

# Package ‘clustifyrdatahub’

October 24, 2025

**Title** External data sets for clustifyr in ExperimentHub

**Version** 1.19.0

**Description** References made from external single-cell mRNA sequencing data sets, stored as average gene expression matrices. For use with clustifyr <<https://bioconductor.org/packages/clustifyr>> to assign cell type identities.

**License** MIT + file LICENSE

**URL** <https://rnabioco.github.io/clustifyrdatahub/>

**Depends** R (>= 4.0), ExperimentHub

**Imports** utils

**Suggests** clustifyr, Seurat, usethis, rmarkdown, knitr, tidyverse, BiocStyle

**VignetteBuilder** knitr

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.1

**biocViews** SingleCellData, SequencingData, MicroarrayData, ExperimentHub, RNASeqData, PackageTypeData, ExpressionData

**BugReports** <https://github.com/rnabioco/clustifyrdatahub/issues>

**Config/Needs/website** pkgdown, rnabioco/rbitemplate

**git\_url** <https://git.bioconductor.org/packages/clustifyrdatahub>

**git\_branch** devel

**git\_last\_commit** ed76112

**git\_last\_commit\_date** 2025-04-15

**Repository** Bioconductor 3.23

**Date/Publication** 2025-10-24

**Author** Rui Fu [aut] (ORCID: <<https://orcid.org/0000-0001-8183-4549>>),  
Kent Riemondy [aut, cre] (ORCID:  
<<https://orcid.org/0000-0003-0750-1273>>),  
RNA Bioscience Initiative [fnd],

Austin Gillen [ctb] (ORCID: <<https://orcid.org/0000-0003-2928-6308>>),  
 Jay Hesselberth [ctb] (ORCID: <<https://orcid.org/0000-0002-6299-179X>>),  
 Sidhant Puntambekar [ctb]

**Maintainer** Kent Riemondy <kent.riemondy@cuanschutz.edu>

## Contents

|                                 |   |
|---------------------------------|---|
| clustifyrdatahub . . . . .      | 2 |
| ref_cortex_dev . . . . .        | 3 |
| ref_hema_microarray . . . . .   | 3 |
| ref_immgen . . . . .            | 4 |
| ref_MCA . . . . .               | 5 |
| ref_moca_main . . . . .         | 5 |
| ref_mouse.rnaseq . . . . .      | 6 |
| ref_mouse_atlas . . . . .       | 7 |
| ref_pan_indrop . . . . .        | 7 |
| ref_pan_smartseq2 . . . . .     | 8 |
| ref_tabula_muris_drop . . . . . | 9 |
| ref_tabula_muris_facs . . . . . | 9 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>11</b> |
|--------------|-----------|

---

|                  |                                                                 |
|------------------|-----------------------------------------------------------------|
| clustifyrdatahub | <i>clustifyrdatahub: reference scRNA-seq data for clustifyr</i> |
|------------------|-----------------------------------------------------------------|

---

## Description

The clustifyrdatahub package provides access to the summarized reference count matrix data from multiple datasets used in the clustifyr manuscript and tutorial. They are provided for future users of clustifyr, for automated cell type classification of scRNA-seq experiments.

## Details

Available datasets:

+ Mouse Cell Atlas - [ref\_MCA()] + Tabula Muris (10X) - [ref\_tabula\_muris\_drop()] + Tabula Muris (SmartSeq2) - [ref\_tabula\_muris\_facs()] + Mouse RNA-seq from 28 cell types - [ref\_mouse.rnaseq()] + Mouse Organogenesis Cell Atlas (main cell types) - [ref\_moca\_main()] + Mouse sorted immune cells - [ref\_immgen()] + Human hematopoietic cell microarray - [ref\_hema\_microarray()] + Human cortex development scRNA-seq - [ref\_cortex\_dev()] + Human pancreatic cell scRNA-seq (inDrop) - [ref\_pan\_indrop()] + Human pancreatic cell scRNA-seq (SmartSeq2) - [ref\_pan\_smartseq2()]

---

|                |                                                                          |
|----------------|--------------------------------------------------------------------------|
| ref_cortex_dev | <i>Matrix of average gene detection on cortex development cell types</i> |
|----------------|--------------------------------------------------------------------------|

---

**Description**

47 distinct annotated clusters of human neuronal subtypes

**Usage**

```
ref_cortex_dev
```

**Format**

An object of class function of length 1.

**Source**

<<https://cells.ucsc.edu/?ds=cortex-dev#>>

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_cortex_dev(metadata = TRUE)
```

---

|                     |                                                                                         |
|---------------------|-----------------------------------------------------------------------------------------|
| ref_hema_microarray | <i>Matrix of average gene detection by microarray on sorted immune cell populations</i> |
|---------------------|-----------------------------------------------------------------------------------------|

---

**Description**

38 distinct purified populations of human hematopoietic cells

**Usage**

```
ref_hema_microarray
```

**Format**

An object of class function of length 1.

**Source**

[<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24759>]

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_hema_microarray(metadata = TRUE)
```

---

|            |                                                           |
|------------|-----------------------------------------------------------|
| ref_immgen | <i>Matrix of average gene expression from immgen data</i> |
|------------|-----------------------------------------------------------|

---

**Description**

253 mouse immune cell types

**Usage**

```
ref_immgen
```

**Format**

An object of class function of length 1.

**Source**

["SingleR"]

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_immgen(metadata = TRUE)
```

---

|         |                                                                   |
|---------|-------------------------------------------------------------------|
| ref_MCA | <i>Matrix of average gene expression for single-cell RNA-seq.</i> |
|---------|-------------------------------------------------------------------|

---

**Description**

from mouse cell atlas project

**Usage**

ref\_MCA

**Format**

An object of class function of length 1.

**Source**

<<https://figshare.com/s/865e694ad06d5857db4b>>

**See Also**

Other ref: [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_MCA(metadata = TRUE)
```

---

|               |                                                                                                                         |
|---------------|-------------------------------------------------------------------------------------------------------------------------|
| ref_moca_main | <i>Matrix of average gene expression per cluster from single-cell sci-RNA-seq3 from Mouse Organogenesis Cell Atlas.</i> |
|---------------|-------------------------------------------------------------------------------------------------------------------------|

---

**Description**

100,000 cells

**Usage**

ref\_moca\_main

**Format**

An object of class function of length 1.

**Source**

<<http://atlas.gs.washington.edu/mouse-rna/>>

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_moca_main(metadata = TRUE)
```

---

|                  |                                                                            |
|------------------|----------------------------------------------------------------------------|
| ref_mouse.rnaseq | <i>Matrix of average gene expression from general mouse cell type data</i> |
|------------------|----------------------------------------------------------------------------|

---

**Description**

28 general cell types

**Usage**

```
ref_mouse.rnaseq
```

**Format**

An object of class function of length 1.

**Source**

["SingleR"]

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_mouse.rnaseq(metadata = TRUE)
```

---

|                 |                                                                                                  |
|-----------------|--------------------------------------------------------------------------------------------------|
| ref_mouse_atlas | <i>Atlas matrix of average gene expression aggregated from scRNA-seq data of mouse organisms</i> |
|-----------------|--------------------------------------------------------------------------------------------------|

---

**Description**

Data aggregated from NCBI Gene Expression Omnibus

**Usage**

```
ref_mouse_atlas
```

**Format**

An object of class function of length 1.

**Source**

<<https://github.com/rnabioco/scRNA-seq-Cell-Ref-Matrix/blob/master/atlas/musMusculus/MouseAtlas.rds>>

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_mouse_atlas(metadata = TRUE)
```

---

|                |                                                                                                                  |
|----------------|------------------------------------------------------------------------------------------------------------------|
| ref_pan_indrop | <i>Matrix of average gene expression per cluster from single-cell RNA-seq pancreas data using indrop method.</i> |
|----------------|------------------------------------------------------------------------------------------------------------------|

---

**Description**

A sample of 8569 pancreas cells

**Usage**

```
ref_pan_indrop
```

**Format**

An object of class function of length 1.

**Source**

<<https://hemberg-lab.github.io/scRNA.seq.datasets/human/pancreas/>>

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_pan_indrop(metadata = TRUE)
```

---

|                   |                                                                                                                     |
|-------------------|---------------------------------------------------------------------------------------------------------------------|
| ref_pan_smartseq2 | <i>Matrix of average gene expression per cluster from single-cell RNA-seq pancreas data using smartseq2 method.</i> |
|-------------------|---------------------------------------------------------------------------------------------------------------------|

---

**Description**

A sample of 2209 pancreas cells

**Usage**

```
ref_pan_smartseq2
```

**Format**

An object of class function of length 1.

**Source**

<<https://hemberg-lab.github.io/scRNA.seq.datasets/human/pancreas/>>

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_pan_smartseq2(metadata = TRUE)
```



---

ref\_tabula\_muris\_drop *Matrix of average gene expression per cluster from single-cell RNA-seq from Tabula Muris (10x genomics).*

---

**Description**

422,803 droplet libraries, 55,656 of which passed a QC cutoff of 500 genes and 1000 UMI

**Usage**

```
ref_tabula_muris_drop
```

**Format**

An object of class function of length 1.

**Source**

<<https://tabula-muris.ds.czbiohub.org/>>

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#), [ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_facs\(\)](#)

**Examples**

```
ref_tabula_muris_drop(metadata = TRUE)
```

---

ref\_tabula\_muris\_facs *Matrix of average gene expression per cluster from single-cell RNA-seq from Tabula Muris (SMART-Seq2).*

---

**Description**

53,760 cells from 20 tissues from 8 mice

**Usage**

```
ref_tabula_muris_facs
```

**Format**

An object of class function of length 1.

**Source**

<<https://tabula-muris.ds.czbiohub.org/>>

**See Also**

Other ref: [ref\\_MCA\(\)](#), [ref\\_cortex\\_dev\(\)](#), [ref\\_hema\\_microarray\(\)](#), [ref\\_immgen\(\)](#), [ref\\_moca\\_main\(\)](#),  
[ref\\_mouse.rnaseq\(\)](#), [ref\\_mouse\\_atlas\(\)](#), [ref\\_pan\\_indrop\(\)](#), [ref\\_pan\\_smartseq2\(\)](#), [ref\\_tabula\\_muris\\_drop\(\)](#)

**Examples**

```
ref_tabula_muris_facs(metadata = TRUE)
```

# Index

## \* datasets

- ref\_cortex\_dev, [3](#)
- ref\_hema\_microarray, [3](#)
- ref\_immgen, [4](#)
- ref\_MCA, [5](#)
- ref\_moca\_main, [5](#)
- ref\_mouse.rnaseq, [6](#)
- ref\_mouse\_atlas, [7](#)
- ref\_pan\_indrop, [7](#)
- ref\_pan\_smartseq2, [8](#)
- ref\_tabula\_muris\_drop, [9](#)
- ref\_tabula\_muris\_facs, [9](#)

## \* internal

- clustifyrdatahub, [2](#)

## \* ref

- ref\_cortex\_dev, [3](#)
- ref\_hema\_microarray, [3](#)
- ref\_immgen, [4](#)
- ref\_MCA, [5](#)
- ref\_moca\_main, [5](#)
- ref\_mouse.rnaseq, [6](#)
- ref\_mouse\_atlas, [7](#)
- ref\_pan\_indrop, [7](#)
- ref\_pan\_smartseq2, [8](#)
- ref\_tabula\_muris\_drop, [9](#)
- ref\_tabula\_muris\_facs, [9](#)

clustifyrdatahub, [2](#)

ref\_cortex\_dev, [3](#), [4–10](#)  
ref\_hema\_microarray, [3](#), [3](#), [4–10](#)  
ref\_immgen, [3](#), [4](#), [4](#), [5–10](#)  
ref\_MCA, [3](#), [4](#), [5](#), [6–10](#)  
ref\_moca\_main, [3–5](#), [5](#), [6–10](#)  
ref\_mouse.rnaseq, [3–6](#), [6](#), [7–10](#)  
ref\_mouse\_atlas, [3–6](#), [7](#), [8–10](#)  
ref\_pan\_indrop, [3–7](#), [7](#), [8–10](#)  
ref\_pan\_smartseq2, [3–8](#), [8](#), [9](#), [10](#)  
ref\_tabula\_muris\_drop, [3–8](#), [9](#), [10](#)  
ref\_tabula\_muris\_facs, [3–9](#), [9](#)