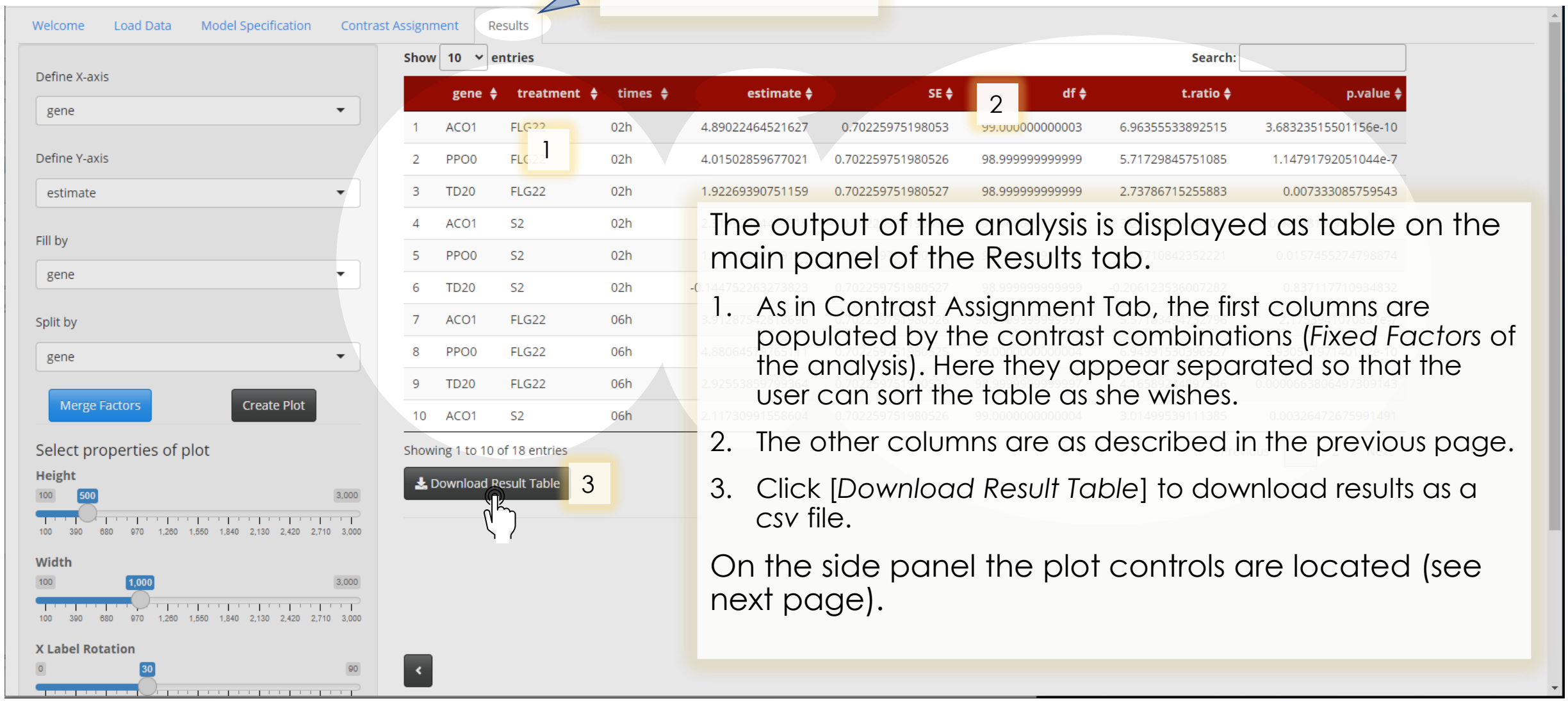


After [*Confirming Selection*] or [*Proceeding without Control*], the main panel will be populated with the results.

1. The first column is the contrast combination. As expected, the reference gene is absent. If the choice in *Contrast Assignment* tab was comparisons <u>with</u> control treatment (a ΔΔct equivalent), then the control treatment will be absent, too.
2. The next column (*estimate*) is the estimated log2FoldChange of each factor combination <u>relative to the reference gene ***and*** the control treatment</u> (if there was one, see 1).
3. Next, the standard error of the estimate (SE), the degrees of freedom (df), the t.ratio and the p.value of the test are shown.
4. Click on *Results* tab or the arrow at the bottom right corner to see this table in interactive and expandable format.

4

Results



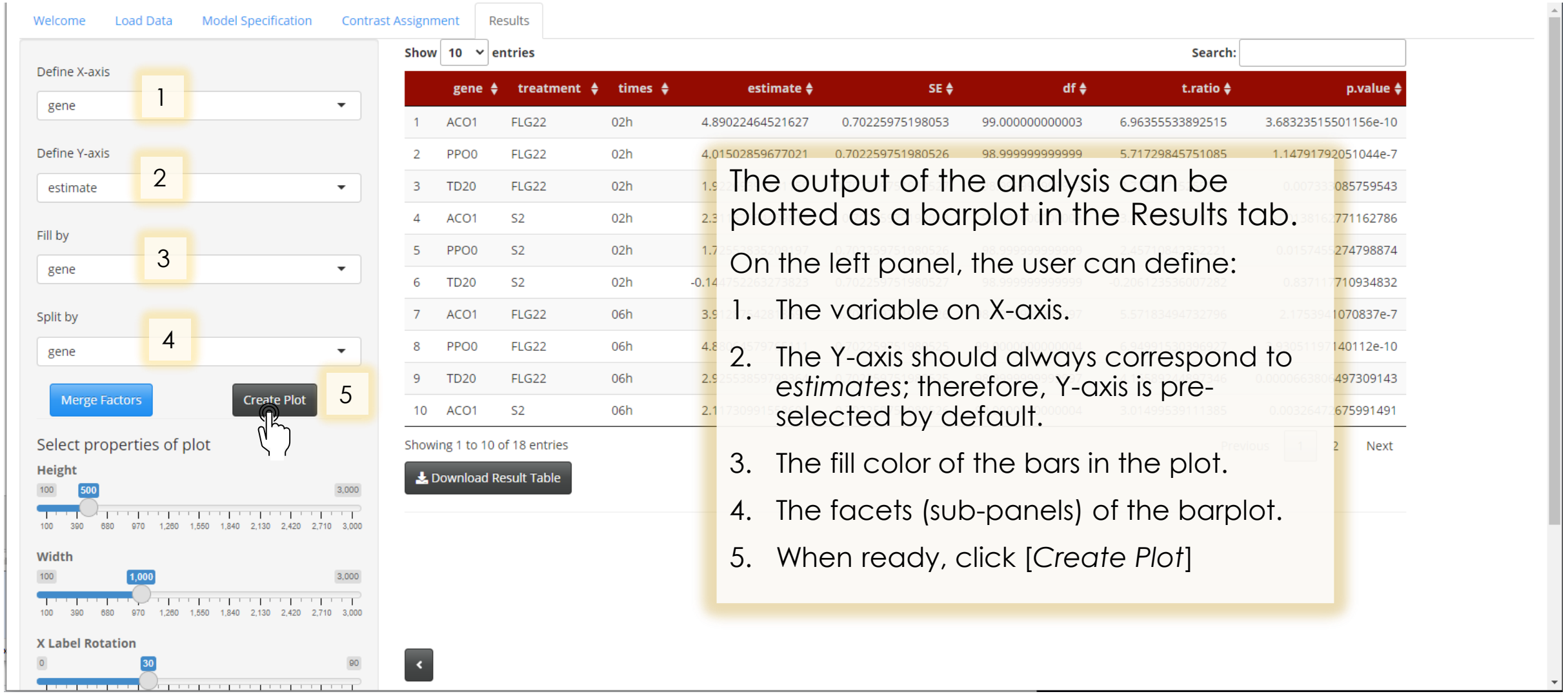
The output of the analysis is displayed as table on the main panel of the Results tab.

1. As in Contrast Assignment Tab, the first columns are populated by the contrast combinations (*Fixed Factors* of the analysis). Here they appear separated so that the user can sort the table as she wishes.
2. The other columns are as described in the previous page.
3. Click [*Download Result Table*] to download results as a *csv* file.

On the side panel the plot controls are located (see next page).

Results

Create Plot



The output of the analysis can be plotted as a barplot in the Results tab.

On the left panel, the user can define:

1. The variable on X-axis.
2. The Y-axis should always correspond to *estimates*; therefore, Y-axis is pre-selected by default.
3. The fill color of the bars in the plot.
4. The facets (sub-panels) of the barplot.
5. When ready, click [*Create Plot*]

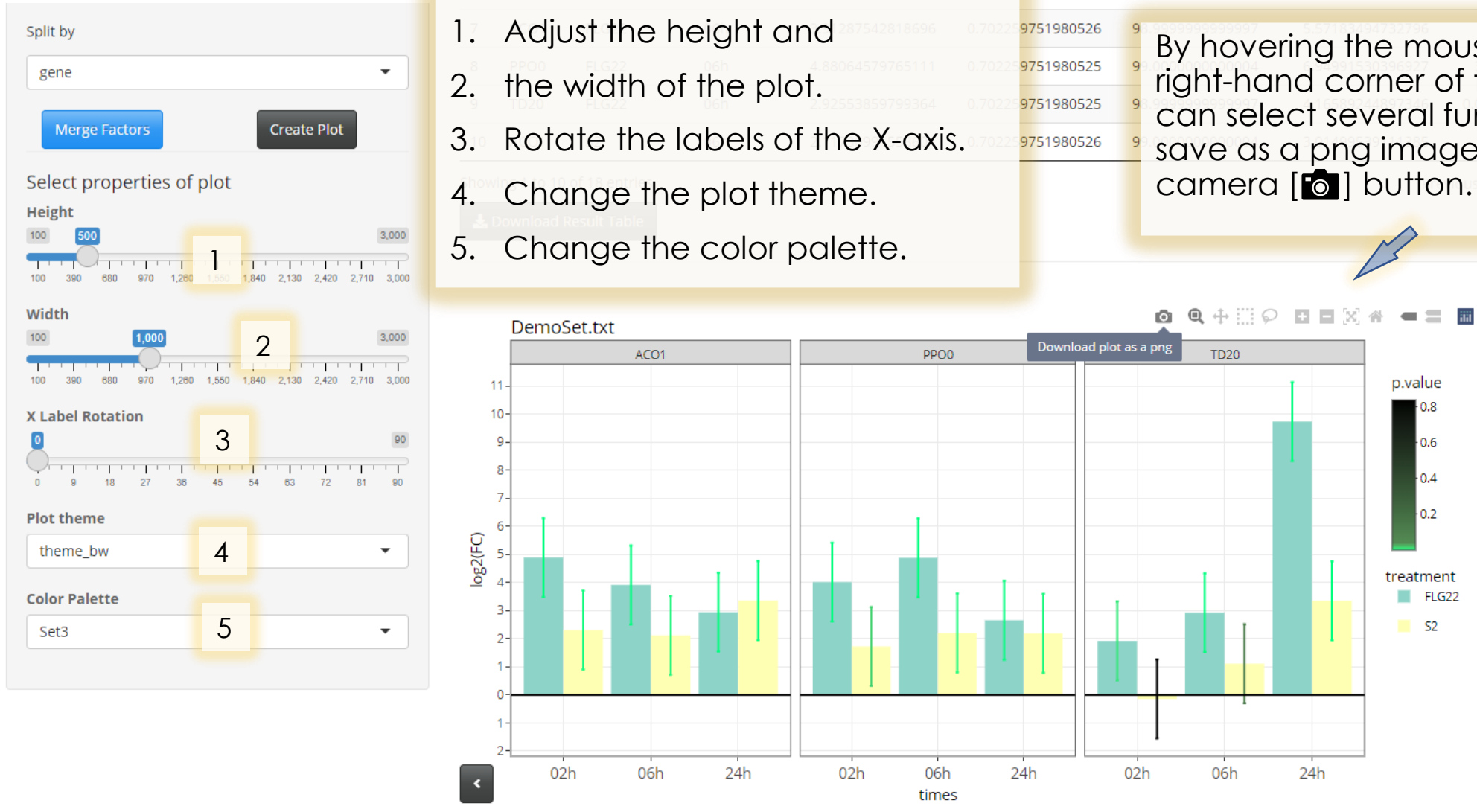
Results

Create Plot

The plot can be further formatted with the Properties panel:

1. Adjust the height and
2. the width of the plot.
3. Rotate the labels of the X-axis.
4. Change the plot theme.
5. Change the color palette.

By hovering the mouse over the top right-hand corner of the plot, the user can select several functionalities. To save as a png image, click the camera [📷] button.



Results

Merge Functionality

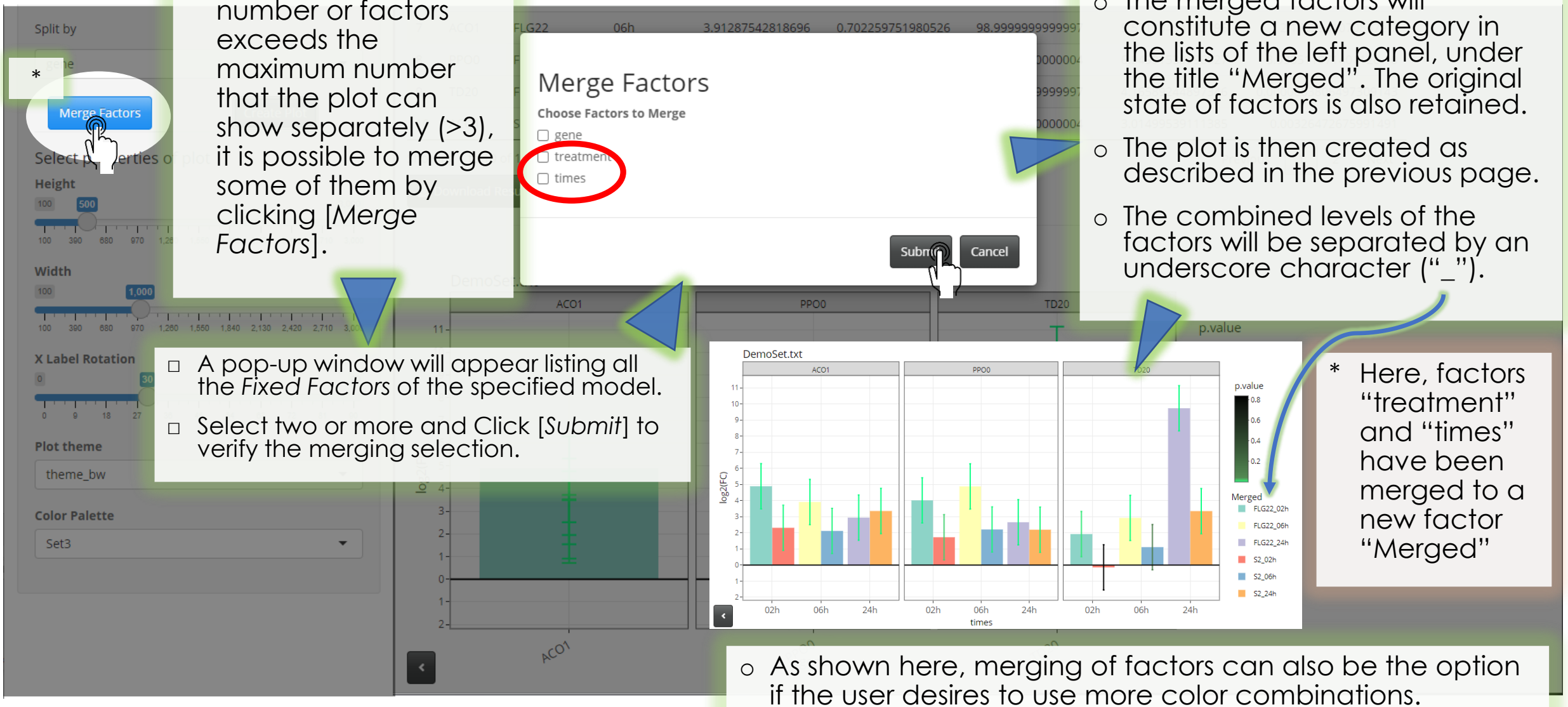
* In cases where the number or factors exceeds the maximum number that the plot can show separately (>3), it is possible to merge some of them by clicking [*Merge Factors*].

- A pop-up window will appear listing all the *Fixed Factors* of the specified model.
- Select two or more and Click [*Submit*] to verify the merging selection.

- The merged factors will constitute a new category in the lists of the left panel, under the title "Merged". The original state of factors is also retained.
- The plot is then created as described in the previous page.
- The combined levels of the factors will be separated by an underscore character ("_").

* Here, factors "treatment" and "times" have been merged to a new factor "Merged"

- As shown here, merging of factors can also be the option if the user desires to use more color combinations.



Credits and Support

Data

Demo data file is a partial subset extracted from:

> Dimopoulou, A., Theologidis, I., Liebmann, B., Kalantidis, K., Vassilakos, N., & Skandalis, N. (2019). *Bacillus amyloliquefaciens* MBI600 differentially induces tomato defense signaling pathways depending on plant part and dose of application. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-55645-2

Source code and Support

- https://github.com/theojohn/QPCRinR

Basic Literature

- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/JSS.V067.I01
- Lenth, R. V., Buerkner, P., Herve, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). CRAN - Package emmeans. https://cran.r-project.org/web/packages/emmeans/index.html
- Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K., & Weinhäusel, A. (2014). A survey of tools for the analysis of quantitative PCR (qPCR) data. In Biomolecular Detection and Quantification (Vol. 1, Issue 1, pp. 23–33). Elsevier GmbH. https://doi.org/10.1016/j.bdq.2014.08.002
- Steibel, J. P., Poletto, R., Coussens, P. M., & Rosa, G. J. M. (2009). A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. Genomics, 94(2), 146–152. https://doi.org/10.1016/j.ygeno.2009.04.008