

# **AnalyCyte: Simplify Your Post-Analysis Reporting for Cytometry**

2024-04-15

# Table of contents

<b>Welcome</b>	<b>5</b>
Conventions Used in This Book . . . . .	5
Development Background . . . . .	5
Acknowledgments . . . . .	5
<b>Preface</b>	<b>6</b>
<b>Key Features</b>	<b>7</b>
<b>I User's Guide</b>	<b>8</b>
<b>1 Installation</b>	<b>9</b>
1.1 Docker . . . . .	9
1.1.1 Deployment (IFB example) . . . . .	9
1.1.2 Development . . . . .	9
<b>2 How to Start</b>	<b>11</b>
2.0.1 Dashboard Page . . . . .	11
2.0.2 The Application left panel . . . . .	11
<b>3 Workflow</b>	<b>14</b>
<b>4 Module list</b>	<b>16</b>
4.1 Batch Correction . . . . .	16
4.2 Extract Clustering Features . . . . .	16
4.3 Import Features . . . . .	17
4.4 Import Concatenated FCS . . . . .	17
4.5 Quality Control . . . . .	18
4.6 Differential Abundance . . . . .	19
4.7 Differential State . . . . .	20
4.8 Import R Object . . . . .	20
4.9 Export to Phantasus . . . . .	20
<b>5 File types terms</b>	<b>21</b>
<b>II Import</b>	<b>22</b>
<b>6 Import Concatened FCS</b>	<b>23</b>
6.1 Clustering selection . . . . .	23

<b>7 Import Features tables</b>	<b>24</b>
<b>8 Import R files</b>	<b>25</b>
<b>III Export</b>	<b>26</b>
<b>9 Export Analysis to disk</b>	<b>27</b>
<b>10 Phantasus Table</b>	<b>28</b>
<b>IV Content Exemple</b>	<b>29</b>
<b>11 Dataset description</b>	<b>30</b>
<b>12 Recurrent element</b>	<b>31</b>
<b>13 QC - control quality of features</b>	<b>32</b>
13.1 Counts . . . . .	32
13.2 Density heatmap . . . . .	32
<b>14 MFI heatmap checks</b>	<b>34</b>
<b>15 MFI / abundance heatmap</b>	<b>37</b>
<b>16 Abundance heatmap</b>	<b>38</b>
16.1 PCA . . . . .	38
16.1.1 Abundance . . . . .	38
16.1.2 MFI . . . . .	43
<b>17 DA - analyse abundancy of clusters - univar</b>	<b>49</b>
17.0.1 DA Tables Columns . . . . .	49
17.0.2 Interactive volcano and abundance plot . . . . .	50
17.0.3 violin plot . . . . .	51
17.1 Consensus results . . . . .	53
17.1.1 Arcsinh Transformed Frequencies . . . . .	53
<b>18 DS - analyse intensity of clusters - univar</b>	<b>55</b>
18.0.1 DS Tables Columns . . . . .	55
18.0.2 Interactive volcano plot . . . . .	55
18.0.3 violin plot . . . . .	56
18.0.4 MFI Marker*Cluster Heatmap . . . . .	56
<b>Glossary</b>	<b>59</b>
Glossary of Terms . . . . .	59
<b>References</b>	<b>61</b>

<b>Appendices</b>	<b>62</b>
<b>A Deploying with IFB</b>	<b>62</b>
<b>B Import data from Flowjo</b>	<b>63</b>
<b>C Import data from OmiQ</b>	<b>64</b>
<b>D Analycyte-verse</b>	<b>65</b>
<b>E YAML Parameters</b>	<b>66</b>
E.1 Explanation of YAML Parameters . . . . .	66

# Welcome

AnalyCyte is an **shiny application** designed to improve post-analysis for cytometry data and user control over this process. This tool simplifies the workflow and facilitate effective decision-making in research and clinical settings. Users can manage their analyses independently, thanks to customized, modular standardized reports.

The tool supports data export from [OMIQ](#) and [FlowJo](#), as well as data import as Excel spreadsheets. All R objects generated by the application can be easily re-executed to incorporate more recent modules.

## Conventions Used in This Book

To comment or ask technical questions about this book, please [file an issue](#).

## Development Background

The AnalyCyte web application was developed by the [CiBi](#) platform at the Centre de Recherche en Cancérologie de Marseille ([CRCM](#)).

We extend our sincere thanks to the Institut Français de Bioinformatique ([IFB](#)) for providing the [deployment solutions](#) that have been instrumental in the successful launch and operation of our work.

## Acknowledgments

Thanks to the `projects` library Krieger, Perzynski, and Dalton (2021) work on [Github](#)

# Preface

Technological advancements in cytometry have significantly broadened the spectrum of available markers, enabling the identification and characterization of a diverse array of cellular populations and states. These advancements have facilitated the analysis of variable sample sizes while ensuring standardized procedures.

Cytometry can now measure an extensive panel of markers on each cell in experiments that may involve dozens to hundreds of samples, each containing approximately a million cells. To manage these large data volumes, established methods such as dimension reduction and clustering have been developed. Dimension reduction techniques create a two-dimensional map that optimally synthesizes the information from multiple markers while clustering efficiently organizes cells into distinct groups. These methods are pivotal for the precise analysis and quantification of cell populations, including the abundance and median intensity of each marker.

The insights gleaned from these analyses are vital for comparing different experimental or clinical conditions and for identifying clusters that exhibit significant variations. However, the current identification methods are often limited, relying on statistical software that is not conducive to automation. This limitation complicates the generation of a list of candidate clusters, particularly when establishing a score that optimally combines multiple pieces of information.

In light of these challenges, post-clustering analysis with AnalyCyte becomes a powerful tool for cytometrists, enabling them to work autonomously and efficiently. By providing standardized reports tailored to specific comparisons, cytometrists can independently manage their analyses, fostering a deeper understanding of their data.

As we move towards more complex, high-dimensional data sets, the ability to consistently interpret and compare results becomes increasingly critical. Standardization paves the way for the development of new computational tools and automated analysis methods, which can handle the complexity of modern cytometry data while maintaining the quality and reliability of the results.

# Key Features

- **Data Import:** Analycyte supports the import of formatted output as Excel files, which include extracted features with tables of abundance and MFI (e.g., from [OMIQ](#)), or concatenated FCS files featuring clustering columns (e.g., from [FlowJo](#)). In both cases, it is necessary to load a “metadata” table specifying sample information. It is also possible to load a table describing the cluster descriptors.
- **Exploratory and Supervised Analyses:** Analycyte facilitates both exploratory and supervised analyses of features, allowing for deeper insights into your cytometry data.
- **Standardized and Reproducible Reporting:** Quickly generate standardized and informative reports to ensure consistency and clarity in your analyses.
  - **Quality Control:** Assess the quality of your data and identify any outliers or anomalies.
  - **Differentially Abundant Clusters Between Groups:** Identify clusters or groups of features showing differential abundance across various experimental conditions or groups.
  - **Differential State Analysis:** Analyze changes in the state of features across different conditions or over time.
- **User Interaction Tracking:** The parameters selected for report generation are recorded in .yaml files in each hashed folder.

**i** tdb

Analyses will be exported along with a spreadsheet that describes the steps followed by each user.

- **Access to Template Documents:** Each analysis folder retrieved from the application contains a progs folder with the .qmd template documents used for report creation.
- **Access to Analysis Environment:** Each analysis folder retrieved from the application is accompanied by an .Rproj file.

**Part I**

**User's Guide**

# 1 Installation

The R Shiny application has numerous R dependencies that can be installed on your local machines. For facilitated deployment, a [Git repository](#) is available containing a Docker-based installation recipe. We recommend using Docker, following the [deployment section](#) provided below:

## 1.1 Docker

The current application is provided as a Dockerfile that could be launched on a cloud computing.

There are two stages, the base image creation and the application image creation. The base image freezes a Linux guest with R and all the packages required by the application. Packages installation is done once. It fasten the startup of an image for developing, but also for teaching.

### 1.1.1 Deployment (IFB example)

Login to IFB, select a Shiny Server Machine, then launch it with advanced options.

Set the following configuration:

```
APP_STACK  
SHINY_REPO
```

The cloud computer starts and runs the Dockerfile recipe that is based on the `analcyte_base` image available on DockerHub. This process provides the application as a Web interface.

If you clone this repo, you can add packages to the guest using `apt.txt` and to the R installation using `Install.R`.

### 1.1.2 Development

Start `analcyte` with the base image and a port for shiny web service. The base image is run in interactive mode, otherwise a shiny server is executed (see the `FROM` item). The base image is pulled from DockerHub on the first run because it does not exist in the local host cache.

```
docker run -it -p 3838:3838 eugloh/analcyte_base bash
```

Then install the latest version of the `analcyte` packages.

```
R -e "  
Sys.setenv(R_REMOTES_UPGRADE = 'never') # do not update packages from CRAN  
remotes::install_gitlab('lohmann/analycyte.projects', host = 'https://gitcrcm.marseille.inserm.fr')  
remotes::install_gitlab('lohmann/analycyte.utils', host = 'https://gitcrcm.marseille.inserm.fr')  
remotes::install_gitlab('lohmann/analycyte', host = 'https://gitcrcm.marseille.inserm.fr')  
"
```

Finally, launch the shiny process (with a log file).

```
R -e "options('shiny.port'=3838,shiny.host='0.0.0.0');library(analycyte);analycyte::run_app()"
```

Once done, exit the guest.

```
exit # or Ctrl-D
```

## 2 How to Start

This step-by-step guide will help users get started with the application, including navigation through the interface.

### **i** Note

In the initial version of **Analycyte**, there are no user accounts or login features. These features may be developed in future updates.

[../video/project\\_creation.webm](#)

### 2.0.1 Dashboard Page

The main application page is located under the **Dashboard** tab.

- **Select a folder button:** Users can select the base of an **analycyte-projects** session by choosing **Select folder**.
  - If a project folder exists, it will be opened.
  - If there is no existing project folder, the user will be prompted to fill out a form about the project.

### **i** No information is mandatory.

Users can choose a project description and the name of a sub-folder. If no 'folder name' is specified, the project folder will increment with a name like 'p000X'.

- **Toolbox Navigation:** Users can navigate through the application using the toolbox.

### 2.0.2 The Application left panel

Once an **analycyte-projects** session is launched:

Tab	Description
<b>Choose a Project</b>	
Cell Set	
Feature Set	
<b>Dashboard</b>	Main interface area.
<b>Dataset Tab</b>	Displays the current dataset.

Tab	Description
<b>Export Project</b>	Allows retrieval of the specified folder through ‘Choose a Project’.
<b>Expert View</b>	A dropdown menu that includes various advanced options:
User Actions	Logs user actions with options to annotate or delete steps. Deleting a step with dependencies will automatically delete those dependencies.
Project List	Lists all projects on the same machine, with options to annotate or delete projects. Deleting a project will erase the folder and its subfolders.
Users Info	Contains tables for authors and affiliations (to be determined).
Available Modules	Lists all actions available to the user.
<b>Session Information Tab</b>	Displays the current version of the tool and the global variables in use.
<b>Closing Session</b>	Sessions can be closed using the ‘Close Session’ button.
<b>Deleting Session</b>	A session and all its contained project folders can be deleted by clicking ‘Delete Session’; a confirmation message will be displayed.
<b>reload symbol</b>	

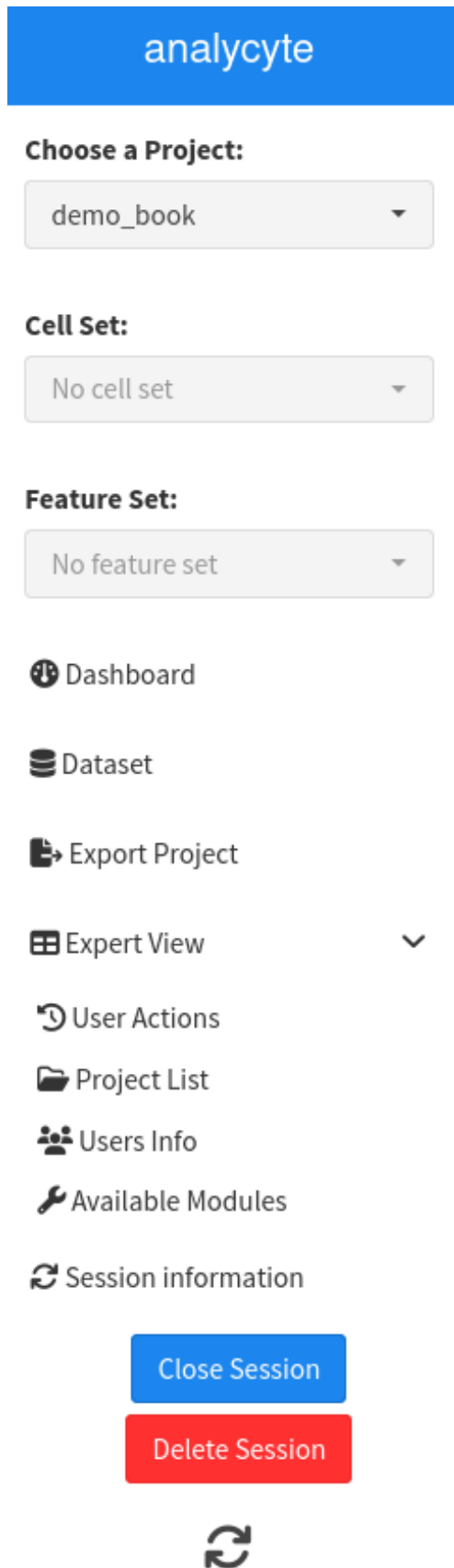
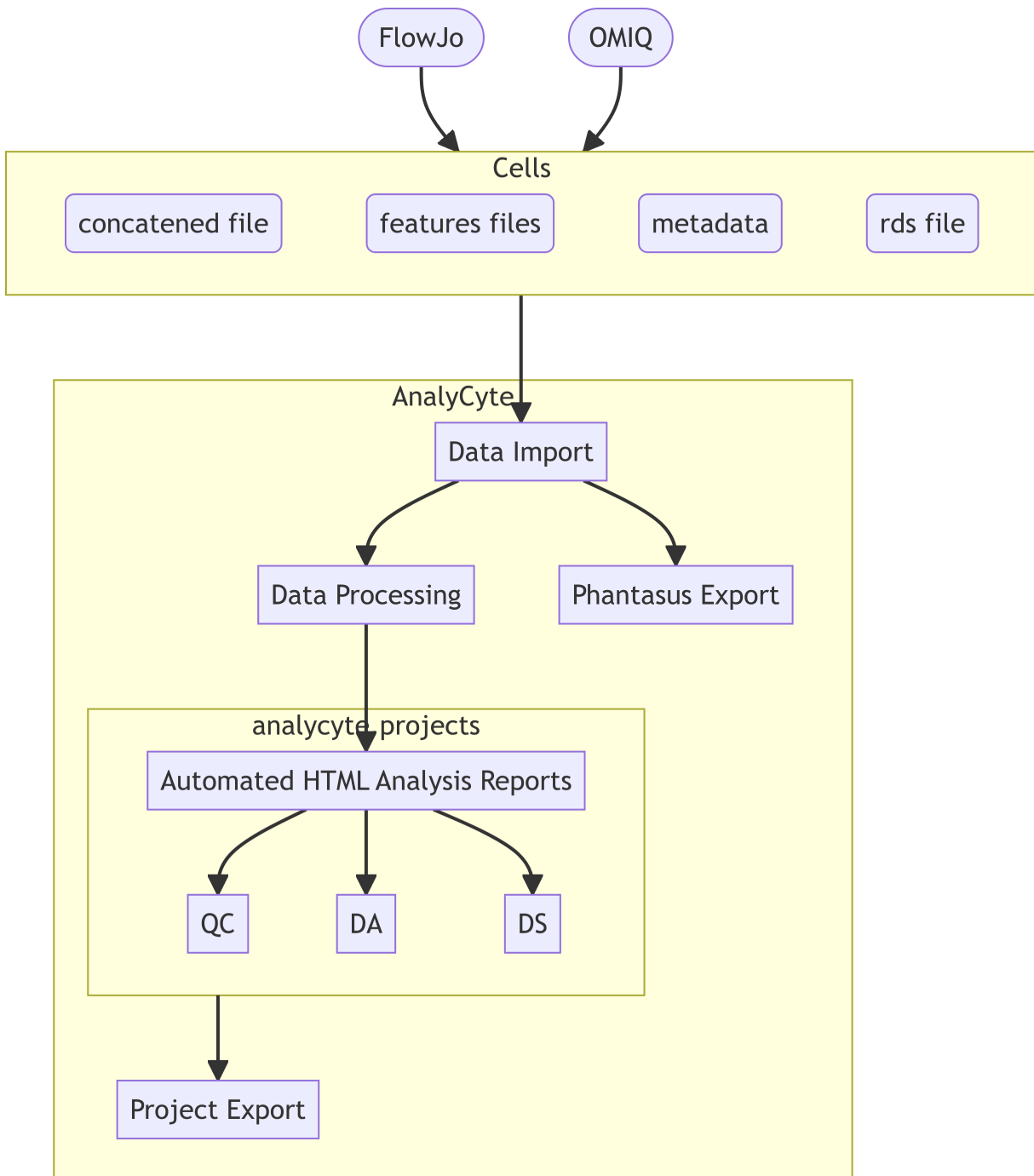


Figure 2.1: left panel

## 3 Workflow

Describe the Shiny framework used in the app and its benefits for creating interactive web applications.



## 4 Module list

List and explanation of the different Shiny modules available in the app, along with their functions and usage.

Table 4.1: Current available modules in app, see [file type](#)

icon	module	description	prerequisite_type	output
	<a href="#">Batch Correction</a>	correct batch effects	2: fb	2: fb
	<a href="#">Extract Clustering Features</a>	extract features from FCS concat with clustering	3: sce	5: fe
	<a href="#">Import Features</a>	import OMIQ FE	0: none	5: fe
	<a href="#">Import Concatenated FCS</a>	import FCS concat	0: none	5: fe
	<a href="#">Quality Control</a>	control quality of features	5: fe	6: ht
	<a href="#">Differential Abundance</a>	analyse abundancy of clusters - univar	5: fe	6: ht
	<a href="#">Differential State</a>	analyse intensity of clusters - univar	5: fe	6: ht
	<a href="#">Import R Object</a>	import .rds object	0: none	5: fe
	<a href="#">Export to Phantasus</a>	export to phantasus	5: fe	4: fil

### 4.1 Batch Correction

**i** `mod_correct_batch`  
tbd

### 4.2 Extract Clustering Features

[../video/v\\_mod\\_extract\\_sce.webm](#)

### **i** mod\_extract\_sce

Functionality Overview: This panel allows users to extract features from a specified cell set. Users can define State Markers and choose their preferred clustering method.

Instructions:

Specify State Markers:

Use the provided input field to specify State Markers.

Select Clustering Method:

Choose a clustering method from the dropdown menu. This selection determines how the cells will be grouped.

Run Feature Extraction:

Once you have specified the State Markers and selected a clustering method, click on the "Extract Features" button to initiate the process.

Additional Information:

Accurate specification of State Markers and the appropriate choice of clustering method are essential for effective feature extraction.

## 4.3 Import Features

[../video/v\\_mod\\_import\\_feature.webm](#)

### **i** mod\_import\_feature

Functionality Overview: The "Import from Feature" panel allows users to import various data tables such as abundance, MFI (Mean Fluorescence Intensity), and metadata tables. Users also have the option to load a marker table. All these tables should be linked by a common ID.

Instructions:

Browse and Select Files:

Click on the 'Browse' button to select the data files you wish to import.

You can import files for abundance data, MFI data, and metadata.

Optionally, you can also load a marker table. (not available yet)

Ensure Common ID Column:

Make sure that each table has a common ID column. This ID is essential for creating the bound between the different tables.

Upload Files:

After selecting the files, complete widgets.

Additional Information:

This import functionality is designed especially for OMIQ exported features.

## 4.4 Import Concatenated FCS

[../video/v\\_mod\\_import\\_concat.webm](#)

### **i** mod\_import\_concat

**Functionality Overview:** This panel allows users to import a concatenated Flow Cytometry Standard (FCS) file representing a complete experiment. Each FCS file within the concatenated file should be annotated and assigned to specific biological samples, studied conditions, etc. Additionally, users must provide a metadata file (in .xlsx or .csv format) containing an FCS ID to identify the origin of cells.

**Instructions:**

**Uploading Files:**

If using this application on a server, first upload your files.

Click on the 'Browse' button (max file size: 500MB) to select and upload your concatenated FCS file and metadata file.

**Selecting FCS Directory:**

After uploading, select the 'uploaded\_fcs' directory containing the fcs file by clicking on 'Select your FCS directory'.

**Metadata File Upload:**

Upload the metadata file that includes the FCS IDs. This file should be in either .xlsx or .csv format.

Ensure that the metadata file correctly maps each FCS ID to its respective sample and condition.

**File Requirements:**

Concatenated FCS file must include properly annotated data with clustering and origine fcs id of each cells.

Metadata file in .xlsx or .csv format, containing FCS IDs for cell origin identification.

**Marker Selection:**

Once the concatenated FCS is loaded, manually select the marker of interest by clicking in the table.

## **4.5 Quality Control**

[../video/v\\_mod\\_qc.webm](#)

### **i** mod\_qc

**Functionality Overview:** The Quality Control Module provides an interactive interface for users to generate customized quality control reports. Users can set parameters, define thresholds, and choose various options for their report.

**Instructions:**

**General Parameters:**

In the ‘General Parameters’ tabset, you can use the widgets to choose general annotations for your report.

**Threshold Settings:**

Navigate to the ‘Threshold’ tabset to set cutoffs for your analysis, like the minimum number of cells for each cluster.

**ON/OFF Options:**

In the ‘ON/OFF’ tabset, use boolean widgets to customize your report.

Options may include showing the code used for analysis or ordering the graphs in a specific manner.

**Run the Report:**

After setting all parameters and options, click on the ”RUN THE REPORT” button located at the top of the page.

This will initiate the generation of your report.

The report will take some time to generate so be patient, it will open in a new window.

## 4.6 Differential Abundance

### **i** mod\_da

**Functionality Overview:** The Differential Analysis Panel is designed for users to conduct comprehensive differential analyses. It includes advanced settings for thresholds, like p-value and fold change, and a dedicated tab for setting univariate parameters to define the differential analysis design.

**Instructions:**

**General Parameters:**

Define the Condition column on which the differential comparison is to be performed.

**Threshold Settings:**

In the ‘Thresholds’ tabset, set critical values for your analysis.

This includes specifying a p-value threshold for statistical significance and a fold change threshold to identify meaningful differences.

**Univariate Parameters:**

Navigate to the new ‘Univariate Parameters’ tab to configure the design of your differential analysis. You will need to confirm the condition the ‘Design columns’ selector.

**Run Analysis:**

After setting all thresholds and univariate parameters, initiate the differential analysis by clicking the ”RUN ANALYSIS” button.

## 4.7 Differential State

### **i** mod\_ds

tbd

## 4.8 Import R Object

[../video/v\\_mod\\_import\\_rds.webm](#)

### **i** mod\_import\_rds

Functionality Overview: This panel enables users to import an existing .rds file that corresponds to a previously processed Flow Cytometry Standard (FCS) file.

Instructions:

Browse File: Click on the 'Browse' button to navigate through your local directory.

Select File: Look for the .rds file you wish to import. These files represent pre-processed FCS data.

Confirmation: Once the file is successfully uploaded, a confirmation message will be displayed.

Next Steps: After importing, you can proceed to new actions in the "Select an action" box.

Additional Information:

This panel is designed to streamline the workflow by allowing easy access to pre-processed data.

## 4.9 Export to Phantasus

### **i** mod\_export\_ph

Functionality Overview: This panel facilitates the export of your data to Excel format for advanced analysis in Phantasus.

Downloadable Excel Files:

MFI Split by Markers: Export the MFI data, providing detailed insights split by each marker.

Median of MFI: Obtain the median MFI values for each FCS file and marker.

Abundance Matrix: Generate an abundance matrix displaying the percentage of cells in each cluster for each FCS file.

## 5 File types terms

Table 5.1: File type description

Term	Description
<b>File type through app:</b>	
0: none	Indicates no file or data input required or expected.
2: fb	File bunch.
3: sce	Stands for ‘SingleCellExperiment’, referring to files formatted to support single-cell data analyses.
4: files	General files, specifics depend on context but typically involve data files necessary for processing, i.e., Excel.
5: fe	Feature element, an R list of abundance and MFI.
6: html	HTML files, used for output that requires web viewing or reporting.

## **Part II**

# **Import**

## 6 Import Concatenated FCS

The app supports the use of a single FCS file containing all the cells, along with additional columns for initial FCS ID, clustering, and possible dimension reduction information (e.g., t-SNE, UMAP). This file will be linked to a metadata file generated by the user, which contains FCS ID information to identify the different initial sample IDs and groups of interest.

To load and concatenate data in the app, follow these steps:

- Upstream analysis:
  - **Prepare your FCS file:** Ensure that your FCS file contains all the cells and the necessary additional columns for initial FCS ID, clustering, and possible dimension reduction information.
  - **Create a metadata file:** Generate a metadata file that contains FCS ID information to identify the different initial sample IDs and groups of interest. This file should be in a compatible format, such as CSV or Excel.
- Through app:
  - Use the **Import Concatenated FCS** module. Ensure that the files are correctly linked during the import process

### 6.1 Clustering selection

- Through app:
  - Once the data is loaded, use the app's concatenation feature **Extract Clustering Features** to combine the data from the different samples and groups. This will create a single dataset that can be used for further analysis and visualization.

## 7 Import Features tables

- Upstream analysis:
  - [Prepare your csv files](#)
- Through app:
  - Use the [Import Concatenated FCS](#) module.

## 8 Import R files

### **i** Note

An dataset can be loaded into a new project using the [Import R Object](#) module, by selecting a `.rds` file.

## **Part III**

# **Export**

## 9 Export Analysis to disk

### **i** Export table

A table allows the user to download the `<project>` project folder `analycyte-projects` to retrieve related analysis files. All quarto `.qmd` templates of generated reports are stored in a **progs** folder.

In each folder (`<hash>`) a yaml file `.yaml` contains the report parameters.

## 10 Phantasm Table

### **i** Phantasm export module

The phantasm import panel has been designed to enable users of the cytometry platform to easily retrieve values to be represented on the [phantasm website](#)  
tbd's objective is to facilitate automatic annotation of metadata in columns and/or rows

**Part IV**

**Content Exemple**

# **11 Dataset description**

ChikH dataset description Michlmayr et al. (2018)

# 12 Recurrent element

Recurrent graphs

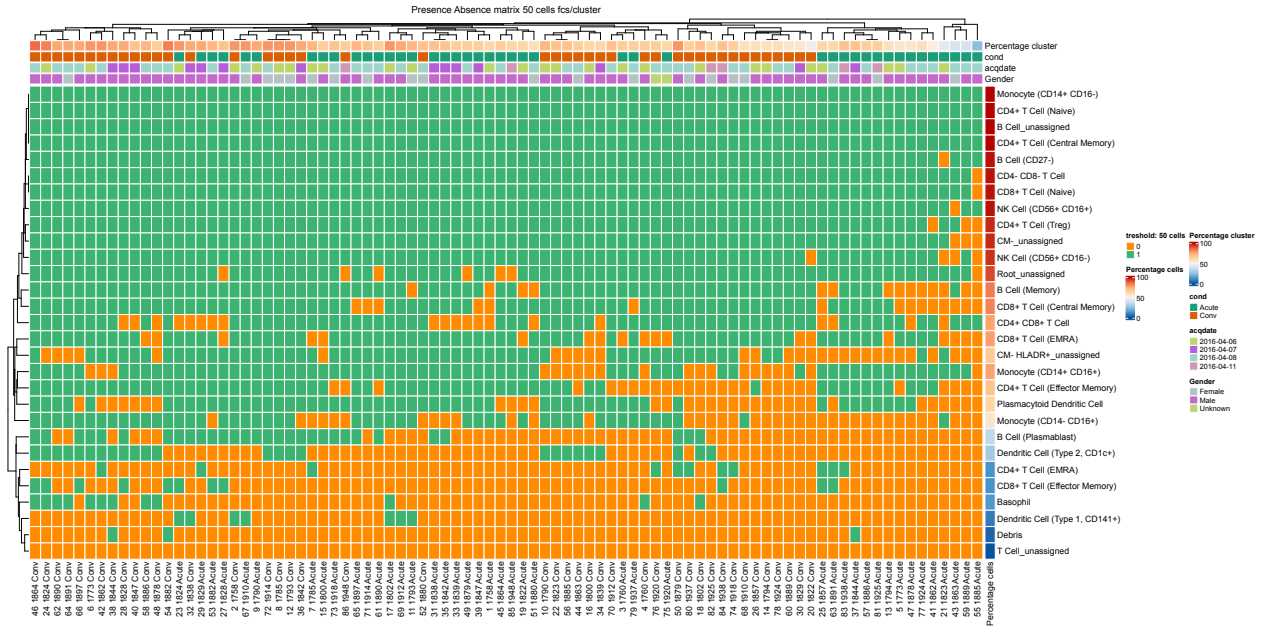


Figure 12.1: Heatmap Cell presence / absence

# 13 QC - control quality of features

## 13.1 Counts

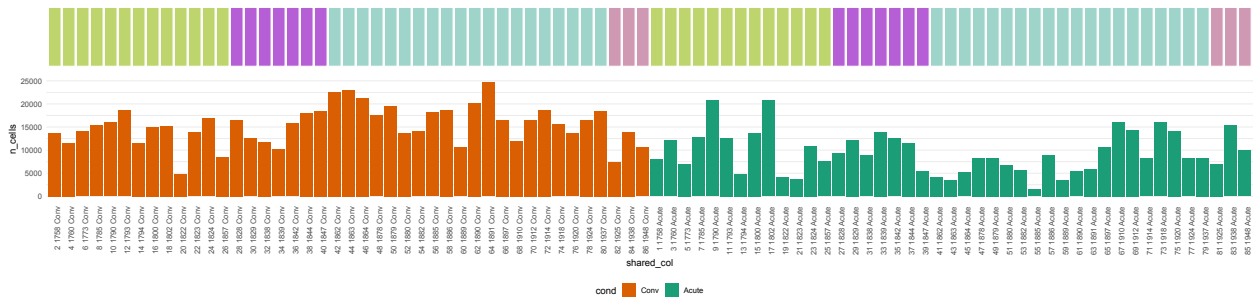


Figure 13.1: Number of cell per fcs

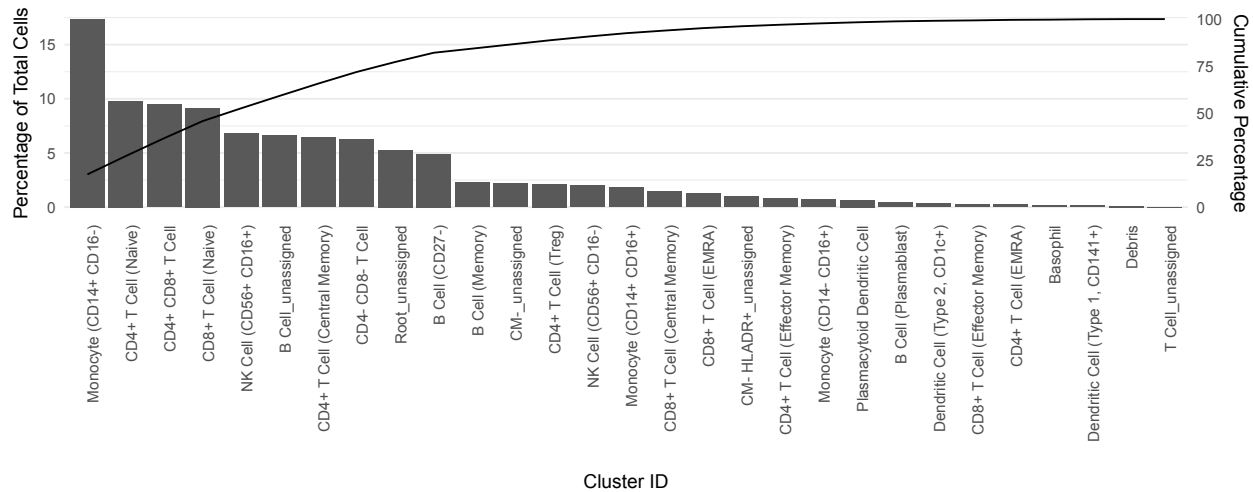


Figure 13.2: Number of cell per clusters/gating

## 13.2 Density heatmap

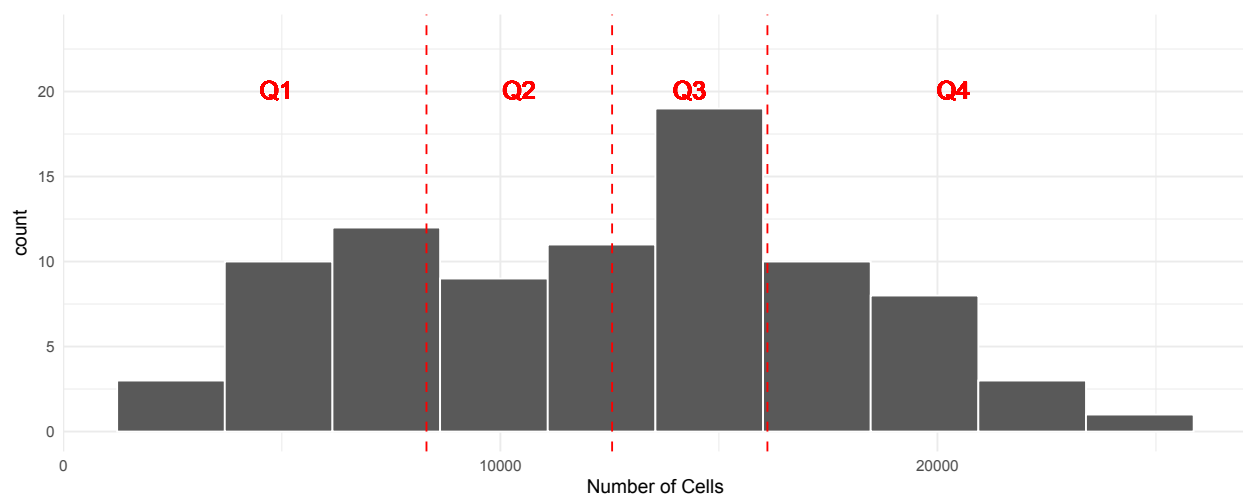


Figure 13.3: Histogram

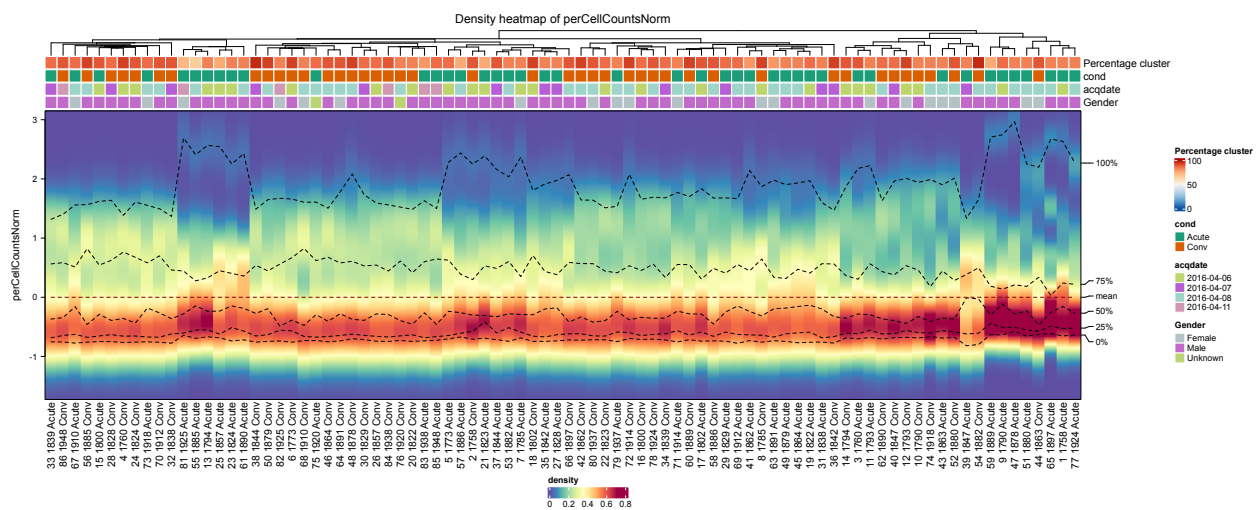
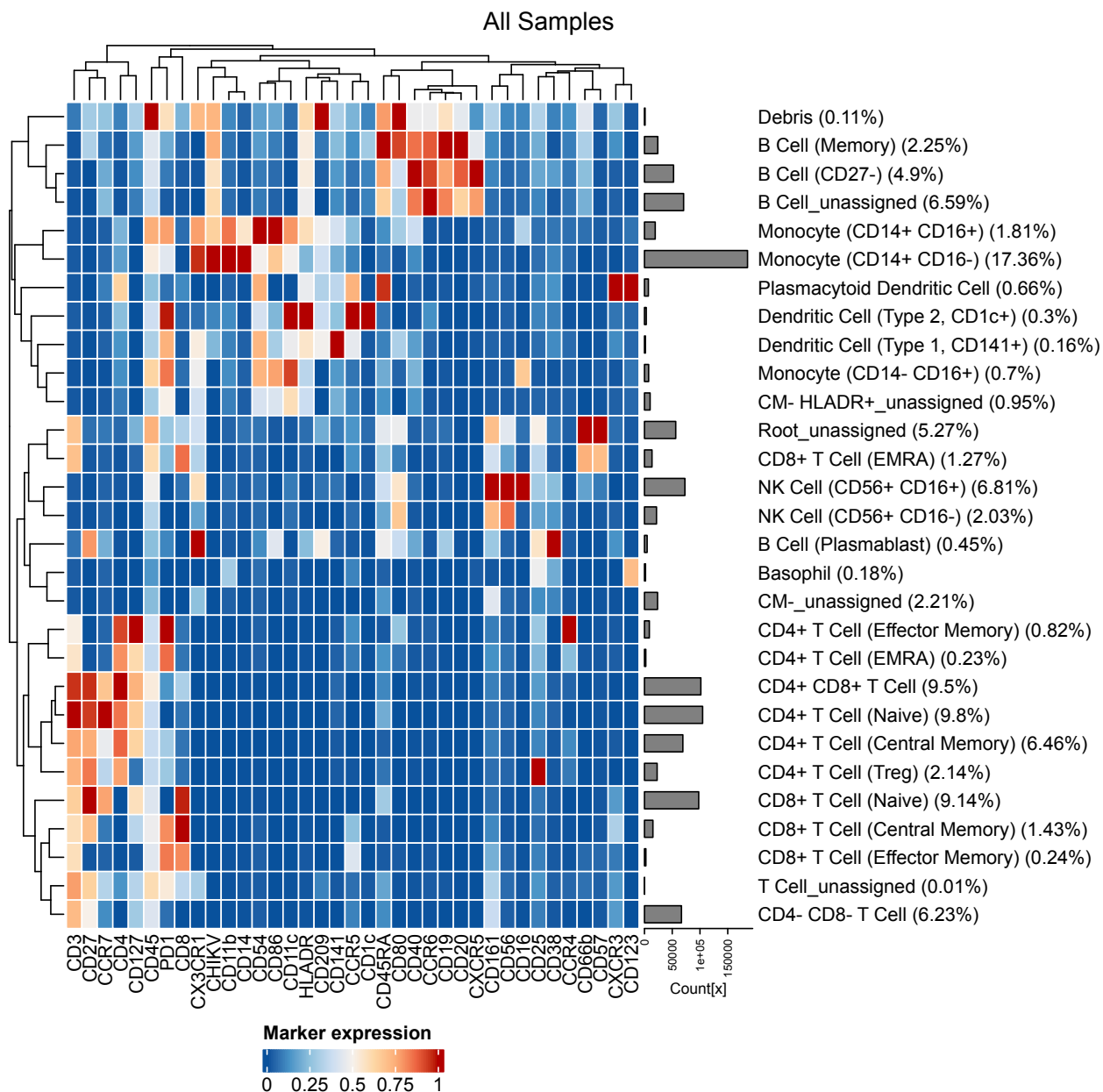
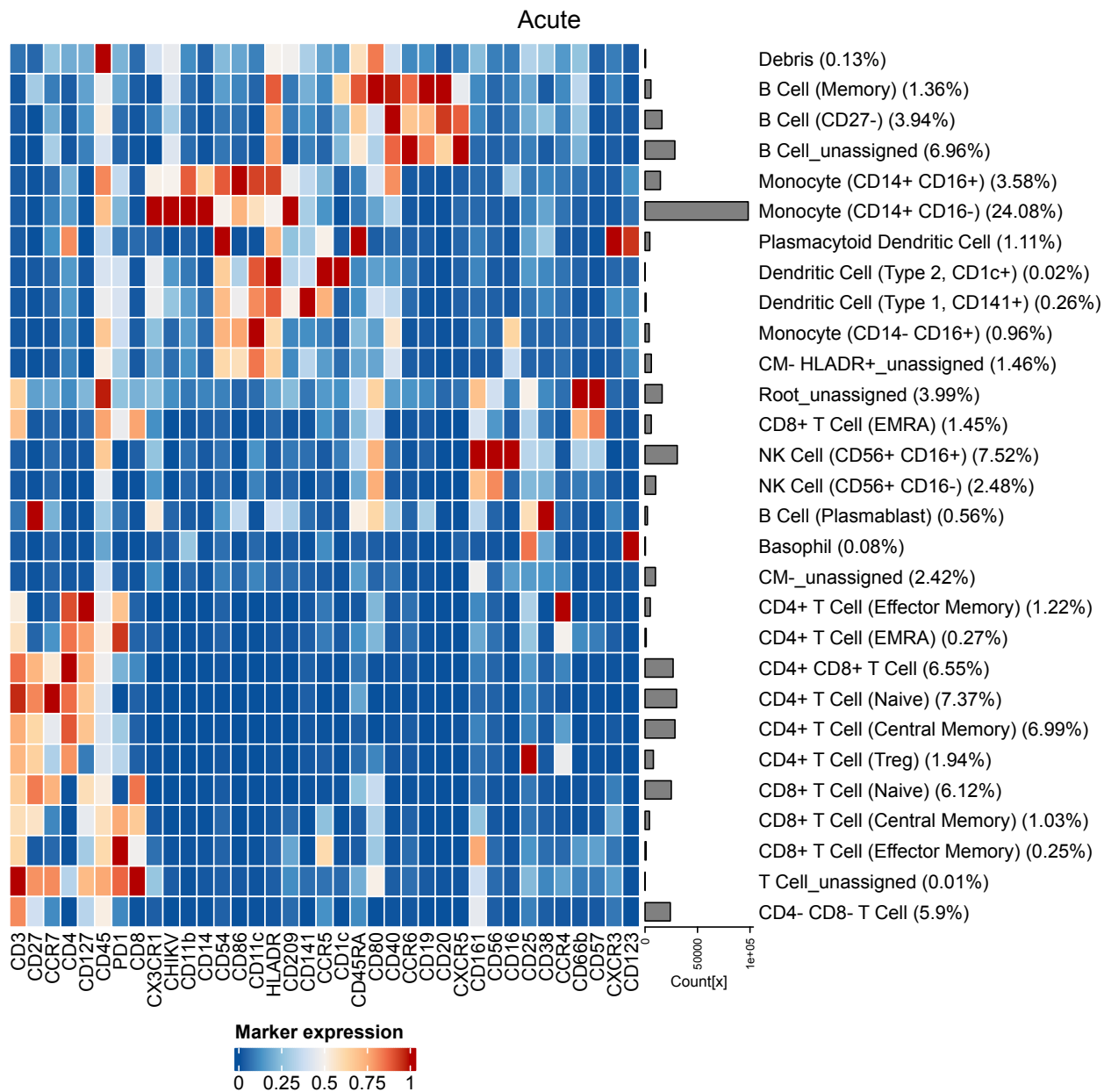
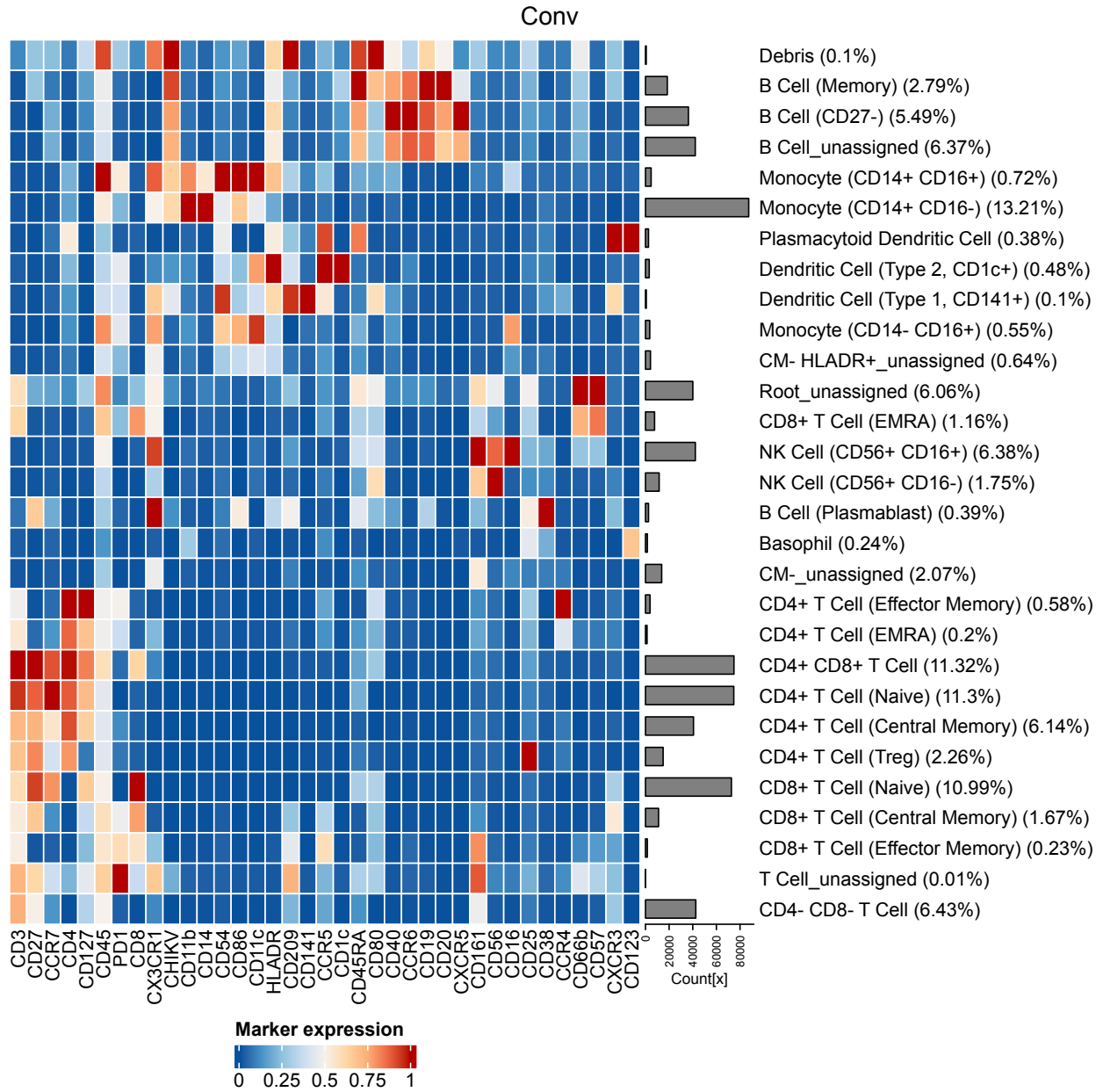


Figure 13.4: Density

# 14 MFI heatmap checks







# 15 MFI / abundance heatmap

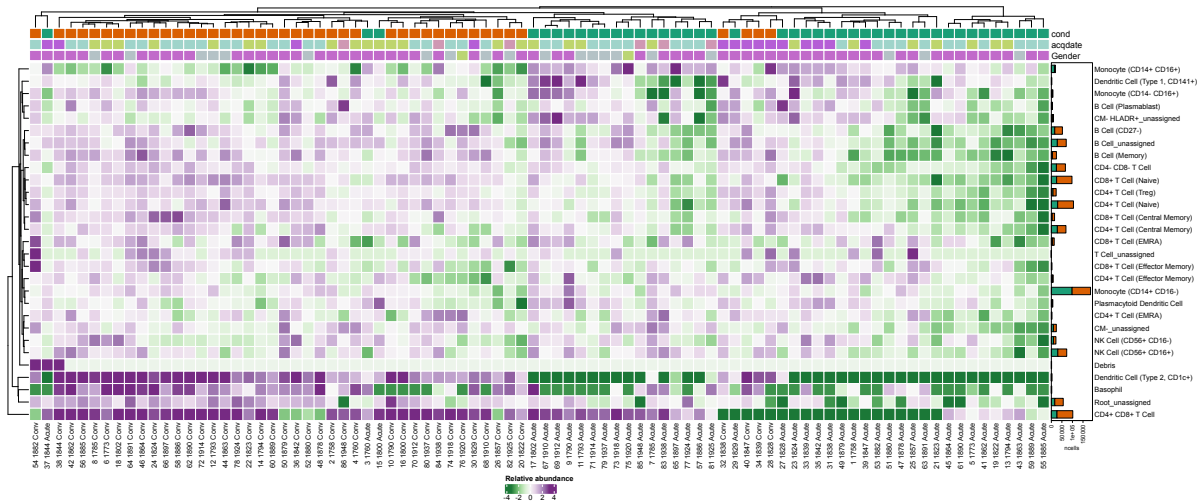
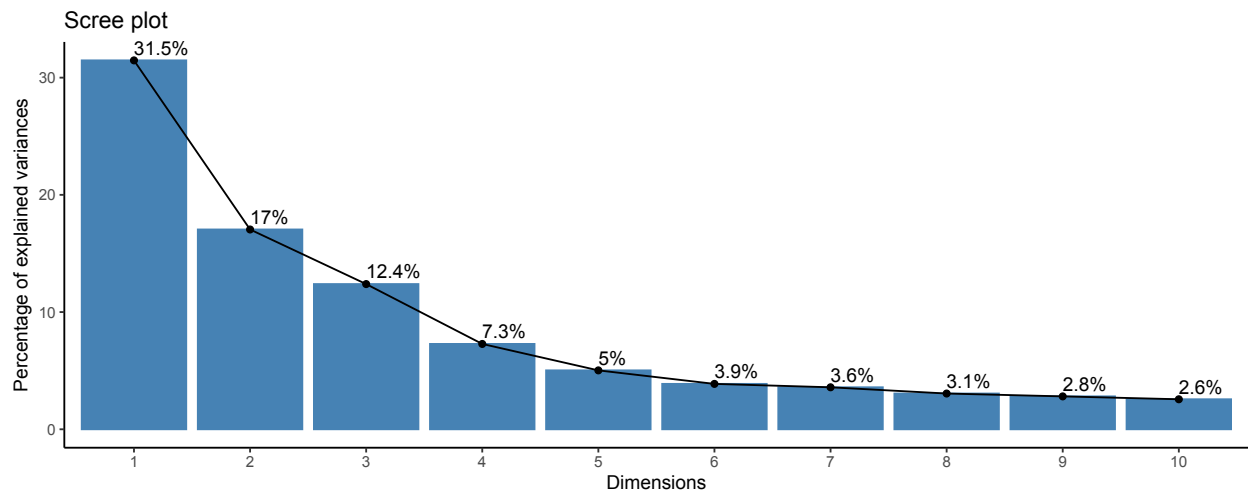


Figure 15.1: MFI / abundance heatmap

# 16 Abundance heatmap

## 16.1 PCA

### 16.1.1 Abundance



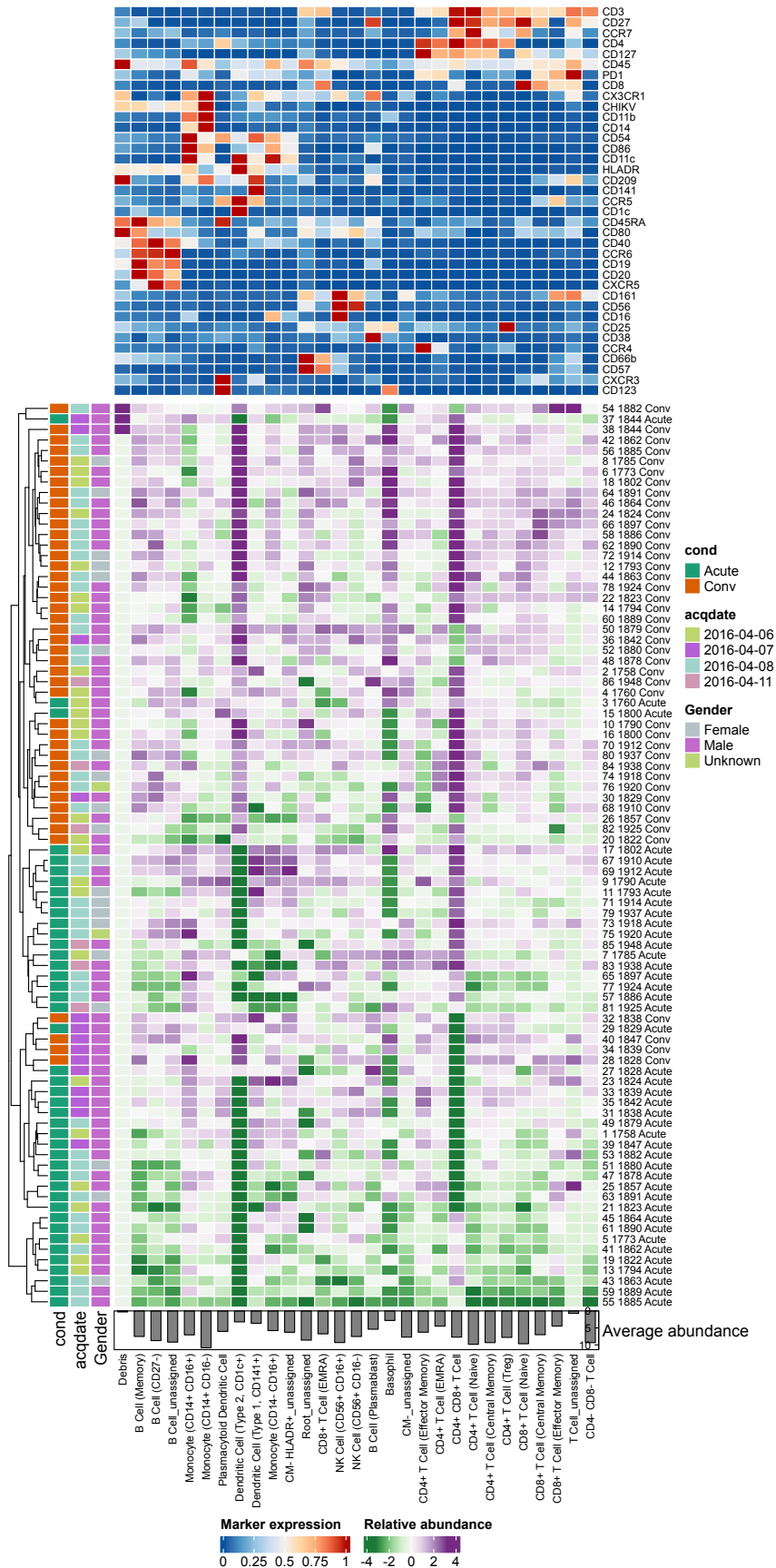
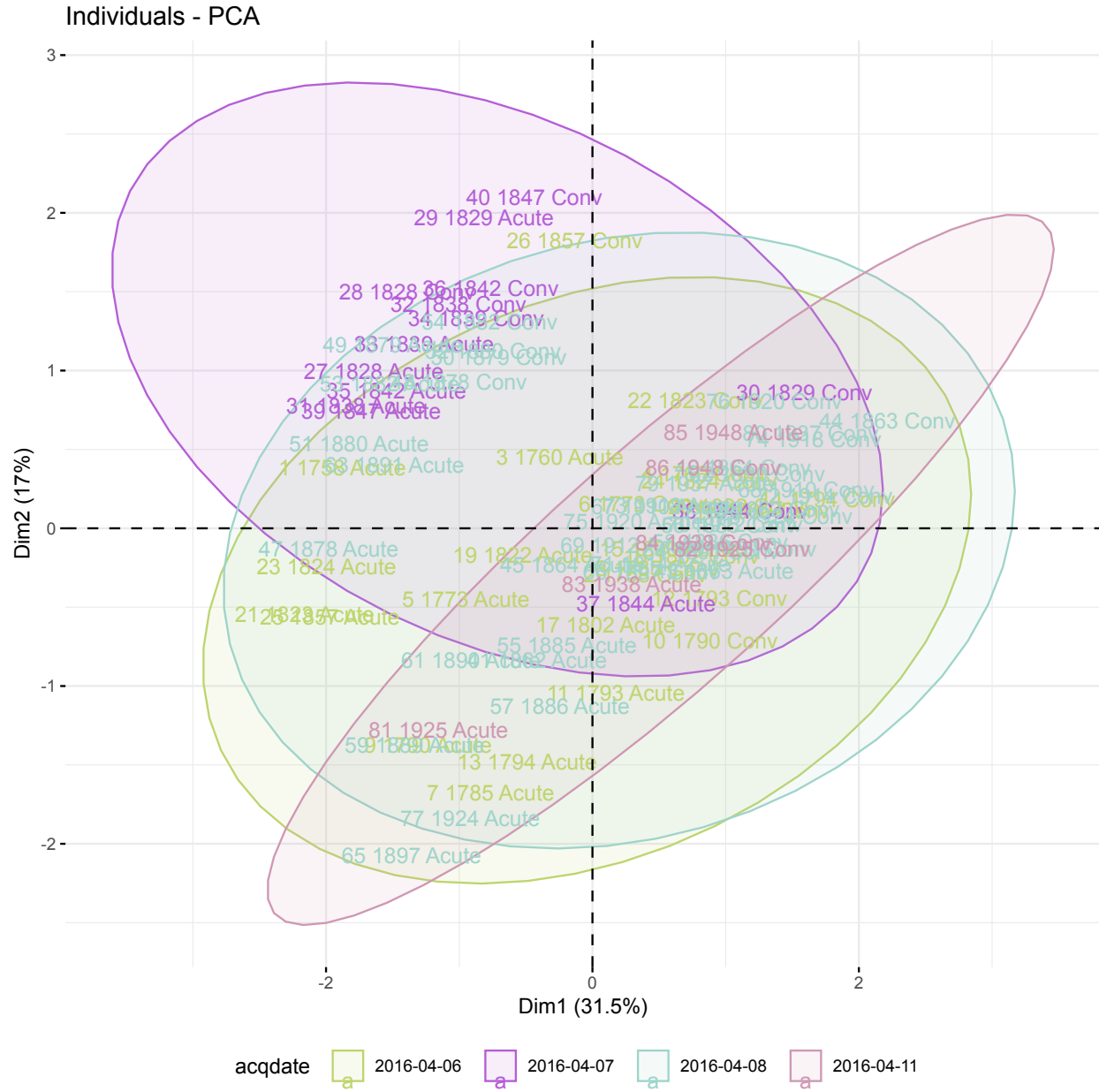
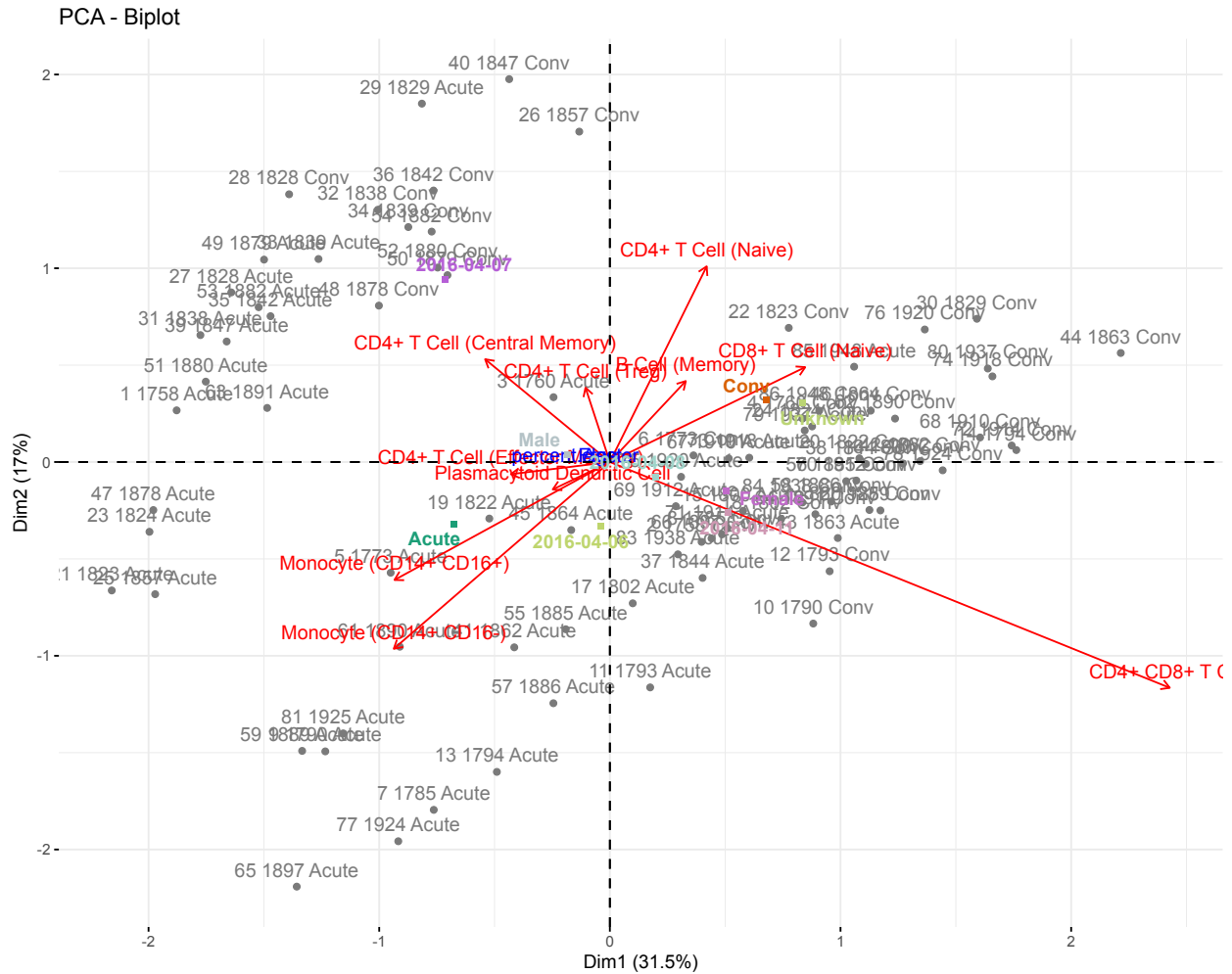


Figure 16.1: Abundance heatmap

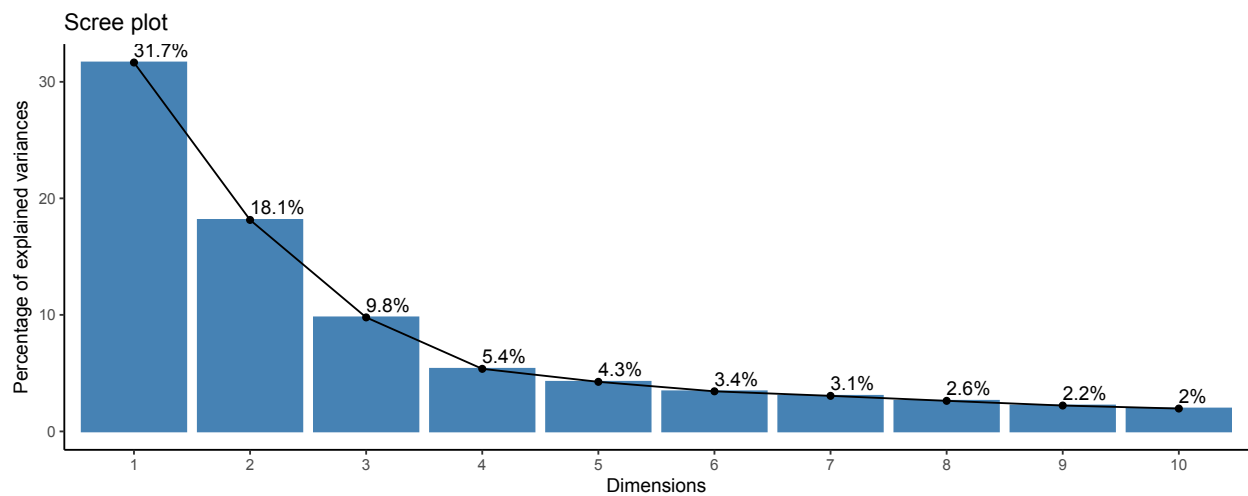


A PCA plot showing the separation of 180 samples (180 patients) based on their condition (Acute vs. Convalescent). The x-axis is labeled 'Dim1 (31.5%)' and ranges from -2 to 2. The y-axis is labeled 'Dim2 (17%)' and ranges from -2 to 2. The plot is divided into two main regions by a vertical dashed line at Dim1 = 0. The left region (Dim1 < 0) is shaded light blue and contains samples labeled 'Acute'. The right region (Dim1 > 0) is shaded light orange and contains samples labeled 'Conv'. Each sample is represented by a point labeled with a number and a condition (e.g., '1828 Acute', '1829 Conv'). The points are clustered within their respective regions, indicating a clear separation between the two conditions. A legend at the bottom indicates 'cond' with 'Acute' in blue and 'Conv' in orange.

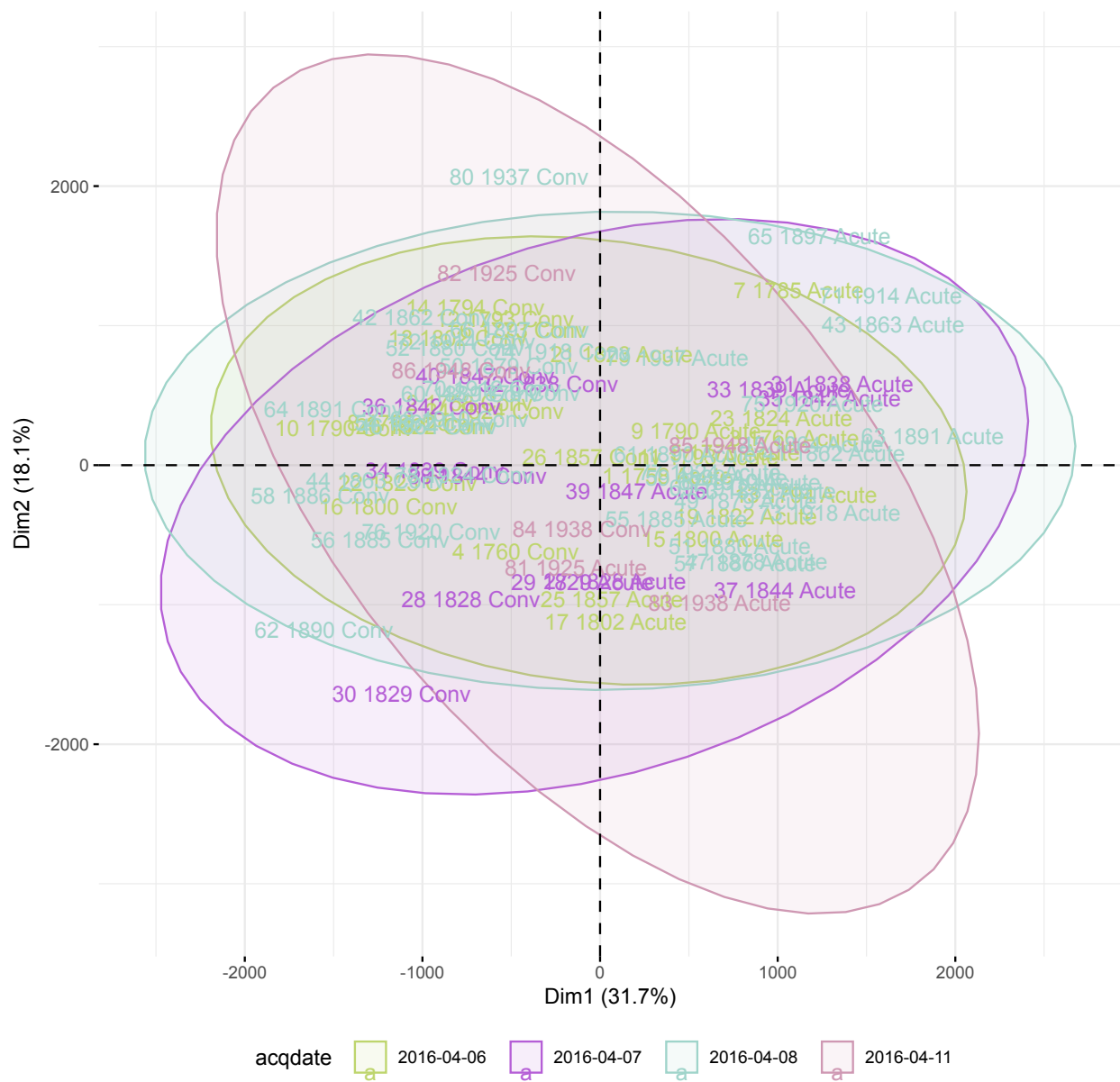


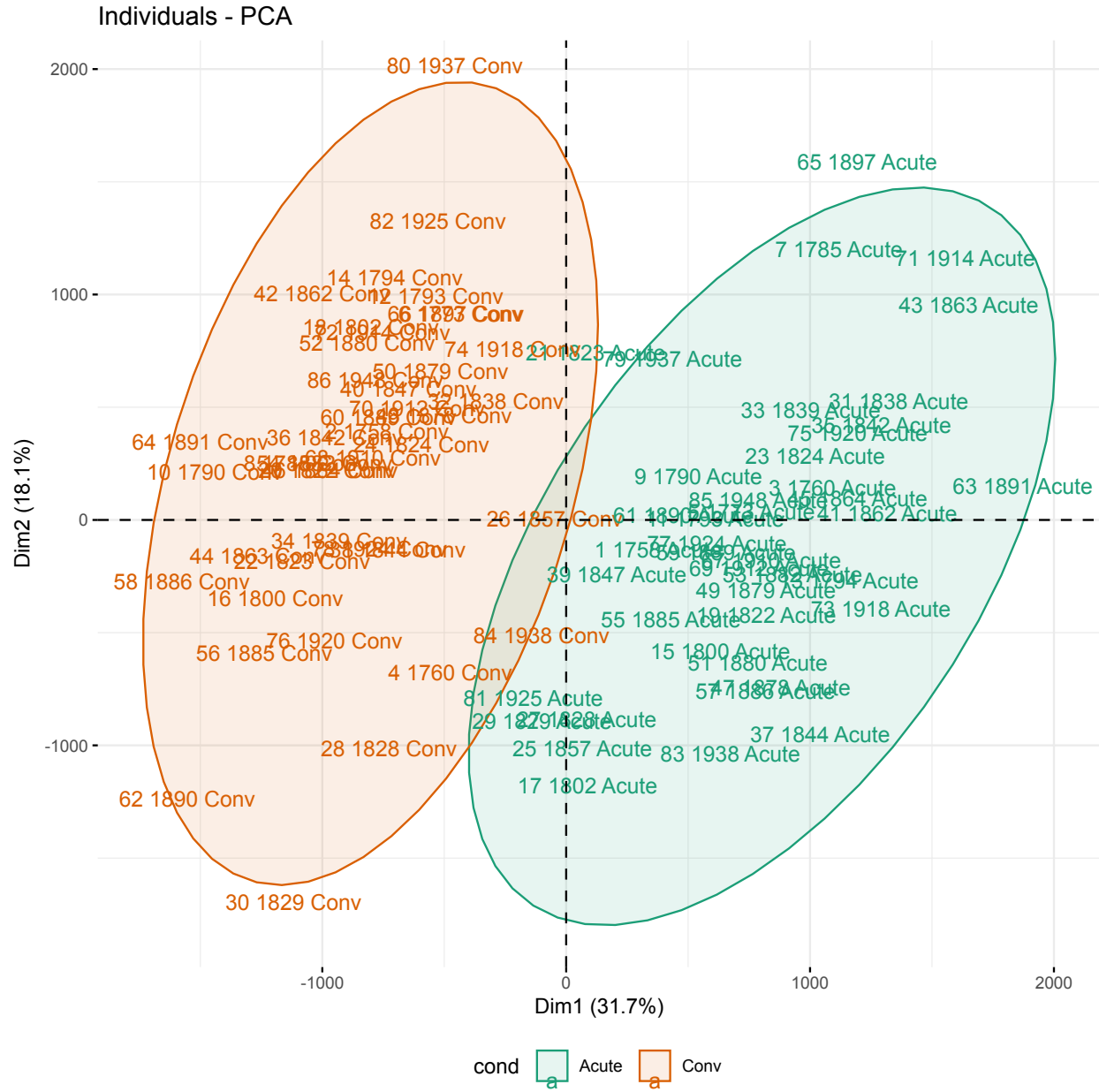


## 16.1.2 MFI

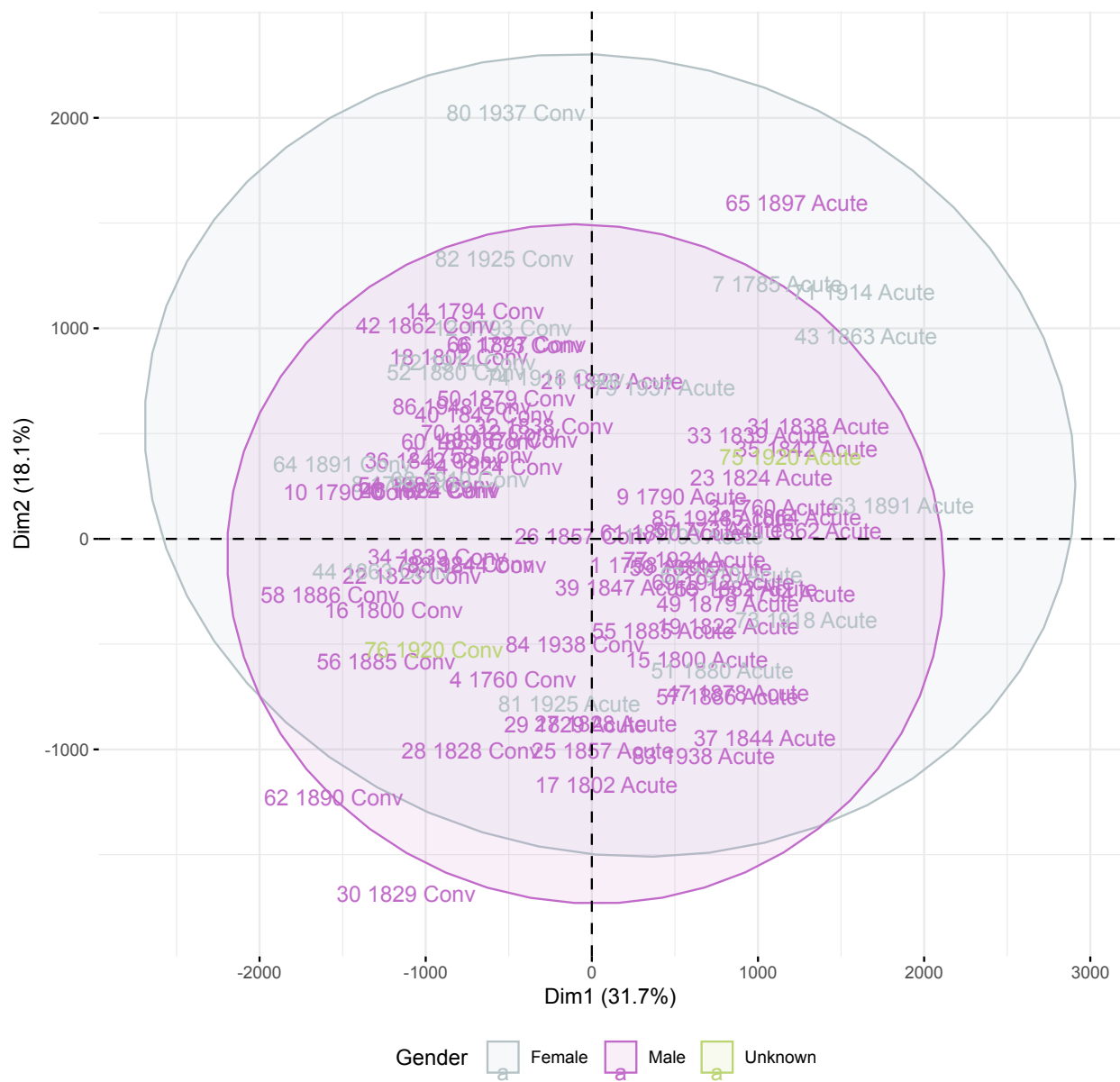


# Individuals - PCA





Individuals - PCA







# 17 DA - analyse abundancy of clusters - univar

## 17.0.1 DA Tables Columns

### **i** Diffcyt methods for differential abundance

The output tables are derived from methods within the diffcyt Weber et al. (2019) package. edgeR outputs is generated by `diffcyt-DA-edgeR`, using the edgeR package Robinson, McCarthy, and Smyth (2010) . VOOM outputs come from `diffcyt-DA-voom`, based on the limma package's VOOM method Ritchie et al. (2015) . GLMM outputs are produced by `diffcyt-DA-GLMM`, employing a GLMM approach detailed by Nowicka et al. (2017).

---

Column

Name    Description

---

**cluster\_id** Unique identifier for each cluster analyzed, linking the data to specific cell populations within the study.

**reference** Name of the reference group or condition used for comparison in the differential abundance analysis.

**contrast** Name of the contrast group or condition compared against the reference.

**mean**    Average percentage of total cells or absolute cell counts in the reference group, reflecting their abundance.

**reference**

**mean**    Average percentage of total cells or absolute cell counts in the contrast group, reflecting their abundance.

**contrast**


**logFC**    Log2 fold change of the mean abundances from the contrast group relative to the reference group. Positive values indicate higher abundance in the contrast group; negative values suggest higher abundance in the reference group.

**explicit** The explicit fold change, often provided directly or calculated as the antilog of logFC, showing the actual change in cell abundance between the contrast and reference groups.

**FC**

**p\_val**    P-value from the statistical test used to determine the significance of the difference in cell abundance between the contrast and reference groups. A lower p-value indicates a more statistically significant difference.

**p\_adj**    Adjusted p-value which accounts for multiple testing corrections, providing a more stringent significance assessment.

Column Name	Description
<b>FDR</b>	False Discovery Rate: Adjusted p-value that accounts for multiple comparisons, representing the expected proportion of falsely rejected hypotheses.
<div>  <b>FDR assumptions</b> <p>Noting that while FDR is theoretically a complex correction, it is used here as a sorting tool under the assumption of similar study group sizes.</p> </div>	
<b>logCPM</b>	Log2 of the counts per million, showing the log-transformed abundance level, normalized by total counts to facilitate comparisons between samples of different sizes.
<b>LR</b>	Likelihood ratio statistic from the statistical test, measuring how well the model fits the data with versus without the variable of interest.
<b>AveExpr</b>	Represents the average abundance of the cells within each cluster across all samples. This measure indicates the general level of cell presence for each cluster in the dataset, providing a baseline for comparing changes in specific groups.
<b>t</b>	The t-statistic from the differential abundance analysis. This statistic assesses the magnitude of the difference in cell abundance relative to the variability observed among samples. A higher absolute value of the t-statistic indicates a more statistically significant difference.
<b>B</b>	The log-odds that the cell population is differentially abundant across the compared groups. A higher B value suggests stronger evidence for differential abundance, combining both the size of the effect and the consistency of this effect across samples.

## 17.0.2 Interactive volcano and abundance plot

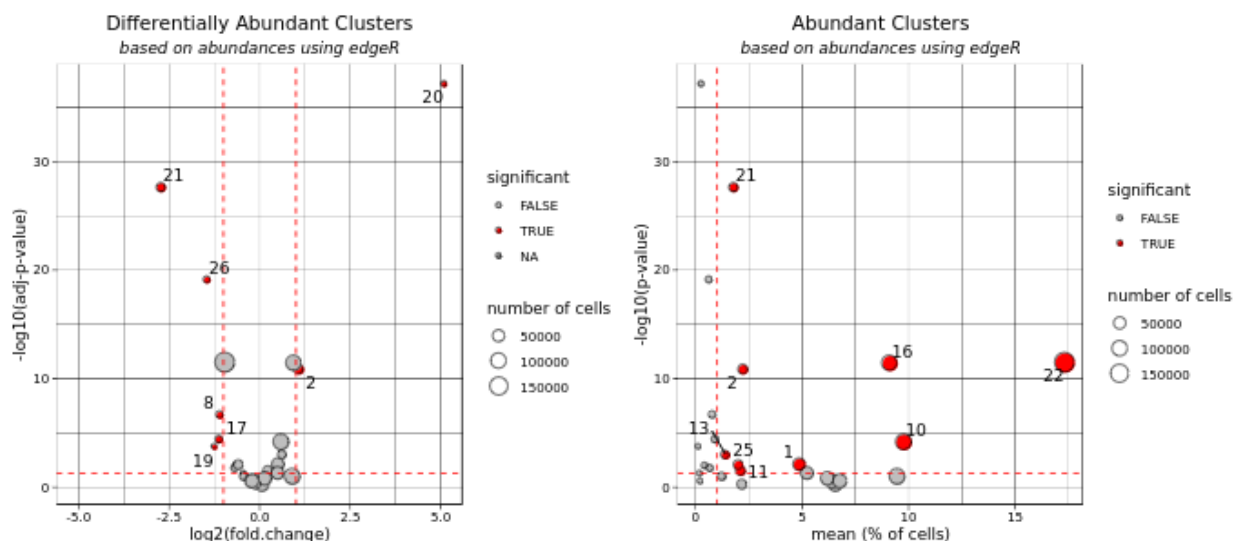


Figure 17.1: Interactive volcano and abundance plot

### 17.0.2.1 volcano plot

This interactive volcano plot is designed to visualize differential abundance comparisons (DAC) across clusters. It plots the  $\log_2(\text{Fold Change})$  on the x-axis against the negative  $\log_{10}$  of the adjusted p-value on the y-axis. The size of each dot on the plot represents the number of cells associated with each cluster, providing a visual scale of abundance.

Significant differences are emphasized with red dashed lines, and the thresholds for these are determined by:

An absolute  $\log_2(\text{Fold Change})$  greater than a specific cutoff ( $\langle \text{cutoff FC} \rangle$ ), An adjusted p-value below a certain threshold ( $\langle \text{cutoff FDR} \rangle$ ). Users can interact with the plot by hovering their mouse over points to overlay additional annotations.

### 17.0.2.2 abundance plot

This plot illustrates the abundance of cells per cluster, focusing on those associated with low adjusted p-values. The x-axis displays the mean percentage of total cells, with clusters exceeding 1% emphasized. The y-axis shows the negative  $\log_{10}$  of the adjusted p-value.

Each dot's size corresponds to the number of cells in the cluster, making it easy to see which clusters have more cells. Clusters meeting specific significance criteria are highlighted in red, according to:

A mean percentage of total cells greater than 1%, An adjusted p-value below a defined threshold ( $\langle \text{cutoff FDR} \rangle$ ). Additionally, when users hover over a dot, the corresponding cluster's data is simultaneously highlighted on this abundance plot and the related volcano graph. This interactive feature aids in correlating significant changes in abundance with their statistical significance across both visual representations.

### 17.0.3 violin plot

This violin plot is designed to illustrate the relative abundance of cells across different experimental groups, with dynamic customization options that adjust to user-defined conditions ( $\langle \text{condition} \rangle$ ) and batch settings ( $\langle \text{batch} \rangle$ ). The plot is marked by a distinctive red dashed line at the zero abscissa.

Each violin's color corresponds to a specific condition, facilitating quick visual comparisons among groups. The shape of the data points within each violin can vary, reflecting the specified batch conditions, which helps in assessing the impact of batching on the cell abundance.

Significance annotations are embedded directly within the plot, categorizing the statistical significance of differences observed between groups based on the adjusted p-values (FDR). The levels of significance are visually encoded as follows:

- \*\*\*\*: Highly significant ( $\text{FDR} \leq 0.0001$ )
- \*\*\*: Very significant ( $0.0001 < \text{FDR} \leq 0.001$ )
- \*\*: Significant ( $0.001 < \text{FDR} \leq 0.01$ )
- \*: Moderately significant ( $0.01 < \text{FDR} \leq 0.05$ )
- .: Suggestive ( $0.05 < \text{FDR} \leq 0.1$ )

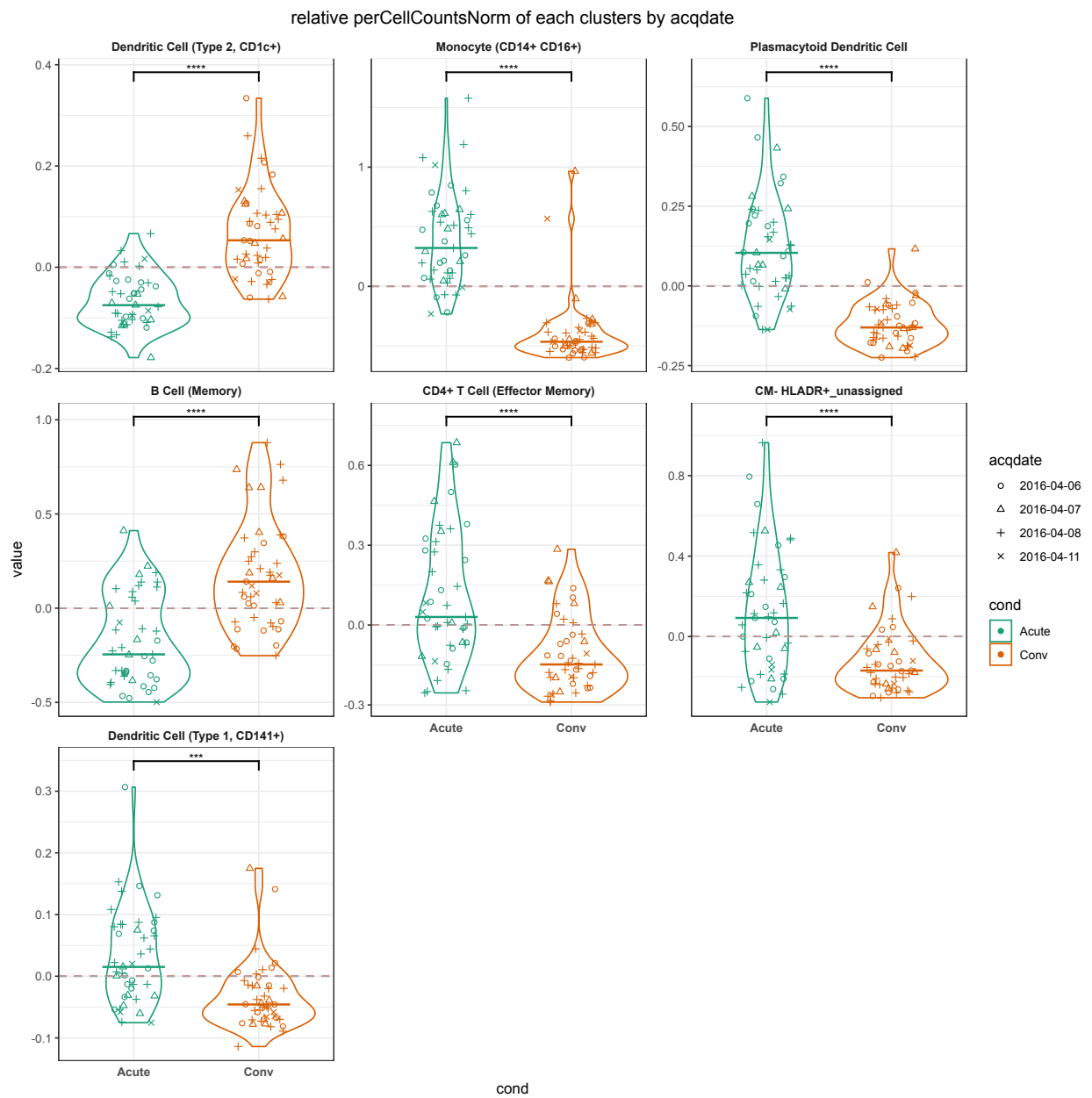


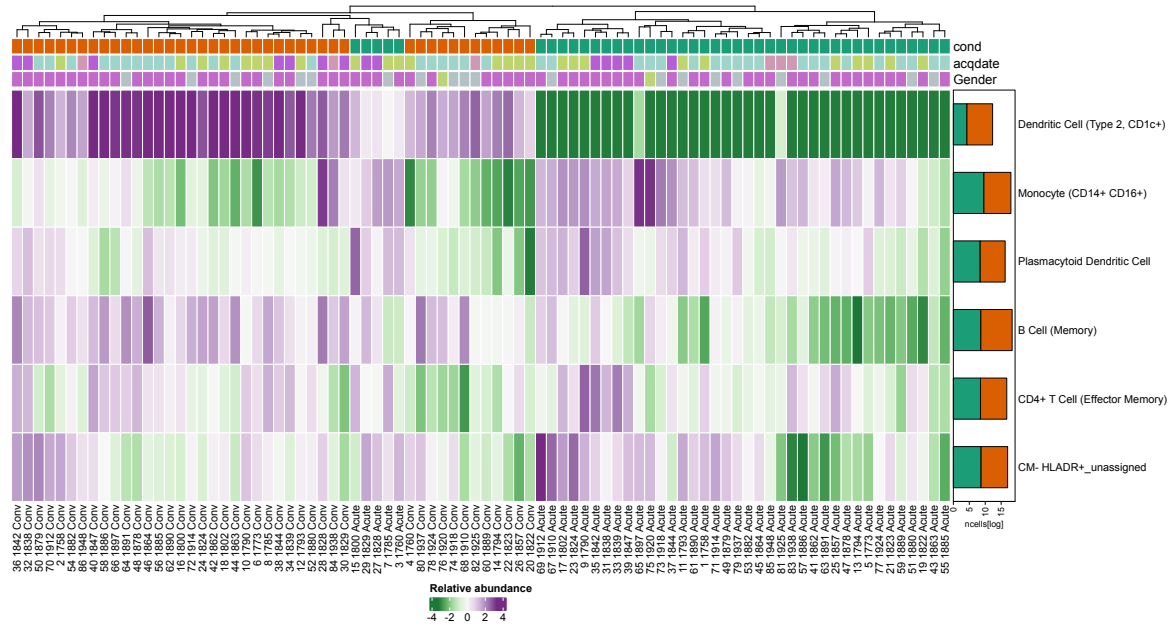
Figure 17.2: violin plot

### 17.0.3.1 Additional Violin Plots for Significant Contrasts

If more than one group contrast shows significant results based on the false discovery rate (FDR) and log fold change (logFC), an extra violin plot is created. This plot only includes groups that meet these significance and effect size thresholds. It helps highlight the most important differences between groups clearly and quickly.

## 17.1 Consensus results

### 17.1.1 Arcsinh Transformed Frequencies



heatmap abundance



# 18 DS - analyse intensity of clusters - univar

## 18.0.1 DS Tables Columns

**i** Diffcyt methods for differential state

## 18.0.2 Interactive volcano plot

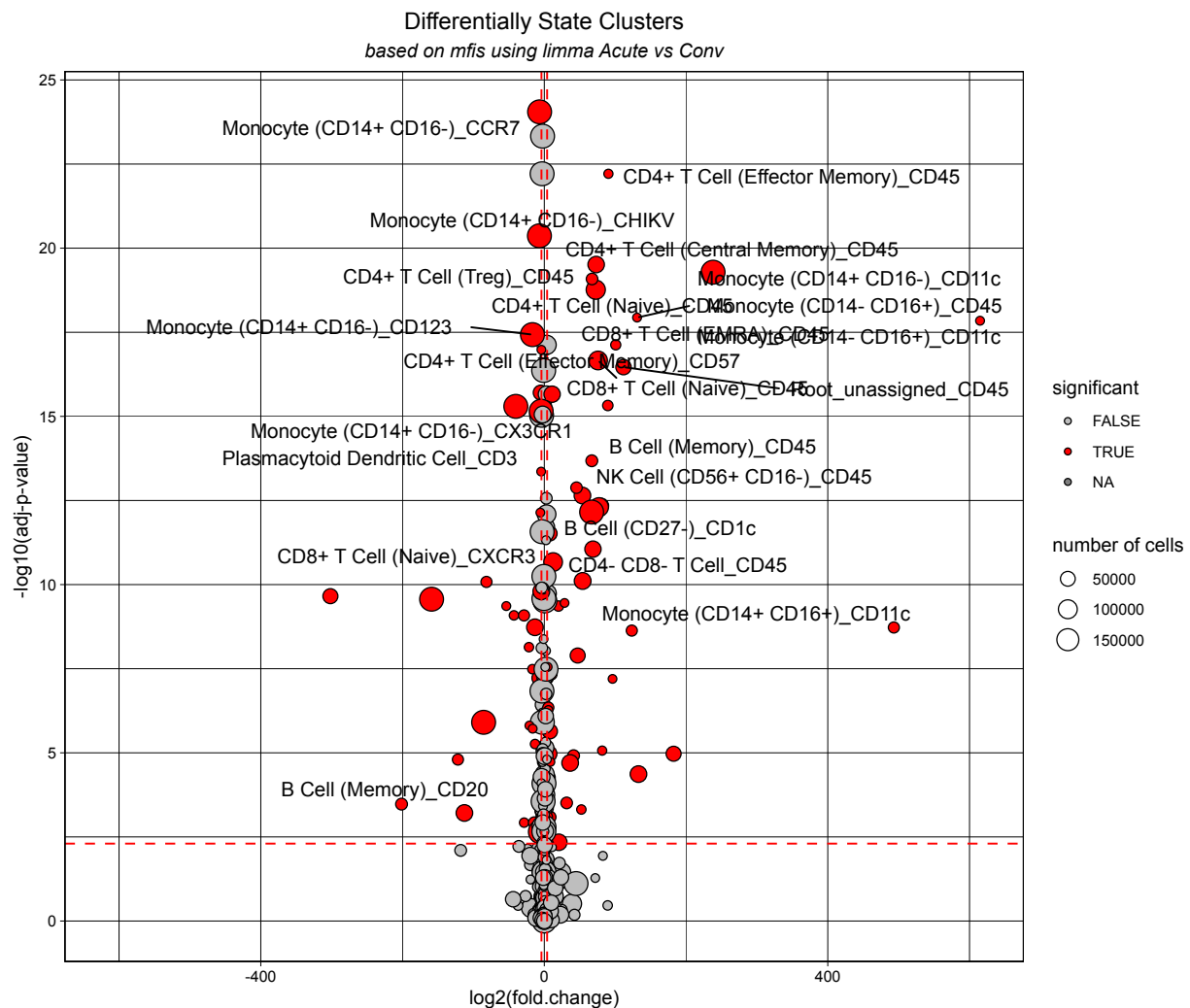


Figure 18.1: Interactive volcano plot

This interactive volcano plot is designed to visualize differential state comparisons (DS) across clusters. It plots the  $\log_2(\text{Fold Change})$  on the x-axis against the negative  $\log_{10}$  of the adjusted p-value on the y-axis. The size of each dot on the plot represents the number of cells associated with each cluster\*marker.

Significant differences are emphasized with red dashed lines, and the thresholds for these are determined by:

An absolute  $\log_2(\text{Fold Change})$  greater than a specific cutoff ( $<\text{cutoff FC}>$ ), An adjusted p-value below a certain threshold ( $<\text{cutoff FDR}>$ ). Users can interact with the plot by hovering their mouse over points to overlay additional annotations.

### 18.0.3 violin plot

This violin plot is designed to illustrate the scaled mfi of marker\*cluster across different experimental groups, with dynamic customization options that adjust to user-defined conditions ( $<\text{condition}>$ ) and batch settings ( $<\text{batch}>$ ). The plot is marked by a distinctive red dashed line at the zero abscissa.

Each violin's color corresponds to a specific condition, facilitating quick visual comparisons among groups. The shape of the data points within each violin can vary, reflecting the specified batch conditions, which helps in assessing the impact of batching on the cell abundance.

Significance annotations are embedded directly within the plot, categorizing the statistical significance of differences observed between groups based on the adjusted p-values (FDR). The levels of significance are visually encoded as follows:

- \*\*\*\*: Highly significant (FDR 0.0001)
- \*\*\*: Very significant ( $0.0001 < \text{FDR} \leq 0.001$ )
- \*\*: Significant ( $0.001 < \text{FDR} \leq 0.01$ )
- \*: Moderately significant ( $0.01 < \text{FDR} \leq 0.05$ )
- .: Suggestive ( $0.05 < \text{FDR} \leq 0.1$ )

### 18.0.4 MFI Marker\*Cluster Heatmap

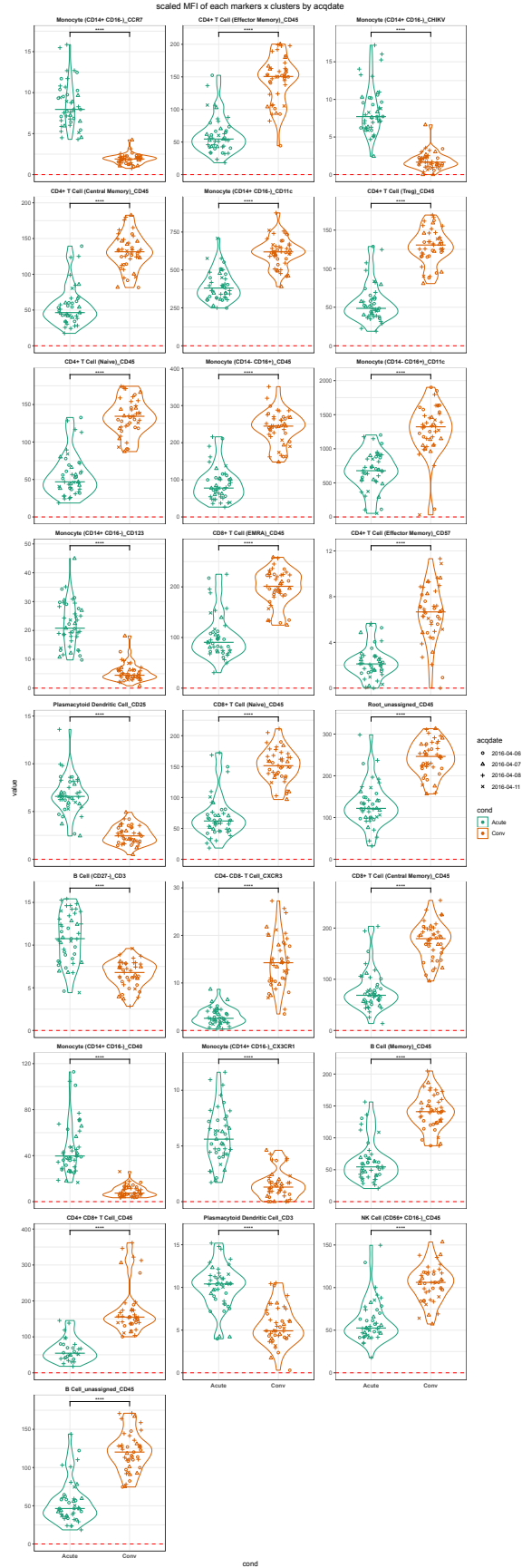


Figure 18.2: violin plot

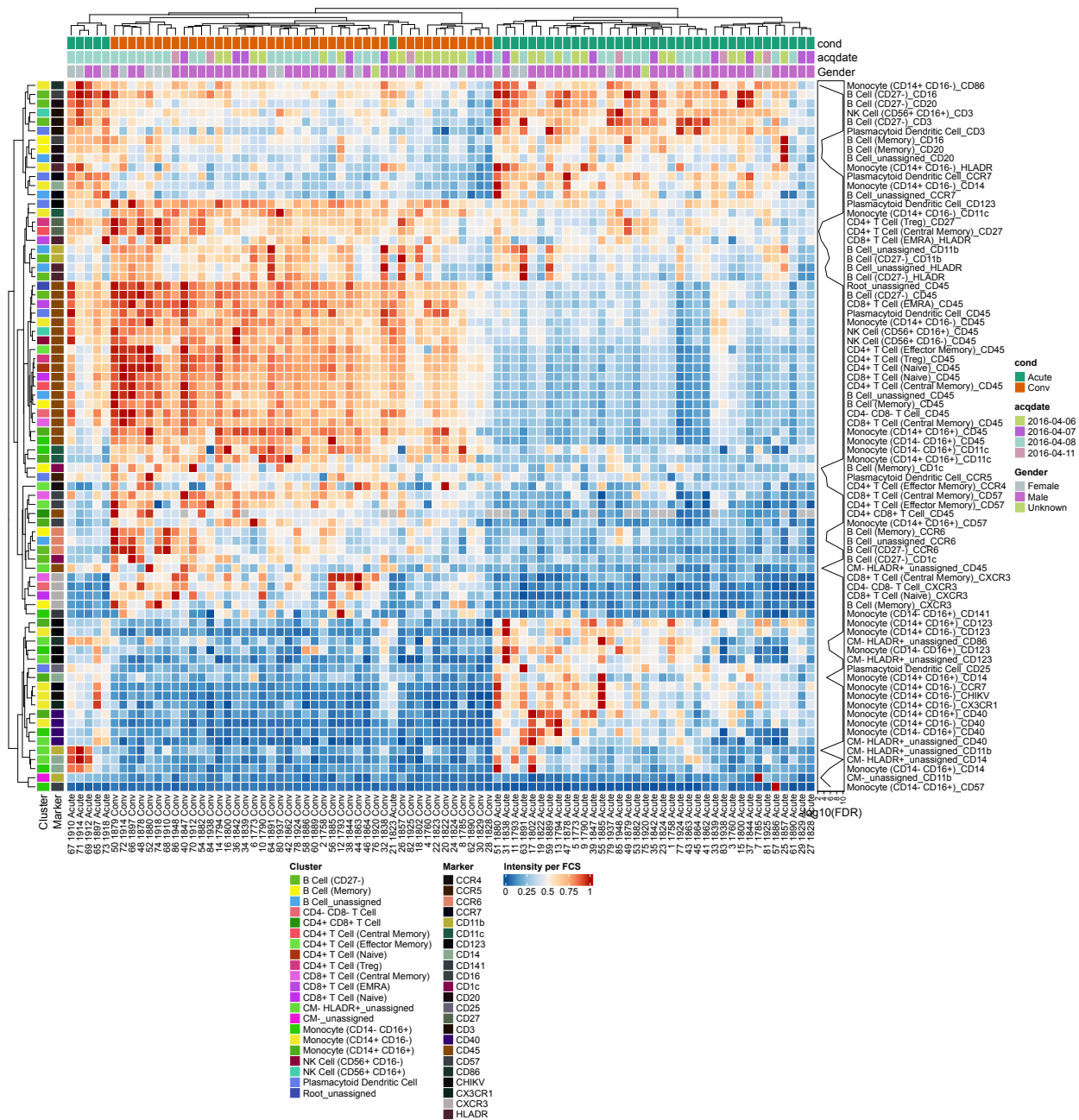


Figure 18.3: MFI Marker\*Cluster Heatmap

# Glossary

## Glossary of Terms

Table 18.1: General terms

Term	Description
<b>Abundance</b>	The quantity or concentration of a specific cell type or marker within a sample, measured per individual FCS file in cytometry data. Often used to describe the relative or absolute presence of cell populations in cytometry.
<b>cluster</b>	Pre-defined groups of similar cells based on specific characteristics or measurements, used to analyze subsets within cytometry data. Clusters are identified prior to analysis in the app.
<b>Differential Abundance</b>	The comparison of abundance levels between different conditions or groups in cytometry data to identify significant changes in cell populations.
<b>Differential State</b>	Refers to the analysis of changes in state or condition of cell populations under different experimental conditions, assessing variations beyond mere abundance.
<b>FCS</b>	Fluorescence-activated Cell Sorting file, a data format used in flow cytometry that contains information about each individual cell measured by the flow cytometer.
<b>Gating</b>	Method used to select cells based on specific criteria in flow cytometry, allowing for the analysis of cell subpopulations. Gating is performed prior to data analysis in the app.
<b>Metadata</b>	
<b>MFI</b>	Mean Fluorescence Intensity, a measure used in flow cytometry to quantify the average fluorescence intensity of cells within a gated population, indicative of protein expression levels.
<b>State Markers</b>	Indicators used in cytometry to identify transient or conditional characteristics of cells, such as activation or stress responses, reflecting the dynamic states of cells.
<b>Type Markers</b>	Specific markers used in cytometry to classify cells into distinct types or classes based on their inherent properties, such as cell lineage or functional attributes.

Table 18.2: Bioinformatic terms

Term	Description
<b>count</b>	
<b>perCellCounts</b>	Frequencies of cells, expressed as abundances per individual cell, allowing for comparison across samples with varying total cell counts.
<b>perCellCountsNorm</b>	Frequencies normalized using the arcsinh transformation (scaled by 0.03) and then centered by the mean per cluster/gating. Commonly used in flow cytometry to stabilize variance across a range of values.
<b>SummarizedExperiment</b>	A class in R used to store experimental data along with annotations about the experiments. It encapsulates data in a way that retains the relationship between data and metadata, useful for complex analyses.
<b>SingleCellExperiment</b>	A specialized version of <b>SummarizedExperiment</b> designed specifically for cell-level data. It facilitates the handling of single-cell specific metrics and annotations.

## References

- Krieger, Nik, Adam Perzynski, and Jarrod Dalton. 2021. *Projects: A Project Infrastructure for Researchers*. <https://CRAN.R-project.org/package=projects>.
- Michlmayr, Daniela, Theodore R. Pak, Adeeb H. Rahman, El-Ad David Amir, Eun-Young Kim, Seunghee Kim-Schulze, Maria Suprun, et al. 2018. “Comprehensive Innate Immune Profiling of Chikungunya Virus Infection in Pediatric Cases.” *Molecular Systems Biology* 14 (8): e7862. <https://doi.org/10.15252/msb.20177862>.
- Nowicka, Malgorzata, Carsten Krieg, Helena L Crowell, Lukas M Weber, Felix J Hartmann, Silvia Guglietta, Burkhard Becher, Mitchell P Levesque, and Mark D Robinson. 2017. “CyTOF Workflow: Differential Discovery in High-Throughput High-Dimensional Cytometry Datasets.” *F1000Research* 6: 748. <https://doi.org/10.12688/f1000research.11622.3>.
- Ritchie, Matthew E, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. “limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47. <https://doi.org/10.1093/nar/gkv007>.
- Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth. 2010. “edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26 (1): 139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
- Weber, Lukas M., Malgorzata Nowicka, Charlotte Soneson, and Mark D. Robinson. 2019. “Difcyt: Differential Discovery in High-Dimensional Cytometry via High-Resolution Clustering.” *Communications Biology* 2 (1): 183. <https://doi.org/10.1038/s42003-019-0415-5>.

# A Deploying with IFB

To deploy your R Shiny app using the IFB platform, follow these steps:

- **Create an IFB account:** Register for an account on the IFB platform.
- **Join a group:** Affiliate yourself with a group on the IFB platform. This will grant you access to the necessary resources for deploying your app.
- **Open a virtual machine (VM):** Once you are affiliated with a group, you can open a VM of the app.
  - Choose an “R Shiny” VM
  - add the `shiny_repo` <https://gitrcm.marseille.inserm.fr/lohmann/deploy>.
  - Run the VM.

## **i** Note

This process typically takes approximately 15 minutes.

## **B Import data from Flowjo**

## **C Import data from OmiQ**

## D Analycyte-verse

- [analycyte](#)
- [analycyte.projects](#)
- [analycyte.utils](#)
- [analycyte.book](#)

# E YAML Parameters

## E.1 Explanation of YAML Parameters

Parameter	Default Value	Description
<code>adcode</code>	<code>false</code>	Whether to show R code.
<code>analyse_type</code>	<code>null</code> <QC,DA,DS>	Specify analysis type.
<code>annotation_column</code>	<code>null</code>	Column used for annotation; defaults to none.
<code>batch</code>	<code>null</code>	Batch metadata column name for the analysis, if applicable.
<code>cluster_id</code>	<code>"cluster_id"</code>	Default identifier for clusters.
<code>col_block</code>	<code>null</code>	<code>block_id</code> 's diffeyt parameter.
<code>cols_design_GLMM_random</code>	<code>null</code>	Columns used in GLMM for random effects.
<code>conditions</code>	<code>null</code>	Conditions metadata column name under which the analysis is performed.
<code>conditions_reference</code>	<code>null</code>	Reference condition for comparison in analyses.
<code>contrast</code>	<code>null</code>	Contrast condition for analysis, each is compare to <code>conditions_reference</code> .
<code>fc</code>	<code>1</code>	Fold change threshold for significance.
<code>log_bar</code>	<code>false</code>	Whether to use logarithmic scaling on bar plots.
<code>min_cells</code>	<code>1</code>	Minimum number of cells required per sample.
<code>min_samples</code>	<code>null</code>	Minimum number of samples required for analysis.
<code>path</code>	<file path>	Path to the data file.
<code>p_value</code>	<code>0.05</code>	P-value threshold for statistical significance.
<code>plot_legends</code>	<code>null</code>	Metadata column name to add on graphs.
<code>project</code>	<code>null</code>	Project description if provided.
<code>rds</code>	<.rds>	RDS file name.
<code>show_parameters</code>	<code>true</code>	Whether to show parameters in outputs.
<code>show_summary</code>	<code>true</code>	Whether to display summary information tables.
<code>show_heatmap_on_off</code>	<code>true</code>	Whether to show or hide the number of cells per fcs heatmaps.
<code>sorted</code>	<code>false</code>	Whether data should be sorted on graph by <code>conditions</code> .
<code>transform</code>	<code>false</code>	Whether data should be transformed on Differential analysis.
<code>subset_column</code>	<code>null</code>	Column used for subsetting data, if applicable.