

# Package ‘HDStIM’

October 12, 2022

**Type** Package

**Title** High Dimensional Stimulation Immune Mapping (‘HDStIM’)

**Version** 0.1.0

**Description** A method for identifying responses to experimental stimulation in mass or flow cytometry that uses high dimensional analysis of measured parameters and can be performed with an end-to-end unsupervised approach. In the context of in vitro stimulation assays where high-parameter cytometry was used to monitor intracellular response markers, using cell populations annotated either through automated clustering or manual gating for a combined set of stimulated and unstimulated samples, ‘HDStIM’ labels cells as responding or non-responding. The package also provides auxiliary functions to rank intracellular markers based on their contribution to identifying responses and generating diagnostic plots.

**License** CC0 | file LICENSE

**Encoding** UTF-8

**LazyData** true

**LazyDataCompression** xz

**RoxygenNote** 7.1.2

**URL** <https://github.com/niaid/HDStIM>, <https://niaid.github.io/HDStIM/>

**BugReports** <https://github.com/niaid/HDStIM/issues>

**Depends** R (>= 3.6.0)

**Imports** tibble, ggplot2, uwot, dplyr, tidyr, broom, tidyselect, ggridges, Boruta, scales

**Suggests** knitr, rmarkdown, testthat

**VignetteBuilder** knitr, rmarkdown

**Language** en-US

**NeedsCompilation** no

**Author** Rohit Farmer [aut, cre] (<<https://orcid.org/0000-0003-4197-3047>>),  
Richard Apps [aut] (<<https://orcid.org/0000-0001-5140-0141>>),  
John Tsang [aut, pdr] (<<https://orcid.org/0000-0003-3186-3047>>)

**Maintainer** Rohit Farmer <[rohit.farmer@gmail.com](mailto:rohit.farmer@gmail.com)>

**Repository** CRAN

**Date/Publication** 2022-06-24 08:10:12 UTC

## R topics documented:

chi11 . . . . .	2
HDStIM . . . . .	3
marker_ranking_boruta . . . . .	4
plot_exprs . . . . .	5
plot_K_Fisher . . . . .	6
plot_umap . . . . .	6
<b>Index</b>	<b>8</b>

---

chi11

*Sample data set for CyTOF Stimulation Assay*

---

### Description

A list with the CyTOF stimulation assay data.

### Usage

chi11

### Format

A list with one tibble containig CyTOF expression data. And four character vectors for arguments in the [HDStIM](#) function.

**chi11\$expr\_data** A 7,000 X 36 tibble. Cells are on the rows and variables on the columns. The first 6 columns contain for each cell `cluster_id` (from FlowSOM clustering), `sample_id` (unique for each FSC file), `condition` (comparison groups), `patient_id` (unique for each subject), `stim_type` (labels for types of stimulation assays including the unstim), `merging1` (meta culster labels from ConsensusClusterPlus). The last 30 columns contain the archsinh transformed CyTOF expression values for the 30 markers (20 type and 10 state) used in the sitmulation panel.

**chi11\$type\_markers** A character vector with the labels for type markers used in the stimulation panel.

**chi11\$state\_markers** A character vector with the labels for state markers used in the stimulation panel.

**chi11\$cluster\_col** A character label of the meta-cluster/cluster ID column in `chi11$expr_data` tibble.

**chi11\$stim\_label** A character vector with the label(s) for the stimulation types corresponding to the labels in the `stim_type` column in `chi11$expr_data`.

**chi11\$unstim\_label** A character label for the unstim cells corresponding to the labels in the `stim_type` column in `chi11$expr_data`.

**Description**

Function to select cells from the stimulated samples that have likely responded to the stimulant.

**Usage**

```
HDStIM(
  dat,
  state_markers,
  cellpop_col,
  stim_lab,
  unstim_lab,
  seed_val = NULL,
  umap = FALSE,
  umap_cells = NULL,
  verbose = FALSE
)
```

**Arguments**

<code>dat</code>	A tibble with the single cell data. Cells on rows and variables/markers on columns.
<code>state_markers</code>	A character vector with the labels of state markers from the stimulation panel.
<code>cellpop_col</code>	Column in the tibble with the cell population IDs.
<code>stim_lab</code>	A character vector of stim label(s).
<code>unstim_lab</code>	A character of unstim label(s).
<code>seed_val</code>	Seed value (integer) for <code>kmeans</code> clustering. Default is <code>NULL</code> for no seed value.
<code>umap</code>	Boolean (T/F) to carry out UMAP on the selected cells. Default is <code>FALSE</code> to skip UMAP calculation.
<code>umap_cells</code>	An integer; for calculating UMAPs take a minimum of <code>umap_cells</code> per cluster or the total number of cells if the cluster size is smaller than <code>umap_cells</code> . Default is <code>NULL</code> .
<code>verbose</code>	Logical. To make function more verbose. Default is <code>FALSE</code> .

**Value**

A list with tibbles for expression data for the selected cells, data to plot stacked bar plots, data to plot UMAP plots, and parameters passed to the function.

## Examples

```
mapped_data <- HDStIM(chi11$expr_data, chi11$state_markers,  
  chi11$cluster_col, chi11$stim_label,  
  chi11$unstim_label, seed_val = 123, umap = FALSE, umap_cells = NULL,  
  verbose = FALSE)
```

---

marker\_ranking\_boruta *Marker Ranking by Boruta*

---

## Description

Function to run Boruta on the stimulation - cell population combinations that passed the Fisher's exact test to rank the markers according to their contribution to the response.

## Usage

```
marker_ranking_boruta(  
  mapped_data,  
  path = NULL,  
  n_cells = NULL,  
  max_runs = 100,  
  seed_val = 123,  
  verbose = 0  
)
```

## Arguments

mapped_data	Returned list from the <a href="#">HDStIM</a> function.
path	Path to the folder to save figures generated by this function.
n_cells	Number of cells to down sample the data. Default is NULL to include all the cells.
max_runs	Maximum number of runs for the random forest algorithm. Default is 100.
seed_val	Seed value for Boruta. Default is 123.
verbose	0, 1, or 2. Default is 0.

## Value

A list with a tibble containing attribute statistics calculated by Boruta and ggplot objects. If the path is not NULL, plots are also rendered and saved in the specified folder in PNG format.

## Examples

```
mapped_data <- HDStIM(chi11$expr_data, chi11$state_markers,
  chi11$cluster_col, chi11$stim_label,
  chi11$unstim_label, seed_val = 123, umap = FALSE, umap_cells = NULL,
  verbose = FALSE)

attribute_stats <- marker_ranking_boruta(mapped_data, path = NULL, n_cells = NULL,
  max_runs = 1000, seed_val = 123,
  verbose = 0)
```

---

plot_exprs	<i>Diagnostic plots showing individual marker distribution before and after mapping by HDStIM</i>
------------	---

---

## Description

Diagnostic plots showing individual marker distribution before and after mapping by HDStIM

## Usage

```
plot_exprs(mapped_data, path = NULL, verbose = FALSE)
```

## Arguments

mapped_data	List output of the <a href="#">HDStIM</a> function.
path	Path to the folder to save figures generated by this function.
verbose	Logical. To make function more verbose. Default is FALSE.

## Value

A list of ggplot objects. If the path is not NULL, PNG files of the plots are saved in the specified folder.

## Examples

```
mapped_data <- HDStIM(chi11$expr_data, chi11$state_markers,
  chi11$cluster_col, chi11$stim_label,
  chi11$unstim_label, seed_val = 123, umap = FALSE, umap_cells = NULL,
  verbose = FALSE)

pe <- plot_exprs(mapped_data, path = NULL, verbose = FALSE)
```

---

plot_K_Fisher	<i>Diagnostic plots explaining K-means clustering and Fisher's exact test carried out by HDStIM</i>
---------------	---

---

**Description**

Diagnostic plots explaining K-means clustering and Fisher's exact test carried out by HDStIM

**Usage**

```
plot_K_Fisher(mapped_data, path = NULL, verbose = FALSE)
```

**Arguments**

mapped_data	Returned list from the <a href="#">HDStIM</a> function.
path	Path to the folder to save figures generated by this function NULL by default.
verbose	Logical. To make function more verbose. Default is FALSE.

**Value**

A list of ggplot objects. If the path is not NULL, PNG files of the plots are saved in the specified folder.

**Examples**

```
mapped_data <- HDStIM(chi11$expr_data, chi11$state_markers,
  chi11$cluster_col, chi11$stim_label,
  chi11$unstim_label, seed_val = 123, umap = FALSE, umap_cells = NULL,
  verbose = FALSE)

pk <- plot_K_Fisher(mapped_data, path = NULL, verbose = FALSE)
```

---

plot_umap	<i>Diagnostic UMAP plots showing the partitioning of cells into responding and non-responding groups by HDStIM</i>
-----------	--

---

**Description**

Diagnostic UMAP plots showing the partitioning of cells into responding and non-responding groups by HDStIM

**Usage**

```
plot_umap(mapped_data, path = NULL, verbose = FALSE)
```

**Arguments**

<code>mapped_data</code>	Returned list from the <code>HStIM</code> function.
<code>path</code>	Path to the folder to save figures generated by this function.
<code>verbose</code>	Logical. To make function more verbose. Default is <code>FALSE</code> .

**Value**

A list of ggplot objects. If the path is not `NULL`, PNG files of the plots are saved in the specified folder.

**Examples**

```
mapped_data <- HStIM(chi11$expr_data, chi11$state_markers,  
                    chi11$cluster_col, chi11$stim_label,  
                    chi11$unstim_label, seed_val = 123, umap = TRUE,  
                    umap_cells = 50, verbose = FALSE)  
  
pu <- plot_umap(mapped_data, path = NULL, verbose = FALSE)
```

# Index

## \* datasets

chi11, 2

chi11, 2

HDStIM, 2, 3, 4–7

kmeans, 3

marker\_ranking\_boruta, 4

plot\_exprs, 5

plot\_K\_Fisher, 6

plot\_umap, 6