THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

# Single cell topic Modeling and Multi-omics interpretation of cell topics.

**Team name:** SINGLE CELL RECITER

Nilesh Kumar[1], Dr. Virginie Grosboillot[2], Hammad Ali Hassan[3], Dr. Shahid Mukhtar[1] (Mentor)

[1]Department of Biology, UAB

[2]Department of Health Sciences and Technology (D-HEST), ETH Zürich. Zurich, Switzerland

[3]Dept of oral and maxillofacial surgery school of dentistry, UAB

# Background, challenge and significance.



- Clustering of cells based on gene expression is one of the major challenge.
- As current approach k-NN and SNN are used for clustering the cells.
- One of the disadvantage of k-NN is that it non-parametric.
- Latent Dirichlet Allocation (LDA) is a popular topic modeling technique to extract topics from a given corpus when it comes to NLP.
- In the case of topic modeling, the process helps in estimating what are the chances of the words/**genes**, which are spread over the document/**cells**, will occur again?

**Parametric** tests usually have more **statistical power** than nonparametric tests, no assumptions are made about the shape of the **decision boundary**. It does not tell us which **predictors are important**.



© UAB. All Rights Reserved.

# Background, challenge and significance.



**The LDA makes two key assumptions:**

1. **Samples**/Documents are a **mixture** of **Cells**/topics, and
2. **Cells**/Topics are a mixture of **Genes** (tokens or words)

Using the Monte Carlo estimate of the gene by topic matrix we can calculate predicted marker gene whereas traditionally deferential gene expression analysis is need to predict marker gene. Additionally, single-cell DE methods are poised to produce false discoveries and marker predict can be improved by using LDA Monte Carlo estimates.



© UAB. All Rights Reserved.

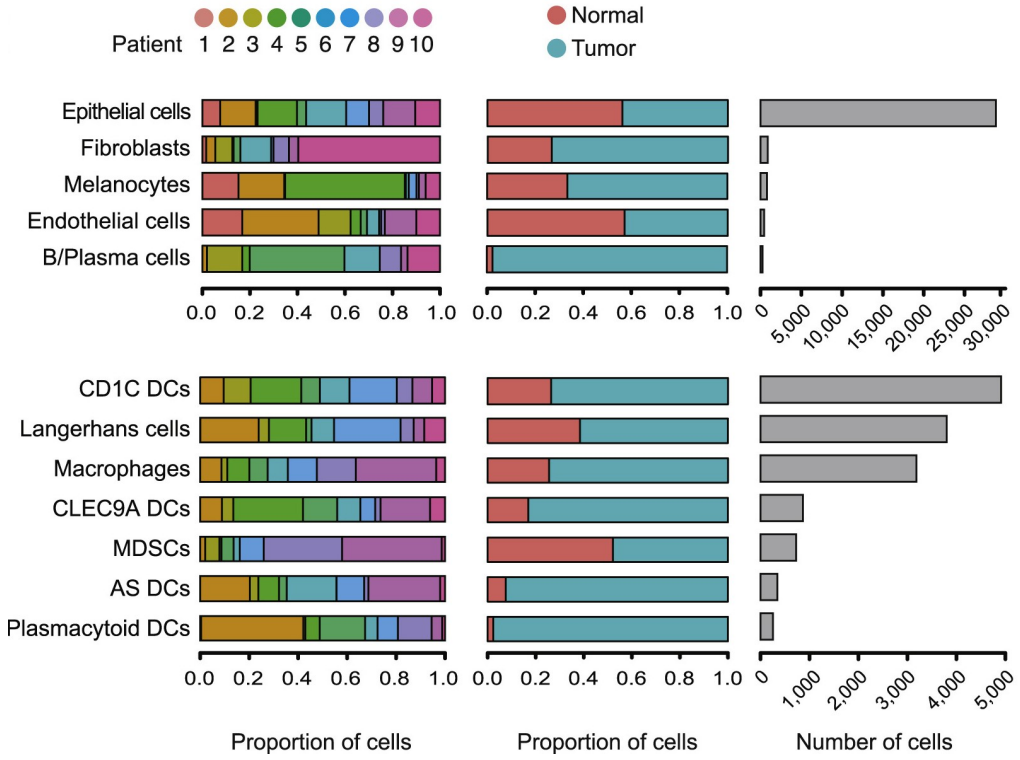
# Dataset

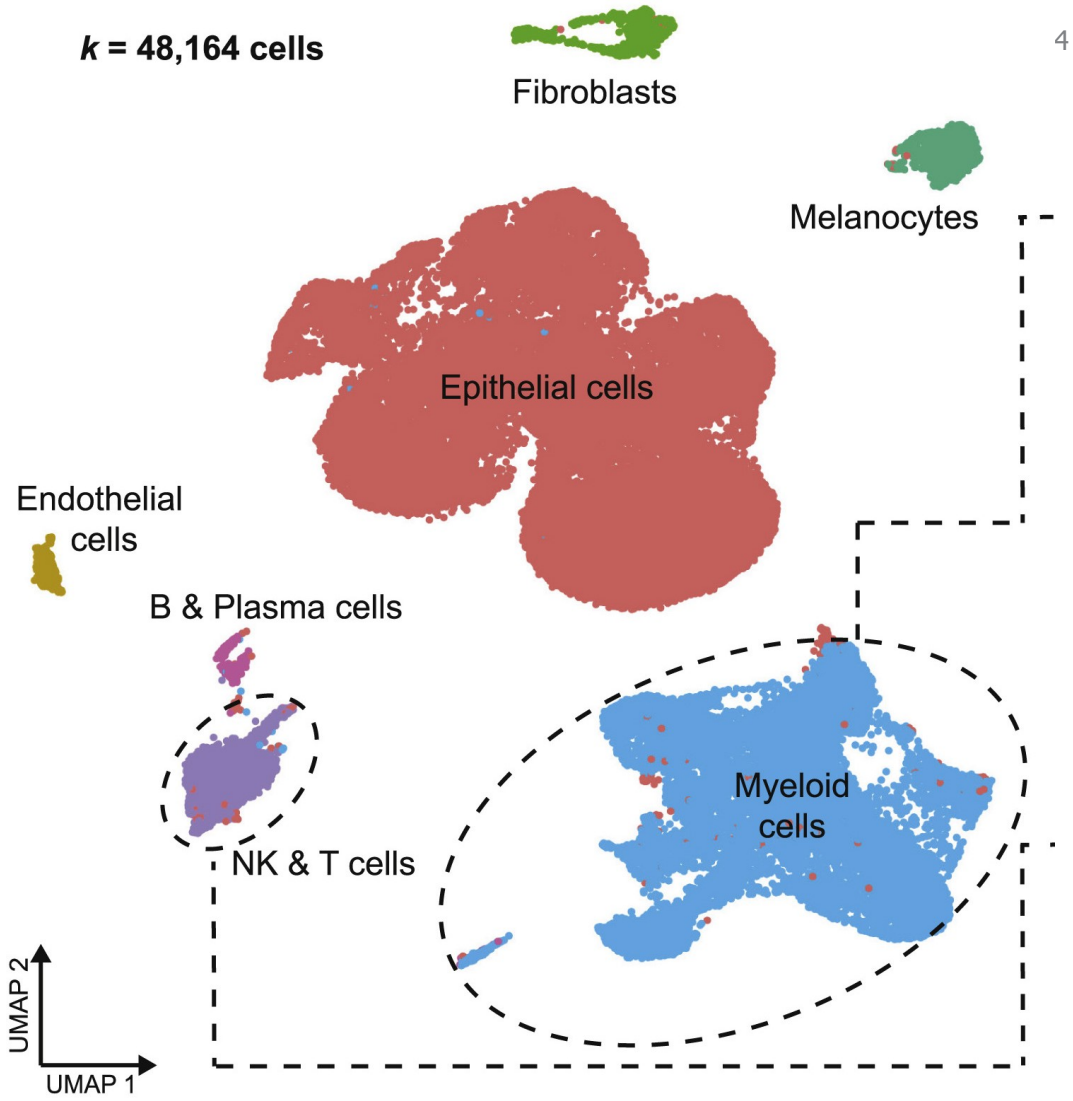
RESOURCE | VOLUME 182, ISSUE 2, P497-514.E22, JULY 23, 2020

## Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma

Andrew L. Ji [8] • Adam J. Rubin [8] • Kim Thrane [8] • ... Garry P. Nolan • Joakim Lundeberg • Paul A. Khavari [9] • Show all authors • Show footnotes

Open Access • Published: June 23, 2020 • DOI: https://doi.org/10.1016/j.cell.2020.05.039 •






© UAB. All Rights Reserved.

# Pipeline

scRNA Data

10x GENOMICS

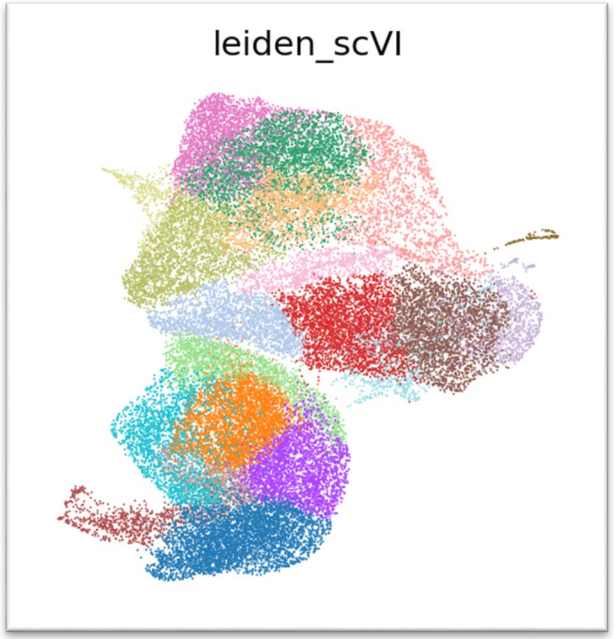
Cell ranger

scvi-tools

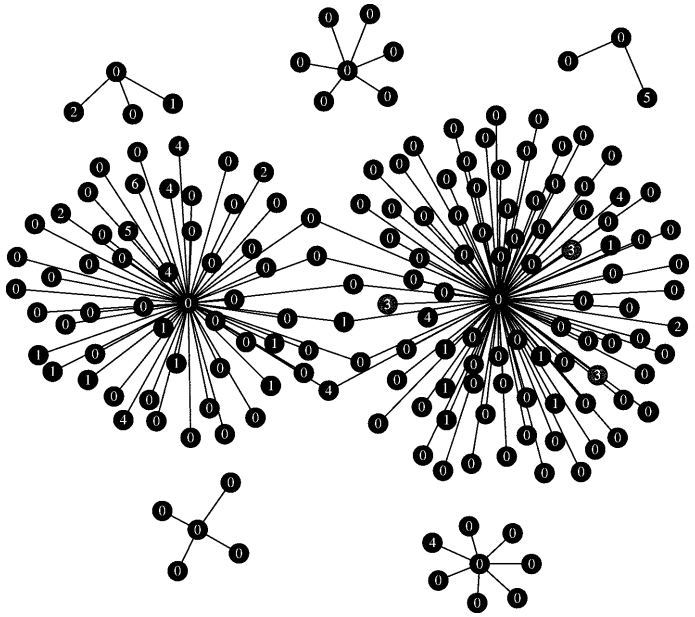
pandas

Topic Modeling

Enrichr




Topic and cells

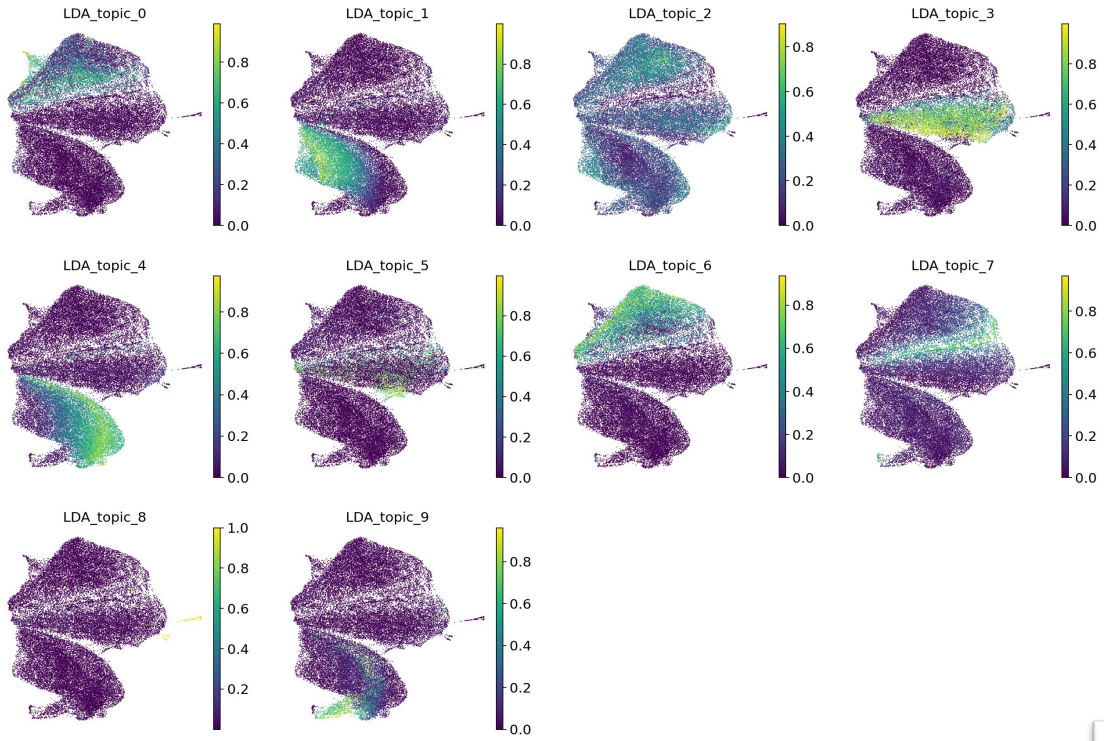


pseudo co-expression network
(Future work)



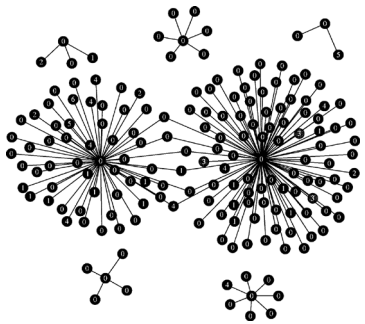
© UAB. All Rights Reserved.

# Results




**Topics cell clusters**

| | topic_0 | topic_0_prop | topic_1 | topic_1_prop | topic_2 | topic_2_prop |
|---|---|---|---|---|---|---|
| 0 | TMSB4X | 0.126198 | LYZ | 0.071308 | LYZ | 0.056180 |
| 1 | TMSB10 | 0.086073 | ACTB | 0.071091 | S100A9 | 0.053984 |
| 2 | ACTB | 0.084189 | HLA-DRA | 0.040470 | FTH1 | 0.050848 |
| 3 | JUNB | 0.035964 | CD74 | 0.036424 | FTL | 0.050564 |
| 4 | FTL | 0.028489 | TMSB4X | 0.035223 | ACTB | 0.034482 |



**Topics gene representations**

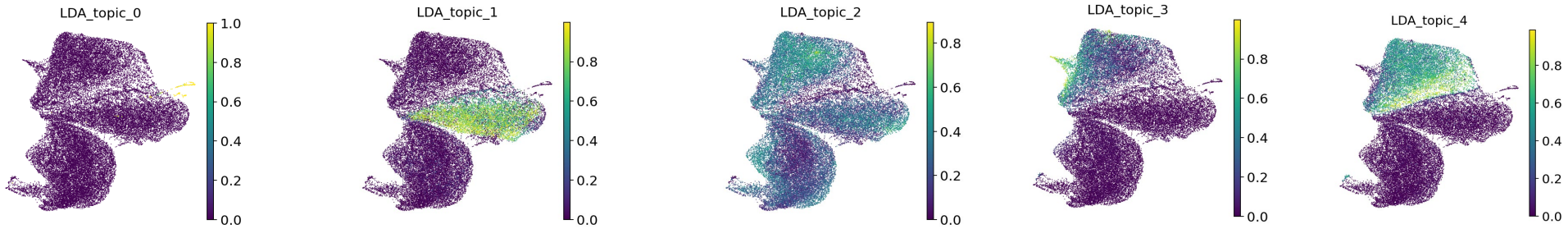
Enrichr



© UAB. All Rights Reserved.

# Results: Multi-omics enrichment





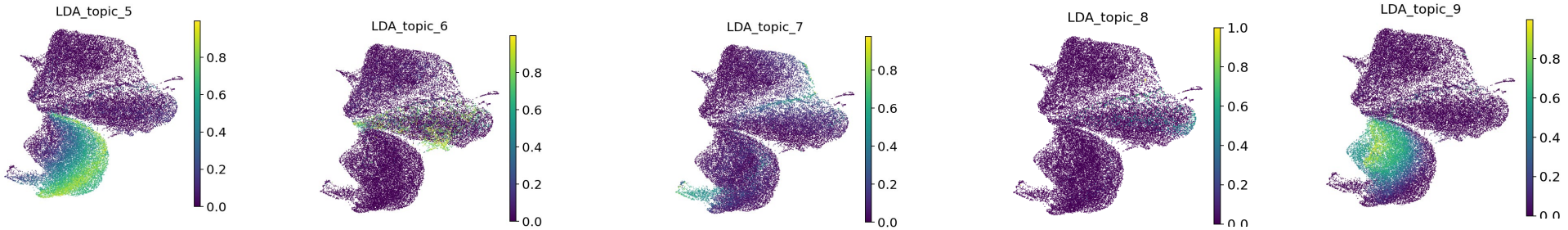

| | | | | | |
|---|---|---|---|---|---|
| Cell Type | CD16+ Monocyte | CD56-bright Natural Killer | Oligodendrocyte Precursor Cell | CD14+ Monocyte | CD56-bright Natural Killer |
| TFs Coexp | | | | EIF3K, HES2, GRHL1 | MSX1, OTX1, STAT4, DMRTA1, SOX21 |
| Kinases Coexp | | | | | LCK |
| lncRNA Coexp | LINC02364 | | | LINC02479 | IGBP1-AS2, LINC02067, LINC02364, TLR8-AS1 |
| ClinVar Coexp | | | | | autosomal dominant non-syndromic sensorineural deafness type dfna |
| PhenGenI Association | Hepatitis C | | | Reverse Transcriptase Inhibitors; Drug Hypersensitivity Syndrome | |



© UAB. All Rights Reserved.

# Results: Multi-omics enrichment






| | | | | | |
|---|---|---|---|---|---|
| Cell Type | CD8+ Naive T | CD14+ Monocyte | Plasmacytoid Dendritic Cell | Layer 5-6 Glutamatergic Neuron, Near-Projectin | Oligodendrocyte Precursor Cell |
| TFs Coexp | | | FOXA2, FOXB1, HOPX, ID2, MGMT | HES5, LENG9 | BARX2, DLX3, FOXN1, GRHL1, GRHL3 |
| Kinases Coexp | CSNK1A1L, IRAK3 | | PIM2 | | |
| lncRNA Coexp | | LINC02647 | LINC02067, LINC02479 | | LINC01181 |
| ClinVar Coexp | | | | | |
| PhenGenI Association | | | | | diamond-blackfan anemia, mitochondrial diseases, retinitis pigmentosa |



© UAB. All Rights Reserved.

# Future directions



- A variation of LDA will used to eliminate the need number of topic parameter.
- scRNA Co-expression network inference.
- scRNA Gene regulatory network inference.



© UAB. All Rights Reserved.

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

# Thank you!