

Package: scmarkeragent (via r-universe)

September 24, 2026

Title Evidence-Grounded Cell-Type Annotation for Seurat Objects

Version 0.1.0

Description Annotates single-cell RNA-seq clusters by measuring curated marker panels in the data and letting a language model decide which identity those measurements support, with every claim resolving to a publication and its source sentence. Reads Seurat .rds objects natively and runs from raw counts, against one shared configuration, one set of prompts and one marker resource.

License file LICENSE

Encoding UTF-8

Depends R (>= 4.2)

Imports curl, data.table, digest, ggplot2, jsonlite, Matrix, presto, scales, Seurat (>= 5.0.0), tools, utils, zip

Suggests ggrastr, httpuv, later, png, promises

Additional_repositories <https://haoran-code.r-universe.dev>

URL <https://github.com/HaoRan-Code/scMarkerAgent>,
<https://haoran-code.r-universe.dev/scmarkeragent>,
<https://doi.org/10.5281/zenodo.22834079>

BugReports <https://github.com/HaoRan-Code/scMarkerAgent/issues>

Config/roxygen2/version 8.0.0

Config/pak/sysreqs cmake libglpk-dev make libicu-dev libpng-dev libuv1-dev libxml2-dev libssl-dev python3 zlib1g-dev

Repository <https://haoran-code.r-universe.dev>

Date/Publication 2026-09-22 20:50:16 UTC

RemoteUrl <https://github.com/HaoRan-Code/scMarkerAgent>

RemoteRef HEAD

RemoteSha 7c7a4f14c95cd43eb9cf16c9db5b6d6ff86d7ebe

RemoteSubdir R

Contents

annotate	2
check_input	3
download_resources	4
resource_index	4
verify_resources	5
Index	6

annotate	<i>Annotate a Seurat .rds dataset</i>
----------	---------------------------------------

Description

Runs the full scMarkerAgent pipeline on one Seurat object: QC, normalization, Leiden clustering (always recomputed), genome-wide differential expression, candidate retrieval from the curated marker resource, the annotating agent and the evidence review that checks it, and the delivered result package (tables, audit sidecar, figures, offline viewer and ZIP).

Usage

```
annotate(  
  input,  
  tag = "dataset",  
  species,  
  tissue,  
  counts_source,  
  disease = "Normal",  
  resource_dir = NULL,  
  work_dir = NULL,  
  clustering_resolution = NULL,  
  offline = FALSE,  
  cluster_annotation = TRUE  
)
```

Arguments

input	Path to a Seurat ‘.rds’ file carrying raw counts.
tag	Short name for the run; names the results directory.
species	“Human”, “Mouse” or “Rat”.
tissue	Free-text tissue, resolved against UBERON. Several names joined by “ ” (e.g. “urinary bladderladipose tissue”) are one operator-fixed list: the marker menu is the union of each listed tissue’s own UBERON closure, every retrieved candidate records which listed tissue admitted it, and the list is written verbatim to the run manifest. A single name is the default behaviour.

counts_source	Where the raw counts live, as "<assay>/<layer>" (for example "RNA/counts"). Required and never guessed.
disease	Disease context; several values may be given as a character vector. Defaults to "Normal".
resource_dir	The downloaded marker resource directory (see [download_resources()]). Verified before the run starts.
work_dir	Cache and results root; defaults to './.scmarkeragent'.
clustering_resolution	Leiden resolution; 'NULL' keeps the shared default.
offline	Run the deterministic stages only, without contacting a model.
cluster_annotation	Set 'FALSE' to skip the annotating agent; the label is then the top of the retrieval order, and every field recording how the label was produced says so.

Value

The parsed 'run_manifest.json', invisibly: run status, per-stage outcomes, and the paths of every delivered file.

check_input	<i>Validate an input file before a run starts</i>
-------------	---

Description

Checks that the file exists and that its format belongs to this arm, and explains the way out when it does not. The check is at the entry point rather than inside the reader because the two failures it prevents look nothing alike to a user: an '.h5ad' handed to this arm dies inside 'readRDS' with a message about an unknown format, and a path typo dies even later, after preprocessing has already written a cache directory. Both are the same mistake, and both are answerable here in one sentence.

Usage

```
check_input(path, expected_arm = "r")
```

Arguments

path	Path to the input object.
expected_arm	Which arm is asking; "r" for this package.

Value

The normalized path, invisibly usable as the validated input.

download_resources	<i>Download the curated marker resource</i>
--------------------	---

Description

Download the curated marker resource

Usage

```
download_resources(dest, url = NULL)
```

Arguments

dest	Directory to unpack into; pass it to a run as 'resource_dir'.
url	Archive URL. Defaults to the one recorded in the resource index; when no URL is configured yet, the function explains how to install the bundle by hand.

Value

The destination path invisibly.

resource_index	<i>The shipped resource index</i>
----------------	-----------------------------------

Description

The shipped resource index

Usage

```
resource_index(path = .index_path())
```

Arguments

path	Optional path to an index file; defaults to the packaged one.
------	---

Value

The index as a list: bundle name, total size and per-file size and SHA-256.

verify_resources	<i>Verify an existing resource directory</i>
------------------	--

Description

Verify an existing resource directory

Usage

```
verify_resources(dest)
```

Arguments

dest	Directory holding the unpacked resource.
------	--

Value

'TRUE' invisibly when every indexed file is present and intact; otherwise it stops with the list of problems.

Index

annotate, [2](#)

check_input, [3](#)

download_resources, [4](#)

resource_index, [4](#)

verify_resources, [5](#)