

Title	Simplicity DiffExpress: A bespoke cloud-based interface for RNA-seq differential expression modeling and analysis
Authors	Palu, Cintia C.;Ribeiro-Alves, Marcelo;Wu, Yanxin;Lawlor, Brendan;Baranov, Pavel V.;Kelly, Brian;Walsh, Paul
Publication date	2019-05-14
Original Citation	Palu, C.C., Ribeiro-Alves, M., Wu, Y., Lawlor, B., Baranov, P.V., Kelly, B. and Walsh, P., 2019. Simplicity DiffExpress: a bespoke cloud-based interface for RNA-seq differential expression modeling and analysis. Frontiers in genetics, 10, (356). DOI:10.3389/fgene.2019.00356
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.frontiersin.org/articles/10.3389/fgene.2019.00356 - 10.3389/fgene.2019.00356
Rights	© 2019 Palu, Ribeiro-Alves, Wu, Lawlor, Baranov, Kelly and Walsh - https://creativecommons.org/licenses/by/4.0/
Download date	2026-09-25 22:01:19
Item downloaded from	https://hdl.handle.net/10468/9068

Supplementary Material

DiffExpress: a bespoke cloud-based interface for RNA-seq differential expression modeling and analysis

Cintia C. Palu^{*}, Marcelo Ribeiro-Alves, Yanxin Wu, Brendan Lawlor, Pavel V. Baranov, Brian Kelly, Paul Walsh^{*}

*** Correspondence:**

Cintia Palu

cintia.palu@nsilico.com

Paul Walsh

paul.walsh@nsilico.com

DiffExpress beta RNA-seq reads differential expression analysis

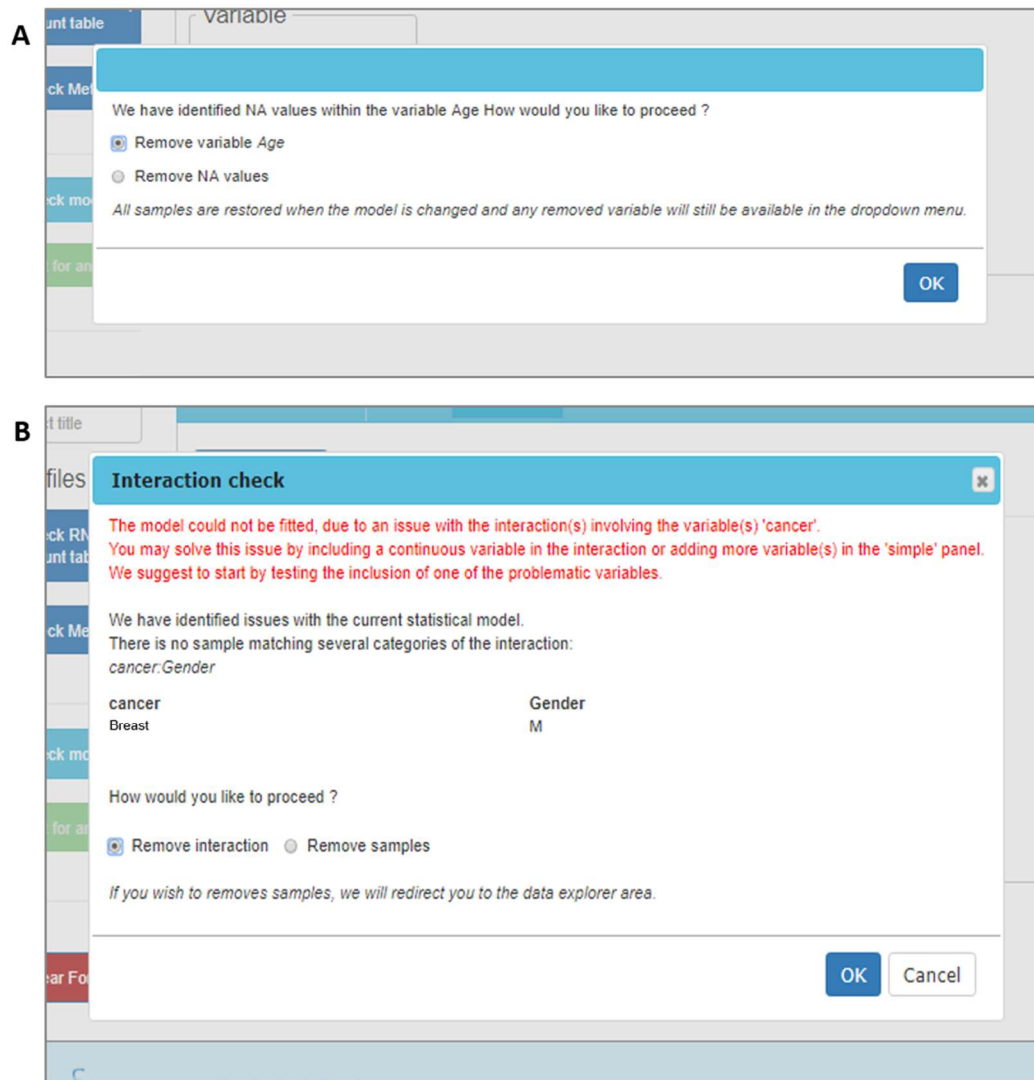
Project title

Input files

Statistical Design | Simple | Interaction | Removed samples | History | Statistics

Variable	Variable	Variable	Variable	Variable
batch	cancer	Gender	Age	Metastasis
☒ I want to remove this effect	☐ I want to remove this effect	☒ I want to remove this effect	☒ I want to remove this effect	☐ I want to remove this effect
☐ Continuous	☐ Continuous	☐ Continuous	☒ Continuous	☐ Continuous
Baseline (control) -- Select Value --	Baseline (control) HD	Baseline (control) -- Select Value --	Baseline (control) -- Select Value --	Baseline (control) N
Remove	Remove	Remove	Remove	Remove

Supplementary Figure 1. Statistical design modeling. It shows the set-up used for the analysis presented as an example in the paper (model $\sim batch + cancer + Gender + Age + Metastasis$). In this example, the focus is to look for genes differentially expressed in the samples from six cancer types and look for genes that could explain the metastasis occurrence. *Batch*, *Gender* and *Age* are considered bias sources and should be in the statistical model, but there is no need to test if there are differentially expressed genes due to those variables. Therefore the option ‘I want to remove this effect’ is selected for the variables. It is also important to inform which variables are continuous like was done *Age*. In the case of categorical variables that are not a source of bias, it is necessary to inform which level is the baseline. In the case of *cancer*, the aim is to identify changes in transcripts in relation to healthy donors, identified as ‘*HD*’ in this dataset. For *Metastasis* the baseline is ‘*N*’ because it indicates that no metastasis was observed in the patient. More details are presented at <https://www.youtube.com/watch?v=QKZu46c4HfU&feature=youtu.be>.



Supplementary Figure 2. Data validation. **(A)** The software can detect inconsistencies on the data upload and **(B)** when checking the model fitness.

A

Statistical Design	Simple	Interaction	Removed samples	History	Statistics
☒ Remove genes with low counts, CPM < 2					
☒ Robustify dispersion estimative against potential outliers					
☐ Compare the gene expression between every category within the selected variables for the simple data model					
Threshold criteria for plot generation					
logFC 1					
Adjusted p-value 0.05					
Reset statistics parameters to default value					

B

Statistical Design	Simple	Interaction	Removed samples	History	Statistics
Samples removed from clinical data (no matching sample found in the count-table) SRR1982702					
Samples removed due to lack of data for the variable(s) listed bellow - Age : 9 samples have been removed - cancer:Gender : 39 samples have been removed					

C

Statistical Design	Simple	Interaction	Removed samples	History	Statistics
- Checked model: ~Age+cancer + cancer:Gender No issues - NA presence detected in variable ~cancer:Gender 39 samples have been removed - Checked model: ~Age + cancer:Gender cancer:Gender is invalid - Checked model: ~Age + cancer:Gender cancer:Gender is invalid - NA presence detected in variable ~Age 9 samples have been removed					

Supplementary Figure 3. Other features. **(A)** It is possible to choose the settings for filtering genes, statistical methods and the criteria for considering a gene differentially expressed. **(B)** The interface keeps track of the samples that needed to be removed and **(C)** tracks the tested models and changes done.

The screenshot shows the 'My Pipelines' menu in the NSilico Simplicity application. The header includes the NSilico logo and navigation links: Home, My Pipelines, New Pipeline, Pipeline Studio, and Support. Below the header is a blue banner with the 'Simplicity' logo. The main content area is titled 'Finished Pipelines' and features a search bar and a 'Show 10 entries' dropdown. A table lists two completed pipelines:

ID	Name	Date Finished	Share	Remove
32825	TEP - SUCCESS	13 th August 2018 [12:33]	[Share Icon]	[X Icon]
32924	CD - SUCCESS	10 th August 2018 [09:36]	[Share Icon]	[X Icon]

Below the table, it indicates 'Showing 1 to 10 of 138 entries' and provides pagination controls: Previous, 1 (selected), 2, 3, 4, 5, ..., 14, Next. The footer contains copyright information: © 2018 - NSilico LifeScience Ltd. - Simplicity™ 1.5, along with social media links and a 'Terms of use' link.

Supplementary Figure 4. *Simplicity's 'My Pipelines' menu.* In this page, all submitted pipelines are listed. The pipeline ID is followed by the Project Title and completion status, with the finish date and the option to share the results with other users and delete the pipeline.