Package 'TBSignatureProfiler'

October 30, 2026

Title Profile RNA-Seq Data Using TB Pathway Signatures

Version 1.25.0

Description Gene signatures of TB progression, TB disease, and other TB disease states have been validated and published previously. This package aggregates known signatures and provides computational tools to enlist their usage on other datasets. The TBSignatureProfiler makes it easy to profile RNA-Seq data using these signatures and includes common signature profiling tools including ASSIGN, GSVA, and ssGSEA. Original models for some gene signatures are also available. A shiny app provides some functionality alongside for detailed command line accessibility.

License MIT + file LICENSE

URL https://github.com/wejlab/TBSignatureProfiler, https://wejlab.github.io/TBSignatureProfiler-docs/

BugReports https://github.com/wejlab/TBSignatureProfiler/issues

Depends R (>= 4.4.0)

Imports ASSIGN (>= 1.23.1), BiocParallel, ComplexHeatmap, DESeq2, DT, edgeR, gdata, ggplot2, glmnet, GSVA (>= 1.51.3), HGNChelper, magrittr, methods, pROC, RColorBrewer, reshape2, ROCit, S4Vectors, singscore, stats, SummarizedExperiment, tibble

Suggests BiocStyle, caret, circlize, class, covr, dplyr, e1071, impute, knitr, lintr, MASS, plyr, randomForest, rmarkdown, shiny, spelling, sva, testthat

VignetteBuilder knitr

biocViews GeneExpression, DifferentialExpression

Encoding UTF-8

Language en-US

LazyData TRUE

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

git_url https://git.bioconductor.org/packages/TBSignatureProfiler

git_branch devel

git_last_commit 3f57ba4

git_last_commit_date 2026-04-28

Repository Bioconductor 3.24

Date/Publication 2026-10-29

Author Kiloni Quiles [cre] (ORCID: <https://orcid.org/0000-0002-7521-5255>),
Aubrey R. Odom [aut, dtm] (ORCID:
<https://orcid.org/0000-0001-7113-7598>),
David Jenkins [aut, org] (ORCID:
<https://orcid.org/0000-0002-7451-4288>),
Xutao Wang [aut],
Yue Zhao [ctb] (ORCID: <https://orcid.org/0000-0001-5257-5103>),
Christian Love [ctb],
W. Evan Johnson [aut]

Maintainer Kiloni Quiles <kiloni.quiles@rutgers.edu>

Contents

.OriginalModel_NoRetraining . . . 3
.OriginalModel_Retraining . . . 4
addTBsignature . . . 5
bootstrapAUC . . . 6
Bootstrap_LOOCV_LR_AUC . . . 7
common_sigAnnotData . . . 8
compareAlgs . . . 9
compareBoxplots . . . 11
COVIDsignatures . . . 12
cv_glmnet_OriginalModel . . . 16
deseq2_norm_rle . . . 16
distinctColors . . . 17
evaluateOriginalModel . . . 17
knn_OriginalModel . . . 19
lda_OriginalModel . . . 19
LOOAUC_simple_multiple_noplot_one_df . . . 20
mkAssay . . . 20
ObtainSampleScore_OriginalModel . . . 22
OriginalTrainingData . . . 23
plotQuantitative . . . 23
randomForest_OriginalModel . . . 25
ref_combat_impute . . . 25
runTBsigProfiler . . . 26
sigAnnotData . . . 29
signatureBoxplot . . . 30
signatureGeneHeatmap . . . 31
signatureHeatmap . . . 33
SignatureQuantitative . . . 36
signatureROCplot . . . 37
signatureROCplot_CI . . . 38
subsetGeneSet . . . 40
SulimanOriginalModel . . . 41
svm_OriginalModel . . . 41
tableAUC . . . 42

TBcommon 43
TBsignatures 51
TBsignaturesSplit 59
TBSPapp 59
TB_hiv 60
TB_indian 61

Index **62**

`.OriginalModel_NoRetraining` *TB gene signatures that do not require retraining.*

Description

A function to obtain predicted score for TB gene signatures that do not need to be retrained.

Usage

```
.OriginalModel_NoRetraining(input, useAssay, geneSignaturesName, BPPARAM)
```

Arguments

`input`	A SummarizedExperiment object with gene symbols as the assay row names.
`useAssay`	A character string or an integer specifying the assay in the `input`. Default is the first assay in the assay list. Used for the test SummarizedExperiment object. Default is `1`, indicating the first assay in the `input`.
`geneSignaturesName`	A character string/vector specifying the signature of interest. If `any(geneSignaturesName == "") == TRUE`, it will run all available gene signatures' original models.
`BPPARAM`	An instance inherited from `bplapply`.

Details

Anderson_42 and Anderson_OD_51 used difference of sums to calculate prediction scores. Difference of sums is obtained by subtracting the sum of the expression of genes within signatures that are down-regulated from the sum of the expression of genes that are up-regulated within signatures. Kaforou_27, Kaforou_OD_44, and Kaforou_OD_53 used difference of arithmetic means to calculate prediction scores. Sweeney_OD_3 used difference of arithmetic mean to calculate prediction score.

Value

A SummarizedExperiment object with predicted scores for each sample obtained from the signature's original model.

`.OriginalModel_Retraining` *TB gene signatures that require retraining.*

Description

A function to obtain predicted score for TB gene signatures that need retraining of original models.

Usage

```
.OriginalModel_Retraining(input, useAssay, geneSignaturesName, adj, BPPARAM)
```

Arguments

`input`	A SummarizedExperiment object with gene symbols as the assay row names.
`useAssay`	A character string or an integer specifying the assay in the `input`. Default is the first assay in the assay list. Used for the test SummarizedExperiment object. Default is `1`, indicating the first assay in the `input`.
`geneSignaturesName`	A character string/vector specifying the signature of interest. If `any(geneSignaturesName == "") == TRUE`, it will run all available gene signatures' original models.
`adj`	A small positive real number used in `ComBat` to solve for genes with 0 counts(rare cases). Default is `1e-3`.
`BPPARAM`	An instance inherited from `bplapply`.

Details

Maertzdorf_4 and Maertzdorf_15 were trained using a random forest to distinguish patients with active TB from healthy controls.

Verhagen_10 was also trained using a random forest to distinguish samples with active TB from either latent infection or healthy controls. The random forest model was build using `randomForest`.

Jacobsen_3 were trained using linear discriminant analysis (LDA) to distinguish samples with active TB from latent infection status.

Sambarey_HIV_10 were also trained using LDA to distinguish samples with active TB from either latent infection, healthy control, or other disease (HIV). The LDA model was built using `lda`.

Berry_OD_86 and Berry_393 were trained using K-nearest neighbors (KNN) model to differentiate samples with active TB from latent infection status. The KNN model was built using `knn`.

Suliman_RISK_4 and Zak_RISK_16 were trained using support vector machines (SVM) to distinguish TB progressor from non-progressors. The input gene expression features for Suliman_RISK_4 used the paired ratio of GAS6/CD1C, SEPTIN4/BLK, SEPTIN4/CD1C, GAS6/BLK. The SVM model was built using `svm`.

Value

A SummarizedExperiment object with predicted scores for each sample obtained from the signature's original model.

addTBsignature *Introduce a new signature into the TBSignatureProfiler.*

Description

This function allows users to integrate new signatures into the TBSP with a function that updates the TBsignatures, TBcommon, sigAnnotData and common_sigAnnotData objects. Users that wish to use this function should do so with the downloaded package as a working directory, and not as a casual package function. This function does not complete all required updates to the package for a signature to be full added; users should check the vignette "Submitting Signatures to the TBSP Package" on the TBSP website for a walkthrough of this complete process. Also note that this function only adds one signature at a time, and must me run multiple times to add subsequent signatures.

Usage

```
addTBsignature(
  sigsymbols,
  authname,
  signame_common = NULL,
  sigtype,
  tissuetype,
  saveobjs = FALSE,
  views = TRUE
)
```

Arguments

sigsymbols	a character vector of the gene symbols that compose the signature to be added. Required.
authname	a character string containing the last name of the primary author of the publication where the signature was first identified. If spaces are present, omit them, and use proper capitalization. Required.
signame_common	a character string of the alternate name of the signature given by the publication, if it exists. If NULL, no assigned name is assumed to exist. Default is null.
sigtype	a character string that gives the context that the signature was developed under. Most commonly, it will distinguish TB from LTBI ("Disease"), TB from some combination of other diseases and possibly LTBI ("Disease/Other Diseases"), TB from Human Immunodeficiency Virus ("Disease/HIV"), TB from pneumonia ("Disease/Pneumonia"), or identify risk of progression to TB ("risk"), risk of TB treatment failure ("failure"), or classify treatment responses (i.e., failures from cures, "response"). Required.
tissuetype	a character string that denotes whether the signature was developed using samples of either whole blood/paxgene ("whole blood") or peripheral blood mononuclear cells ("PBMC"). Due to the manipulation of cells inherently required to obtain PBMCs, many scientists prefer to use only whole blood samples for analysis. Accepts "whole blood", "PBMC" or "mixed". Required.

saveobjs — logical. If TRUE, the contents of the data file (TBsignatures, TBcommon, sigAnnotData, common_sigAnnotData) will be overwritten and updated to include the new signature. If FALSE, no files will be overwritten, but you can check function output for errors before writing RDS objects by setting views = TRUE. Default is saveobjs = FALSE.

views — logical. If TRUE, all objects will be sent to a data view in a new window to check for errors. Default is TRUE.

Value

Either data objects TBsignatures, TBcommon, sigAnnotData, and common_sigAnnotData will be updated with the new signature and overwritten if saveobjs = FALSE, or no output will be produced except errors and messages for checking that the function runs correctly given the inputs.

Examples

```
# Mock example signature
TBSignatureProfiler:::addTBsignature(sigsymbols = c("GBP5", "BATF2", "GZMA"),
                             authname = "Odom",
                             signame_common = NULL,
                             sigtype = "Disease/HIV",
                             tissuetype = "PBMC",
                             saveobjs = FALSE,
                             views = FALSE)
```

---

bootstrapAUC *Bootstrap the AUC and conduct T-Tests for a collection of signatures.*

Description

Run bootstrapping of the AUC and derive the p-value for a 2-sample t-test for all signatures tested on a given dataset.

Usage

```
bootstrapAUC(
  SE_scored,
  annotationColName,
  signatureColNames,
  num.boot = 100,
  pb.show = TRUE
)
```

Arguments

SE_scored — a SummarizedExperiment object with genes as the row features and signature scores in the colData. There should also be a column of annotation data. Required.

annotationColName — a character string giving the column name in colData that contains the annotation data. Required.

signatureColNames
: a vector of column names in the colData that contain the signature score data. Required.

num.boot
: integer. The number of times to bootstrap the data. The default is 100.

pb.show
: logical for whether to show a progress bar while running code. The default is TRUE.

Value

A list of length 5 returning a vector of p-values for a 2-sample t-test, bootstrapped AUC values, an AUC value for using all scored values for all signatures specified in signatureColNames, and values for the lower and upper bounds of a bootstrapped AUC confidence interval using pROC::roc().

Examples

```
# Run signature profiling
choose_sigs <- list("madeupsig" = c("FCRL3", "OAS2", "IFITM3"))
prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                algorithm = "ssGSEA",
                                combineSigAndAlgorithm = TRUE,
                                signatures = choose_sigs,
                                parallel.sz = 1)
# Bootstrapping
booted <- bootstrapAUC(SE_scored = prof_indian, annotationColName = "label",
                       signatureColNames = names(choose_sigs), num.boot = 2)
booted
```

Bootstrap_LOOCV_LR_AUC

Bootstrap on Leave-one-out CV with Logistic Regression.

Description

Bootstrap on Leave-one-out CV with Logistic Regression.

Usage

```
Bootstrap_LOOCV_LR_AUC(df, targetVec, nboot)
```

Arguments

df
: a data.frame of gene expression count data. Required.

targetVec
: a binary vector of the response variable. Should be the same number of rows as df. Required.

nboot
: an integer specifying the number of bootstrap iterations.

Value

A list of length 2 with elements

auc	A vector the length of nboot with the AUC from each bootstrap iteration.
byClass	A dataframe with number of rows equal to nboot. Each row contains the sensitivity, specificity, positive predictive value, negative predictive value, precision, recall, F1, prevalence, detection rate, detection prevalence and balanced accuracy for that bootstrap iteration.

---

common_sigAnnotData *Annotation information for published TB signatures.*

Description

A data.frame of annotation information for published tuberculosis signatures. This table differs from that of sigAnnotData as it refers to signatures via the name given in scientific publications, and via a consistent naming system otherwise. Currently, this table includes two variables, disease and tissue type.

Usage

```
common_sigAnnotData
```

Format

data.frame

Details

The disease variable indicates whether the signature was developed to distinguish TB from LTBI ("Disease"), TB from some combination of other diseases and possibly LTBI ("OD"), TB from Human Immunodeficiency Virus ("HIV"), TB from pneumonia ("PNA"), or identify risk of progression to TB ("RISK"), risk of TB treatment failure ("FAIL"), or classify treatment responses (i.e., failures from cures, "RES").

The tissue type variable denotes whether the signature was developed using samples of either whole blood/paxgene or peripheral blood mononuclear cells (PBMCs). Due to the manipulation of cells inherently required to obtain PBMCs, many scientists prefer to use only whole blood samples for analysis.

Source

See ?TBcommon for reference information.

Examples

```
data("common_sigAnnotData")
```

compareAlgs *Compare scoring algorithms on a single signature via heatmap or boxplot.*

Description

It may be useful to compare the results of scoring across several different scoring algorithms via a method of visualization, such as a heatmap. The compareSigs function allows the input of a SummarizedExperiment data object and conducts profiling on each signature desired, and outputting a heatmap or boxplot for each signature.

Usage

```
compareAlgs(
  input,
  signatures = NULL,
  annotationColName,
  useAssay = "counts",
  algorithm = c("GSVA", "ssGSEA", "ASSIGN", "PLAGE", "Zscore", "singscore"),
  showColumnNames = TRUE,
  showRowNames = TRUE,
  scale = FALSE,
  colorSets = c("Set1", "Set2", "Set3", "Pastel1", "Pastel2", "Accent", "Dark2",
    "Paired"),
  choose_color = c("blue", "gray95", "red"),
  colList = list(),
  show.pb = FALSE,
  parallel.sz = 0,
  output = "heatmap",
  num.boot = 100,
  column_order = NULL
)
```

Arguments

input	an input data object of the class "SummarizedExperiment". Required.
signatures	a list of signatures to run with their associated genes. This list should be in the same format as TBsignatures, included in the TBSignatureProfiler package. If signatures = NULL, the default set of signatures TBsignatures list is used. For details, run ?TBsignatures. If <2 genes in a signature are present in the sample, that signature will not be evaluated and will not be present in the resulting SE object. The default is NULL.
annotationColName	a character string giving the column name in colData that contains the annotation data. Required.
useAssay	a character string specifying the assay to use for signature profiling when input is a SummarizedExperiment. Required only for input data of the class SummarizedExperiment. If null, the assay used will be "counts". The default is NULL.

algorithm	a vector of algorithms to run, or character string if only one is desired. The default is c("GSVA", "ssGSEA", "ASSIGN","PLAGE", "Zscore", "singscore"). NOTE: ASSIGN takes a long time to run and is not recommended for efficient use.
showColumnNames	logical. Setting showColumnNames = TRUE will show the column names (i.e. sample names) on the heatmap. The default is TRUE.
showRowNames	logical. Setting showColumnNames = TRUE will show the row names (i.e. signature names) on the heatmap. The default is TRUE.
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
colorSets	a vector of names listing the color sets in the order that they should be used in creating the heatmap. By default, this function will use the color sets in the order listed in Usage for annotation information. You may replace the default with the same collection of sets in order that you want to use them, or provide custom color sets with the colList parameter.
choose_color	a vector of color names to be interpolated for the heatmap gradient, or a colorRamp function produced by circlize::colorRamp2. The default is c("blue", "gray95", "red").
colList	a named list of named vectors specifying custom color information to pass to ComplexHeatmap::Heatmap(). The list should have as many elements as there are annotation columns, and each element name should correspond exactly with the name of each annotation column. The colors in the vector elements should be named according to the levels of the factor in that column's annotation data if the annotation is discrete, or it should be produced with circlize::colorRamp2 if the annotation is continuous. By default, ColorBrewer color sets will be used. See the the parameter colorSets for additional details.
show.pb	logical, whether warnings and other output from the profiling should be suppressed (including progress bar output). Default is FALSE.
parallel.sz	an integer identifying the number of processors to use when running the calculations in parallel for the GSVA and ssGSEA algorithms. If parallel.sz = 0, all cores are used. The default is 0.
output	a character string specifying whether the outputted plot should be a "heatmap" or "boxplot". The default is "heatmap".
num.boot	an integer indicating the number of times to bootstrap the data.
column_order	a vector of character strings indicating the order in which to manually arrange the heatmap columns. Default is NULL, such that column order is automatically determined via clustering.

Value

A heatmap or boxplot for each signature specified comparing the enumerated algorithms.

Examples

```
compareAlgs(TB_indian,
            signatures = TBsignatures[c("Gliddon_OD_3")],
            annotationColName = "label",
            algorithm = c("ssGSEA", "PLAGE"),
            scale = TRUE, parallel.sz = 1, output = "heatmap")
```

compareBoxplots *Create a comparison plot of boxplots for bootstrapped AUC values.*

Description

Present the results of AUC bootstrapping for a collection of scored signatures via boxplots.

Usage

```
compareBoxplots(
  SE_scored,
  annotationColName,
  signatureColNames,
  num.boot = 100,
  name = "Boxplot Comparison of Signature AUCs",
  pb.show = TRUE,
  abline.col = "red",
  fill.col = "gray79",
  outline.col = "black",
  rotateLabels = FALSE,
  violinPlot = FALSE
)
```

Arguments

SE_scored	a SummarizedExperiment object with genes as the row features and signature scores in the colData. There should also be a column of annotation data. Required.
annotationColName	a character string giving the column name in colData that contains the annotation data. Required.
signatureColNames	a vector of column names in the colData that contain the signature score data. Required.
num.boot	an integer indicating the number of times to bootstrap the data.
name	a character string giving the overall title for the plot. The default is "Boxplot Comparison of Signature AUCs".
pb.show	logical for whether to show a progress bar while running code. Default is TRUE.
abline.col	the color to be used for the dotted line at AUC = 0.5 (the chance line). The default is "red".
fill.col	the color to be used to fill the boxplots. The default is "white".
outline.col	the color to be used for the boxplot outlines. The default is "black".
rotateLabels	If TRUE, rotate labels. Default is FALSE.
violinPlot	logical. Setting violinPlot = TRUE creates violin plots in place of boxplots. The mean and +/- 1 standard deviation are added to the violin plot interior for each signature. The default is FALSE.

Value

A plot with side-by-side boxplots of bootstrapped AUC values for each specified signature.

Examples

```
# Run signature profiling
 choose_sigs <- TBsignatures[c("Zak_RISK_16", "Zhao_NANO_6")]
 prof_indian <- runTBsigProfiler(TB_indian[seq_len(25), ],
                                 useAssay = "logcounts",
                                 algorithm = "ssGSEA",
                                 signatures = choose_sigs,
                                 parallel.sz = 1)
 # Create boxplots
 compareBoxplots(prof_indian, annotationColName = "label",
                 signatureColNames = names(choose_sigs), rotateLabels = TRUE)
```

COVIDsignatures *A list of published/pre-print COVID-19 signatures.*

Description

A set of 47 COVID-19 gene signatures from various single-cell and bulk RNA-seq publications and preprint manuscripts from early- to mid-2020. This set of signatures uses gene symbols.

Usage

```
COVIDsignatures
```

Format

list

Details

Signature names are composed of the last name of the primary author, followed by the type of sequencing data from which the signature was derived, the tissue from which the signature was derived, and ending with a reference to the cell type, infection status, or disease to which the signature belongs, as defined in the original publication.

Note that in some cases signatures will be positive identifiers of COVID-19 as positive markers of immune cell clusters are often provided for single-cell RNA-seq data; this should be taken into account when creating ROC curves and computing any AUC or disease risk estimates.

Source

- **Wilk_sc_PBMC_monocytes_up**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "A single-cell atlas of the peripheral immune response in patients with severe COVID-19." Nature Medicine 26 (7): 1070-1076. https://doi.org/10.1038/s41591-020-0944-y
- **Wilk_sc_PBMC_monocytes_up**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "A single-cell atlas of the peripheral immune response in patients with severe COVID-19." Nature Medicine 26 (7): 1070-1076. https://doi.org/10.1038/s41591-020-0944-y

- **Wilk_sc_PBMC_monocytes_down**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "A single-cell atlas of the peripheral immune response in patients with severe COVID-19." Nature Medicine 26 (7): 1070-1076. https://doi.org/10.1038/s41591-020-0944-y
- **Wilk_sc_PBMC_NK_cells_up**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "A single-cell atlas of the peripheral immune response in patients with severe COVID-19." Nature Medicine 26 (7): 1070-1076. https://doi.org/10.1038/s41591-020-0944-y
- **Wilk_sc_PBMC_NK_cells_down**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "A single-cell atlas of the peripheral immune response in patients with severe COVID-19." Nature Medicine 26 (7): 1070-1076. https://doi.org/10.1038/s41591-020-0944-y
- **Wilk_sc_PBMCs_ISG_signature**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "A single-cell atlas of the peripheral immune response in patients with severe COVID-19." Nature Medicine 26 (7): 1070-1076. https://doi.org/10.1038/s41591-020-0944-y
- **Wilk_sc_PBMC_activated_granulocytes**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "A single-cell atlas of the peripheral immune response in patients with severe COVID-19." Nature Medicine 26 (7): 1070-1076. https://doi.org/10.1038/s41591-020-0944-y
- **Huang_sc_PBMC_IFN_signature**: Wilk, A.J., Rustagi, A., Zhao, N.Q. et al. 2020. "Blood single cell immune profiling reveals the interferon-MAPK pathway mediated adaptive immune response for COVID-19." medRxiv.org: https://doi.org/10.1101/2020.03.15.20033472
- **Wen_sc_PBMC_monocytes**: Wen, W., Su, W., Tang, H. et al. 2020. "Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing." Cell Discovery 6 (31). https://doi.org/10.1038/s41421-020-0168-9
- **Wen_sc_PBMC_NK_cells**: Wen, W., Su, W., Tang, H. et al. 2020. "Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing." Cell Discovery 6 (31). https://doi.org/10.1038/s41421-020-0168-9
- **Wen_sc_PBMC_CD4_T_cells**: Wen, W., Su, W., Tang, H. et al. 2020. "Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing." Cell Discovery 6 (31). https://doi.org/10.1038/s41421-020-0168-9
- **Wen_sc_PBMC_CD8_T_cells**: Wen, W., Su, W., Tang, H. et al. 2020. "Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing." Cell Discovery 6 (31). https://doi.org/10.1038/s41421-020-0168-9
- **Wen_sc_PBMC_B_cells**: Wen, W., Su, W., Tang, H. et al. 2020. "Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing." Cell Discovery 6 (31). https://doi.org/10.1038/s41421-020-0168-9
- **Xiong_bulk_PBMC_gene_signature_up**: Xiong Y, Liu Y, Cao L, et al. 2020. "Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients." Emerging Microbes & Infections 9 (1):761-770. https://doi/org/10.1080/22221751.2020.1747363
- **Xiong_bulk_PBMC_gene_signature_down**: Xiong Y, Liu Y, Cao L, et al. 2020. "Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients." Emerging Microbes & Infections 9 (1):761-770. https://doi/org/10.1080/22221751.2020.1747363
- **Xiong_sc_PBMC_cytokines_up**: Xiong Y, Liu Y, Cao L, et al. 2020. "Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients." Emerging Microbes & Infections 9 (1):761-770. https://doi/org/10.1080/22221751.2020.1747363
- **Xiong_sc_PBMC_cytokines_down**: Xiong Y, Liu Y, Cao L, et al. 2020. "Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients." Emerging Microbes & Infections 9 (1):761-770. https://doi/org/10.1080/22221751.2020.1747363

- **Liao_sc_BALF_G1_macrophages**: Liao, M., Liu, Y., Yuan, J. et al. 2020. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature Medicine 26 (6): 842-844. https://doi.org/10.1038/s41591-020-0901-9
- **Liao_sc_BALF_G1_2_macrophages**: Liao, M., Liu, Y., Yuan, J. et al. 2020. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature Medicine 26 (6): 842-844. https://doi.org/10.1038/s41591-020-0901-9
- **Liao_sc_BALF_G2_macrophages**: Liao, M., Liu, Y., Yuan, J. et al. 2020. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature Medicine 26 (6): 842-844. https://doi.org/10.1038/s41591-020-0901-9
- **Liao_sc_BALF_G3_macrophages**: Liao, M., Liu, Y., Yuan, J. et al. 2020. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature Medicine 26 (6): 842-844. https://doi.org/10.1038/s41591-020-0901-9
- **Liao_sc_BALF_G4_macrophages**: Liao, M., Liu, Y., Yuan, J. et al. 2020. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature Medicine 26 (6): 842-844. https://doi.org/10.1038/s41591-020-0901-9
- **Liao_sc_BALF_CD8_T_cells_up**: Liao, M., Liu, Y., Yuan, J. et al. 2020. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature Medicine 26 (6): 842-844. https://doi.org/10.1038/s41591-020-0901-9
- **Liao_sc_BALF_CD8_T_cells_down**: Liao, M., Liu, Y., Yuan, J. et al. 2020. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature Medicine 26 (6): 842-844. https://doi.org/10.1038/s41591-020-0901-9
- **Hadjadj_nanostring_WB_gene_signature_up**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_WB_gene_signature_down**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_WB_ISG_signature**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_WB_mild_moderate_up**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_WB_mild_moderate_down**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_WB_severe_up**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_WB_severe_down**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_critical_up**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027
- **Hadjadj_nanostring_critical_down**: Hadjadj J, Yatim N, Barnabei L, et al. 2020. "Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients." Science 369 (6504): 718-724. https://doi.org/10.1126/science.abc6027

- **Wei_sc_PBMC_inactivated_monocytes**: Wei et al. 2020. "Viral Invasion and Type I Interferon Response Characterize the Immunophenotypes during COVID-19 Infection." SSRN: https://dx.doi.org/10.2139/ssrn.3555695
- **Wei_sc_PBMC_classical_monocytes**: Wei et al. 2020. "Viral Invasion and Type I Interferon Response Characterize the Immunophenotypes during COVID-19 Infection." SSRN: https://dx.doi.org/10.2139/ssrn.3555695
- **Wei_sc_PBMCs_T_cells**: Wei et al. 2020. "Viral Invasion and Type I Interferon Response Characterize the Immunophenotypes during COVID-19 Infection." SSRN: https://dx.doi.org/10.2139/ssrn.3555695
- **Wei_sc_PBMC_B_cells**: Wei et al. 2020. "Viral Invasion and Type I Interferon Response Characterize the Immunophenotypes during COVID-19 Infection." SSRN: https://dx.doi.org/10.2139/ssrn.3555695
- **Silvin_sc_WB_combined_signature**: Silvin A, Chapuis N, Dunsmore G, et al. 2020. "Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19." Cell 182 (6): 1401-1418.E18. https://doi.org/10.1016/j.cell.2020.08.002
- **Silvin_sc_WB_monocytes_up**: Silvin A, Chapuis N, Dunsmore G, et al. 2020. "Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19." Cell 182 (6): 1401-1418.E18. https://doi.org/10.1016/j.cell.2020.08.002
- **Silvin_sc_WB_monocytes_down**: Silvin A, Chapuis N, Dunsmore G, et al. 2020. "Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19." Cell 182 (6): 1401-1418.E18. https://doi.org/10.1016/j.cell.2020.08.002
- **Silvin_sc_WB_neutrophils_up**: Silvin A, Chapuis N, Dunsmore G, et al. "Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19." Cell 182 (6): 1401-1418.E18. https://doi.org/10.1016/j.cell.2020.08.002
- **Silvin_sc_WB_neutrophils_down**: Silvin A, Chapuis N, Dunsmore G, et al. "Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19." Cell 182 (6): 1401-1418.E18. https://doi.org/10.1016/j.cell.2020.08.002
- **Arunachalam_bulk_PBMC_blood_modules**: Arunachalam PS, Wimmers F, Mok CKP, et al. 2020. "Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans." Science 369 (6508): 1210-1220. https://doi.org/10.1126/science.abc6261
- **Arunachalam_bulk_PBMC_covid_combined**: Arunachalam PS, Wimmers F, Mok CKP, et al. 2020. "Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans." Science 369 (6508): 1210-1220. https://doi.org/10.1126/science.abc6261
- **Arunachalam_bulk_PBMC_moderate**: Arunachalam PS, Wimmers F, Mok CKP, et al. 2020. "Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans." Science 369 (6508): 1210-1220. https://doi.org/10.1126/science.abc6261
- **Arunachalam_bulk_PBMC_severe**: Arunachalam PS, Wimmers F, Mok CKP, et al. 2020. "Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans." Science 369 (6508): 1210-1220. https://doi.org/10.1126/science.abc6261
- **Arunachalam_bulk_PBMC_intensive_care**: Arunachalam PS, Wimmers F, Mok CKP, et al. 2020. "Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans." Science 369 (6508): 1210-1220. https://doi.org/10.1126/science.abc6261
- **Dunning_bulk_WB_flu**: Dunning J, Blankley S, Hoang LT, et al. 2018. "Progression of whole-blood transcriptional signatures from interferon-induced to neutrophil-associated patterns in severe influenza." Nature Immunology 19 (6): 625-635. https://doi.org/10.1038/s41590-018-0111-5

Examples

```
data("COVIDsignatures")
```

`cv_glmnet_OriginalModel` *Train original model for gene signatures Leong_24, Leong_RISK_29, Zhao_NANO_6 using lasso logistic regression.*

Description

Train original model for gene signatures Leong_24, Leong_RISK_29, Zhao_NANO_6 using lasso logistic regression.

Usage

```
cv_glmnet_OriginalModel(dat_list, dat_test_sig)
```

Arguments

`dat_list`	A `list` contains training data and disease status outcomes from the discovery data of corresponding gene signatures.
`dat_test_sig`	A `data frame` contains corresponding gene sets from the `input`.

Value

The predicted score for each sample in the test study.

`deseq2_norm_rle` *Normalize gene expression count data.*

Description

Normalize gene expression count data.

Usage

```
deseq2_norm_rle(inputData)
```

Arguments

`inputData`	a `data.frame` or `matrix` of gene expression count data. Required.

Value

A `data.frame` or `matrix` of normalized count data.

Examples

```
## Example using the counts assay from a SummarizedExperiment
data_in <- SummarizedExperiment::assay(TB_indian, "counts")
res <- deseq2_norm_rle(data_in)
```

distinctColors *Generate a distinct palette for coloring different clusters.*

Description

Create a distinct palette for coloring different heatmap clusters. The function returns colors for input into `ComplexHeatmap:Heatmap()`, `signatureGeneHeatmap()` and `signatureHeatmap()`.

Usage

```
distinctColors(
  n,
 hues = c("red", "cyan", "orange", "blue", "yellow", "purple", "green", "magenta"),
  saturation.range = c(0.7, 1),
  value.range = c(0.7, 1)
)
```

Arguments

`n`	an integer describing the number of colors to generate. Required.
`hues`	a vector of character strings indicating the R colors available from the `colors()` function. These will be used as the base colors for the clustering scheme. Different saturations and values (i.e. darkness) will be generated for each hue. Default is `c("red", "cyan", "orange","blue", "yellow", "purple", "green", "magenta")`
`saturation.range`	a numeric vector of length 2 with values between 0 and 1 giving the range of saturation. The default is `c(0.25, 1)`.
`value.range`	a numeric vector of length 2 with values between 0 and 1 giving the range of values. The default is `c(0.5, 1)`.

Value

A vector of distinct colors that have been converted to HEX from HSV.

Examples

```
distinctColors(10)
```

evaluateOriginalModel *A function that implements the original methods for multiple TB signatures.*

Description

This function computes prediction for multiple TB signatures based on their training models/methods. To avoid naming issues, the gene names for both training data and input gene sets have been updated using the checkGeneSymbols. TB signatures with available original models are: Anderson_42, Anderson_OD_51, Kaforou_27, Kaforou_OD_44, Kaforou_OD_53, Sweeney_OD_3, Maertzdorf_4, Verhagen_10, Jacobsen_3, Sambarey_HIV_10, Leong_24, Berry_OD_86, Berry_393, Bloom_OD_144, Suliman_RISK_4, Zak_RISK_16, Leong_RISK_29, and Zhao_NANO_6. The predicted score for each signature has been stored in the column data section of the input SummarizedExperiment study.

Usage

```
evaluateOriginalModel(
  input,
  geneSignaturesName,
  useAssay = 1,
  adj = 0.001,
  BPPARAM = BiocParallel::SerialParam(progressbar = TRUE)
)
```

Arguments

input	A SummarizedExperiment object with gene symbols as the assay row names.
geneSignaturesName	A character string/vector specifying the signature of interest. If any(geneSignaturesName == "") == TRUE, it will run all available gene signatures' original models.
useAssay	A character string or an integer specifying the assay in the input. Default is the first assay in the assay list. Used for the test SummarizedExperiment object. Default is 1, indicating the first assay in the input.
adj	A small positive real number used in ComBat to solve for genes with 0 counts(rare cases). Default is 1e-3.
BPPARAM	An instance inherited from bplapply.

Value

A SummarizedExperiment object with predicted scores for each sample obtained from the signature's original model.

Examples

```
re <- evaluateOriginalModel(input = TB_indian,
                            geneSignaturesName = c("Anderson_42"),
                            useAssay = "counts")
re$Anderson_42_OriginalModel
```

knn_OriginalModel *Train original model for gene signatures Berry_393 and Berry_OD_86.*

Description

Train original model for gene signatures Berry_393 and Berry_OD_86.

Usage

```
knn_OriginalModel(dat_list, dat_test_sig)
```

Arguments

dat_list	A list contains training data and disease status outcomes from the discovery data of corresponding gene signatures.
dat_test_sig	A data frame contains corresponding gene sets from the input.

Value

The predicted score for each sample in the test study.

lda_OriginalModel *Train original model for gene signatures Jacobsen_3 and Sambarey_HIV_10.*

Description

Train original model for gene signatures Jacobsen_3 and Sambarey_HIV_10.

Usage

```
lda_OriginalModel(dat_list, dat_test_sig)
```

Arguments

dat_list	A list contains training data and disease status outcomes from the discovery data of corresponding gene signatures.
dat_test_sig	A data frame contains corresponding gene sets from the input.

Value

The predicted score for each sample in the test study.

LOOAUC_simple_multiple_noplot_one_df

Perform Leave-one-out CV with Logistic Regression.

Description

Perform Leave-one-out CV with Logistic Regression.

Usage

```
LOOAUC_simple_multiple_noplot_one_df(df, targetVec)
```

Arguments

df	a data.frame of gene expression count data. Required.
targetVec	a binary vector of the response variable. Should be the same number of samples as in df. Required.

Value

A list of length 3 with elements

auc	The AUC from the LOOCV procedure.
byClass	A vector containing the sensitivity, specificity, positive predictive value, negative predictive value, precision, recall, F1, prevalence, detection rate, detection prevalence and balanced accuracy.
prob	A vector of the test prediction probabilities.

mkAssay

Add SummarizedExperiment assays to the data structure.

Description

Given a SummarizedExperiment input with a counts or CPM assay, this function creates additional assays for by computing the CPM, log, or both of the input assay to be used in further analysis.

Usage

```
mkAssay(
  SE_obj,
  input_name = "counts",
  output_name = NULL,
  log = FALSE,
  counts_to_CPM = TRUE,
  prior_counts = 3
)
```

Arguments

SE_obj	a SummarizedExperiment object containing count or CPM data. Required.
input_name	a character string specifying the name of the assay to be referenced for creating additional assays. Default is "counts".
output_name	a character string to use in place of the input_name. If NULL, then input_name will be substituted. Default is NULL. See Return details for how names are altered.
log	logical. Indicate whether an assay returned should be the log of whichever assay is specified in "output_name". If counts_to_CPM = TRUE as well, then a log CPM assay will also be created. Default is FALSE.
counts_to_CPM	logical. This argument only applies if the input_type is a counts assay. If TRUE, then the output assays will include a normalized CPM assay. If log = TRUE as well, then a log CPM assay will also be created. Default is TRUE.
prior_counts	a small integer specifying the average count to be added to each observation to avoid taking the log of zero. Used only if log = TRUE. The default is 3.

Value

This function returns a SummarizedExperiment object with up to 3 additional assay types attached to the original inputted object.

output_name_cpm
: Counts per million

log_output_name_cpm
: Log counts per million

log_output_name
: Log of original input assay.

Author(s)

Aubrey Odom-Mabey

Examples

```
# Create a log assay of the original assay input
# TB_hiv dataset already has counts data
log_only <- mkAssay(TB_hiv, log = TRUE, counts_to_CPM = FALSE)
log_only

# Create a CPM assay
CPM_only <- mkAssay(TB_hiv)
CPM_only

# Create a logCPM, logcounts, and CPM assay
all_assays <- mkAssay(TB_hiv, log = TRUE)
all_assays
```

`ObtainSampleScore_OriginalModel` *Obtain training data, testing data, and train signature's original model.*

Description

Obtain training data, testing data, and train signature's original model.

Usage

```
ObtainSampleScore_OriginalModel(
  theObject_train,
  useAssay,
  gene_set,
  input,
  SigName,
  obtainDiagnosis,
  annotationColName,
  FUN,
  adj
)
```

Arguments

`theObject_train`	A SummarizedExperiment object that has been pre-stored in the data file: OriginalTrainingData.
`useAssay`	A character string or an integer specifying the assay in the `input`. Used for the test SummarizedExperiment object. Default is 1, indicating the first assay in the test SummarizedExperiment object.
`gene_set`	A character vector that includes gene symbols for selected gene signature.
`input`	A SummarizedExperiment object with gene symbols as the assay row names.
`SigName`	Optional. A character string that indicates the name for `gene_set`. `SigName` is used to provide information when gene signatures were missing in the test data.
`obtainDiagnosis`	Boolean. Used to create training data if TRUE. Default is FALSE
`annotationColName`	A character string specifying the column name of disease status. Only used when creating training data. Default is NULL.
`FUN`	A character string specifying the function name of the corresponding signature's original model.
`adj`	A small real number used in combat to solve for genes with 0 counts in rare cases. Not required for most of cases.

Value

The predicted score for each sample in the test study using corresponding gene signature's original model.

OriginalTrainingData *Discovery datasets for corresponding gene signatures.*

Description

Discovery datasets for corresponding gene signatures.

Usage

```
OriginalTrainingData
```

Format

list

Source

See `?TBsignatures` for reference information.

Examples

```
data("OriginalTrainingData")
```

plotQuantitative *Create a boxplot using logistic regression and bootstrap LOOCV to evaluate signatures.*

Description

This function takes as input a `data.frame` with genetic expression count data, and uses a bootstrapped leave-one-out cross validation procedure with logistic regression to allow for numeric and graphical comparison across any number of genetic signatures. It creates a boxplot of bootstrapped AUC values.

Usage

```
plotQuantitative(
  df.input,
  targetVec.num,
  signature.list = NULL,
  signature.name.vec = NULL,
  num.boot = 100,
  pb.show = TRUE,
  name = "Signature Evaluation: Bootstrapped AUCs",
  fill.col = "white",
  outline.col = "black",
  abline.col = "red",
  rotateLabels = FALSE
)
```

Arguments

df.input	a data.frame of gene expression count data. Required.
targetVec.num	a numeric binary vector of the response variable. The vector should be the same number of rows as df. Required.
signature.list	a list of signatures to run with their associated genes. This list should be in the same format as TBsignatures, included in the TBSignatureProfiler package. If signature.list = NULL, the default set of signatures TBsignatures list is used. For details, run ?TBsignatures.
signature.name.vec	A vector specifying the names of the signatures to be compared. This should be the same length as signature.list. If signature.name.vec = NULL, the default set of signatures TBsignatures list is used.
num.boot	an integer specifying the number of bootstrap iterations.
pb.show	logical. If TRUE then a progress bar for the bootstrapping procedure will be displayed as output. The default is TRUE.
name	a character string giving a name for the outputted boxplot of bootstrapped AUCs. The default is "Signature Evaluation: Bootstrapped AUCs".
fill.col	the color to be used to fill the boxplots. The default is "white".
outline.col	the color to be used for the boxplot outlines. The default is "black".
abline.col	the color to be used for the dotted line at AUC = 0.5 (the chance line). The default is "red".
rotateLabels	logical. If TRUE, the x-axis labels will be rotated. The default is FALSE.

Value

a boxplot comparing the bootstrapped AUCs of inputted signatures

Examples

```
inputTest <- matrix(rnorm(1000), 100, 20,
                    dimnames = list(paste0("gene", seq.int(1, 100)),
                                    paste0("sample", seq.int(1, 20))))
inputTest <- as.data.frame(inputTest)
targetVec <- sample(c(0,1), replace = TRUE, size = 20)
signature.list <- list(sig1 = c("gene1", "gene2", "gene3"),
                       sig2 = c("gene4", "gene5", "gene6"))
signature.name.vec <- c("sig1", "sig2")
num.boot <- 5
plotQuantitative(inputTest, targetVec.num = targetVec,
                 signature.list = signature.list,
                 signature.name.vec = signature.name.vec,
                 num.boot = num.boot, rotateLabels = FALSE)
```

`randomForest_OriginalModel` *Train original model for gene signatures Maertzdorf_4, Maertzdorf_15, Verhagen_10, and LauxdaCosta_OD_3.*

Description

Train original model for gene signatures Maertzdorf_4, Maertzdorf_15, Verhagen_10, and LauxdaCosta_OD_3.

Usage

```
randomForest_OriginalModel(dat_list, dat_test_sig)
```

Arguments

`dat_list`	A `list` contains training data and disease status outcomes from the discovery data of corresponding gene signatures.
`dat_test_sig`	A `data frame` contains corresponding gene sets from the `input`.

Value

The predicted score for each sample in the test study.

`ref_combat_impute` *A function for reference batch correction and imputation.*

Description

A function used to perform reference batch correction and imputation in the testing data for gene signatures that require retraining of the model. We used the k-nearest neighbors to impute the expression values for missing gene(s). The imputation operation is achieved using `impute.knn`. Since the computational time for the imputation step can be excessive for large number of missing genes. We made some constrains to prevent the overflow of imputation operation. The evaluation will not run if more than `geneMax`*100\ of the genes are not found for the corresponding gene signature in the input study. By default `geneMax` = 0.8, so the evaluation will not run if more than 80\ of the genes are missing when matching the input study to the reference data.

Usage

```
ref_combat_impute(
  theObject_train,
  useAssay,
  gene_set,
  input,
  SigName,
  adj,
  geneMax = 0.8
)
```

Arguments

theObject_train	A SummarizedExperiment object that has been pre-stored in the data file: OriginalTrainingData.
useAssay	A character string or an integer specifying the assay in the input. Used for the test SummarizedExperiment object. Default is 1, indicating the first assay in the test SummarizedExperiment object.
gene_set	A character vector that includes gene symbols for selected gene signature.
input	A SummarizedExperiment object with gene symbols as the assay row names.
SigName	Optional. A character string that indicates the name for gene_set. SigName is used to provide information when gene signatures were missing in the test data.
adj	A small real number used in combat to solve for genes with 0 counts in rare cases. Not required for most of cases.
geneMax	A real number between 0 and 1. This is used to detect the maximum percent missing genes allowed in the evaluated signatures. See impute.knn for details. The default value is 0.8.

Value

Gene set subset

---

runTBsigProfiler *Run TB gene signature profiling.*

Description

Using some subset of the signatures listed in TBsignatures and specified scoring algorithms, this function runs gene signature profiling on an input gene expression dataset. It allows for scores to be computed for these signatures which can be compared using various visualization tools also provided in the TBSignatureProfiler package.

Usage

```
runTBsigProfiler(
  input,
  useAssay = NULL,
  signatures = NULL,
  algorithm = c("GSVA", "ssGSEA", "ASSIGN", "PLAGE", "Zscore", "singscore"),
  combineSigAndAlgorithm = FALSE,
  assignDir = NULL,
  outputFormat = NULL,
  parallel.sz = 0,
  ASSIGNiter = 1e+05,
  ASSIGNburnin = 50000,
  ssgsea_norm = TRUE,
  update_genes = TRUE
)
```

Arguments

input	an input data object of the class SummarizedExperiment, data.frame, or matrix containing gene expression data. Required.
useAssay	a character string specifying the assay to use for signature profiling when input is a SummarizedExperiment. Required only for input data of the class SummarizedExperiment. If null, the assay used will be "counts". The default is NULL.
signatures	a list of signatures to run with their associated genes. This list should be in the same format as TBsignatures, included in the TBSignatureProfiler package. If signatures = NULL, the default set of signatures TBsignatures list is used. For details, run ?TBsignatures. If <2 genes in a signature are present in the sample, that signature will not be evaluated and will not be present in the resulting SE object. The default is NULL.
algorithm	a vector of algorithms to run, or character string if only one is desired. The default is c("GSVA", "ssGSEA", "ASSIGN","PLAGE", "Zscore", "singscore"). NOTE: ASSIGN takes a long time to run and is not recommended for efficient use.
combineSigAndAlgorithm	logical, if TRUE, output row names will be of the form _. It must be set to TRUE if the ouputFormat will be a SummarizedExperiment and length(algorithm) > 1. It will always be FALSE if only one algorithm is selected. If FALSE, there will be a column named algorithm' that lists which algorithm is used, and a column named 'pathway' that lists the signature profiled. If NULL, and one algorithm was used, the algorithm will not be listed. The default is FALSE.
assignDir	a character string naming a directory to save intermediate ASSIGN results if algorithm specifies "ASSIGN". The default is NULL, in which case intermediate results will not be saved.
outputFormat	a character string specifying the output data format. Possible values are "SummarizedExperiment", "matrix", or "data.frame". The default is to return the same type as the input object.
parallel.sz	an integer identifying the number of processors to use when running the calculations in parallel for the GSVA and ssGSEA algorithms. If parallel.sz = 0, all cores are used. The default is 0.
ASSIGNiter	an integer indicating the number of iterations to use in the MCMC for the ASSIGN algorithm. The default is 100,000.
ASSIGNburnin	an integer indicating the number of burn-in iterations to use in the MCMC for the ASSIGN algorithm. These iterations are discarded when computing the posterior means of the model parameters. The default is 50,000.
ssgsea_norm	logical, passed to GSVA::gsva(). When parameter algorithm = "ssgsea",the profiler runs the SSGSEA method from Barbie et al. (2009) normalizing the scores by the absolute difference between the minimum and the maximum, as described in their paper. When ssgsea.norm = FALSE, this last normalization step is skipped. The default is TRUE.
update_genes	logical, denotes whether gene names from signatures and the rownames of input should be checked for accuracy using HGNChelper::checkGeneSymbols(). The mapping assumes genes are from humans and will keep unmapped genes as the original input gene name. Default is TRUE.

Value

A SummarizedExperiment object, data.frame, or matrix of signature profiling results. The returned object will be of the format specified in outputFormat. If input is a SummarizedExperiment and outputFormat = "SummarizedExperiment", then the output will retain any input information stored in the input colData. In general, if outputFormat = "SummarizedExperiment" then columns in the colData will include the scores for each desired signature with samples on the rows. If input is a data.frame or matrix, then the returned object will have signatures on the rows and samples on the columns.

Source

Profiling for the Z-Score, PLAGE, GSVA, ssGSEA algorithms are all conducted with the Bioconductor GSVA package. Profiling for the singscore algorithm is conducted with the Bioconductor singscore package.

References

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112. doi: 10.1038/nature08460.

Foroutan, M. et al. (2018). Single sample scoring of molecular phenotypes. *BMC Bioinformatics*, **19**. doi: 10.1186/s12859-018-2435-4.

Lee, E. et al. (2008). Inferring pathway activity toward precise disease classification. *PLoS Comp Biol*, 4(11):e1000217. doi: 10.1371/journal.pcbi.1000217

Shen, Y. et al. (2015). ASSIGN: context-specific genomic profiling of multiple heterogeneous biological pathways. *Bioinformatics*, **31**, 1745-1753. doi: 10.1093/bioinformatics/btv031.

Subramanian, A. et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, **102**, 15545-15550. doi: 10.1073/pnas.0506580102.

Tomfohr, J. et al. (2005). Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics*, 6:225. doi: 10.1186/1471-2105-6-225

Examples

```
## Using a data.frame input/output
 # Create some toy data to test Zak_RISK_16 signature, using 5 samples with low
 # expression & five samples with high expression of the signatures genes.
df_testdata <- as.data.frame(rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                              dimnames = list(TBsignatures$Zak_RISK_16,
                              paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                              dimnames = list(paste0("gene", seq_len(100)),
                              paste0("sample", seq_len(10))))))
res <- runTBsigProfiler(input = df_testdata,
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"),
                        combineSigAndAlgorithm = FALSE,
                        parallel.sz = 1)
subset(res, res$pathway == "Zak_RISK_16")

## Using a SummarizedExperiment input/output
 # The TB_indian SummarizedExperiment data is included in the package.
GSVA_res <- runTBsigProfiler(input = TB_indian,
                              useAssay = "logcounts",
```

```
                            signatures = TBsignatures["Zak_RISK_16"],
                            algorithm = c("GSVA"),
                            combineSigAndAlgorithm = FALSE,
                            parallel.sz = 1)
GSVA_res$Zak_RISK_16
```

`sigAnnotData` *Annotation information for published TB signatures.*

Description

A data.frame of annotation information for published tuberculosis signatures. Currently, this table includes two variables, `disease` and `tissue type`.

Usage

```
sigAnnotData
```

Format

data.frame

Details

The `disease` variable indicates whether the signature was developed to distinguish TB from LTBI ("Disease"), TB from some combination of other diseases and possibly LTBI ("OD"), TB from Human Immunodeficiency Virus ("HIV"), TB from pneumonia ("PNA"), or identify risk of progression to TB ("RISK"), risk of TB treatment failure ("FAIL"), or classify treatment responses (i.e., failures from cures, "RES").

The `tissue type` variable denotes whether the signature was developed using samples of either whole blood/paxgene or peripheral blood mononuclear cells (PBMCs). Due to the manipulation of cells inherently required to obtain PBMCs, many scientists prefer to use only whole blood samples for analysis.

Source

See `?TBsignatures` for reference information.

Examples

```
data("sigAnnotData")
```

signatureBoxplot *Plot a boxplot of signature genes.*

Description

Plot a boxplot of signature genes.

Usage

```
signatureBoxplot(
  inputData,
  annotationData,
  signatureColNames,
  annotationColName,
  name = "Signatures",
  scale = FALSE,
  violinPlot = FALSE,
  includePoints = TRUE,
  notch = FALSE,
  rotateLabels = FALSE,
  nrow = NULL,
  ncol = NULL,
  fill_colors = NULL
)
```

Arguments

inputData	an input data object. It should either be of the class SummarizedExperiment and contain the profiled signature data and annotation data as columns in the colData, or alternatively be of the classes data.frame or matrix and contain only the gene expression data. Required.
annotationData	a data.frame or matrix of annotation data, with one column. Only required if inputData is a data.frame or matrix of signature data.
signatureColNames	a vector of the column names in colData that contain the signature data. Only required if inputData is a SummarizedExperiment object.
annotationColName	a character string naming the column name in the colData that contains the annotation data to be used in making the boxplot. Only required if inputData is a SummarizedExperiment object.
name	a character string giving the title of the boxplot. The default is "Signatures".
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
violinPlot	logical. Setting violinPlot = TRUE creates violin plots in place of boxplots. The default is FALSE.
includePoints	logical. If TRUE, points will be included over the boxplots. The default is TRUE.
notch	logical. Notches are used to compare groups; if the notches of two boxes do not overlap, this suggests that the medians are significantly different. If TRUE, the boxplot will be notched. The default is FALSE.

rotateLabels	logical. If TRUE, the x-axis labels will be rotated. The default is FALSE.
nrow	integer giving the number of rows in the resulting array.
ncol	integer giving the number of columns in the resulting array.
fill_colors	a vector of color names to be used as the fill colors for the boxplot. If NULL, colors will be supplied via RColorBrewer. The default is fill_colors = NULL.

Value

A ggplot2 boxplot of the signature data using the provided annotation information.

Examples

```
library(SummarizedExperiment)

# Generate some artificial data that shows a difference in Zak_RISK_16
mat_testdata <- rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                             dimnames = list(TBsignatures$Zak_RISK_16,
                                             paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                             dimnames = list(paste0("gene", seq_len(100)),
                                             paste0("sample", seq_len(10)))))

# Create a SummarizedExperiment object that contains the data
testdataSE <- SummarizedExperiment(assays = SimpleList(data = mat_testdata),
                                   colData = DataFrame(sample =
                                                         c(rep("down", 5),
                                                           rep("up", 5))))

# Run profiler using GSVA and ssGSEA on Zak_RISK_16 signature
res <- runTBsigProfiler(testdataSE, useAssay = "data",
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"), parallel.sz = 1,
                        combineSigAndAlgorithm = TRUE)
signatureBoxplot(res, signatureColNames = c("GSVA_Zak_RISK_16",
                                            "ssGSEA_Zak_RISK_16"),
                 annotationColName = "sample", name = "Zak_RISK_16 Signature")
```

---

signatureGeneHeatmap *Plot a heatmap of a single signature score with individual gene expression levels.*

Description

This function takes the profiled gene expression data for a single signature and creates a heatmap based on the expression scores.

Usage

```
signatureGeneHeatmap(
  inputData,
  useAssay,
```

```
  sigGenes,
  name = "Signature",
  signatureColNames = NULL,
  annotationColNames = NULL,
  scale = TRUE,
  showColumnNames = TRUE,
  showRowNames = TRUE,
  colList = list(),
  colorSets = c("Set1", "Set2", "Set3", "Pastel1", "Pastel2", "Accent", "Dark2",
    "Paired"),
  choose_color = c("blue", "gray95", "red"),
  column_order = NULL,
  ...
)
```

Arguments

inputData	a SummarizedExperiment object containing the profiled signature data and annotation data as columns in the colData. Required.
useAssay	a character string specifying the assay to use for the gene expression data. Required.
sigGenes	a vector identifying the genes in the signature to use in the heatmap. For inbuilt signatures, you can use TBsignatures (e.g., TBsignatures[["ACS_COR"]]). Required.
name	a character string with the plot title of the heatmap. The default is "Signatures".
signatureColNames	a vector of the column names in the colData that contain the signature data. Required.
annotationColNames	a vector of the column names in the colData that contain the annotation data. If NULL, no annotation bar besides those of the scoring algorithms will be drawn on the heatmap. The default is NULL.
scale	logical. Setting scale = TRUE scales the signature data. The default is TRUE.
showColumnNames	logical. Setting showColumnNames = TRUE will show the column names (i.e. sample names) on the heatmap. The default is TRUE.
showRowNames	logical. Setting showColumnNames = TRUE will show the row names (i.e. signature names) on the heatmap. The default is TRUE.
colList	a named list of named vectors specifying custom color information to pass to ComplexHeatmap::Heatmap(). The list should have as many elements as there are annotation columns and gene signatures (i.e. sigGenes), and each element name should correspond exactly with the name of each annotation column/signature. The colors in the vector elements should be named according to the levels of the factor in that column's annotation data if the annotation is discrete, or it should be produced with circlize::colorRamp2 if the annotation/gene is continuous. By default, ColorBrewer color sets will be used. See the the parameter colorSets for additional details.
colorSets	a vector of names listing the color sets in the order that they should be used in creating the heatmap. By default, this function will use the color sets in the order listed in Usage for annotation information. You may replace the default

with the same collection of sets in order that you want to use them, or provide custom color sets with the colList parameter.

choose_color — a vector of color names to be interpolated for the heatmap gradient, or a colorRamp function produced by circlize::colorRamp2. The default is c("blue", "gray95", "red").

column_order — a vector of character strings indicating the order in which to manually arrange the heatmap columns. Default is NULL, such that column order is automatically determined via clustering.

... — Additional arguments to be passed to ComplexHeatmap::Heatmap().

Value

A ComplexHeatmap plot.

Examples

```
library(SummarizedExperiment)
# Generate some artificial data that shows a difference in Zak_RISK_16
mat_testdata <- rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                             dimnames = list(TBsignatures$Zak_RISK_16,
                                             paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                             dimnames = list(paste0("gene", seq_len(100)),
                                             paste0("sample", seq_len(10)))))

# Create a SummarizedExperiment object that contains the data
testdataSE <- SummarizedExperiment(assays = SimpleList(data = mat_testdata),
                                   colData = DataFrame(sample =
                                                         c(rep("down", 5),
                                                           rep("up", 5))))
# Run profiler using GSVA and ssGSEA on Zak_RISK_16
res <- runTBsigProfiler(testdataSE, useAssay = "data",
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"), parallel.sz = 1,
                        combineSigAndAlgorithm = TRUE)

# Plot a heatmap of signature genes and pathway predictions
signatureGeneHeatmap(res, useAssay = "data",
                     sigGenes = TBsignatures[["Zak_RISK_16"]],
                     signatureColNames = c("GSVA_Zak_RISK_16",
                                           "ssGSEA_Zak_RISK_16"),
                     annotationColNames = c("sample"), showColumnNames = FALSE,
                     name = "Zak_RISK_16")
```

---

signatureHeatmap *Plot a heatmap of signature scores.*

Description

This function takes a dataset of scored gene expression data as an input and returns a ComplexHeatmap plot for for visual comparison of signature performance. The function takes arguments listed here as well as any others to be passed on to ComplexHeatmap::Heatmap().

Usage

```
signatureHeatmap(
  inputData,
  annotationData = NULL,
  name = "Signatures",
  signatureColNames,
  annotationColNames = NULL,
  colList = list(),
  scale = FALSE,
  showColumnNames = TRUE,
  showRowNames = TRUE,
  colorSets = c("Set1", "Set2", "Set3", "Pastel1", "Pastel2", "Accent", "Dark2",
    "Paired"),
  choose_color = c("blue", "gray95", "red"),
  split_heatmap = "none",
  annotationSignature = sigAnnotData,
  column_order = NULL,
  cluster_columns = TRUE,
  ...
)
```

Arguments

inputData	an input data object. It should either be of the class SummarizedExperiment and contain the profiled signature data and annotation data as columns in the colData, or alternatively be of the classes data.frame or matrix and contain only the gene expression data. Required.
annotationData	a data.frame or matrix of annotation data, with one column. Only required if inputData is a data.frame or matrix of signature data. The row names must equal those of the inputData column names. Default is NULL.
name	a character string with the plot title of the heatmap. The default is "Signatures".
signatureColNames	a vector of the column names in colData that contain the signature data. Only required if inputData is a SummarizedExperiment object.
annotationColNames	a vector of the column names in colData that contain the annotation data. Only required if inputData is a SummarizedExperiment. Default is NULL.
colList	a named list of named vectors specifying custom color information to pass to ComplexHeatmap::Heatmap(). The list should have as many elements as there are annotation columns, and each element name should correspond exactly with the name of each annotation column. The colors in the vector elements should be named according to the levels of the factor in that column's annotation data if the annotation is discrete, or it should be produced with circlize::colorRamp2 if the annotation is continuous. By default, ColorBrewer color sets will be used. See the the parameter colorSets for additional details.
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
showColumnNames	logical. Setting showColumnNames = TRUE will show the column names (i.e. sample names) on the heatmap. The default is TRUE.
showRowNames	logical. Setting showColumnNames = TRUE will show the row names (i.e. signature names) on the heatmap. The default is TRUE.

colorSets — a vector of names listing the color sets in the order that they should be used in creating the heatmap. By default, this function will use the color sets in the order listed in Usage for annotation information. You may replace the default with the same collection of sets in order that you want to use them, or provide custom color sets with the colList parameter.

choose_color — a vector of color names to be interpolated for the heatmap gradient, or a colorRamp function produced by circlize::colorRamp2. The default is c("blue", "gray95", "red").

split_heatmap — a character string either giving the column title of annotationSignature containing annotation data for which to split the heatmap rows (i.e., signatures), or "none" if no split is desired. To split based on the type of signature, set split_heatmap = "disease". The default is "none".

annotationSignature — a data.frame or matrix with information to be used in splitting the heatmap. The first column should signature names. The column of annotation information should be specified in split_heatmap. Other columns will be ignored. The default is sigAnnotData.

column_order — a vector of character strings indicating the order in which to manually arrange the heatmap columns. Default is NULL, such that column order is automatically determined via clustering.

cluster_columns — A logical indicating whether columns (samples) should be clustered together. Must be FALSE if user supplies column_order. Default is TRUE.

... — Additional arguments to be passed to ComplexHeatmap::Heatmap().

Details

If both annotationData = NULL and annotationColNames = NULL, no annotation bar will be drawn on the heatmap.

Value

A ComplexHeatmap plot.

Examples

```
library(SummarizedExperiment)
# Generate some artificial data that shows a difference in Zak_RISK_16
mat_testdata <- rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                             dimnames = list(TBsignatures$Zak_RISK_16,
                                             paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                             dimnames = list(paste0("gene", seq_len(100)),
                                             paste0("sample", seq_len(10)))))
# Create a SummarizedExperiment object that contains the data
testdataSE <- SummarizedExperiment(assays = SimpleList(data = mat_testdata),
                                   colData = DataFrame(sample =
                                                         c(rep("down", 5),
                                                           rep("up", 5))))
res <- runTBsigProfiler(testdataSE, useAssay = "data",
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"), parallel.sz = 1,
                        combineSigAndAlgorithm = TRUE)
```

```
signatureHeatmap(res, signatureColNames = c("GSVA_Zak_RISK_16",
                                            "ssGSEA_Zak_RISK_16"),
                 annotationColNames = "sample", scale = TRUE,
                 showColumnNames = FALSE, split_heatmap = "none")

# Example using custom colors for the annotation information
color2 <- stats::setNames(c("purple", "black"), c("down", "up"))
color.list <- list("sample" = color2)

signatureHeatmap(res, signatureColNames = c("GSVA_Zak_RISK_16",
                                            "ssGSEA_Zak_RISK_16"),
                 annotationColNames = "sample", scale = TRUE,
                 showColumnNames = FALSE,
                 colList = color.list, split_heatmap = "none")
```

`SignatureQuantitative` *Use logistic regression and bootstrap LOOCV to evaluate signatures.*

Description

This function takes as input a `data.frame` with genetic expression count data, and uses a bootstrapped leave-one-out cross validation procedure with logistic regression to allow for numeric and graphical comparison across any number of genetic signatures.

Usage

```
SignatureQuantitative(
  df.input,
  targetVec.num,
  signature.list = NULL,
  signature.name.vec = NULL,
  num.boot = 100,
  pb.show = TRUE
)
```

Arguments

`df.input`	a `data.frame` of gene expression count data. Required.
`targetVec.num`	a numeric binary vector of the response variable. The vector should be the same number of rows as `df`. Required.
`signature.list`	a `list` of signatures to run with their associated genes. This list should be in the same format as `TBsignatures`, included in the TBSignatureProfiler package. If `signature.list = NULL`, the default set of signatures `TBsignatures` list is used. For details, run `?TBsignatures`.
`signature.name.vec`	A vector specifying the names of the signatures to be compared. This should be the same length as `signature.list`. If `signature.name.vec = NULL`, the default set of signatures `TBsignatures` list is used.
`num.boot`	an integer specifying the number of bootstrap iterations.

pb.show — logical. If TRUE then a progress bar for the bootstrapping procedure will be displayed as output. The default is TRUE.

name — a character string giving a name for the outputted boxplot of bootstrapped AUCs. The default is "Quantitative Evaluation of Signatures via Bootstrapped AUCs".

Value

the AUC, sensitivity and specificity

Examples

```
inputTest <- matrix(rnorm(1000), 100, 20,
                    dimnames = list(paste0("gene", seq.int(1, 100)),
                                    paste0("sample", seq.int(1, 20))))
inputTest <- as.data.frame(inputTest)
targetVec <- sample(c(0,1), replace = TRUE, size = 20)
signature.list <- list(sig1 = c("gene1", "gene2", "gene3"),
                       sig2 = c("gene4", "gene5", "gene6"))
signature.name.vec <- c("sig1", "sig2")
num.boot <- 2
SignatureQuantitative(inputTest, targetVec.num = targetVec,
                      signature.list = signature.list,
                      signature.name.vec = signature.name.vec,
                      num.boot = num.boot)
```

---

signatureROCplot *Create an array of ROC plots to compare signatures.*

Description

Create an array of ROC plots to compare signatures.

Usage

```
signatureROCplot(
  inputData,
  annotationData,
  signatureColNames,
  annotationColName,
  scale = FALSE,
  choose_colors = c("cornflowerblue", "gray24"),
  name = "Signatures",
  nrow = NULL,
  ncol = NULL
)
```

Arguments

inputData	an input data object. It should either be of the class SummarizedExperiment and contain the profiled signature data and annotation data as columns in the colData, or alternatively be of the classes data.frame or matrix and contain only the gene expression data. Required.
annotationData	a data.frame or matrix of annotation data, with one column. Only required if inputData is a data.frame or matrix of signature data.
signatureColNames	a vector of the column names of inputData that contain the signature data. If inputData is a SummarizedExperiment object, these are the column names of the object colData.
annotationColName	a character string naming the column name in the colData that contains the annotation data to be used in making the boxplot. Only required if inputData is a SummarizedExperiment object.
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
choose_colors	a vector of length 2 defining the colors to be used in the ROC plots. The default is c("cornflowerblue", "gray24").
name	a character string giving the title of the boxplot. The default is "Signatures".
nrow	integer giving the number of rows in the resulting array.
ncol	integer giving the number of columns in the resulting array.

Value

An array of ROC plots.

Examples

```
# Run signature profiling
 choose_sigs <- subset(TBsignatures,
                       !(names(TBsignatures) %in% c("Lee_4", "Roe_OD_4")))[c(1,2)]
 prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                 algorithm = "ssGSEA",
                                 signatures = choose_sigs,
                                 parallel.sz = 1)

# Create ROC plots
signatureROCplot(prof_indian, signatureColNames = names(choose_sigs),
                 annotationColName = "label")
```

---

signatureROCplot_CI *Create an array of ROC plots with confidence interval bands to compare signatures.*

Description

Create an array of ROC plots with confidence interval bands to compare signatures.

Usage

```
signatureROCplot_CI(
  inputData,
  annotationData,
  signatureColNames,
  annotationColName,
  scale = FALSE,
  choose_colors = c("cornflowerblue", "gray50", "gray79"),
  name = NULL,
  nrow = NULL,
  ncol = NULL,
  ci.lev = 0.95,
  pb.show = TRUE
)
```

Arguments

inputData	an input data object. It should either be of the class SummarizedExperiment and contain the profiled signature data and annotation data as columns in the colData, or alternatively be of the classes data.frame or matrix and contain only the gene expression data. Required.
annotationData	a data.frame or matrix of annotation data, with one column. Only required if inputData is a data.frame or matrix of signature data.
signatureColNames	a vector of the column names of inputData that contain the signature data. If inputData is a SummarizedExperiment object, these are the column names of the object colData.
annotationColName	a character string naming the column name in the colData that contains the annotation data to be used in making the boxplot. Only required if inputData is a SummarizedExperiment object.
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
choose_colors	a vector of length 3 defining the colors to be used in the ROC plots. The default is c("cornflowerblue","gray50", "gray79").
name	a character string giving the title of the ROC plot. If NULL, the plot title will be "ROC Plots for Gene Signatures, <ci.lev>% Confidence". The default is NULL.
nrow	integer giving the number of rows in the resulting array.
ncol	integer giving the number of columns in the resulting array.
ci.lev	a number between 0 and 1 giving the desired level of confidence for computing ROC curve estimations.
pb.show	logical for whether to show a progress bar while running code. The default is TRUE.

Value

An array of ROC plots.

Examples

```
# Run signature profiling

 choose_sigs <- TBsignatures[c(1, 2)]
 prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                 algorithm = "Zscore",
                                 signatures = choose_sigs,
                                 parallel.sz = 1)

# Create ROC plots with confidence intervals
signatureROCplot_CI(prof_indian, signatureColNames = names(choose_sigs),
                    annotationColName = "label")
```

`subsetGeneSet` *Filter gene expression value matrix based on certain gene sets.*

Description

A function used to subset gene expression value matrix based on certain gene sets.

Usage

```
subsetGeneSet(
  theObject,
  gene_set,
  useAssay,
  obtainDiagnosis = FALSE,
  annotationColName = NULL
)
```

Arguments

`theObject`	A SummarizedExperiment object that has been pre-stored in OriginalTrainingData.RDA
`gene_set`	A character vector that includes gene symbols for gene signatures.
`useAssay`	A character string or an integer specifying the assay in the `theObject` that will be selected.
`obtainDiagnosis`	Boolean. Usually used to create training data if TRUE. Default is FALSE
`annotationColName`	A character string specifying the column name of disease status. Only used when creating training data. Default is NULL.

Value

A `matrix` with selected gene expression value if `obtainDiagnosis == FALSE`. If `obtainDiagnosis == TRUE`, return a `list` contains the selected gene expression value and diagnosis results for each sample.

SulimanOriginalModel *Train original model gene signature Suliman_RISK_4.*

Description

Train original model gene signature Suliman_RISK_4.

Usage

```
SulimanOriginalModel(dat_list, dat_test_sig)
```

Arguments

dat_list	A list contains training data and disease status outcomes from the discovery data of corresponding gene signatures.
dat_test_sig	A data frame contains corresponding gene sets from the input.

Value

The predicted score for each sample in the test study.

svm_OriginalModel *Train original model for gene signatures Bloom_OD_144 and Zak_RISK_16.*

Description

Train original model for gene signatures Bloom_OD_144 and Zak_RISK_16.

Usage

```
svm_OriginalModel(dat_list, dat_test_sig)
```

Arguments

dat_list	A list contains training data and disease status outcomes from the discovery data of corresponding gene signatures.
dat_test_sig	A data frame contains corresponding gene sets from the input.

Value

The predicted score for each sample in the test study.

tableAUC *Create a table of results for t-tests and bootstrapped AUCs for multiple scored signatures.*

Description

This function collects the results of bootstrapping and t-tests for a scored gene expression dataset and presents them using a JavaScript table with an R interface, or as a data.frame.

Usage

```
tableAUC(
  SE_scored,
  annotationColName,
  signatureColNames,
  num.boot = 100,
  pb.show = TRUE,
  output = "DataTable"
)
```

Arguments

SE_scored	a SummarizedExperiment object with genes as the row features and signature scores in the colData. There should also be a column of annotation data. Required.
annotationColName	a character string giving the column name in colData that contains the annotation data. Required.
signatureColNames	a vector of column names in the colData that contain the signature score data. Required.
num.boot	integer. The number of times to bootstrap the data. The default is 100.
pb.show	logical for whether to show a progress bar while running code. The default is TRUE.
output	a character string indicating the table output format. Possible values are DataTable and data.frame. The default is DataTable.

Value

A JavaScript table with an R interface using the DT package.

Examples

```
# Run signature profiling
 choose_sigs <- TBsignatures[c(1, 2)]
 prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                 algorithm = "ssGSEA",
                                 signatures = choose_sigs,
                                 parallel.sz = 1,
                                 update_genes = FALSE)
 # Create table
```

```
 tableAUC(SE_scored = prof_indian, annotationColName = "label",
          signatureColNames = names(choose_sigs))

 # Create data.frame object
h <-  tableAUC(SE_scored = prof_indian, annotationColName = "label",
               signatureColNames = names(choose_sigs),
               output = "data.frame",
               num.boot = 5)
head(h)
```

TBcommon *A list of published TB signatures, using author-given names.*

Description

A set of Tuberculosis gene signatures from various publications. This set of signatures uses gene symbols. Attempts have been made to use updated gene symbols and remove symbols that did not match the most recent annotation. Additional sets for Entrez IDs and Ensembl IDs are forthcoming.

Usage

```
TBcommon
```

Format

list

Details

This list differs from `TBsignatures` in that signatures with names specified in their originating publication (or that of a peer) are given that common name rather than using the `TBSignatureProfiler` naming system. Otherwise, signature names are composed of the last name of the primary author, followed by a possible context for the signature, and ending with either the number of gene transcripts or genes in the signature with respect to however it was described in the original publication.

Possible signature contexts:

- <blank>: TB vs LTBI or Healthy Controls
- OD: Other diseases
- HIV: Human Immunodeficiency Virus
- PNA: Pneumonia
- RISK: Risk of developing active TB
- RES: Response to TB treatment
- FAIL: Failure of TB treatment

Note that in some cases signatures will be positive identifiers of TB whereas others are negative identifiers; this should be taken into account when creating ROC curves and computing any AUC estimates.

Source

- **Anderson_42**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. https://dx.doi.org/10.1056/NEJMoa130365710.1056/NEJMoa1303657
- **Anderson_OD_51**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. https://dx.doi.org/10.1056/NEJMoa130365710.1056/NEJMoa1303657
- **Berry_393**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. https://dx.doi.org/10.1038/nature0924710.1038/nature09247
- **Berry_OD_86**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. https://dx.doi.org/10.1038/nature0924710.1038/nature09247
- **Blankley_380**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. https://dx.doi.org/10.1183/13993003.02121-201510.1183/13993003.02121-2015
- **Blankley_5**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. https://dx.doi.org/10.1183/13993003.02121-201510.1183/13993003.02121-2015
- **Bloom_OD_144**: Bloom, Chloe I., Christine M. Graham, Matthew P. R. Berry, Fotini Rozakeas, Paul S. Redford, Yuanyuan Wang, Zhaohui Xu, et al. 2013. "Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers." PloS One 8 (8): e70630. https://dx.doi.org/10.1371/journal.pone.007063010.1371/journal.pone.0070630
- **Bloom_RES_268**: Bloom CI, Graham CM, Berry MP, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One. 2012;7(10):e46191. https://dx.doi.org/10.3389/fmicb.2021.65056710.3389/fmicb.2021.650567
- **Bloom_RES_558**: Bloom CI, Graham CM, Berry MP, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One. 2012;7(10):e46191. https://dx.doi.org/10.3389/fmicb.2021.65056710.3389/fmicb.2021.650567
- **Chen_5**: Chen L, Hua J, He X. Coexpression Network Analysis-Based Identification of Critical Genes Differentiating between Latent and Active Tuberculosis. Dis Markers. 2022;2022:2090560. http://dx.doi.org/10.1155/2022/209056010.1155/2022/2090560
- **Chen_HIV_4**: Chen Y, Wang Q, Lin S, et al. Meta-Analysis of Peripheral Blood Transcriptome Datasets Reveals a Biomarker Panel for Tuberculosis in Patients Infected With HIV. Front Cell Infect Microbiol. 2021;11:585919. Published 2021 Mar 19. https://dx.doi.org/10.3389/fcimb.2021.58591910.3389/fcimb.2021.585919

- **Chendi_HIV_2**: Chendi BH, Tveiten H, Snyders CI, et al. CCL1 and IL-2Ra differentiate Tuberculosis disease from latent infection Irrespective of HIV infection in low TB burden countries. J Infect. 2021;S0163-4453(21)00379-0. https://dx.doi.org/10.1016/j.jinf.2021.07.03610.1016/j.jinf.2021.07.036
- **RISK11**: Darboe, F. et al. Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells. Tuberculosis 108, 124-126 (2018). https://dx.doi.org/10.1016/j.tube.2017.11.00110.1016/j.tube.2017.11.001
- **Dawany_HIV_251**: Dawany, N. et al. Identification of a 251 gene expression signature that can accurately detect M. tuberculosis in patients with and without HIV co-infection. PLoS One 9, (2014). https://dx.doi.org/10.1371/journal.pone.008992510.1371/journal.pone.0089925
- **CMTB_CT**: Duffy FJ, Olson GS, Gold ES, Jahn A, Aderem A, Aitchison J, Rothchild AC, Diercks AH, Nemeth J. A contained Mycobacterium tuberculosis mouse infection model predicts active disease and containment in humans. The Journal of Infectious Diseases. 2021 Mar 10. https://dx.doi.org/10.1093/infdis/jiab13010.1093/infdis/jiab130
- **Esmail_203**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. https://dx.doi.org/10.1073/pnas.171185311510.1073/pnas.1711853115
- **Esmail_82**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. https://dx.doi.org/10.1073/pnas.171185311510.1073/pnas.1711853115
- **Esmail_OD_893**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. https://dx.doi.org/10.1073/pnas.171185311510.1073/pnas.1711853115
- **Estevez_133**: Estévez O, Anibarro L, Garet E, et al. An RNA-seq Based Machine Learning Approach Identifies Latent Tuberculosis Patients With an Active Tuberculosis Profile. Front Immunol. 2020;11:1470. Published 2020 Jul 14. https://dx.doi.org/10.3389/fimmu.2020.0147010.3389/fimmu.2020.01470
- **Estevez_259**: Estévez O, Anibarro L, Garet E, et al. An RNA-seq Based Machine Learning Approach Identifies Latent Tuberculosis Patients With an Active Tuberculosis Profile. Front Immunol. 2020;11:1470. Published 2020 Jul 14. https://dx.doi.org/10.3389/fimmu.2020.0147010.3389/fimmu.2020.01470
- **Francisco_OD_2**: Francisco NM, Fang YM, Ding L, et al. Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. J Infect. 2017;75(6):499-510. http://dx.doi.org/10.1016/j.jinf.2017.09.01210.1016/j.jinf.2017.09.012
- **Gjoen_10**: Gjøen, J.E., Jenum, S., Sivakumaran, D. et al. 'Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children.' Sci Rep 7, 5839 (2017). https://doi.org/10.1038/s41598-017-05057-x10.1038/s41598-017-05057-x

- **Gjoen_7**: Gjøen, J.E., Jenum, S., Sivakumaran, D. et al. 'Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children.' Sci Rep 7, 5839 (2017). https://doi.org/10.1038/s41598-017-05057-x10.1038/s41598-017-05057-x
- **Gliddon_2_OD_4**: Gliddon HD, Kaforou M, Alikian M, et al. Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. Front Immunol. 2021;12:637164. Published 2021 Mar 2. https://dx.doi.org/10.3389/fimmu.2021.63716410.3389/fimmu.2021.637164
- **Gliddon_HIV_3**: Gliddon HD, Kaforou M, Alikian M, et al. Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. Front Immunol. 2021;12:637164. Published 2021 Mar 2. https://dx.doi.org/10.3389/fimmu.2021.63716410.3389/fimmu.2021.637164
- **Gliddon_OD_3**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org: . https://dx.doi.org/10.1101/58367410.1101/583674
- **Gliddon_OD_4**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org: . https://dx.doi.org/10.1101/58367410.1101/583674
- **Gong_OD_4**: Gong Z, Gu Y, Xiong K, Niu J, Zheng R, Su B, Fan L and Xie J (2021) The Evaluation and Validation of Blood-Derived Novel Biomarkers for Precise and Rapid Diagnosis of Tuberculosis in Areas With High-TB Burden. Front. Microbiol. 12:650567. https://dx.doi.org/10.3389/fmicb.2021.65056710.3389/fmicb.2021.650567
- **Heycken_FAIL_22**: Heyckendorf J, Marwitz S, Reimann M, et al. Prediction of anti-tuberculosis treatment duration based on a 22-gene transcriptomic model. Eur Respir J. 2021;2003492. https://dx.doi.org/10.1183/13993003.03492-202010.1183/13993003.03492-2020
- **Hoang_OD_13**: Hoang, Long & Jain, Pooja & Pillay, Timesh & Tolosa-Wright, Mica & Niazi, Umar & Takwoingi, Yemisi & Halliday, Alice & Berrocal-Almanza, Luis & Deeks, Jonathan & Beverley, Peter & Kon, Onn & Lalvani, Ajit. (2021). Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. The Lancet Infectious Diseases. https://dx.doi.org/10.1016/S1473-3099(20)30928-210.1016/S1473-3099(20)30928-2
- **Hoang_OD_20**: Hoang, Long & Jain, Pooja & Pillay, Timesh & Tolosa-Wright, Mica & Niazi, Umar & Takwoingi, Yemisi & Halliday, Alice & Berrocal-Almanza, Luis & Deeks, Jonathan & Beverley, Peter & Kon, Onn & Lalvani, Ajit. (2021). Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. The Lancet Infectious Diseases. https://dx.doi.org/10.1016/S1473-3099(20)30928-210.1016/S1473-3099(20)30928-2
- **Hoang_OD_3**: Hoang, Long & Jain, Pooja & Pillay, Timesh & Tolosa-Wright, Mica & Niazi, Umar & Takwoingi, Yemisi & Halliday, Alice & Berrocal-Almanza, Luis & Deeks, Jonathan & Beverley, Peter & Kon, Onn & Lalvani, Ajit. (2021). Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. The Lancet Infectious Diseases. https://dx.doi.org/10.1016/S1473-3099(20)30928-210.1016/S1473-3099(20)30928-2

- **Huang_13**: Huang, Hai-Hui et al. 'Identification of 13 Blood-based Gene Expression Signatures to Accurately Distinguish Tuberculosis from Other Pulmonary Diseases and Healthy Controls'. 1 Jan. 2015 : S1837 - S1843.https://doi.org/10.3233/BME-15148610.3233/BME-151486
- **Jacobsen_3**: Jacobsen, Marc, Dirk Repsilber, Andrea Gutschmidt, Albert Neher, Knut Feldmann, Hans J. Mollenkopf, Andreas Ziegler, and Stefan H. E. Kaufmann. 2007. "Candidate Biomarkers for Discrimination between Infection and Disease Caused by Mycobacterium Tuberculosis." Journal of Molecular Medicine 85 (6): 613-21. https://dx.doi.org/10.1007/s00109-007-0157-610.1007/s00109-007-0157-6
- **Jenum_8**: Jenum, S., Dhanasekaran, S., Lodha, R. et al. Approaching a diagnostic point-of-care test for pediatric tuberculosis through evaluation of immune biomarkers across the clinical disease spectrum. Sci Rep 6, 18520 (2016). https://doi.org/10.1038/srep1852010.1038/srep18520
- **Kaforou_27**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. https://dx.doi.org/10.1371/journal.pmed.100153810.1371/journal.pmed.1001538
- **Kaforou_OD_44**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. https://dx.doi.org/10.1371/journal.pmed.100153810.1371/journal.pmed.1001538
- **Kaforou_OD_53**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. https://dx.doi.org/10.1371/journal.pmed.100153810.1371/journal.pmed.1001538
- **Kaul_3**: Kaul S, Nair V, Birla S, et al. Latent Tuberculosis Infection Diagnosis among Household Contacts in a High Tuberculosis-Burden Area: a Comparison between Transcript Signature and Interferon Gamma Release Assay. Microbiol Spectr. 2022;10(2):e0244521. http://dx.doi.org/10.1128/spectrum.02445-2110.1128/spectrum.02445-21
- **Kulkarni_HIV_2**: Kulkarni V, Queiroz ATL, Sangle S, et al. A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. Front Immunol. 2021;12:631165. Published 2021 Feb 22. https://dx.doi.org/10.3389/fimmu.2021.63116510.3389/fimmu.2021.631165
- **Kwan_186**: Kwan PKW, Periaswamy B, De Sessions PF, et al. A blood RNA transcript signature for TB exposure in household contacts. BMC Infect Dis. 2020;20(1):403. http://dx.doi.org/10.1186/s12879-020-05116-110.1186/s12879-020-05116-1
- **LauxdaCosta_OD_3**: Laux da Costa L, Delcroix M, Dalla Costa ER, et al. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. Tuberculosis (Edinb). 2015;95(4):421-425. https://dx.doi.org/10.1016/j.tube.2015.04.00810.1016/j.tube.2015.04.008
- **Lee_4**: Lee, Shih-Wei, Lawrence Shih-Hsin Wu, Guan-Mau Huang, Kai-Yao Huang, Tzong-Yi Lee, and Julia Tzu-Ya Weng. 2016. "Gene Expression Profiling Identifies Candidate Biomarkers for Active and Latent Tuberculosis." BMC Bioinformatics 17 Suppl 1 (January): 3. https://dx.doi.org/10.1186/s12859-015-0848-x10.1186/s12859-015-0848-x
- **Leong_24**: Leong, Samantha, Yue Zhao, Noyal M. Joseph, Natasha S. Hochberg, Sonali Sarkar, Jane Pleskunas, David Hom, et al. 2018. "Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India."

Tuberculosis (109): 41-51. https://doi.org/10.1016/j.tube.2018.01.00210.1016/j.tube.2018.01.002

- **PREDICT29**: Leong, S., Zhao, Y., Ribeiro-Rodrigues, R., Jones-López, E. C., Acuña-Villaorduña, C., Rodrigues, P. M., Palaci, M., Alland, D., Dietze, R., Ellner, J. J., Johnson, W. E., Salgame, P., Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB. Tuberculosis (Edinb). 2020 Jan;120:101898. https://dx.doi.org/10.1016/j.tube.2020.10189810.1016/j.tube.2020.101898
- **Li_3**: Li M, Qiu Y, Guo M, Qu R, Tian F, Wang G, Wang Y, Ma J, Liu S, Takiff H, Tang Y, Gao Q.2023.Evaluation of the Cepheid 3-gene host response blood test for tuberculosis diagnosis and treatment response monitoring in a primary-level clinic in rural China. J Clin Microbiol61:e00911-23. http://dx.doi.org/10.1128/jcm.00911-23
- **Long_RES_10**: Long NP, Phat NK, Yen NTH, et al. A 10-gene biosignature of tuberculosis treatment monitoring and treatment outcome prediction. Tuberculosis (Edinb). 2021;131:102138. http://dx.doi.org/10.1016/j.tube.2021.10213810.1016/j.tube.2021.102138
- **Maertzdorf_15**: Maertzdorf J, McEwen G, Weiner J 3rd, et al. Concise gene signature for point-of-care classification of tuberculosis. EMBO Mol Med. 2016;8(2):86-95. https://dx.doi.org/10.15252/emmm.20150579010.15252/emmm.201505790
- **DIAG4**: Maertzdorf, Jeroen, Gayle McEwen, January Weiner 3rd, Song Tian, Eric Lader, Ulrich Schriek, Harriet Mayanja-Kizza, Martin Ota, John Kenneth, and Stefan He Kaufmann. 2016. "Concise Gene Signature for Point-of-Care Classification of Tuberculosis." EMBO Molecular Medicine 8 (2): 86-95. https://dx.doi.org/10.15252/emmm.20150579010.15252/emmm.201505790
- **Maertzdorf_OD_100**: Maertzdorf, Jeroen, January Weiner 3rd, Hans-Joachim Mollenkopf, TBornot TB Network, Torsten Bauer, Antje Prasse, Joachim Müller-Quernheim, and Stefan H. E. Kaufmann. 2012. "Common Patterns and Disease-Related Signatures in Tuberculosis and Sarcoidosis." Proceedings of the National Academy of Sciences of the United States of America 109 (20): 7853-58. https://dx.doi.org/10.1073/pnas.112107210910.1073/pnas.1121072109
- **Natarajan_7**: Natarajan S, Ranganathan M, Hanna LE, Tripathy S. Transcriptional Profiling and Deriving a Seven-Gene Signature That Discriminates Active and Latent Tuberculosis: An Integrative Bioinformatics Approach. Genes (Basel). 2022;13(4):616. Published 2022 Mar 29. http://dx.doi.org/10.3390/genes1304061610.3390/genes13040616
- **RISK6**: Penn-Nicholson, A. et al. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. Sci. Rep. 10, (2020). https://dx.doi.org/10.1038/s41598-020-65043-810.1038/s41598-020-65043-8
- **Qian_OD_17**: Qian, Zhongqing et al. "Expression of nuclear factor, erythroid 2-like 2-mediated genes differentiates tuberculosis." Tuberculosis (Edinburgh, Scotland) vol. 99 (2016): 56-62. https://doi.org/10.1016/j.tube.2016.04.00810.1016/j.tube.2016.04.008
- **Rajan_HIV_5**: Rajan, Jayant V., Semitala, Fred C., Kamya, Moses R., Yoon, Christina., Mehta, Tejas., Cattamanchi, Adithya., Seielstad, Mark., Montalvo, Lani., Andama, Alfred., Katende, Jane., Asege, Lucy., Nakaye, Martha., Mwebe, Sandra. 2018 "A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus" Clinical Infectious Diseases: 1-7. https://doi.org/10.1093/cid/ciy83510.1093/cid/ciy835
- **Roe_3**: Roe, Jennifer, Venturini, Cristina, Gupta, Rishi K., Gurry, Celine, Chain, Benjamin M., Sun, Yuxin, Southern, Jo, Jackson, Charlotte, Lipman, Marc, C., Miller, Robert F., Martineau, Adrian R., Abubakar, Ibrahim, Noursadeghi, Mahdad. 2019 "T1 Blood transcriptomic stratification of short-term risk in contacts of tuberculosis": . https://doi.org/10.1093/cid/ciz25210.1093/cid/ciz252

- **Roe_OD_4**: Roe, Jennifer K., Niclas Thomas, Eliza Gil, Katharine Best, Evdokia Tsaliki, Stephen Morris-Jones, Sian Stafford, et al. 2016. "Blood Transcriptomic Diagnosis of Pulmonary and Extrapulmonary Tuberculosis." JCI Insight 1 (16): e87238. https://dx.doi.org/10.1172/jci.insight.8723810.1172/jci.insight.87238
- **Sambarey_HIV_10**: Sambarey, Awanti, Abhinandan Devaprasad, Abhilash Mohan, Asma Ahmed, Soumya Nayak, Soumya Swaminathan, George D'Souza, et al. 2017. "Unbiased Identification of Blood-Based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks." EBioMedicine 15 (February): 112-26. https://dx.doi.org/10.1016/j.ebiom.2016.12.00910.1016/j.ebiom.2016.12.009
- **Singhania_OD_20**: Singhania, Akul, Raman Verma, Christine M. Graham, Jo Lee, Trang Tran, Matthew Richardson, Patrick Lecine, et al. 2018. "A Modular Transcriptional Signature Identifies Phenotypic Heterogeneity of Human Tuberculosis Infection." Nature Communications 9 (1): 2308. https://dx.doi.org/10.1038/s41467-018-04579-w10.1038/s41467-018-04579-w
- **Sivakumaran_11**: Sivakumaran D, Ritz C, Gjøen JE, et al. Host Blood RNA Transcript and Protein Signatures for Sputum-Independent Diagnostics of Tuberculosis in Adults. Front Immunol. 2021;11:626049. Published 2021 Feb 4. https://dx.doi.org/10.3389/fimmu.2020.62604910.3389/fimmu.2020.626049
- **Sloot_HIV_2**: Sloot, Rosa, Maarten F. Schim van der Loeff, Erik W. van Zwet, Mariëlle C. Haks, Sytze T. Keizer, Maarten Scholing, Tom H. M. Ottenhoff, Martien W. Borgdorff, and Simone A. Joosten. 2015. "Biomarkers Can Identify Pulmonary Tuberculosis in HIV-Infected Drug Users Months Prior to Clinical Diagnosis." EBioMedicine 2 (2): 172-79. https://dx.doi.org/10.1016/j.ebiom.2014.12.00110.1016/j.ebiom.2014.12.001
- **Suliman_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. https://dx.doi.org/10.1164/rccm.201711-2340OC10.1164/rccm.201711-2340OC
- **Suliman_RISK_2**: Suliman, S. et al. Four-gene pan-African blood signature predicts progression to tuberculosis. Am. J. Respir. Crit. Care Med. 197, 1198-1208 (2018). https://dx.doi.org/10.1164/rccm.201711-2340OC10.1164/rccm.201711-2340OC
- **RISK4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. https://dx.doi.org/10.1164/rccm.201711-2340OC10.1164/rccm.201711-2340OC
- **DIAG3**: Sweeney, Timothy E., Lindsay Braviak, Cristina M. Tato, and Purvesh Khatri. 2016. "Genome-Wide Expression for Diagnosis of Pulmonary Tuberculosis: A Multicohort Analysis." The Lancet. Respiratory Medicine 4 (3): 213-24. https://dx.doi.org/10.1016/S2213-2600(16)00048-510.1016/S2213-2600(16)00048-5
- **TB12**: Tabone O, Verma R, Singhania A, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. J Exp Med. 2021;218(10):e20210915. http://dx.doi.org/10.1084/jem.2021091510.1084/jem.20210915
- **EarlyRESP-TB25**: Tabone O, Verma R, Singhania A, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. J Exp Med. 2021;218(10):e20210915. http://dx.doi.org/10.1084/jem.2021091510.1084/jem.20210915
- **TREAT-TB27**: Tabone O, Verma R, Singhania A, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. J Exp Med. 2021;218(10):e20210915. http://dx.doi.org/10.1084/jem.2021091510.1084/jem.20210915

- **DISEASE**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. https://dx.doi.org/10.1016/j.tube.2017.08.00410.1016/j.tube.2017.08.004
- **FAILURE**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. https://dx.doi.org/10.1016/j.tube.2017.08.00410.1016/j.tube.2017.08.004
- **RESPONSE5**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. https://dx.doi.org/10.1016/j.tube.2017.08.00410.1016/j.tube.2017.08.004
- **Tornheim_71**: Tornheim, Jeffrey A., Anil K. Madugundu, Mandar Paradkar, Kiyoshi F. Fukutani, Artur TL Queiroz, Nikhil Gupte, Akshay N. Gupte et al. 2020. "Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children." The Journal of Infectious Diseases 221(10): 1647-1658. https://doi.org/10.1093/infdis/jiz63910.1093/infdis/jiz639
- **Tornheim_RES_25**: Tornheim, Jeffrey A., Anil K. Madugundu, Mandar Paradkar, Kiyoshi F. Fukutani, Artur TL Queiroz, Nikhil Gupte, Akshay N. Gupte et al. 2020. "Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children." The Journal of Infectious Diseases 221(10): 1647-1658. https://doi.org/10.1093/infdis/jiz63910.1093/infdis/jiz639
- **Vargas_18**: Vargas R, Abbott L, Bower D, Frahm N, Shaffer M, Yu WH. Gene signature discovery and systematic validation across diverse clinical cohorts for TB prognosis and response to treatment. PLoS Comput Biol. 2023;19(7):e1010770. http://dx.doi.org/10.1371/journal.pcbi.101077010.1371/journal.pcbi.1010770
- **Vargas_42**: Vargas R, Abbott L, Bower D, Frahm N, Shaffer M, Yu WH. Gene signature discovery and systematic validation across diverse clinical cohorts for TB prognosis and response to treatment. PLoS Comput Biol. 2023;19(7):e1010770. http://dx.doi.org/10.1371/journal.pcbi.101077010.1371/journal.pcbi.1010770
- **Verhagen_10**: Verhagen, L.M., Zomer, A., Maes, M. et al. A predictive signature gene set for discriminating active from latent tuberculosis in Warao Amerindian children. BMC Genomics 14, 74 (2013). https://doi.org/10.1186/1471-2164-14-7410.1186/1471-2164-14-74
- **Walter_51**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. https://dx.doi.org/10.1128/JCM.01990-1510.1128/JCM.01990-15
- **Walter_PNA_119**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. https://dx.doi.org/10.1128/JCM.01990-1510.1128/JCM.01990-15
- **Walter_PNA_47**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. https://dx.doi.org/10.1128/JCM.01990-1510.1128/JCM.01990-15

- **ACS_COR**: Zak, Daniel E., Adam Penn-Nicholson, Thomas J. Scriba, Ethan Thompson, Sara Suliman, Lynn M. Amon, Hassan Mahomed, et al. 2016. "A Blood RNA Signature for Tuberculosis Disease Risk: A Prospective Cohort Study." The Lancet 387 (10035): 231222. https://dx.doi.org/10.1016/S0140-6736(15)01316-110.1016/S0140-6736(15)01316-1
- **Zhao_NANO_6**: Kaipilyawar, Vaishnavi, et al. Development and Validation of a Parsimonious Tuberculosis Gene Signature Using the digital NanoString nCounter Platform. Clinical Infectious Diseases 75.6 (2022): 1022-1030. http://dx.doi.org/10.1093/cid/ciac010

Examples

```
data("TBcommon")
```

`TBsignatures` *A list of published TB signatures.*

Description

A set of Tuberculosis gene signatures compiled from various publications. This set of signatures uses gene symbols. Attempts have been made to use updated gene symbols and remove symbols that did not match the most recent annotation. Additional sets for Entrez IDs and Ensembl IDs are forthcoming.

Usage

```
TBsignatures
```

Format

list

Details

Signature names are composed of the last name of the primary author, followed by a possible context for the signature, and ending with either the number of gene transcripts or genes in the signature, with respect to however it was described in the signature's original publication.

Possible signature contexts:

- <blank>: TB vs LTBI or Healthy Controls
- OD: Other diseases
- HIV: Human Immunodeficiency Virus
- PNA: Pneumonia
- RISK: Risk of developing active TB
- RES: Response to TB treatment
- FAIL: Failure of TB treatment

Note that in some cases signatures will be positive identifiers of TB whereas others are negative identifiers; this should be taken into account when creating ROC curves and computing any AUC estimates.

Source

- **Anderson_42**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. https://dx.doi.org/10.1056/NEJMoa130365710.1056/NEJMoa1303657
- **Anderson_OD_51**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. https://dx.doi.org/10.1056/NEJMoa130365710.1056/NEJMoa1303657
- **Berry_393**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. https://dx.doi.org/10.1038/nature0924710.1038/nature09247
- **Berry_OD_86**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. https://dx.doi.org/10.1038/nature0924710.1038/nature09247
- **Blankley_380**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. https://dx.doi.org/10.1183/13993003.02121-201510.1183/13993003.02121-2015
- **Blankley_5**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. https://dx.doi.org/10.1183/13993003.02121-201510.1183/13993003.02121-2015
- **Bloom_OD_144**: Bloom, Chloe I., Christine M. Graham, Matthew P. R. Berry, Fotini Rozakeas, Paul S. Redford, Yuanyuan Wang, Zhaohui Xu, et al. 2013. "Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers." PloS One 8 (8): e70630. https://dx.doi.org/10.1371/journal.pone.007063010.1371/journal.pone.0070630
- **Bloom_RES_268**: Bloom CI, Graham CM, Berry MP, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One. 2012;7(10):e46191. https://dx.doi.org/10.3389/fmicb.2021.65056710.3389/fmicb.2021.650567
- **Bloom_RES_558**: Bloom CI, Graham CM, Berry MP, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One. 2012;7(10):e46191. https://dx.doi.org/10.3389/fmicb.2021.65056710.3389/fmicb.2021.650567
- **Chen_5**: Chen L, Hua J, He X. Coexpression Network Analysis-Based Identification of Critical Genes Differentiating between Latent and Active Tuberculosis. Dis Markers. 2022;2022:2090560. http://dx.doi.org/10.1155/2022/209056010.1155/2022/2090560
- **Chen_HIV_4**: Chen Y, Wang Q, Lin S, et al. Meta-Analysis of Peripheral Blood Transcriptome Datasets Reveals a Biomarker Panel for Tuberculosis in Patients Infected With HIV. Front Cell Infect Microbiol. 2021;11:585919. Published 2021 Mar 19. https://dx.doi.org/10.3389/fcimb.2021.58591910.3389/fcimb.2021.585919

- **Chendi_HIV_2**: Chendi BH, Tveiten H, Snyders CI, et al. CCL1 and IL-2Ra differentiate Tuberculosis disease from latent infection Irrespective of HIV infection in low TB burden countries. J Infect. 2021;S0163-4453(21)00379-0. https://dx.doi.org/10.1016/j.jinf.2021.07.03610.1016/j.jinf.2021.07.036
- **Darboe_RISK_11**: Darboe, F. et al. Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells. Tuberculosis 108, 124-126 (2018). https://dx.doi.org/10.1016/j.tube.2017.11.00110.1016/j.tube.2017.11.001
- **Dawany_HIV_251**: Dawany, N. et al. Identification of a 251 gene expression signature that can accurately detect M. tuberculosis in patients with and without HIV co-infection. PLoS One 9, (2014). https://dx.doi.org/10.1371/journal.pone.008992510.1371/journal.pone.0089925
- **Duffy_23**: Duffy FJ, Olson GS, Gold ES, Jahn A, Aderem A, Aitchison J, Rothchild AC, Diercks AH, Nemeth J. A contained Mycobacterium tuberculosis mouse infection model predicts active disease and containment in humans. The Journal of Infectious Diseases. 2021 Mar 10. https://dx.doi.org/10.1093/infdis/jiab13010.1093/infdis/jiab130
- **Esmail_203**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. https://dx.doi.org/10.1073/pnas.171185311510.1073/pnas.1711853115
- **Esmail_82**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. https://dx.doi.org/10.1073/pnas.171185311510.1073/pnas.1711853115
- **Esmail_893**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. https://dx.doi.org/10.1073/pnas.171185311510.1073/pnas.1711853115
- **Estevez_133**: Estévez O, Anibarro L, Garet E, et al. An RNA-seq Based Machine Learning Approach Identifies Latent Tuberculosis Patients With an Active Tuberculosis Profile. Front Immunol. 2020;11:1470. Published 2020 Jul 14. https://dx.doi.org/10.3389/fimmu.2020.0147010.3389/fimmu.2020.01470
- **Estevez_259**: Estévez O, Anibarro L, Garet E, et al. An RNA-seq Based Machine Learning Approach Identifies Latent Tuberculosis Patients With an Active Tuberculosis Profile. Front Immunol. 2020;11:1470. Published 2020 Jul 14. https://dx.doi.org/10.3389/fimmu.2020.0147010.3389/fimmu.2020.01470
- **Francisco_OD_2**: Francisco NM, Fang YM, Ding L, et al. Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. J Infect. 2017;75(6):499-510. http://dx.doi.org/10.1016/j.jinf.2017.09.01210.1016/j.jinf.2017.09.012
- **Gjoen_10**: Gjøen, J.E., Jenum, S., Sivakumaran, D. et al. 'Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children.' Sci Rep 7, 5839 (2017). https://doi.org/10.1038/s41598-017-05057-x10.1038/s41598-017-05057-x

- **Gjoen_7**: Gjøen, J.E., Jenum, S., Sivakumaran, D. et al. 'Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children.' Sci Rep 7, 5839 (2017). https://doi.org/10.1038/s41598-017-05057-x10.1038/s41598-017-05057-x
- **Gliddon_2_OD_4**: Gliddon HD, Kaforou M, Alikian M, et al. Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. Front Immunol. 2021;12:637164. Published 2021 Mar 2. https://dx.doi.org/10.3389/fimmu.2021.63716410.3389/fimmu.2021.637164
- **Gliddon_HIV_3**: Gliddon HD, Kaforou M, Alikian M, et al. Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. Front Immunol. 2021;12:637164. Published 2021 Mar 2. https://dx.doi.org/10.3389/fimmu.2021.63716410.3389/fimmu.2021.637164
- **Gliddon_OD_3**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org:https://dx.doi.org/10.1101/58367410.1101/583674
- **Gliddon_OD_4**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org:https://dx.doi.org/10.1101/58367410.1101/583674
- **Gong_OD_4**: Gong Z, Gu Y, Xiong K, Niu J, Zheng R, Su B, Fan L and Xie J (2021) The Evaluation and Validation of Blood-Derived Novel Biomarkers for Precise and Rapid Diagnosis of Tuberculosis in Areas With High-TB Burden. Front. Microbiol. 12:650567. https://dx.doi.org/10.3389/fmicb.2021.65056710.3389/fmicb.2021.650567
- **Heycken_FAIL_22**: Heyckendorf J, Marwitz S, Reimann M, et al. Prediction of anti-tuberculosis treatment duration based on a 22-gene transcriptomic model. Eur Respir J. 2021;2003492. https://dx.doi.org/10.1183/13993003.03492-202010.1183/13993003.03492-2020
- **Hoang_OD_13**: Hoang, Long & Jain, Pooja & Pillay, Timesh & Tolosa-Wright, Mica & Niazi, Umar & Takwoingi, Yemisi & Halliday, Alice & Berrocal-Almanza, Luis & Deeks, Jonathan & Beverley, Peter & Kon, Onn & Lalvani, Ajit. (2021). Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. The Lancet Infectious Diseases. https://dx.doi.org/10.1016/S1473-3099(20)30928-210.1016/S1473-3099(20)30928-2
- **Hoang_OD_20**: Hoang, Long & Jain, Pooja & Pillay, Timesh & Tolosa-Wright, Mica & Niazi, Umar & Takwoingi, Yemisi & Halliday, Alice & Berrocal-Almanza, Luis & Deeks, Jonathan & Beverley, Peter & Kon, Onn & Lalvani, Ajit. (2021). Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. The Lancet Infectious Diseases. https://dx.doi.org/10.1016/S1473-3099(20)30928-210.1016/S1473-3099(20)30928-2
- **Hoang_OD_3**: Hoang, Long & Jain, Pooja & Pillay, Timesh & Tolosa-Wright, Mica & Niazi, Umar & Takwoingi, Yemisi & Halliday, Alice & Berrocal-Almanza, Luis & Deeks, Jonathan & Beverley, Peter & Kon, Onn & Lalvani, Ajit. (2021). Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. The Lancet Infectious Diseases. https://dx.doi.org/10.1016/S1473-3099(20)30928-210.1016/S1473-3099(20)30928-2

- **Huang_13**: Huang, Hai-Hui et al. 'Identification of 13 Blood-based Gene Expression Signatures to Accurately Distinguish Tuberculosis from Other Pulmonary Diseases and Healthy Controls'. 1 Jan. 2015 : S1837 - S1843.https://doi.org/10.3233/BME-15148610.3233/BME-151486
- **Jacobsen_3**: Jacobsen, Marc, Dirk Repsilber, Andrea Gutschmidt, Albert Neher, Knut Feldmann, Hans J. Mollenkopf, Andreas Ziegler, and Stefan H. E. Kaufmann. 2007. "Candidate Biomarkers for Discrimination between Infection and Disease Caused by Mycobacterium Tuberculosis." Journal of Molecular Medicine 85 (6): 613-21. https://dx.doi.org/10.1007/s00109-007-0157-610.1007/s00109-007-0157-6
- **Jenum_8**: Jenum, S., Dhanasekaran, S., Lodha, R. et al. Approaching a diagnostic point-of-care test for pediatric tuberculosis through evaluation of immune biomarkers across the clinical disease spectrum. Sci Rep 6, 18520 (2016). https://doi.org/10.1038/srep1852010.1038/srep18520
- **Kaforou_27**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. https://dx.doi.org/10.1371/journal.pmed.100153810.1371/journal.pmed.1001538
- **Kaforou_OD_44**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. https://dx.doi.org/10.1371/journal.pmed.100153810.1371/journal.pmed.1001538
- **Kaforou_OD_53**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. https://dx.doi.org/10.1371/journal.pmed.100153810.1371/journal.pmed.1001538
- **Kaul_3**: Kaul S, Nair V, Birla S, et al. Latent Tuberculosis Infection Diagnosis among Household Contacts in a High Tuberculosis-Burden Area: a Comparison between Transcript Signature and Interferon Gamma Release Assay. Microbiol Spectr. 2022;10(2):e0244521. http://dx.doi.org/10.1128/spectrum.02445-2110.1128/spectrum.02445-21
- **Kulkarni_HIV_2**: Kulkarni V, Queiroz ATL, Sangle S, et al. A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. Front Immunol. 2021;12:631165. Published 2021 Feb 22. https://dx.doi.org/10.3389/fimmu.2021.63116510.3389/fimmu.2021.631165
- **Kwan_186**: Kwan PKW, Periaswamy B, De Sessions PF, et al. A blood RNA transcript signature for TB exposure in household contacts. BMC Infect Dis. 2020;20(1):403. http://dx.doi.org/10.1186/s12879-020-05116-110.1186/s12879-020-05116-1
- **LauxdaCosta_OD_3**: Laux da Costa L, Delcroix M, Dalla Costa ER, et al. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. Tuberculosis (Edinb). 2015;95(4):421-425. https://dx.doi.org/10.1016/j.tube.2015.04.00810.1016/j.tube.2015.04.008
- **Lee_4**: Lee, Shih-Wei, Lawrence Shih-Hsin Wu, Guan-Mau Huang, Kai-Yao Huang, Tzong-Yi Lee, and Julia Tzu-Ya Weng. 2016. "Gene Expression Profiling Identifies Candidate Biomarkers for Active and Latent Tuberculosis." BMC Bioinformatics 17 Suppl 1 (January): 3. https://dx.doi.org/10.1186/s12859-015-0848-x10.1186/s12859-015-0848-x
- **Leong_24**: Leong, Samantha, Yue Zhao, Noyal M. Joseph, Natasha S. Hochberg, Sonali Sarkar, Jane Pleskunas, David Hom, et al. 2018. "Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India."

Tuberculosis (109): 41-51. https://doi.org/10.1016/j.tube.2018.01.00210.1016/j.tube.2018.01.002

- **Leong_RISK_29**: Leong, S., Zhao, Y., Ribeiro-Rodrigues, R., Jones-López, E. C., Acuña-Villaorduña, C., Rodrigues, P. M., Palaci, M., Alland, D., Dietze, R., Ellner, J. J., Johnson, W. E., Salgame, P., Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB. Tuberculosis (Edinb). 2020 Jan;120:101898. https://dx.doi.org/10.1016/j.tube.2020.10189810.1016/j.tube.2020.101898
- **Li_3**: Li M, Qiu Y, Guo M, Qu R, Tian F, Wang G, Wang Y, Ma J, Liu S, Takiff H, Tang Y, Gao Q.2023.Evaluation of the Cepheid 3-gene host response blood test for tuberculosis diagnosis and treatment response monitoring in a primary-level clinic in rural China. J Clin Microbiol61:e00911-23. http://dx.doi.org/10.1128/jcm.00911-23
- **Long_RES_10**: Long NP, Phat NK, Yen NTH, et al. A 10-gene biosignature of tuberculosis treatment monitoring and treatment outcome prediction. Tuberculosis (Edinb). 2021;131:102138. http://dx.doi.org/10.1016/j.tube.2021.10213810.1016/j.tube.2021.102138
- **Maertzdorf_15**: Maertzdorf J, McEwen G, Weiner J 3rd, et al. Concise gene signature for point-of-care classification of tuberculosis. EMBO Mol Med. 2016;8(2):86-95. https://dx.doi.org/10.15252/emmm.20150579010.15252/emmm.201505790
- **Maertzdorf_4**: Maertzdorf, Jeroen, Gayle McEwen, January Weiner 3rd, Song Tian, Eric Lader, Ulrich Schriek, Harriet Mayanja-Kizza, Martin Ota, John Kenneth, and Stefan He Kaufmann. 2016. "Concise Gene Signature for Point-of-Care Classification of Tuberculosis." EMBO Molecular Medicine 8 (2): 86-95. https://dx.doi.org/10.15252/emmm.20150579010.15252/emmm.201505790
- **Maertzdorf_OD_100**: Maertzdorf, Jeroen, January Weiner 3rd, Hans-Joachim Mollenkopf, TBornot TB Network, Torsten Bauer, Antje Prasse, Joachim Müller-Quernheim, and Stefan H. E. Kaufmann. 2012. "Common Patterns and Disease-Related Signatures in Tuberculosis and Sarcoidosis." Proceedings of the National Academy of Sciences of the United States of America 109 (20): 7853-58. https://dx.doi.org/10.1073/pnas.112107210910.1073/pnas.1121072109
- **Natarajan_7**: Natarajan S, Ranganathan M, Hanna LE, Tripathy S. Transcriptional Profiling and Deriving a Seven-Gene Signature That Discriminates Active and Latent Tuberculosis: An Integrative Bioinformatics Approach. Genes (Basel). 2022;13(4):616. Published 2022 Mar 29. http://dx.doi.org/10.3390/genes1304061610.3390/genes13040616
- **PennNich_RISK_6**: Penn-Nicholson, A. et al. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. Sci. Rep. 10, (2020). https://dx.doi.org/10.1038/s41598-020-65043-810.1038/s41598-020-65043-8
- **Qian_OD_17**: Qian, Zhongqing et al. "Expression of nuclear factor, erythroid 2-like 2-mediated genes differentiates tuberculosis." Tuberculosis (Edinburgh, Scotland) vol. 99 (2016): 56-62. https://doi.org/10.1016/j.tube.2016.04.00810.1016/j.tube.2016.04.008
- **Rajan_HIV_5**: Rajan, Jayant V., Semitala, Fred C., Kamya, Moses R., Yoon, Christina., Mehta, Tejas., Cattamanchi, Adithya., Seielstad, Mark., Montalvo, Lani., Andama, Alfred., Katende, Jane., Asege, Lucy., Nakaye, Martha., Mwebe, Sandra. 2018 "A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus" Clinical Infectious Diseases: 1-7. https://doi.org/10.1093/cid/ciy83510.1093/cid/ciy835
- **Roe_3**: Roe, Jennifer, Venturini, Cristina, Gupta, Rishi K., Gurry, Celine, Chain, Benjamin M., Sun, Yuxin, Southern, Jo, Jackson, Charlotte, Lipman, Marc, C., Miller, Robert F., Martineau, Adrian R., Abubakar, Ibrahim, Noursadeghi, Mahdad. 2019 "T1 Blood transcriptomic stratification of short-term risk in contacts of tuberculosis": . https://doi.org/10.1093/cid/ciz25210.1093/cid/ciz252

- **Roe_OD_4**: Roe, Jennifer K., Niclas Thomas, Eliza Gil, Katharine Best, Evdokia Tsaliki, Stephen Morris-Jones, Sian Stafford, et al. 2016. "Blood Transcriptomic Diagnosis of Pulmonary and Extrapulmonary Tuberculosis." JCI Insight 1 (16): e87238. https://dx.doi.org/10.1172/jci.insight.8723810.1172/jci.insight.87238
- **Sambarey_HIV_10**: Sambarey, Awanti, Abhinandan Devaprasad, Abhilash Mohan, Asma Ahmed, Soumya Nayak, Soumya Swaminathan, George D'Souza, et al. 2017. "Unbiased Identification of Blood-Based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks." EBioMedicine 15 (February): 112-26. https://dx.doi.org/10.1016/j.ebiom.2016.12.00910.1016/j.ebiom.2016.12.009
- **Singhania_OD_20**: Singhania, Akul, Raman Verma, Christine M. Graham, Jo Lee, Trang Tran, Matthew Richardson, Patrick Lecine, et al. 2018. "A Modular Transcriptional Signature Identifies Phenotypic Heterogeneity of Human Tuberculosis Infection." Nature Communications 9 (1): 2308. https://dx.doi.org/10.1038/s41467-018-04579-w10.1038/s41467-018-04579-w
- **Sivakumaran_11**: Sivakumaran D, Ritz C, Gjøen JE, et al. Host Blood RNA Transcript and Protein Signatures for Sputum-Independent Diagnostics of Tuberculosis in Adults. Front Immunol. 2021;11:626049. Published 2021 Feb 4. https://dx.doi.org/10.3389/fimmu.2020.62604910.3389/fimmu.2020.626049
- **Sloot_HIV_2**: Sloot, Rosa, Maarten F. Schim van der Loeff, Erik W. van Zwet, Mariëlle C. Haks, Sytze T. Keizer, Maarten Scholing, Tom H. M. Ottenhoff, Martien W. Borgdorff, and Simone A. Joosten. 2015. "Biomarkers Can Identify Pulmonary Tuberculosis in HIV-Infected Drug Users Months Prior to Clinical Diagnosis." EBioMedicine 2 (2): 172-79. https://dx.doi.org/10.1016/j.ebiom.2014.12.00110.1016/j.ebiom.2014.12.001
- **Suliman_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. https://dx.doi.org/10.1164/rccm.201711-2340OC10.1164/rccm.201711-2340OC
- **Suliman_RISK_2**: Suliman, S. et al. Four-gene pan-African blood signature predicts progression to tuberculosis. Am. J. Respir. Crit. Care Med. 197, 1198-1208 (2018). https://dx.doi.org/10.1164/rccm.201711-2340OC10.1164/rccm.201711-2340OC
- **Suliman_RISK_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. https://dx.doi.org/10.1164/rccm.201711-2340OC10.1164/rccm.201711-2340OC
- **Sweeney_OD_3**: Sweeney, Timothy E., Lindsay Braviak, Cristina M. Tato, and Purvesh Khatri. 2016. "Genome-Wide Expression for Diagnosis of Pulmonary Tuberculosis: A Multicohort Analysis." The Lancet. Respiratory Medicine 4 (3): 213-24. https://dx.doi.org/10.1016/S2213-2600(16)00048-510.1016/S2213-2600(16)00048-5
- **Tabone_OD_11**: Tabone O, Verma R, Singhania A, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. J Exp Med. 2021;218(10):e20210915. http://dx.doi.org/10.1084/jem.2021091510.1084/jem.20210915
- **Tabone_RES_25**: Tabone O, Verma R, Singhania A, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. J Exp Med. 2021;218(10):e20210915. http://dx.doi.org/10.1084/jem.2021091510.1084/jem.20210915
- **Tabone_RES_27**: Tabone O, Verma R, Singhania A, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. J Exp Med. 2021;218(10):e20210915. http://dx.doi.org/10.1084/jem.2021091510.1084/jem.20210915

- **Thompson_9**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. https://dx.doi.org/10.1016/j.tube.2017.08.00410.1016/j.tube.2017.08.004
- **Thompson_FAIL_13**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. https://dx.doi.org/10.1016/j.tube.2017.08.00410.1016/j.tube.2017.08.004
- **Thompson_RES_5**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. https://dx.doi.org/10.1016/j.tube.2017.08.00410.1016/j.tube.2017.08.004
- **Tornheim_71**: Tornheim, Jeffrey A., Anil K. Madugundu, Mandar Paradkar, Kiyoshi F. Fukutani, Artur TL Queiroz, Nikhil Gupte, Akshay N. Gupte et al. 2020. "Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children." The Journal of Infectious Diseases 221(10): 1647-1658. https://doi.org/10.1093/infdis/jiz63910.1093/infdis/jiz639
- **Tornheim_RES_25**: Tornheim, Jeffrey A., Anil K. Madugundu, Mandar Paradkar, Kiyoshi F. Fukutani, Artur TL Queiroz, Nikhil Gupte, Akshay N. Gupte et al. 2020. "Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children." The Journal of Infectious Diseases 221(10): 1647-1658. https://doi.org/10.1093/infdis/jiz63910.1093/infdis/jiz639
- **Vargas_18**: Vargas R, Abbott L, Bower D, Frahm N, Shaffer M, Yu WH. Gene signature discovery and systematic validation across diverse clinical cohorts for TB prognosis and response to treatment. PLoS Comput Biol. 2023;19(7):e1010770. http://dx.doi.org/10.1371/journal.pcbi.101077010.1371/journal.pcbi.1010770
- **Vargas_42**: Vargas R, Abbott L, Bower D, Frahm N, Shaffer M, Yu WH. Gene signature discovery and systematic validation across diverse clinical cohorts for TB prognosis and response to treatment. PLoS Comput Biol. 2023;19(7):e1010770. http://dx.doi.org/10.1371/journal.pcbi.101077010.1371/journal.pcbi.1010770
- **Verhagen_10**: Verhagen, L.M., Zomer, A., Maes, M. et al. A predictive signature gene set for discriminating active from latent tuberculosis in Warao Amerindian children. BMC Genomics 14, 74 (2013). https://doi.org/10.1186/1471-2164-14-7410.1186/1471-2164-14-74
- **Walter_51**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. https://dx.doi.org/10.1128/JCM.01990-1510.1128/JCM.01990-15
- **Walter_PNA_119**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. https://dx.doi.org/10.1128/JCM.01990-1510.1128/JCM.01990-15
- **Walter_PNA_47**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. https://dx.doi.org/10.1128/JCM.01990-1510.1128/JCM.01990-15

- **Zak_RISK_16**: Zak, Daniel E., Adam Penn-Nicholson, Thomas J. Scriba, Ethan Thompson, Sara Suliman, Lynn M. Amon, Hassan Mahomed, et al. 2016. "A Blood RNA Signature for Tuberculosis Disease Risk: A Prospective Cohort Study." The Lancet 387 (10035): 231222. https://dx.doi.org/10.1016/S0140-6736(15)01316-110.1016/S0140-6736(15)01316-1
- **Zhao_NANO_6**: Kaipilyawar, Vaishnavi, et al. Development and Validation of a Parsimonious Tuberculosis Gene Signature Using the digital NanoString nCounter Platform. Clinical Infectious Diseases 75.6 (2022): 1022-1030. http://dx.doi.org/10.1093/cid/ciac010

Examples

```
data("TBsignatures")
```

`TBsignaturesSplit` *Up/Down-regulated genes information for selected TB signatures.*

Description

Up/Down-regulated genes information for selected TB signatures.

Usage

```
TBsignaturesSplit
```

Format

list

Source

See `?TBsignatures` for reference information.

Examples

```
data("TBsignaturesSplit")
```

`TBSPapp` *Run the TBSignatureProfiler Shiny application.*

Description

Use this function to run the TBSignatureProfiler application.

Usage

```
TBSPapp()
```

Value

The Shiny application will open.

Examples

```
# Upload data through the app
if (interactive()) {
  TBSPapp()
}
```

`TB_hiv` *An example TB dataset with TB/HIV data.*

Description

An example dataset containing the gene expression and metadata in a SummarizedExperiment object for 31 subjects with HIV and/or Tuberculosis diseases. Information on subject infection status can be accessed with `TB_hiv$Disease`. Samples with both TB and HIV contamination are marked as `tb_hiv`, while samples with HIV and no TB are marked as `hiv_only`.

Usage

```
TB_hiv
```

Format

SummarizedExperiment

Details

This dataset was published as part of a study to assess whether gene expression signatures and cytokine levels would distinguish active TB in advanced HIV in a cohort residing in Sub-Saharan Africa (Verma et. al 2018). Participants were severely immunosuppressed TB-HIV patients who had not yet received TB treatment or anti-retroviral therapy (ART). The dataset included in this package has been lightly edited from the originally published dataset due to the removal of one participant who was HIV positive, on ART and developed TB during follow-up. Whole blood RNA-Seq analysis was performed on all 31 participants.

References

Verma S., Du P., et. al. (2018). Tuberculosis in advanced HIV infection is associated with increased expression of IFN and its downstream targets. *BMC Infectious Diseases* **18:220**. doi: https://doi.org/10.1186/s12879-018-3127-410.1186/s12879-018-3127-4.

Examples

```
data("TB_hiv")
```

TB_indian *An example TB dataset with Indian population data.*

Description

An example dataset containing the gene expression and metadata in a SummarizedExperiment object for an Indian population. Active TB contamination of the 44 subjects is denoted for each as a "1"(active) or "0" (latent/not present), and can be accessed via `TB_indian$label`. The SummarizedExperiment object contains 2 assays (counts and log(counts)), and the column names give the unique subject identification number along with the subject's gender.

Usage

```
TB_indian
```

Format

SummarizedExperiment

Details

This dataset was published as part of a study to assess performance of published TB signatures in a South Indian population (Leong et. al 2018). RNA sequencing was performed on whole blood PAX gene samples collected from 28 TB patients and 16 latent TB infected (LTBI) subjects enrolled as part of an ongoing household contact study. Whole blood RNA-Seq analysis was performed on all 44 participants.

References

Leong S., Zhao Y., et. al. (2018). Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India. *Tuberculosis* **109**, 41-51. doi: https://doi.org/10.1016/j.tube.2018.01.00210.1016/j.tube.2018.01.002.

Examples

```
data("TB_indian")
```

Index

∗ **datasets**
common_sigAnnotData, 8
COVIDsignatures, 12
OriginalTrainingData, 23
sigAnnotData, 29
TB_hiv, 60
TB_indian, 61
TBcommon, 43
TBsignatures, 51
TBsignaturesSplit, 59
.OriginalModel_NoRetraining, 3
.OriginalModel_Retraining, 4

addTBsignature, 5

Bootstrap_LOOCV_LR_AUC, 7
bootstrapAUC, 6
bplapply, *3, 4*, *18*

checkGeneSymbols, *18*
ComBat, *4*, *18*
common_sigAnnotData, 8
compareAlgs, 9
compareBoxplots, 11
COVIDsignatures, 12
cv_glmnet_OriginalModel, 16

deseq2_norm_rle, 16
distinctColors, 17

evaluateOriginalModel, 17

impute.knn, *25, 26*

knn, *4*
knn_OriginalModel, 19

lda, *4*
lda_OriginalModel, 19
LOOAUC_simple_multiple_noplot_one_df, 20

mkAssay, 20

ObtainSampleScore_OriginalModel, 22
OriginalTrainingData, 23

plotQuantitative, 23

randomForest, *4*
randomForest_OriginalModel, 25
ref_combat_impute, 25
runTBsigProfiler, 26

sigAnnotData, 29
signatureBoxplot, 30
signatureGeneHeatmap, 31
signatureHeatmap, 33
SignatureQuantitative, 36
signatureROCplot, 37
signatureROCplot_CI, 38
subsetGeneSet, 40
SulimanOriginalModel, 41
svm, *4*
svm_OriginalModel, 41

tableAUC, 42
TB_hiv, 60
TB_indian, 61
TBcommon, 43
TBsignatures, 51
TBsignaturesSplit, 59
TBSPapp, 59