

# Package ‘HCA TonsilData’

April 2, 2026

**Title** Provide programmatic access to the tonsil cell atlas datasets

**Version** 1.8.0

**Description** This package provides access to the scRNA-seq, scATAC-seq, multiome, CITE-seq and spatial transcriptomics (Visium) data generated by the tonsil cell atlas in the context of the Human Cell Atlas (HCA). The data is provided via the Bioconductor project in the form of SingleCellExperiments. Additionally, information on the whole compendium of identified cell types is provided in form of a glossary.

**License** MIT + file LICENSE

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.2.3

**Imports** ExperimentHub, SingleCellExperiment, SpatialExperiment, HDF5Array, SummarizedExperiment, S4Vectors, htmltools, rmarkdown, base64enc, utils

**Suggests** knitr, ggplot2, testthat (>= 3.0.0), scater, Seurat, Signac, zellkonverter, iSEE, ggspavis, kableExtra, BiocStyle

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**URL** <https://github.com/massonix/HCA TonsilData>

**BugReports** <https://github.com/massonix/HCA TonsilData/issues>

**biocViews** SingleCellData, ExperimentData, RNASeqData, ExperimentHub, ExpressionData, SpatialData

**Depends** R (>= 4.3.0)

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

**git\_branch** RELEASE\_3\_22

**git\_last\_commit** a4a1e25

**git\_last\_commit\_date** 2025-10-29

**Repository** Bioconductor 3.22

**Date/Publication** 2026-04-02

**Author** Ramon Massoni-Badosa [aut, cre] (ORCID: <https://orcid.org/0000-0001-7115-8145>),  
 Federico Marini [aut] (ORCID: <https://orcid.org/0000-0003-3252-7758>),  
 Alan O'Callaghan [aut],  
 Helena L. Crowell [aut] (ORCID: <https://orcid.org/0000-0002-4801-1767>)

**Maintainer** Ramon Massoni-Badosa <ramonmassoni@gmail.com>

## Contents

annotations_dictionary . . . . .	2
colors_20230508 . . . . .	3
donor_metadata . . . . .	3
HCA_TonsilData . . . . .	3
listCellTypes . . . . .	4
NBC_MBC_annotation_df . . . . .	5
TonsilData_cellinfo . . . . .	5
TonsilData_cellinfo_html . . . . .	6
TonsilData_glossary . . . . .	7
updateAnnotation . . . . .	7
<b>Index</b>	<b>9</b>

---

annotations\_dictionary

*Annotation dictionary to keep track of annotations changes and versions*

---

## Description

Annotations are opinionated and dynamic by nature. They are subjected to change as more experts look into the data and more evidence is published. HCA\_TonsilData has been designed to account for that. We timestamp annotations and plan to include additional annotations if users propose them and we validate them.

## Format

A list of named vectors with the correspondence between time-stamped annotations

## Source

This file was created with the script in "inst/scripts/make-annotation-dictionary.R"

---

colors_20230508	<i>Color palettes</i>
-----------------	-----------------------

---

**Description**

Color palettes used in the paper for all cell types

**Format**

A list with the HEX color codes for each cell type

**Source**

This file was created with the script in "inst/scripts/make-annotation-colors.R"

---

donor_metadata	<i>Donor metadata</i>
----------------	-----------------------

---

**Description**

Data frame that contains all metadata information for the 17 donors included in the tonsil atlas.

**Format**

A dataframe with 17 observations and 8 variables

**Source**

Check the tonsil atlas publication

---

HCATonsilData	<i>Access the Tonsil Atlas data (RNA, ATAC, Multiome, CITE, Spatial)</i>
---------------	--

---

**Description**

The data was downloaded from Zenodo <https://zenodo.org/record/8373756>

**Usage**

```
HCATonsilData(  
  assayType = c("RNA", "ATAC", "CITE", "Spatial"),  
  cellType = "All",  
  version = "2.0",  
  processedCounts = TRUE  
)
```

**Arguments**

assayType	One of 'RNA', 'ATAC', 'Multiome', 'CITE' or 'Spatial'.
cellType	A character vector of length 1 with the desired cell type. A list of available cell types can be obtained using <code>listCellTypes(assay_type)</code> .
version	Version of the tonsil atlas data to retrieve: "1.0" (preprint) or "2.0" (publication, default)
processedCounts	Logical scalar. If TRUE, include the processed (normalized) counts in addition to the raw counts in the <code>SingleCellExperiment</code> object.

**Value**

A `SingleCellExperiment` object for the cellType requested. For scATAC-seq, Multiome and CITE we provide the instructions for downloading the Seurat objects in Zenodo (see vignette)

**Author(s)**

Ramon Massoni-Badosa

**Examples**

```
# retrieve the epithelial scRNA-seq dataset
sce_epithelial <- HCATonsilData(
  assayType = "RNA",
  cellType = "epithelial"
)
sce_epithelial
```

---

listCellTypes	<i>List available cell types for a dataset of the tonsil cell atlas</i>
---------------	---

---

**Description**

List available cell types for a dataset of the tonsil cell atlas

**Usage**

```
listCellTypes(
  assayType = c("RNA", "ATAC", "Multiome", "CITE", "Spatial"),
  version = "2.0"
)
```

**Arguments**

assayType	Either 'RNA', 'ATAC', 'Multiome', 'CITE', or 'Spatial'
version	Version of the tonsil atlas data: '1.0' (preprint) or '2.0' or '2.0' (publication, default). Note that, for version 2.0, CD8-T and ILC-NK are combined in a single 'Cytotoxic' object.

**Value**

A character vector with the available cell types for the indicated dataset.

**Examples**

```
listCellTypes(assayType = "RNA", version = "2.0")
```

---

NBC\_MBC\_annotation\_df *Annotation dictionary for naive and memory B cells (NBC-MBC)*

---

**Description**

In version 1 of the atlas, we changed the clusters and annotation for NBC-MBC prior to publishing the preprint. Since, there is not a 1:1 mapping between clusters, here we provide the correspondence between the annotations in February 2022 and the preprint (June 2022) for each NBC-MBC cell barcode.

**Format**

A dataframe with 112478 NBC-MBC and 5 variables

**barcode** cell barcode

**names\_level\_5\_clusters\_eta** Cell annotation given by the B-cell annotation team in February 2022

**annotation\_20220619** Cell annotation given by the B-cell annotation team June 2022 (preprint)

**UMAP\_1** UMAP1 coordinates

**UMAP\_2** UMAP2 coordinates

**Details**

This data is used by the updateAnnotation function in B cells.

**Source**

<https://zenodo.org/record/8373756>

---

TonsilData\_cellinfo *TonsilData\_cellinfo*

---

**Description**

Query the glossary to traceback the rationale for each annotation of the tonsil atlas

**Usage**

```
TonsilData_cellinfo(cellType = NULL)
```

**Arguments**

**cellType** String character, used to define the cell type for which the information will be displayed. Defaults to NULL - if left unspecified, the function returns a list of the possible options

**Value**

Invisible NULL - the information is displayed as a message in the console.

**Examples**

```
TonsilData_cellinfo()  
  
TonsilData_cellinfo("PDC")
```

---

```
TonsilData_cellinfo_html  
      TonsilData_cellinfo_html
```

---

**Description**

TonsilData\_cellinfo\_html

**Usage**

```
TonsilData_cellinfo_html(  
  cellType = NULL,  
  display_plot = TRUE,  
  output_to = c("single_page", "html_to_embed")  
)
```

**Arguments**

cellType	String character, used to define the cell type for which the information will be displayed. Defaults to NULL - if left unspecified, the function returns a list of the possible options
display_plot	Logical value, defines whether to include or not a plot for the UMAP of the selected cell type
output_to	Character value, one of "single_page" or "html_to_embed". Defines in which form the information is returned, either as an individual page or simply as HTML code to directly embed into other documents.

**Value**

Either the HTML code generated to be embedded, or a standalone HTML page is created - and the location to this temp file is returned as a character (default).

**Examples**

```
TonsilData_cellinfo_html("PDC")  
TonsilData_cellinfo_html("preB")  
TonsilData_cellinfo_html("preT")  
TonsilData_cellinfo_html("preT", output_to = "html_to_embed")
```

---

TonsilData\_glossary      *TonsilData\_glossary*

---

### Description

Convenience function to read directly in the file provided as extdata

### Usage

```
TonsilData_glossary()
```

### Value

Data frame containing the info on the cell types included in the TonsilDataAtlas

### Examples

```
glossary_df <- TonsilData_glossary()
head(glossary_df)
```

---

updateAnnotation      *Update cell type annotation*

---

### Description

Annotations are dynamic by nature. As more experts look into the data and newer literature comes out, we expect annotations to be refined over time. We have accommodated this by allowing us and users to add new annotations to the SingleCellExperiment objects. If you want to propose a new annotation based on your experience or new evidence, please open an issue at <https://github.com/massonix/HCATonsilData/issues>.

### Usage

```
updateAnnotation(sce, refAnnotation = "20220215", newAnnotation = "20220619")
```

### Arguments

sce                      A SingleCellExperiment object obtained using HCATonsilData function.  
refAnnotation      string specifying the date of the annotation to use as reference.  
newAnnotation      string specifying the suffix to add to the new column (annotation\_\*).

### Value

A SingleCellExperiment object with additional an additional column (annotation\_\*) that contains more annotations.

**Examples**

```
# update the annotation from preprint (version 1.0) to publication (version 2.0)
## Not run:
sce <- updateAnnotation(
  sce,
  refAnnotation = "20220619",
  newAnnotation = "20230508"
)

## End(Not run)
```

# Index

[annotations\\_dictionary](#), 2

[colors\\_20230508](#), 3

[donor\\_metadata](#), 3

[HCA\\_TonsilData](#), 3

[listCellTypes](#), 4

[NBC\\_MBC\\_annotation\\_df](#), 5

[SingleCellExperiment](#), 4

[TonsilData\\_cellinfo](#), 5

[TonsilData\\_cellinfo\\_html](#), 6

[TonsilData\\_glossary](#), 7

[updateAnnotation](#), 7