Pseudotime analysis with Stream

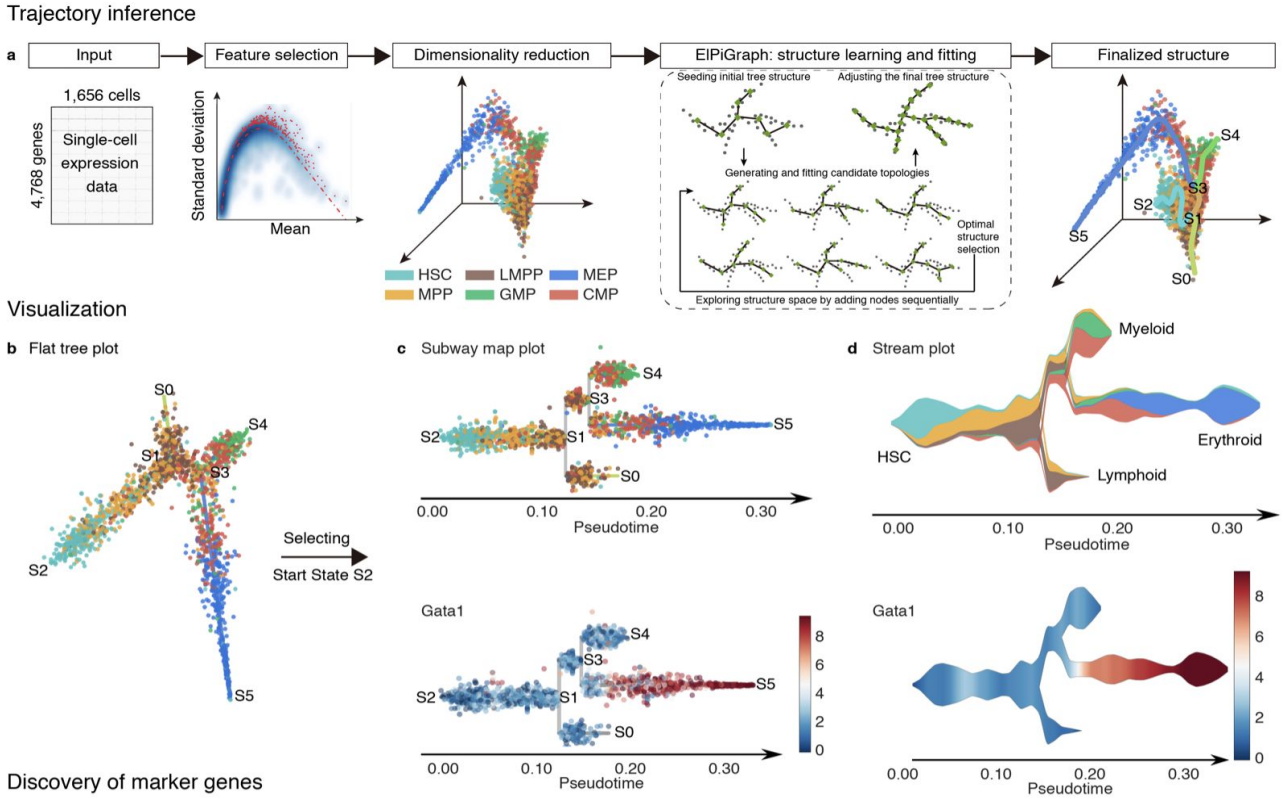
DJP Lab

Sara Masarone

27/02/2020

Stream (2019)

Pinello Lab



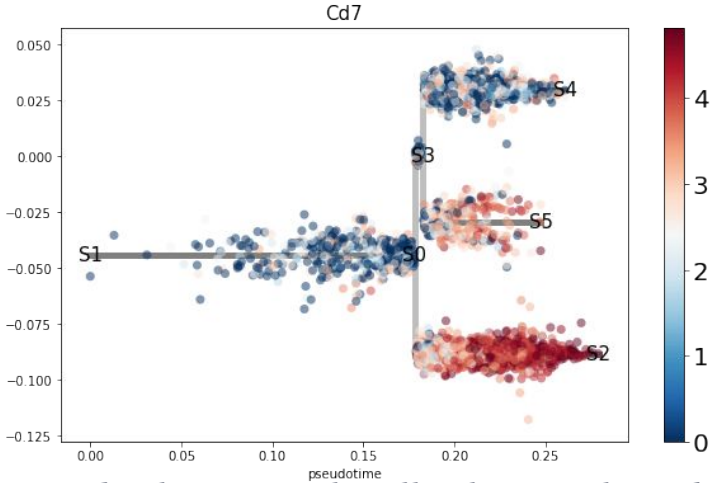
Stream

STREAM (Single-cell Trajectories Reconstruction, Exploration And Mapping) is an interactive pipeline capable of disentangling and visualizing complex branching (closed) trajectories from both single-cell transcriptomic and epigenomic data.

- captures dynamic state of cells such as differentiation, maturation etc, that cannot be captured by discrete clustering
- Single cells are ordered along a deterministic or probabilistic trajectories and a numeric value, called pseudotime, is assigned to each cell to indicate how far it has progressed along the dynamic process of interest.

Why STREAM?

Stream allows the user to reconstruct robust cell trajectories for RNASeq data and provides tools to explore relevant genes/biomarkers in each branch



- Current methods focus on displaying single cells/clusters along the pseudotime
- hard to study subpopulation composition and continuous transition along trajectories
- no trajectory inference method provides the possibility to map new cells to previously obtained reference trajectories without pooling cells and re-computing trajectories.

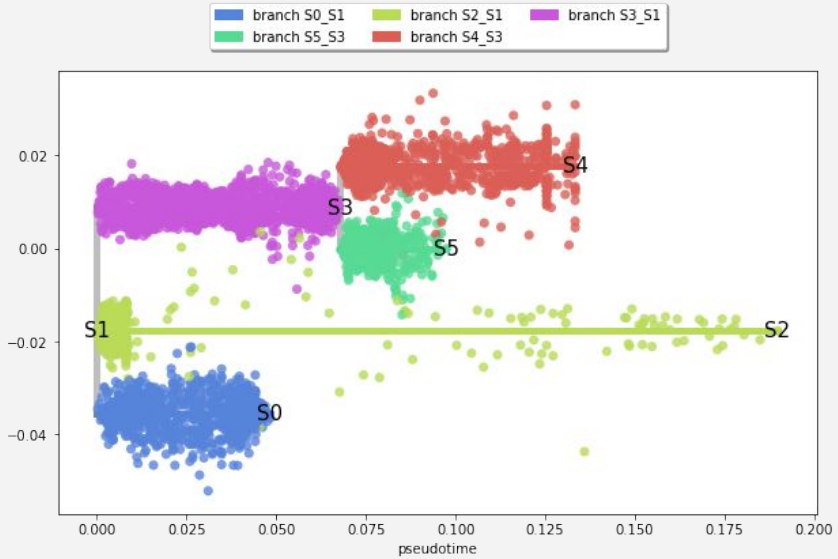
Analytical and computational challenges

Challenges of scRnaSeq data:

- cell to cell variation
- biological and technical noise
- sparsity of data
- even after feature selection, each cell still has hundreds of components, making it difficult to reliably assess similarity or distances between cells (a problem often referred as the “curse of dimensionality”)

Input: Single cell data (matrix, rows = genes, columns = cells)

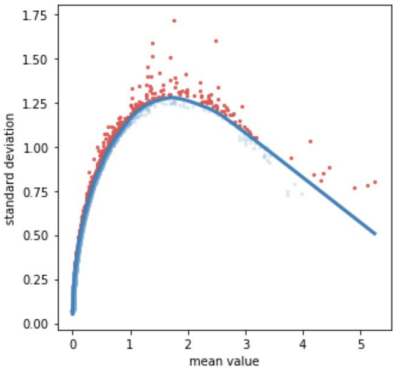
Output: a principal graph (3-4d to approximates the structure of the data)



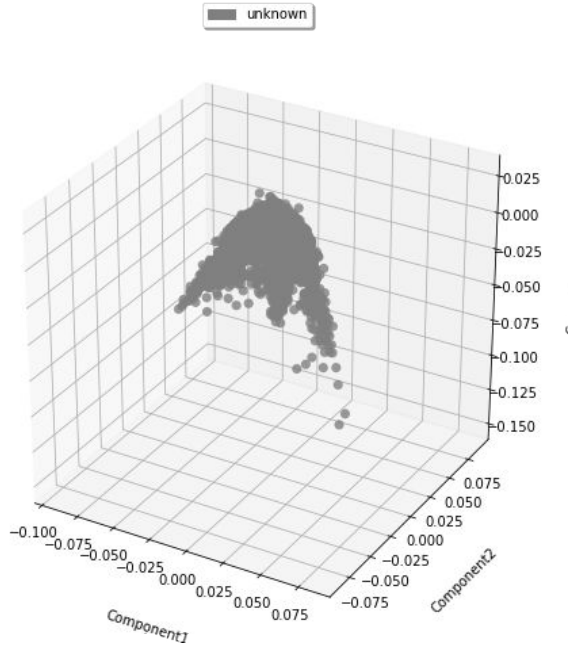
1) Identify salient features (variable genes/ top PCs)
2) Project features to a low dimensional space using non-linear dimensionality reduction called MLLE (Modified Locally Linear Embedding which preserves local structures)
3) Infers cellular trajectories using Elastic Principal Graph based on the elastic matrix Laplacian, trimmed squared error and control of overall topological complexity

Filtering/dimensionality reduction

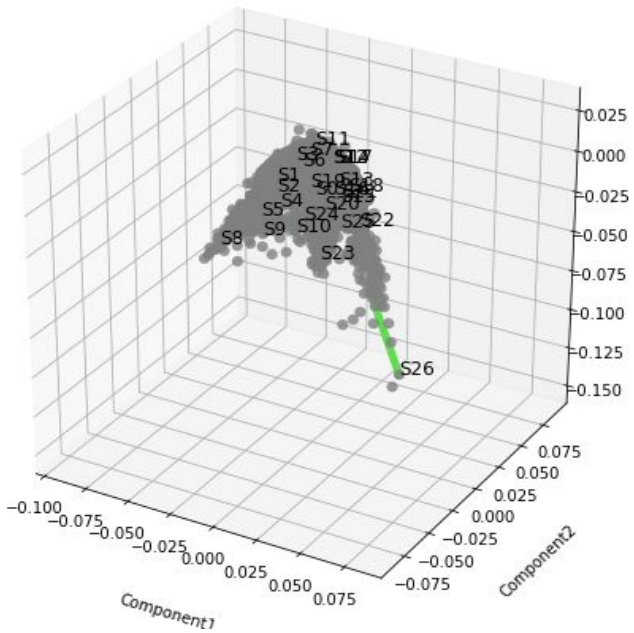
```
In [8]: #Filtering genes
        st.filter_genes(input_data, min_num_cells = 5)

        Filter genes based on min_num_cells
        After filtering out low-expressed genes:
        4038 cells, 11746 genes

In [9]: st.select_variable_genes(input_data,loess_frac=0.01, n_genes=500)

        500 variable genes are selected
```



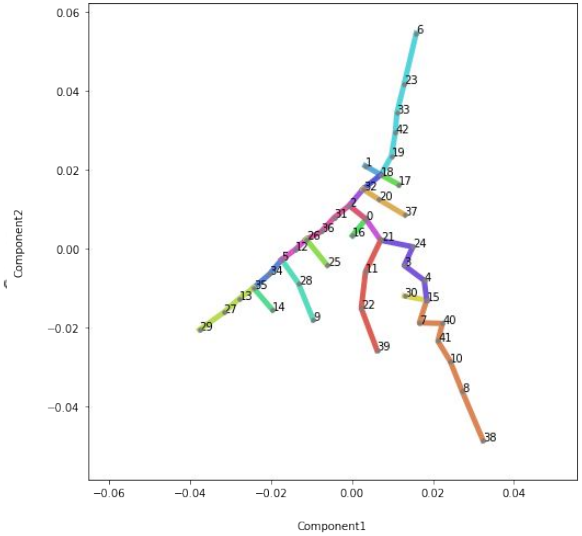
1) Filtering/feature selection



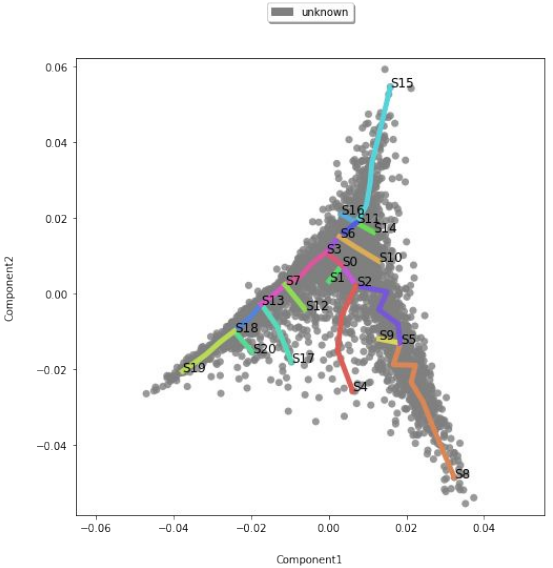
2) Dimensionality reduction (Modified Locally Linear Embedding)



3) Structure learning and fitting



Structure only

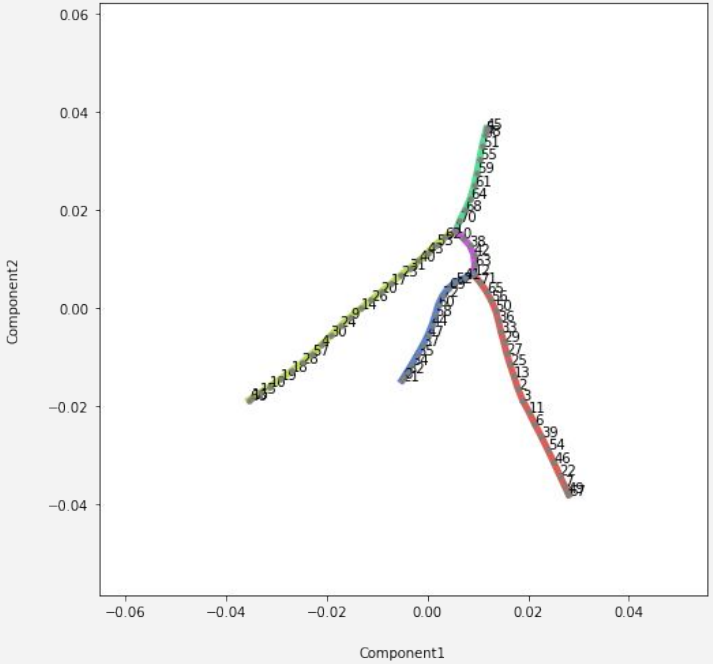


structure and density

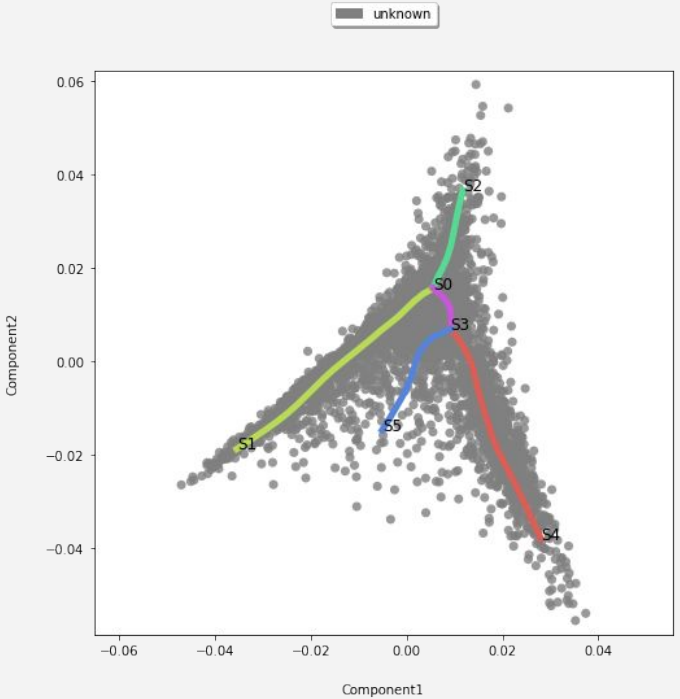
4) Tree structure learning and fitting by ElPiGraph.

- Optimal structure selected based on elastic energy minimization among a set of candidate structures that are constructed every time a tree node is added.
- The final tree is interpreted as a set of connected curves representing different trajectories.

Visualisation

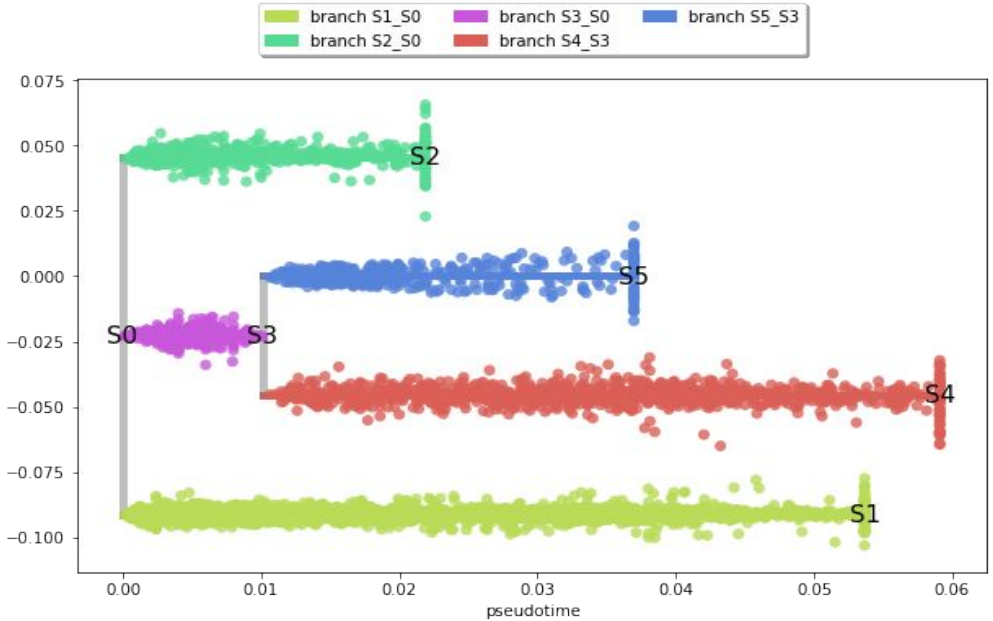


3) Tree representation of elastic graph in low dimensional space



3) Elastic graph in low dimensional space with density information

STREAM Subway plot

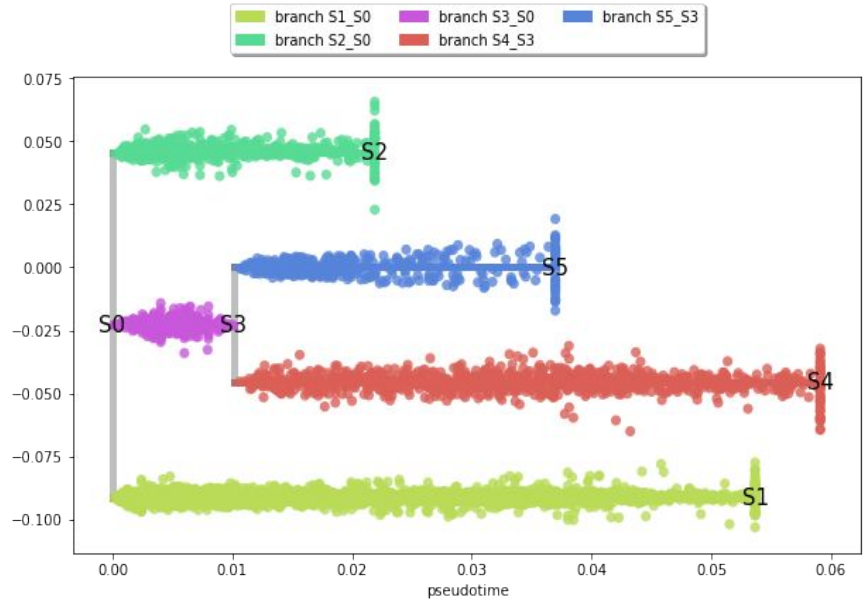


In **subway plots**, branches are represented as straight lines and each circle represents a single-cell. The lengths of the branches and the distances between cells and their assigned branches are preserved from the space where trajectories were inferred. In subway plots you need to select a starting point.

```
In [31]: st.subwaymap_plot(data_low,percentile_dist=100, color_by = 'branch')
```

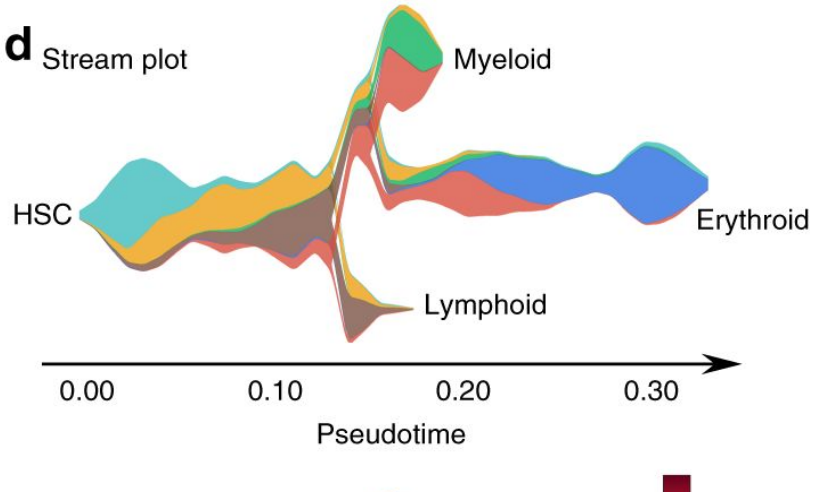
```
In [28]: st.subwaymap_plot(adata,root='S8')
```

Subway plot



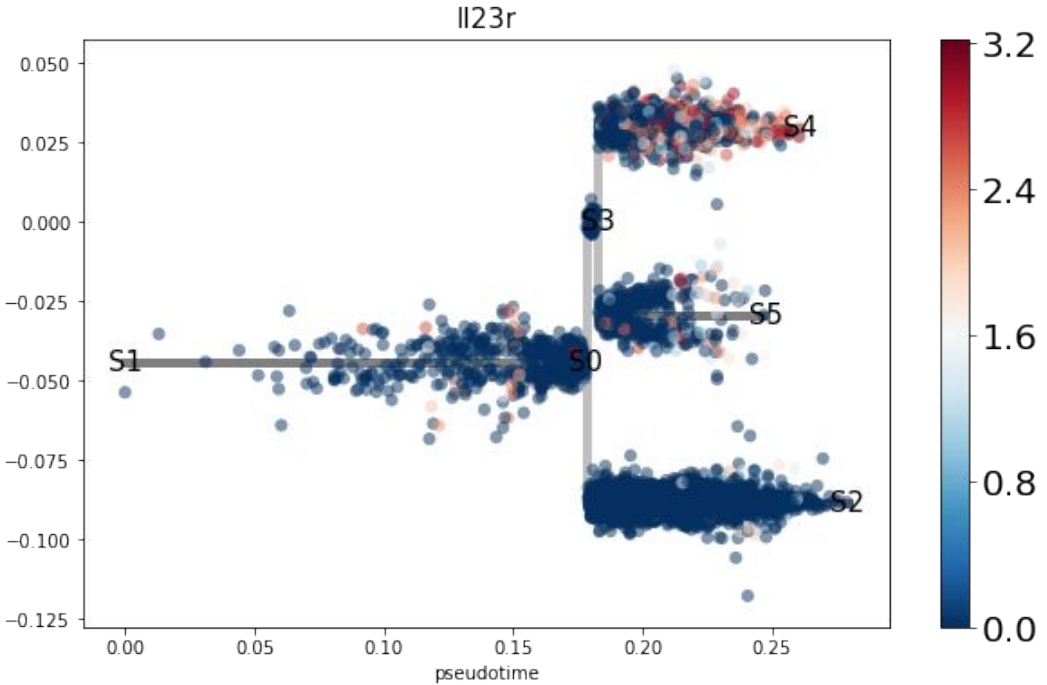
Subway plots capture trajectories and branching points, but they are not informative of the density and composition of cell types along pseudotime

Stream plot



- Having labelled cells you can explore how the cell populations change in density as the cells differentiate
- abundance and type of cells

Expression of key genes



Expression of key genes can be tracked along the pseudotime

Documentation/Resources

paper: https://www.nature.com/articles/s41467-019-09670-4#code-availability

github account: https://github.com/pinellolab/STREAM

web-app: http://stream.pinellolab.org