Package: thanatome (via r-universe)

November 10, 2026

Title Donor Death Signatures for Post-Mortem Control Tissue

Version 0.99.0

Description Annotates each sample of a post-mortem tissue cohort (GTEx or similar) with its position along two transcriptional axes of donor death derived from GTEx v8 across ten tissues, ventilator (perfused versus post-mortem) and prolonged agony, plus the ischemia component of the ventilator axis. For every sample it returns a signature score per axis, a position scaled between the two reference poles of the same tissue, a state label and technical contamination flags, and optionally the fraction of the sample's deviation from the tissue centroid that lies along the axes. Samples can then be plotted, selected with explicit criteria, and the per-sample scores used as covariates in differential-expression designs or batch-correction tools. compare_controls() repeats a case-versus-control differential expression with each candidate control subset against random subsets of the same size, optionally carrying a master-regulator analysis; annotate_de() and annotate_mra() annotate existing differential-expression and regulator tables with the axes their genes fall in; audit_report() writes the summary for a supplement.

License MIT + file LICENSE

Encoding UTF-8

biocViews Software, QualityControl, Transcriptomics, RNASeq, GeneExpression, BatchEffect, Classification, DifferentialExpression, Sequencing

Roxygen list(markdown = TRUE)

Depends R (>= 4.5.0), SummarizedExperiment

Imports BiocParallel, S4Vectors, edgeR (>= 4.0.0), ggplot2, methods, stats, utils

Suggests BiocStyle, DESeq2, GSEABase, knitr, rmarkdown, testthat (>= 3.1.5), limma, AnnotationDbi, org.Hs.eg.db

Config/testthat/edition 3

VignetteBuilder knitr

URL https://github.com/hachepunto/thanatome

BugReports https://github.com/hachepunto/thanatome/issues

LazyData false

Date 2026-11-09

RoxygenNote 7.3.3

Config/pak/sysreqs zlib1g-dev

Repository https://biocstaging.r-universe.dev

Date/Publication 2026-11-09 23:33:37 UTC

RemoteUrl https://github.com/BiocStaging/thanatome

RemoteRef HEAD

RemoteSha 2abb737d23d7d0df1124caa5af3213ef055280bc

Contents

thanatome-package . . . 3
add_thanatome_score . . . 4
annotate_de . . . 5
annotate_mra . . . 6
as_de_table . . . 7
audit_report . . . 8
compare_controls . . . 9
contamination_flags . . . 11
death_covariates . . . 12
death_gene_sets . . . 13
death_reference . . . 14
death_signatures . . . 15
hardy_layers . . . 15
plot_thanatome . . . 16
select_samples . . . 17
thanatome_example . . . 18
thanatome_references . . . 19
thanatome_score . . . 20
thanatome_signatures . . . 23

Index **25**

thanatome-package *thanatome: donor death signatures for post-mortem control tissue*

Description

Annotates each sample of a post-mortem tissue cohort with its position along two transcriptional axes of donor death derived from GTEx v8 in ten tissues (ventilator / perfusion and prolonged agony), plus the ischemia component of the ventilator axis, lets the user plot and select samples with explicit criteria, exports the per-sample scores as covariates, and annotates existing differential-expression tables with the axes their genes fall in.

Details

The main entry point is `thanatome_score()`. See the vignette `vignette("thanatome")` for a worked example.

Author(s)

Maintainer: Hugo Tovar <hatovar@inmegen.gob.mx> (ORCID) [copyright holder]

Authors:

- Daniel P<U+00E9>rez-Calixto <dperez@inmegen.gob.mx> (ORCID) [copyright holder]
- Enrique Hernandez-Lemus <ehernandez@inmegen.gob.mx> (ORCID) [copyright holder]

Other contributors:

- Camila Barrios-Cervantes <paulacamilabarrioscervantes@gmail.com> (ORCID) [contributor, copyright holder]
- Instituto Nacional de Medicina Gen<U+00F3>mica (ROR) (INMEGEN-2026-14, Proyecto 81) [funder]

See Also

Useful links:

- https://github.com/hachepunto/thanatome
- Report bugs at https://github.com/hachepunto/thanatome/issues

add_thanatome_score *Add the death-axis annotation to the colData of a SummarizedExperiment*

Description

Convenience wrapper around `thanatome_score()` that returns `x` itself, with the per-sample annotation appended as `colData` columns, so that the scores travel with the counts into `DESeqDataSet`, `DGEList` or any other downstream object. The metadata of the score (reference, method, poles and the internal check against Hardy) is stored in `metadata(x)$thanatome`.

Usage

```
add_thanatome_score(x, reference, ..., prefix = "")
```

Arguments

`x`	A `SummarizedExperiment` with raw counts (see `thanatome_score()`).
`reference, ...`	Passed to `thanatome_score()`.
`prefix`	Character string prepended to the new column names, to avoid clashes with existing `colData` columns. Default `""`.

Value

`x` with the columns of `thanatome_score()` added to its `colData` (existing columns with the same name are replaced, with a message) and `metadata(x)$thanatome`.

See Also

`thanatome_score()`

Examples

```
data(thanatome_example)
ctl <- thanatome_example[, thanatome_example$group == "control"]
ctl <- add_thanatome_score(ctl, reference = "ovary")
table(ctl$state)
```

annotate_de *Annotate a differential-expression table with the death axes*

Description

For each gene, adds the axis and direction in which it falls in the consensus signatures, the consensus log fold change, and the change in the case-versus-control log fold change expected when the control set moves from one pole to the other: from all non-ventilated to all ventilated donors (ventilator axis) or from all fast to all slow deaths (agony axis). Within GTEx that extreme change equals the contrast between the two subsets of controls, which the consensus of the other tissues predicts with a slope close to -1; the prediction is therefore `-lfc_consensus`. A move from a mixed control set to one pole is a fraction of it: scale by the share of controls that change pole (`scale`). The flag `sign_depends_on_control` marks the genes whose (scaled) predicted change exceeds their own log fold change. The ischemia component is not annotated by default: its consensus is per standard deviation of ischemic time and is not a pole-to-pole contrast.

Usage

```
annotate_de(de, axes = c("ventilator", "agony"), min_tissues = NULL, scale = 1)
```

Arguments

de	A DE result accepted by `as_de_table()`. Gene ids are Ensembl (with or without version); if fewer than half look like Ensembl they are taken as HGNC symbols and mapped with `org.Hs.eg.db` (Suggests). The original ids are kept in the output.
axes	Axes to annotate (default the two axes).
min_tissues	Minimum number of concordant tissues for a gene to be annotated (default: the signature threshold, 5).
scale	Factor applied to the predicted change (default 1, the full pole-to-pole swing; use for example 0.4 if 40 percent of the controls would change pole).

Value

The `DataFrame` from `as_de_table()` with the extra columns `axis_<axis>`, `lfc_consensus_<axis>`, `delta_lfc_<axis>` and `sign_depends_on_control`.

Examples

```
de <- data.frame(
    logFC = c(1.2, -0.8, 0.3), adj.P.Val = c(0.01, 0.2, 0.5),
    row.names = c("ENSG00000166189", "ENSG00000172409", "ENSG00000000003")
)
annotate_de(de)
```

`annotate_mra` *Annotate a master-regulator table with the death axes*

Description

For each regulator, reports which fraction of its regulon falls in each death axis, whether that fraction is larger than expected (Fisher test against the regulon universe) and a death-direction score: the mean, over the targets that belong to the axis, of the product of the regulon mode (sign of `tfmode`) and the sign of the consensus log fold change. A score near +1 means the regulon's activated targets lie on the "up" side of the axis, near -1 that they lie on the "down" side. The sign of "up" is defined per axis: higher in ventilated donors (Hardy 0, the perfused pole) for `ventilator`, higher in slow deaths for `agony`, and higher with longer ischemic time for `ischemia`, as in `death_signatures()`. Regulators with a significant enrichment are flagged as ones whose differential activity may reflect the state of the controls rather than the cases.

Usage

```
annotate_mra(
  nes,
  regulons,
  axes = c("ventilator", "agony"),
  min_targets = 10,
  universe = NULL,
  padj_threshold = 0.05
)
```

Arguments

`nes`	A named numeric vector of regulator activities (NES from viper, NaRnEA or similar), or a `data.frame` with columns `regulator` and `nes`.
`regulons`	A list named by regulator. Each element is either a list with a named numeric `tfmode` (viper format; `likelihood` is ignored), a named numeric vector of modes, or a character vector of target gene ids (mode +1). Gene ids are Ensembl, with or without version; HGNC symbols are mapped to Ensembl with `org.Hs.eg.db` (Suggests) when the ids do not look like Ensembl, with a message, and the universe is mapped the same way.
`axes`	Axes to annotate (default the two axes; `"ischemia"` adds the component).
`min_targets`	Regulators with fewer targets in the universe are dropped. Default 10.
`universe`	Gene universe for the Fisher test; by default the union of all regulon targets.
`padj_threshold`	Adjusted p-value at or below which an axis enrichment raises `flag_control_state`. Default 0.05.

Value

A `DataFrame` with one row per regulator: `nes`, `n_targets`, and per axis `frac_<axis>`, `fisher_padj_<axis>`, `direction_<axis>` (+1: the regulon's activated targets are higher in ventilated donors for `ventilator`, in slow deaths for `agony`, with longer ischemic time for `ischemia`), plus `flag_control_state`.

Examples

```
sig <- death_signatures("ventilator")
up <- sig$ensembl[sig$direction == "up"][1:30]
other <- paste0("ENSG", seq_len(30))
regs <- list(TF_vent = list(tfmode = setNames(rep(1, 30), up)),
             TF_other = list(tfmode = setNames(rep(1, 30), other)))
annotate_mra(c(TF_vent = 3.2, TF_other = -1.1), regs)
```

`as_de_table` *Coerce a differential-expression result to a standard table*

Description

Accepts a `DESeqResults` or any `data.frame`-like object from DESeq2, `limma::topTable()` or `edgeR::topTags()` and returns a `DataFrame` with the columns `gene` (Ensembl id without version), `log2FoldChange` and `padj`, keeping the other columns.

Usage

```
as_de_table(de)
```

Arguments

`de`	The result object or table.

Value

A `DataFrame`.

Examples

```
de <- data.frame(
    logFC = c(1.2, -0.8), adj.P.Val = c(0.01, 0.2),
    row.names = c("ENSG00000166189.7", "ENSG00000172409.5")
)
as_de_table(de)
```

audit_report *Plain-text audit report of the control samples*

Description

Summarises, in text and tables that can be pasted into a supplement, the per-sample annotation (thanatome_score()), the selection applied (select_samples()) and the sensitivity of the differential expression to the control subset (compare_controls()).

Usage

```
audit_report(scores, selection = NULL, comparison = NULL, file = NULL)

## S3 method for class 'thanatome_report'
print(x, ...)
```

Arguments

scores	The DataFrame from thanatome_score().
selection	Optional result of select_samples().
comparison	Optional result of compare_controls().
file	Optional path; the report is written there as well.
x	A thanatome_report.
...	Ignored.

Value

An object of class thanatome_report: the character vector with the report lines, printed line by line when shown.

Examples

```
data(thanatome_example)
ctl <- thanatome_example[, thanatome_example$group == "control"]
sc <- thanatome_score(ctl, reference = "ovary")
audit_report(sc, selection = select_samples(sc, preset = "no_agony"))
```

`compare_controls` *Repeat a differential-expression analysis with explicit control subsets*

Description

Scores the controls, applies each selection criterion, repeats the case-versus-control differential expression (edgeR quasi-likelihood) with the controls each criterion keeps, and compares every result with the one obtained with all controls and with `n_random` random subsets of the same size. The random subsets are the yardstick: a criterion that loses or gains no more genes than a random subset of the same size is acting through power, not through the death axes. The function is the packaged form of the sensitivity analysis behind the package and replaces the tempting loop "filter, look at the DE, filter again", which selects on the outcome. Decide the candidate criteria first, run them all once, and choose with their cost in view.

Usage

```
compare_controls(
  x,
  group,
  control = "control",
  criteria = list(),
  scores = NULL,
  reference = NULL,
  n_random = 10,
  padj = 0.05,
  lfc = 1,
  min_controls = 8,
  mra = NULL,
  flags = TRUE,
  assay = "counts",
  lib_size = NULL,
  design = NULL,
  BPPARAM = BiocParallel::SerialParam()
)
```

Arguments

`x`	A `SummarizedExperiment` or an edgeR `DGEList` with raw counts of cases and controls. For a `DGEList` the counts are `x$counts`, the library sizes `x$samples$lib.size` (`norm.factors` are ignored: TMM is recomputed in every fit), `group` may name a column of `x$samples` and `design` refers to the columns of `x$samples`.
`group`	Name of the `colData` column with the group, or a vector.
`control`	Value of `group` that marks the controls.
`criteria`	Named list of criteria; each element is a list of arguments for `select_samples()` (for example `list(preset = "perfused")` or `list(agony = c(-Inf, 0.5))`). The comparison with all controls is always added as `all`.

`scores` Optional `DataFrame` from `thanatome_score()` for the controls; computed from `reference` when `NULL`.

`reference` Reference for `thanatome_score()` when `scores` is `NULL` (a tissue name or the controls themselves). The controls are then scored with the defaults of `thanatome_score()` (for example `normalize = "auto"`); to score them otherwise, pass `scores`.

`n_random` Number of random control subsets of the same size per criterion. Default 10. The subsets are drawn with `sample()`, so the results depend on the random number generator: call `set.seed()` beforehand to reproduce them.

`padj`, `lfc` Thresholds that define a differentially expressed gene.

`min_controls` Minimum controls a criterion must keep. Default 8.

`mra` Optional function taking the DE table of one run (a `DataFrame` with `gene`, `log2FoldChange`, `stat`, `padj`) and returning a named numeric vector of regulator activities (for example a wrapper around viper or NaRnEA). When given, the activities of each criterion are compared with those of `all` by Spearman correlation and overlap of the top 25 regulators. It is always called serially.

`flags` Logical; `TRUE` (default) removes the controls with a technical flag from every subset, including `all` and the random ones, so that all comparisons share the same pool.

`assay`, `lib_size` Passed to the normalisation; see `thanatome_score()`. `x` must hold raw counts (edgeR needs them). With fewer than 5000 genes in `x`, the library sizes are used as given and no TMM factors are computed, because TMM assumes that most genes do not change.

`design` Optional one-sided formula over the sample annotation (`colData(x)`, or `x$samples` for a `DGEList`) with the covariates to include in every DE fit, for example `~ batch + sex`. The case/control indicator is added as the last term and is the tested coefficient; `NULL` (default) fits `~ group`. The covariates should be the same as in your real DE analysis, so that the comparison reproduces it. The function stops if a variable is missing, has missing values, or the design matrix of a fit is not of full column rank (for example a covariate confounded with the group, or constant in the samples of a fit).

`BPPARAM` A BiocParallel::BiocParallelParam object. The edgeR fits (one per criterion plus the random-subset fits) are run with `BiocParallel::bplapply()`. The random subsets are always drawn first, in the main process, so the results do not depend on the backend. Default `BiocParallel::SerialParam()`.

Value

A list with `summary` (one row per criterion: controls kept, DEGs, lost and gained with respect to `all`, the same for the random subsets, Jaccard, standard deviation of the change in log fold change, fraction of DEGs that are signature genes, sign flips, and the MRA agreement when `mra` is given), `de` (the DE table of each criterion), `kept` (the control names kept by each criterion), `scores` and `mra`.

Examples

```
data(thanatome_example)
set.seed(1)
cc <- compare_controls(thanatome_example, group = "group",
                       control = "control",
                       criteria = list(no_agony = list(preset = "no_agony")),
                       reference = "ovary", n_random = 2)
cc$summary

## Covariates of the real DE in every fit, and the fits spread over
## workers (MulticoreParam(4), for example, outside Windows)
x <- thanatome_example
x$batch <- factor(rep(c("a", "b"), length.out = ncol(x)))
set.seed(1)
cc2 <- compare_controls(x, group = "group", control = "control",
    criteria = list(no_agony = list(preset = "no_agony")),
    reference = "ovary", n_random = 2, design = ~ batch,
    BPPARAM = BiocParallel::SerialParam())
cc2$summary[, c("criterion", "n_controls", "n_deg", "lost")]
```

---

contamination_flags *Technical contamination flags per sample*

Description

Flags samples whose mean log2-CPM over the pancreatic enzyme genes (PNLIP, CELA3A, PRSS1, CPA1, CELA2A, CTRB1; contamination by neighbouring tissue, Nieuwenhuis et al. 2020) or over the surfactant genes (SFTPC, SFTPB, SFTPA1, SFTPA2) exceeds an absolute threshold. The threshold is a heuristic on absolute expression, not a z-score against the reference, because the reference tissues do not express these genes and a z-score against a near-zero standard deviation is meaningless. A flag is not raised for the tissue that expresses its genes physiologically: the surfactant flag for lung and the pancreatic flag for pancreas, when `tissue` says so.

Usage

```
contamination_flags(
  x,
  threshold = 4,
  tissue = NULL,
  assay = "counts",
  lib_size = NULL,
  log = FALSE
)
```

Arguments

x	A `SummarizedExperiment` or matrix of raw counts (or log2-CPM with `log = TRUE`), as in `thanatome_score()`.

threshold	Mean log2-CPM above which the flag is raised. Default 4 (about 16 CPM).
tissue	Optional tissue of the samples (case-insensitive; names starting with "lung" disable the surfactant flag and names starting with "pancrea" the pancreatic flag). thanatome_score() passes the tissue of the packaged reference unless told otherwise.
assay, lib_size, log	Passed to the normalisation; see thanatome_score().

Value

A data.frame with logical columns pancreas and surfactant, the mean log2-CPM of each set (pancreas_logcpm, surfactant_logcpm) and the number of genes found (pancreas_n, surfactant_n). When a set has no gene in x the flag is NA and a message says so.

Examples

```
data(thanatome_example)
fl <- contamination_flags(thanatome_example)
table(fl$pancreas, useNA = "ifany")
```

---

death_covariates *Per-sample death scores as covariates for designs and batch correction*

Description

Returns a numeric matrix (one row per sample, one column per axis) ready to be passed as covariates to limma::removeBatchEffect(), added to a model.matrix() for limma, edgeR or DESeq2, or given to sva as known surrogate variables. ComBat takes only categorical batches; the state label can be used there, but a continuous adjustment is the honest use of these scores.

Usage

```
death_covariates(
  scores,
  axes = c("ventilator", "agony"),
  center = TRUE,
  quiet = FALSE
)
```

Arguments

scores	The DataFrame returned by thanatome_score().
axes	Columns to export (default the two axes). The ischemia component can be added explicitly, but it is nearly collinear with the ventilator score (about -0.8 in GTEx) and only informative among perfused samples.
center	Logical; centre each column (default TRUE).
quiet	Logical; TRUE suppresses the cautions.

Details

Two cautions are printed on each call unless `quiet = TRUE`. First, the axes vary only among the post-mortem samples: in a tumour-versus-control design every tumour sits at the same (undefined) value, so adjusting across groups partially confounds the death axes with the group effect; adjust within controls, or include the covariates in a design where the group term is kept. Second, the adjustment centres the controls at their own mean, not at the living pole: it removes the variation between donors, not the distance between post-mortem and living tissue.

Value

A numeric matrix with the samples in rows.

Examples

```
data(thanatome_example)
ctl <- thanatome_example[, thanatome_example$group == "control"]
sc <- thanatome_score(ctl, reference = "ovary")
cv <- death_covariates(sc)
head(cv)
## e.g. limma::removeBatchEffect(logcpm, covariates = cv)
```

`death_gene_sets` *Export the death signatures as gene sets*

Description

Returns the consensus signatures as gene sets for the gene-set tools of Bioconductor (for example `fgsea`, `limma::camera()`, `GSVA`, `singscore` or `clusterProfiler`). By default every axis is split by direction, because the signatures are signed: `ventilator_up` holds the genes higher in ventilated donors (Hardy 0) than in the rest, `agony_up` the genes higher in slow deaths (Hardy 3-4) than in fast deaths (Hardy 1-2), and `ischemia_up` the genes that rise with post-mortem ischemic time. With `split = FALSE` each axis becomes a single unsigned set, suitable only for tests that ignore direction.

Usage

```
death_gene_sets(
  axes = c("ventilator", "agony"),
  id = c("ensembl", "symbol"),
  split = TRUE,
  format = c("list", "GeneSetCollection")
)
```

Arguments

`axes`	Axes to export: `"ventilator"`, `"agony"` and/or `"ischemia"`. Default the two axes.
`id`	Gene identifiers: `"ensembl"` (unversioned, default) or `"symbol"` (genes without a symbol are dropped).

split	Logical; TRUE (default) gives one set per axis and direction (<axis>_up, <axis>_down), FALSE one set per axis.
format	"list" (default, a named list of character vectors) or "GeneSetCollection" (package GSEABase).

Value

The gene sets in the requested format.

See Also

death_signatures(), annotate_de()

Examples

```
gs <- death_gene_sets()
lengths(gs)
death_gene_sets("ischemia", id = "symbol")$ischemia_up[1:5]
if (requireNamespace("GSEABase", quietly = TRUE)) {
    death_gene_sets(format = "GeneSetCollection")
}
```

---

death_reference *Access the GTEx reference of one tissue*

Description

Access the GTEx reference of one tissue

Usage

```
death_reference(tissue)
```

Arguments

tissue	One of names(thanatome_references).

Value

The reference list of that tissue (mean, sd, expressed, poles, hardy, ..., and qn with the statistics after quantile normalisation); see thanatome_references.

Examples

```
death_reference("ovary")$poles
```

death_signatures *Access the consensus death signatures*

Description

Access the consensus death signatures

Usage

```
death_signatures(axes = c("ventilator", "agony", "ischemia"))
```

Arguments

axes	Which signatures to return: "ventilator", "agony" and/or "ischemia" (the ischemia component of the ventilator axis; see thanatome_signatures).

Value

A data.frame with one row per gene and signature (columns ensembl, symbol, axis, lfc_consensus, z_stouffer, n_tissues, direction, in_ventilator).

Examples

```
head(death_signatures("agony"))
```

hardy_layers *Map the GTEx Hardy scale to the death axes*

Description

Map the GTEx Hardy scale to the death axes

Usage

```
hardy_layers(hardy)
```

Arguments

hardy	Vector of Hardy scale values (0-4; 0 = ventilator case, 1 = violent and fast death, 2 = fast death of natural causes, 3 = intermediate, 4 = slow death). Integers, numerics, characters or a factor with those labels are accepted.

Value

A data.frame with hardy as integer, the logical labels ventilator (Hardy 0) and agony (Hardy 3-4 versus 1-2; NA for ventilator cases, which were not used for that contrast) and the state label (perfused, postmortem_rapid, postmortem_agony).

Examples

```
hardy_layers(c(0, 1, 2, 3, 4, NA))
hardy_layers(factor(c("0", "4")))
```

`plot_thanatome` *Plot the per-sample death scores*

Description

`type = "scatter"` draws the samples on two axes (by default the scaled positions on the ventilator and agony axes) with the reference poles at 0 and 1, coloured by state label; `type = "ranked"` draws every sample as a point on each axis, ordered by position, with the state threshold, so that individual samples can be found. Both return a `ggplot` object that can be modified further.

Usage

```
plot_thanatome(
  scores,
  type = c("scatter", "ranked"),
  x = "position_ventilator",
  y = "position_agony",
  colour = "state",
  hardy = NULL,
  label = FALSE
)
```

Arguments

`scores`	The `DataFrame` returned by `thanatome_score()`.
`type`	`"scatter"` or `"ranked"`.
`x, y`	Column names used by the scatter plot.
`colour`	Column used for colour (default `state`); `NULL` for none.
`hardy`	Optional vector of Hardy scale values (same order as `scores`) drawn as point shape, to see the agreement with the label.
`label`	Logical; `TRUE` writes the sample names next to the points in the scatter plot (useful for small cohorts); ignored by `"ranked"`.

Value

A `ggplot` object.

Examples

```
data(thanatome_example)
ctl <- thanatome_example[, thanatome_example$group == "control"]
sc <- thanatome_score(ctl, reference = "ovary")
plot_thanatome(sc, hardy = ctl$hardy)
plot_thanatome(sc, type = "ranked")
```

`select_samples` *Select samples by explicit criteria on the death axes*

Description

Returns the samples whose scaled positions, ischemia component, state label and technical flags satisfy the criteria given, together with a table of what was excluded and why. No criterion on the death axes is applied by default (technical flags are excluded unless `flags = FALSE`): the analyses behind this package found that no subset of post-mortem controls is closer to surgical tissue than the full set except, slightly, the ventilated donors, and that excluding samples costs power. Named presets are offered for the two criteria the evidence does support as explicit choices; each is documented with its cost.

Usage

```
select_samples(
  scores,
  ventilator = NULL,
  agony = NULL,
  ischemia = NULL,
  state = NULL,
  flags = TRUE,
  preset = c("none", "perfused", "no_agony")
)
```

Arguments

`scores`	The `DataFrame` returned by `thanatome_score()`.
`ventilator, agony`	Numeric length-2 intervals (closed) on `position_ventilator` and `position_agony`; `NULL` means no criterion. The agony axis is defined only among non-perfused donors, so the `agony` interval is applied to non-perfused samples only; samples labelled `perfused` pass it.
`ischemia`	Numeric length-2 interval on the raw ischemia component (meaningful among perfused samples only).
`state`	Character vector of allowed state labels.
`flags`	Logical; `TRUE` excludes samples with any technical flag.
`preset`	`"none"` (default), `"perfused"` (keep only samples labelled `perfused`: closest to living tissue, but in the GTEx references it discards the non-ventilated donors, 22-62 percent of the samples depending on the tissue) or `"no_agony"` (exclude samples labelled `postmortem_agony`: the only axis that displaces controls coherently with respect to adjacent normal tissue, at a cost of the Hardy 3-4 donors, 9-17 percent of the samples, with a label that, built from the joint-regression positions in tissue not seen by the signatures, was right for 61-84 percent of samples). `"no_agony"` keeps every other sample (including those without a label)

and warns, excluding nothing on agony, when no sample carries an agony label (agony axis missing). A preset only acts when `state` is `NULL`; the other criteria are combined with it.

Value

A list with `keep` (sample names retained), `excluded` (a `DataFrame` with the sample and the first criterion it failed) and `summary` (counts per criterion).

Examples

```
data(thanatome_example)
ctl <- thanatome_example[, thanatome_example$group == "control"]
sc <- thanatome_score(ctl, reference = "ovary")
sel <- select_samples(sc, agony = c(-Inf, 0.5))
sel$summary
head(sel$excluded)
```

---

thanatome_example *Example data: GTEx ovary controls and TCGA-OV tumours*

Description

A `SummarizedExperiment` with raw read counts of the signature and flag genes for 74 GTEx v8 ovary samples (stratified by Hardy scale) and 40 TCGA-OV primary tumours, both from recount3 (same Monorail pipeline), with the full library size (`lib_size`, needed because only a subset of genes is kept), `group` (`control` or `tumor`), donor, Hardy scale, ischemic time and RIN in `colData`. The GTEx samples entered the consensus, the references and the poles, so any accuracy measured on them is in-sample. Built by `inst/scripts/build_example.R`.

Usage

```
data(thanatome_example)
```

Format

A `SummarizedExperiment` with 1663 genes and 114 samples.

Source

recount3 projects OVARY (GTEx v8, RNASEQ freeze) and OV (TCGA primary tumours), gene read counts, GENCODE v26.

Examples

```
data(thanatome_example)
table(thanatome_example$group, thanatome_example$hardy, useNA = "ifany")
```

`thanatome_references` *Per-tissue GTEx references for centring new samples*

Description

For each of the ten GTEx v8 tissues used to derive the signatures: the mean and standard deviation of log2-CPM (prior count 1, full library size) of every signature and flag gene across the GTEx samples of that tissue, which of those genes are expressed there (mean log2-CPM above 0; only these are used for scoring), and the poles of each axis (median signature score of the ventilated and non-ventilated donors for `ventilator`; of fast and slow deaths for `agony`) that define the scaled position returned by `thanatome_score()`. Counts come from recount3 (Monorail pipeline, GENCODE v26); the references apply to data quantified the same way. Built by `inst/scripts/build_references.R`.

Usage

```
data(thanatome_references)
```

Format

A named list, one element per tissue (`ovary`, `breast`, `thyroid`, `lung`, `colon_sigmoid`, `colon_transverse`, `liver`, `prostate`, `uterus`, `skin_leg`), each a list with `tissue`, `gtex_name`, `pipeline`, `n_samples`, `hardy` (table of samples by Hardy scale), `mean`, `sd`, `expressed` (named vectors over Ensembl ids), `poles` (data.frame), `ischemic_time_range` (minutes) and `qn`. `qn` holds what `thanatome_score(normalize = "quantile")` uses: `genes` (the gene universe, genes with mean log2-CPM above 0 in the tissue), `target` (the target distribution as 2000 quantiles of the mean sorted log2-CPM profile), and `mean`, `sd`, `expressed` and `poles` computed as above after quantile normalisation of the samples to `target`, over the signature and flag genes that belong to `genes`.

Source

recount3 (https://rna.recount.bio; Wilks C et al. (2021) Genome Biology 22:323 doi:10.1186/s13059021025336), GTEx projects OVARY, BREAST, THYROID, LUNG, COLON, LIVER, PROSTATE, UTERUS and SKIN, RNASEQ freeze samples, one run per sample.

Examples

```
data(thanatome_references)
names(thanatome_references)
thanatome_references$ovary$poles
```

thanatome_score *Annotate each sample with its position along the death axes*

Description

For every sample, computes a signature score per death axis (mean of the per-gene z-scores against the reference, with the sign of the consensus), the position scaled between the two poles of the reference tissue (0 = ventilated donors / fast deaths, 1 = non-ventilated donors / slow deaths; values outside 0-1 are allowed and meaningful: living tissue falls below 0 on the ventilator axis), a state label and technical contamination flags. With `method = "regression"` it adds the coefficients of the joint regression of the sample's z profile on the consensus vectors and the fraction of the sample's deviation from the tissue centroid explained by the axes.

Usage

```
thanatome_score(
  x,
  reference,
  axes = c("ventilator", "agony"),
  method = c("signature", "regression"),
  assay = "counts",
  log = FALSE,
  lib_size = NULL,
  normalize = c("auto", "quantile", "none"),
  drop_unexpressed = TRUE,
  tissue = NULL,
  min_pole = 8,
  state_threshold = 0.5,
  flag_threshold = 4
)
```

Arguments

x	A `SummarizedExperiment` with raw counts in `assay(x, assay)`, an edgeR `DGEList` (its library sizes are used; normalisation factors are ignored, as in the references), a limma `EList` (its log2-CPM `E` is used) or a numeric matrix of raw counts (genes in rows, Ensembl ids with or without version, unique sample names). Set `log = TRUE` if the values are already log2-CPM (`edgeR::cpm(log = TRUE, prior.count = 1)` on the full library size; edgeR scales the prior count by the library size, so lowly expressed genes depend slightly on the depth of the cohort); VST or rlog values are not comparable with the reference.
reference	Either the name of a GTEx tissue (see `names(thanatome_references)`), or a `SummarizedExperiment` / count matrix of the same tissue to be used as cohort reference.
axes	Axes to score: `"ventilator"`, `"agony"` or both (default). The ischemia component is always reported.

`method` `"signature"` (default) or `"regression"`. The regression adds `beta_ventilator`, `beta_agony`, `r2_signature` (over the signature genes) and `r2_total` (over all genes; `NA` unless both the reference, as a cohort, and `x` carry at least 10 000 genes). Positions and labels always come from the signature score.

`assay` Name of the assay with counts when `x` is a `SummarizedExperiment`.

`log` Logical; `TRUE` if `x` already holds log2-CPM values.

`lib_size` Optional library sizes of the samples of `x` (one per column) for the CPM normalisation, needed when `x` holds only a subset of the genes (as the example data do). If `NULL`, a `lib_size` column of `colData(x)` is used when present, and the column sums otherwise; a warning is given when `x` has fewer than 10 000 genes and no library sizes. A cohort reference always uses its own `lib_size` column or its column sums.

`normalize` `"auto"` (default), `"quantile"` or `"none"`. `"quantile"` quantile-normalises the samples to the distribution of the reference before scoring (see *External cohorts*). It needs the unfiltered matrix: at least 10 000 of the genes expressed in the reference must be in `x` (error otherwise), and a warning is given when `x` holds less than 95% of them, as after filtering lowly expressed genes. `"auto"` uses it when both conditions hold, with a message, and `"none"` otherwise (subsets of genes, such as the example data, or filtered matrices). The choice made is stored in `metadata(result)$normalize`.

`drop_unexpressed` Logical; `TRUE` (default) leaves out the genes absent from `x`: zero counts in more than half of its samples (for log input, log2-CPM below log2(0.1) in more than half), when `x` has at least 5 samples (with fewer, an absent gene cannot be told apart from a lowly expressed one and all genes are kept). The criterion is far from the reference's own expression cut, so a sample's score does not depend on the other samples through it. The genes left out are listed in `metadata()$unexpressed_signature_genes`.

`tissue` Optional tissue of the samples, used only to switch off the contamination flag that is physiological there (surfactant in lung, pancreatic enzymes in pancreas); by default the tissue of the packaged reference, and none for a cohort reference.

`min_pole` For a cohort reference, minimum number of samples in each pole. Poles with fewer samples, or whose score separates the two Hardy groups with an internal AUC below 0.6, are not used: positions and state labels of that axis are `NA`, with a warning, and `metadata()$poles` records the counts, the AUC and `ok`. Default 8.

`state_threshold` Position above which a sample is labelled non-perfused (ventilator axis) or in agony (agony axis). Default 0.5.

`flag_threshold` Mean log2-CPM above which a contamination flag is raised; see `contamination_flags()`.

Details

Reference. A reference of the same tissue is mandatory: a sample has no position on its own, since z-scores are computed per gene against a centroid. The packaged references (`names(thanatome_references)`) were built from GTEx v8 counts quantified by the recount3 Monorail pipeline (GENCODE v26)

and apply to data quantified the same way; for another pipeline, platform or protocol pass your own cohort (a `SummarizedExperiment` or count matrix, at least 20 samples) as `reference`: the centroid and, if its `colData` carries a `hardy` (or `DTHHRDY`) column, the poles are estimated from it. Without Hardy the positions and labels are `NA`. Scores are rank-stable across pipelines; positions are not.

External cohorts. Against a packaged reference, a cohort quantified or sequenced differently carries a batch effect. Two safeguards apply. Genes absent from `x` (zero counts in most of its samples) are left out by default, since an absent gene is not a down-regulated one (`drop_unexpressed`). When `x` holds the whole, unfiltered transcriptome, the samples are first quantile-normalised to the distribution of the reference (the GTEx tissue or the cohort) and scored against its equally normalised statistics and poles (`normalize = "auto"`, the default). Without it, autopsies of external cohorts could land on the perfused side of the ventilator axis; within GTEx it barely changes the scores. In external GTEx-independent cohorts (living colon biopsies, accident autopsies) the side of the poles on which a quantile-normalised cohort falls was stable across gene filters and batch handling, whereas its distance beyond the poles was not: read positions outside 0-1 as a direction, not a magnitude.

Ischemia. Ischemia is not a third axis. Its signature (86 genes) is mostly a subset of the ventilator signature with the opposite sign, and only 10-23 of its genes were expressed per tissue in the analysis (15-29 in the packaged references). It is reported as the column `ischemia`, the component of the ventilator axis sensitive to post-mortem ischemic time, and it tracks ischemic time only among perfused (ventilated) donors (partial Spearman correlation with ischemic time given the ventilator score 0.20-0.50 within GTEx Hardy 0). It never enters the joint regression as a coefficient.

Reliability. With signatures derived without the scored tissue, the ventilator score recovers the Hardy 0 label with AUC 0.87-0.94 and the agony score the Hardy 3-4 label with AUC 0.69-0.90. The state label built from the joint-regression positions, in tissue not seen by the signatures, was right for 61-84 percent of samples against 45-80 percent for the majority class; its weak side is rapid versus agony.

Extreme genes. The per-gene z-scores are not truncated. Clipping them at +/-5 or +/-3, or replacing the mean by the median, left the leave-one-tissue-out AUC unchanged within GTEx (maximum |z| 7.4) and did not improve the discrimination of external cohorts (|z| up to 22); clipping at 3 only compressed the positions beyond the poles, and the median lost part of the ischemia signal.

Value

A `DataFrame` (one row per sample) with the signature scores (`ventilator`, `agony`, `ischemia`), `position_ventilator`, `position_agony`, `state` (`perfused`, `postmortem_rapid`, `postmortem_agony`, or `postmortem` when the agony axis is not available), `flag_pancreas`, `flag_surfactant`, the regression columns when requested, and `n_genes` used per score. The raw `ventilator` score grows towards the perfused pole; the position inverts it so that 1 is post-mortem. `metadata()` records the reference, the method, the poles and, when `x` carries a `hardy` column, the AUC of each score against the Hardy label and the agreement of the state label, as an internal check (in-sample if the samples are GTEx).

See Also

`select_samples()`, `plot_thanatome()`, `death_covariates()`, `compare_controls()`

Examples

```
data(thanatome_example)
ctl <- thanatome_example[, thanatome_example$group == "control"]
sc <- thanatome_score(ctl, reference = "ovary")
head(sc)
S4Vectors::metadata(sc)$auc
```

`thanatome_signatures` *Consensus death signatures from GTEx v8*

Description

Genes with a concordant sign of differential expression in at least 5 of 10 GTEx v8 tissues for each of three contrasts within the GTEx donors: `ventilator` (Hardy scale 0, death on a ventilator, versus Hardy 1-4; |median log2 fold change| > 0.5), `agony` (slow deaths, Hardy 3-4, versus fast deaths, Hardy 1-2, among non-ventilated donors; > 0.5) and `ischemia` (post-mortem ischemic time as a continuous covariate, adjusted for ventilator; > 0.25 because the contrast is per standard deviation of ischemic time). Ischemia is not a third axis: 61 of its 86 genes belong to the ventilator signature with the opposite sign (`in_ventilator`) and only 10-23 were expressed per tissue in the analysis (15-29 in the packaged recount3 references); it is the component of the ventilator axis sensitive to ischemic time. Built by `inst/scripts/build_signatures.R` (see `system.file("scripts", package = "thanatome")`).

Usage

```
data(thanatome_signatures)
```

Format

A list with four elements:

signatures data.frame with one row per gene and signature: `ensembl` (without version), `symbol`, `axis`, `lfc_consensus` (median log2 fold change across tissues), `z_stouffer`, `n_tissues` (concordant tissues), `direction` (`up` or `down`) and `in_ventilator`.

flags data.frame of technical flag genes: pancreatic enzymes (contamination by neighbouring tissue) and surfactant proteins.

thresholds `min_tissues` and the per-axis `min_lfc`.

source one line describing the origin.

Source

GTEx v8 gene read counts from the GTEx portal (https://gtexportal.org, dbGaP phs000424.v8), ten tissues: ovary, breast, thyroid, lung, sigmoid colon, transverse colon, liver, prostate, uterus and skin (lower leg); DESeq2 per tissue and Stouffer consensus, analysis repository https://github.com/hachepunto/thanatome-manuscript.

References

GTEx Consortium (2020) Science 369:1318-1330 doi:10.1126/science.aaz1776; Ferreira PG et al. (2018) Nature Communications 9:490 doi:10.1038/s4146701702772x; Nieuwenhuis TO et al. (2020) Nature Communications 11:1933 doi:10.1038/s41467020158219.

Examples

```
data(thanatome_signatures)
sg <- thanatome_signatures$signatures
table(sg$axis, sg$direction)
```

Index

∗ **datasets**
    thanatome_example, 18
    thanatome_references, 19
    thanatome_signatures, 23

add_thanatome_score, 4
annotate_de, 5
annotate_de(), *14*
annotate_mra, 6
as_de_table, 7
as_de_table(), *5*
audit_report, 8

BiocParallel::BiocParallelParam, *10*
BiocParallel::bplapply(), *10*
BiocParallel::SerialParam(), *10*

compare_controls, 9
compare_controls(), *8*, *22*
contamination_flags, 11
contamination_flags(), *21*

death_covariates, 12
death_covariates(), *22*
death_gene_sets, 13
death_reference, 14
death_signatures, 15
death_signatures(), *6*, *14*

hardy_layers, 15

plot_thanatome, 16
plot_thanatome(), *22*
print.thanatome_report *(*audit_report*)*, 8

sample(), *10*
select_samples, 17
select_samples(), *8*, *9*, *22*
set.seed(), *10*

thanatome *(*thanatome-package*)*, 3
thanatome-package, 3
thanatome_example, 18
thanatome_references, *14*, 19
thanatome_score, 20
thanatome_score(), *3*, *4*, *8*, *10–12*, *16*, *17*, *19*
thanatome_signatures, *15*, 23