

Package: rsynthbio (via r-universe)

October 16, 2026

Type Package

Title Synthesize Bio API Wrapper

Version 4.2.3

Description Access Synthesize Bio models from their API
<<https://app.synthesize.bio/>> using this wrapper that provides
a convenient interface to the Synthesize Bio API, allowing
users to generate realistic gene expression data based on
specified biological conditions. This package enables
researchers to easily access AI-generated transcriptomic data
for various modalities including bulk RNA-seq, single-cell
RNA-seq, microarray data, and more.

URL <https://github.com/synthesizebio/rsynthbio>

BugReports <https://github.com/synthesizebio/rsynthbio/issues>

Imports getPass, keyring, jsonlite, http

Suggests rmarkdown, knitr, testthat (>= 3.0.2), mockery, arrow

Config/testthat/edition 3

Encoding UTF-8

RoxygenNote 7.3.3

LazyData true

VignetteBuilder knitr

License MIT + file LICENSE

Config/pak/sysreqs libsecret-1-dev libssl-dev

Repository <https://synthesizebio.r-universe.dev>

Date/Publication 2026-10-16 23:33:41 UTC

RemoteUrl <https://github.com/synthesizebio/rsynthbio>

RemoteRef HEAD

RemoteSha 9343cda9cd171d6650a2b8c6644b3ff3ebab0257

Contents

API_BASE_URL	2
clear_synthesize_token	2
DEFAULT_POLL_INTERVAL_SECONDS	3
DEFAULT_POLL_TIMEOUT_SECONDS	3
DEFAULT_TIMEOUT	4
get_example_query	4
has_synthesize_token	5
list_models	6
load_synthesize_token_from_keyring	7
predict_query	7
SELF_HOSTED_TIMEOUT	9
set_synthesize_token	10
Index	11

API_BASE_URL	<i>API Base URL</i>
--------------	---------------------

Description

Base URL for the Synthesize Bio API

Usage

API_BASE_URL

Format

An object of class character of length 1.

clear_synthesize_token	<i>Clear Synthesize Bio API Token</i>
------------------------	---------------------------------------

Description

Clears the Synthesize Bio API token from the environment for the current R session. This is useful for security purposes when you’ve finished working with the API or when switching between different accounts.

Usage

clear_synthesize_token(remove_from_keyring = FALSE)

Arguments

`remove_from_keyring`

Logical, whether to also remove the token from the system keyring if it's stored there. Defaults to FALSE.

Value

Invisibly returns TRUE.

Examples

```
## Not run:
# Clear token from current session only
clear_synthesize_token()

# Clear token from both session and keyring
clear_synthesize_token(remove_from_keyring = TRUE)

## End(Not run)
```

DEFAULT_POLL_INTERVAL_SECONDS

Default Poll Interval

Description

Default polling interval (seconds) for async model queries

Usage

DEFAULT_POLL_INTERVAL_SECONDS

Format

An object of class `numeric` of length 1.

DEFAULT_POLL_TIMEOUT_SECONDS

Default Poll Timeout

Description

Default maximum timeout (seconds) for async model queries

Usage

DEFAULT_POLL_TIMEOUT_SECONDS

Format

An object of class `numeric` of length 1.

DEFAULT_TIMEOUT	<i>Default Timeout</i>
-----------------	------------------------

Description

Default timeout (seconds) for outbound HTTP requests

Usage

DEFAULT_TIMEOUT

Format

An object of class `numeric` of length 1.

get_example_query	<i>Get Example Query for Model</i>
-------------------	------------------------------------

Description

Retrieves an example query structure for a specific model. This provides a template that can be modified for your specific needs.

Usage

get_example_query(model_id, api_base_url = NULL, self_hosted = NULL)

Arguments

model_id	Character string specifying the model ID (e.g., "gem-1-bulk", "gem-1-sc").
api_base_url	The base URL for the API server. When <code>NULL</code> (default), it is resolved from the 'SYNTHESIZE_API_BASE_URL' environment variable, falling back to the production default (<code>API_BASE_URL</code>). Point this at a self-hosted model container to fetch its example query.
self_hosted	Logical; when <code>TRUE</code> , the request targets a self-hosted container and does not require an API key (one is only sent if set). When <code>NULL</code> (default), it is resolved from the 'SYNTHESIZE_SELF_HOSTED' environment variable (truthy for 1/true/yes/on), defaulting to <code>FALSE</code> .

Value

A list representing a valid query structure for the specified model.

Examples

```
## Not run:
# Get example query for bulk RNA-seq model
query <- get_example_query(model_id = "gem-1-bulk")$example_query

# Get example query for single-cell model
query_sc <- get_example_query(model_id = "gem-1-sc")$example_query

# Modify the query structure
query$inputs[[1]]$num_samples <- 10

# Fetch from a self-hosted container (no API key required)
query <- get_example_query(
  model_id = "gem-1-bulk",
  api_base_url = "https://gem-1-bulk.internal.partner.example",
  self_hosted = TRUE
)$example_query

## End(Not run)
```

has_synthesize_token *Check if Synthesize Bio API Token is Set*

Description

Checks whether a Synthesize Bio API token is currently set in the environment. Useful for conditional code that requires an API token.

Usage

```
has_synthesize_token()
```

Value

Logical, TRUE if token is set, FALSE otherwise.

Examples

```
## Not run:
# Check if token is set
if (!has_synthesize_token()) {
  # Prompt for token if not set
  set_synthesize_token()
}

## End(Not run)
```

list_models	<i>List Available Models</i>
-------------	------------------------------

Description

Returns a list of all models available in the Synthesize Bio API. Each model has a unique ID that can be used with `predict_query()` and `get_example_query()`.

Usage

```
list_models(api_base_url = NULL, self_hosted = NULL)
```

Arguments

api_base_url	The base URL for the API server. When NULL (default), it is resolved from the 'SYNTHESIZE_API_BASE_URL' environment variable, falling back to the production default (API_BASE_URL). Point this at a self-hosted model container to list its models.
self_hosted	Logical; when TRUE, the request targets a self-hosted container and does not require an API key (one is only sent if set). When NULL (default), it is resolved from the 'SYNTHESIZE_SELF_HOSTED' environment variable (truthy for 1/true/yes/on), defaulting to FALSE.

Value

A list or data frame containing available models with their IDs and metadata.

Examples

```
## Not run:
# Get all available models
models <- list_models()
print(models)

# List models from a self-hosted container (no API key required)
models <- list_models(
  api_base_url = "https://gem-1-bulk.internal.partner.example",
  self_hosted = TRUE
)

## End(Not run)
```

`load_synthesize_token_from_keyring`*Load Synthesize Bio API Token from Keyring*

Description

Loads the previously stored Synthesize Bio API token from the system keyring and sets it in the environment for the current session.

Usage

```
load_synthesize_token_from_keyring()
```

Value

Invisibly returns TRUE if successful, FALSE if token not found in keyring.

Examples

```
## Not run:  
# Load token from keyring  
load_synthesize_token_from_keyring()  
  
## End(Not run)
```

`predict_query`*Predict Gene Expression*

Description

Sends a query to the Synthesize Bio API for prediction and retrieves gene expression samples. This function sends the query to the API and processes the response into usable data frames.

Usage

```
predict_query(  
  query,  
  model_id,  
  api_base_url = NULL,  
  poll_interval_seconds = DEFAULT_POLL_INTERVAL_SECONDS,  
  poll_timeout_seconds = DEFAULT_POLL_TIMEOUT_SECONDS,  
  return_download_url = FALSE,  
  raw_response = FALSE,  
  self_hosted = NULL,  
  ...  
)
```

Arguments

query	<p>A list representing the query data to send to the API. Use <code>'get_example_query()'</code> to generate an example. The query supports additional optional fields:</p> <ul style="list-style-type: none"> • <code>'total_count'</code> (integer): Library size used when converting predicted log CPM back to raw counts. Higher values scale counts up proportionally. • <code>'deterministic_latents'</code> (logical): If TRUE, the model uses the mean of each latent distribution instead of sampling, producing deterministic outputs for the same inputs. Useful for reproducibility. • <code>'seed'</code> (integer): Random seed for reproducibility.
model_id	Character string specifying the model ID (e.g., "gem-1-bulk", "gem-1-sc"). Use <code>'list_models()'</code> to see available models.
api_base_url	<p>The base URL for the API server. When NULL (default), it is resolved in order from the per-model environment variable <code>'SYNTHESIZE_API_BASE_URL__<MODEL>'</code> (e.g. <code>'SYNTHESIZE_API_BASE_URL__GEM_1_BULK'</code>), then the global <code>'SYNTHESIZE_API_BASE_URL'</code>, then the production default (<code>API_BASE_URL</code>). The per-model variable lets you point each self-hosted model at its own container once and omit <code>'api_base_url'</code> on every call.</p>
poll_interval_seconds	Seconds between polling attempts of the status endpoint. Default is <code>DEFAULT_POLL_INTERVAL_SECONDS</code> (2).
poll_timeout_seconds	Maximum total seconds to wait before timing out. Default is <code>DEFAULT_POLL_TIMEOUT_SECONDS</code> (900 = 15 minutes).
return_download_url	Logical, if TRUE, returns a list containing the signed download URL instead of parsing into data frames. Default is FALSE.
raw_response	Logical, if TRUE, returns the raw (unformatted) response from the API without applying any output transformers. For the production path this is the parsed JSON; for <code>'self_hosted = TRUE'</code> it is the parsed Arrow <code>'Table'</code> together with its schema metadata. Default is FALSE.
self_hosted	<p>Logical, if TRUE, sends a single synchronous request to a self-hosted model container that returns predictions as an Apache Arrow IPC stream (no polling, no download URL). Requires the optional <code>'arrow'</code> package and an <code>'api_base_url'</code> pointing at the container. Unlike the production path, no API key is required (one is only sent if configured). When NULL (default), it is resolved from the <code>'SYNTHESIZE_SELF_HOSTED'</code> environment variable (truthy for 1/true/yes/on), defaulting to FALSE.</p>
...	Additional parameters to include in the query body. These are passed directly to the API and validated server-side.

Value

A list. For the production path, if `'return_download_url'` is `'FALSE'` (default) the list contains `'metadata'` and `'expression'` data frames; if `'TRUE'` it contains `'download_url'` and empty data frames. For `'self_hosted = TRUE'`, the list contains the transformed data frames (`'metadata'`, `'expression'`, and `'latents'`; plus `'classifier_probs'` for metadata-prediction models) with `'model_version'` and `'request_type'` attached as attributes.

Examples

```

# Set your API key (in practice, use a more secure method)
## Not run:

# To start using rsynthbio, first you need to have an account with synthesize.bio.
# Go here to create one: https://app.synthesize.bio/

set_synthesize_token()

# Get available models
models <- list_models()

# Create a query for a specific model
query <- get_example_query(model_id = "gem-1-bulk")$example_query

# Request raw counts
result <- predict_query(query, model_id = "gem-1-bulk")

# Access the results
metadata <- result$metadata
expression <- result$expression

# Explore the top expressed genes in the first sample
head(sort(expression[1, ], decreasing = TRUE))

# Use deterministic latents for reproducible results
query$deterministic_latents <- TRUE
result_det <- predict_query(query, model_id = "gem-1-bulk")

# Specify a custom total count (library size)
query$total_count <- 5000000
result_custom <- predict_query(query, model_id = "gem-1-bulk")

# Self-hosted container returning a synchronous Apache Arrow IPC stream
result_sh <- predict_query(
  query,
  model_id = "gem-1-bulk",
  api_base_url = "https://gem-1-bulk.internal.partner.example",
  self_hosted = TRUE
)

## End(Not run)

```

SELF_HOSTED_TIMEOUT	<i>Self-Hosted Timeout</i>
---------------------	----------------------------

Description

Timeout (seconds) for synchronous self-hosted container predictions. These run on the partner's GPU box and can take minutes for large sample counts, so they use a longer timeout than hosted control-plane calls.

Usage

```
SELF_HOSTED_TIMEOUT
```

Format

An object of class `numeric` of length 1.

```
set_synthesize_token
```

Set Synthesize Bio API Token

Description

Securely prompts for and stores the Synthesize Bio API token in the environment. This function uses `getPass` to securely handle the token input without displaying it in the console. The token is stored in the `SYNTHESIZE_API_KEY` environment variable for the current R session.

Usage

```
set_synthesize_token(use_keyring = FALSE, token = NULL)
```

Arguments

<code>use_keyring</code>	Logical, whether to also store the token securely in the system keyring for future sessions. Defaults to <code>FALSE</code> .
<code>token</code>	Character, optional. If provided, uses this token instead of prompting. This parameter should only be used in non-interactive scripts.

Value

Invisibly returns `TRUE` if successful.

Examples

```
# Interactive prompt for token
## Not run:
set_synthesize_token()

# Provide token directly (less secure, not recommended for interactive use)
set_synthesize_token(token = "your-token-here")

# Store in system keyring for future sessions
set_synthesize_token(use_keyring = TRUE)

## End(Not run)
```

Index

* **datasets**

- API_BASE_URL, [2](#)
- DEFAULT_POLL_INTERVAL_SECONDS, [3](#)
- DEFAULT_POLL_TIMEOUT_SECONDS, [3](#)
- DEFAULT_TIMEOUT, [4](#)
- SELF_HOSTED_TIMEOUT, [9](#)

API_BASE_URL, [2](#)

clear_synthesize_token, [2](#)

DEFAULT_POLL_INTERVAL_SECONDS, [3](#)

DEFAULT_POLL_TIMEOUT_SECONDS, [3](#)

DEFAULT_TIMEOUT, [4](#)

get_example_query, [4](#)

has_synthesize_token, [5](#)

list_models, [6](#)

load_synthesize_token_from_keyring, [7](#)

predict_query, [7](#)

SELF_HOSTED_TIMEOUT, [9](#)

set_synthesize_token, [10](#)