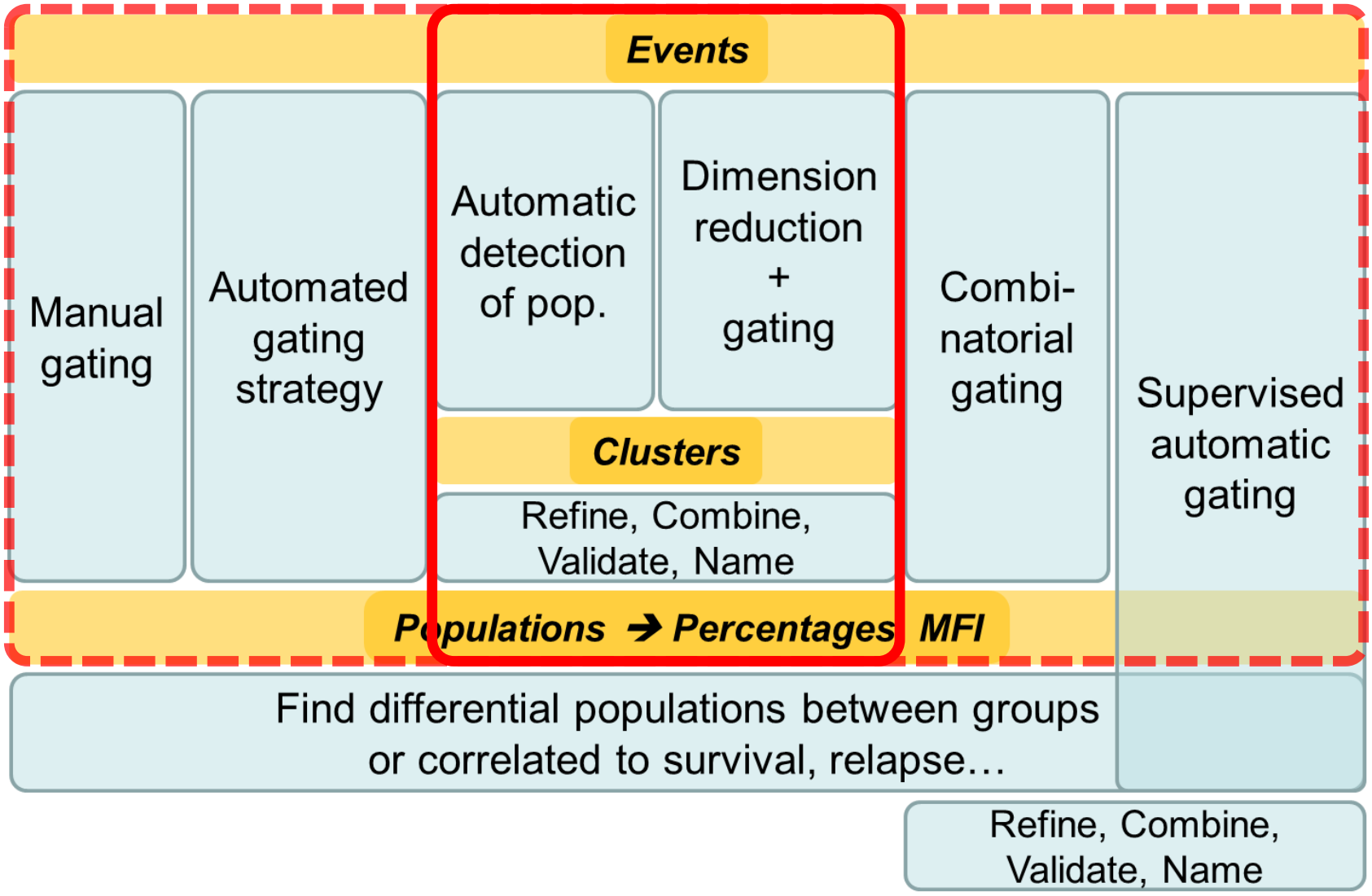
Cytofkit – Cytofast Analyses

Unsupervised Analyses

S. Granjeaud, CRCM

2022-10-19

Events analysis



Cytofkit – Cytofast Analysis

- http://i-cyto.github.io

- cytofkit
  - Cytofkit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline

- Cytofast
  - Cytofast: A workflow for visual and quantitative analysis of flow and mass cytometry data to discover immune signatures and correlations

Installation

- Open R/RStudio
- Do copy/paste/run commands **line by line**
- Whenever you get a message 'Update all/some/none? [a/s/n]:', answer **n**

```
# Install Bioconductor stuff
install.packages(c("BiocManager", "remotes"))
BiocManager::install(update = FALSE)  # verify

# Install requirements
options(install.packages.check.source = "no")
BiocManager::install("flowCore", update = FALSE)
BiocManager::install("uwot", update = FALSE)
BiocManager::install("cytofast", update = FALSE)
BiocManager::install(c("RANN", "igraph"), update = FALSE)

# Install cytofkitlab packages for Windows
install.packages("https://github.com/i-cyto/Rphenograph/releases/download/Rphenograph_0.99.1.9003/Rphenograph_0.99.1.9003.zip",
       repos = NULL, type = "win.binary")
BiocManager::install("i-cyto/cytofkitlab", update = FALSE)

# Install cytofkitlab packages for Mac/Linux
BiocManager::install("i-cyto/Rphenograph", type = "source", update = FALSE)
BiocManager::install("i-cyto/cytofkitlab", update = FALSE)
```

Outline

- Question: Find a group of cells that differ in abundance between two groups of patients

- Discover cytofkit
- Run calculations
- Use graphical interface to view results
- Use a cytofast script to get nicer figures

ORGANIZE DATA

Organize data

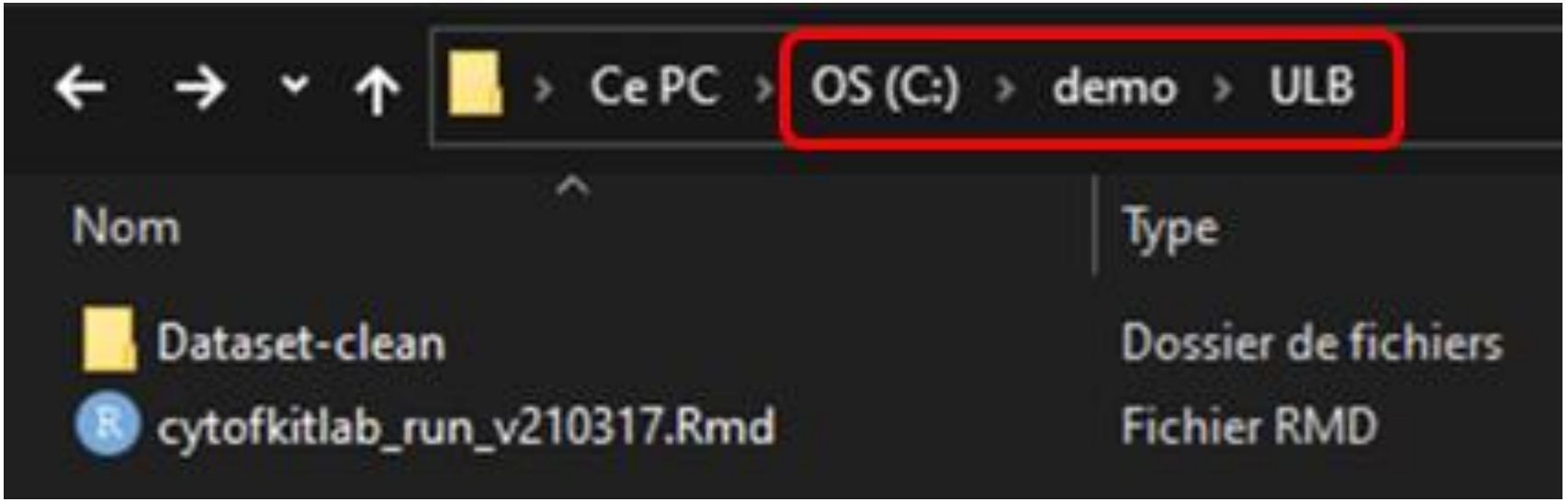
- Create a folder and copy
  - FCS files in a data folder
    FCS files to analyze
  - cytofkitlab_run_v210317.Rmd
    template to run commands
- Start a RStudio project
  from this existing folder on disk

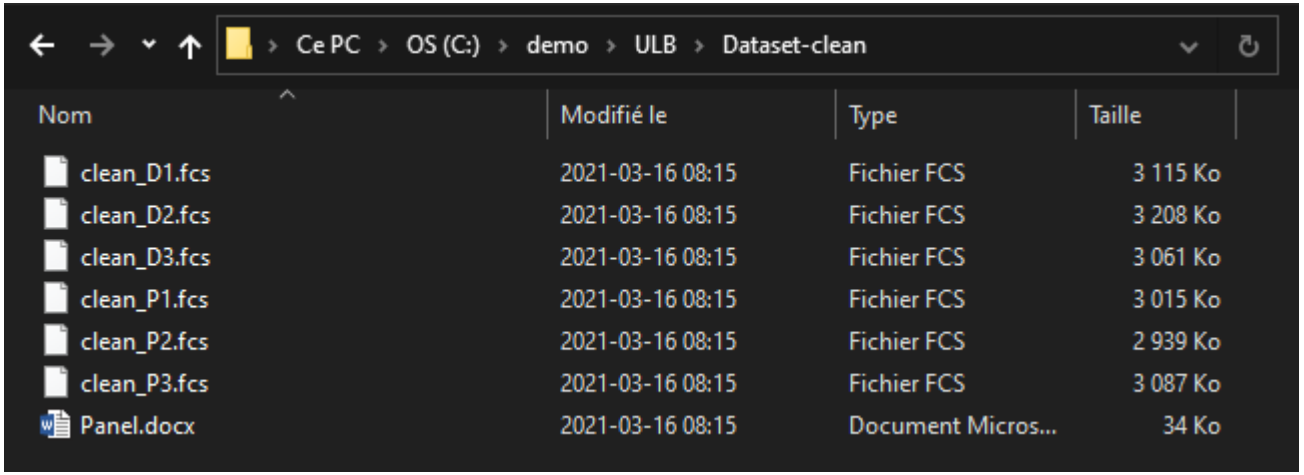
Prepare data

C:\demo\ULB

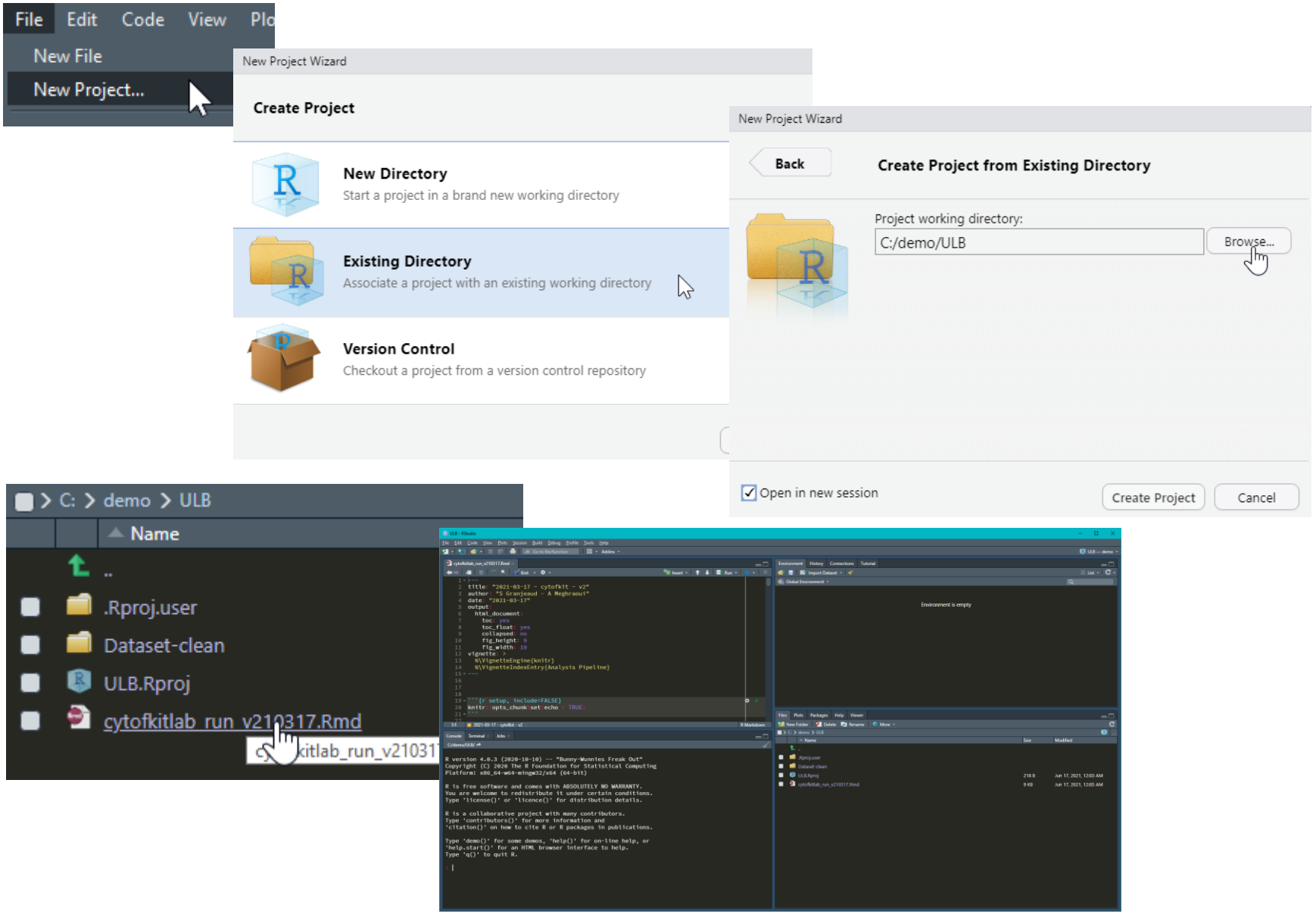
=> D:\0_bootcamp\cytofkit

Prepare data

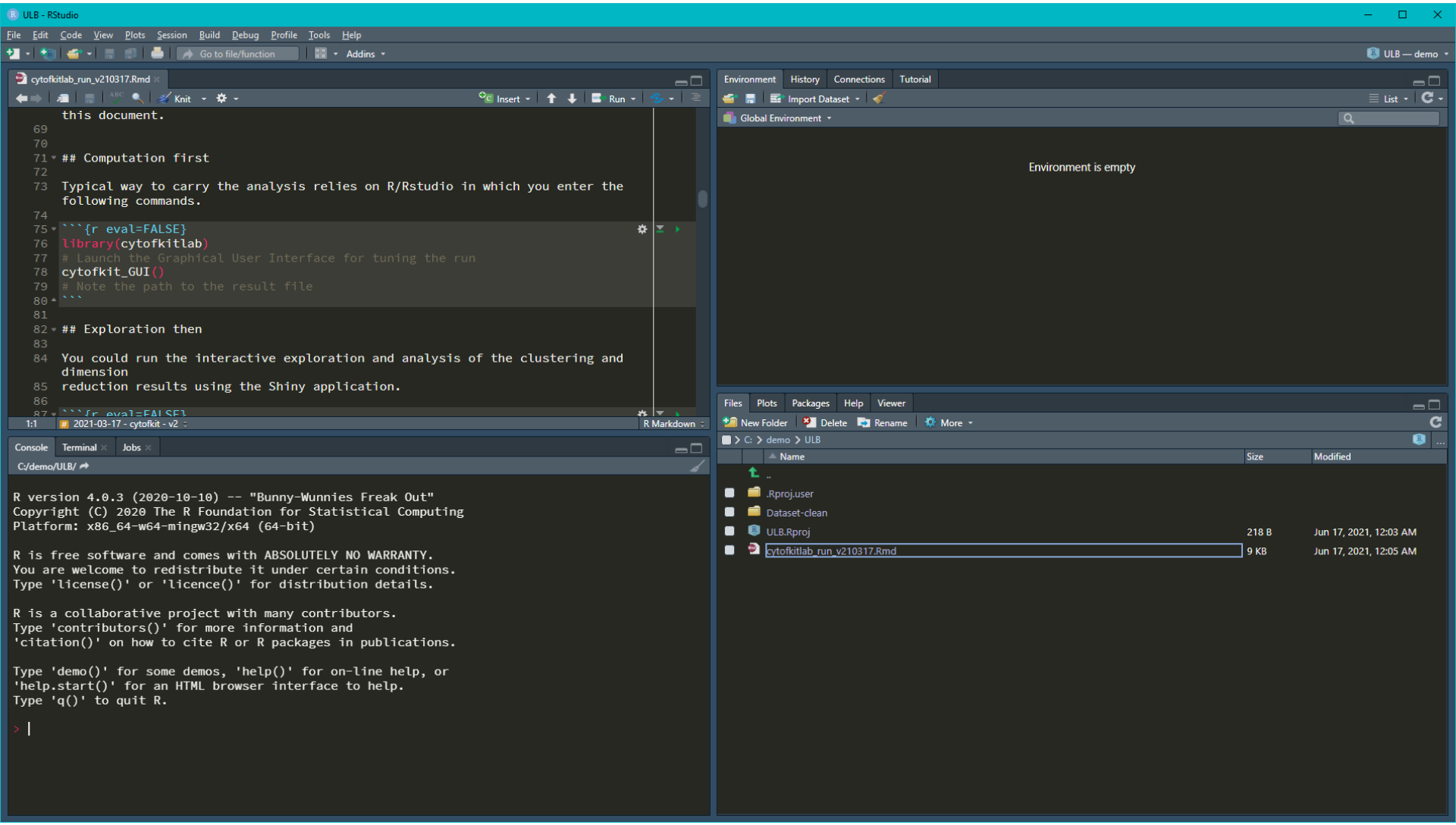




Create project



Guided analysis

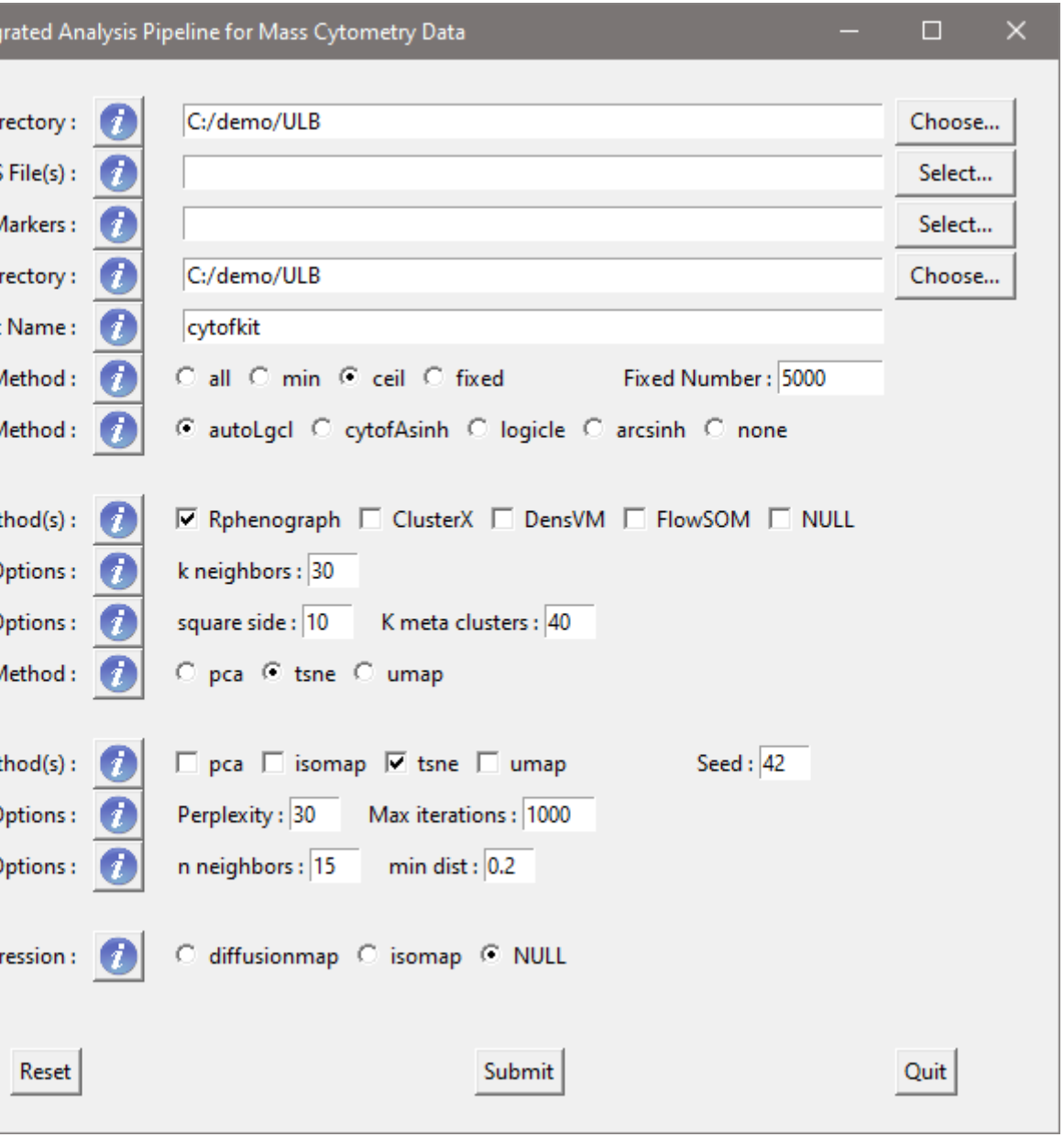


RUN CALCULATIONS CYTOFKIT

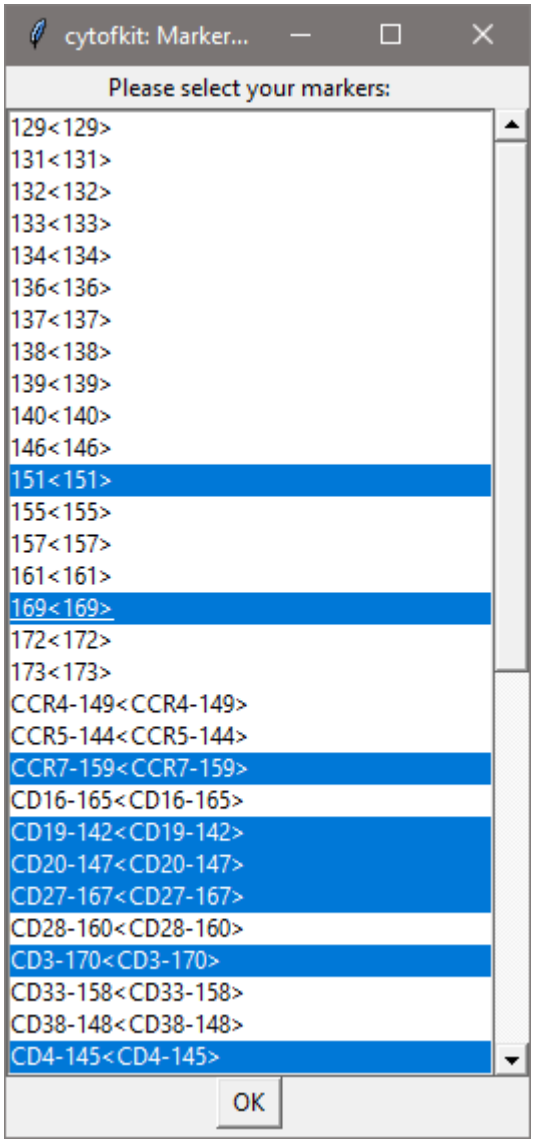
Run calculations with cytofkit

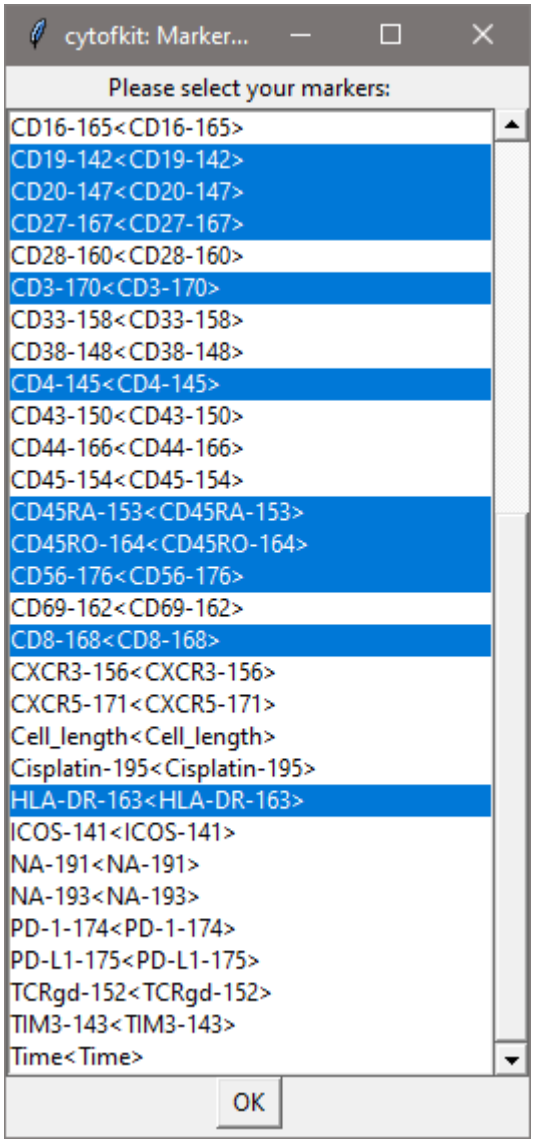
```
71 ## Computation first
72
73 Typical way to carry the analysis relies on R/Rstudio in which you enter the
   following commands.
74
75 ```{r eval=FALSE}
76 library(cytofkitlab)
77 # Launch the Graphical User Interface for tuning the run
78 cytofkit_GUI()
79 # Note the path to the result file
80 ```
```

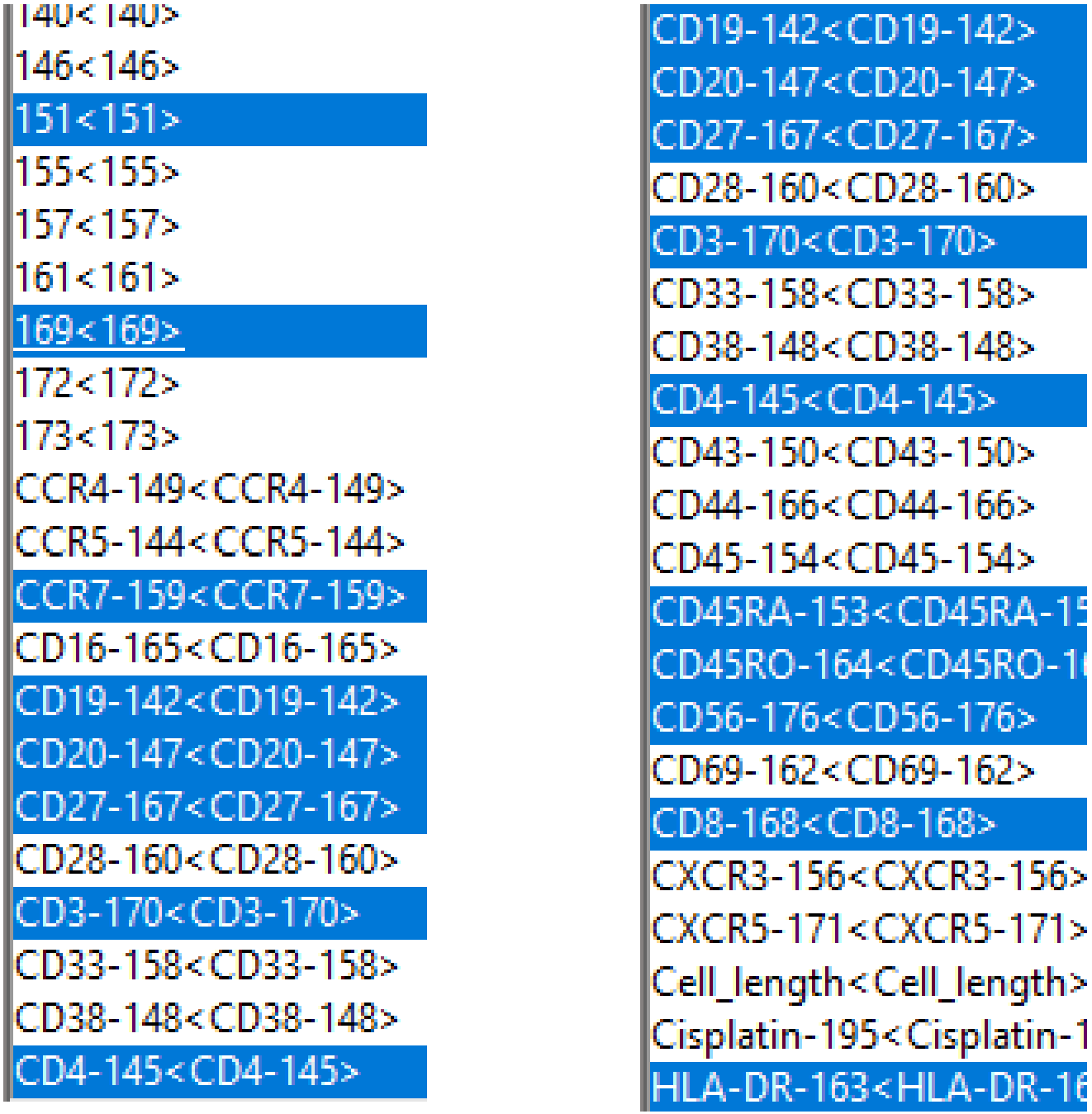
Calculations GUI



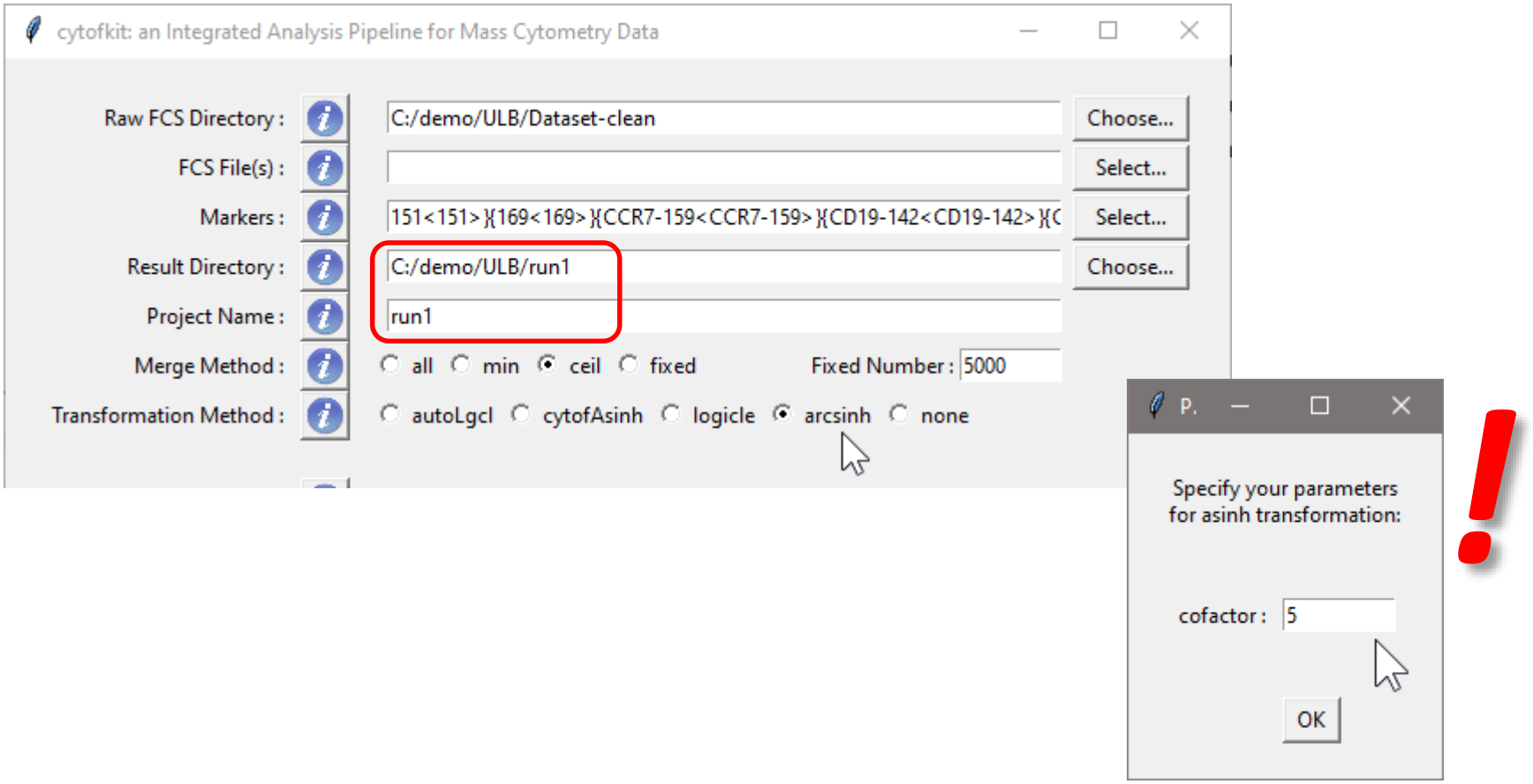
Calculations GUI



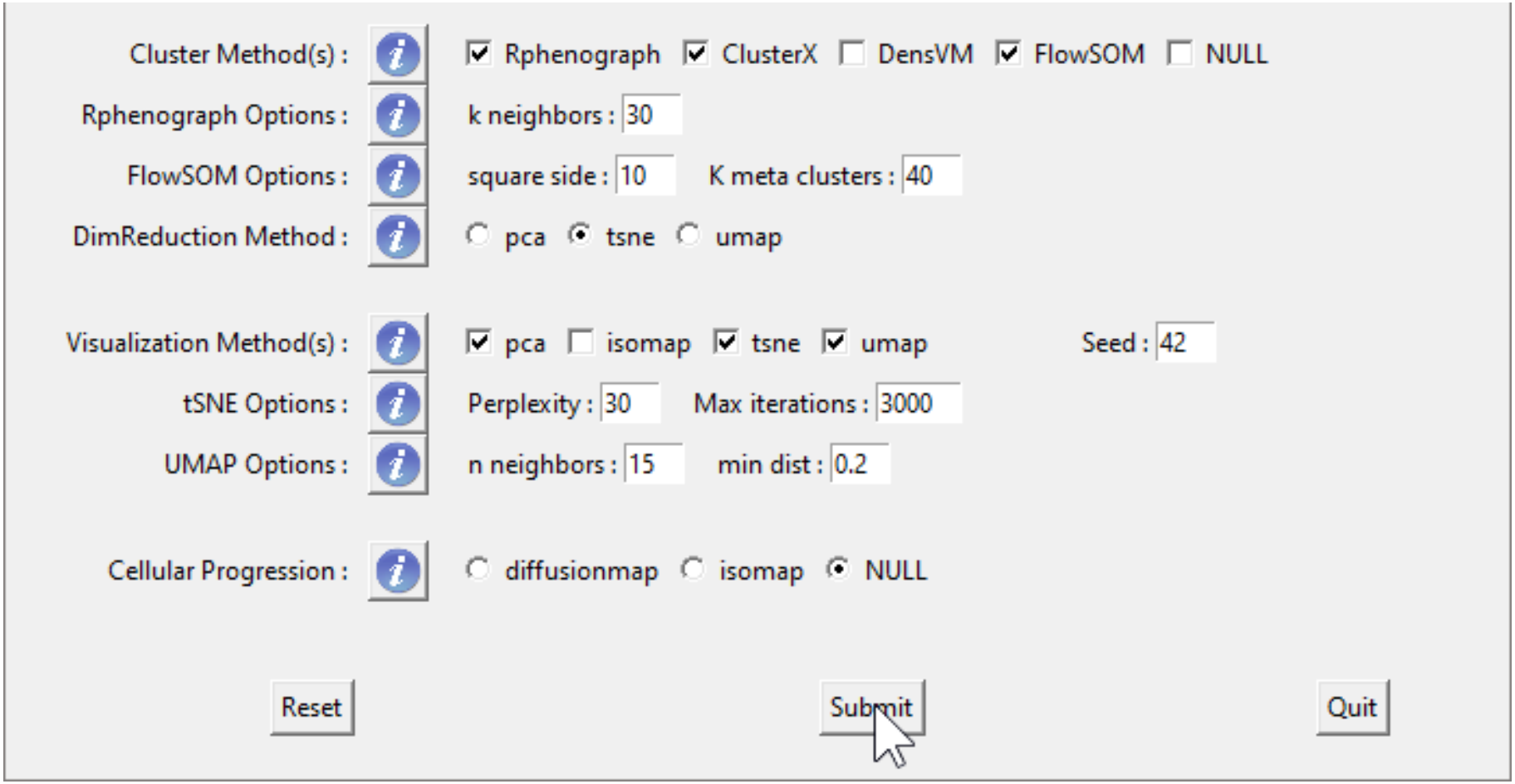



140<140>
146<146>
151<151>
155<155>
157<157>
161<161>
169<169>
172<172>
173<173>
CCR4-149<CCR4-149>
CCR5-144<CCR5-144>
CCR7-159<CCR7-159>
CD16-165<CD16-165>
CD19-142<CD19-142>
CD20-147<CD20-147>
CD27-167<CD27-167>
CD28-160<CD28-160>
CD3-170<CD3-170>
CD33-158<CD33-158>
CD38-148<CD38-148>
CD4-145<CD4-145>
CD19-142<CD19-142>
CD20-147<CD20-147>
CD27-167<CD27-167>
CD28-160<CD28-160>
CD3-170<CD3-170>
CD33-158<CD33-158>
CD38-148<CD38-148>
CD4-145<CD4-145>
CD43-150<CD43-150>
CD44-166<CD44-166>
CD45-154<CD45-154>
CD45RA-153<CD45RA-15
CD45RO-164<CD45RO-1
CD56-176<CD56-176>
CD69-162<CD69-162>
CD8-168<CD8-168>
CXCR3-156<CXCR3-156>
CXCR5-171<CXCR5-171>
Cell_length<Cell_length>
Cisplatin-195<Cisplatin-1
HLA-DR-163<HLA-DR-16

Calculations GUI



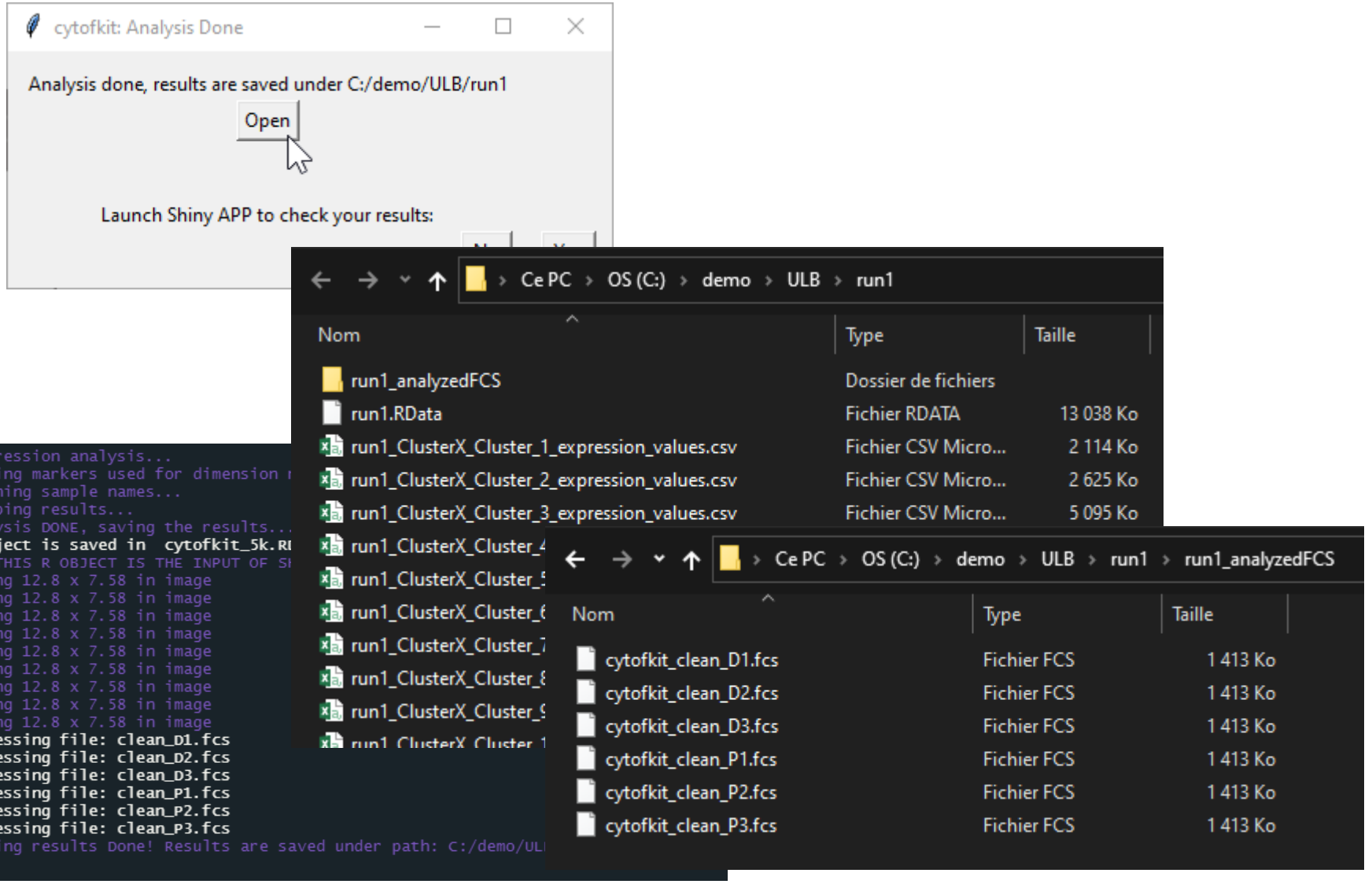
Calculations GUI



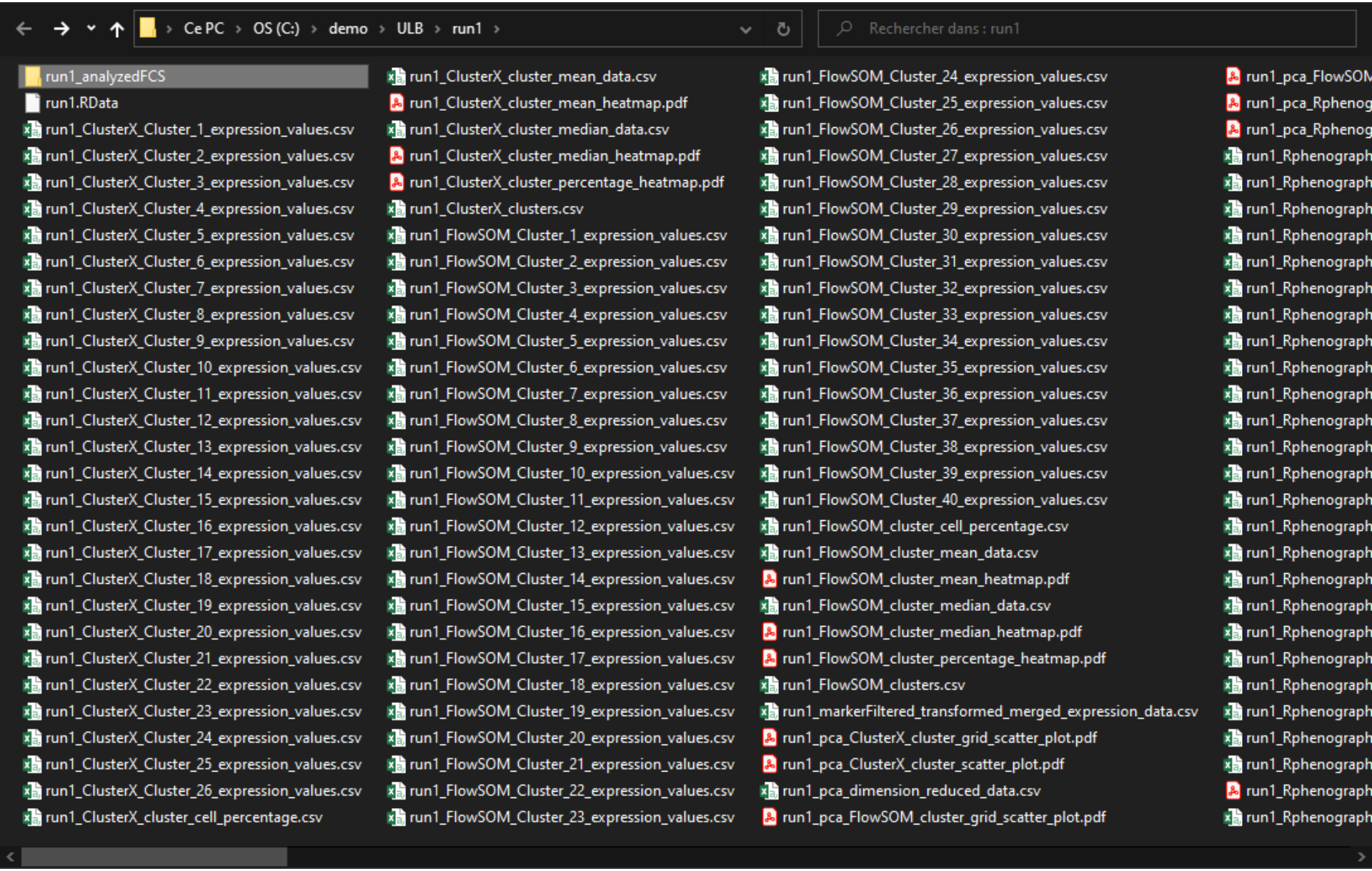
Calculating…

```
> setwd("C:/demo/ULB/")
> library(cytofkitlab)
> # Launch the Graphical User Interface for tuning the run
> cytofkit_GUI()
Input arguments:
* Project Name: cytofkit_5k
* Input FCS files for analysis (6):
    clean_D1.fcs
    clean_D2.fcs
    clean_D3.fcs
    clean_P1.fcs
    clean_P2.fcs
    clean_P3.fcs
* Markers (13):
    151<151>
    169<169>
    CCR7-159<CCR7-159>
    CD19-142<CD19-142>
    CD20-147<CD20-147>
    CD27-167<CD27-167>
    CD3-170<CD3-170>
    CD4-145<CD4-145>
    CD45RA-153<CD45RA-153>
    CD45RO-164<CD45RO-164>
    CD56-176<CD56-176>
    CD8-168<CD8-168>
    HLA-DR-163<HLA-DR-163>
* Data merging method: ceil
* Data transformation method: arcsinh
* Data normalization method: default
* Dimensionality reduction method: tsne
* Data clustering method(s): Rphenograph ClusterX FlowSOM
* Data visualization method(s): pca tsne umap
* Subset progression analysis method: NULL

Extract expression data...
   30000  x  50  data was extracted!
Dimension reduction...
  Running PCA...
  DONE in 0.02 mins
  Running t-SNE...with seed 42
Extract expression data...
   30000  x  50  data was extracted!
Dimension reduction...
  Running PCA...
  DONE in 0.02 mins
  Running t-SNE...with seed 42
  DONE in 0.9 mins
  Running UMAP...with seed  42
  DONE in 0.5 mins
Run clustering...
  Running PhenoGraph...
Run Rphenograph starts:
  -Input data of 30000 rows and 13 columns
  -k is set to 30
  Finding nearest neighbors...DONE ~ 2.08 s
  Compute jaccard coefficient between nearest-neighbor sets...DONE ~ 0.29 s
  Build a graph from the weighted links...
  Links are not combined...DONE ~ 3.22 s
  Run  clustering on the graph ...DONE ~ 5.12 s
Run Rphenograph DONE, totally takes 10.71 s.
  Return a community class
  -Modularity value: 0.8789248
  -Number of clusters: 22

  DONE in 0.22 mins
  Running ClusterX...    Calculate cutoff distance...5.36
    Calculate local Density...DONE!
    Detect nearest neighbour with higher density...DONE!
    Peak detection...DONE!
    Cluster assigning...DONE!
    Noise cluster removed

  DONE in 3.5 mins
  Running FlowSOM...
    Building SOM...
    Meta clustering to 40 clusters...
  DONE in 0.11 mins
Progression analysis...
Listing markers used for dimension reduction...
Stashing sample names...
Wrapping results...
Analysis DONE, saving the results...
R object is saved in  cytofkit_5k.RData
  **THIS R OBJECT IS THE INPUT OF SHINY APP!**
```

Results of calculations

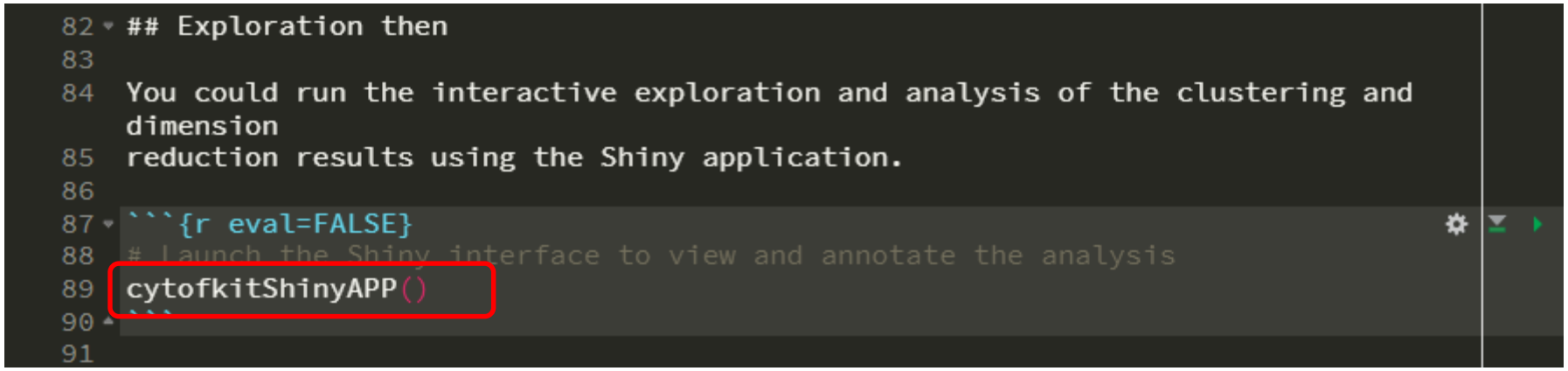


Results of calculations

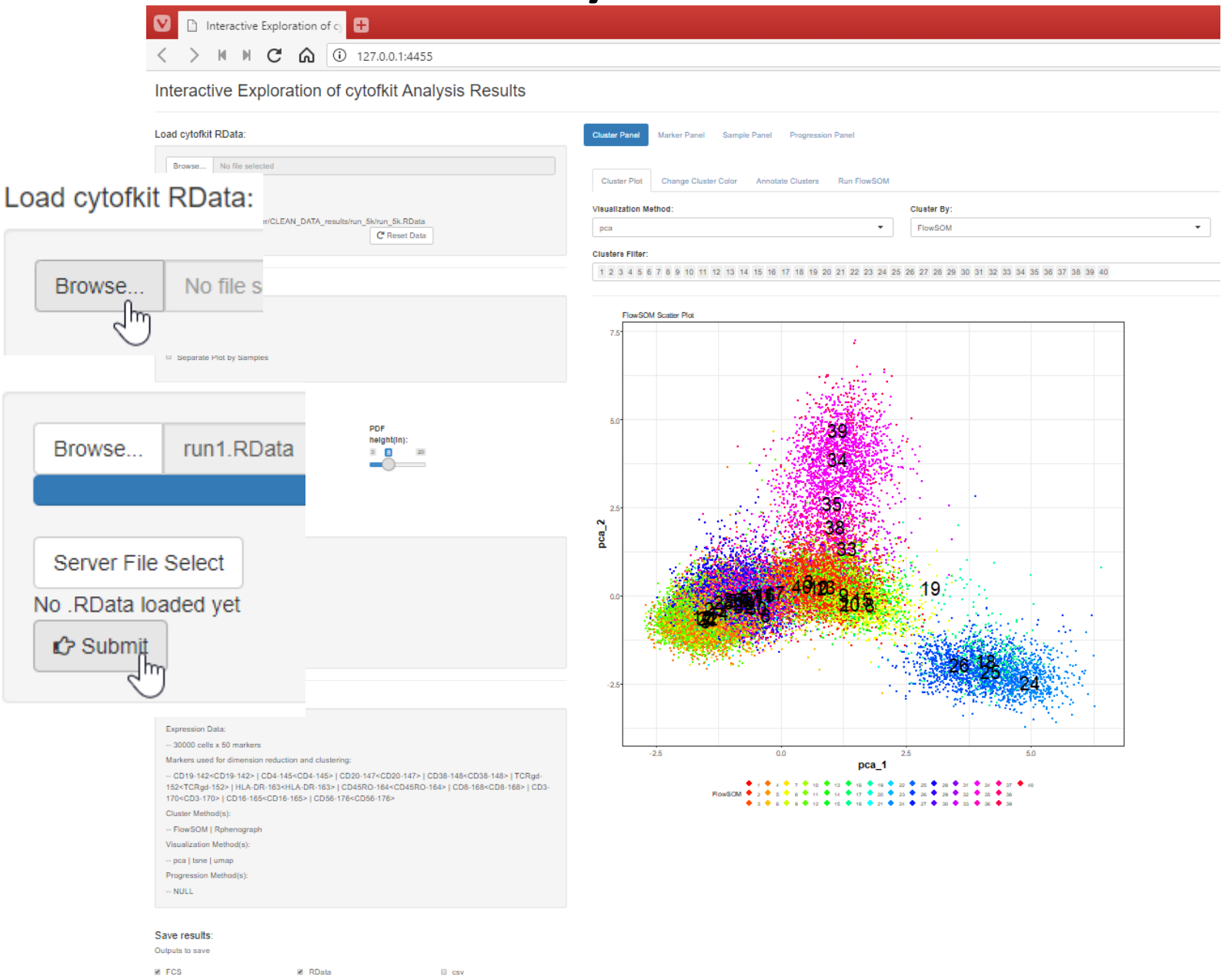


EXPLORE RESULTS
CYTOFKIT

Start analysis interface



Shiny interface



Load cytofkit RData:

Browse... No file selected

Server File Select

Loaded:
c:/demo/aida_atelier/CLEAN_DATA_results/run_5k/run_5k.RData

Submit Reset Data

Plot Control:

- [x] Add Cluster Labels
- [ ] Repel Cluster Labels
- [ ] Separate Plot by Samples

Download Cluster Plot in PDF

PDF width(in): 3 8 20

PDF height(in): 3 8 20

Open download folder

Sample Filter:

- [x] Select/Deselect All
- [x] clean_D1
- [x] clean_D2
- [x] clean_D3
- [x] clean_P1
- [x] clean_P2
- [x] clean_P3

Data Summary:

Save results:

Outputs to save

- [x] FCS
- [x] RData
- [ ] csv

Save Data

Cluster Panel | Marker Panel | Sample Panel | Progression Panel

Cluster Plot | Change Cluster Color | Annotate Clusters | Run FlowSOM

Visualization Method: pca

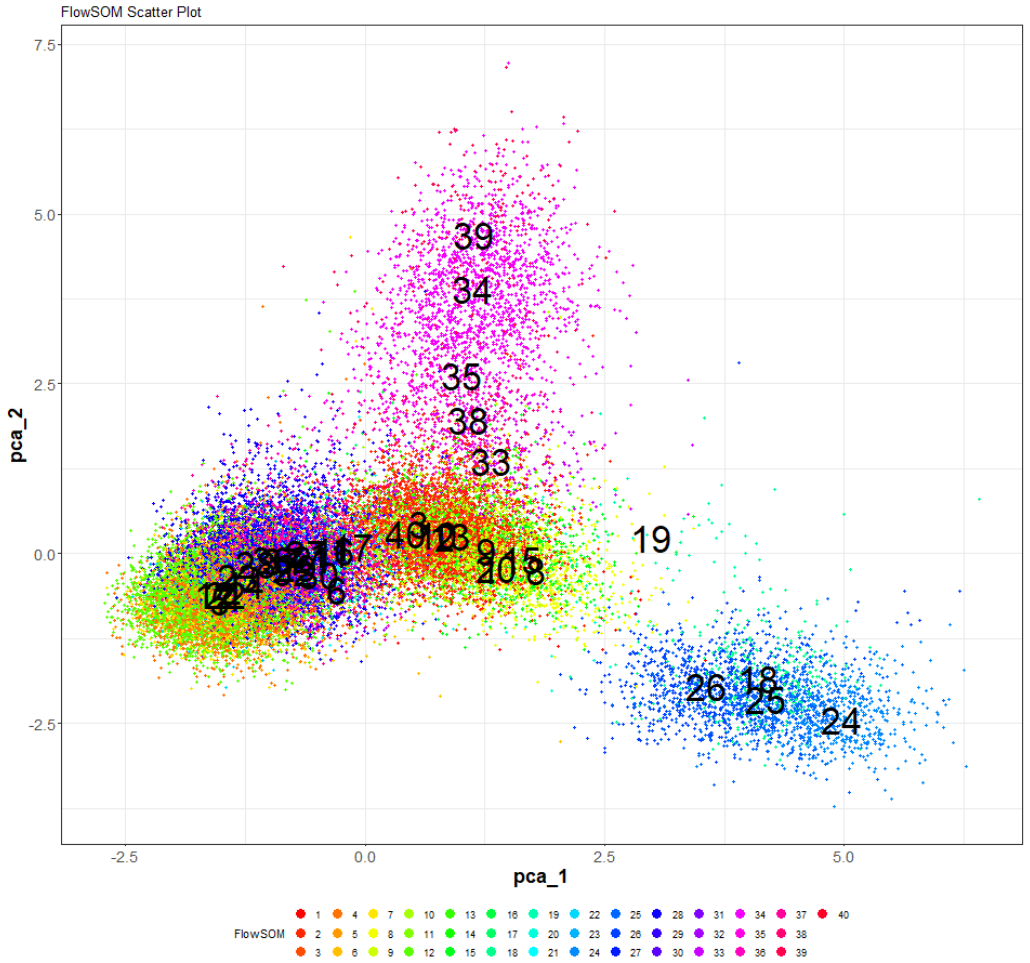
Cluster By: FlowSOM

Point Size: 1

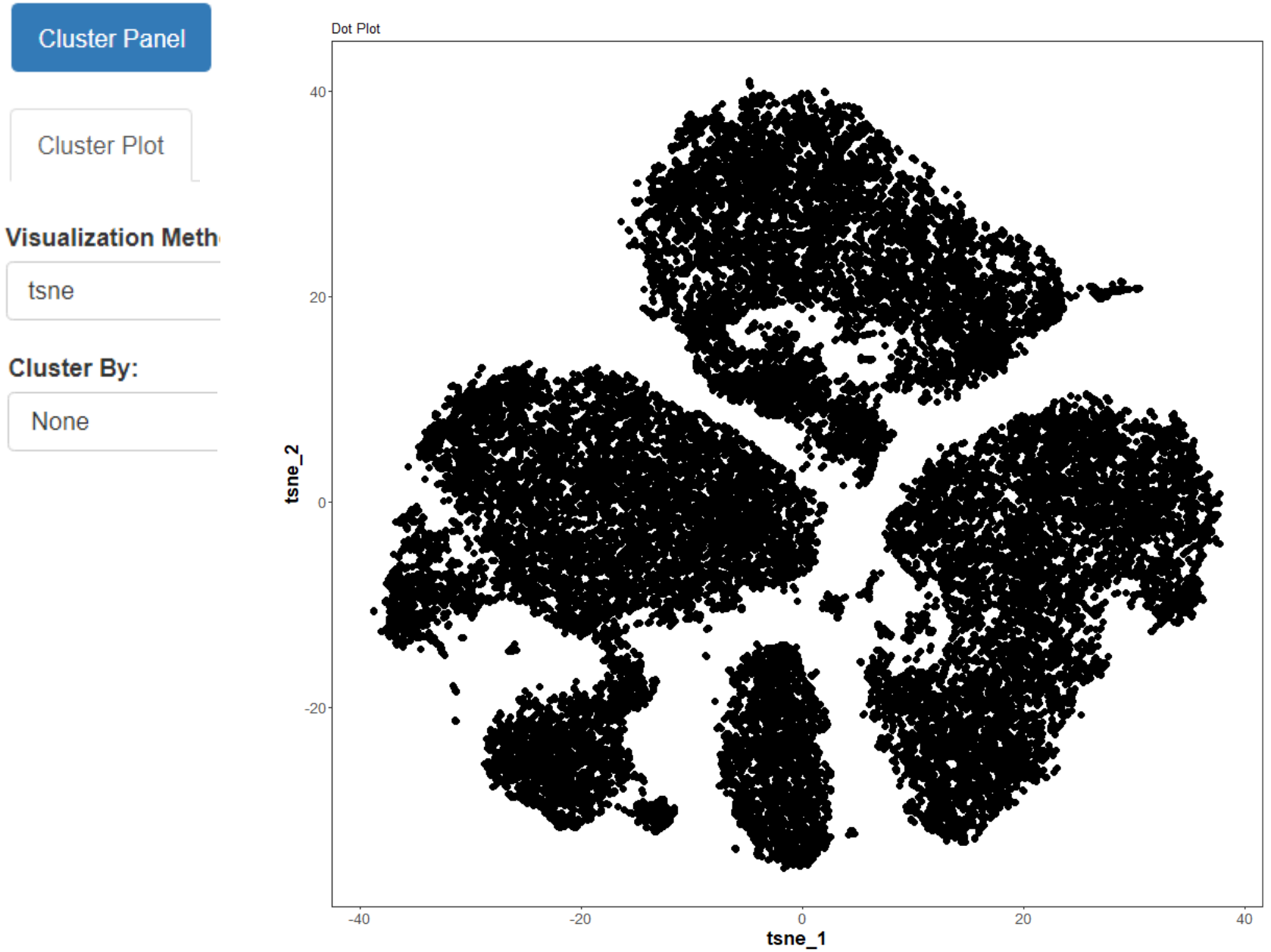
Label Size: 12

Clusters Filter:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40



tSNE reduction into 2D



Cell density on tSNE map

Cluster Panel

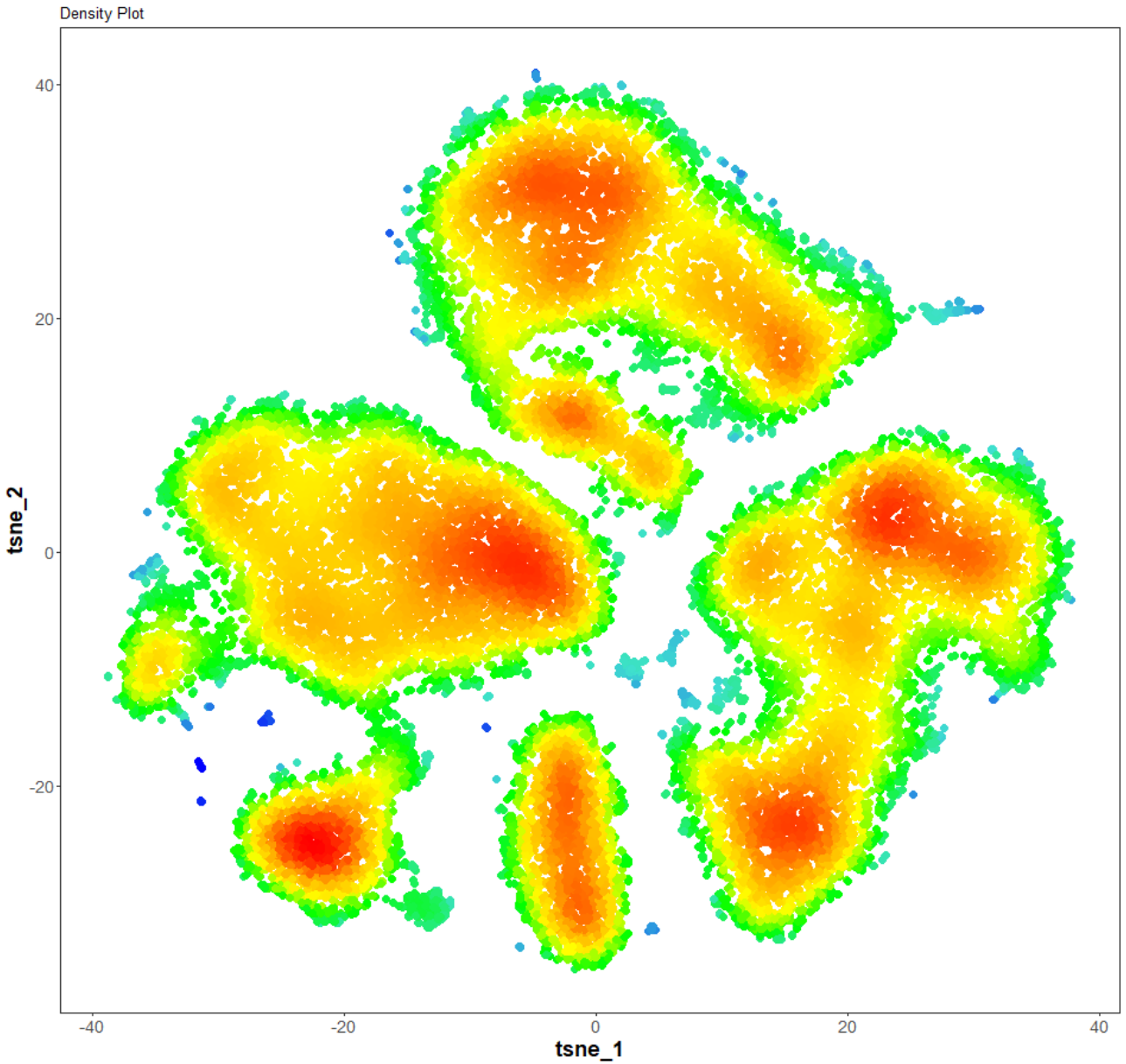
Cluster Plot

Visualization Meth

tsne

Cluster By:

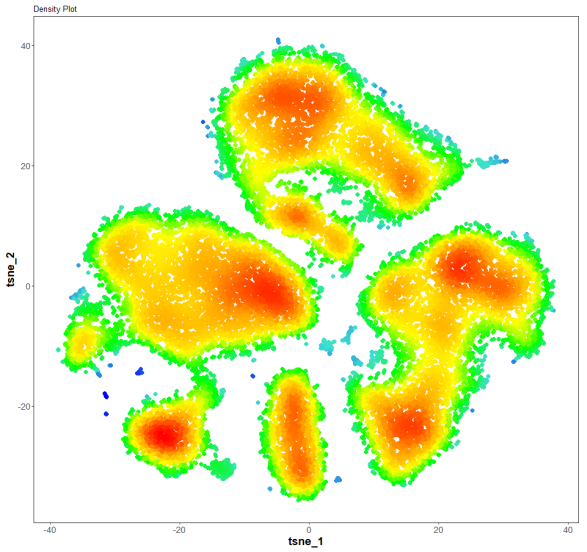
Density



Quizz

- How to make sense of the aggregates?

- Without clustering yet!



Markers on tSNE

Marker Panel

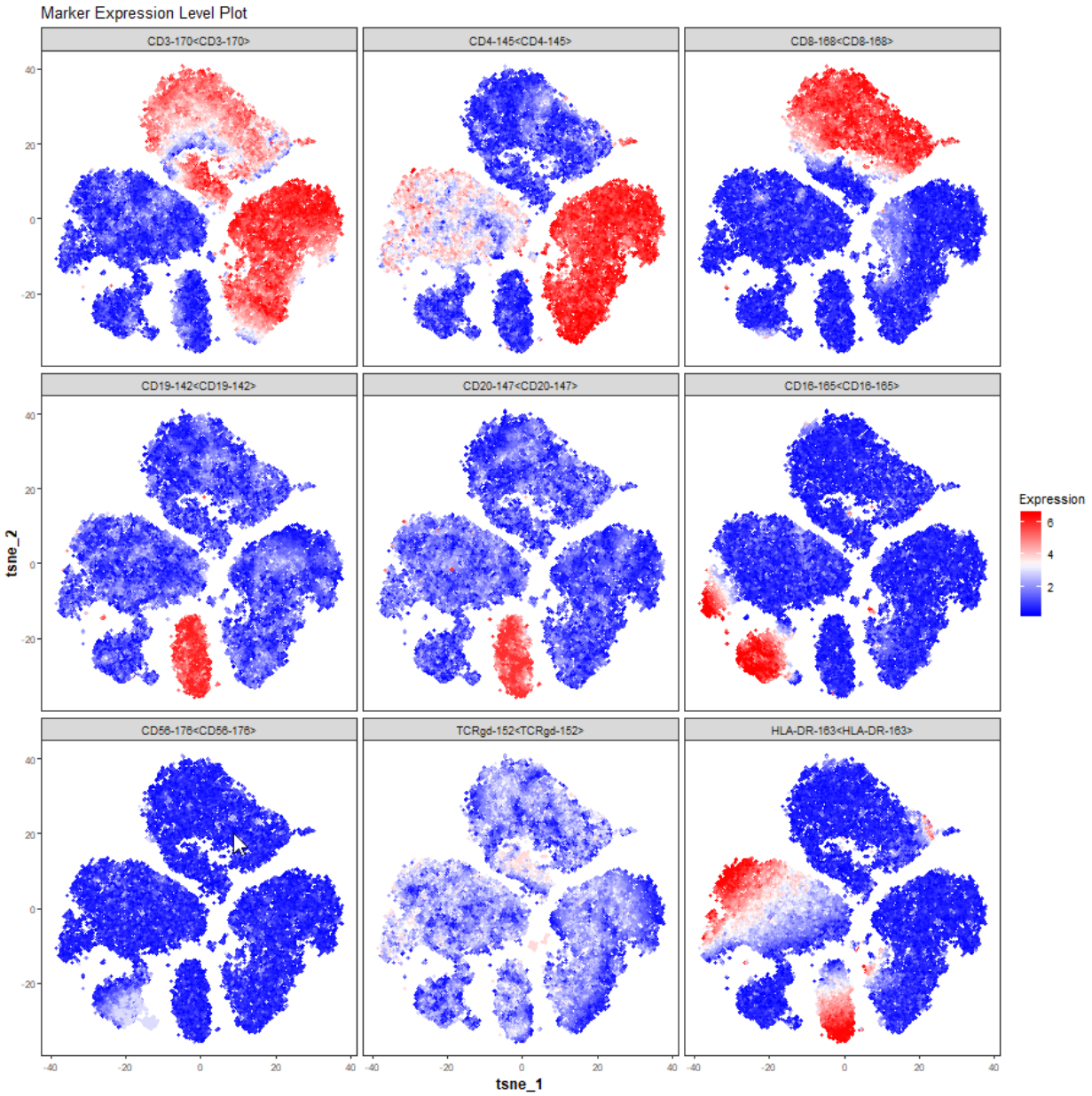
Expression Level Plot

Plot Marker:

- CD3-170<CD3-170>
- CD4-145<CD4-145>
- CD8-168<CD8-168>
- CD19-142<CD19-142>
- CD20-147<CD20-147>
- CD16-165<CD16-165>
- CD56-176<CD56-176>
- TCRgd-152<TCRgd-152>
- HLA-DR-163<HLA-DR-163>

All Markers

Update Plot



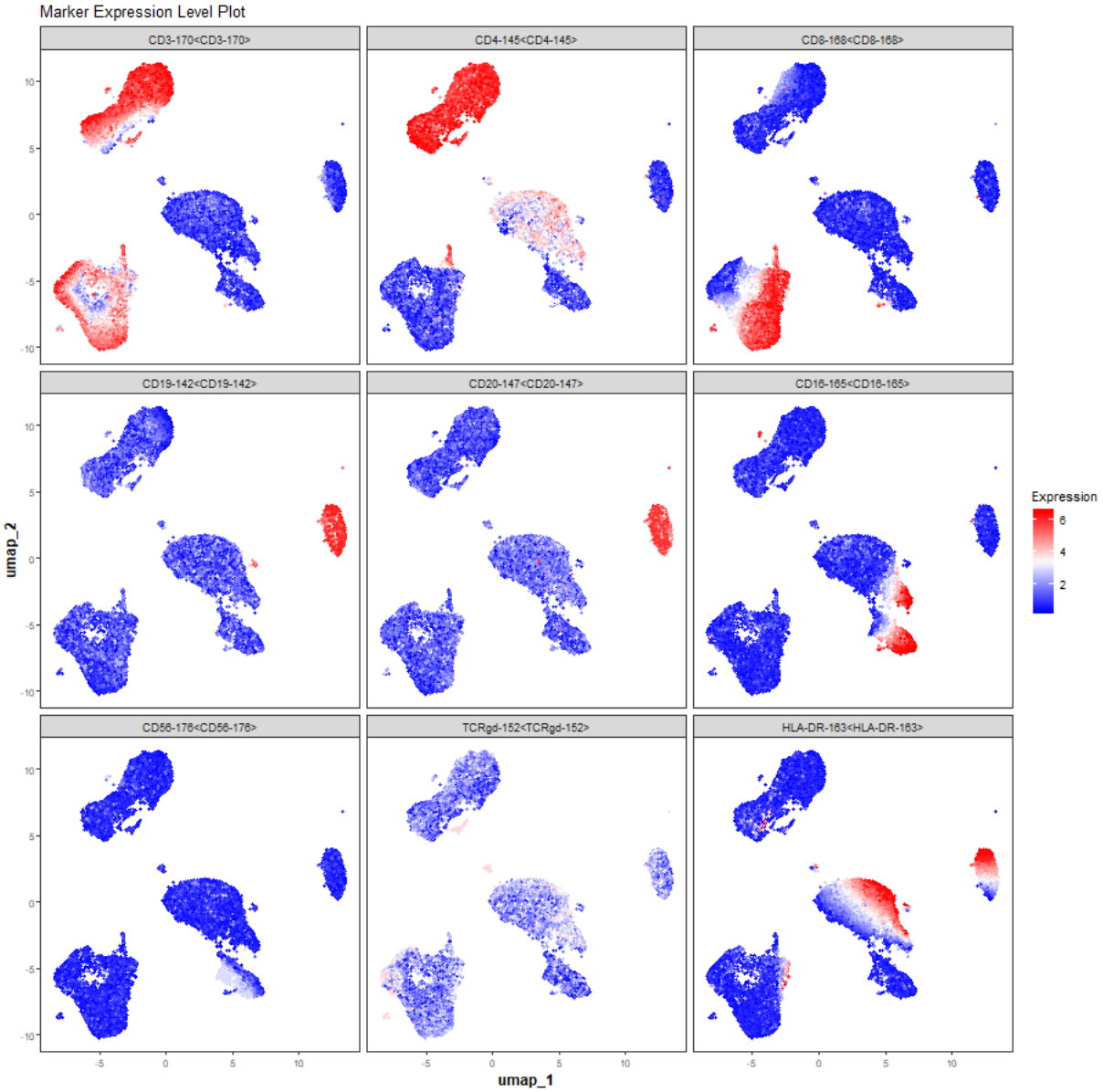
Markers on UMAP

Cluster Panel | Marker Panel | Samp

Expression Heat Map | Expression Le

Visualization Method:

umap



What about clusterings?

- FlowSOM
- Phenograph
- ClusterX

- By definition, in cytofkit, what is the difference of ClusterX vs FlowSOM & Phenograph?

- Let’s overlay FlowSOM on top of tSNE

Clusters on tSNE

Cluster Panel

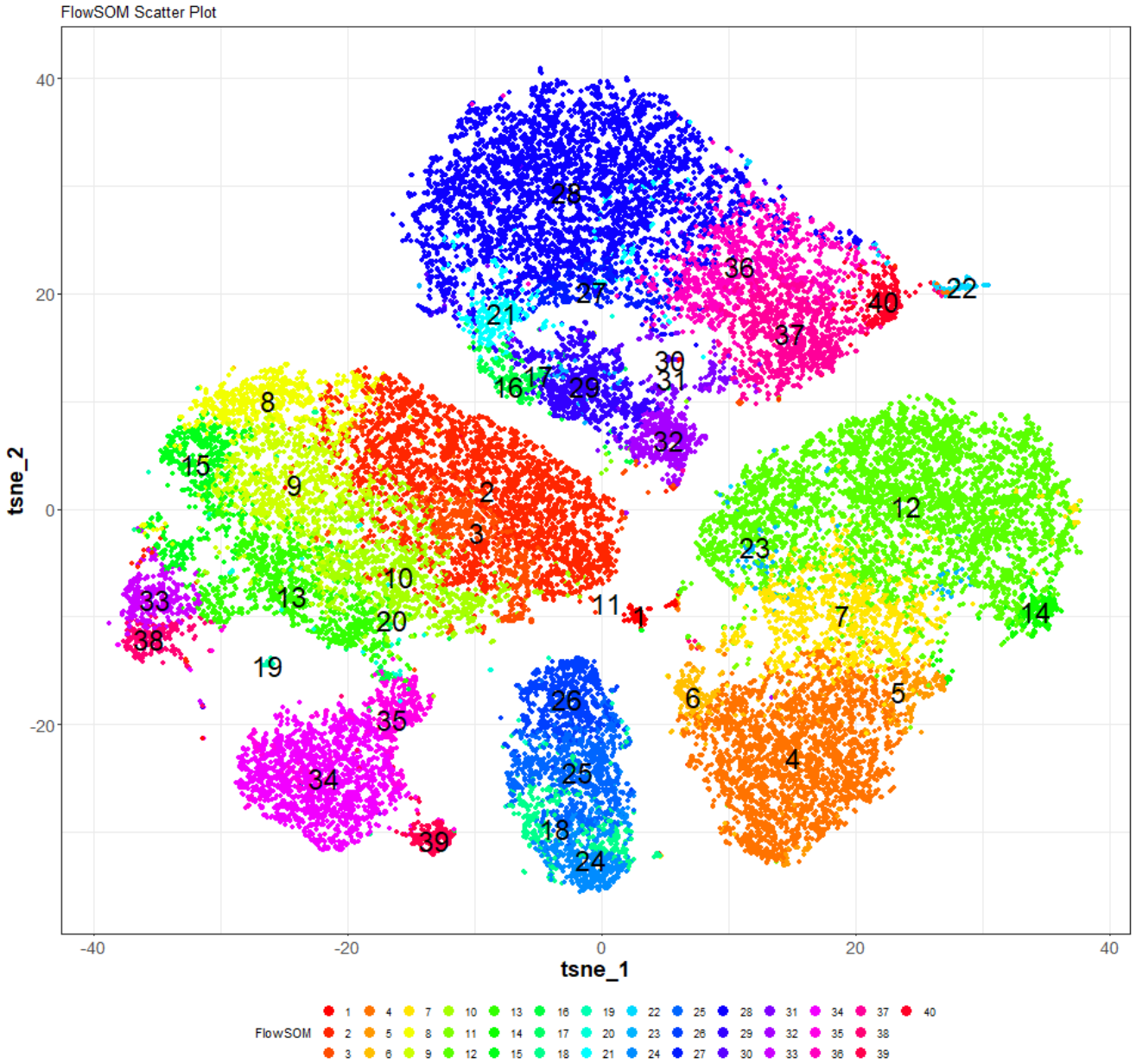
Cluster Plot

Visualization Meth

tsne

Cluster By:

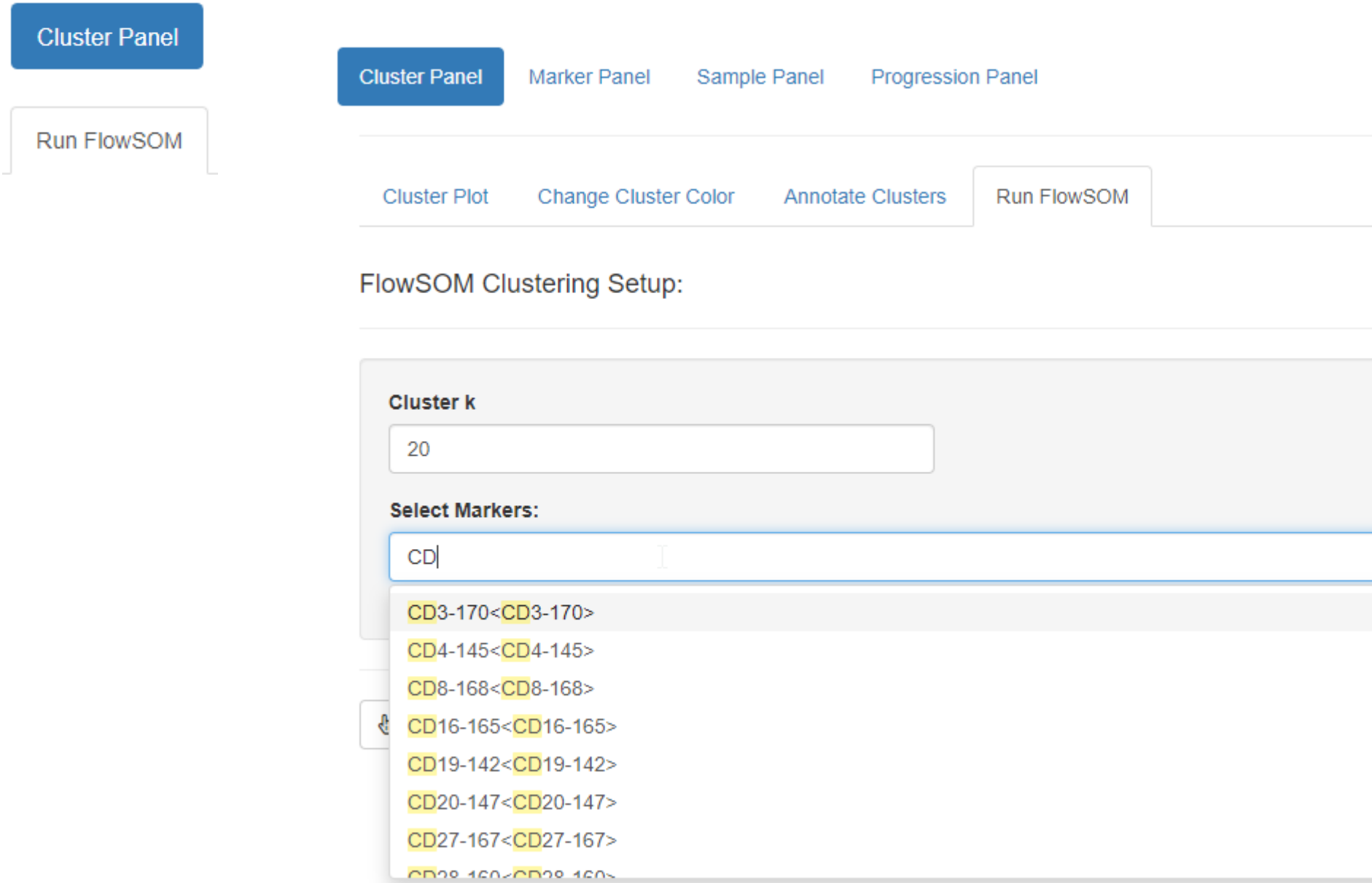
FlowSOM



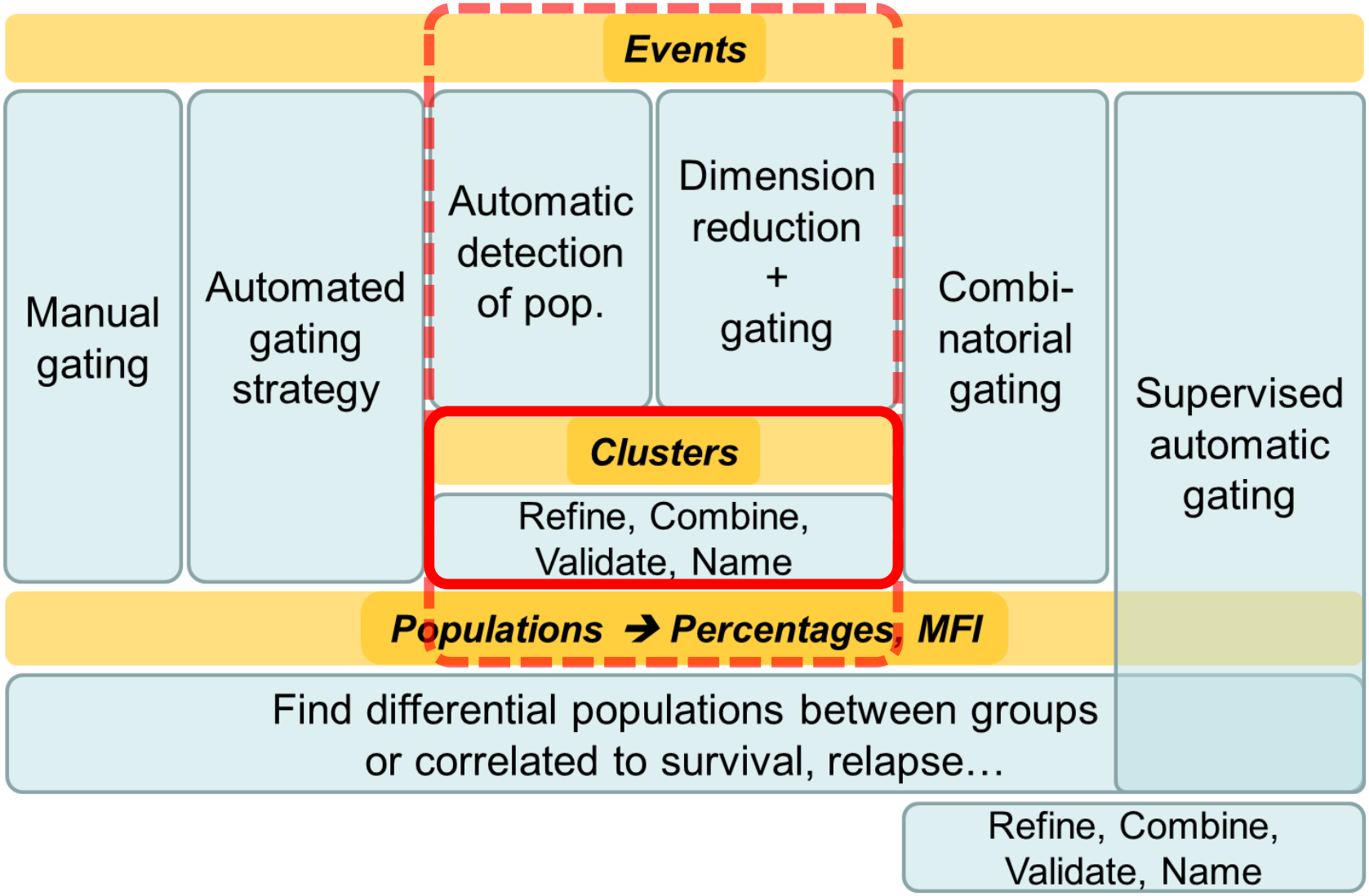
Hands on

- Overlay other clusterings
- Use UMAP also

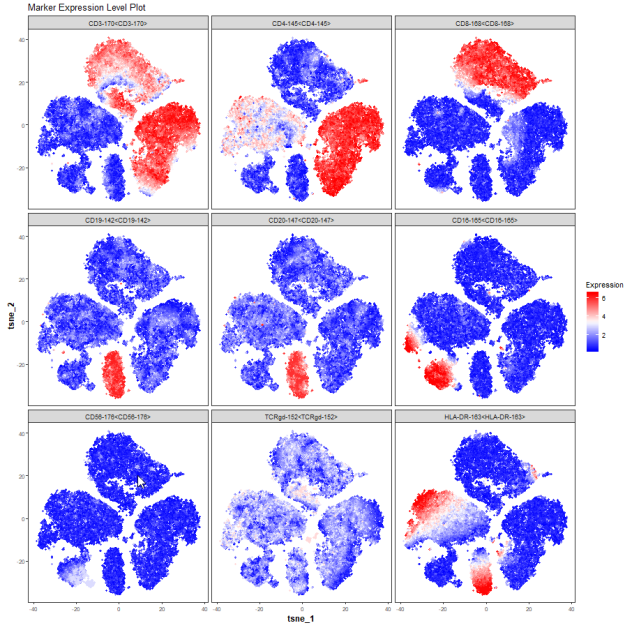
FlowSOM is fast… do it again!

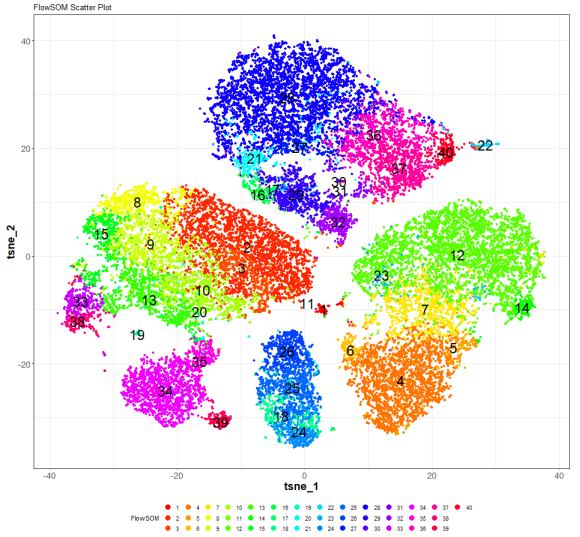


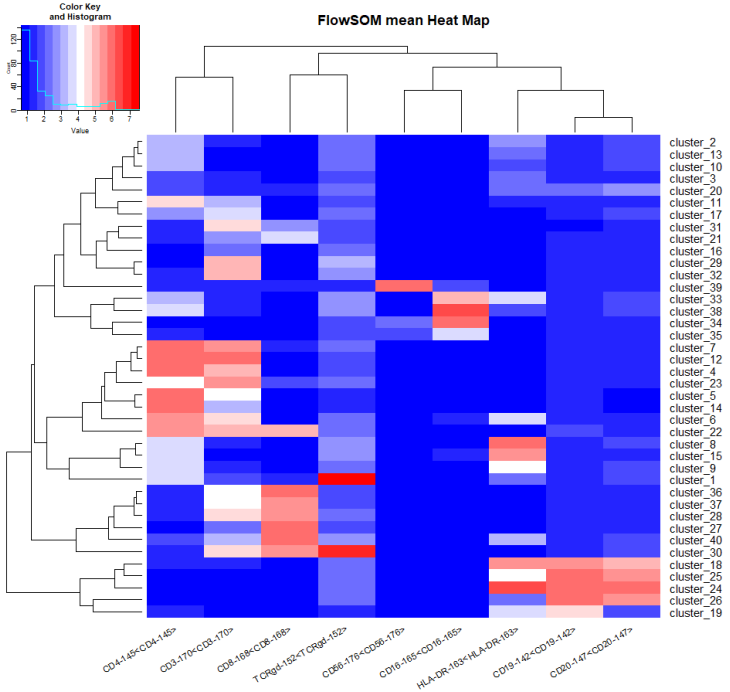
Features analysis



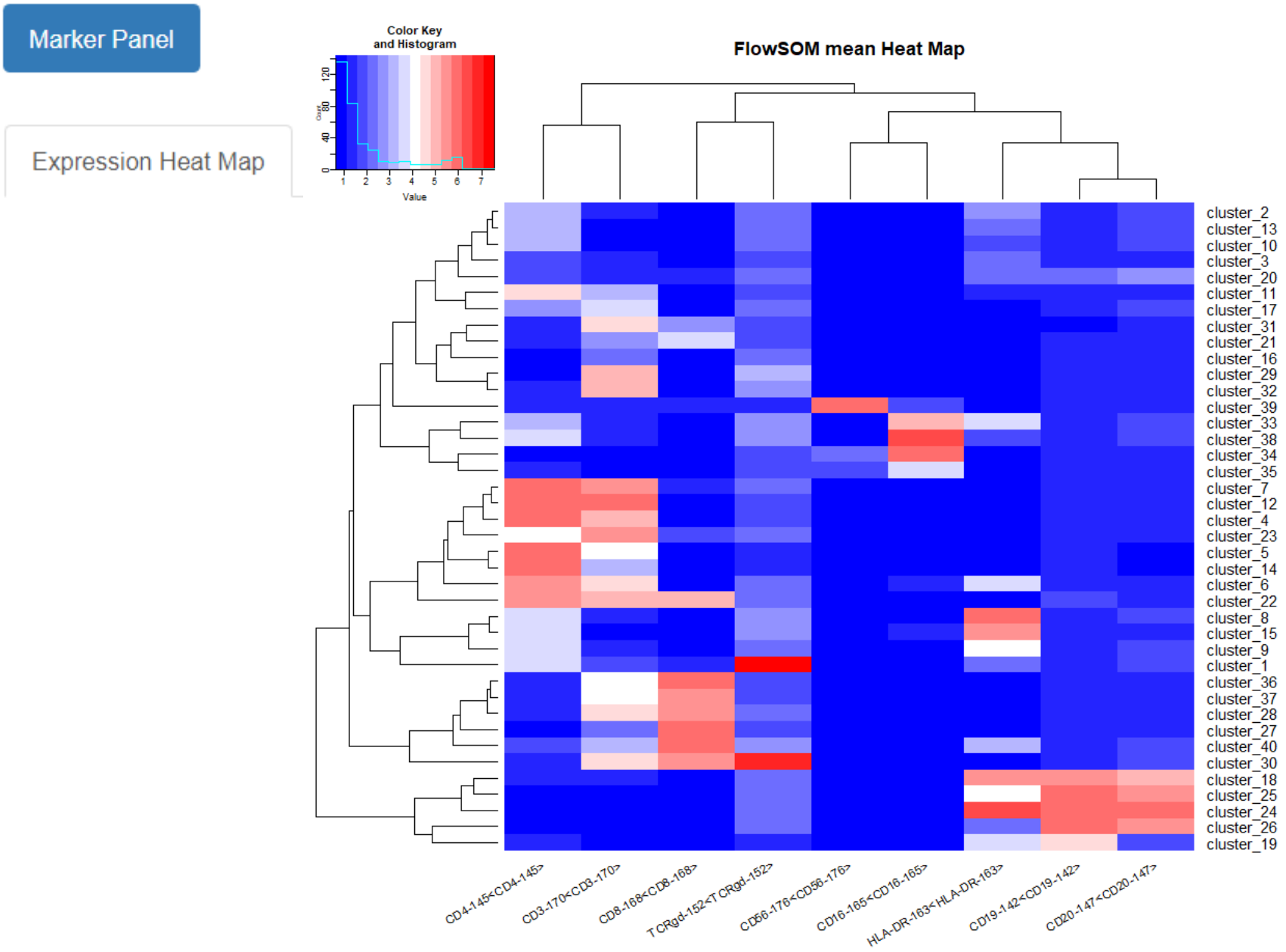
Annotating clusters



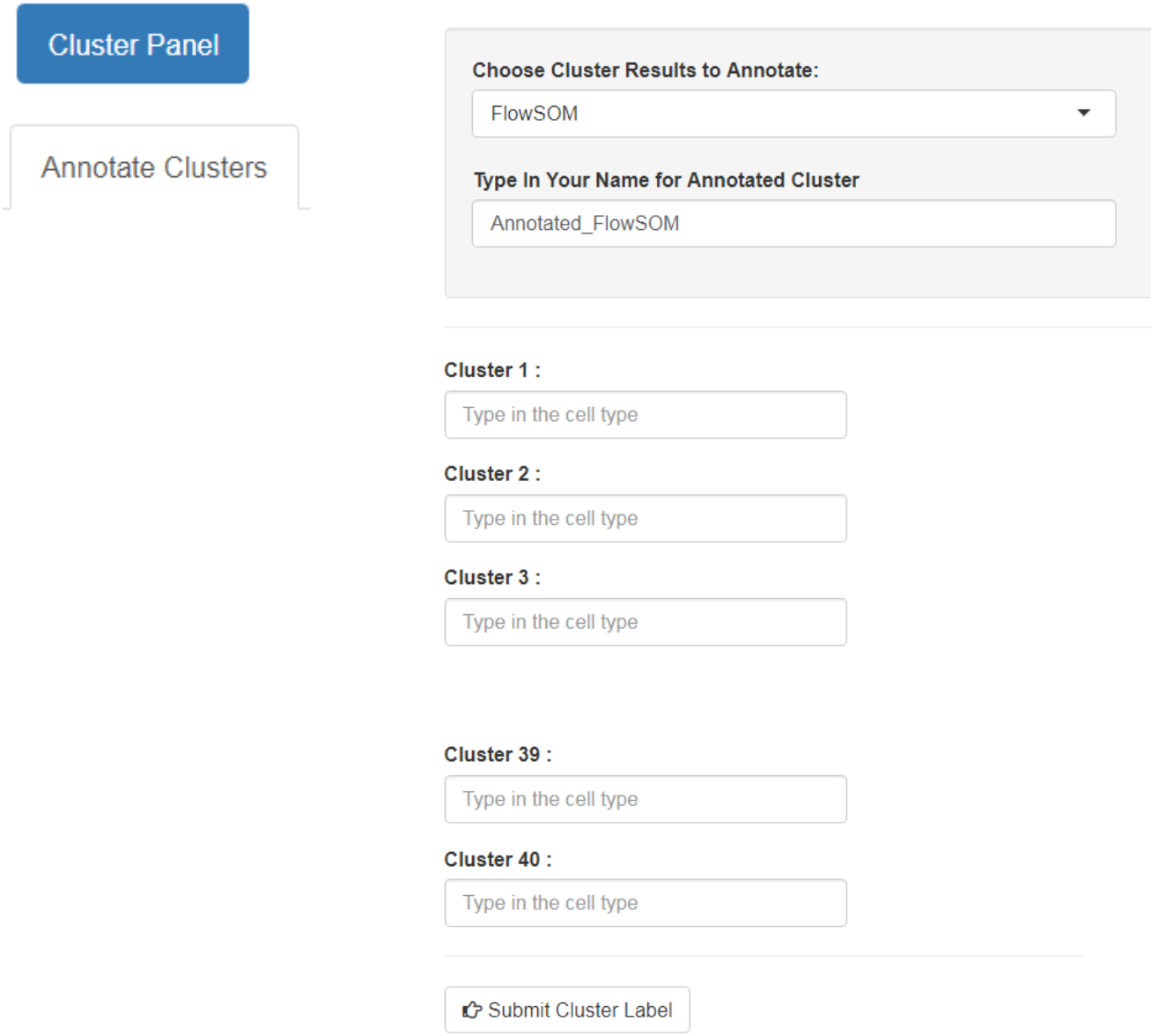




Expression heatmap of clusters



Cytofkit - Analysis



Hands on

- Merge a few clusters

Outline

- ☑ Reduce dimensions
- ☑ Clusterize cells
- ☑ Merge clusters
- ☑ Annotate clusters

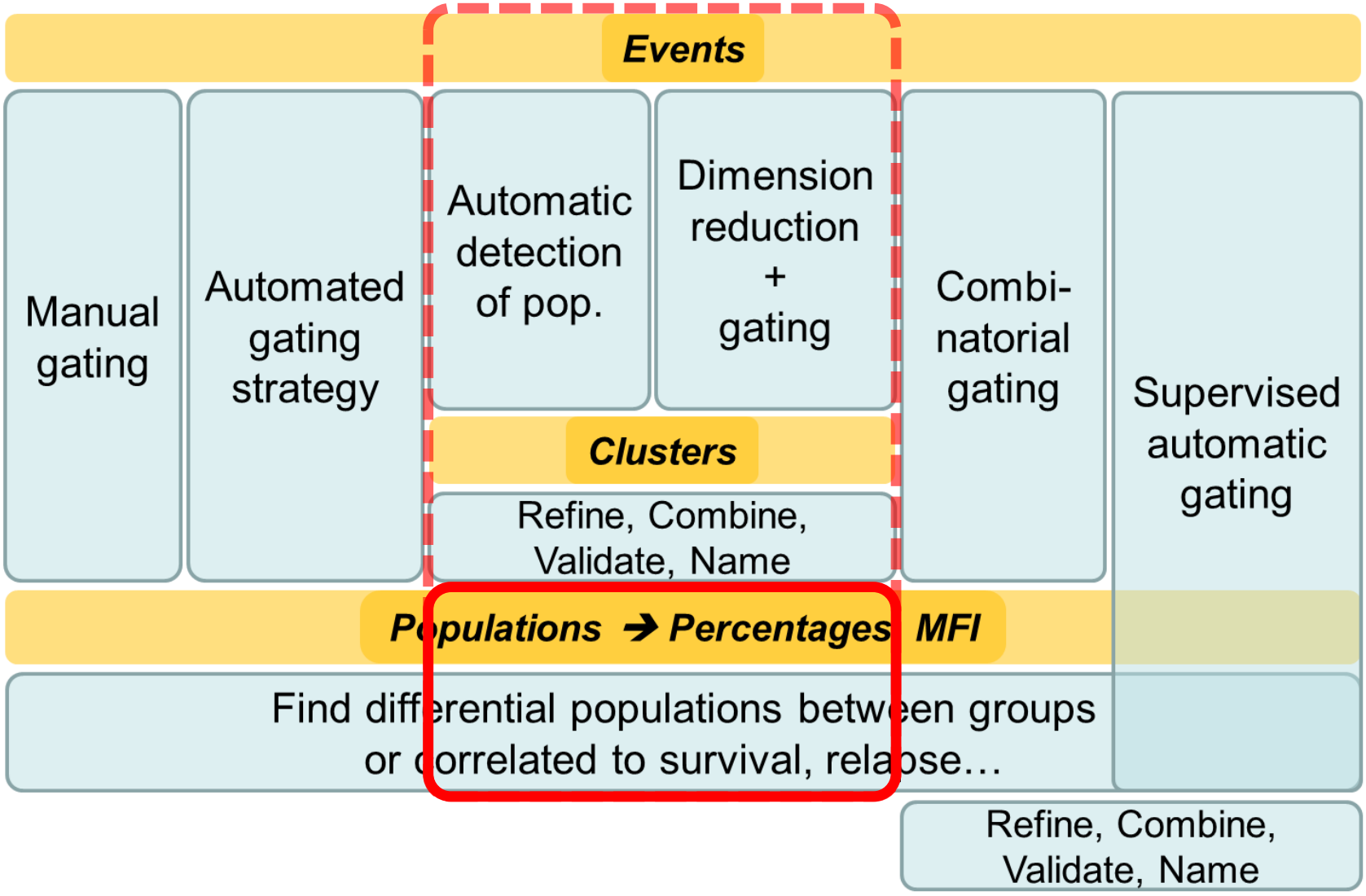
What have we done?

Outline

- ☑ Reduce dimensions
- ☑ Clusterize cells
- ☑ Merge clusters
- ☑ Annotate clusters
- ☛ Unsupervised gating
- ☛ Data driven cell populations with expert annotation

What's next?

Features analysis



Back to the question

- Find a group of cells that differ in abundance between two groups of patients

- Try to view differences on tSNE (or UMAP)

tSNE split by samples

Cluster Panel

Cluster Plot

Visualization Meth

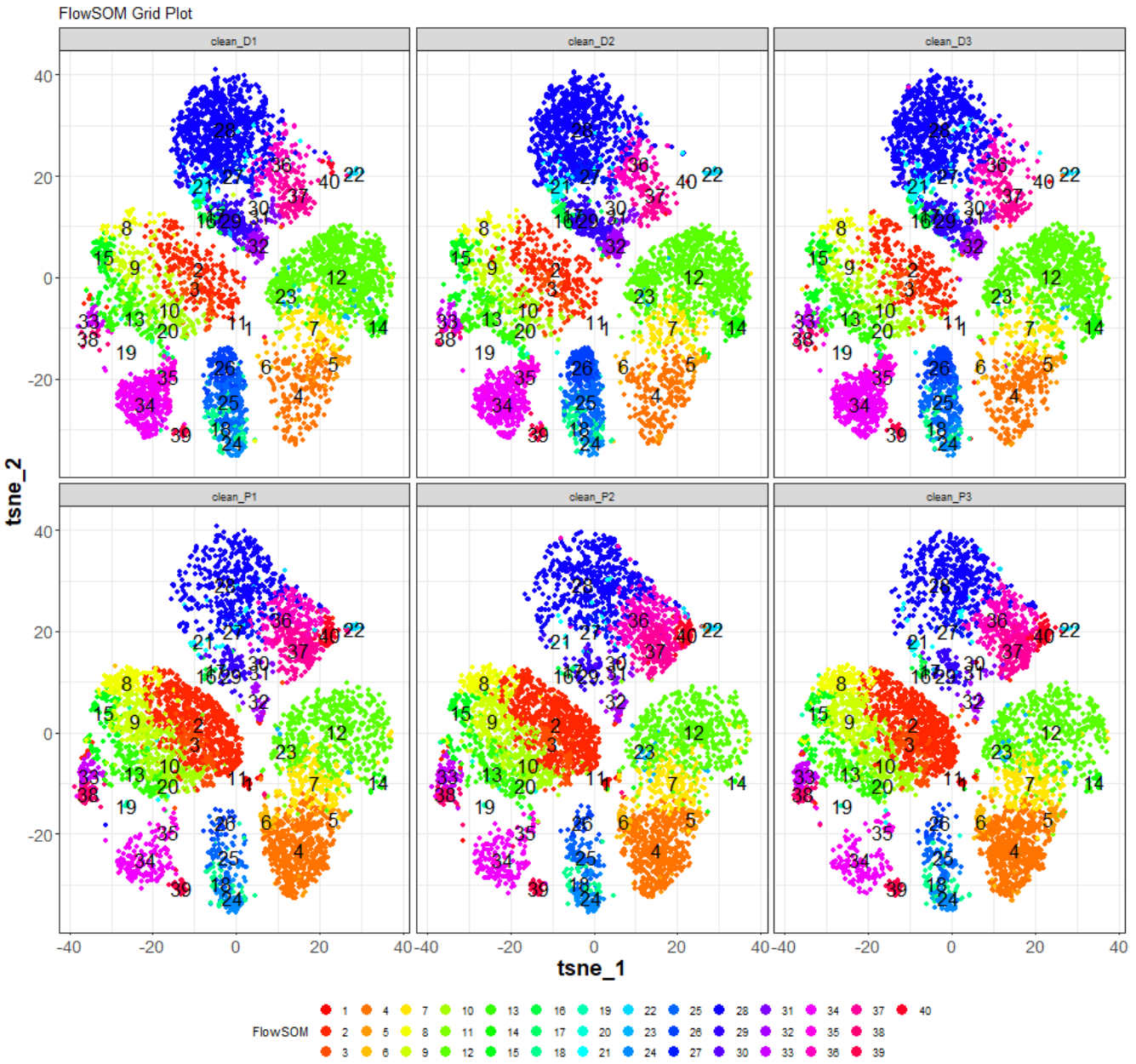
tsne

Cluster By:

FlowSOM

Plot Control:

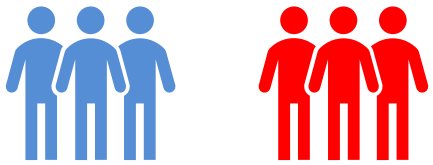
- [x] Add Cluster Labels
- [ ] Repel Cluster Labels
- [x] Separate Plot by Samples



Back to the question

- Find a group of cells that differ in abundance between two groups of patients

- What would you like to see/do?



Pool samples

Sample Panel

Regroup Samples

Type in the Group Name for Each Sample:

clean_D1 :

D_status

clean_D2 :

D_status

clean_D3 :

D_status

clean_P1 :

P_status

clean_P2 :

P_status

clean_P3 :

P_status

Group Name Levels: (to order the group names)

Type in group names in order, seperated by semicolon(;)

Submit New Sample Groups

Revert to Old Sample Names

Hands on

- Pool samples in order to get two meta-samples
- View side-by-side meta-samples

- Visualize differences on tSNE (or UMAP)

Side-by-side tSNE

Cluster Panel

Cluster Plot

Visualization Meth

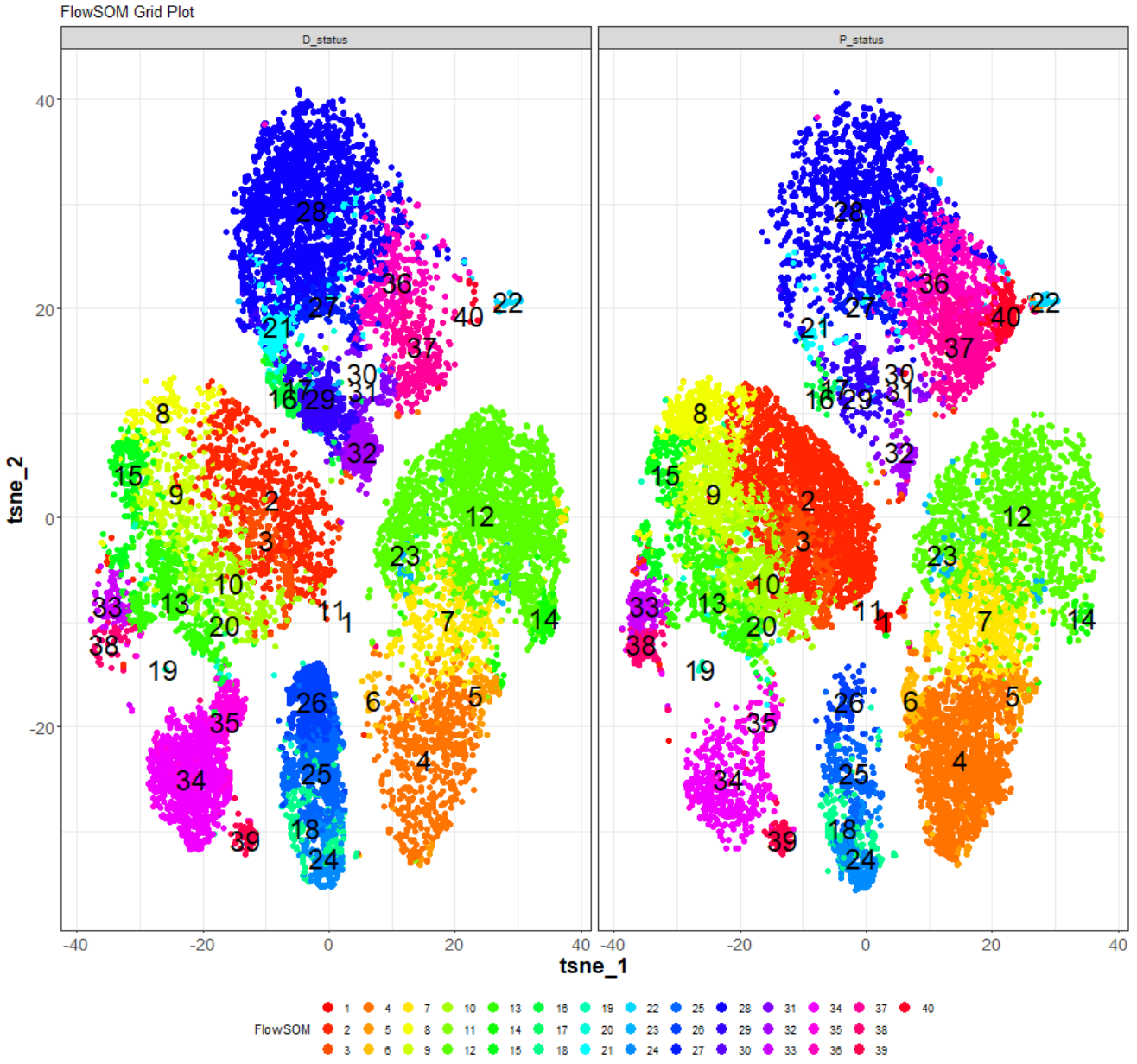
tsne

Cluster By:

FlowSOM

Plot Control:

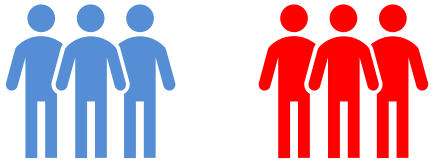
- [x] Add Cluster Labels
- [ ] Repel Cluster Labels
- [x] Separate Plot by Samples



Outline

- ☑ Reduce dimensions
- ☑ Clusterize cells
- ☑ Merge clusters
- ☑ Annotate clusters
- ☛ Unsupervised gating, cell populations

- ☛ What's next?



Back to the objective

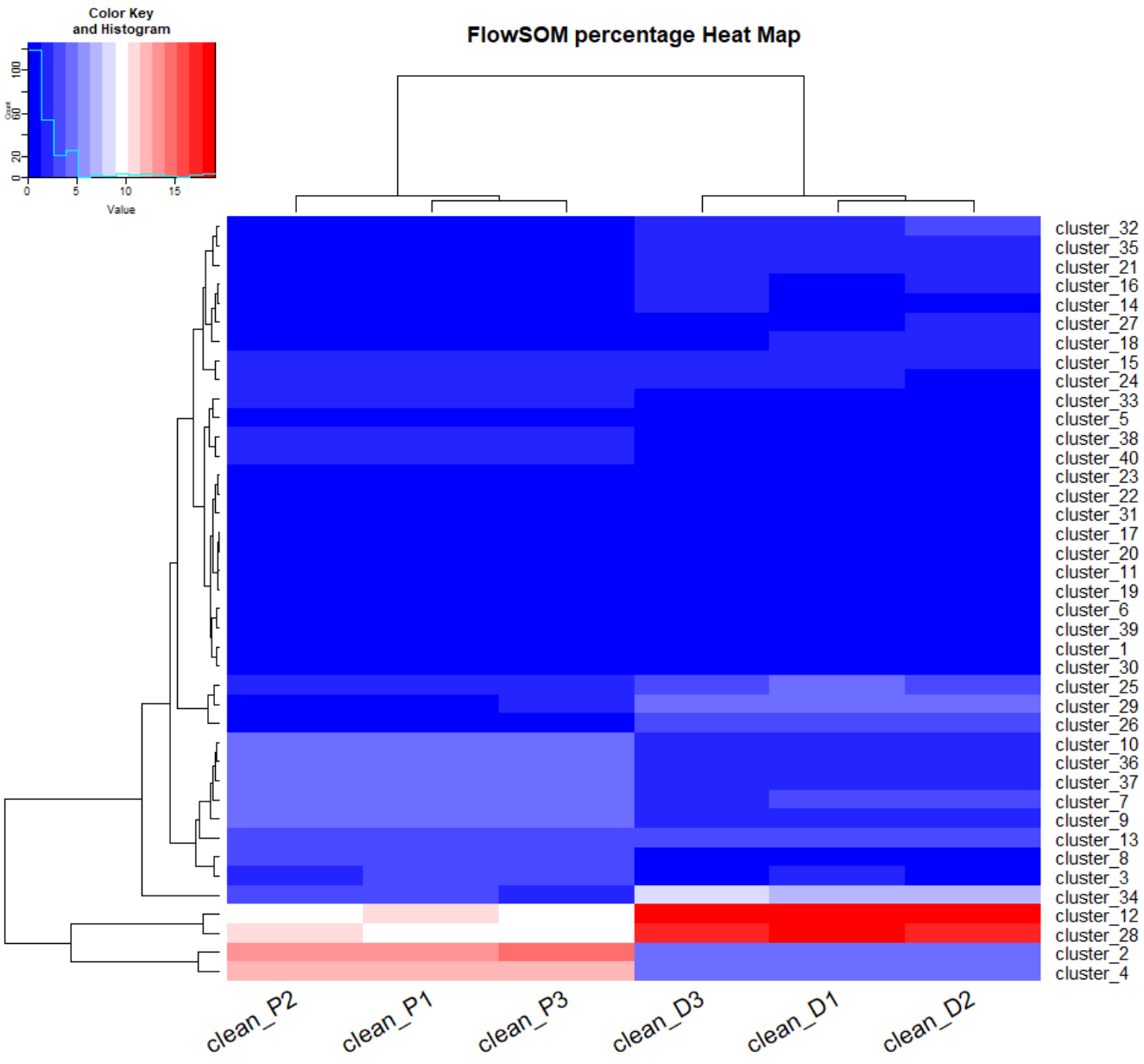
- Find a group of cells that differ in abundance between two groups of patients

- How to quantify the differences?
- Which feature to extract?

Cell percentage heatmap

Sample Panel

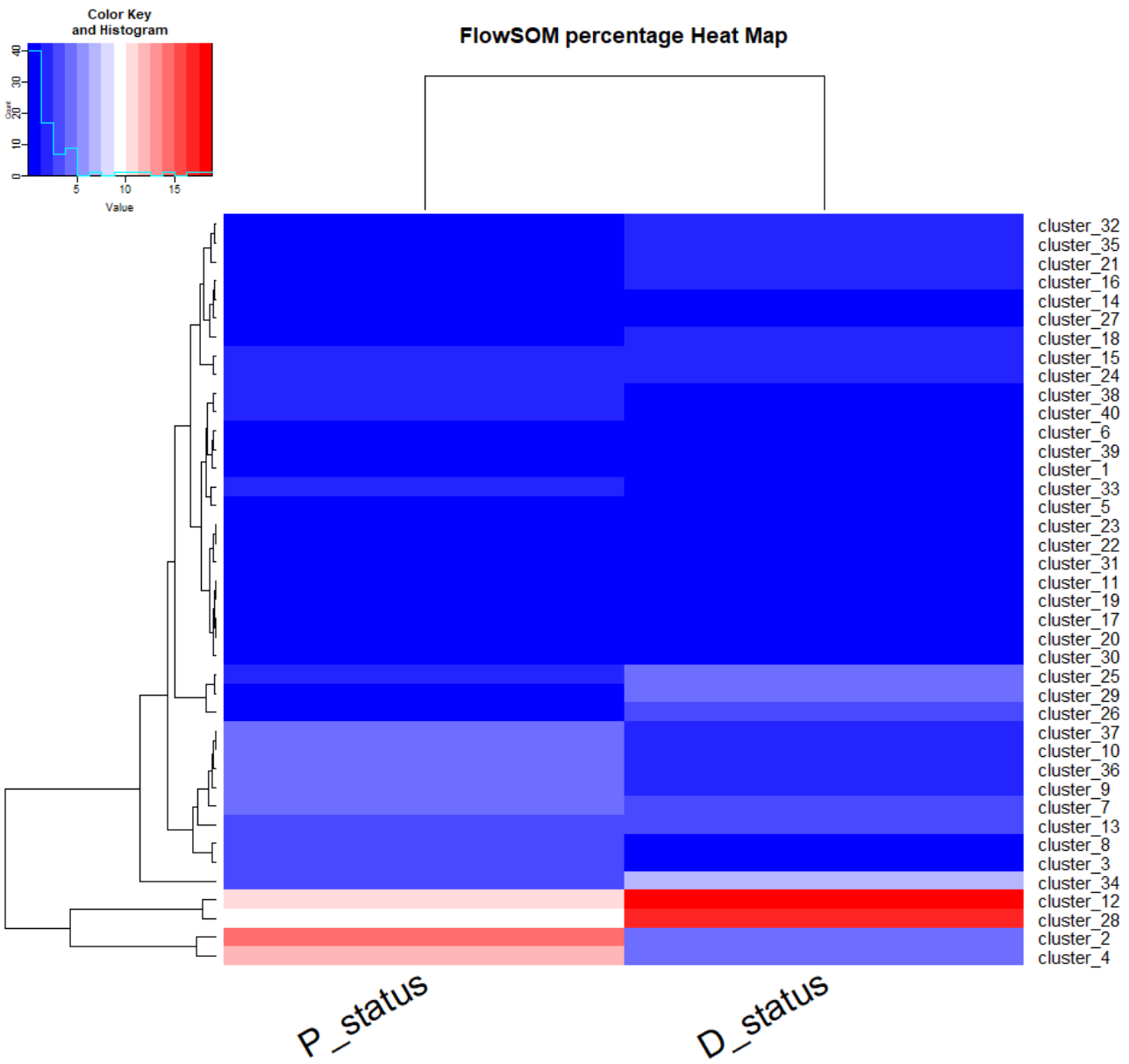
Cell Percentage Heatmap



Cell percentage heatmap

Sample Panel

Cell Percentage Heatmap

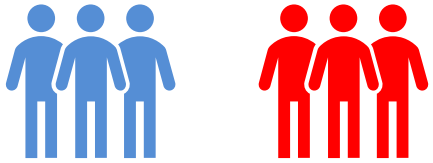


Outline

- ☛ Unsupervised gating, cell populations

- ☛ Cell percentage analysis
- ☑ Extract percentages
- ☑ Visualize percentages as heatmap
- ☑ Clusterize percentages of cell populations

- ☛ What's next?

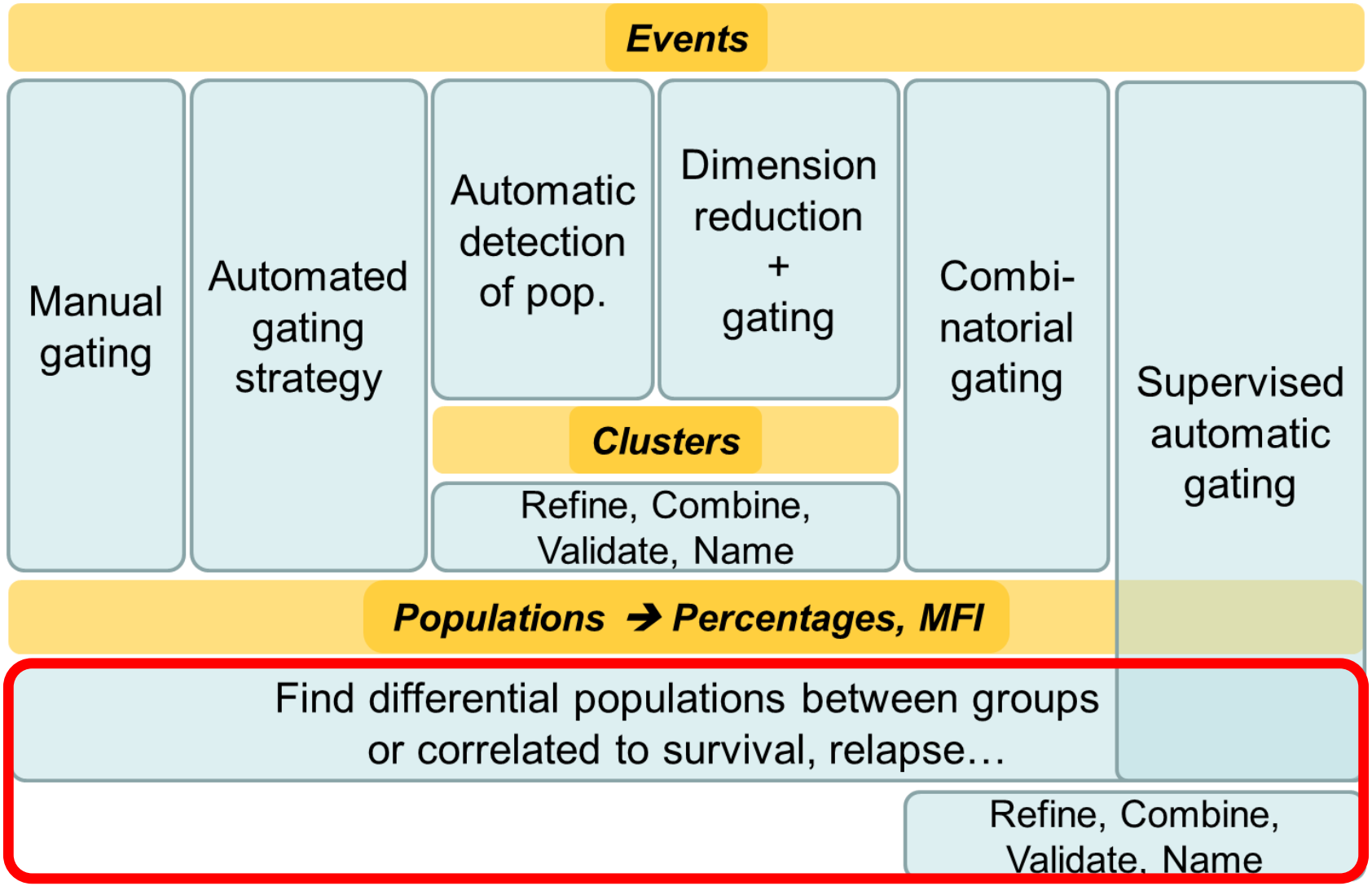


Outline

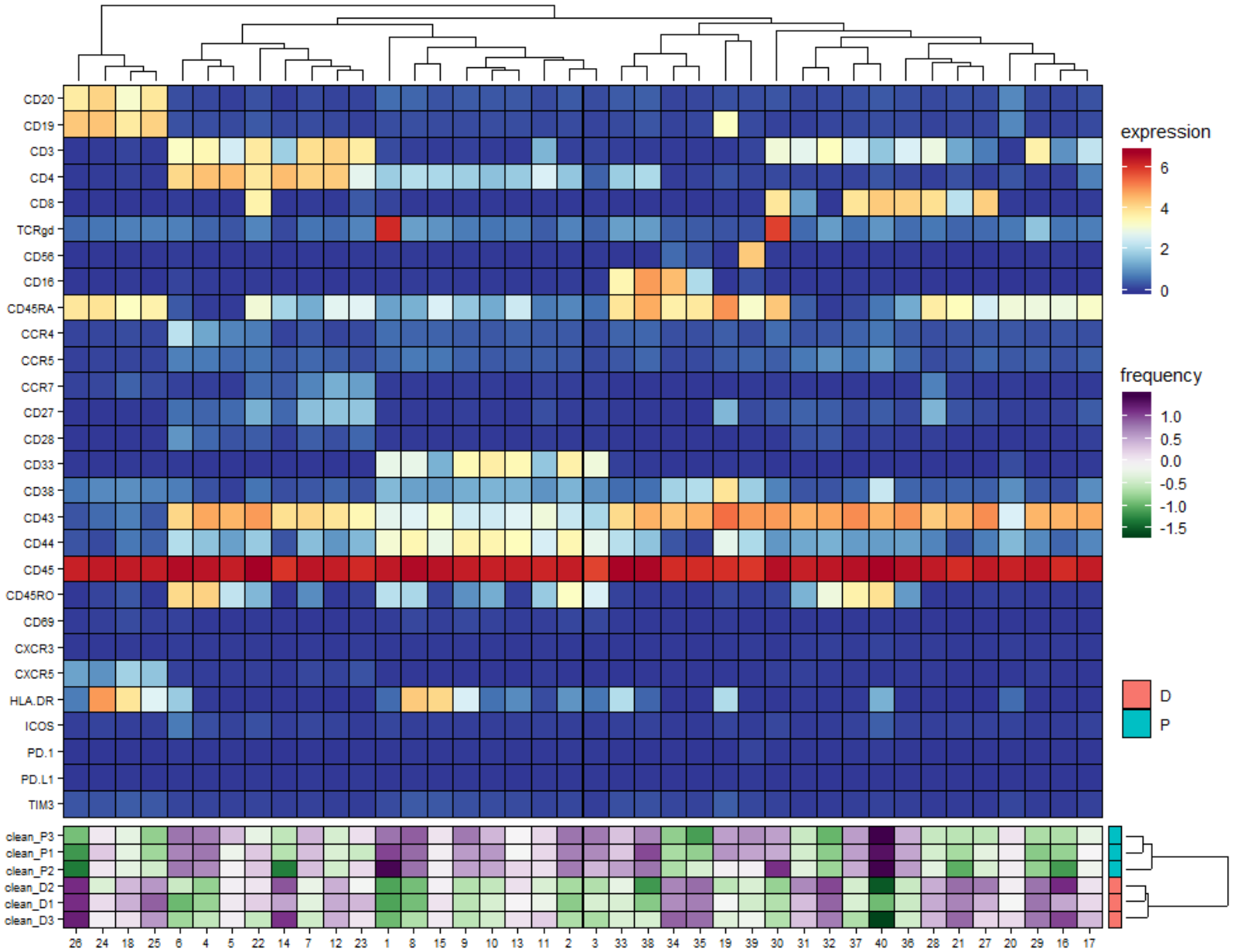
- ☛ Unsupervised gating, cell populations
- ☛ Cell percentage analysis
- ☛ What's next?
- ☐ Publication ready figures
- ☐ Box-plots, p-values
- ☐ What else?

PERCENTAGES ANALYSIS CYTOFAST

Features analysis



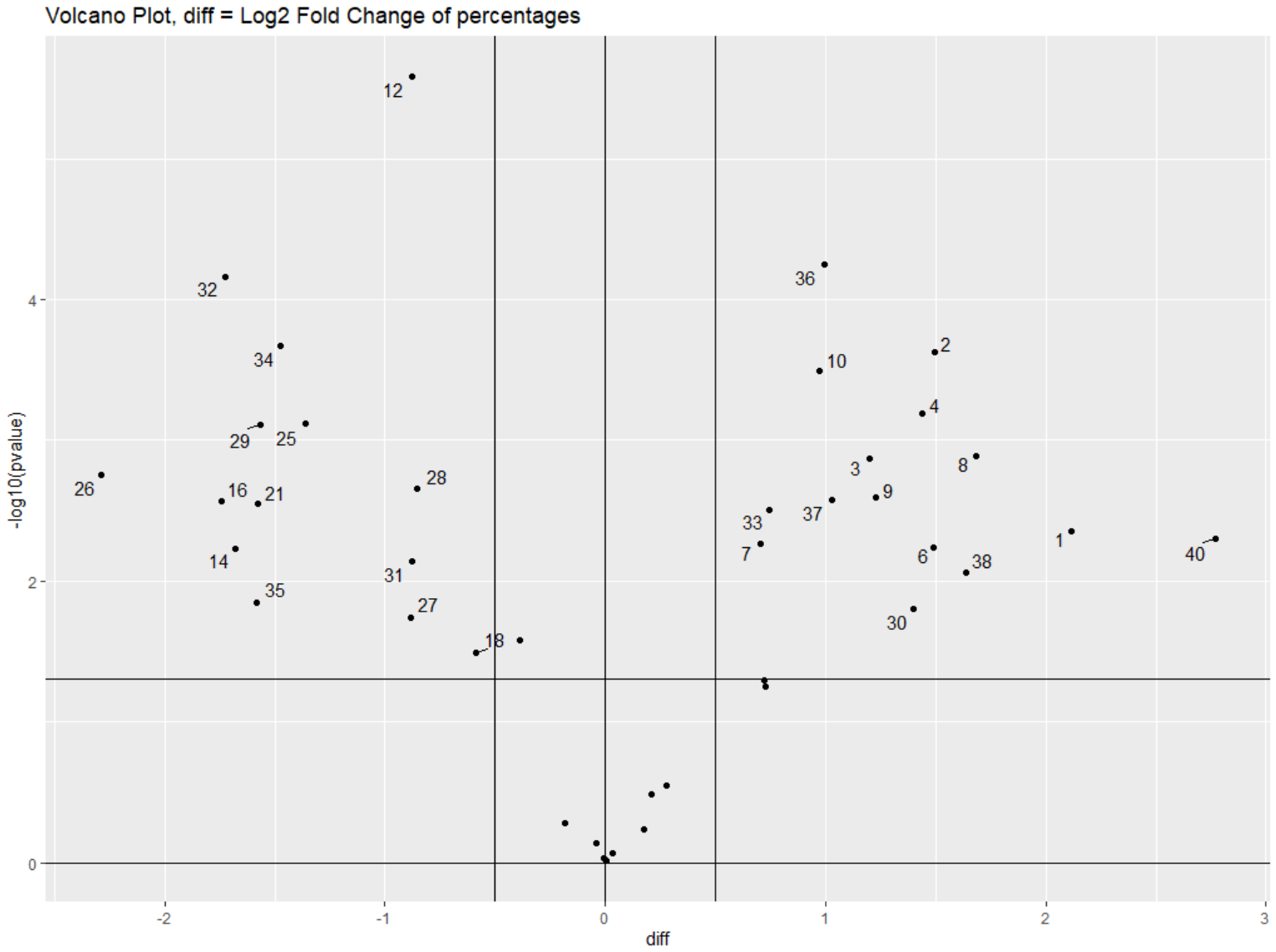
Prettier heatmaps



Boxplots


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
value
grouping
grouping
D
P

Multiple testing



Statistical tests

	pvalue	adjusted	fold	diff	clean_D1	clean_D2	clean_D3	clean_P1	clean_P2	clean_P3
1	0.0044447	0.0093572	4.33	2.1132617	0	0	1	33	46	27
2	0.0002351	0.0018806	2.82	1.4959178	227	247	254	687	701	718
3	0.0013465	0.0048965	2.29	1.1976641	66	55	53	152	124	163
4	0.0006430	0.0034625	2.71	1.4362518	207	205	242	628	580	609
6	0.0057602	0.0103445	2.81	1.4921239	9	16	12	50	53	54
7	0.0054537	0.0103445	1.63	0.7038460	130	138	116	211	211	221
8	0.0012854	0.0048965	3.22	1.6858110	40	40	50	160	156	167
9	0.0025327	0.0066377	2.34	1.2273065	88	85	88	208	204	241
10	0.0003242	0.0021612	1.96	0.9721761	113	103	110	238	224	207
12	0.0000026	0.0001030	-1.84	-0.8776517	956	939	932	517	499	509
14	0.0059481	0.0103445	-3.21	-1.6819130	43	61	69	14	5	15
16	0.0027074	0.0066377	-3.35	-1.7449611	46	77	70	13	8	15
18	0.0321336	0.0428447	-1.50	-0.5888729	79	76	60	44	44	44
21	0.0028210	0.0066377	-2.98	-1.5757889	86	95	97	26	19	29
25	0.0007619	0.0034625	-2.57	-1.3591659	233	187	186	68	85	65
26	0.0017456	0.0058185	-4.88	-2.2860022	175	176	186	28	24	35
27	0.0182757	0.0261081	-1.85	-0.8836710	56	73	46	30	27	23
28	0.0021898	0.0066377	-1.81	-0.8560822	901	890	846	474	525	447
29	0.0007791	0.0034625	-2.96	-1.5641741	204	205	202	59	61	67
30	0.0157821	0.0233808	2.64	1.3991602	1	0	1	14	27	15
31	0.0071710	0.0119516	-1.83	-0.8747789	30	40	29	13	15	12
32	0.0000683	0.0009111	-3.31	-1.7275874	120	132	109	30	32	26
33	0.0031106	0.0069124	1.67	0.7438357	37	42	37	66	81	68
34	0.0002145	0.0018806	-2.78	-1.4731193	346	357	407	129	132	119
35	0.0142354	0.0219006	-2.99	-1.5813353	106	114	114	30	42	22
36	0.0000561	0.0009111	1.99	0.9931444	101	100	104	212	220	205
37	0.0026399	0.0066377	2.04	1.0262677	114	111	94	230	217	232
38	0.0087517	0.0140028	3.11	1.6349654	16	11	25	83	71	66
40	0.0049719	0.0099439	6.81	2.7685644	8	2	1	76	83	84

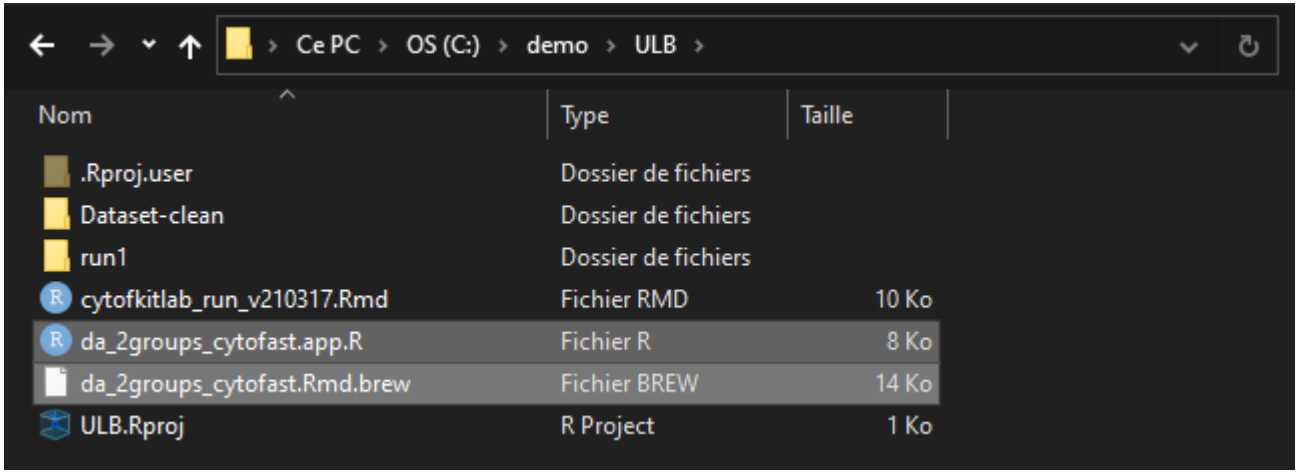
Cytofast - Analysis

- Designed for cytosplore
- Available as R commands
- Read FCS files with cell annotations
  - new channels holding tSNE, UMAP dimensions
  - new channels holding clustering results

- Adapted to read cytofkit results
- Analysis available as a R MarkDown file
- Diff Abund, 2 groups

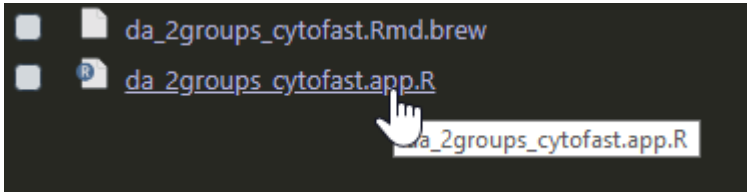
Prepare analysis

Copy the two files to carry out a differential abundance analysis in the Rstudio project folder where the FCS files and the cytofkit results are.

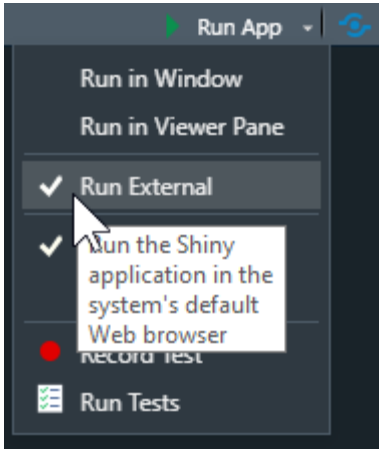


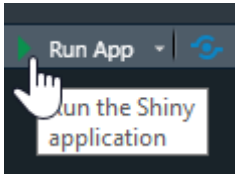
Prepare analysis

Run the Shiny application to start the analysis



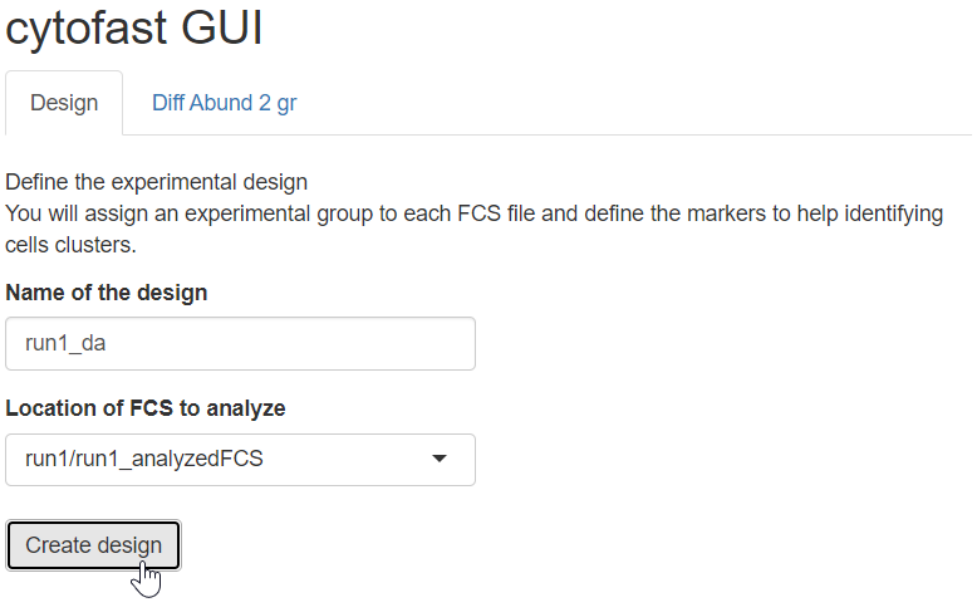
Verify the application will run in an external Web browser (prefer Firefox or Chrome), then run

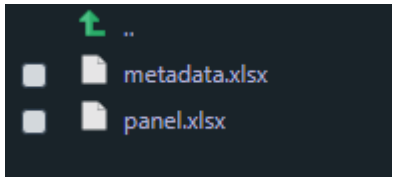


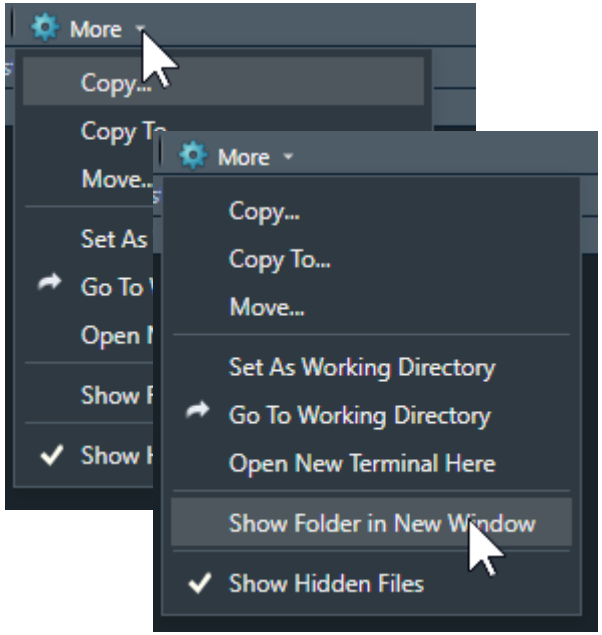


Design the analysis

Create a template design, then edit it to fit your experimental design







Design

panel.xlsx

Edit the two files to inform about the experimental design:

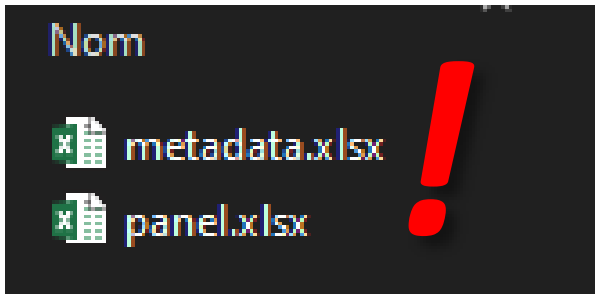
- metadata: fulfill the condition column,
- panel: remove unneeded markers/rows, change names in the antigen column

metadata.xlsx

file_name	sample_id	patient_id	condition
run1/run1_analyzedFCS/cytofkit_clean_D1.fcs	cytofkit_clean_D1	1	control
run1/run1_analyzedFCS/cytofkit_clean_D2.fcs	cytofkit_clean_D2	2	control
run1/run1_analyzedFCS/cytofkit_clean_D3.fcs	cytofkit_clean_D3	3	control
run1/run1_analyzedFCS/cytofkit_clean_P1.fcs	cytofkit_clean_P1	1	treated
run1/run1_analyzedFCS/cytofkit_clean_P2.fcs	cytofkit_clean_P2	2	treated
run1/run1_analyzedFCS/cytofkit_clean_P3.fcs	cytofkit_clean_P3	3	treated

	A	B	C
1	fcs_colname	antigen	marker_class
2	146	CD64	type
3	151	CD14	type
4	161	CD32	type
5	169	CD25	type
6	172	CD57	type
7	CCR4-149	CCR4-149	type
8	CCR5-144	CCR5-144	type
9	CCR7-159	CCR7-159	type
10	CD16-165	CD16-165	type
11	CD19-142	CD19-142	type
12	CD20-147	CD20-147	type
13	CD27-167	CD27-167	type
14	CD28-160	CD28-160	type
15	CD3-170	CD3-170	type
16	CD33-158	CD33-158	type
17	CD38-148	CD38-148	type
18	CD4-145	CD4-145	type
19	CD43-150	CD43-150	type
20	CD44-166	CD44-166	type
21	CD45-154	CD45-154	type
22	CD45RA-153	CD45RA-153	type
23	CD45RO-164	CD45RO-164	type
24	CD56-176	CD56-176	type
25	CD69-162	CD69-162	type
26	CD8-168	CD8-168	type
27	CXCR3-156	CXCR3-156	type
28	CXCR5-171	CXCR5-171	type
29	HLA-DR-163	HLA-DR-163	type
30	ICOS-141	ICOS-141	type
31	NA-191	NA-191	type
32	NA-193	NA-193	type
33	PD-1-174	PD-1-174	type
34	PD-L1-175	PD-L1-175	type
35	TCRgd-152	TCRgd-152	type
36	TIM3-143	TIM3-143	type
37	pca_1	<NA>	dimred

Create script and report



cytofast GUI

Design | Diff Abund 2 gr

Set up the differential abundance analysis
Define the parameters of the analysis for grouping samples, reporting cells clusters statistica

Name of the analysis

da2

Design to use

0480D5CC-da_2gr_1

Asinh cofactor for scaling

5

Clustering to report

Rphenograph_clusterIDs

Column to group FCS

condition

Reference group

control

P-value threshold

0.05

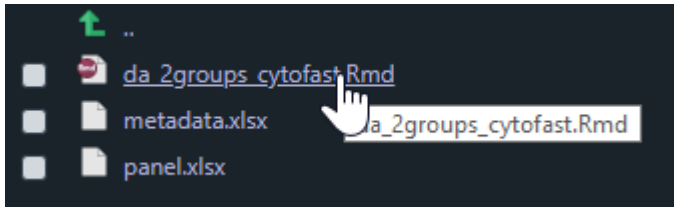
Logfold threshold

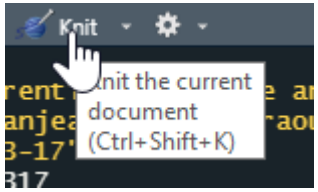
1

Create script

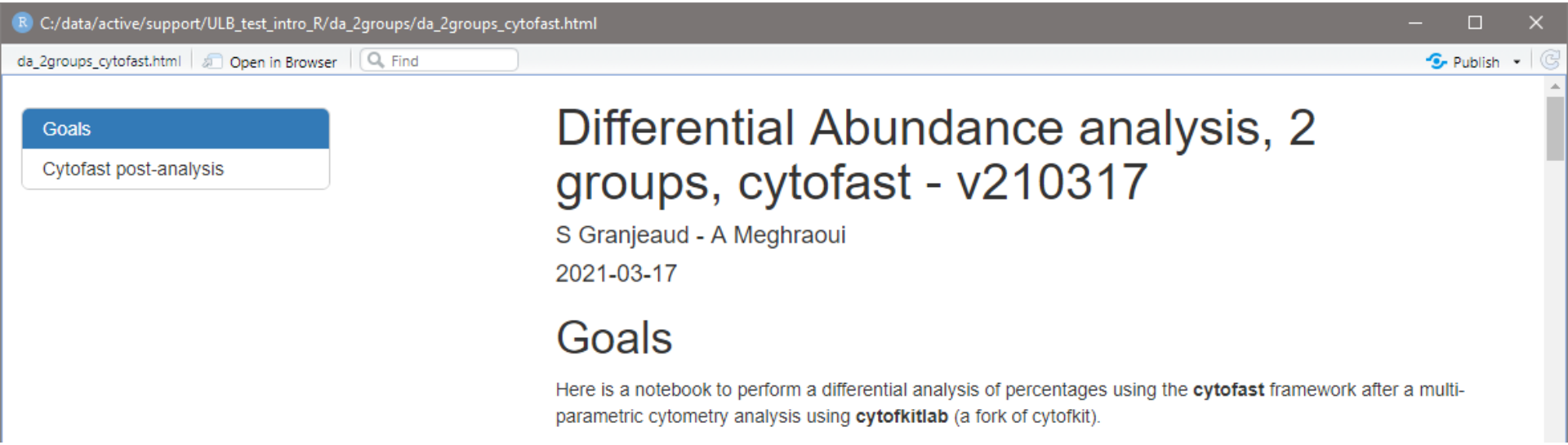
Download report

Knit script



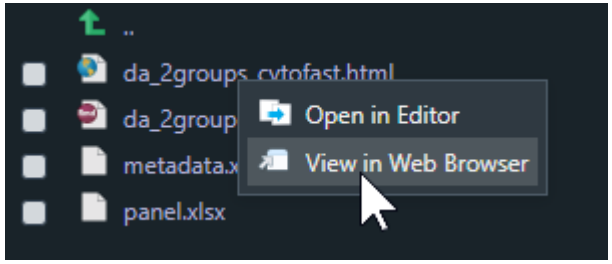


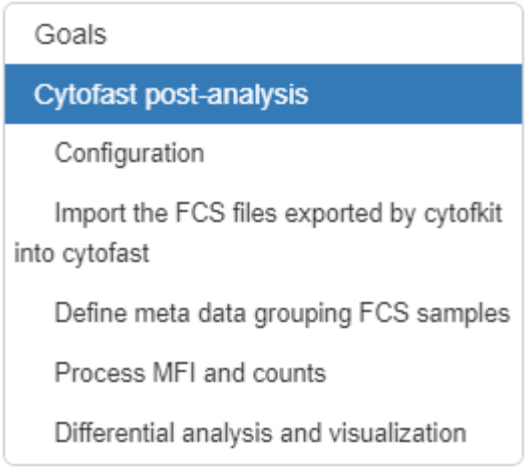
The HTML report is automatically created with the Shiny application.
You can tweak the Rmd script if needed.
Knit it to generate a new HTML report.



View results

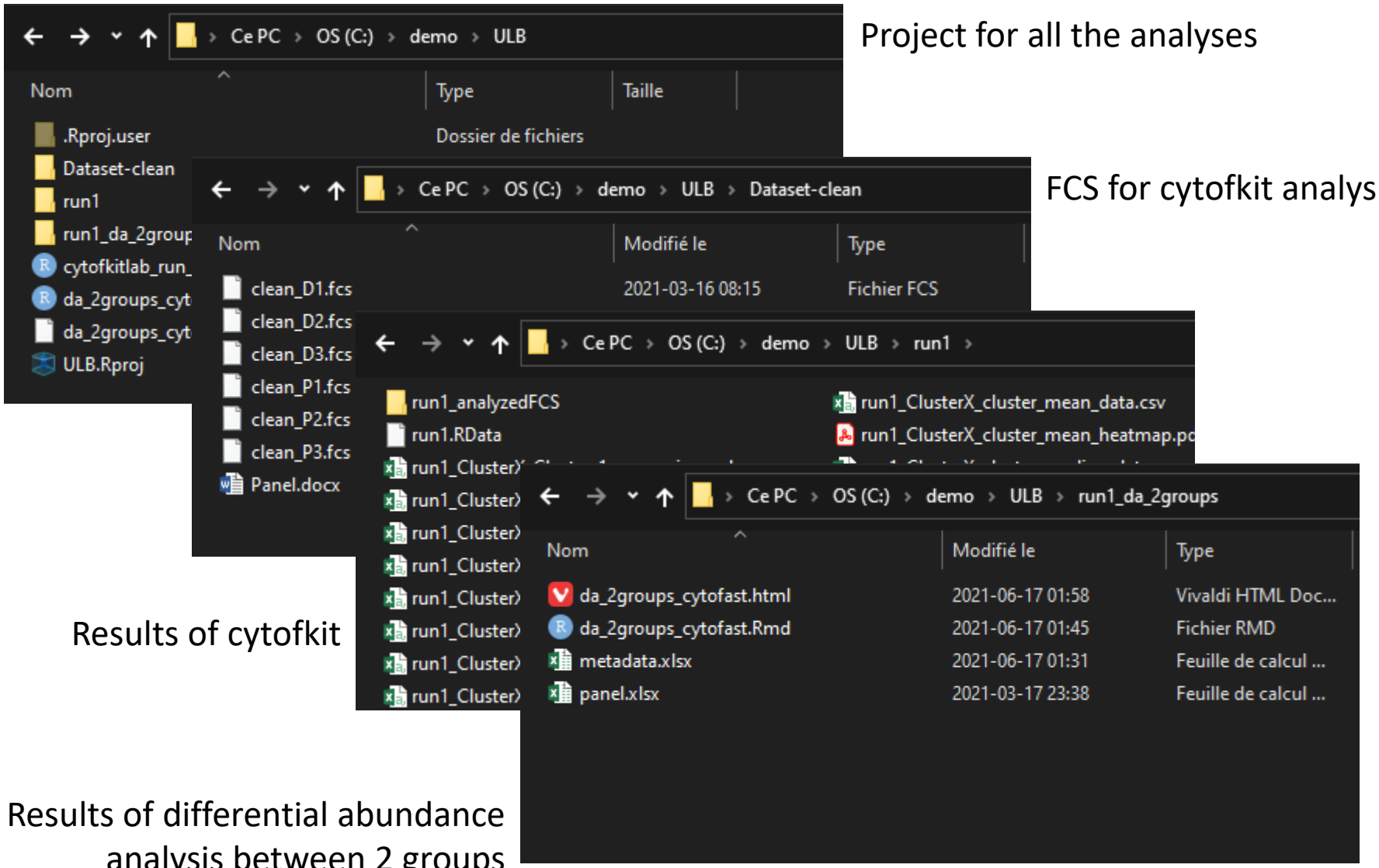
Open the HTML report in a Web browser and navigate in the table of content





Stop the R Shiny application

Analysis folder



Project for all the analyses

FCS for cytofkit analysis

Results of cytofkit

Results of differential abundance analysis between 2 groups

Final word

- ☛ Unsupervised gating, cell populations
- ☛ Cell percentage analysis
- ☛ Better and deeper analyses with cytofast
- ☐ What else?