

# Package ‘tuberculosis’

April 10, 2025

**Title** Tuberculosis Gene Expression Data for Machine Learning

**Description** The tuberculosis R/Bioconductor package features tuberculosis gene expression data for machine learning. All human samples from GEO that did not come from cell lines, were not taken postmortem, and did not feature recombination have been included. The package has more than 10,000 samples from both microarray and sequencing studies that have been processed from raw data through a hyper-standardized, reproducible pipeline.

**biocViews** ExperimentHub, ExpressionData, Homo\_sapiens\_Data, MicroarrayData, SequencingData, ReproducibleResearch

**Version** 1.13.0

**License** Artistic-2.0

**Depends** R (>= 4.1.0), SummarizedExperiment

**Imports** AnnotationHub, ExperimentHub, S4Vectors, dplyr, magrittr, purrr, rlang, stringr, tibble, tidyr

**Suggests** BiocStyle, ggplot2, hrbrthemes, knitr, readr, rmarkdown, scater, usethis, utils

**URL** <https://github.com/schifferl/tuberculosis>

**BugReports** <https://github.com/schifferl/tuberculosis/issues>

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.1

**VignetteBuilder** knitr

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

**git\_branch** devel

**git\_last\_commit** 777ac2e

**git\_last\_commit\_date** 2024-10-29

**Repository** Bioconductor 3.21

**Date/Publication** 2025-04-10

**Author** Lucas Schiffer [aut, cre] (ORCID:  
<<https://orcid.org/0000-0003-3628-0326>>)

**Maintainer** Lucas Schiffer <[schiffer.lucas@gmail.com](mailto:schiffer.lucas@gmail.com)>

## Contents

|                        |          |
|------------------------|----------|
| tuberculosis . . . . . | 2        |
| <b>Index</b>           | <b>3</b> |

---

|              |   |
|--------------|---|
| tuberculosis | <i>find/get tuberculosis gene expression data</i> |
|--------------|---|

---

### Description

To find or get tuberculosis gene expression data, users will use the `tuberculosis` function. The `dryrun` argument allows users to test a query prior to returning resources. When `dryrun = TRUE`, the function will print the names of matching resources as a message and return them invisibly as a character vector. When `dryrun = FALSE`, the function will either download resources from ExperimentHub or load them from the user's local cache. If a resource has multiple creation dates, the most recent is selected by default; add a date to override this behavior.

### Usage

```
tuberculosis(pattern, dryrun = TRUE)
```

### Arguments

|                      |  |
|----------------------|--|
| <code>pattern</code> | regular expression pattern to look for in the titles of resources available in tuberculosis; "." will return all resources   |
| <code>dryrun</code>  | if TRUE (the default), a character vector of resource names is returned invisibly; if FALSE, a list of resources is returned |

### Value

if `dryrun = TRUE`, a character vector of resource names is returned invisibly; if `dryrun = FALSE`, a list of resources is returned

### Examples

```
tuberculosis("GSE103147")
```

# Index

tuberculosis, [2](#)