Package 'velociraptor'

March 30, 2021

Title Toolkit for Single-Cell Velocity

Version 1.0.0

Date 2020-10-10

Description This package provides Bioconductor-friendly wrappers for RNA velocity calculations in single-cell RNA-seq data. We use the basilisk package to manage Conda environments, and the zellkonverter package to convert data structures between SingleCellExperiment (R) and AnnData (Python). The information produced by the velocity methods is stored in the various components of the SingleCellExperiment class.

Depends SummarizedExperiment

Imports methods, stats, Matrix, BiocGenerics, reticulate, S4Vectors, DelayedArray, basilisk, zellkonverter, scuttle, SingleCellExperiment, BiocParallel, BiocSingular

Suggests BiocStyle, testthat, knitr, rmarkdown, scran, scater, scRNAseq, ggplot2, Rtsne

StagedInstall no

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.1

URL https://github.com/kevinrue/velociraptor

BugReports https://github.com/kevinrue/velociraptor/issues

biocViews SingleCell, GeneExpression, Sequencing, Coverage

VignetteBuilder knitr

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

git_branch RELEASE_3_12

git_last_commit 7cd4624

git_last_commit_date 2020-10-27

Date/Publication 2021-03-29

Author Kevin Rue-Albrecht [aut, cre] (<https://orcid.org/0000-0003-3899-3872>), Aaron Lun [aut] (<https://orcid.org/0000-0002-3564-4813>), Charlotte Soneson [aut] (<https://orcid.org/0000-0003-3833-2169>)

Maintainer Kevin Rue-Albrecht <kevinrue67@gmail.com>

R topics documented:

`embedVelocity` *Project velocities onto an embedding*

Description

Project the velocity vector for each cell onto an existing low-dimensional embedding.

Usage

```
embedVelocity(x, vobj, ...)

## S4 method for signature 'ANY'
embedVelocity(x, vobj, ...)

## S4 method for signature 'SingleCellExperiment'
embedVelocity(x, vobj, ..., use.dimred = 1)
```

Arguments

`x`	A numeric matrix of low-dimensional coordinates, e.g., after t-SNE. Alternatively, a SingleCellExperiment containing such coordinates in its `reducedDims`.
`vobj`	A SingleCellExperiment containing the output of the velocity calculations, typically after running `scvelo`.
`...`	For the generic, further arguments to pass to specific methods. For the ANY method, further arguments to pass to the `velocity_embedding` Python function from **scVelo**. For the SingleCellExperiment method, further arguments to pass to the ANY method.
`use.dimred`	String or integer scalar specifying the reduced dimensions to retrieve from `x`.

Details

This is a simple wrapper around the `scvelo.tools.velocity_embedding` function. Briefly, we construct a cell-cell transition matrix where a cell is more likely to transition to one of its neighbors if its velocity vector is pointing in the same direction as that neighbor. The resulting matrix is then used to compute a weighted average of the positions in `x`, allowing us to compute a velocity in the low-dimensional embedding.

Value

A numeric matrix of the same dimensions as `x`, containing the projected velocity vectors in that embedding.

Author(s)

Aaron Lun

Examples

```
example(scvelo, echo=FALSE) # recycling that example.

# Making up a new embedding.
tsne.results <- matrix(rnorm(2*ncol(out)), ncol=2)

# Projecting the future state of each cell:
projected <- embedVelocity(tsne.results, out)
```

---

gridVectors *Summarize vectors into a grid*

Description

Summarize the velocity vectors into a grid, usually for easy plotting.

Usage

```
gridVectors(x, embedded, ...)

## S4 method for signature 'ANY'
gridVectors(x, embedded, resolution = 40, scale = TRUE, as.data.frame = TRUE)

## S4 method for signature 'SingleCellExperiment'
gridVectors(x, embedded, ..., use.dimred = 1)
```

Arguments

x	A numeric matrix of low-dimensional coordinates, e.g., after t-SNE. Alternatively, a SingleCellExperiment containing such coordinates in its reducedDims.
embedded	A low-dimensional projection of the velocity vectors into the embedding of x. This should be of the same dimensions as x and is typically produced by embedVelocity.
...	For the generic, further arguments to pass to specific methods. For the SingleCellExperiment method, further arguments to pass to the ANY method.
resolution	Integer scalar specifying the resolution of the grid, in terms of the number of grid intervals along each axis.
scale	Logical scalar indicating whether the averaged vectors should be scaled by the grid resolution.
as.data.frame	Logical scalar indicating whether the output should be a data.frame. If FALSE, a list of two matrices is returned.
use.dimred	String or integer scalar specifying the reduced dimensions to retrieve from x.

Details

This partitions the bounding box of `x` into a grid with `resolution` units in each dimension. The locations and vectors of all cells in each block are averaged to obtain a representative of that block. This is most obviously useful for visualization to avoid overplotting of velocity vectors.

If `scale=TRUE`, per-block vectors are scaled so that the median vector length is comparable to the spacing between blocks. This improves visualization when the scales of `x` and `embedded` are not immediately comparable.

Value

If `as.data.frame=FALSE`, a list is returned containing `start` and `end`, two numeric matrices with one row per non-empty block in the grid and one column per column in `x`. `start` contains the mean location of all cells inside that block, and `end` contains the endpoint after adding the (scaled) average of the block's cell's velocity vectors.

If `as.data.frame=TRUE`, a data.frame is returned with numeric columns of the same contents as the list above. Column names are prefixed by `start.*` and `end.*`.

Author(s)

Aaron Lun

See Also

`embedVelocity`, to generate `embedded`.

Examples

```
tsne.results <- matrix(rnorm(10000), ncol=2)
tsne.vectors <- matrix(rnorm(10000), ncol=2)

out <- gridVectors(tsne.results, tsne.vectors)

# Demonstration for plotting.
plot(tsne.results[,1], tsne.results[,2], col='grey')
arrows(out$start.1, out$start.2, out$end.1, out$end.2, length=0.05)
```

`scvelo` *RNA velocity with* **scVelo**

Description

Perform RNA velocity calculations with the **scVelo** package.

Usage

```
scvelo(x, ...)

## S4 method for signature 'ANY'
scvelo(
  x,
```

```
  subset.row = NULL,
  sf.X = NULL,
  sf.spliced = NULL,
  sf.unspliced = NULL,
  use.theirs = FALSE,
  mode = c("steady_state", "deterministic", "stochastic", "dynamical"),
  scvelo.params = list(),
  dimred = NULL,
  ncomponents = 30,
  BPPARAM = SerialParam(),
  BSPARAM = bsparam()
)

## S4 method for signature 'SummarizedExperiment'
scvelo(
  x,
  ...,
  assay.X = "counts",
  assay.spliced = "spliced",
  assay.unspliced = "unspliced"
)

## S4 method for signature 'SingleCellExperiment'
scvelo(x, ..., sf.X = sizeFactors(x), dimred = NULL, use.dimred = NULL)
```

Arguments

x	A named list of three matrices of the same dimensions where genes are in rows and cells are in columns. The list should contain "spliced" and "unspliced" entries containing spliced and unspliced counts, respectively. It should also contain an "X" entry containing the "usual" count matrix, see details below. Alternatively, a SummarizedExperiment object containing three such matrices among its assays.
...	For the generic, further arguments to pass to specific methods. For the SummarizedExperiment and SingleCellExperiment methods, further arguments to pass to the ANY method.
subset.row	A character, integer or logical vector specifying the genes to use for the velocity calculations. Defaults to all genes.
sf.X	A numeric vector containing size factors for usual count matrix. Defaults to librarySizeFactors on the "X" matrix in x.
sf.spliced	A numeric vector containing size factors for the spliced counts for each cell. Defaults to librarySizeFactors on the "spliced" matrix in x.
sf.unspliced	A numeric vector containing size factors for the unspliced counts for each cell. Defaults to librarySizeFactors on the "unspliced" matrix in x.
use.theirs	Logical scalar indicating whether **scVelo**'s gene filtering and normalization should be used.
mode	String specifying the method to use to estimate the transcriptional dynamics.
scvelo.params	List of lists containing arguments for individual **scVelo** functions, see details below.

dimred	A low-dimensional representation of the cells with number of rows equal to the number of cells in x, used to find the nearest neighbors.
ncomponents	Numeric scalar indicating the number of principal components to obtain. Only used if use.theirs=FALSE and dimred=NULL.
BPPARAM	A BiocParallelParam object specifying whether the PCA calculations should be parallelized. Only used if use.theirs=FALSE and dimred=NULL.
BSPARAM	A BiocSingularParam object specifying which algorithm should be used to perform the PCA. Only used if use.theirs=FALSE and dimred=NULL.
assay.X	An integer scalar or string specifying the assay of x containing the usual count matrix.
assay.spliced	An integer scalar or string specifying the assay of x containing the spliced counts.
assay.unspliced	An integer scalar or string specifying the assay of x containing the unspliced counts.
use.dimred	String naming the entry of reducedDims(x) to use for nearest neighbor calculations. Ignored if dimred is supplied.

Details

This function uses the **scVelo** Python package (https://pypi.org/project/scvelo/) for RNA velocity calculations. The main difference from the original **velocyto** approach is that the dynamical model of **scVelo** does not rely on the presence of observed steady-state populations, which should improve the reliability of the velocity calculations in general applications.

For consistency with other Bioconductor workflows, we perform as many standard steps in R as we can before starting the velocity calculations with **scVelo**. This involves:

1. Size factor-based normalization with sf.* values and normalizeCounts. For "X", log-transformation is performed as well, while for the others, only scaling normalization is performed.
2. Subsetting all matrices to subset.row, most typically to a subset of interest, e.g., highly variable genes. Note that, if set, any subsetting is done *after* normalization so that library sizes are correctly computed.
3. If dimred=NULL, the PCA step on the log-expression values derived from the "X" matrix, using the specified BSPARAM to obtain the first ncomponents PCs.

This allows us to guarantee that, for example, the log-expression matrix of HVGs or the PCA coordinates are the same as that used in other applications like clustering or trajectory reconstruction.

Nonetheless, one can set use.theirs=TRUE to directly use the entire **scVelo** normalization and filtering pipeline. This ignores all of the size factors arguments (sf.*), all of the PCA-related arguments (ncomponents, BSPARAM) and subset.row. However, if a low-dimensionality result is supplied via dimred or use.dimred, the **scVelo** PCA will always be omitted.

Upon first use, this function will instantiate a conda environment containing the **scVelo** package. This is done via the **basilisk** package - see the documentation for that package for trouble-shooting.

Value

A SingleCellExperiment is returned containing the output of the velocity calculations. Of particular interest are:

- the velocity_pseudotime field in the colData, containing the velocity pseudotime for each cell.

- the velocity entry of the assays, containing the velocity vectors for each cell.

The output will always have number of columns equal to the number of cells supplied in x, though the number of rows will depend on whether any subsetting (if subset.row is supplied) or feature selection (if use.theirs=TRUE) was performed.

Comments on the three matrices

Strictly speaking, only the spliced and unspliced matrices are necessary for the velocity calculations. However, it is often the case that the spliced matrix is not actually the same as a "usual" count matrix (e.g., generated by summing counts across all exons for all mapped genes). This is due to differences in the handling of ambiguous reads that map across exon-intron boundaries, or to genomic regions that can be either exonic or intronic depending on the isoform; the spliced count matrix is more likely to exclude such reads.

We request the usual count matrix as the "X" entry of x to ensure that the PCA and nearest neighbor detection in **scVelo** are done on the same data as that used in other steps of the large analysis (e.g., clustering, visualization, trajectory reconstruction). In practice, if the usual count matrix is not available, one can often achieve satisfactory results by simply re-using the spliced count matrix as both the "X" and "spliced" entries of x.

Note that if reduced dimensions are supplied in dimred, any "X" entry is only used to create the AnnData object and is not used in any actual calculations.

Additional arguments to Python

Additional arguments to **scVelo** functions are provided via scvelo.params. This is a named list where each entry is named after a function and is itself a named list of arguments for that function. The following function names are currently recognized:

- "filter_and_normalize", for gene selection and normalization. This is not used unless use.theirs=TRUE.
- "moments", for PCA and nearest neighbor detection. The PCA is not performed if dimred or use.dimred is already supplied.
- "recover_dynamics"
- "velocity"
- "velocity_graph"
- "velocity_pseudotime"
- "latent_time"
- "velocity_confidence"

See the **scVelo** documentation for more details about the available arguments.

Author(s)

Aaron Lun, Charlotte Soneson

References

Bergen V. et al. (2019). Generalizing RNA velocity to transient cell states through dynamical modeling. bioxRiv, https://www.biorxiv.org/content/10.1101/820936v1

Examples

```
# Using mock data to demonstrate the process:
library(scuttle)
sce1 <- mockSCE()
sce2 <- mockSCE()

spliced <- counts(sce1)
unspliced <- counts(sce2)

out <- scvelo(list(X=spliced, spliced=spliced, unspliced=unspliced))
```

Index