

Video Article

Visualization and Quantification of High-Dimensional Cytometry Data using Cytofast and the Upstream Clustering Methods FlowSOM and Cytosplore

Guillaume Beyrend¹, Koen Stam², Ferry Ossendorp¹, Ramon Arens¹¹Department of Immunohematology and Blood Transfusion, Leiden University Medical Center²Department of Parasitology, Leiden University Medical CenterCorrespondence to: Guillaume Beyrend at g.beyrend@lumc.nl, Ramon Arens at r.arenslumc.nlURL: <https://www.jove.com/video/60525>DOI: [doi:10.3791/60525](https://doi.org/10.3791/60525)

Keywords: Immunology and Infection, Issue 154, flow and mass cytometry analysis, R package, high-dimensional analysis, single-cell analysis, visualization tool, data mining

Date Published: 12/12/2019

Citation: Beyrend, G., Stam, K., Ossendorp, F., Arens, R. Visualization and Quantification of High-Dimensional Cytometry Data using Cytofast and the Upstream Clustering Methods FlowSOM and Cytosplore. *J. Vis. Exp.* (154), e60525, doi:10.3791/60525 (2019).

Abstract

The complexity of data generated by mass cytometry has necessitated new tools to rapidly visualize analytic outcomes. Clustering methods like *Cytosplore* or *FlowSOM* are used for the visualization and identification of cell clusters. For downstream analysis, a newly developed R package, *Cytofast*, can generate a rapid visualization of results from clustering methods. *Cytofast* takes into account the phenotypic characterization of cell clusters, calculates the cell cluster abundance, then quantitatively compares groups. This protocol explains the applications of *Cytofast* to the use of mass cytometry data based on modulation of the immune system in the tumor microenvironment (i.e., the natural killer [NK] cell response) upon tumor challenge followed by immunotherapy (PD-L1 blockade). Demonstration of the usefulness of *Cytofast* with *FlowSOM* and *Cytosplore* is shown. *Cytofast* rapidly generates visual representations of group-related immune cell clusters and correlations with immune system composition. Differences are observed in the clustering analysis, but separation between groups are visible with both clustering methods. *Cytofast* visually shows the patterns induced by PD-L1 treatment that include a higher abundance of activated NK cell subsets, expressing a higher intensity of activation markers (i.e., CD54 or CD11c).

Video Link

The video component of this article can be found at <https://www.jove.com/video/60525/>

Introduction

Mass cytometry (cytometry by time-of-flight, or CyTOF) allows detection of a wide range of intracellular or extracellular biomarkers in millions of single cells. The high-dimensional nature of mass cytometry data necessitates certain analysis tools, such as cell clustering techniques like *SPADE*¹, *FlowMaps*², *FlowSOM*³, *Phenograph*⁴, *VorteX*⁵, and *scaffold maps*⁶. In addition, various dimensionality reduction-based techniques have been developed (i.e., principal component analysis [PCA]⁷, t-distributed stochastic neighbor embedding [t-SNE]⁸, hierarchical stochastic neighbor embedding [HSNE]⁹, uniform manifold approximation and projection [UMAP]¹⁰, and diffusion maps¹¹) to improve the speed, interpretation and visualization of high-dimensional datasets.

Downstream analysis of high-dimensional flow and mass cytometric data often lacks automatic processes to perform statistical tests on cluster frequency and links with clinical outcomes. Previously, we developed an R-based workflow known as *Cytofast*¹², which allows for visual and quantitative downstream analyses of clustering techniques by *Cytosplore* or *FlowSOM*.

The protocol described here clarifies the use of *Cytofast* in R and shows how to generate quantitative and qualitative heatmaps and graphs. Furthermore, it facilitates the determination of connections between observed immune phenotypes and clinical outcomes. This report also describes the analysis of a specific mass cytometry dataset using two different clustering procedures: *FlowSOM* and *Cytosplore*. By using *Cytofast* with both clustering methods, it is correspondingly shown that the activation phenotype of NK cells is influenced by PD-L1 immune checkpoint blockade.

Protocol

All animal experiments were approved by the Animal Experiments Committee of LUMC and were executed according to the animal experimentation guidelines of the LUMC in compliance with the guidelines of Dutch and European committees.

NOTE: For experimental set-up, C57BL/6 mice were subcutaneously inoculated on the right flank with the murine colon tumor MC38 at a concentration of 0.3×10^6 cells/200 μ L of phosphate-buffered saline (PBS). After 10 days, when tumors were palpable, the mice were treated with PD-L1 blocking antibodies (clone MIH-5, 200 μ g/mouse, intra-peritoneal injection) or were mock-treated. The tumors were resected 3 days later after PD-L1 injection, processed ex vivo, and analyzed by CyTOF mass cytometry using 38 markers¹³.

1. Equipment and Software for Data Analysis

NOTE: Use a computer (Windows 7 or newer) and processor I5 at 2.4 GHz or equivalent, installed memory RAM 6 GB, and 10 GB of free hard drive space. The R package *Cytofast* uses existing functions: mainly *flowCore*, *heatmap*, and *ggplot*. The command lines to be executed in R are included in the protocol. The resource for R instructions can be found at <https://education.rstudio.com/>.

- To install the *Cytofast* package, start R (version "3.6") and install *Bioconductor* version 3.9 by entering the following code:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("cytofast")
```
- Ensure that the package is loaded in the desired environment by running the following:

```
library(cytofast)
```

2. Creating Clusters

NOTE: To showcase the two clustering methods *Cytosplore* and *FlowSOM* with *Cytofast*, the NK cells (CD161⁺) in the tumor micro-environment 3 days after PD-L1 treatment are analyzed.

- Clustering performed by *Cytosplore*
 - After downloading and installing *Cytosplore*, hosted at <www.Cytosplore.org>, upload the .fcs files (**Supplementary Files 1.1–1.8** [Cytosplore input files]) by clicking **File | Open FCS File(s)** in *Cytosplore*. Add a unique sample tag as channel by clicking **Add unique sample tag as channel** when prompted and select a cofactor for hyperbolic arcsinh transformation (default is 5).
 - Select **Run H-SNE**, run a HSNE level of 3, and wait for the map to be generated.
NOTE: This step may require some time, depending on the number of cells analyzed and level of HSNE chosen.
 - On the first HSNE level, check the cells that are positive for CD161. Select the CD161⁺ cells and right-click **Zoom into selection**. At the second level, repeat the procedure to reach the third level with only CD161⁺ events.
 - Once the last tSNE map is generated, save the clusters defined by *Cytosplore* by right-clicking on the tSNE map and choose **Save Clusters**. Choose the directory of the output files as prompted by *Cytosplore* and note this location, because this directory will then be used to load the .fcs files into R.
NOTE: The number of subsets can be manually changed by changing the sigma value. The sigma value is set by default at 30; however, the actual number of subsets depends on the input. Here, *Cytosplore* detected 10 different subsets, with each file representing one subset.
 - Use a simple name (characters only) when renaming the output files, which will make identification and further handling easier. Save the output files by selecting **Save**.
NOTE: After saving, a folder is being created by *Cytosplore* with the .fcs files, with each file corresponding to the identified clusters in *Cytosplore*. The next step will be loading the files into R with the help of *Cytofast*. Here, the generated output files are provided in **Supplementary Files 2.1–2.10** (Cytosplore output files).
 - Load the output files generated by *Cytosplore* into R with the designated function: `readCytosploreFCS`.

```
dirFCS <- "C:\\Users\\username\\Desktop\\tostudy"
cfData <- readCytosploreFCS(dir = dirFCS, colNames = "description")
```
 - Clean the data by removing some parameters such as "Time" and "Background". Check the position of the column related to its unnecessary parameters and remove it from the matrix.

```
colnames(cfData@expr)
cfData@expr <- cfData@expr[, -c(3,4,6,8:10,46:49,51:54)]
```

NOTE: The unnecessary columns can be seen by reading the column names of the generated matrix. By running `colnames(cfData@expr)` once more, ensure that only the desired parameters are obtained.
 - Re-order the markers so that lineage markers are displayed first, followed by functional markers.

```
cfData@expr <- cfData@expr[, c(1,2,3,35,36,31,9,10,18,8,37,20,
29,40,5,30,33,11,34,14,19,
32,28,6,7,4,12,13,17,16,15,
21,22,24,25,26,27,38,39)]
```

NOTE: Step 2.1.8 is optional.
 - Link the meta datafile to the generated data from *Cytosplore* by uploading the spreadsheet metafile containing clinical information (**Supplementary File 3**).

```
library(readxl)
meta <- read_excel("C:\\Users\\username\\Desktop\\sample_id.xlsx")
cfData@samples <- data.frame(meta)
```

NOTE: The clustering being performed by *Cytosplore* is now finished. An alternative clustering option to *Cytosplore* is *FlowSOM* and is described in section 2.2. Once one of the two clustering steps is performed, proceed with the visualization step (section 3).
- Clustering performed by *FlowSOM*
 - First install *FlowSOM* in R by running the following command:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("FlowSOM")
library(FlowSOM)
```
 - Install the *flowCore* package using a similar method and load it in the environment by running the following:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
```

- ```

BiocManager::install("flowCore")
library(FlowSOM)

```
- Load the raw data provided in **Supplementary Files 4.1–4.8** (FCS FlowSOM input), which were previously gated on CD161+ events, in R with the `read.flowSet` function.

```

fcs_raw <- read.flowSet(path="C:\\Users\\username\\Desktop\\tcluster", pattern = ".fcs", transformation = FALSE, truncate_max_range = FALSE, seed=123)

```
  - Select the relevant biological markers (remove "Background" or "Time") by selecting the proper columns and transform the data in an `arcsinh5` manner as shown in the code below (here, remove columns 1, 2, 4, 5, 6, 17, 21, 24, 25, 34, 35, 37, 38, 51, which do not correspond to any biological marker). Apply a cofactor of 5 as previously applied with *Cytosplore*, by choosing **cofactor=5** in the function below.

```

fcs_raw <- fsApply(fcs_raw, function(x, cofactor=5){
 colnames(x) <- fcs_raw[[1]]@parameters@data$desc
 expr <- exprs(x)
 expr <- asinh(expr[,c(1,2,4,5,6,17,21,24,25,34,35,37,38,51)]/ cofactor)
 exprs(x) <- expr
 return(x)})

```
  - Cluster the data using the *FlowSOM* function. To compare *FlowSOM* and *Cytosplore*, choose to cluster the data in ten subsets as the output previously yielded by *Cytosplore*.

```

fsom <- FlowSOM(fcs_raw, transformFunction = FALSE, scale = FALSE,
 scaled.center = FALSE, scaled.scale = FALSE, silent = FALSE, colsToUse = c(1:37),
 nClus = 10, maxMeta = 10, importance = NULL, seed = 123)

```

**NOTE:** This can be changed manually by the user.
  - Assign each cell to its identified subset and sample ID.

```

subset_id <- as.factor(fsom$FlowSOM$map$mapping[,1])
levels(subset_id) <- fsom$metaclustering
head(subset_id)

```
  - Load the metadata file in R (available in **Supplementary File 5**) containing the group assignment and link it to the .fcs files.

```

sampleid <- read_excel("C:\\Users\\username\\Desktop\\sample_id.xlsx")
sampleid <- na.omit(sampleid)
sampleid$sampleID <- as.factor(sampleid$sampleID)
sampleid$group <- as.factor(sampleid$group)
sampleid$CSPLR_ST <- as.factor(sampleid$CSPLR_ST)
sampleid <- as.data.frame(sampleid)
names(sampleid)[3] <- "sampleID"
sampleID <- lapply(fsom$FlowSOM$metaData, function(x){rep(x[1], each = length(x[1]:x[2]))})
attr(sampleID, 'names') <- NULL
sampleID <- as.factor(unlist(sampleID))
sampleid <- data.frame(sampleid)
levels(sampleID) <- paste("ID", 1:dim(sampleid)[1], sep=" ")
df <- data.frame(subset_id, sampleID, fsom$FlowSOM$data[, c(1:37)])
rename <- data.frame(colpar=fcs_raw[[1]]@parameters@data$desc)
colnames(df) <- c("clusterID", "sampleID", rename$colpar[c(1:37)])
df$ clusterID <- as.factor(df$ clusterID)
df$sampleID <- as.factor(df$sampleID)

```
  - Create a `cfList` based on the dataframe obtained from *FlowSOM* by running the following script:

```

cfData <- cfList(samples = sampleid,
 expr = df)

```
  - Re-order the markers to appear similarly to the output from the *Cytosplore* analysis.

```

cfData@expr <- cfData@expr[,c(1,2,34,36,37,10,23,24,31,22,
 38,15,8,3,27,9,11,28,35,26,
 14,33,17,20,21,18,25,29,13,
 30,12,16,32,4,5,6,7,19,39)]

```

**NOTE:** The clustering by *FlowSOM* is now finished. Next, perform visualization of the clustering output.

### 3. Visualization: Post-processing Clustering Analysis

**NOTE:** This step is a method that is common to both clustering methods. Therefore, it can be performed after clustering either with *FlowSOM* or *Cytosplore*.

- Before creating the heatmaps, generate the count table per sample by using the function `cellCounts` as shown in the code below. Since some clusters contain fewer cells than others, scale the data per cluster by specifying "scale = TRUE" inside the function `cellCounts`, so that the dispersion among samples can be easily seen.

```

cfData <- cellCounts(cfData, frequency = TRUE, scale = TRUE)

```

**NOTE:** The data can now be visualized.
- Visualize with heatmap.

**NOTE:** One of the main functions of this package is `cytoHeatmaps`, which is used to visualize the phenotype of the created clusters as well as their heterogeneity with respect to the samples.

```

cytoHeatmaps(cfData, group="group", legend=TRUE)

```

3. Visualization with boxplots

**NOTE:** Data can be represented in a quantitative way by calling the function `cytoBoxplots`. The output of this function represents the proportion of each sample for every cluster.

1. Generate the cell count as done in step 3.1, but do not scale the data to obtain the frequency of each cluster.  
`cfData <- cellCounts(cfData, frequency = TRUE, scale = FALSE)`  
`cytoBoxplots(cfData, group = "group")`

4. Visualization with median intensity signal histogram

**NOTE:** The data can also be visualized by representing the median intensity signal histogram.

1. Visualize the expression intensity of three markers: CD45, CD11c, and CD54.
2. Check the names of the desired markers by calling the following line. Note the marker names and include it in the `msiPlot` function.  
`names(cfData@expr)`

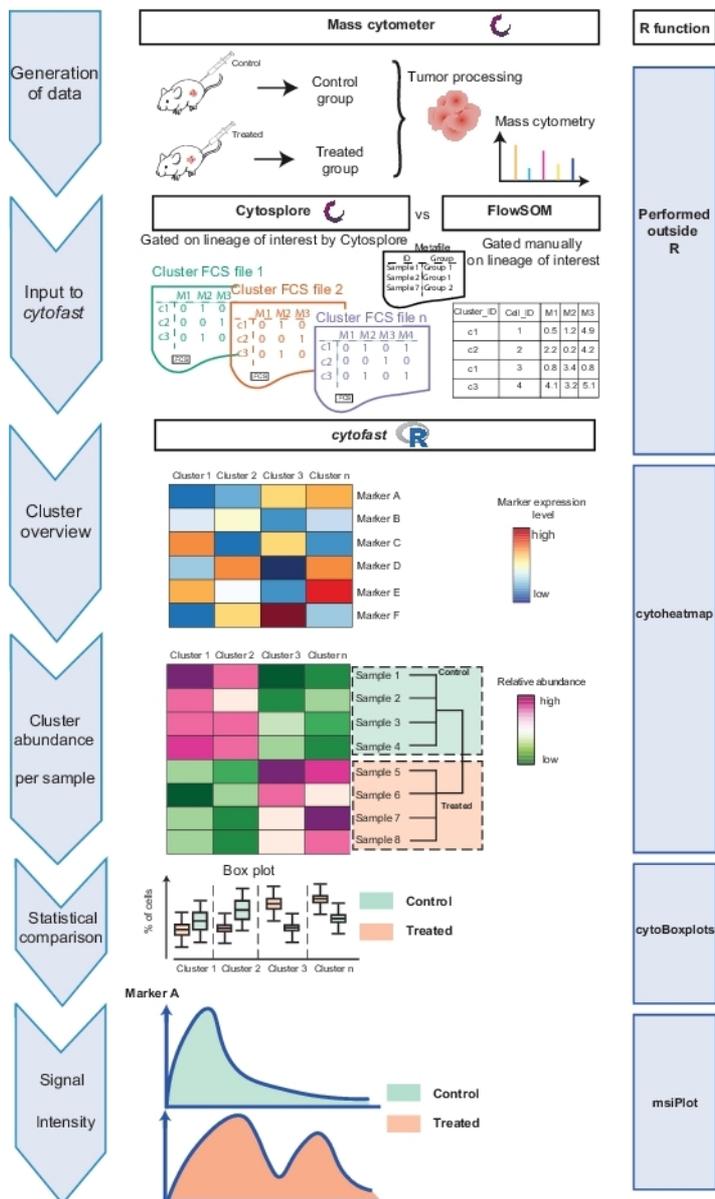
**NOTE:** Here, focus on CD45, CD11c, and CD54. Check the exact spelling of the markers and adjust if needed:  
`msiPlot(cfData, markers = c("89Y_CD45", "167Er_CD11c", "164Dy_CD54"), byGroup='group')`

## Representative Results

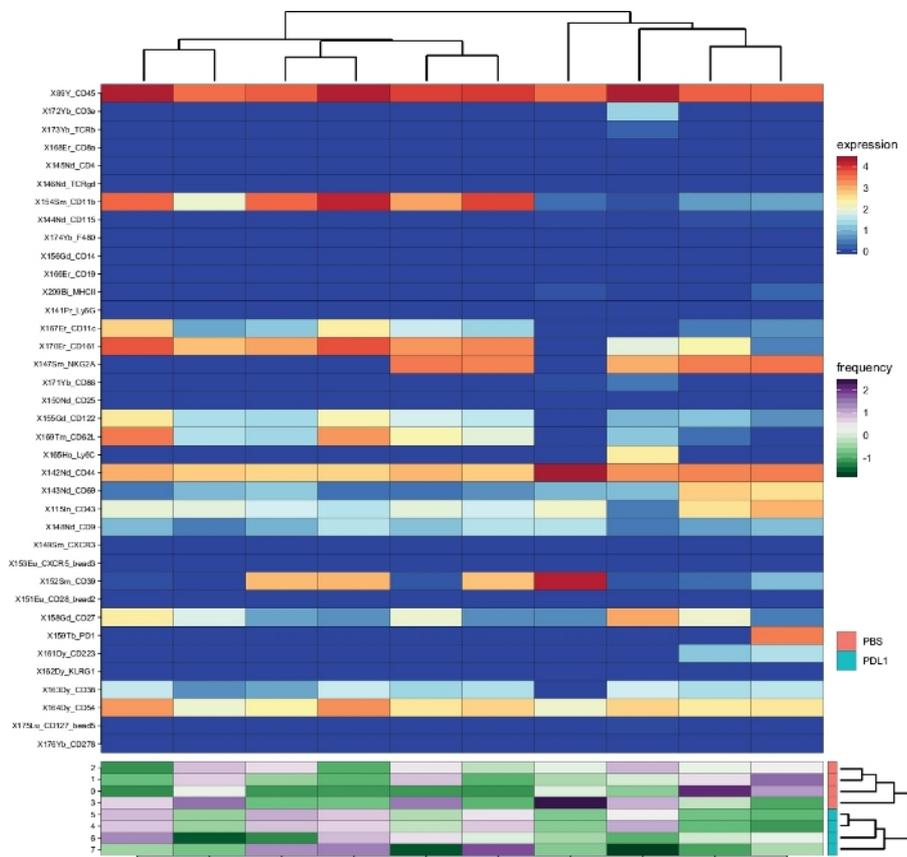
The workflow *Cytofast* (**Figure 1**) is meant to provide a quantitative and qualitative overview of the data originally clustered by analysis software (i.e., *FlowSOM* or *Cytosplore*). *Cytofast* runs several possible outputs, including the heatmap of all clusters identified in the analysis and based on marker expression (**Figure 2** and **Figure 3**). The dendrogram on the top represents the hierarchical similarity between the identified clusters. The upper panel displays another heatmap showing the relative quantity of corresponding subsets in each sample. The dendrogram on the right shows the similarity between samples and is based on hierarchical clustering performed on the Euclidean distances between samples. The combined heatmaps are shown for *FlowSOM* followed by *Cytofast* in **Figure 2** and for *Cytosplore* followed by *Cytofast* in **Figure 3**. *Cytofast* can also be used to present the data quantitatively and display the results in boxplots (by using `cytoBoxplots` function), as shown in **Figure 4** and **Figure 5**.

Similar clusters were found between the two different methods (e.g., cluster 8 from *Cytosplore* corresponds to cluster 10 from *FlowSOM*), and co-expression of some inhibitory markers like PD-1 and LAG-3 were still visible in both methods). Both clustering methods allowed discrimination between PD-L1 vs. PBS treated mice. In contrast, some differences between both methods can be highlighted. *FlowSOM* identifies 2 clusters (MHC-II<sup>+</sup>), whereas *Cytosplore* shows only one cluster (MHC-II<sup>+dim</sup>). This is due to the initial gating strategy in which NK cells were manually gated on CD161<sup>+</sup> cells, then further processed by *FlowSOM*. However, *Cytosplore* automatically gated cells from the CD45<sup>+</sup> population on the first HSNE level, which were then clustered in a higher hierarchical level. Thus, *Cytosplore* defined the NK cell subsets more precisely than how manual gating focused on CD161. Nevertheless, hierarchical clustering of the samples was preserved, as shown in the dendrogram on the right, indicating that segregation between the two groups (PD-L1 and PBS) was not dependent on the chosen clustering method.

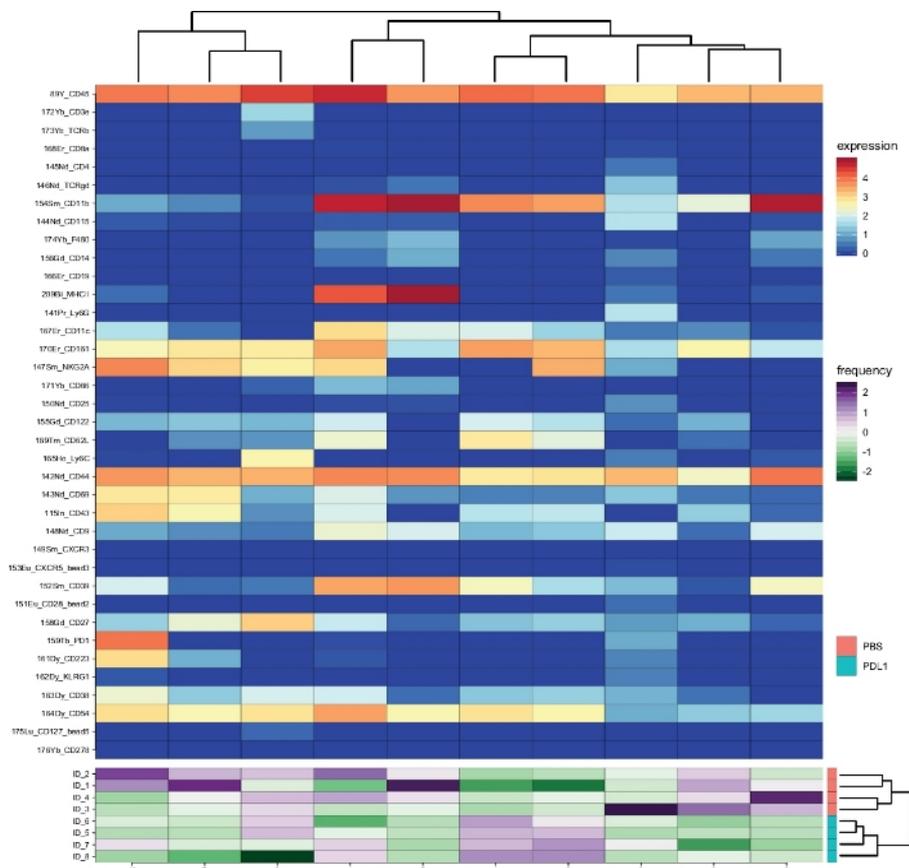
The number of clusters can be manually defined using both methods. *Cytofast* enables the user to assess the heterogeneity of their data and can provide insight into how to choose the number of clusters into which the data should be divided. Other features are included in the *Cytofast* package, such as the `msiPlot` function (step 3.4.2), showing the median signal Intensity (MSI) plot of every marker per group (**Figure 6** and **Figure 7**). This function allows detection of global changes, such as increases in the expression of CD54 or CD11c in NK cells of the PD-L1-treated group. Optional features can be incorporated in the *Cytofast* package, such as displaying data in bar graphs and other methods of data representation. The latter requires the addition of `ggplot` tools, which can be generated by R.



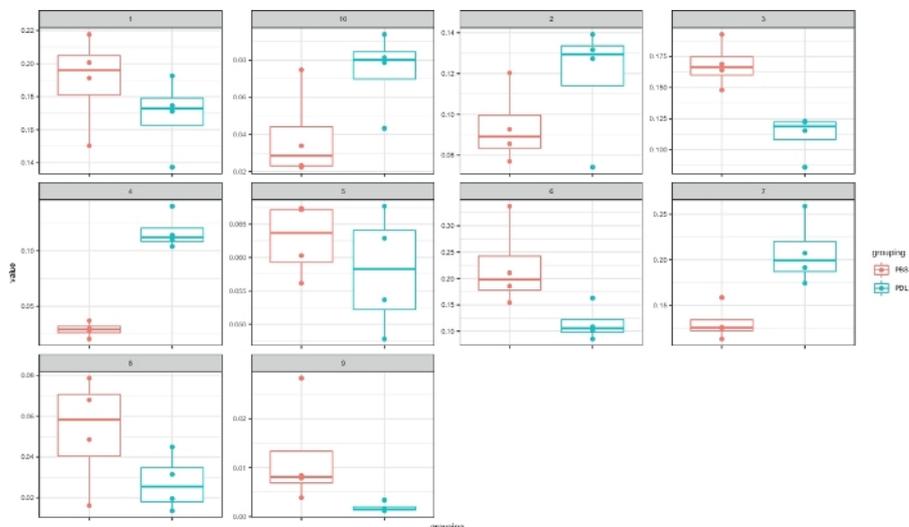
**Figure 1: Workflow of Cytofast package.** The data were generated by mass cytometry from a tumor 3 days after treatment with immunotherapy or left untreated. Two different clustering techniques were compared: *Cytosplore* and *FlowSOM*. *Cytofast* was used to visualize differences between the two techniques. [Please click here to view a larger version of this figure.](#)



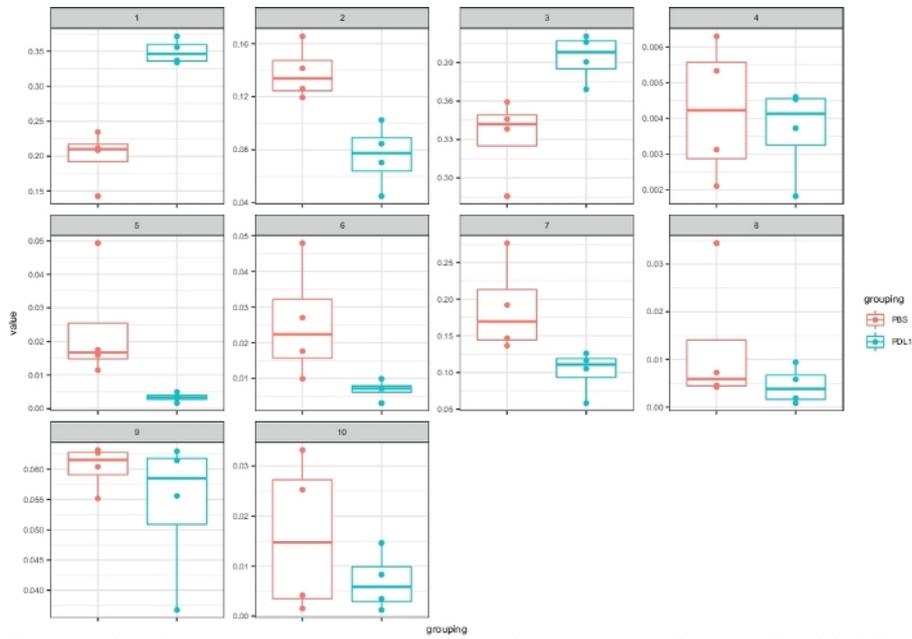
**Figure 2: Cluster overview and cluster abundance per group as analyzed by Cytofast following Cytosplore.** Heatmap of all NK cell clusters (CD161<sup>+</sup> cells defined automatically by Cytosplore), which were identified 3 days after immunotherapy (PD-L1). Data shown is based on Cytosplore clustering and pooled from the untreated and PD-L1 treated groups. Levels of ArcSinh5-transformed expression marker are displayed on a rainbow scale. On the lower panel, the relative abundance of each sample is represented by the green-to-purple scale. The dendrogram on the right represents the similarity between samples based on subset frequencies. The frequency scale represents the dispersion of the mean. A low or a high frequency is represented by a green or purple color, respectively. [Please click here to view a larger version of this figure.](#)



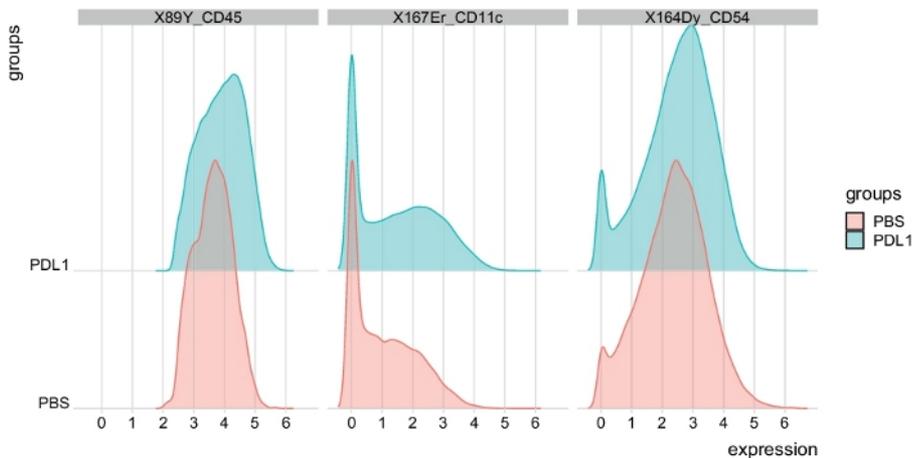
**Figure 3: Cluster overview and cluster abundance per group as analyzed by Cytofast following FlowSOM.** Heatmap of all NK cell clusters (pre-gated on CD161<sup>+</sup> events), which were identified 3 days after immunotherapy (PD-L1). Data shown is based on FlowSOM clustering and pooled from the untreated and PD-L1 treated groups. Levels of ArcSinh5-transformed expression marker are displayed on a rainbow scale. On the lower panel, the relative abundance of each sample is represented by the green-to-purple scale. The dendrogram on the right represents the similarity between samples based on subset frequencies. The frequency scale represents the dispersion of the mean. A low or a high frequency is represented by a green or purple color, respectively. [Please click here to view a larger version of this figure.](#)



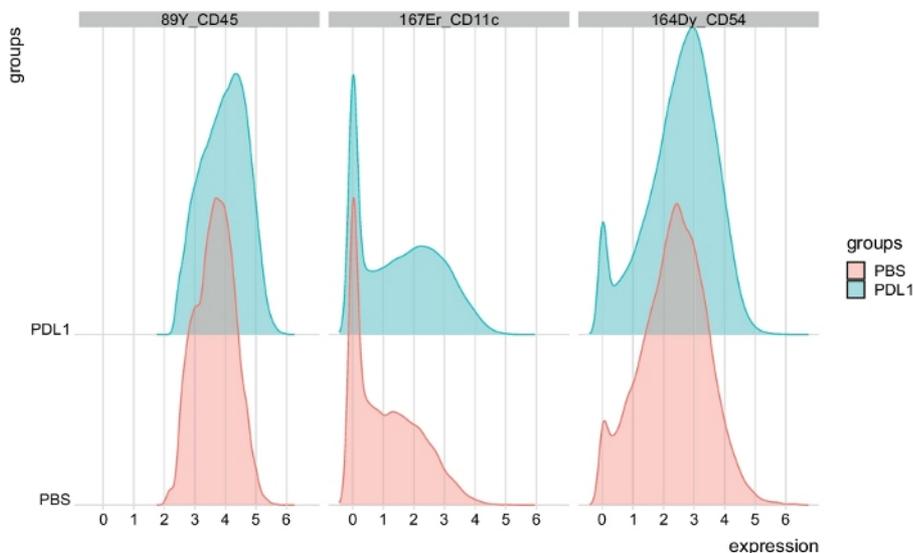
**Figure 4: Cytofast representation with boxplots of the clusters defined by Cytosplore.** The frequency of each cluster is represented in a boxplot, separated into the two groups (PBS and PD-L1). One individual dot corresponds to one mouse. [Please click here to view a larger version of this figure.](#)



**Figure 5: Cytofast representation with boxplots of the clusters defined by FlowSOM.** The frequency of each cluster is represented in a boxplot, separated into the two groups (PBS and PD-L1). One individual dot corresponds to one mouse. [Please click here to view a larger version of this figure.](#)



**Figure 6: Distribution of signal intensity plots from NK cells automatically gated by Cytosplore.** Distribution of signal intensities are shown in a histogram for three specific markers: CD45, CD11c, and CD54. [Please click here to view a larger version of this figure.](#)



**Figure 7: Distribution of signal intensity plots from NK cells automatically gated by *FlowSOM*.** Distribution of signal intensities are shown in a histogram for three specific markers: CD45, CD11c, and CD54, segregated by groups PBS and PD-L1. [Please click here to view a larger version of this figure.](#)

**Supplementary Files 1.1–1.8.** [Please click here to view this file \(Right click to download\).](#)

**Supplementary Files 2.1–2.10.** [Please click here to view this file \(Right click to download\).](#)

**Supplementary File 3.** [Please click here to view this file \(Right click to download\).](#)

**Supplementary Files 4.1–4.8.** [Please click here to view this file \(Right click to download\).](#)

**Supplementary File 5.** [Please click here to view this file \(Right click to download\).](#)

## Discussion

*Cytofast* is a rapid computational tool that provides a quick and global exploration of cytometric data by highlighting and quantifying treatment-specific cellular subsets. The protocol described aims to further process clustering analyses with *Cytosplore* or *FlowSOM*. Other clustering analysis tools are suitable for *Cytofast*, but this requires the use of *Cytofast* to assign each cell to a subset. *Cytofast*, however, is not a clustering method, and therefore requires clustering procedures before use.

The analysis performed here showed that certain CD161<sup>+</sup> NK cell subsets in the tumor microenvironment were sensitive to a PD-L1 blockade. This was evidenced by changes in their phenotype and abundance, which were observed using both *Cytosplore* and *FlowSOM* as clustering methods. Both methods distinguished the main NK cell cluster (CD11b<sup>+</sup> NKG2A<sup>+</sup>) with slightly different frequencies (15%–20% for *Cytosplore*, 30%–40% for *FlowSOM*). The differences in abundance and this approximation did not affect the global pattern, because both dendrograms displayed in the right panels of **Figure 2** and **Figure 3** showed similar results. By using *Cytofast*, it is thus possible (independent from the clustering method chosen) to segregate PD-L1-treated and untreated mice based on analyses of NK cell cluster phenotype and abundance.

Depending on the recorded parameters, modifications to the protocol are needed. Specifically, certain parameters such as time and background must be removed while performing the clustering analysis. In addition, it is important that each cell is assigned to a subset. The *cfData* function will simply add the raw cell counts per cluster per sample into the *cfList*. From this step, the cytoheatmap can be built as explained in section 3.

*Cytofast* has been successfully used as a visualization and quantification tool to compare different clustering methods<sup>13</sup>. This R package is also compatible with advanced features, such as the *globaltest*<sup>14</sup>, which can test associations between groups of clusters using clinical variables. In the future, the *globaltest* tool and other algorithms can be integrated with *Cytofast* for more in-depth visualization and quantification.

## Disclosures

The authors have nothing to disclose.

## Acknowledgments

We acknowledge funding from the European Commission of a H2020 MSCA award under proposal number 675743 (ISPIC). We thank Tetje van der Sluis and Iris Pardieck for testing the protocol.

## References

1. Anchang, B. et al. Visualization and cellular hierarchy inference of single-cell data using SPADE. *Nature Protocols*. **11** (7), 1264-1279 (2016).
2. Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M., Nolan, G. P. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell*. **16** (3), 323-337 (2015).
3. Van Gassen, S. et al. *FlowSOM*: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A*. **87** (7), 636-645 (2015).
4. Levine, J. H. et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*. **162** (1), 184-197 (2015).
5. Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L., Nolan, G. P. Automated mapping of phenotype space with single-cell data. *Nature Methods*. **13** (6), 493-496 (2016).
6. Spitzer, M. H. et al. IMMUNOLOGY. An interactive reference framework for modeling a dynamic immune system. *Science*. **349** (6244), 1259425, (2015).
7. Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. **24** (6), 417-441 (1933).
8. van der Maaten, L., Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research*. (2008).
9. Pezzotti, N., Holtt, T., Lelieveldt, B., Eisemann, E., Vilanova, A. Hierarchical Stochastic Neighbor Embedding. *Computer Graphics Forum*. **35** (3), 21-30 (2016).
10. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*. **37**, 38 (2018).
11. Haghverdi, L., Buettner, F., Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*. **31** (18), 2989-2998 (2015).
12. Beyrend, G., Stam, K., Höllt, T., Ossendorp, F., Arens, R. *Cytofast*: A workflow for visual and quantitative analysis of flow and mass cytometry data to discover immune signatures and correlations. *Computational and Structural Biotechnology Journal*. **16** 435-442 (2018).
13. Beyrend, G. et al. PD-L1 blockade engages tumor-infiltrating lymphocytes to co-express targetable activating and inhibitory receptors. *Journal for ImmunoTherapy of Cancer*. **7** (1), 217 (2019).
14. Goeman, J. J., van de Geer, S. A., de Kort, F., van Houwelingen, H. C. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*. **20** (1), 93-99 (2004).