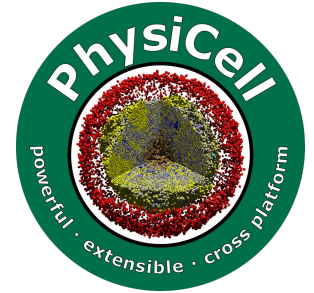


# pcdl, pandas, and anndata data analysis plots

Elmar Bucher

PhD Student 2013-09-17

## PhysiCell Project



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

**PhysiCell.org**

 **@PhysiCell**

# PhysiCell data analysis with one lines

- PhysiCell sample project 2D: interaction.
- PhysiCell sample project 3D: cancer-immune.
- Tutorial:  
<https://github.com/elmbeech/physicelldataloader/blob/master/man/TUTORIAL.md>
- Jupyter notebook:  
[https://github.com/elmbeech/physicelldataloader/blob/master/man/jupyter/pcdl\\_repl\\_programming.ipynb](https://github.com/elmbeech/physicelldataloader/blob/master/man/jupyter/pcdl_repl_programming.ipynb)



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

**PhysiCell.org**

 **@PhysiCell**

# pcdl dataset fetching functions

core python:

- `mcdsts.get_cell_df_states()` # python dictionary with **feature to states** or state count mapping.
- `mcdsts.get_conc_df_states()` # python dictionary with **feature to states** or state count mapping.

pandas:

- `mcds.get_cell_df()` # pandas DataFrame with **cell agent to categorical and numerical data** mapping.
- `mcds.get_conc_df()` # pandas DataFrame with **voxel to substrate concentration** mapping.
- `mcds.get_unit_se()` # pandas Series with **parameter to unit** mapping.

scverse:

- `mcds.get_anndata()` # transform PhysiCell **TimeStep** into **anndata** object.
- `mcdsts.get_anndata()` # transform PhysiCell **TimeSeries** into **anndata** object.



**LUDDY**

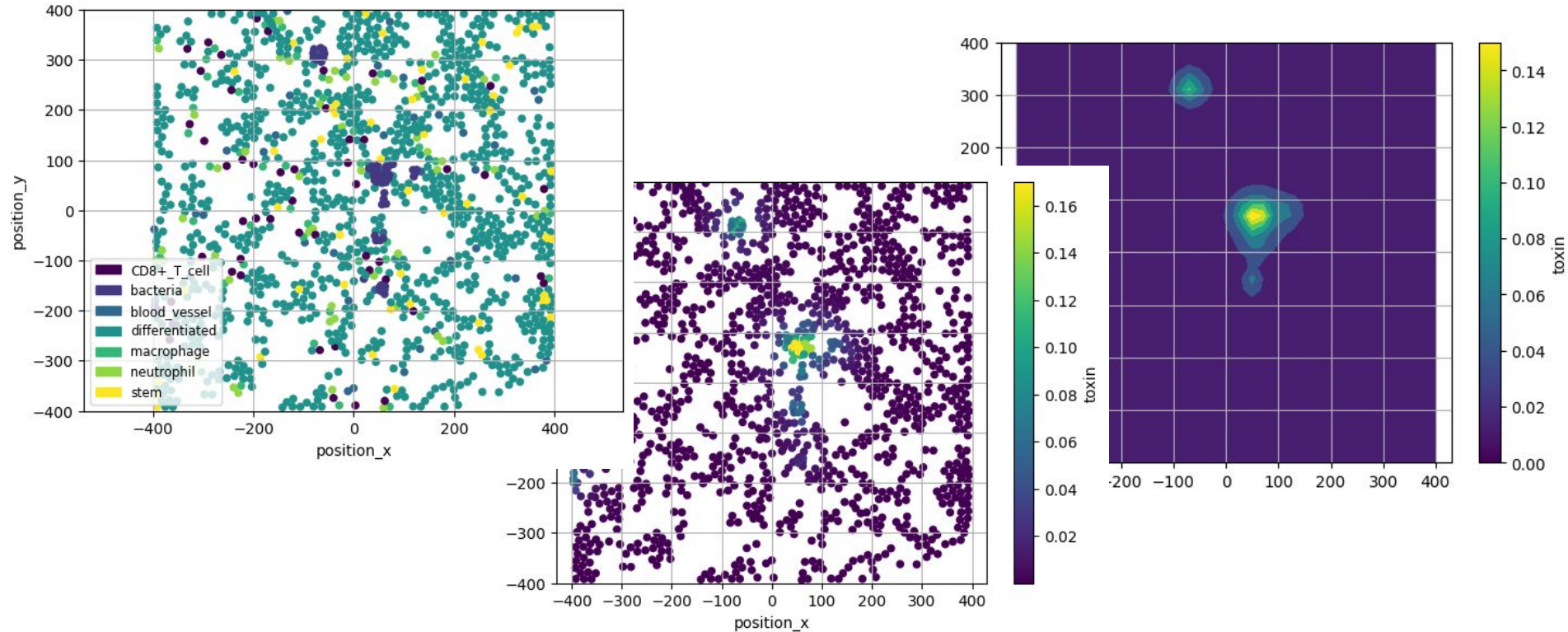
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

**PhysiCell.org**

 **@PhysiCell**

# TimeStep pcdl: get\_scatter() & get\_contour()



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# pcdl TimeSeries plotting functions

- `mcdsts.make_imgcell()`
- `mcdsts.make_imgconc()`
- `mcdsts.make_gif()`
- `mcdsts.make_movie()`



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

**PhysiCell.org**

 **@PhysiCell**



**LUDDY**

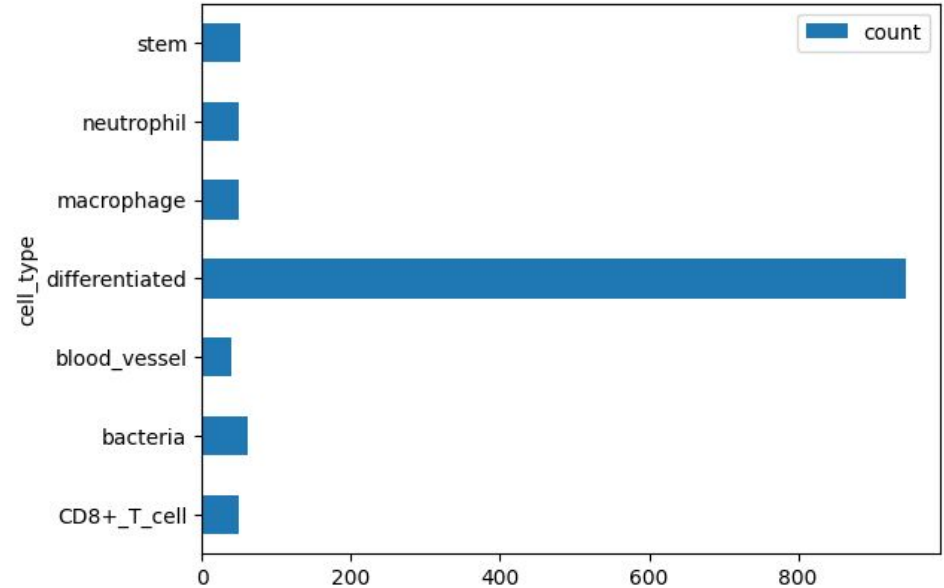
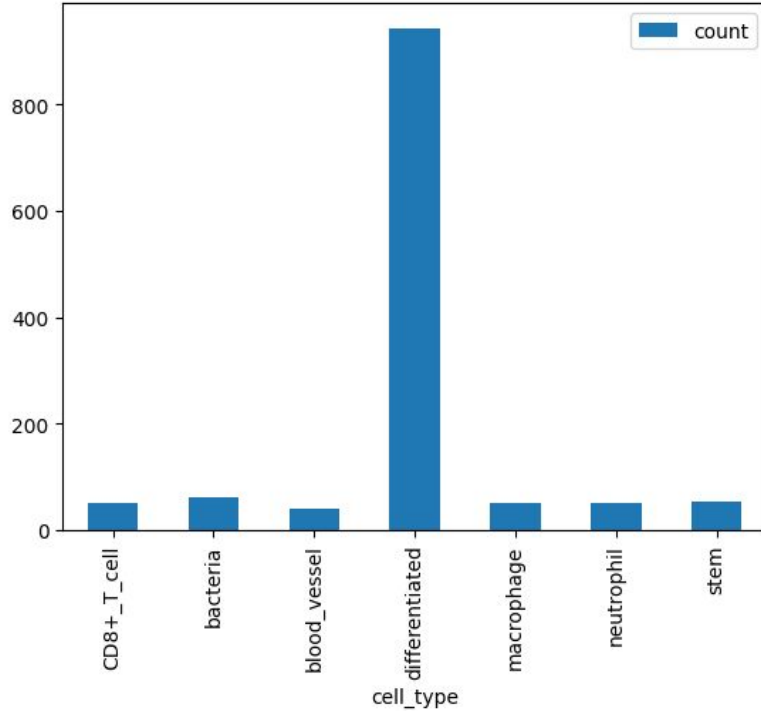
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](http://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# TimeStep pandas: bar & barh



**LUDDY**

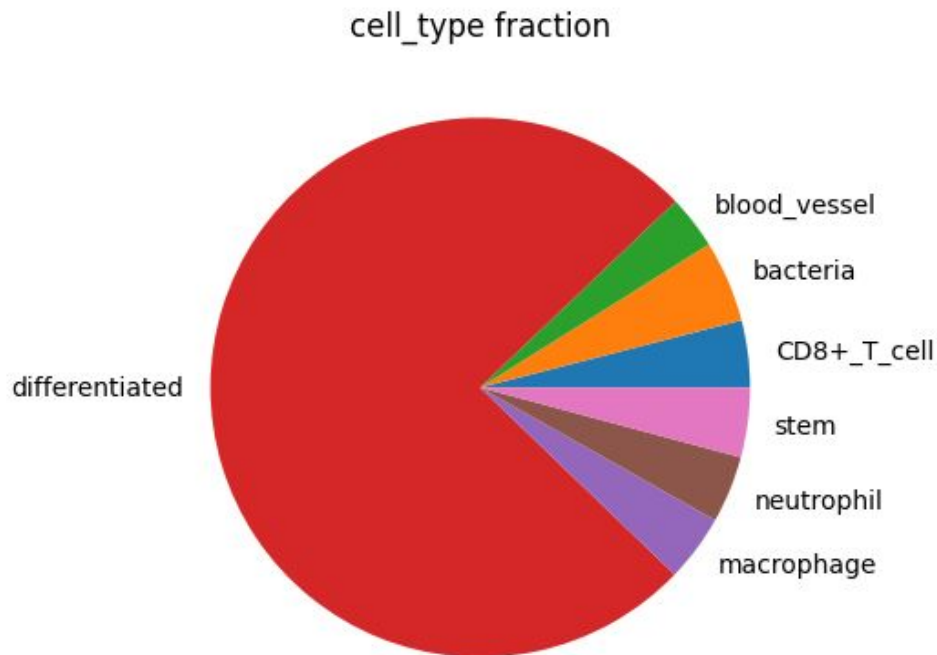
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# TimeStep pandas: pie



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

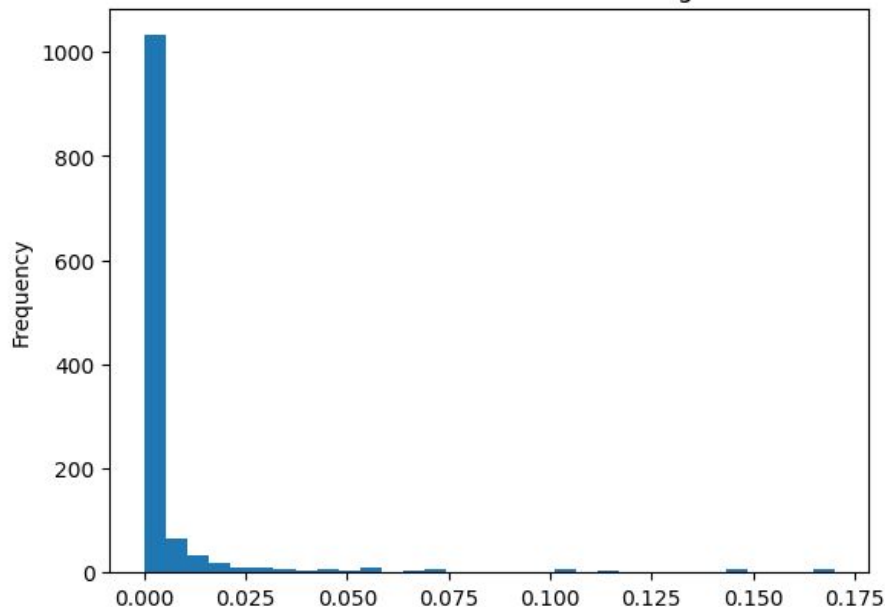
[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

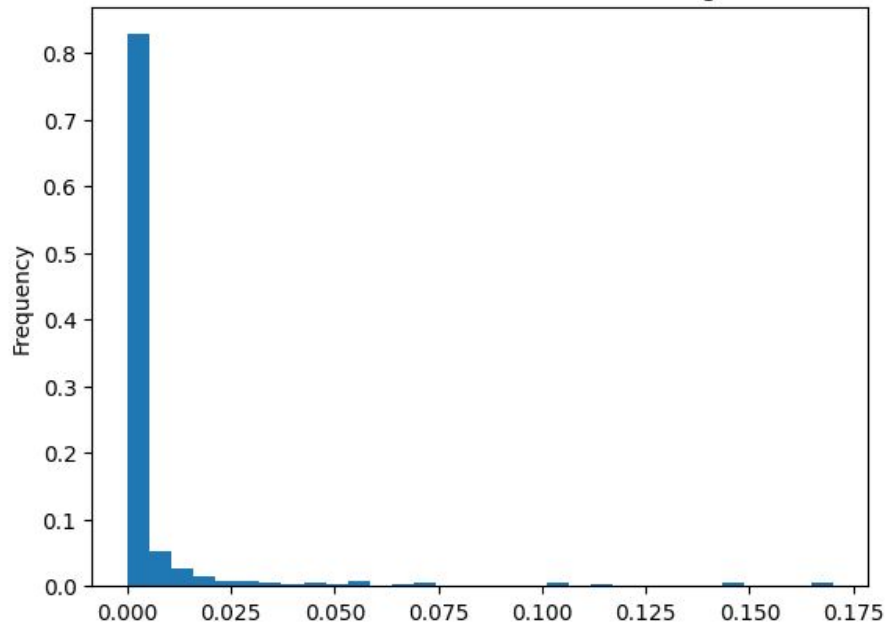


# TimeStep pandas: histogram

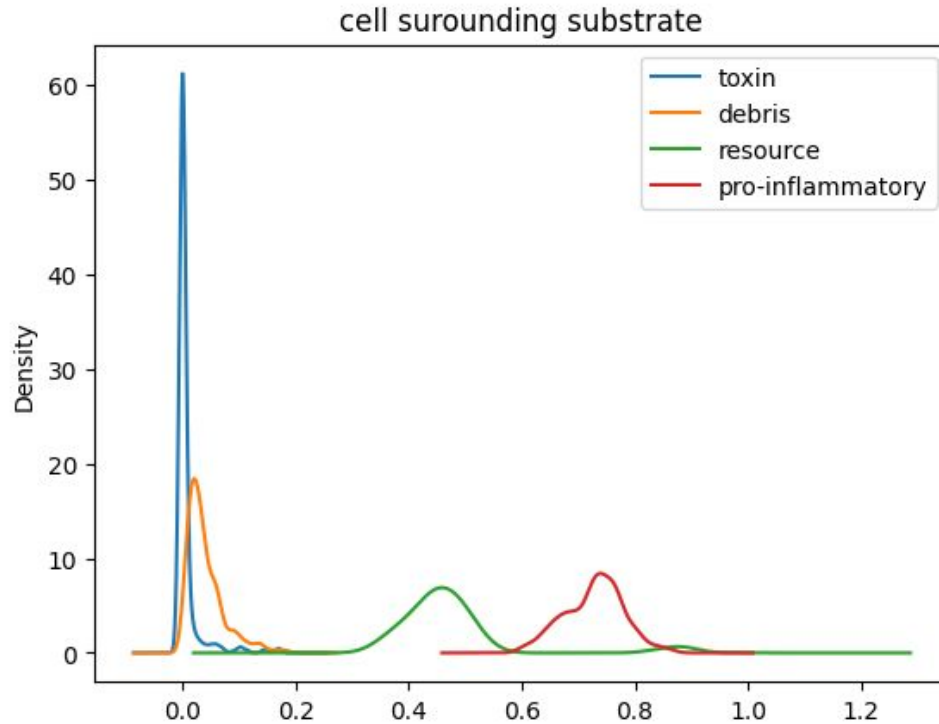
cell count toxin cell surrounding



cell fraction toxin in cell surrounding



# TimeStep pandas: kernel density estimation



**LUDDY**

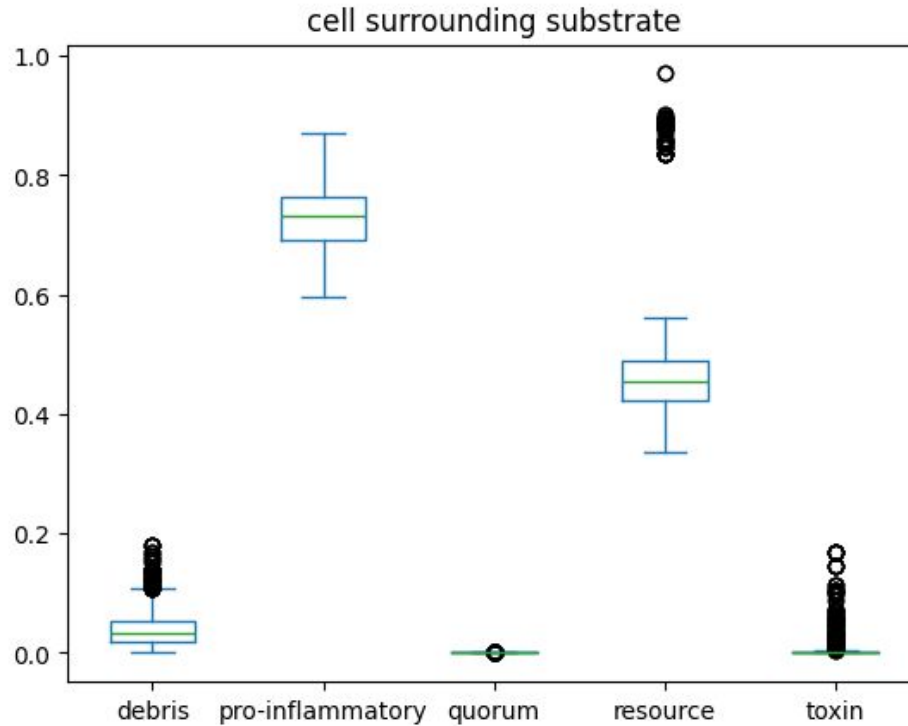
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# TimeStep pandas: box



**LUDDY**

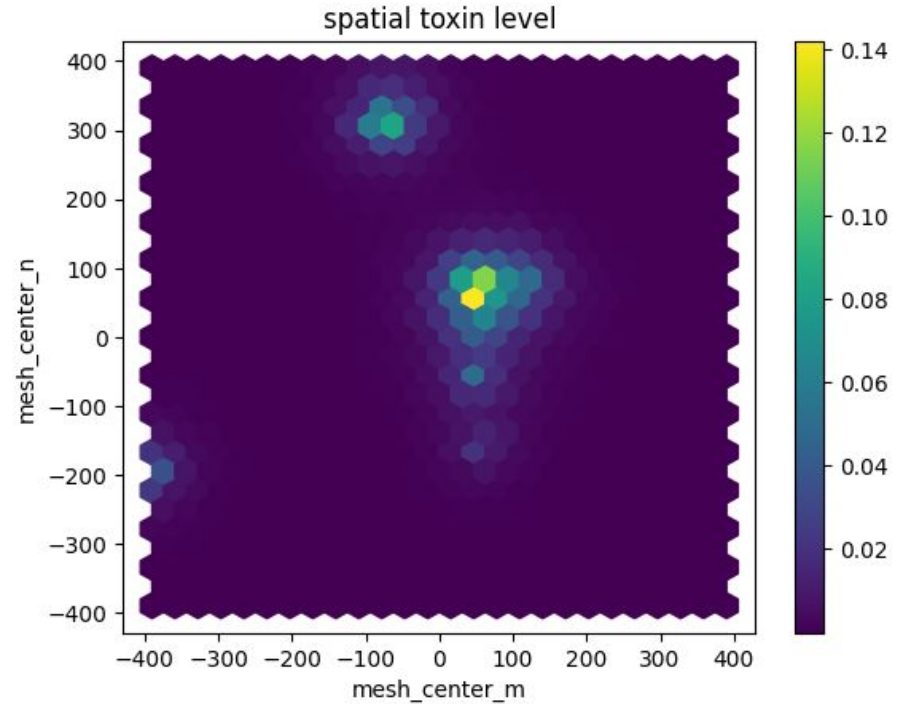
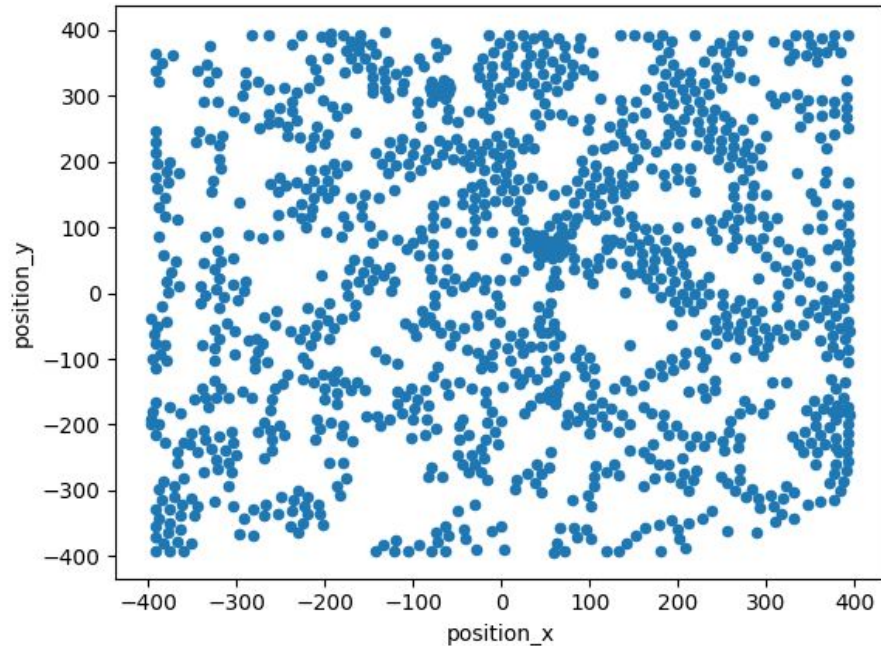
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

PhysiCell.org

@PhysiCell

# TimeStep pandas: scatter & hexbin



**LUDDY**

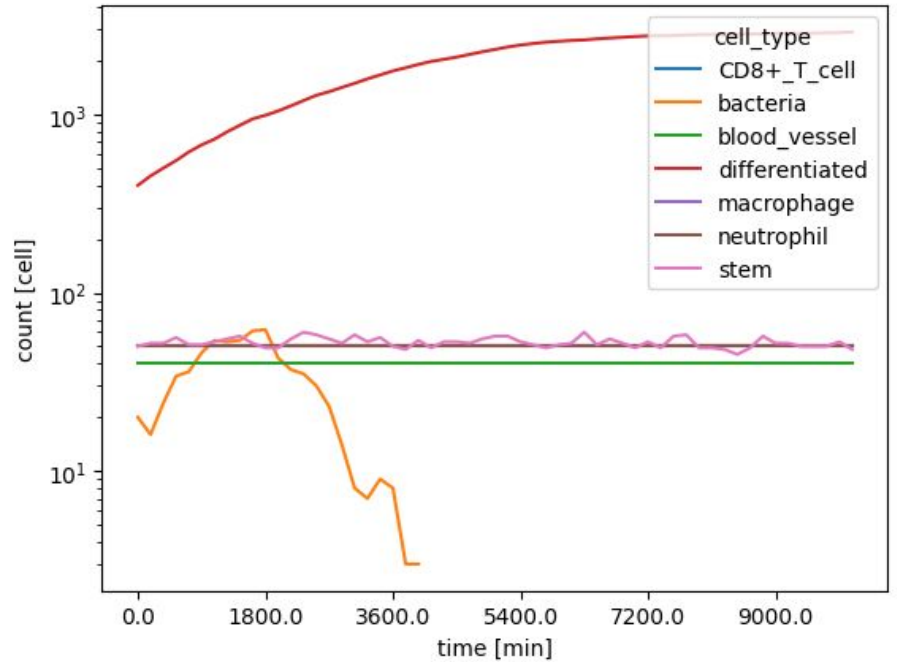
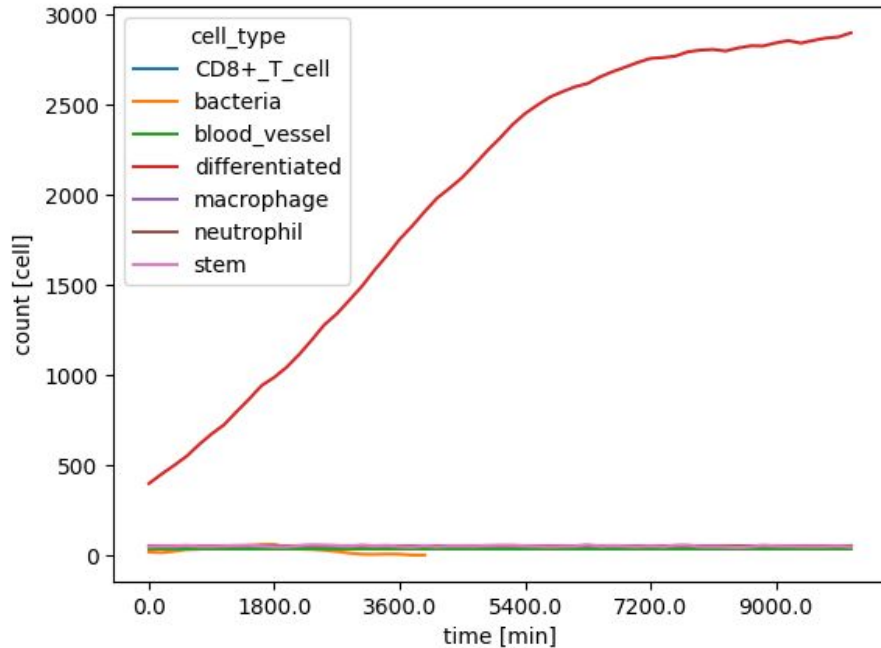
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# TimeSeries pandas: line



**LUDDY**

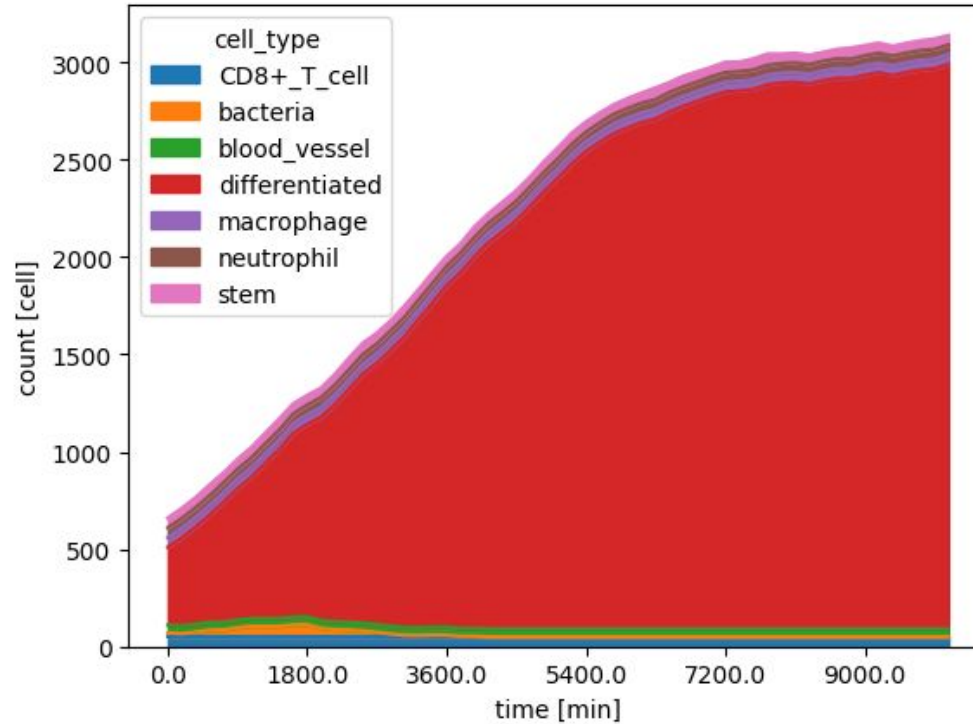
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# TimeSeries pandas: area



**LUDDY**

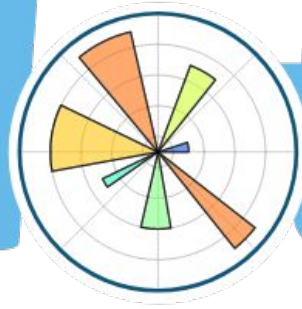
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

PhysiCell.org

@PhysiCell

# matplotlib



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

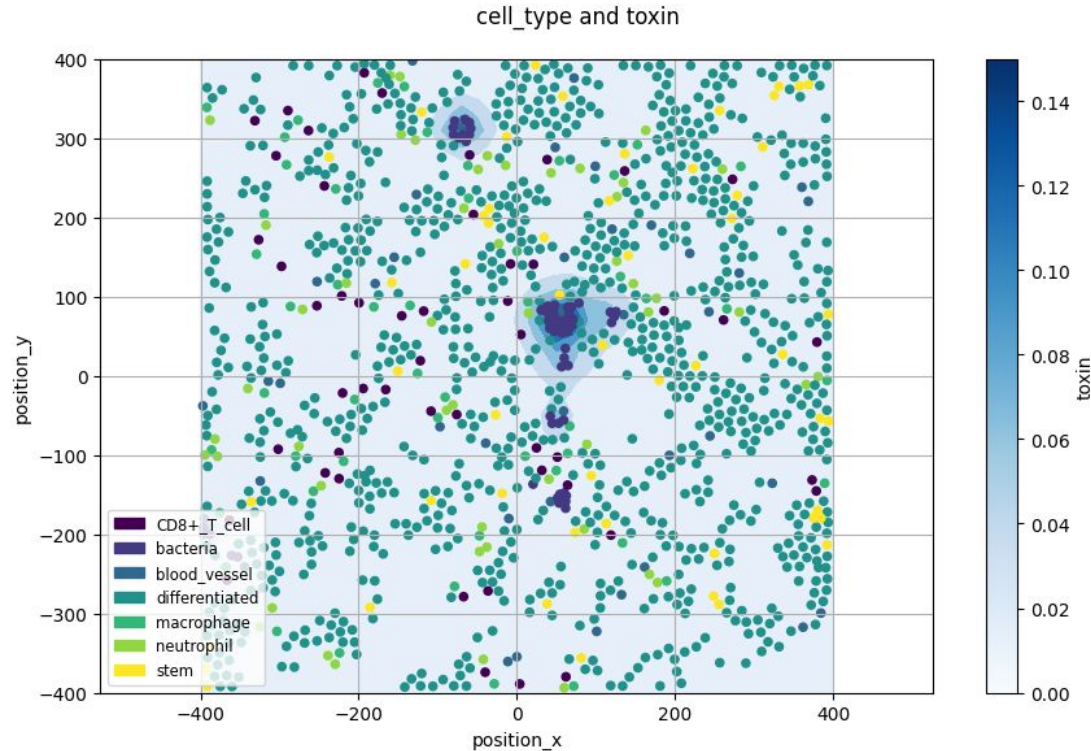
PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

 [@PhysiCell](https://twitter.com/PhysiCell)



# matplotlib embedding



**LUDDY**

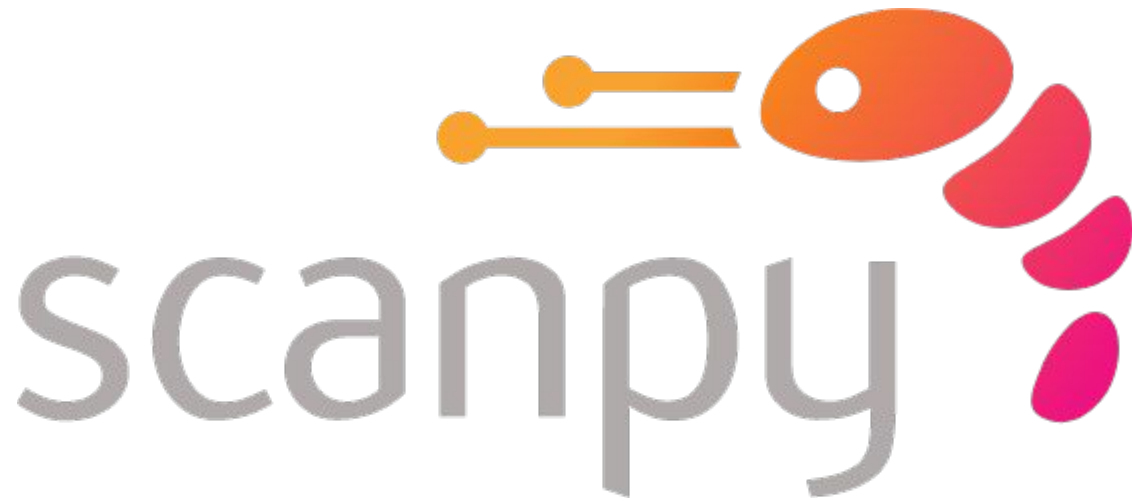
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)





**LUDDY**

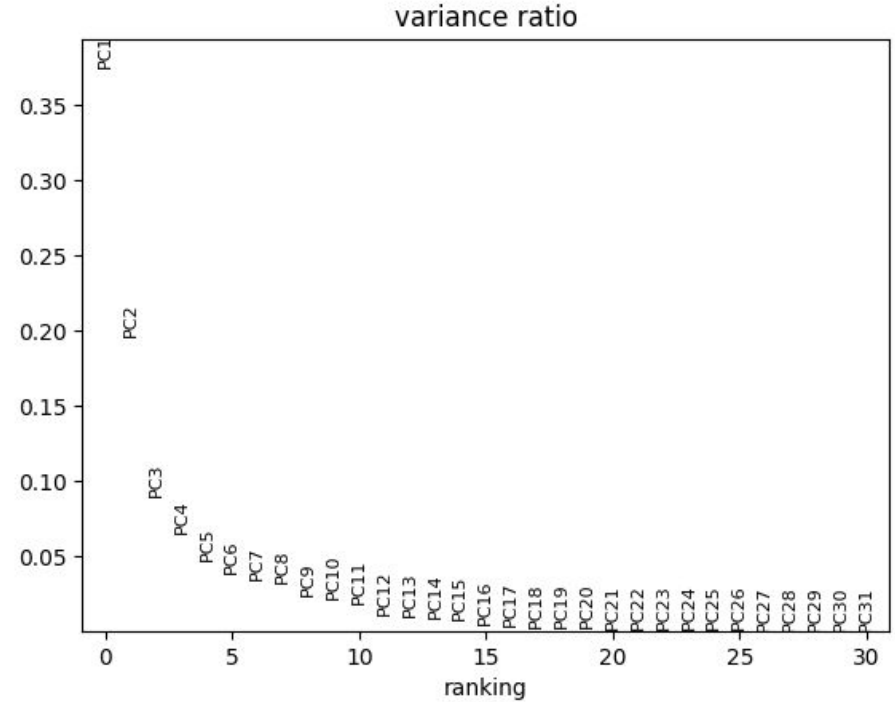
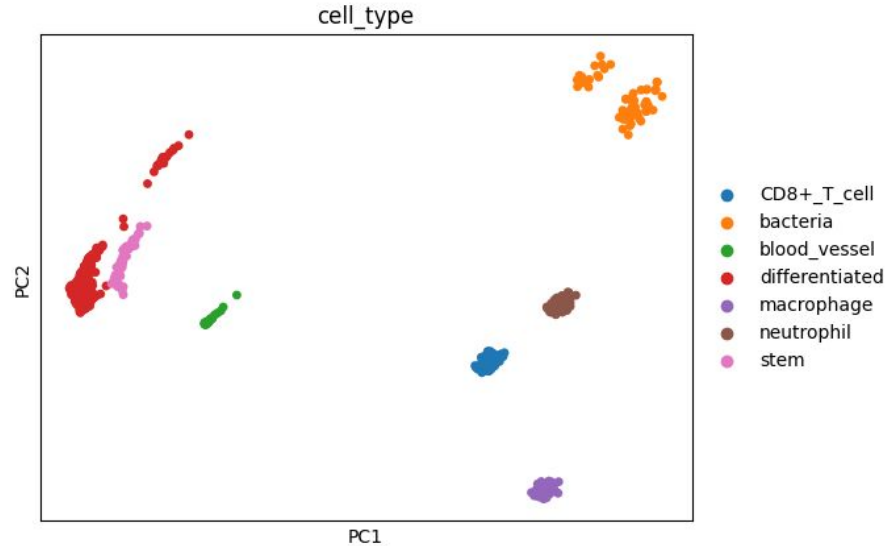
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

**PhysiCell.org**

 **@PhysiCell**

# scanpy: pca & pca\_variance\_ratio



**LUDDY**

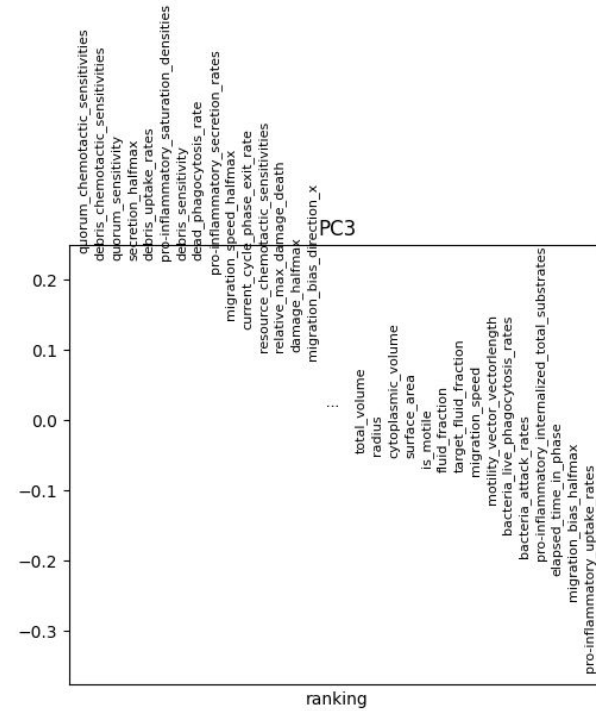
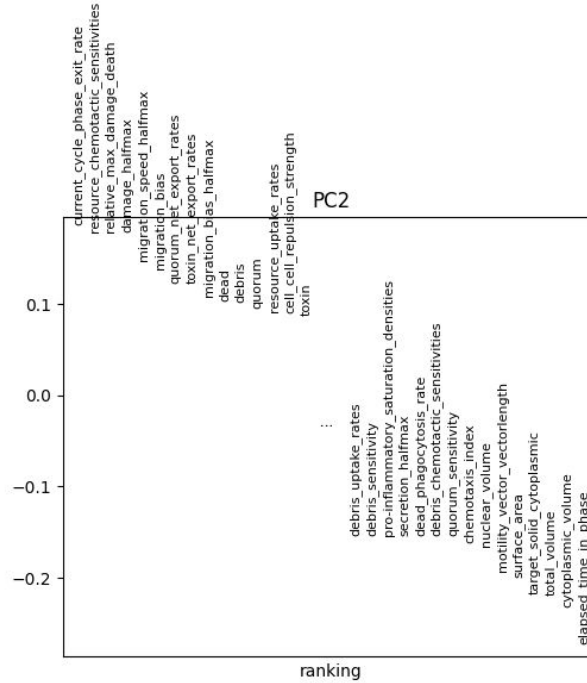
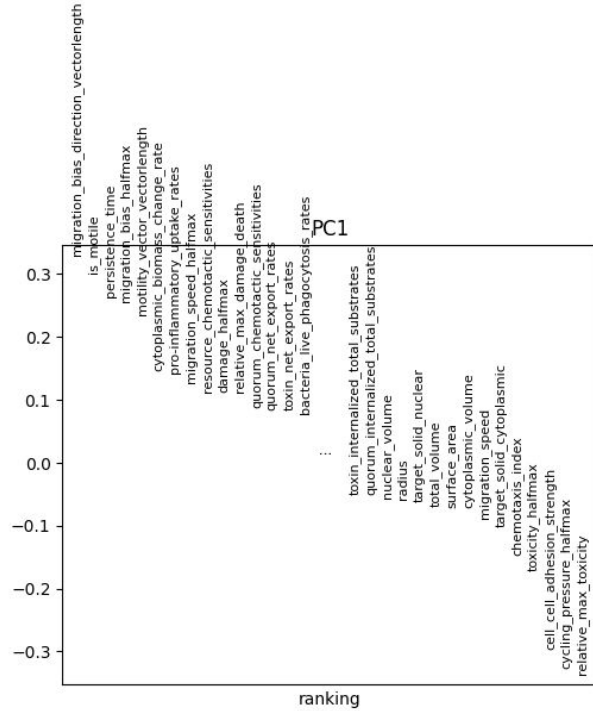
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# scanpy: pca\_loadings



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

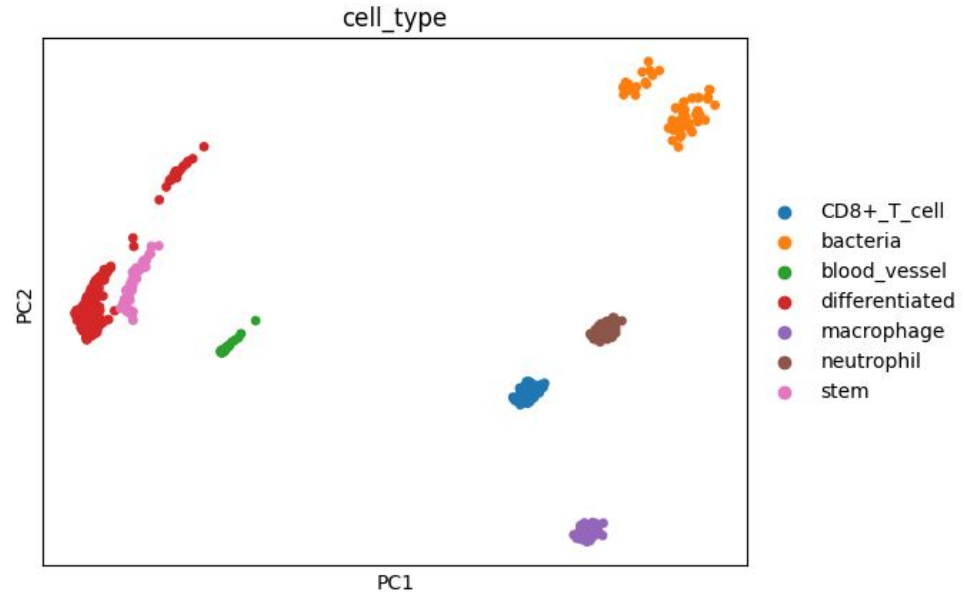
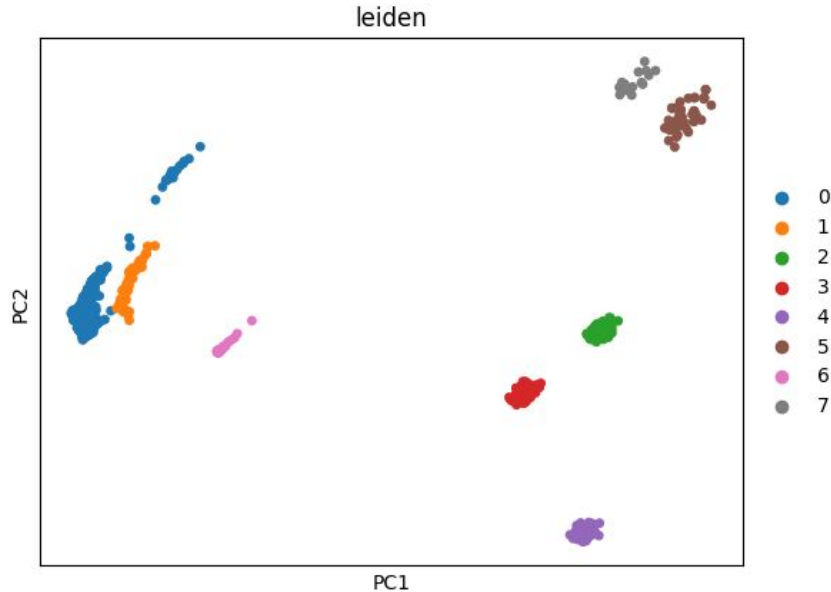
PhysiCell Project

PhysiCell.org

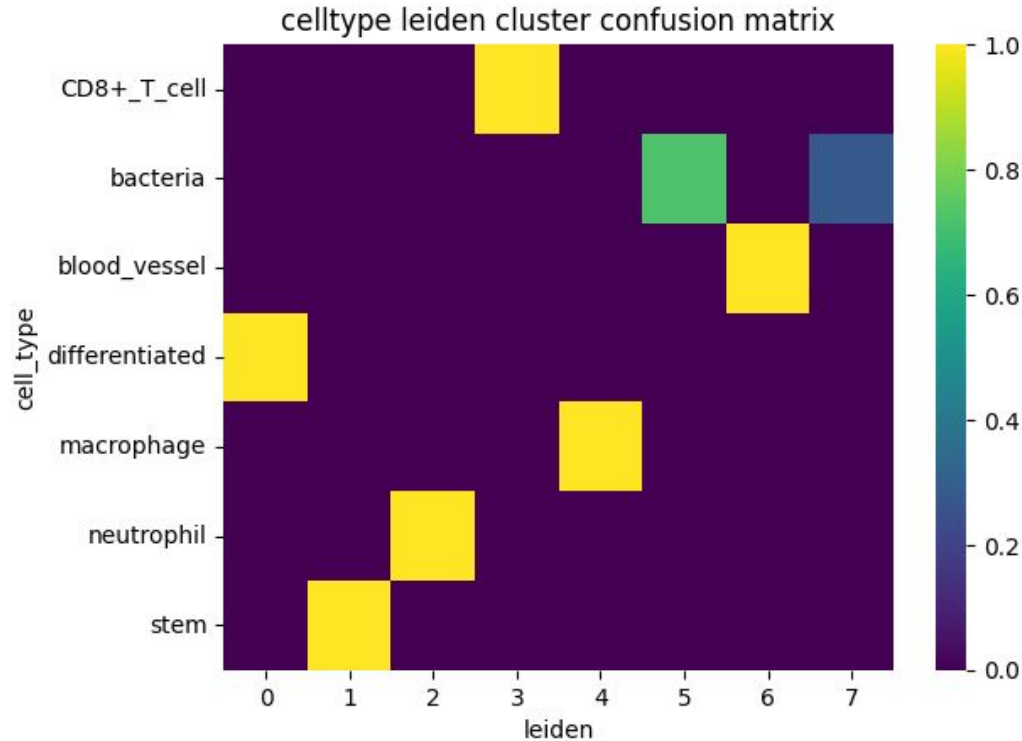
@PhysiCell

# scanpy: leiden nearest neighbour clustering

```
sc.pp.neighbors(ann, n_neighbors=15) # compute the neighborhood graph.  
sc.tl.leiden(ann, resolution=0.01) # cluster the neighborhood graph.
```



# sns scanpy: confusion matrix



**LUDDY**

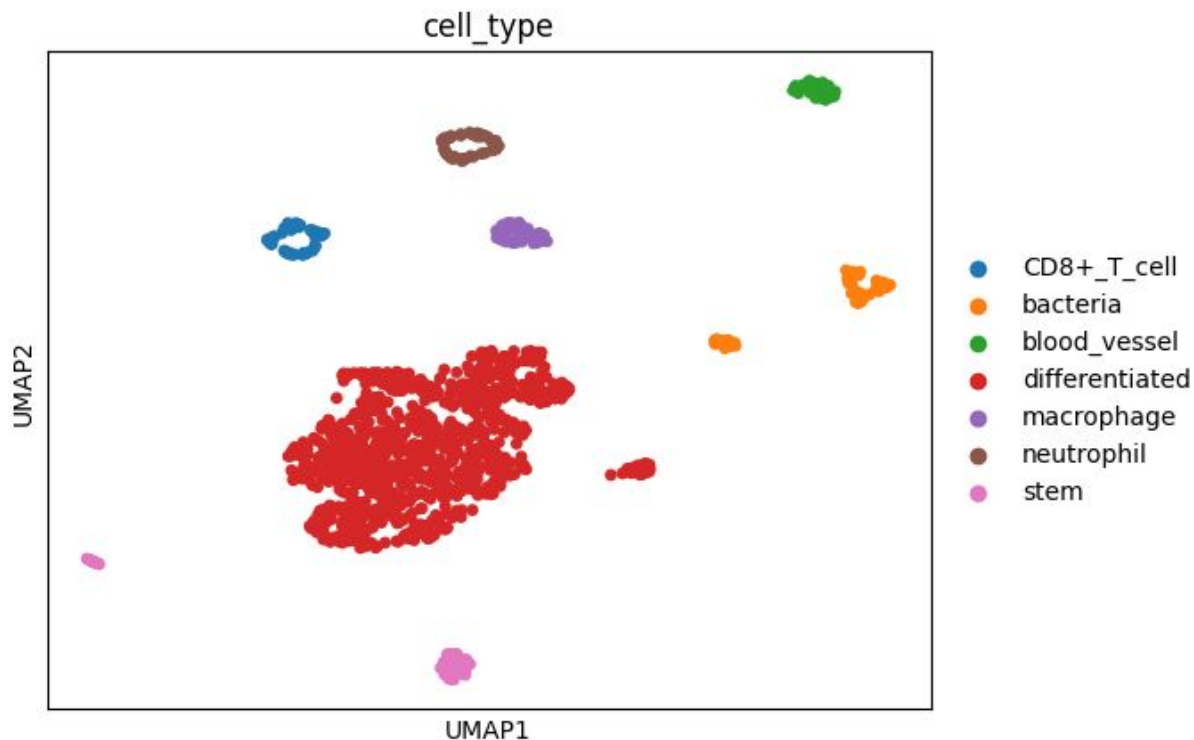
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# scanpy: umap



**LUDDY**

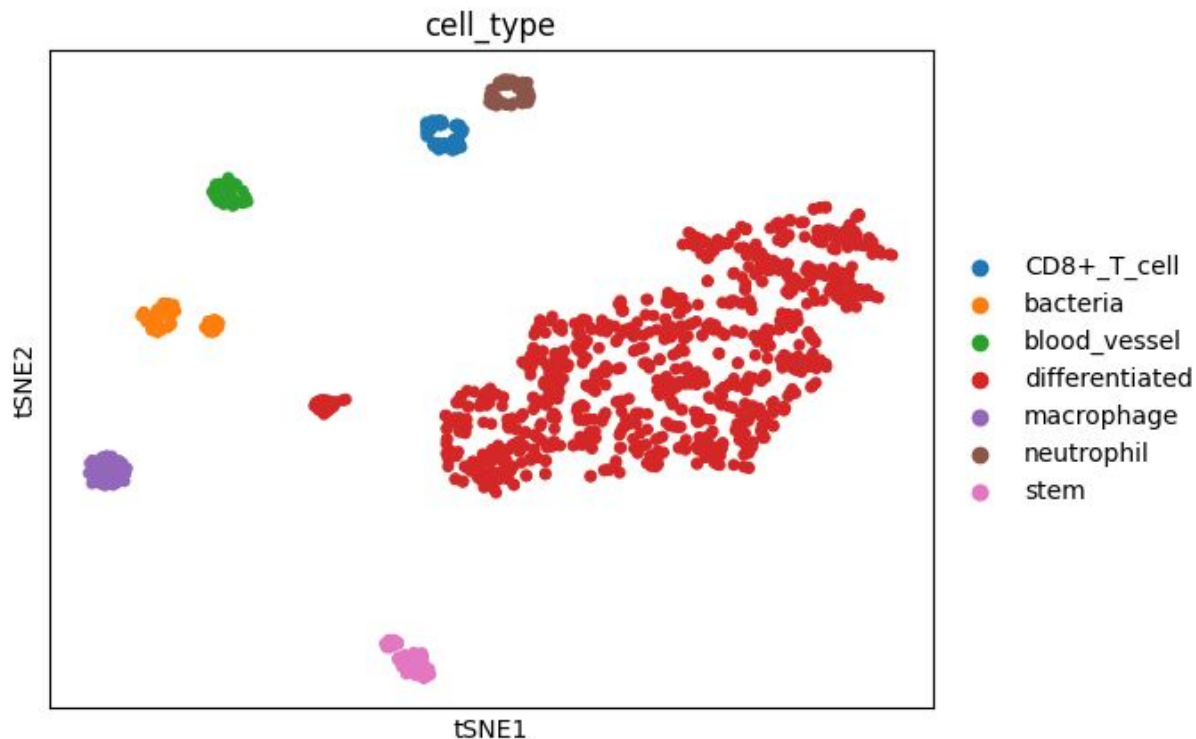
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# scanpy: tSNE



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# scanpy: other dimensional reduction

- fa: ForceAtlas2 force-directed graph.
- fr: Fruchterman Reingold.
- grid\_fr: Grid Fruchterman Reingold.
- kk: Kamadi Kawai.
- lgl: Large Graph Layout.
- drl: Distributed Recursive Layout.
- rt: Reingold Tilford tree layout.
- rt\_circular: Reingold Tilford circular layout.



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

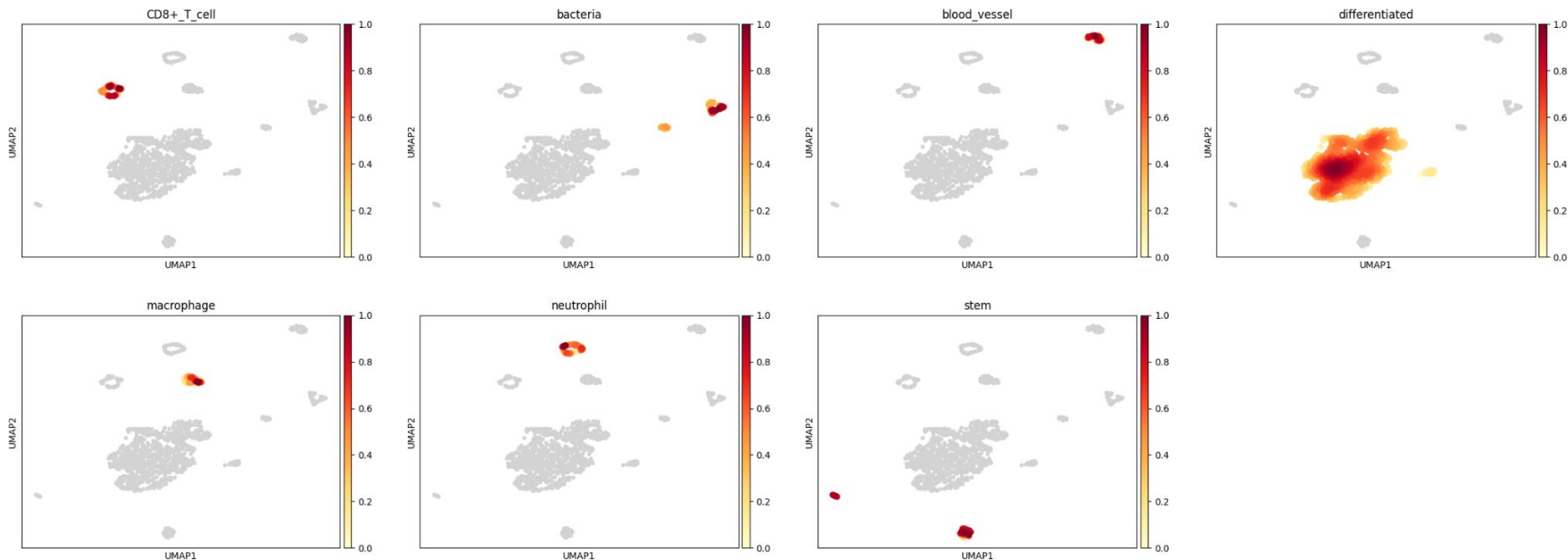
PhysiCell Project

**PhysiCell.org**

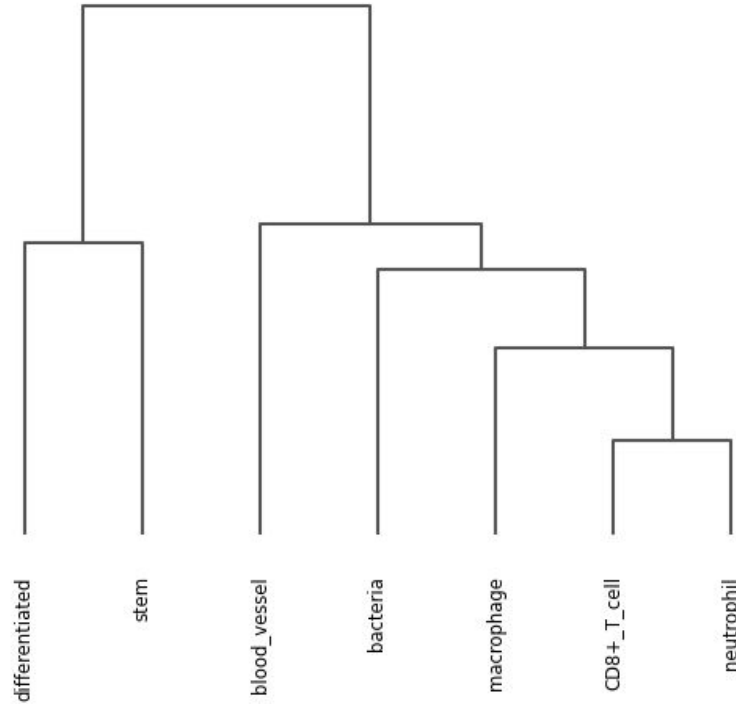
 **@PhysiCell**



# scanpy: embedding\_density



# scanpy: dendrogram



**LUDDY**

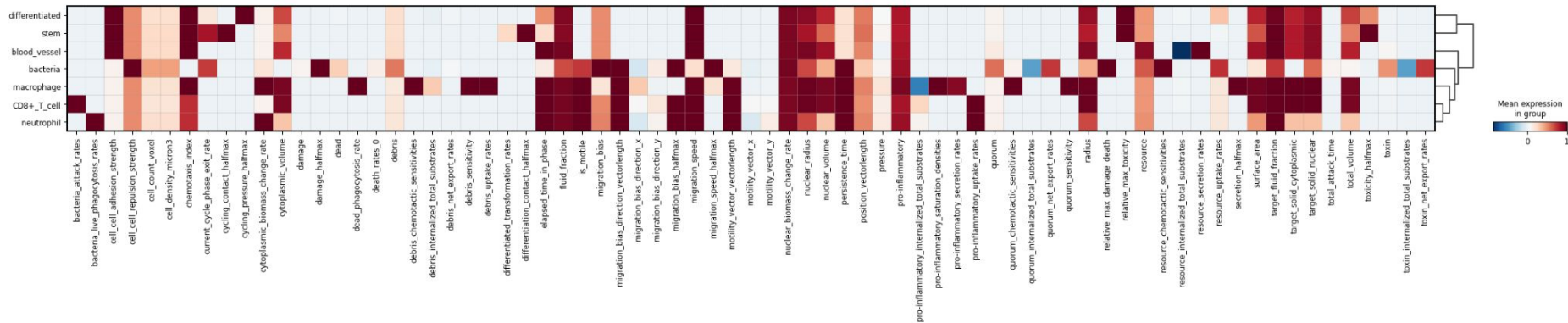
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# scanpy: matrixplot



**LUDDY**

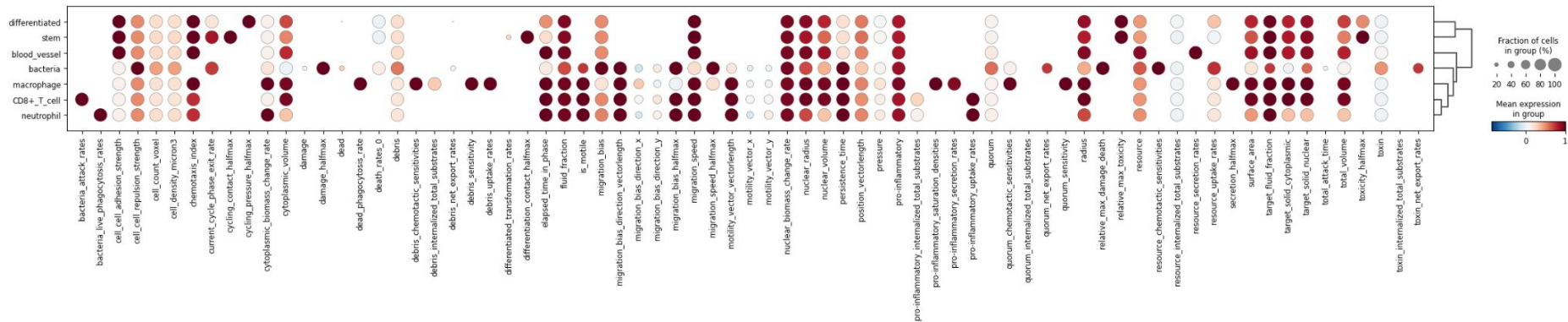
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

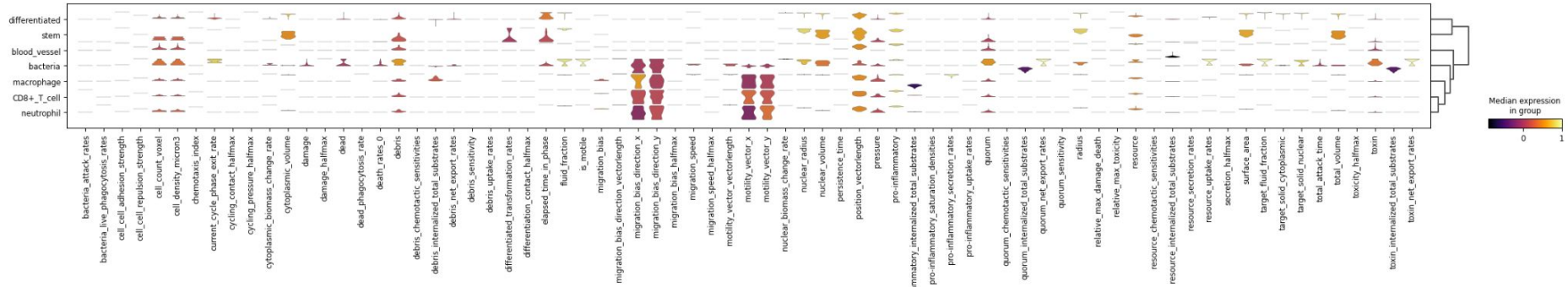
PhysiCell.org

@PhysiCell

# scanpy: dotplot



# scanpy: stacked\_violin



**LUDDY**

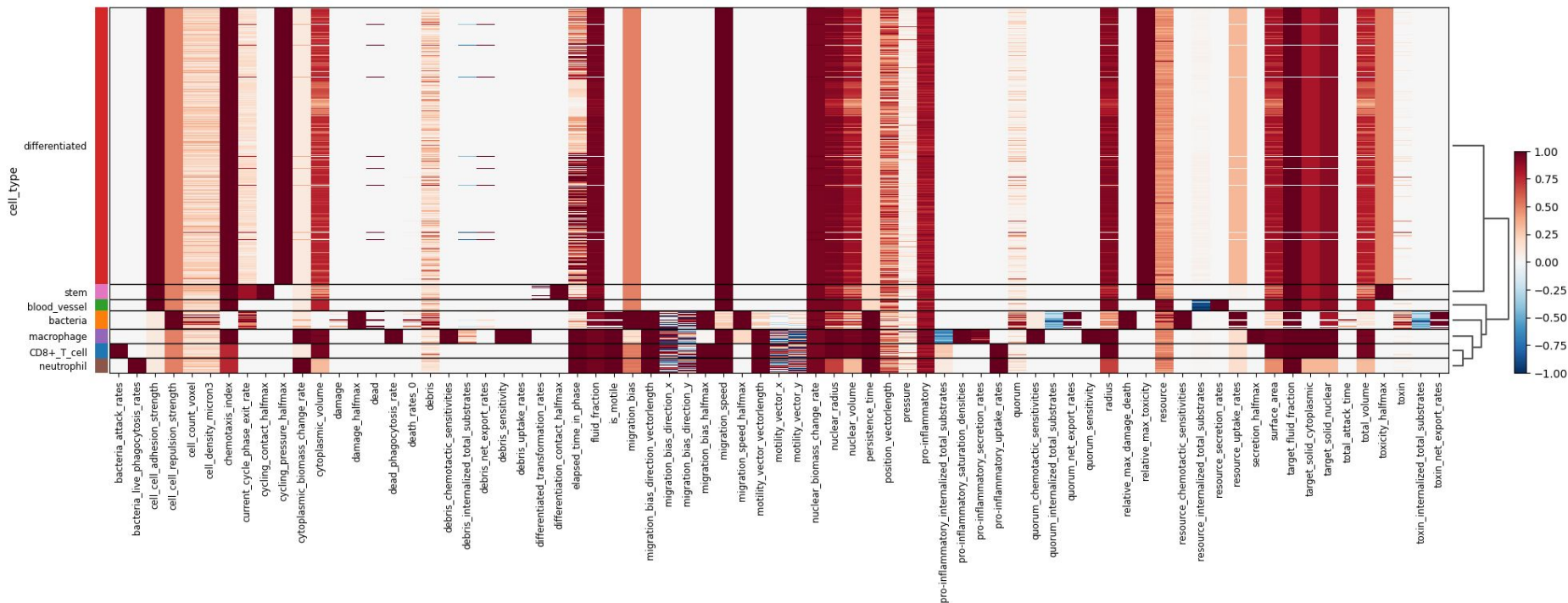
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

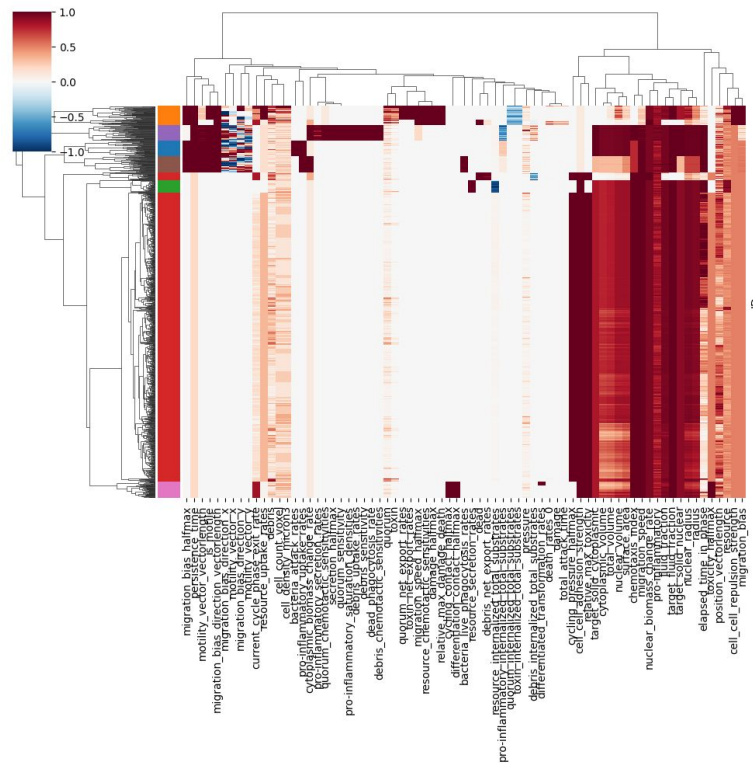
PhysiCell.org

@PhysiCell

# scanpy: heatmap

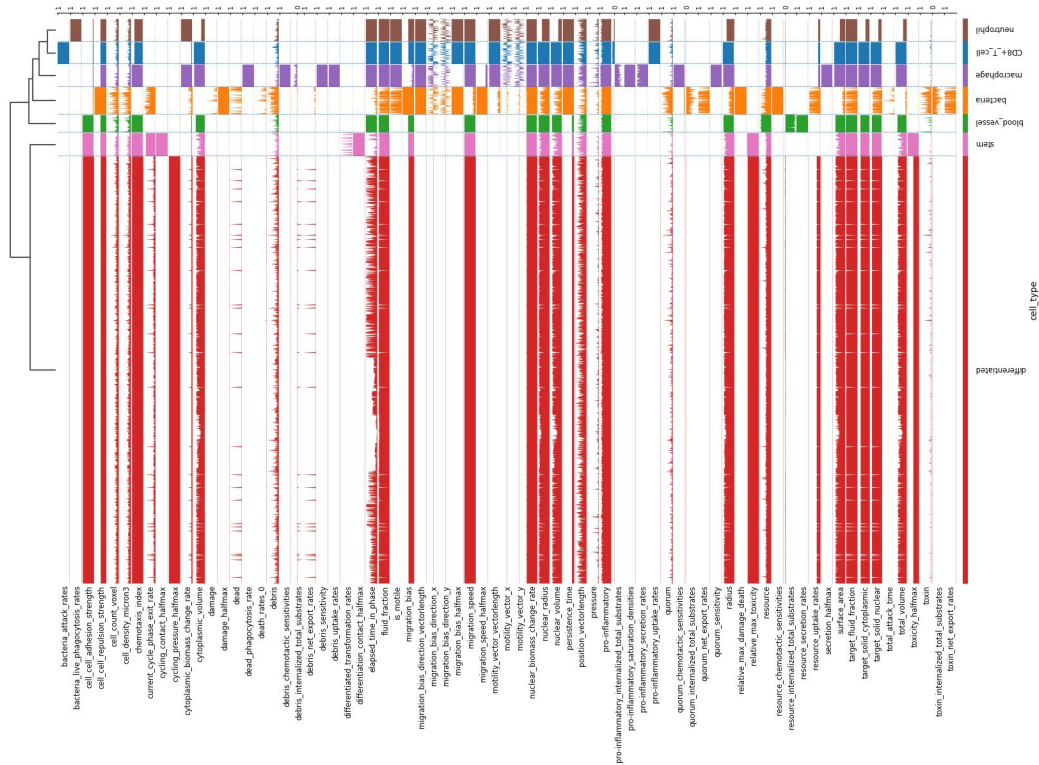


# scanpy: clustermmap



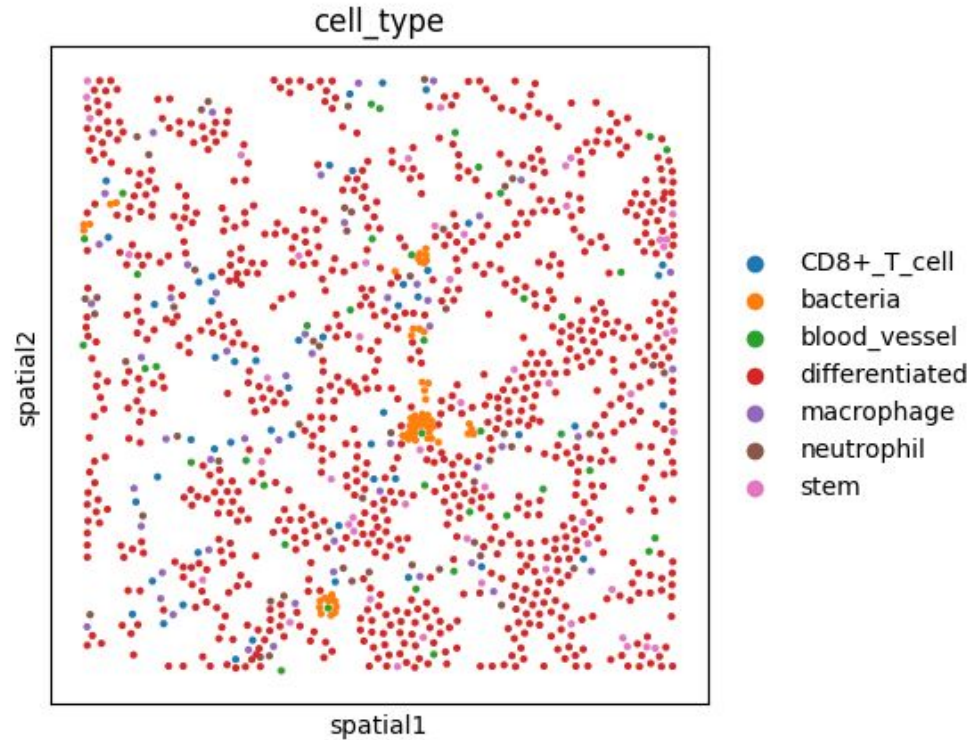


# scanpy: trackplot





# scanpy: spatial



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

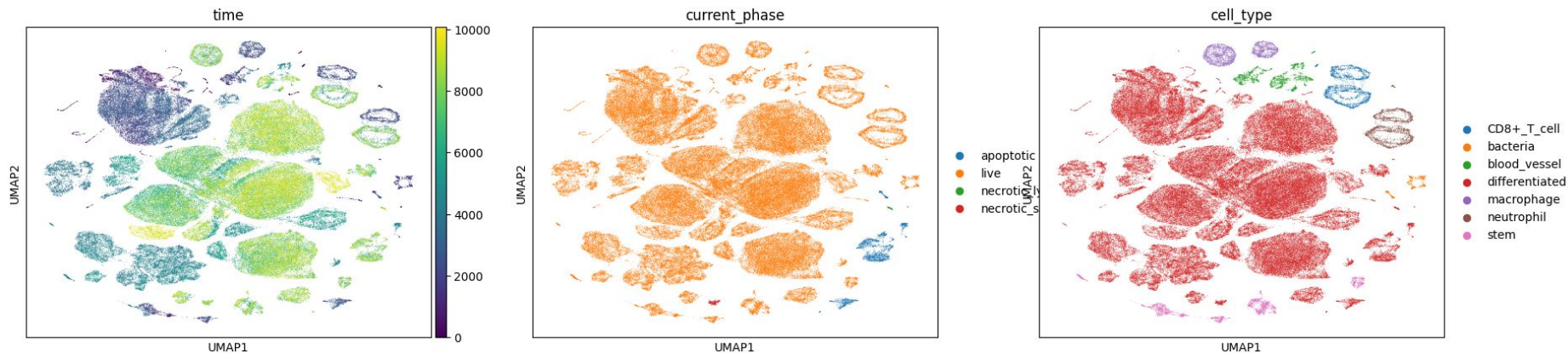
# scanpy: global autocorrelation statistic

## Moran's I and Geary's C

	bacteria_attack_rates	bacteria_live_phagocytosis_rates	cell_cell_adhesion_strength	cell_cell_repulsion_strength	cell_count_voxel	cell_density_micron3	chemotaxis_index
gc	0.000000	0.000000	0.000000	0.000000	0.295582	0.295582	0.000000
mi	0.941978	0.938326	0.959807	0.959389	0.617114	0.617114	0.96043

2 rows × 73 columns

# anndata & scanpy & timeseries



# scanpy: “irrelevant” functions

- `pp.calculate_qc_metric` # scRNAseq related
- `pp.filter_cells` # scRNAseq related
- `pp.filter_genes` # scRNAseq related
- `pp.highly_variable_genes` # scRNAseq related
- `pp.log1p` # scRNAseq related
- `pp.normalize_total` # scRNAseq related
- `pp.regress_out` # scRNAseq related
- `pp.scale` # `get_anndata` takes care of this.
- `pp.subsample` # maybe useful.
- `pp.downsample_counts` # scRNAseq related
- `pp.combat` # batch effect.
- `tl.diffmap` # we have time series data, we don't need to infer.
- `tl.dpt` # we have time series data, we don't need to infer.
- `tl.ingest` # map labels and embedding from reference to new data.
- `tl.rank_gene_groups` # scRNAseq related
- `tl.filter_rank_genes_groups` # scRNAseq related
- `tl.mark_gene_overlap` # scRNAseq related
- `tl.score_gene` # scRNAseq related
- `tl.score_gene_cell_cycle` # scRNAseq related
- `tl.sim` # simulate dynamic gene expression data



**LUDDY**

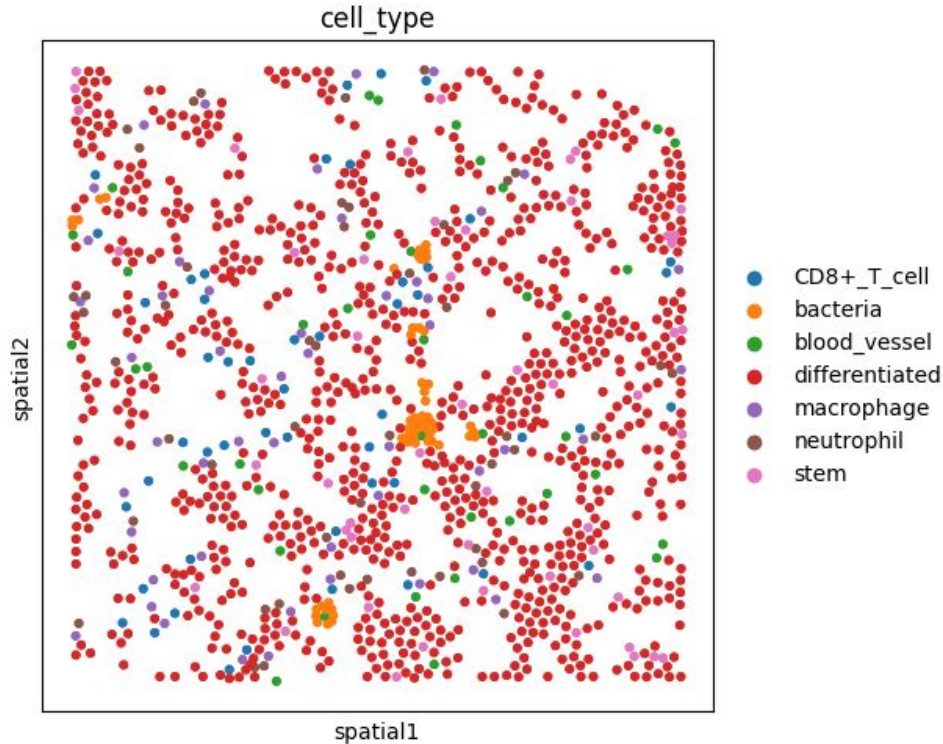
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

**PhysiCell.org**

 **@PhysiCell**

# squidpy: spatial\_scatter



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

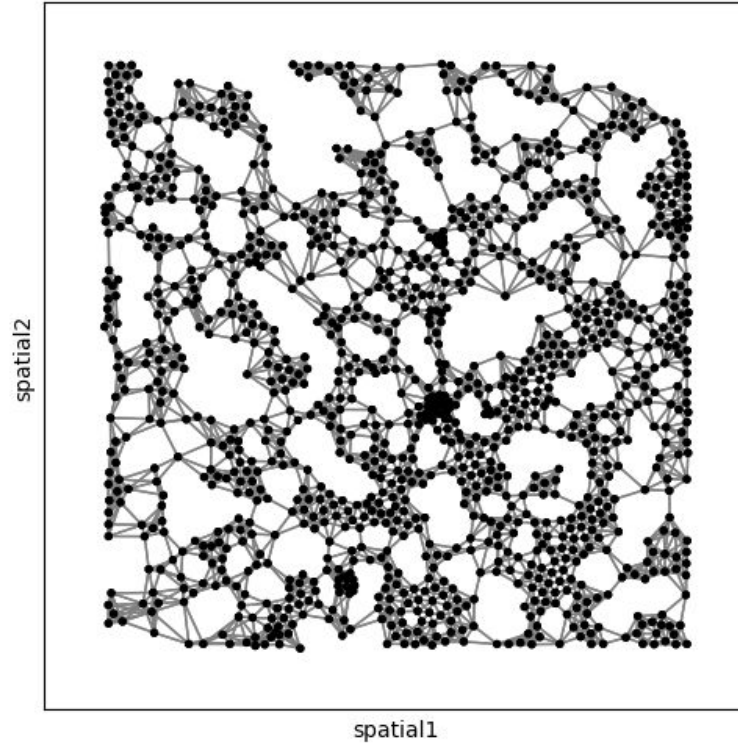
PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)



# squidpy: spatial\_neighbours



**LUDDY**

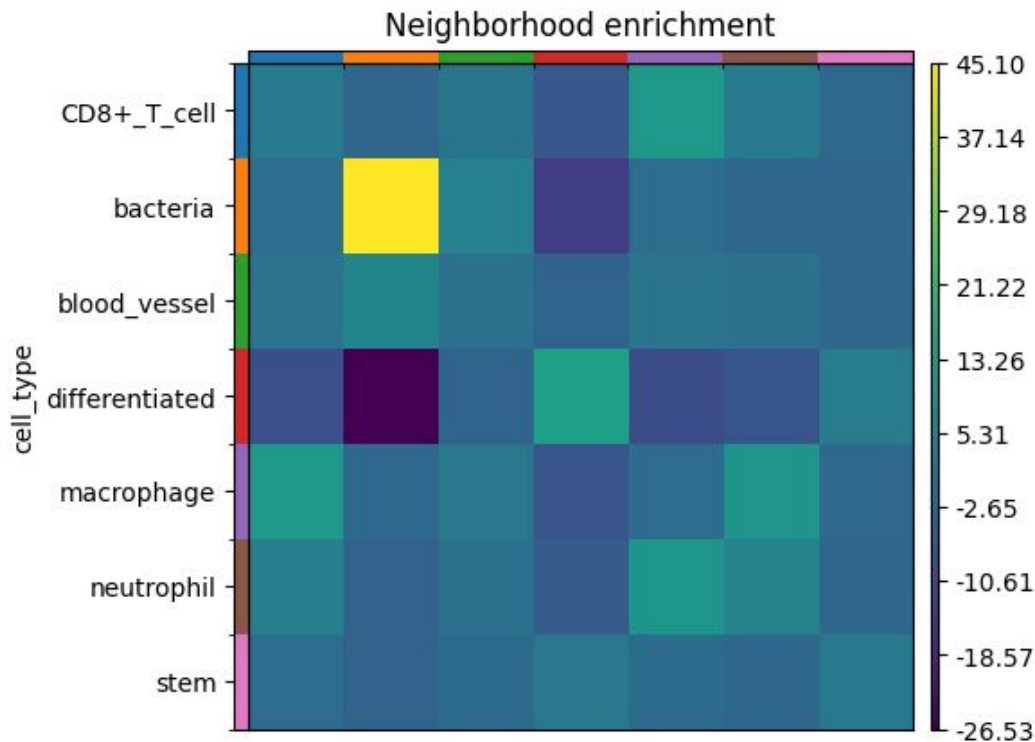
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# squidpy: nhood\_enrichment



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

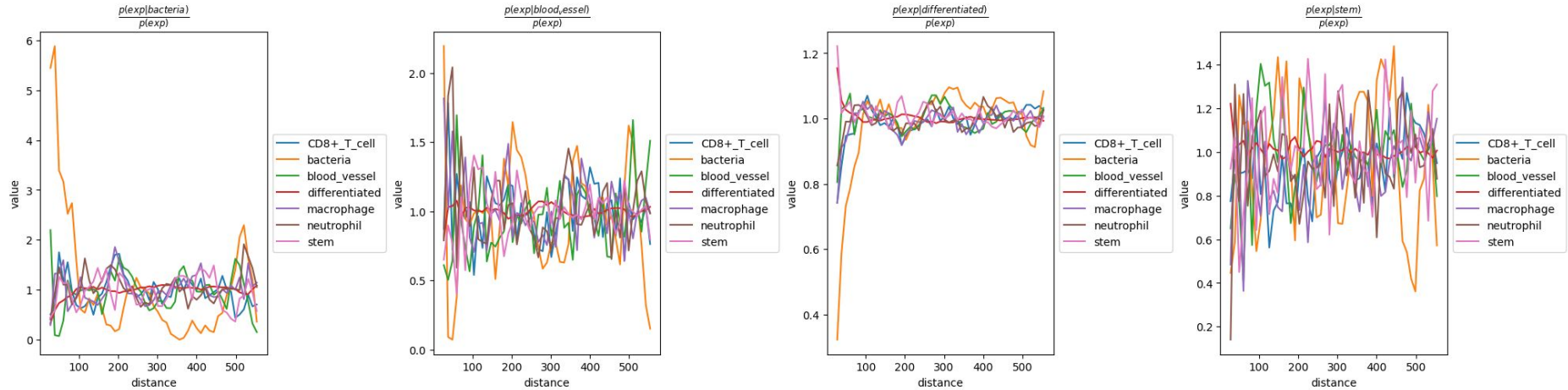
PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)



# squidpy: co\_occurrence



**LUDDY**

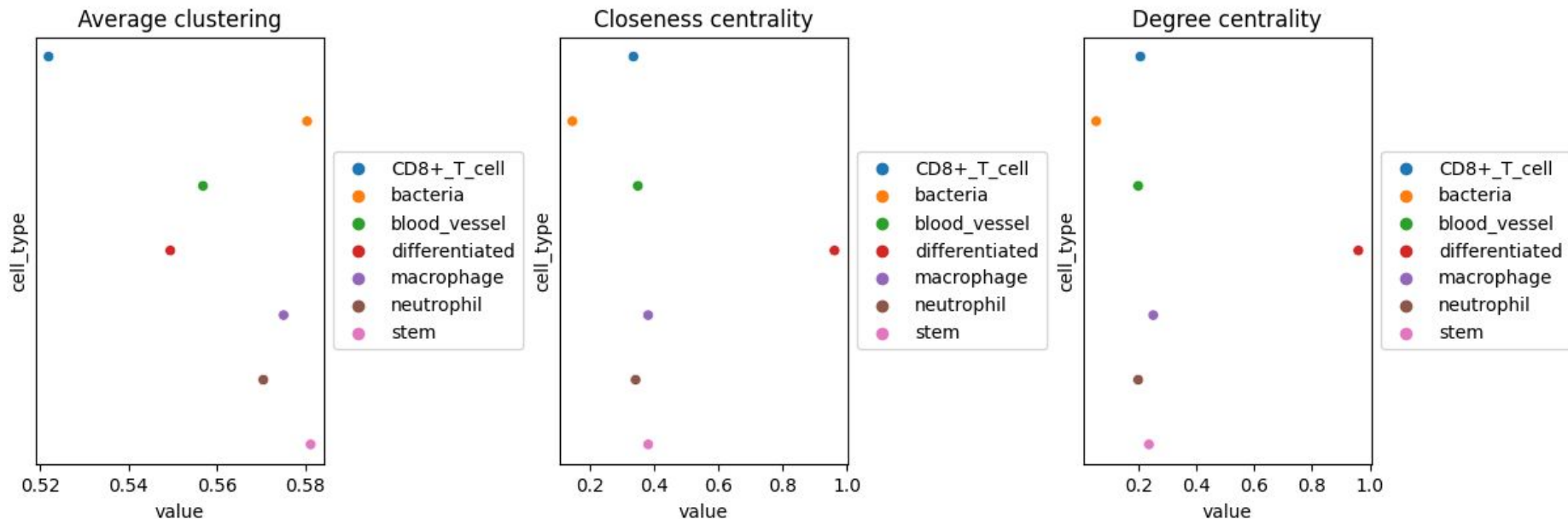
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

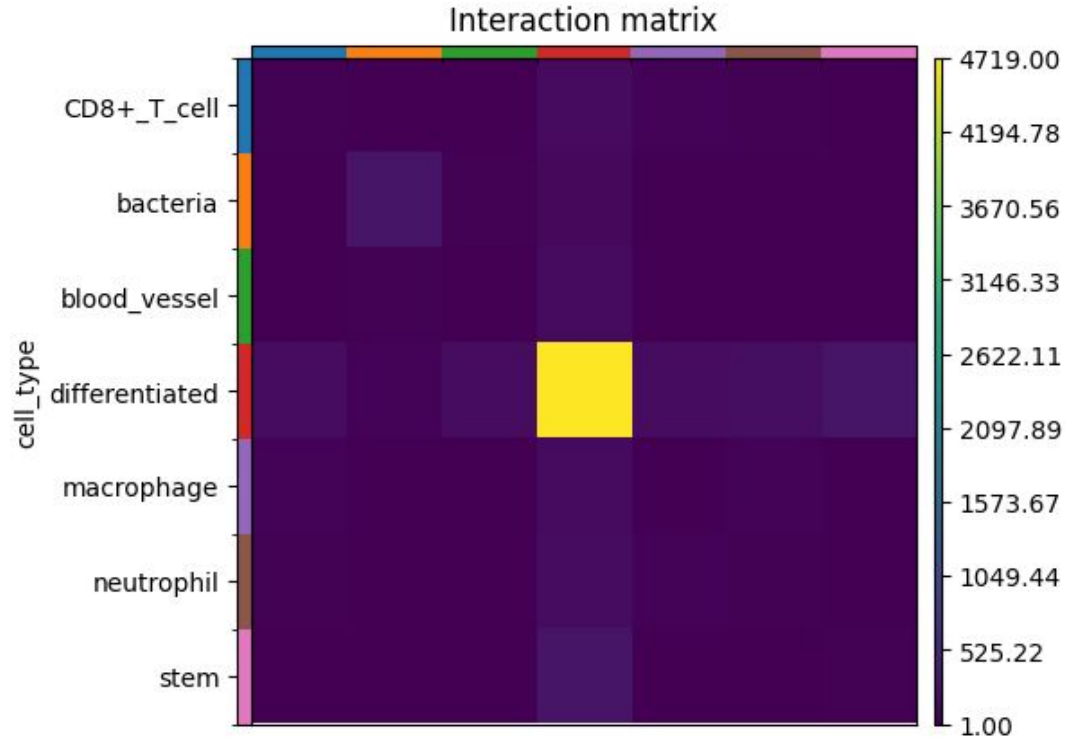
**PhysiCell.org**

 @PhysiCell

# squidpy: centrality\_scores



# squidpy: interaction\_matrix



**LUDDY**

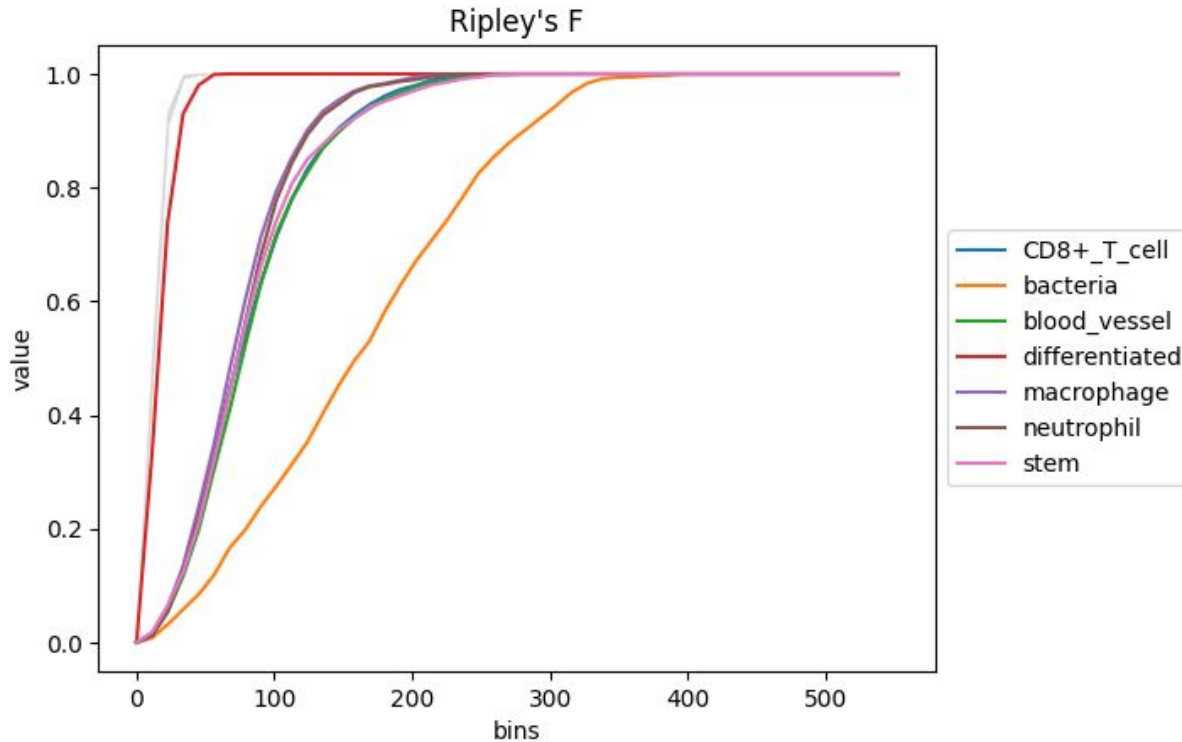
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

[@PhysiCell](https://twitter.com/PhysiCell)

# squidpy: Ripley's statistics for point processes



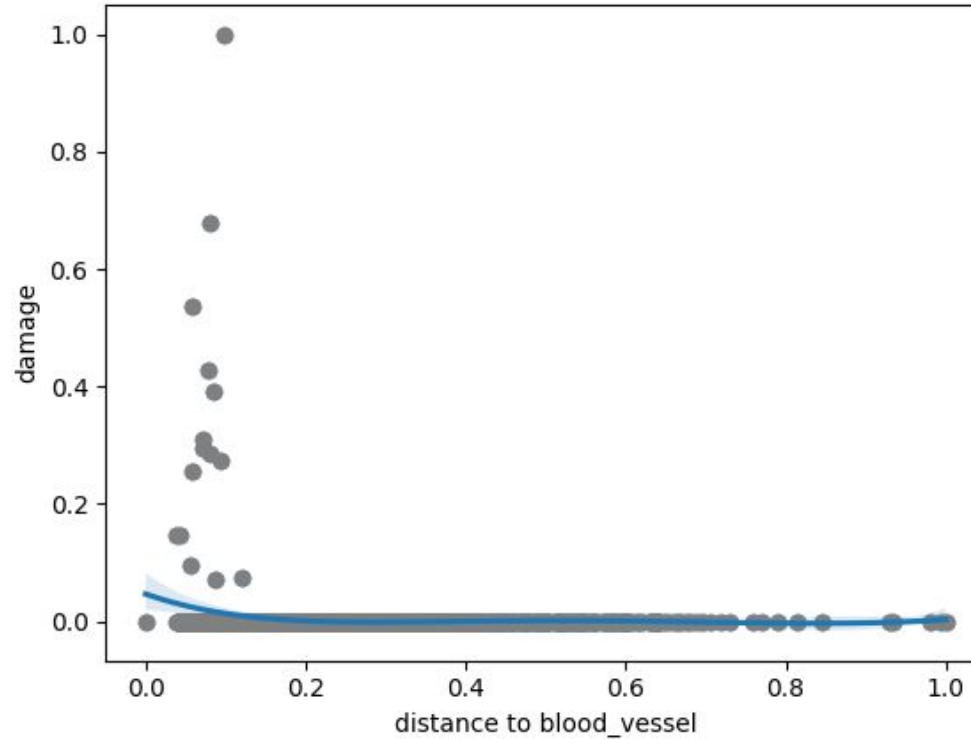
# squidpy: global autocorrelation statistic

## Moran's I and Geary's C

	I	pval_norm	var_norm	pval_norm_fdr_bh
position_vectorlength	0.992993	0.0	0.000273	0.0
quorum	0.934202	0.0	0.000273	
debris	0.927604	0.0	0.000273	
pro-inflammatory	0.870561	0.0	0.000273	
toxin	0.853246	0.0	0.000273	
damage_halfmax	0.603295	0.0	0.000273	
cell_cell_repulsion_strength	0.603295	0.0	0.000273	
relative_max_damage_death	0.603295	0.0	0.000273	
resource_chemotactic_sensitivities	0.603295	0.0	0.000273	
migration_speed	0.595514	0.0	0.000273	

	C	pval_norm	var_norm	pval_norm_fdr_bh
position_vectorlength	0.009540	0.0	0.000273	0.0
quorum	0.030820	0.0	0.000273	0.0
debris	0.062550	0.0	0.000273	0.0
toxin	0.080504	0.0	0.000273	0.0
pro-inflammatory	0.128708	0.0	0.000273	0.0
cell_cell_repulsion_strength	0.340886	0.0	0.000273	0.0
relative_max_damage_death	0.340886	0.0	0.000273	0.0
resource_chemotactic_sensitivities	0.340886	0.0	0.000273	0.0
damage_halfmax	0.340886	0.0	0.000273	0.0
migration_speed	0.346057	0.0	0.000273	0.0

# squidpy: var\_by\_distance



**LUDDY**

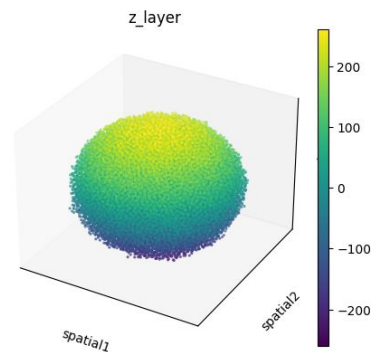
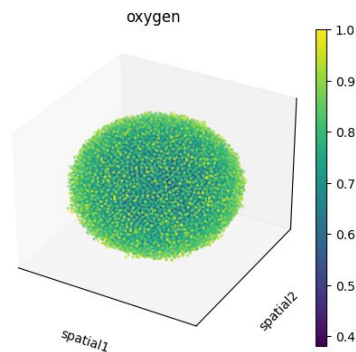
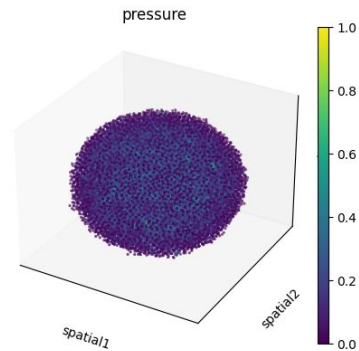
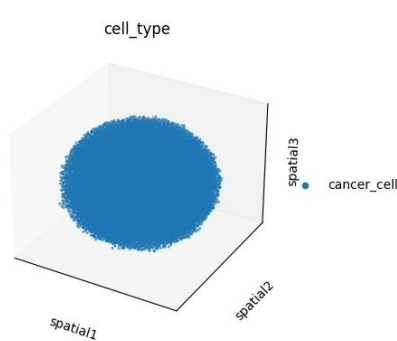
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

[PhysiCell.org](https://PhysiCell.org)

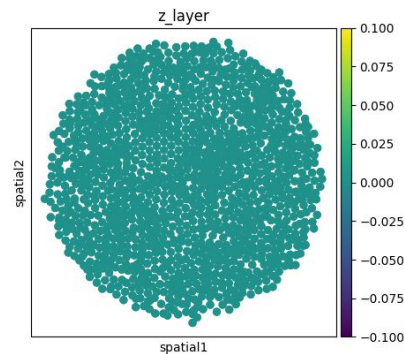
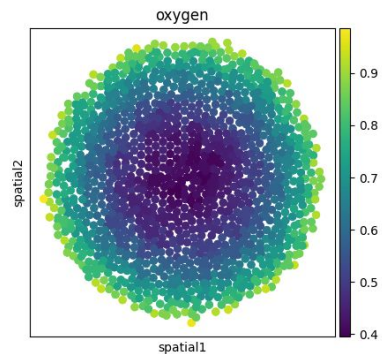
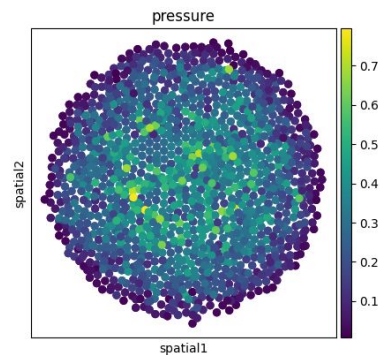
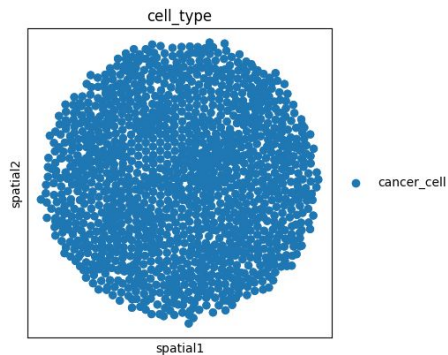
[@PhysiCell](https://twitter.com/PhysiCell)

# anndata & squidpy & 3D





# anndata & squidpy & 3D





# squidpy: “irrelevant” functions

- `gr.ligrec` # ligand receptor related
- `gr.sepal` # to identify spatially variable genes, but works only on square and hex matrices.
- `im.process` # h&e and microscopy image related
- `im.segment` # h&e and microscopy image related
- `im.calculate_image_features` # h&e and microscopy image related
- `pl.spatial_segment` # h&e and microscopy image related
- `sq.pl.extract` # make a temporary anndata object for plotting



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

PhysiCell Project

**PhysiCell.org**

 **@PhysiCell**



- deep learning.
- scvi-tools: single-cell variational inference tools.
- probabilistic models for single-cell omics data.
- publication: <https://doi.org/10.1038/s41587-021-01206-w>
- manual: <https://docs.scvi-tools.org/en/stable/index.html>
- pytorch: <https://pytorch.org/>
- pytorch lightning: <https://lightning.ai/>
- pyro: <https://pyro.ai/>



**LUDDY**

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

**PhysiCell Project**

**PhysiCell.org**

**@PhysiCell**



# more from scverse

homepage

- <https://scverse.org/>

data analysis

- **muon**: multimodal omics analysis.
- **scirpy**: T cell receptor (TCR) or B cell receptor (BCR) repertoires analysis from scRNA-seq data.
- scverse **ecosystem**: <https://scverse.org/packages/#ecosystem>



data formats

- **anndata**: <https://anndata.readthedocs.io/en/latest/>
- **spatialdata**: <https://spatialdata.scverse.org/en/latest/>
- **mudata**: <https://mudata.readthedocs.io/en/latest/>

