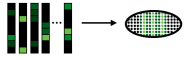

novosparc Documentation

Nikos Karaiskos, Mor Nitzan

Apr 20, 2021

Contents

1	Version 0.4.3 20 April 2021	3
2	Version 0.4.2 03 April 2021	5
3	Version 0.4.1 24 August 2020	7
4	Version 0.3.11 27 April 2020	9
5	Version 0.3.10 07 February 2020	11
6	Version 0.3.7 29 October 2019	13
7	Version 0.3.5 13 June 2019	15
8	Version 0.3.4 27 February 2019	17
	Bibliography	21



novoSpaRc predicts locations of single cells in space by solely using single-cell RNA sequencing data. An existing reference database of marker genes is not required, but significantly enhances performance if available.

novoSpaRc accompanies the following publication:

Gene Expression Cartography

M. Nitzan, N. Karaikos, N. Friedman, N. Rajewsky

CHAPTER 1

Version 0.4.3 20 April 2021

Fixed bugs. Added self consistency analysis and updated tutorials.

CHAPTER 2

Version 0.4.2 03 April 2021

Improved package structure, fixed minor performance issues, and fixed bugs. Added two new tutorials (corti & osteosarcoma) and updated the previous tutorials with validation analyses.

CHAPTER 3

Version 0.4.1 24 August 2020

Changed the package structure and run flow of the scripts. Added anndata and scanpy support. Updated tutorials and implemented basic target geometries.

CHAPTER 4

Version 0.3.11 27 April 2020

Moran's I algorithm for spatially informative genes is implemented and removed pysal dependency.

CHAPTER 5

Version 0.3.10 07 February 2020

Added Moran's I algorithm to detect spatially informative genes.

CHAPTER 6

Version 0.3.7 29 October 2019

Updated computation of shortest paths that significantly reduces running time.

CHAPTER 7

Version 0.3.5 13 June 2019

Fixed a bug that was prone to produce infinities during reconstruction. Improved plotting functions and added new ones for plotting mapped cells.

novoSpaRc reconstructs single-cell gene expression without relying on existing reference markers and makes great use of such information if available.

8.1 Installation

A working Python 3.5 installation and the following libraries are required: matplotlib, numpy, sklearn, scipy, ot and networkx.

The code is partially based on adjustments of the [POT \(Python Optimal Transport\)](#) library.

novoSpaRc requires a working Python 3.4 installation.

8.1.1 PyPI

To install novoSpaRc try:

```
pip install novoSpaRc
```

8.1.2 Trouble shooting

If you do not have sudo rights (you get a `Permission denied error`):

```
pip install --user novosparc
```

If installation through pip fails try installing the pot library first:

```
pip install cython
pip install pot
```

and then novoSpaRc:

```
pip install novosparc
```

8.2 General usage

To spatially reconstruct gene expression, `novoSpaRC` performs the following steps:

1. Read the gene expression matrix.
 - a. *Optional*: select a random set of cells for the reconstruction.
 - b. *Optional*: select a small set of genes (e.g. highly variable).
2. Construct the target space.
3. Setup the optimal transport reconstruction.
 - a. *Optional*: use existing information of marker genes, if available.
4. Perform the spatial reconstruction.
 - a. Assign cells a probability distribution over the target space.
 - b. Derive a virtual in situ hybridization (vISH) for all genes over the target space.
5. Write outputs to file for further use, such as the spatial gene expression matrix and the target space coordinates.
6. *Optional*: plot spatial gene expression patterns.
7. *Optional*: identify and plot spatial archetypes.

8.2.1 Demonstration

We provide scripts that spatially reconstruct two of the tissues presented in the paper: the intestinal epithelium [Moor18] and the stage 6 *Drosophila* embryo [BDTNP].

See also our [tutorial](#) on reconstructing the *Drosophila* embryo.

8.2.2 The intestinal epithelium

The `reconstruct_intestine_denovo.py` script reconstructs the crypt-to-villus axis of the mammalian intestinal epithelium, based on data from [Moor18]. The reconstruction is performed *de novo*, without using any marker genes. The script outputs plots of (a) a histogram showing the distribution of assignment values over embedded zones for each original villus zone, and (b) average spatial gene expression over the original villus zones and embedded zones of 4 gene groups.

Running time on a standard computer is under a minute.

8.2.3 The *Drosophila* embryo

The `reconstruct_bdtnp_with_markers.py` script reconstructs the early *Drosophila* embryo with only a handful of markers, based on the [BDTNP] dataset. All cells are used and a random set of 1-4 markers is selected. The script outputs plots of gene expression for a list of genes, as well as Pearson correlations of the reconstructed and original expression values for all genes. Notice that the results depend on which marker genes are selected. In the manuscript we averaged the results over many different choices of marker genes.

Running time on a standard desktop computer is around 6-7 minutes.

8.2.4 Running novoSpaRc on your data

A template file for running novoSpaRc on custom datasets is provided (`reconstruct_tissue.py`). To successfully run novoSpaRc modify the template file accordingly.

8.2.5 Constructing different grid shapes

We advise to use `novoSpaRc` with diverse target spaces to assess how robust the spatial reconstructions are. A straightforward way to create a target space which is more interesting than a square grid, is to have a simple image with the target space painted in black on it, such as the one below:



Then use the function `create_target_space_from_image` from the `geometry` module to read the image and create a target space out of it. It is advisable to sample a number of all the read locations and not use them all.

8.3 References

Bibliography

- [BDTNP] BDTNP, *Berkeley Drosophila Transcription Network Project*, bdtnp.lbl.gov.
- [Halpern17] Halpern *et al.* (2017), *Single-cell spatial reconstruction reveals global division of labour in the mammalian liver*, [Nature](#).
- [Karaiskos17] Karaiskos *et al.* (2017), *The Drosophila embryo at single-cell transcriptome resolution*, [Science](#).
- [Moor18] Moor *et al.* (2018), *Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis*, [Cell](#).
- [Nitzan18] Nitzan *et al.* (2018), *Charting tissues from single-cell transcriptomes*, [bioRxiv](#).