Package: STACAS (via r-universe)

May 27, 2026

Type Package

Title STACAS: Sub-Type Anchoring Correction for Alignment in Seurat

Version 2.4.1

Description This package implements methods for batch correction and integration of scRNA-seq datasets, based on the popular Seurat anchor-based integration framework. In particular, STACAS is optimized for the integration of heterogenous datasets with only limited overlap between cell sub-types (e.g. TIL sets of CD8 from tumor with CD8/CD4 T cells from lymphnode), for which the default Seurat alignment methods would tend to over-correct biological differences. The 2.0 version of our package allows to the users to incorporate explicit information about cell-types in order to assist the integration process.

Depends R (>= 4.0.0)

Imports Seurat(>= 5.0.0), SeuratObject(>= 5.0.0), data.table, pbapply, ggridges, colorspace, R.utils, ggplot2, stats, BiocNeighbors, BiocParallel

Suggests testthat (>= 3.0.0), sf

License GPL-3 + file LICENSE

Encoding UTF-8

LazyData false

RoxygenNote 7.3.2

Config/testthat/edition 3

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

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

Date/Publication 2026-01-27 09:46:28 UTC

RemoteUrl https://github.com/carmonalab/STACAS

RemoteRef HEAD

RemoteSha ea34e824e17df6316823db0a9b6322af0833e501

Contents

`annotate.by.neighbors` *Annotate by neighbors*

Description

Given a partially annotated dataset, propagate labels to un-annotated cells (NA values) by similarity with annotated cells. This can be useful after integration of fully annotated datasets with other dataset that lack cell type annotation. Propagation of labels is done by K-nearest neighbors with annotated cells in a given dimensionality reduction (e.g. PCA space).

Usage

```
annotate.by.neighbors(
  obj,
  ref.cells = NULL,
  reduction = "pca",
  ndim = NULL,
  k = 20,
  ncores = 1,
  bg.pseudocount = 10^9,
  labels.col = "functional.cluster"
)
```

Arguments

`obj`	A Seurat object
`ref.cells`	Barcode of the cells to be used as reference to annotate all remaining cells. By default uses all annotated cells as reference (i.e. all cells with metadata column 'labels.col != NA').
`reduction`	Dimensionality reduction to be used for knn calculation

ndim	Number of dimensions to use in given reduction (by default use all dimensions)
k	Number of nearest neighbors for knn calculation
ncores	Number of cores for multi-thread execution
bg.pseudocount	Background counts for cell type frequency estimation
labels.col	Metadata column that stores cell type annotations to be propagated

Value

Returns a Seurat object with standard gene names. Genes not found in the standard list are removed. Synonyms are accepted when the conversion is not ambiguous.

Examples

```
# Fully annotate object, where partial annotations are stored in metadata column "celltype"
## Not run:
obj.full <- annotate.by.neighbors(obj.partial, labels.col="celltype")

## End(Not run)
```

EnsemblGeneTable.Hs *Standardized gene list from ENSEMBL (human)*

Description

A reference of stable gene names for Homo Sapiens

Usage

```
EnsemblGeneTable.Hs
```

Format

A dataframe of ENSEMBL and gene symbols

Source

https://www.ensembl.org/Homo_sapiens/Info/Index

EnsemblGeneTable.Mm *Standardized gene list from ENSEMBL (mouse)*

Description

A reference of stable gene names for Mus Musculus

Usage

```
EnsemblGeneTable.Mm
```

Format

A dataframe of ENSEMBL and gene symbols

Source

https://www.ensembl.org/Mus_musculus/Info/Index

FindAnchors.STACAS *Find integration anchors using STACAS*

Description

This function computes anchors between datasets for single-cell data integration. It is based on the Seurat function `FindIntegrationAnchors`, but is optimized for integration of heterogenous data sets containing only partially overlapping cells subsets. It also computes a measure of distance between candidate anchors (rPCA), which is combined with the Seurat's anchor weight by the factor `alpha`. Prior knowledge about cell types can optionally be provided to guide anchor finding. Give this information in the `cell.labels` metadata column. This annotation level, which can be incomplete (set to NA for cells of unknown type), is used to penalize anchor pairs with inconsistent annotation. The set of anchors returned by this function can then be passed to `IntegrateData.STACAS` for dataset integration.

Usage

```
FindAnchors.STACAS(
  object.list = NULL,
  assay = NULL,
  reference = NULL,
  min.sample.size = 100,
  max.seed.objects = 10,
  anchor.features = 1000,
  genesBlockList = "default",
  dims = 30,
  k.anchor = 5,
```

```
  k.score = 30,
  alpha = 0.8,
  anchor.coverage = 0.5,
  correction.scale = 2,
  cell.labels = NULL,
  label.confidence = 1,
  scale.data = FALSE,
  seed = 123,
  verbose = TRUE
)
```

Arguments

object.list	A list of Seurat objects. Anchors will be determined between pairs of objects, and can subsequently be used for Seurat dataset integration.
assay	A vector containing the assay to use for each Seurat object in object.list. If not specified, uses the default assay.
reference	A vector specifying the object/s to be used as a reference during integration. If NULL (default), all pairwise anchors are found (no reference/s). If not NULL, the corresponding objects in object.list will be used as references. When using a set of specified references, anchors are first found between each query and each reference. The references are then integrated through pairwise integration. Each query is then mapped to the integrated reference.
min.sample.size	Minimum number of cells per sample. Objects with fewer than this number of cells are not integrated.
max.seed.objects	Number of objects to use as seeds to build the integration tree. Automatically chooses the largest max.seed.objects datasets; the remaining datasets will be added sequentially to the reference.
anchor.features	Can be either: • A numeric value. This will call FindVariableFeatures.STACAS to identify anchor.features that are consistently variable across datasets • A pre-calculated vector of integration features to be used for anchor search.
genesBlockList	If anchor.features is numeric, genesBlockList optionally takes a (list of) vectors of gene names. These genes will be removed from the integration features. If set to "default", STACAS uses its internal list data("genes.blocklist"). This is useful to mitigate effect of genes associated with technical artifacts or batch effects (e.g. mitochondrial, heat-shock response).
dims	The number of dimensions used for PCA reduction
k.anchor	The number of neighbors to use for identifying anchors
k.score	The number of neighbors to use for scoring anchors
alpha	Weight on rPCA distance for rescoring (between 0 and 1).
anchor.coverage	Center of logistic function, based on quantile value of rPCA distance distribution

correction.scale
: Scale factor for logistic function (multiplied by SD of rPCA distance distribution)

cell.labels
: A metadata column name, storing cell type annotations. These will be taken into account for semi-supervised alignment (optional). Note that not all cells need to be annotated - please set unannotated cells as NA or 'unknown' for this column. Cells with NA or 'unknown' cell labels will not be penalized in semi-supervised alignment.

label.confidence
: How much you trust the provided cell labels (from 0 to 1).

scale.data
: Whether to rescale expression data before PCA reduction.

seed
: Random seed for probabilistic anchor acceptance

verbose
: Print all output

Value

Returns an AnchorSet object, which can be passed to `IntegrateData.STACAS`

Examples

```
data(sampleObj)
library(Seurat)
obj.list <- SplitObject(sampleObj, split.by="donor")
anchors <- FindAnchors.STACAS(obj.list, min.sample.size=10, k.score=5, dims=3)
```

`FindVariableFeatures.STACAS`

FindVariableFeatures.STACAS

Description

Select highly variable genes (HVG) from an expression matrix. Genes from a blocklist (e.g. cell cycling genes, mitochondrial genes) can be excluded from the list of variable genes, as well as genes with very low or very high average expression

Usage

```
## S3 method for class 'STACAS'
FindVariableFeatures(
  obj,
  nfeat = 1500,
  genesBlockList = "default",
  min.exp = 0.01,
  max.exp = 3
)
```

Arguments

obj	A Seurat object containing an expression matrix
nfeat	Number of top HVG to be returned
genesBlockList	Optionally takes a list of vectors of gene names. These genes will be removed from initial HVG set. If set to "default", STACAS uses its internal list `data("genes.blocklist")`. This is useful to mitigate effect of genes associated with technical artifacts or batch effects (e.g. mitochondrial, heat-shock response). If set to 'NULL' no genes will be excluded
min.exp	Minimum average normalized expression for HVG. If lower, the gene will be excluded
max.exp	Maximum average normalized expression for HVG. If higher, the gene will be excluded

Value

Returns a list of highly variable genes

Examples

```
data(sampleObj)
hvg <- FindVariableFeatures.STACAS(sampleObj)
```

---

genes.blocklist *Genes blocklists for excluding HVGs*

Description

A list of gene signatures, including cycling, heat-shock response, mitochondrial and risobomal genes, interferon response; for mouse and human. Derived from the SignatuR R package: https://github.com/carmonalab/SignatuR

Usage

```
genes.blocklist
```

Format

A list of gene signatures

Source

https://github.com/carmonalab/SignatuR

IntegrateData.STACAS *IntegrateData.STACAS*

Description

Integrate a list of datasets using STACAS anchors. Based on the IntegrateData function from Seurat. This function requires that you have calculated a set of integration anchors using FindAnchors.STACAS. To perform semi-supervised integration, run FindAnchors.STACAS with cell type annotations labels. Integration anchors with inconsistent cell type will be excluded from integration, providing an integrated space that is partially guided by prior information.

Usage

```
IntegrateData.STACAS(
  anchorset,
  new.assay.name = "integrated",
  features.to.integrate = NULL,
  dims = 30,
  k.weight = 100,
  sample.tree = NULL,
  hclust.method = c("single", "complete", "ward.D2", "average"),
  semisupervised = TRUE,
  verbose = TRUE
)
```

Arguments

anchorset	A set of anchors calculated using FindAnchors.STACAS
new.assay.name	Assay to store the integrated data
features.to.integrate	Which genes to include in the corrected integrated space (def. variable genes)
dims	Number of dimensions for local anchor weighting
k.weight	Number of neighbors for local anchor weighting. Set k.weight="max" to disable local weighting
sample.tree	Specify the order of integration. See SampleTree.STACAS to calculate an integration tree.
hclust.method	Clustering method for integration tree (single, complete, average, ward)
semisupervised	Whether to use cell type label information (if available)
verbose	Print progress bar and output

Value

Returns a Seurat object with a new integrated Assay, with batch-corrected expression values

Examples

```
data(sampleObj)
library(Seurat)
obj.list <- SplitObject(sampleObj, split.by="donor")
anchors <- FindAnchors.STACAS(obj.list, min.sample.size=10, k.score=5, dims=3)
integrated <- IntegrateData.STACAS(anchors, dims=3)
```

PlotAnchors.STACAS *PlotAnchors.STACAS*

Description

Plot distribution of rPCA distances between pairs of datasets

Usage

```
PlotAnchors.STACAS(ref.anchors = NULL, obj.names = NULL, anchor.coverage = 0.5)
```

Arguments

ref.anchors	A set of anchors calculated using FindAnchors.STACAS, containing the pairwise distances between anchors.
obj.names	Vector of object names, one for each dataset in ref.anchors
anchor.coverage	Quantile of rPCA distance distribution

Value

A plot of the distribution of rPCA distances

Run.STACAS *Run the STACAS integration pipeline*

Description

This function is a wrapper for running the several steps required to integrate single-cell datasets using STACAS: 1) Finding integration anchors; 2) Calculating the sample tree for the order of dataset integration; 3) Dataset batch effect correction and integration

Usage

```
Run.STACAS(
  object.list = NULL,
  assay = NULL,
  new.assay.name = "integrated",
  reference = NULL,
  max.seed.objects = 10,
  min.sample.size = 100,
  anchor.features = 1000,
  genesBlockList = "default",
  dims = 30,
  k.anchor = 5,
  k.score = 30,
  k.weight = 100,
  alpha = 0.8,
  anchor.coverage = 0.5,
  correction.scale = 2,
  cell.labels = NULL,
  label.confidence = 1,
  scale.data = FALSE,
  hclust.method = c("single", "complete", "ward.D2", "average"),
  seed = 123,
  verbose = FALSE
)
```

Arguments

`object.list`	A list of Seurat objects. Anchors will be determined between pairs of objects, and can subsequently be used for Seurat dataset integration.
`assay`	A vector containing the assay to use for each Seurat object in object.list. If not specified, uses the default assay.
`new.assay.name`	Assay to store the integrated data
`reference`	A vector specifying the object/s to be used as a reference during integration. If NULL (default), all pairwise anchors are found (no reference/s). If not NULL, the corresponding objects in `object.list` will be used as references. When using a set of specified references, anchors are first found between each query and each reference. The references are then integrated through pairwise integration. Each query is then mapped to the integrated reference.
`max.seed.objects`	Number of objects to use as seeds to build the integration tree. Automatically chooses the largest max.seed.objects datasets; the remaining datasets will be added sequentially to the reference.
`min.sample.size`	Minimum number of cells per sample. Objects with fewer than this number of cells are not integrated.
`anchor.features`	Can be either:

- A numeric value. This will call `Seurat::SelectIntegrationFeatures` to identify `anchor.features` genes for anchor finding.
- A pre-calculated vector of integration features to be used for anchor search.

`genesBlockList` If `anchor.features` is numeric, `genesBlockList` optionally takes a list of vectors of gene names. These genes will be removed from the integration features. If set to "default", STACAS uses its internal list `data("genes.blocklist")`. This is useful to mitigate effect of genes associated with technical artifacts or batch effects (e.g. mitochondrial, heat-shock response).

`dims` The number of dimensions used for PCA reduction

`k.anchor` The number of neighbors to use for identifying anchors

`k.score` The number of neighbors to use for scoring anchors

`k.weight` Number of neighbors for local anchor weighting. Set `k.weight="max"` to disable local weighting

`alpha` Weight on rPCA distance for rescoring (between 0 and 1).

`anchor.coverage`
Center of logistic function, based on quantile value of rPCA distance distribution

`correction.scale`
Scale factor for logistic function (multiplied by SD of rPCA distance distribution)

`cell.labels` A metadata column name, storing cell type annotations. These will be taken into account for semi-supervised alignment (optional). Cells annotated as NA or NULL will not be penalized in semi-supervised alignment

`label.confidence`
How much you trust the provided cell labels (from 0 to 1).

`scale.data` Whether to rescale expression data before PCA reduction.

`hclust.method` Clustering method for integration tree (single, complete, average, ward)

`seed` Random seed for probabilistic anchor acceptance

`verbose` Print all output

Value

Returns a `Seurat` object with a new integrated Assay. Also, centered, scaled variable features data are returned in the scale.data slot, and the pca of these batch-corrected scale data in the pca 'reduction' slot

Examples

```
data(sampleObj)
library(Seurat)
obj.list <- SplitObject(sampleObj, split.by="donor")
integrated <- Run.STACAS(obj.list, min.sample.size=10, k.score=5, dims=3)
```

sampleObj *Sample dataset to test STACAS installation*

Description

A Seurat object containing single-cell transcriptomes (scRNA-seq) for 50 cells and 20729 genes. Single-cell UMI counts were normalized using a standard log-normalization: counts for each cell were divided by the total counts for that cell and multiplied by 10,000, then natural-log transformed using 'log1p'.

This a subsample of 25 predicted B cells and 25 predicted NK cells from the large scRNA-seq PBMC dataset published by Hao et al. (doi:10.1016/j.cell.2021.04.048) and available as UMI counts at https://atlas.fredhutch.org/data/nygc/multimodal/pbmc_multimodal.h5seurat

Usage

```
sampleObj
```

Format

A sparse matrix of 50 cells and 20729 genes.

Source

doi:10.1016/j.cell.2021.04.048

SampleTree.STACAS *Integration tree generation*

Description

Build an integration tree by clustering samples in a hierarchical manner. Cumulative scoring among anchor pairs will be used as pairwise similarity criteria of samples.

Usage

```
SampleTree.STACAS(
  anchorset,
  obj.names = NULL,
  hclust.method = c("single", "complete", "ward.D2", "average"),
  usecol = c("score", "dist.mean"),
  method = c("weight.sum", "counts"),
  semisupervised = TRUE,
  plot = TRUE
)
```

Arguments

anchorset	Scored anchorsobtained from FindAnchors.STACAS and FilterAnchors.STACAS function
obj.names	Option vector of names for objects in anchorset
hclust.method	Clustering method to be used (single, complete, average, ward)
usecol	Column name to be used to compute sample similarity. Default "score"
method	Aggregation method to be used among anchors for sample similarity computation. Default: weight.sum
semisupervised	Whether to use cell type label information (if available)
plot	Logical indicating if dendrogram must be plotted

Value

An integration tree to be passed to the integration function.

Examples

```
data(sampleObj)
library(Seurat)
obj.list <- SplitObject(sampleObj, split.by="donor")
anchors <- FindAnchors.STACAS(obj.list, min.sample.size=10, k.score=5, dims=3)
tree <- SampleTree.STACAS(anchors)
```

StandardizeGeneSymbols

Standardize gene symbols

Description

Converts gene names of a Seurat single-cell object to a dictionary of standard symbols. This function is useful prior to integration of datasets from different studies, where gene names may be inconsistent.

Usage

```
StandardizeGeneSymbols(
  obj,
  assay = NULL,
  slots = c("counts", "data"),
  EnsemblGeneTable = NULL,
  EnsemblGeneFile = NULL
)
```

Arguments

`obj`	A Seurat object
`assay`	Assay where gene names should be translated
`slots`	Slots where gene names should be translated
`EnsemblGeneTable`	A data frame of gene name mappings. This should have the format of Ensembl BioMart tables with fields "Gene name", "Gene Synonym" and "Gene stable ID" (and optionally "NCBI gene (formerly Entrezgene) ID"). See also the default conversion table in STACAS with `data(EnsemblGeneTable.Mm)`
`EnsemblGeneFile`	If `EnsemblGeneTable==NULL`, read a gene mapping table from this file

Value

Returns a Seurat object with standard gene names. Genes not found in the standard list are removed. Synonyms are accepted when the conversion is not ambiguous.

Examples

```
data(EnsemblGeneTable.Mm)
data(sampleObj)
sampleObj <- StandardizeGeneSymbols(sampleObj, EnsemblGeneTable=EnsemblGeneTable.Mm)
```

Index