Machine Learning In Computational Biology

A comprehensive single cell transcriptional landscape of human hematopoietic progenitors.

Final Project Presentation

Presented by
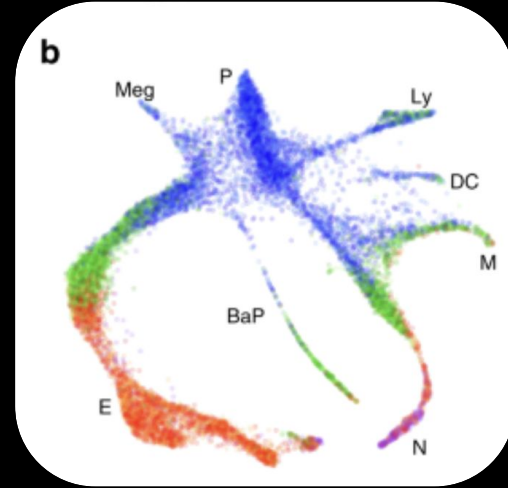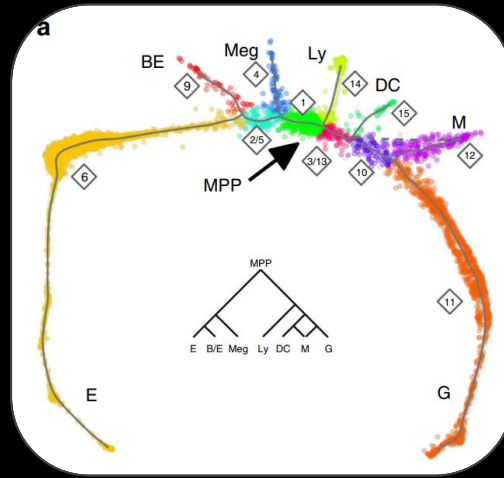
Aris Podotas   Rafail Adam   Georgios Leventis

Biological problem:

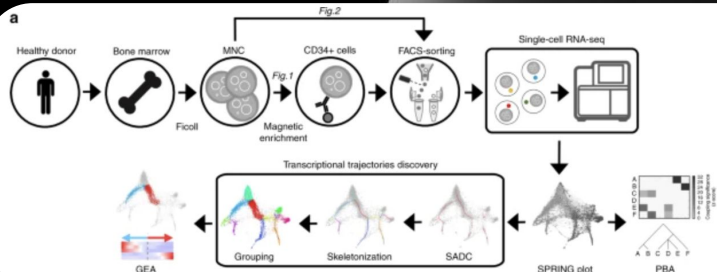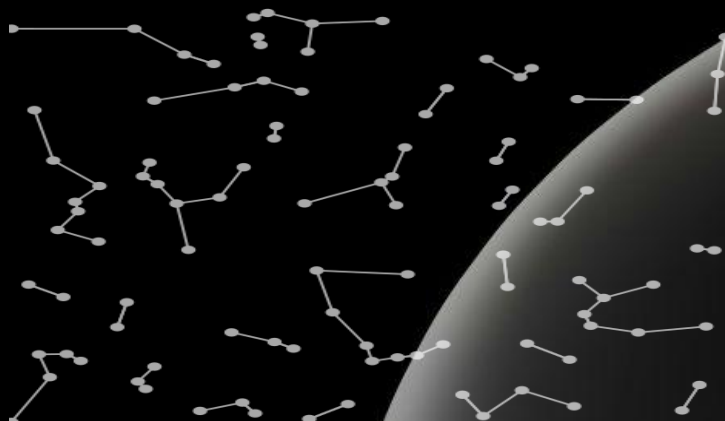Define a comprehensive differentiation trajectory for progenitor cells of of the hematopoietic lineage

- Many different progenitors
- Focus on CD34 is not enough

Importance:

Increase resolution regarding early differentiation by analysing whole bone marrow fraction

Method used

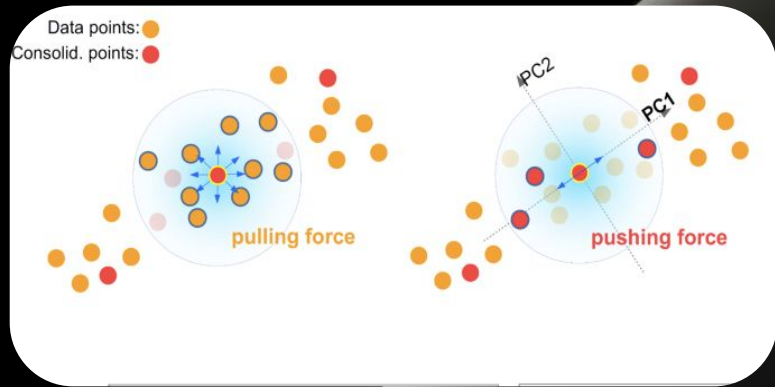Single-cell RNA-Seq on cells from the entire Bone Marrow + fractionation based on CD164

Filtering out cells based on mitochondrial gene expression threshold and normalizing expression values
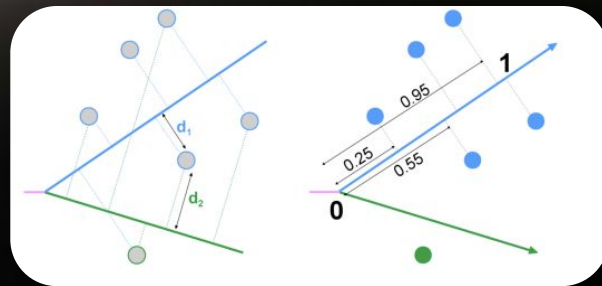
KNN graph (k=4), branching reconstruction and ordering of cells based on the structure
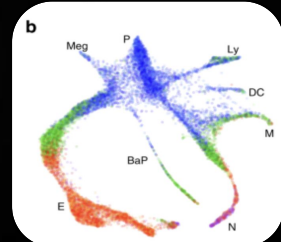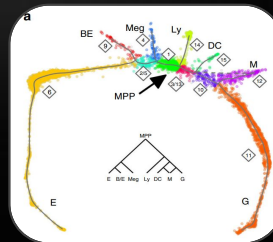
Structure-Aware filtering

Branch reconstruction:
Minimum spanning tree

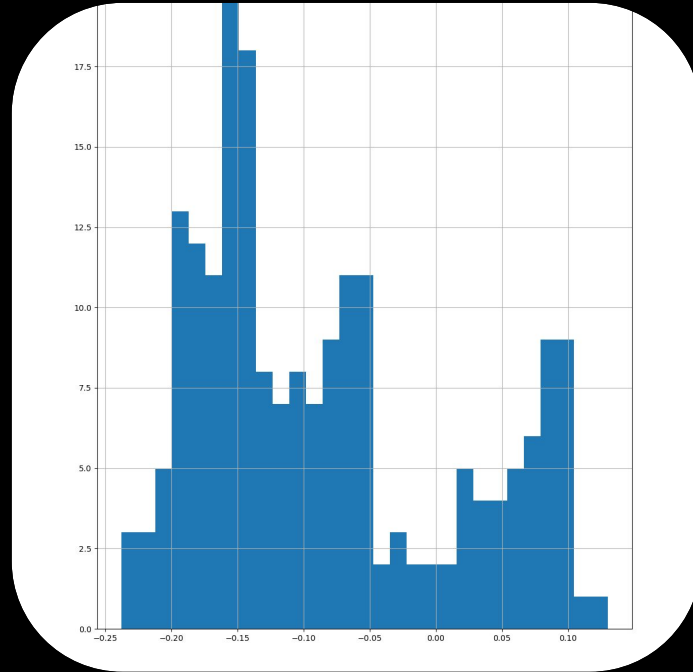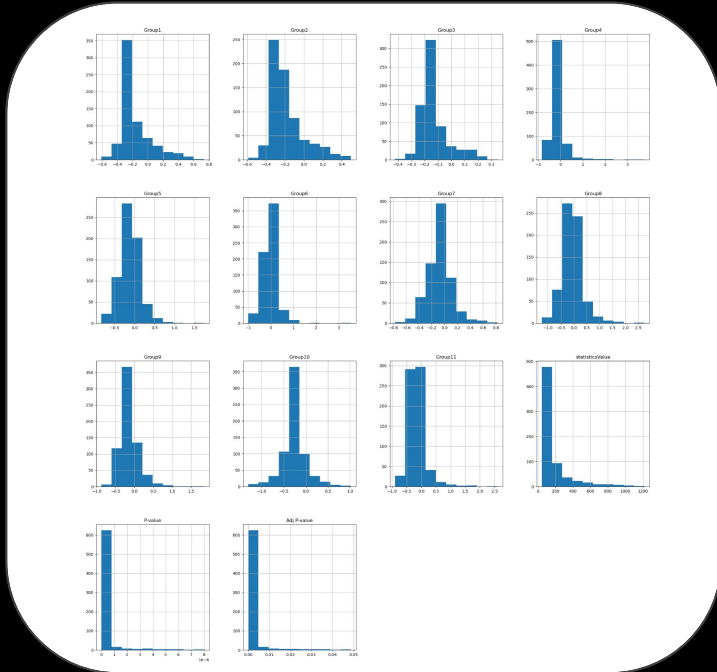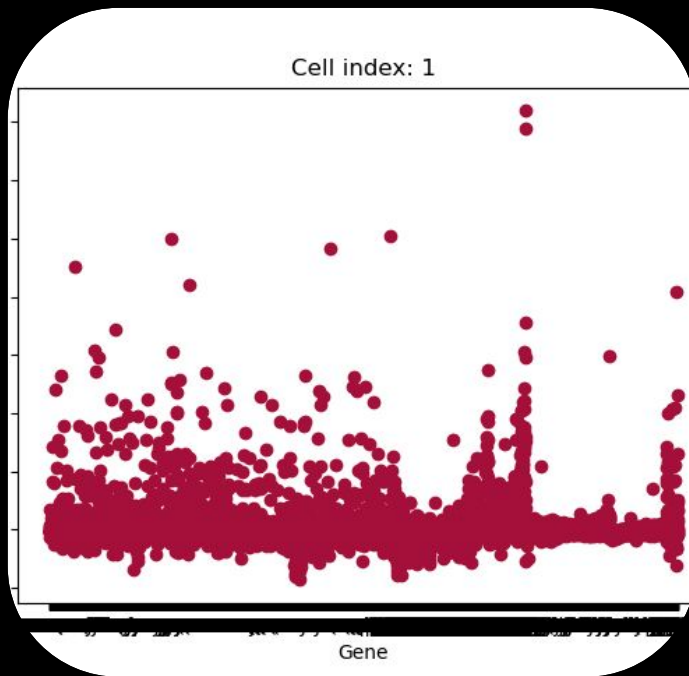# Branch association

Measure distances between
branches for ordering
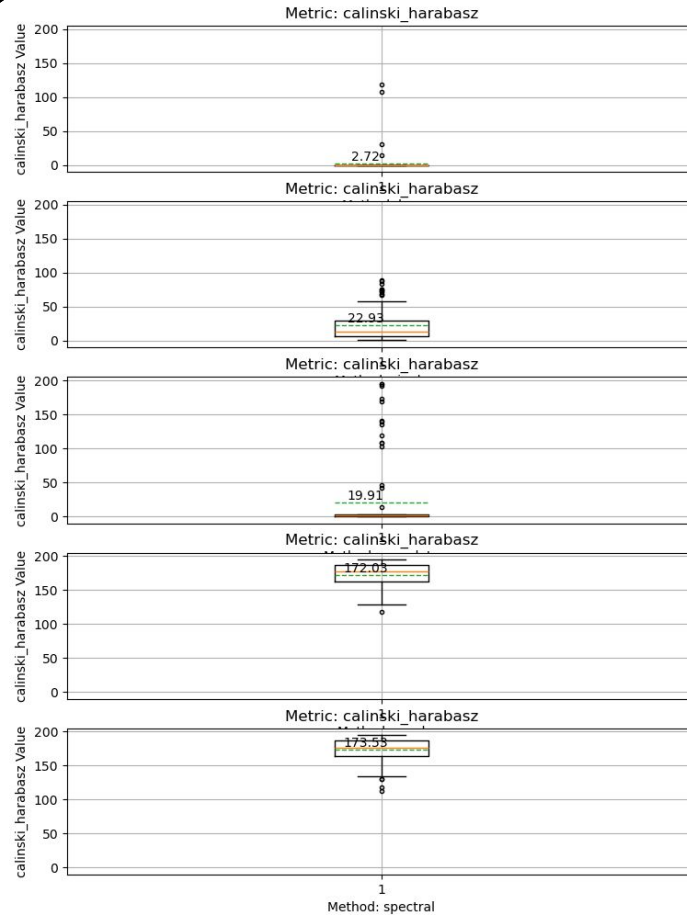
Representing expression domains
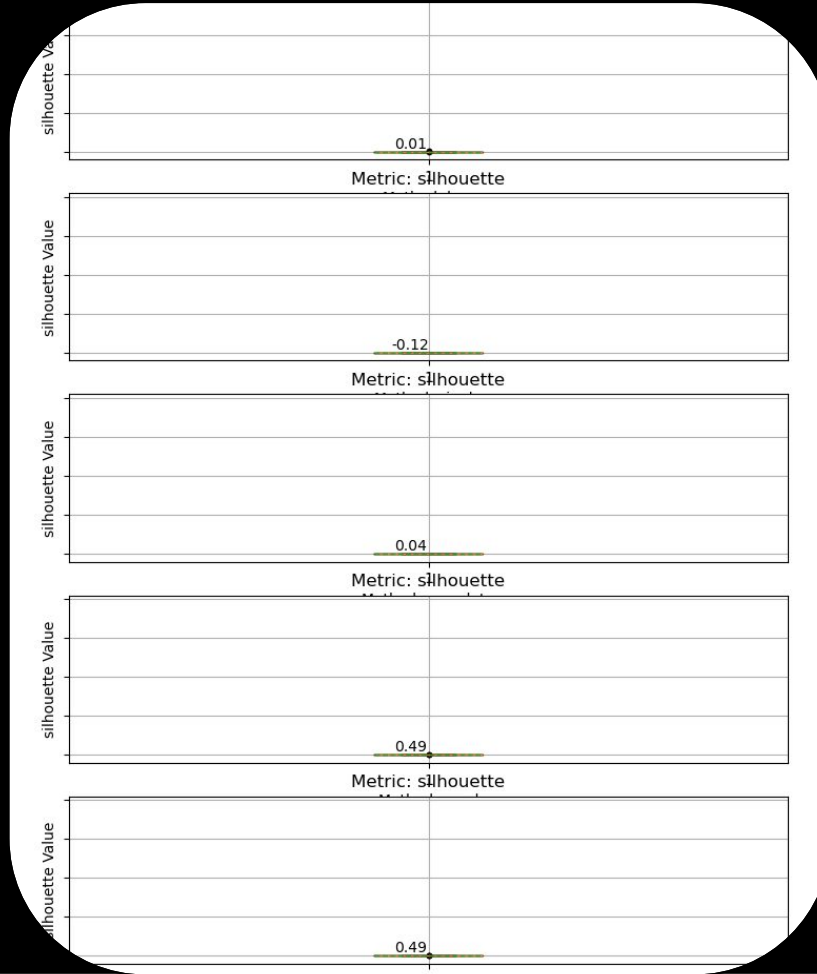
What the new cells look like

Results

Significance

KNN
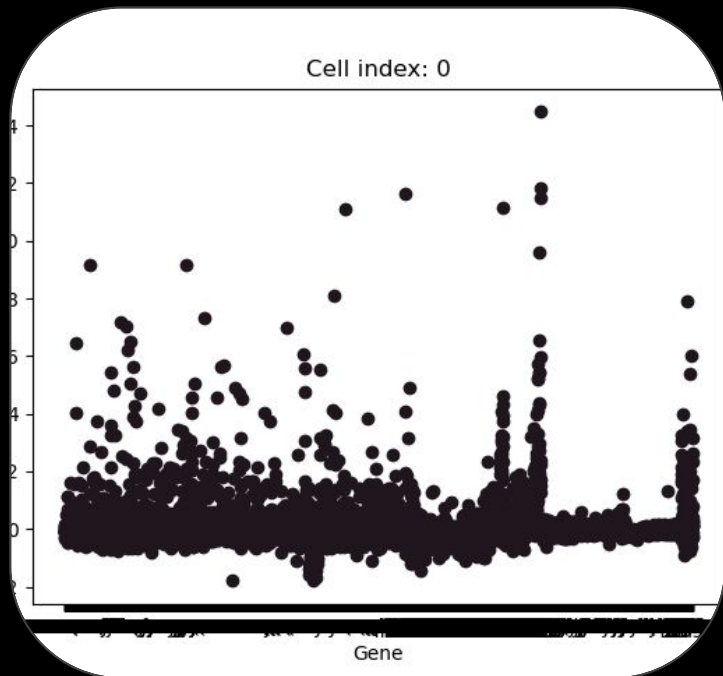
Single Link

Complete Link

Ward Algorithm
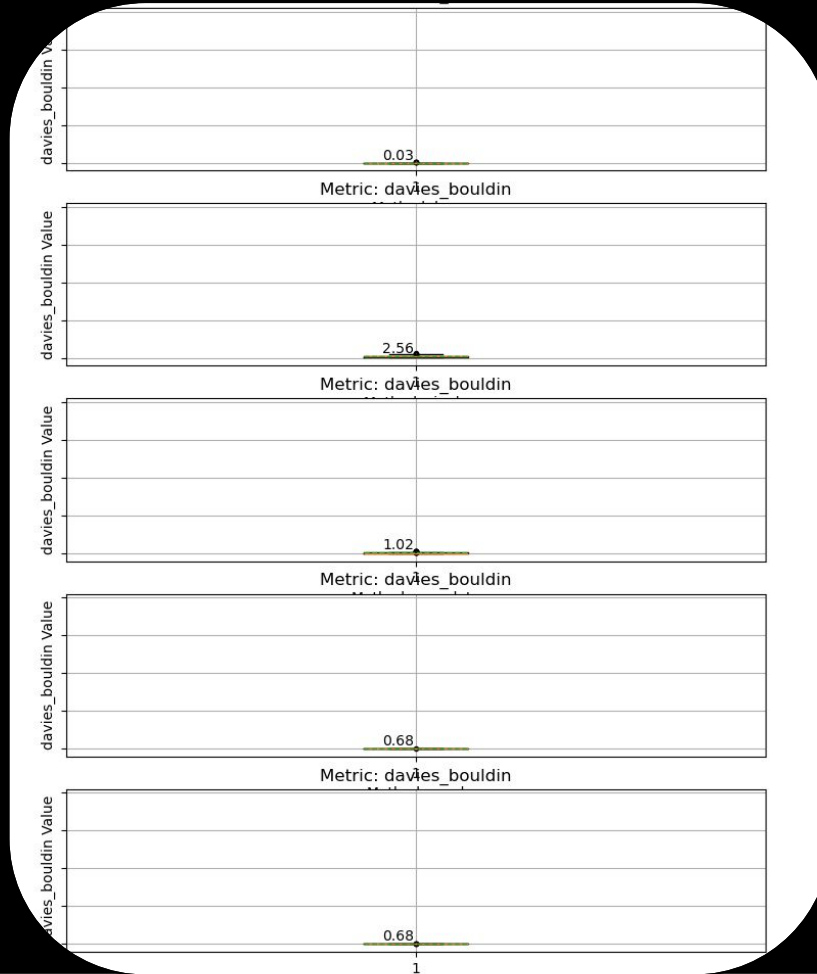
Spectral

Significance

KNN

Single Link

Complete Link

Ward Algorithm

Spectral

Metric: silhouette

0.01

-0.12

0.04

0.49

0.49

DSIT

2024 ⬛ 2025

Significance

KNN

Single Link

Complete Link

Ward Algorithm

Spectral

■ 2024 =2025

Still pending

Altering the KNN

Testing more values of K for the KNN (especially odd values)

Pending 1

Emulating PCA

The original citation used PCA with 40 components in the final data. This is a significant modifications since the original data has more than 5800 genes and reducing this matrix would yield large benefits in runtime

Pending 2

Gene Hierarchy

Using LDA to get the feature significance to know what genes contributed to formation of clusters the most

Pending 3

Citations

[1] D. Pellin *et al.*, "A comprehensive single cell transcriptional landscape of human hematopoietic progenitors," *Nat. Commun.*, vol. 10, no. 1, p. 2395, Jun. 2019, doi: 10.1038/s41467-019-10291-0.

[2] S. Wu, P. Bertholet, H. Huang, D. Cohen-Or, M. Gong, and M. Zwicker, "Structure-Aware Data Consolidation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2529–2537, Oct. 2018, doi: 10.1109/TPAMI.2017.2754254.

[3] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.

[4] A. P. Chatzigeorgiou and E. S. Manolakos, "MLscAN A flexible tool for single-cell data analysis pipelines and model selection using unsupervised machine learning methods".