



האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

Artificial Intelligence in Medicine

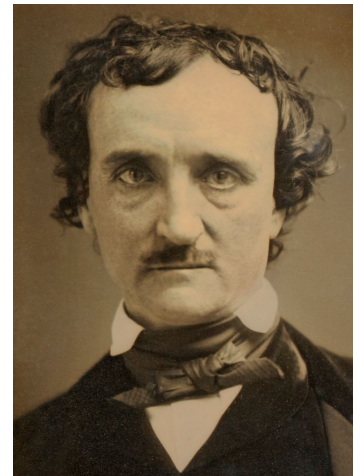
Visualizations and Dimensionality Reduction

Nir Friedman and Tommy Kaplan

26/2/24

“Believe half of what you see...”

Edgar Allan Poe



תוכן

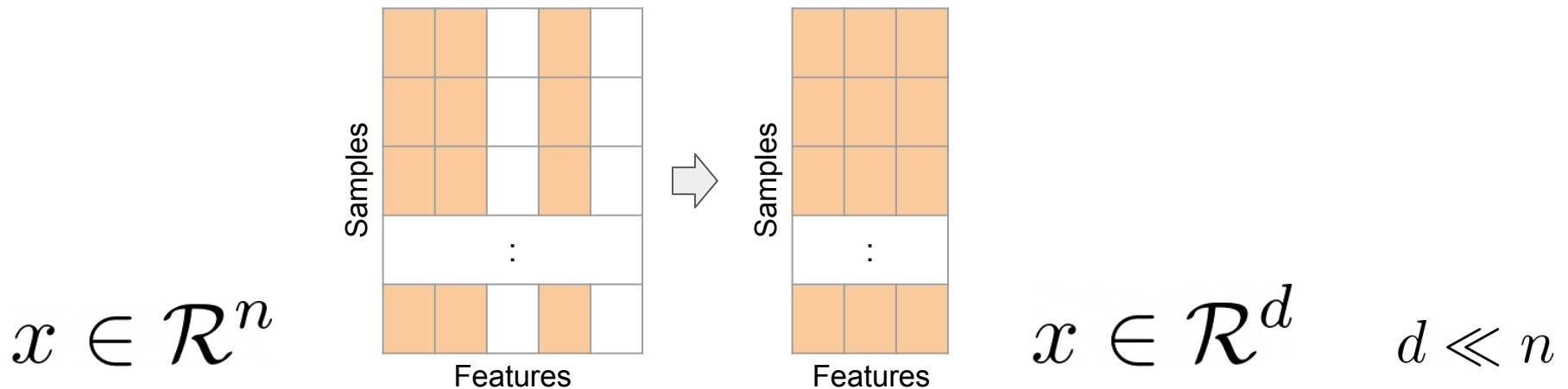
- למה להוריד מימד
 - להיפטר ממדידות לא רלוונטיות
 - לאחד מדידות קורלטיביות
 - להתגבר על רעש, אובר-פיטינג, curse of dimensionality
 - לאפשר אינטרפרטציה ותצוגה של הדוגמאות (קלאסטרिंग, רצף)
- הורדת מימד כרוכה בעיות - כמה מדוייק הייצוג החדש?
- מיזעור העיות הכולל - הורדת מימד לינארית PCA
- מיזעור העיות בין שכנים (קרובים) - הורדת מימד tSNE
- דוגמא כצירוף ארכיטיפים - זיהוי פנים, ניתוח מסמכים

Why dimensionality reduction?

- Represent data with fewer dimensions

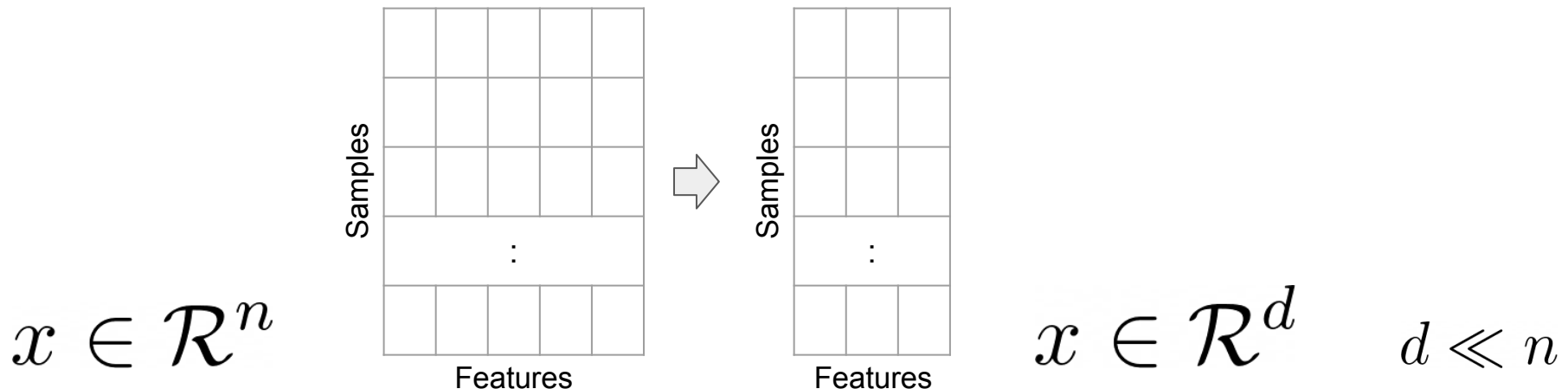
Why dimensionality reduction?

- Represent data with fewer dimensions
- Remove irrelevant data features

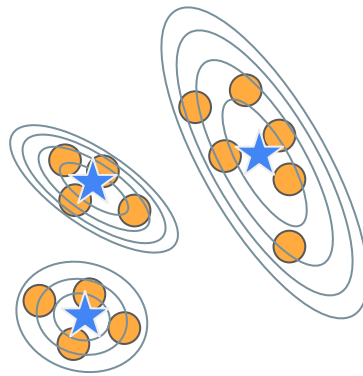
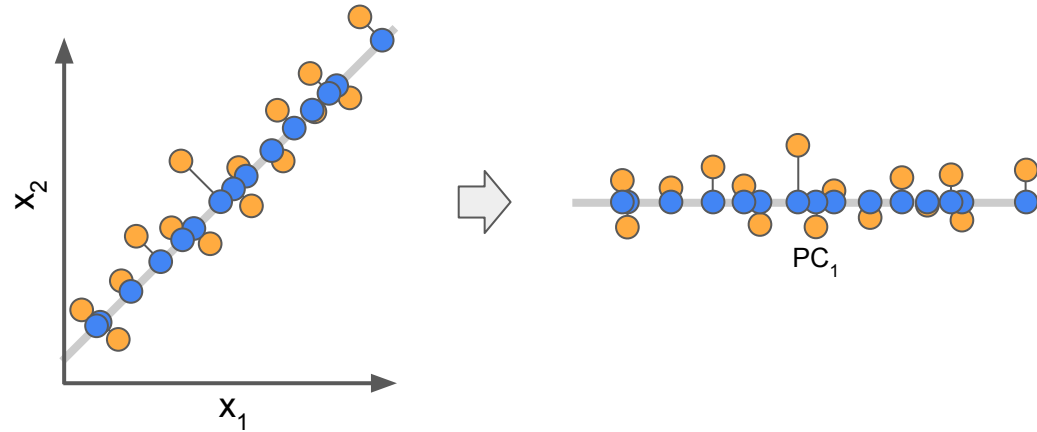


Why dimensionality reduction?

- Represent data with fewer dimensions
- Remove irrelevant data features
- Merge correlated/coordinated measures
- Reduce noise
- Avoid over-fitting (the Curse of Dimensionality)
- Cluster
- Find trajectory of disease, continuous state



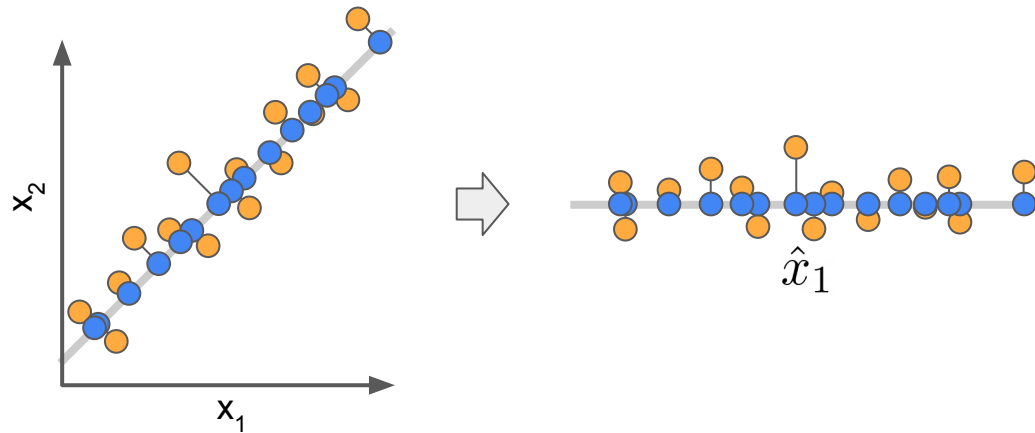
Why dimensionality reduction?



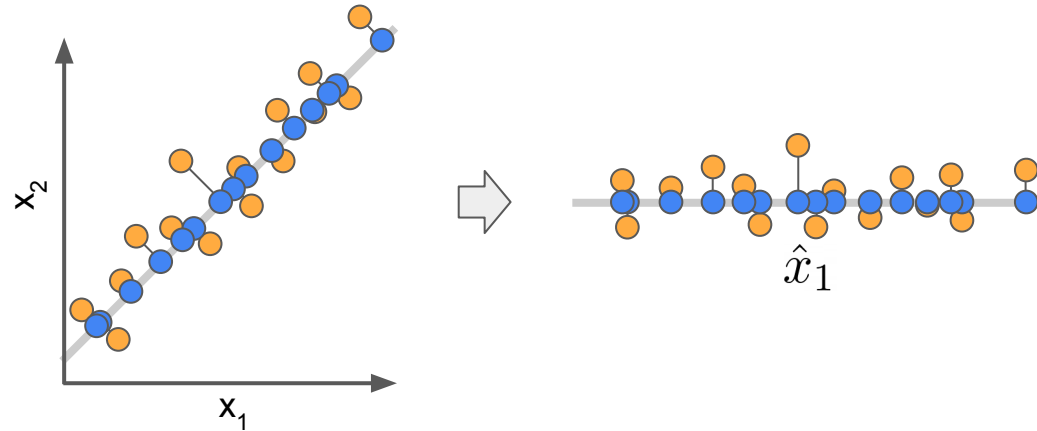
Linear projections

- Use a linear function to transform data to new space

$$\hat{x}_1 = w_1 x_1 + w_2 x_2$$



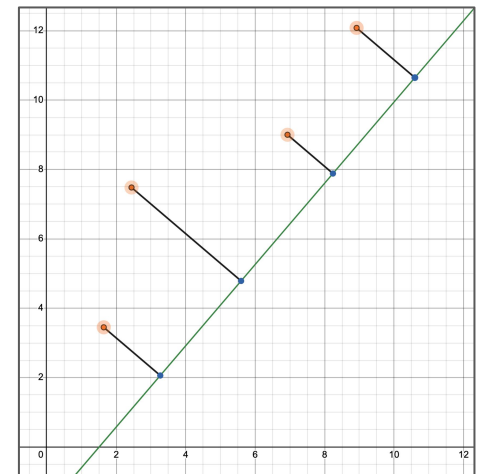
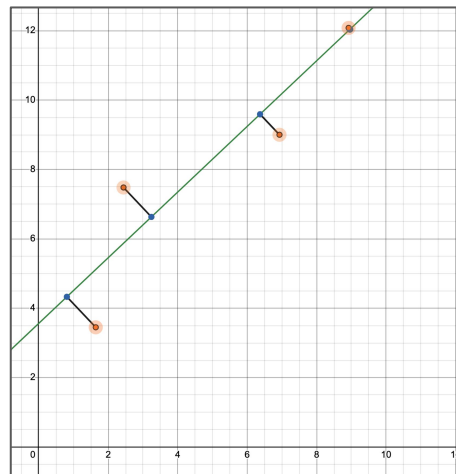
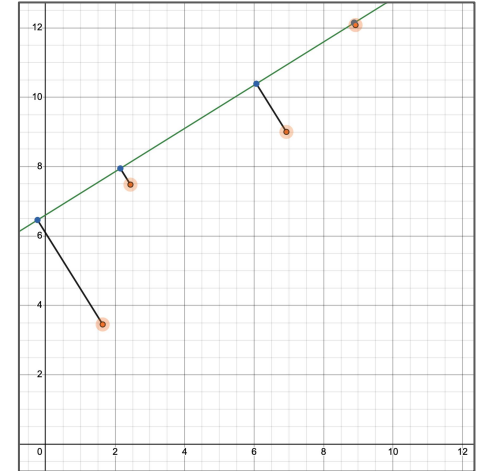
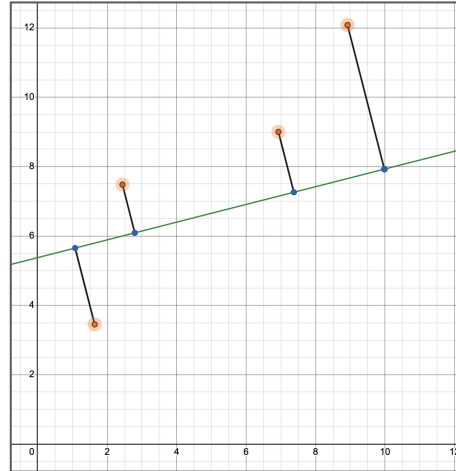
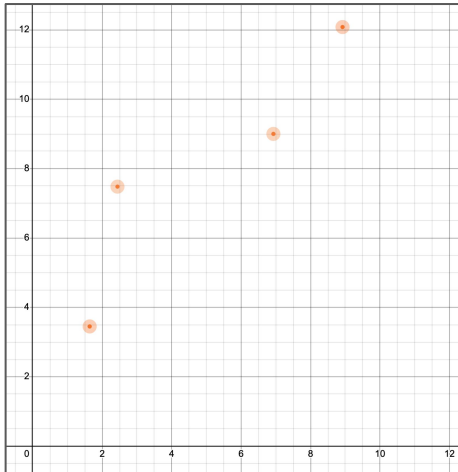
Which projection is “optimal”?



- Low-dimensional representation often introduces errors
- Aim for the most “accurate” new representation

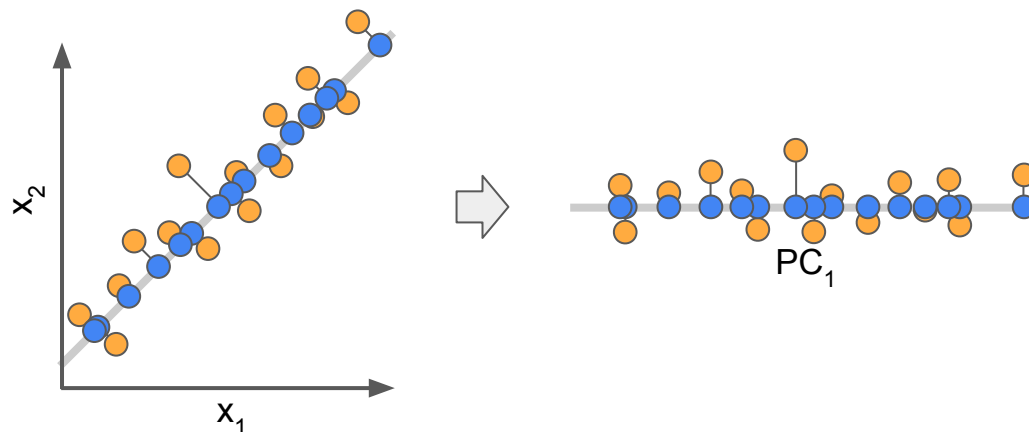
$$\arg \min \sum_i \|x_i - \hat{x}_i\|_2^2$$

Find the axis (PC) that minimizes the error



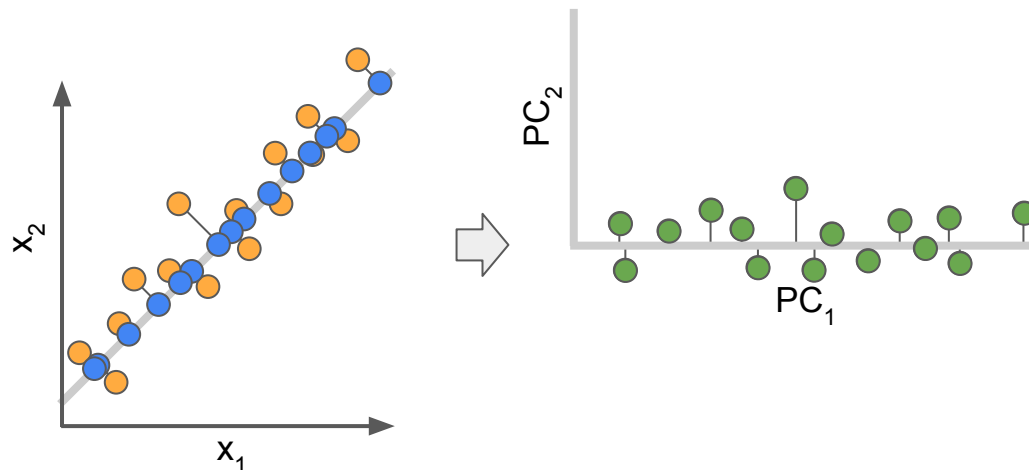
PCA - principal component analysis

- PC_1 is the direction that minimizes the errors on a 1-dimensional projection
- $\Rightarrow PC_1$ maximizes the variance
- $\Rightarrow$ More information is conserved
- $\Rightarrow$ **The “optimal” 1D projection**



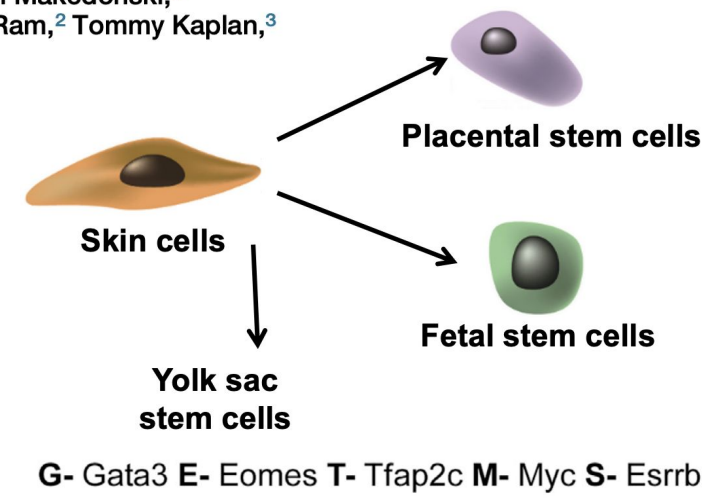
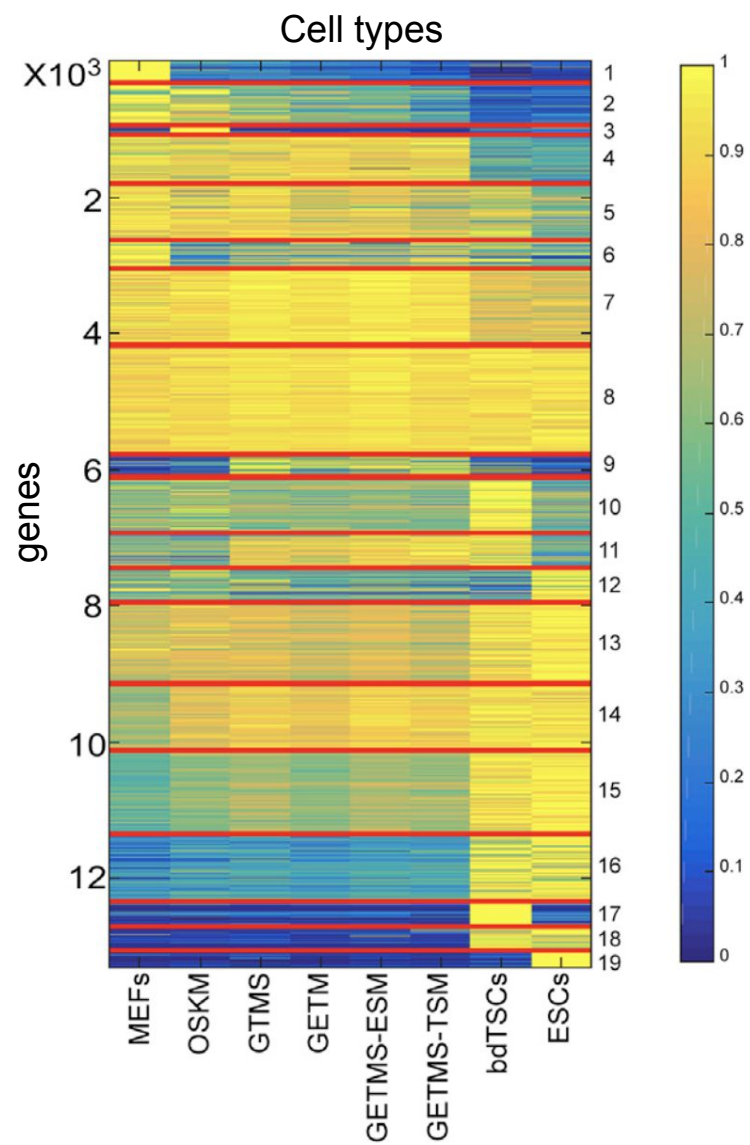
PCA - principal component analysis

- PC_1 is the direction that minimizes the errors on a 1-dimensional projection
- PC_2 minimizes the remaining errors



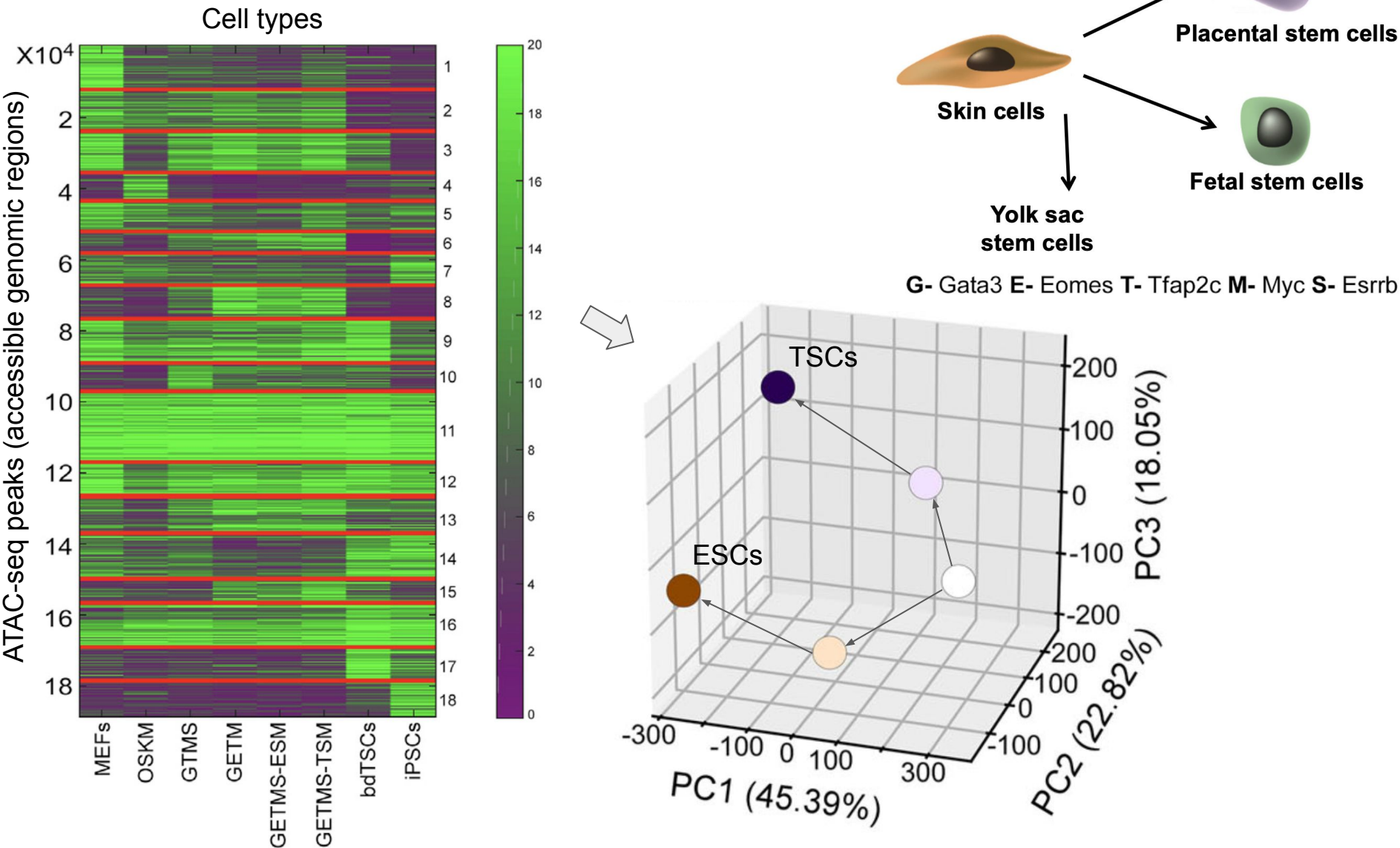
Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts

Hana Benchetrit,^{1,4} Mohammad Jaber,^{1,4} Valery Zayat,^{1,4} Shulamit Sebban,¹ Avital Pushett,¹ Kirill Makedonski,¹ Zvi Zakheim,¹ Ahmed Radwan,¹ Noam Maoz,¹ Rachel Lasry,¹ Noa Renous,¹ Michal Inbar,¹ Oren Ram,² Tommy Kaplan,³ and Yosef Buganim^{1,5,*}



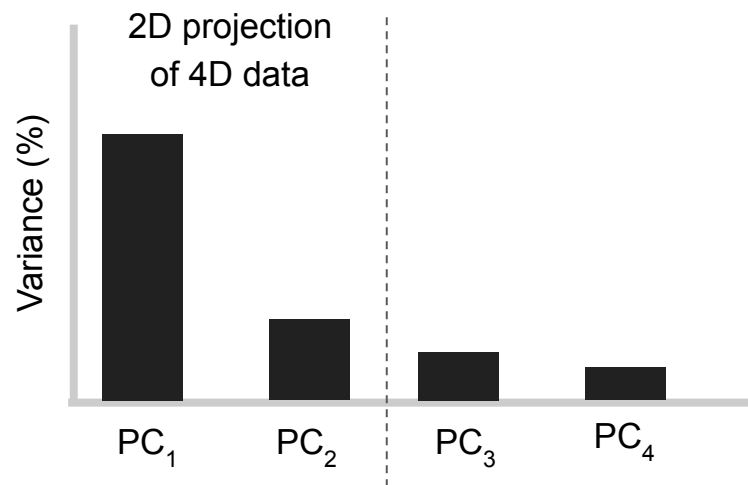
Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts

Hana Benchetrit,^{1,4} Mohammad Jaber,^{1,4} Valery Zayat,^{1,4} Shulamit Sebban,¹ Avital Pushett,¹ Kirill Makedonski,¹ Zvi Zakheim,¹ Ahmed Radwan,¹ Noam Maoz,¹ Rachel Lasry,¹ Noa Renous,¹ Michal Inbar,¹ Oren Ram,² Tommy Kaplan,³ and Yosef Buganim^{1,5,*}



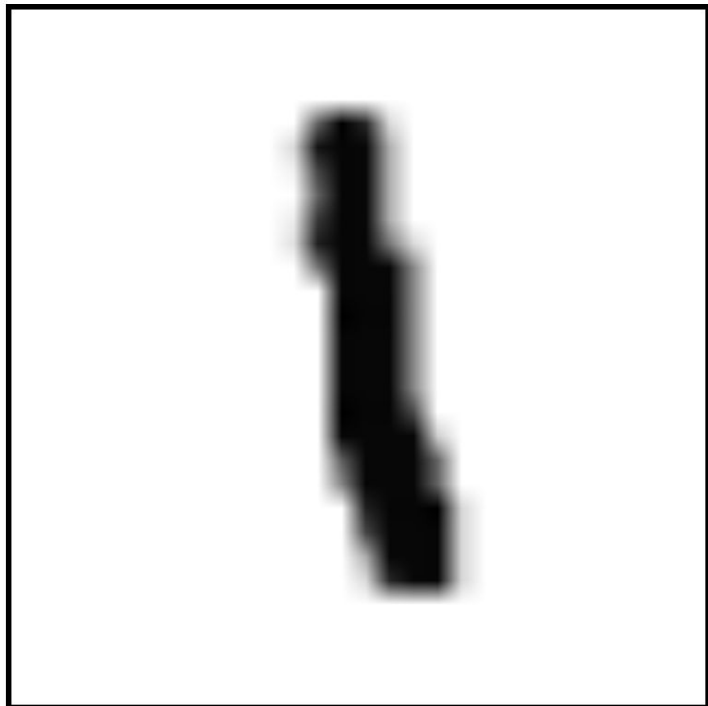
PCA - principal component analysis

- PC_1 is the direction that minimizes the errors on a 1-dimensional projection
- PC_2 minimizes the remaining errors
- PC_3 minimizes the remaining errors
- and so on...
- We then ignore components that are less important



PCA can be applied to different types of data

- Images are points in high-dimensional space
- A hand-drawn digit “1”, shown as a 14x14 matrix
- A point in a 196-dimensional space



12

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	.6	.8	0	0	0	0	0	0	0
0	0	0	0	0	0	.7	1	0	0	0	0	0	0	0
0	0	0	0	0	0	.7	1	0	0	0	0	0	0	0
0	0	0	0	0	0	.5	1	.4	0	0	0	0	0	0
0	0	0	0	0	0	0	1	.4	0	0	0	0	0	0
0	0	0	0	0	0	0	1	.4	0	0	0	0	0	0
0	0	0	0	0	0	0	1	.7	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	.9	1	.1	0	0	0	0	0
0	0	0	0	0	0	0	.3	1	.1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Every value actually limited to one of 256 binary values encoded using 8 bits

PCA can be applied to different types of data

- Faces, for images
- Each picture is a 256x256 pixel image



Input faces



Principal components

PCA can be applied to different types of data

PC 1

PC1 + PC2

PCs 1-3

PCs 1-4

PCs 1-5



PCs 1 through 13
(13D projection of 256x256 data)

Similar methods applied to text analysis

- Each document viewed as a weighted list of words

Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer^{*,†}, Hongyue Dai^{†,‡}, Marc J. van de Vijver^{*,†}, Yudong D. He[‡], Augustinus A. M. Hart^{*}, Mao Mao[‡], Hans L. Peterse^{*}, Karin van der Kooy^{*}, Matthew J. Marton[‡], Anke T. Witteveen^{*}, George J. Schreiber[‡], Ron M. Kerkhoven^{*}, Chris Roberts[‡], Peter S. Linsley[‡], René Bernards^{*} & Stephen H. Friend[‡]

** Divisions of Diagnostic Oncology, Radiotherapy and Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, 121 Plesmanlaan, 1066 CX Amsterdam, The Netherlands*

† Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034,

<https://www.nature.com/articles/415530a>

Word List

964 words in your word list

Frequency	Word
56	genes
54	tumours
46	patients
42	prognosis
32	expression
30	breast
30	er

☒ Remove Stop Words ☐ Remove Numbers ☒ Remove Special Characters

Similar methods applied to text analysis

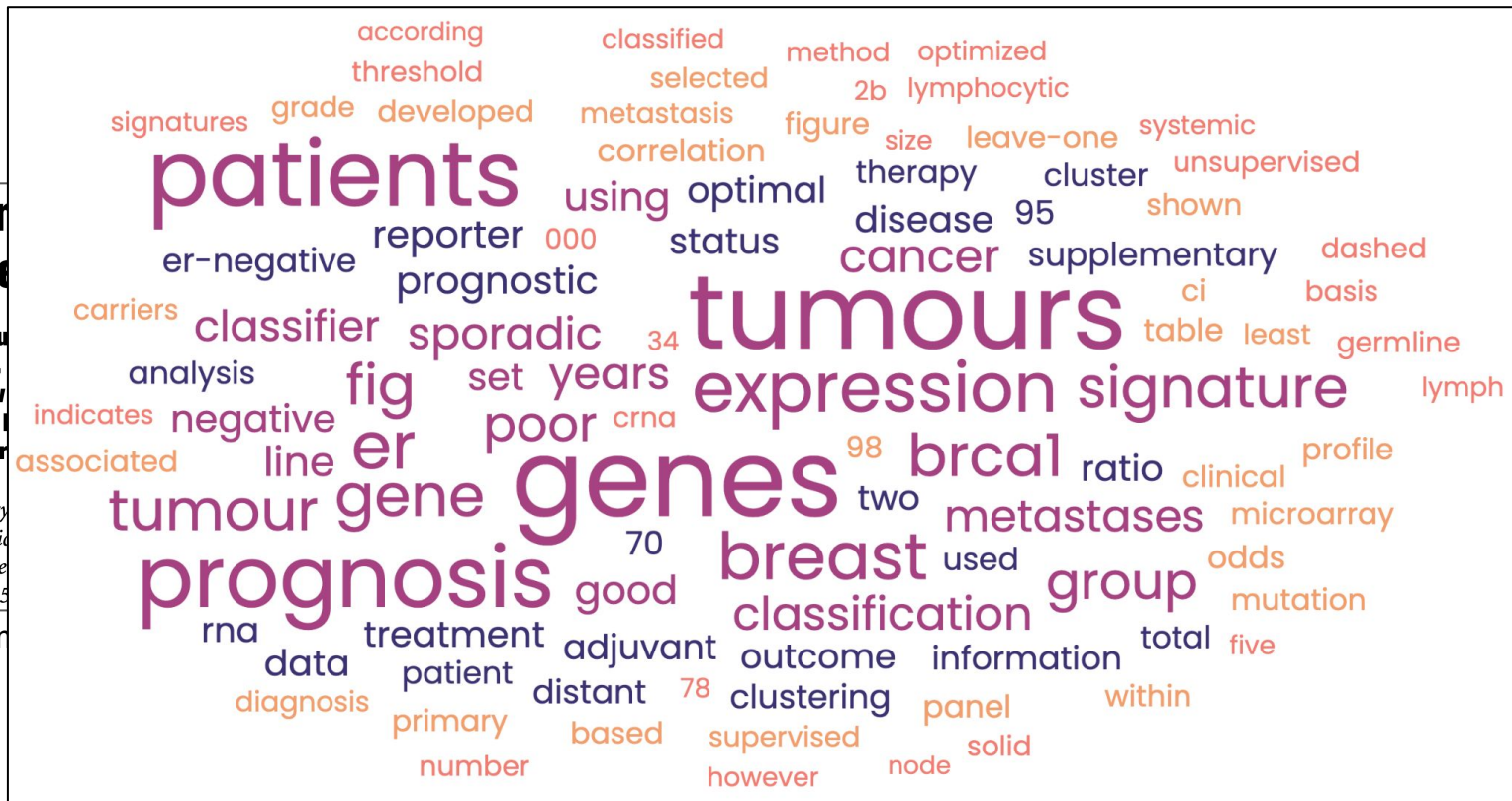
- Each document viewed as a weighted list of words

Gene expression clinical outcome

Laura J. van 't Veer^{*,†}, Hongyu
Yudong D. He[‡], Augustinus A.
Karin van der Kooy^{*}, Matthew
George J. Schreiber[‡], Ron M. I.
Peter S. Linsley[‡], René Bernar

^{*} Divisions of Diagnostic Oncology
and Center for Biomedical Genetic
121 Plesmanlaan, 1066 CX Amste
[‡] Rosetta Inpharmatics, 12040 115

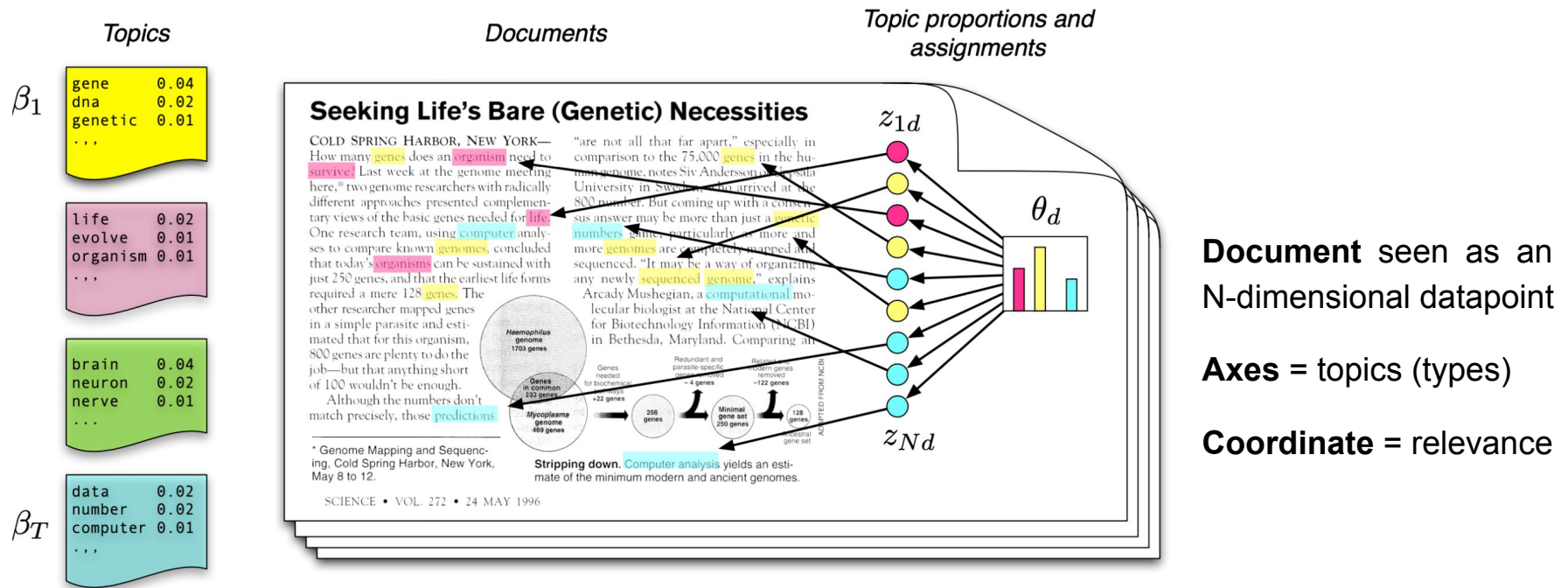
<https://www.nature.com>



Similar methods applied to text analysis

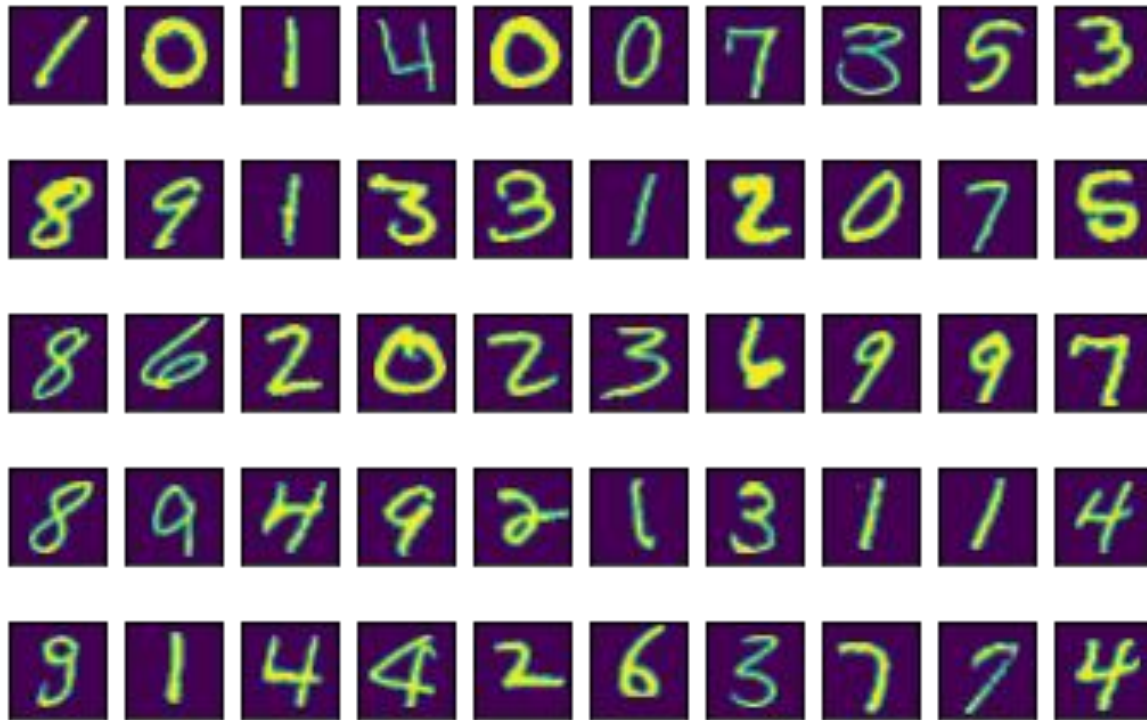
- Each document viewed as a weighted list of words
- Train on huge corpus, define typical “topics”

*a bit different than PCA, as topics are not orthogonal, but idea is similar



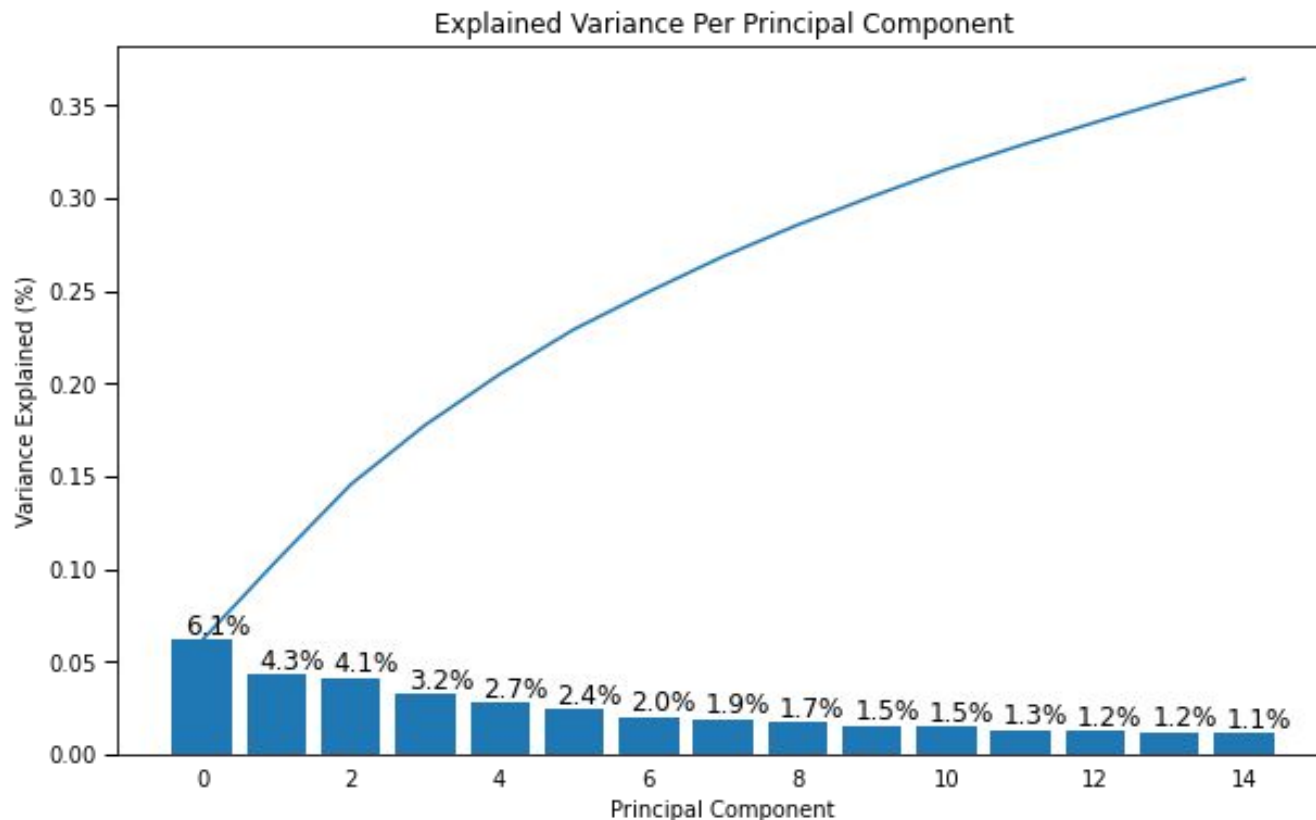
PCA can be applied to different types of data

- MNIST dataset of handwritten digits
- Each sample is a 28x28 pixel image

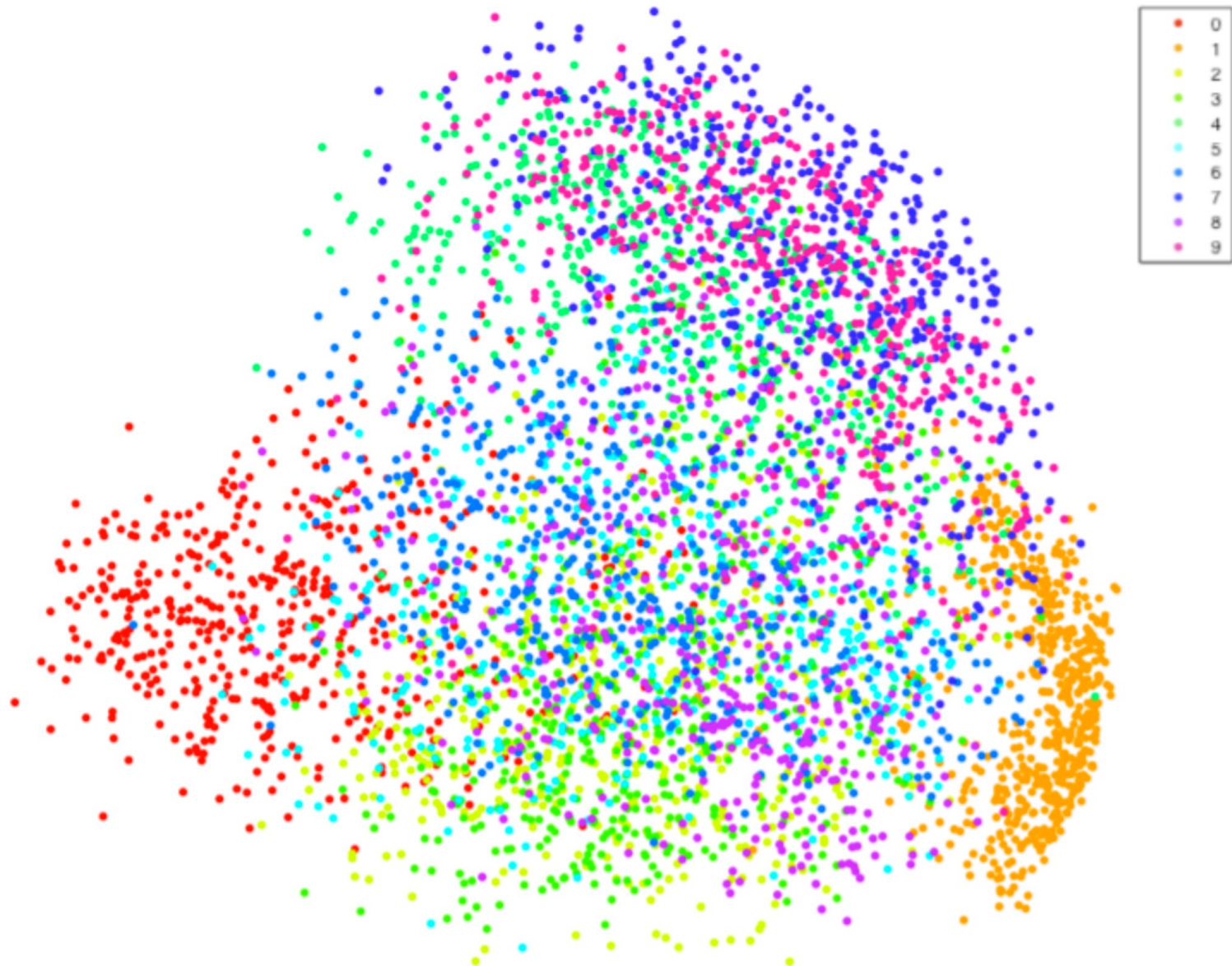


PCA applied to MNIST

- First projections do not capture much of variance

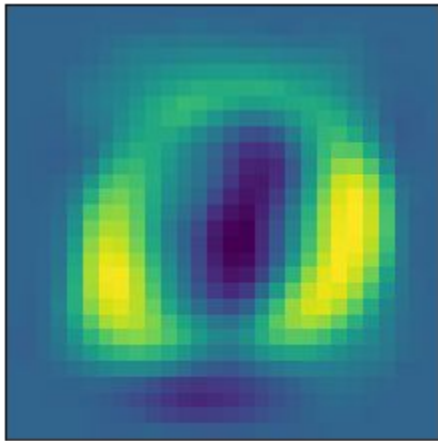


PCA applied to MNIST

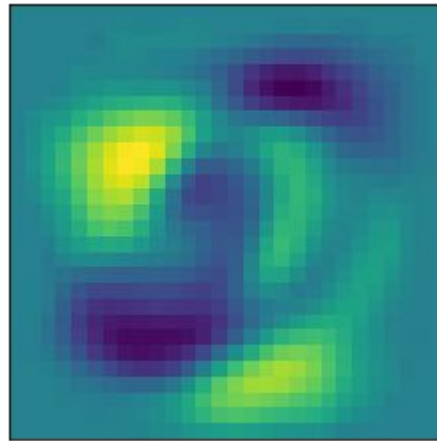


PCA applied to MNIST

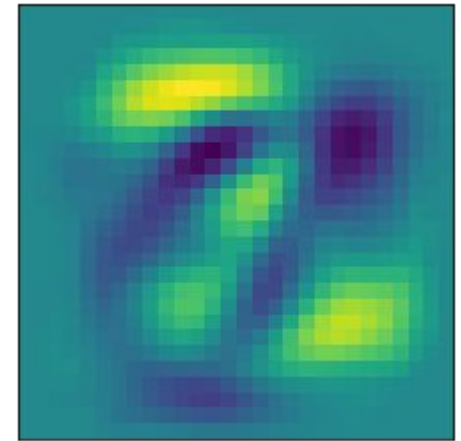
- Simple projections do not capture much of variance
- Nor do they make much sense



PC₁ (PVE=6.1%)



PC₂ (PVE=4.3%)

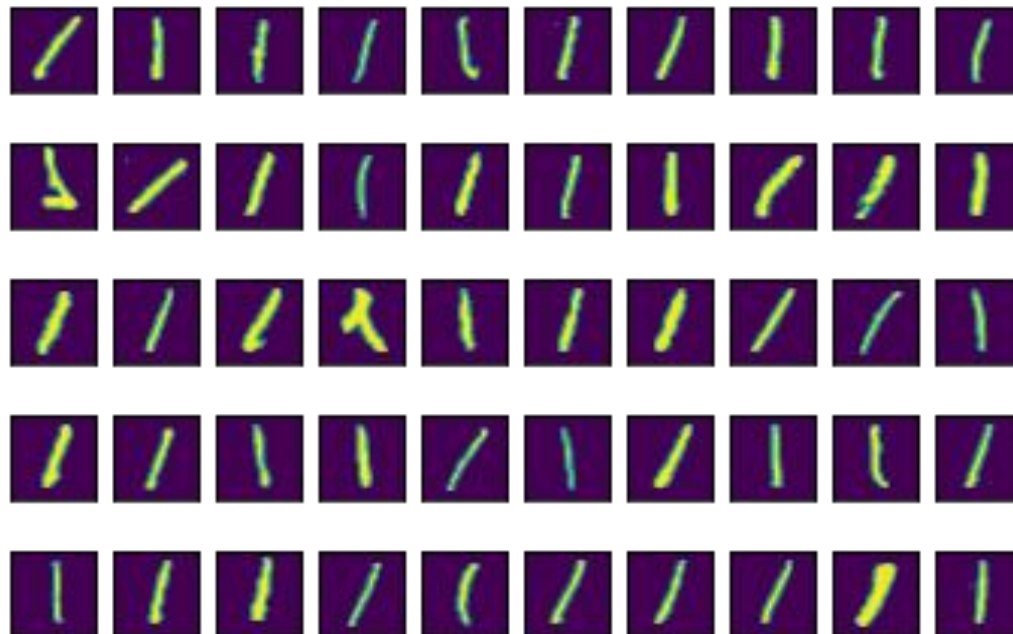


PC₃ (PVE=4.1%)

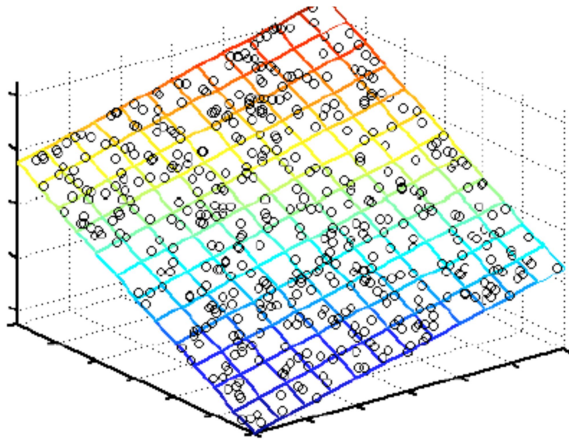
Always check percent of variance explained (PVE) by PCs

PCA can be applied to different types of data

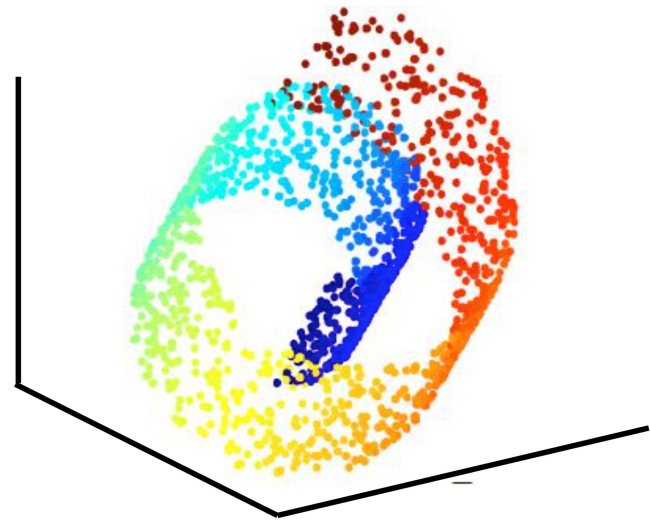
- Why has it failed?
- In this representation, digits are not similar enough (different regions/pixels are “on” - no archetypes)



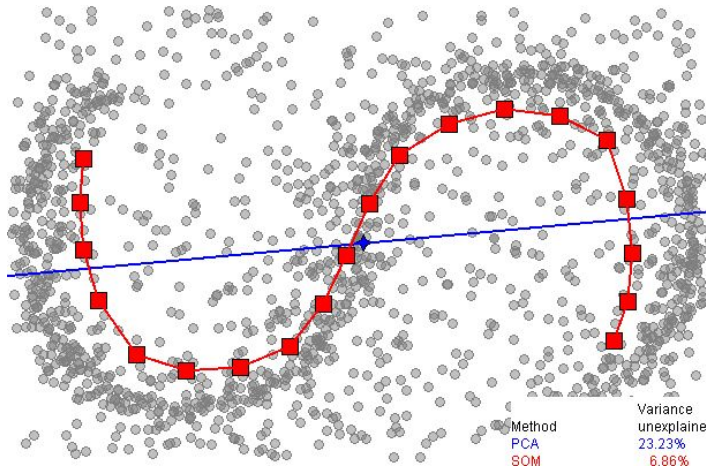
Beyond linear dimensionality reduction



Linear models are limited to hyperplanes
(or lower dimensional subspaces)



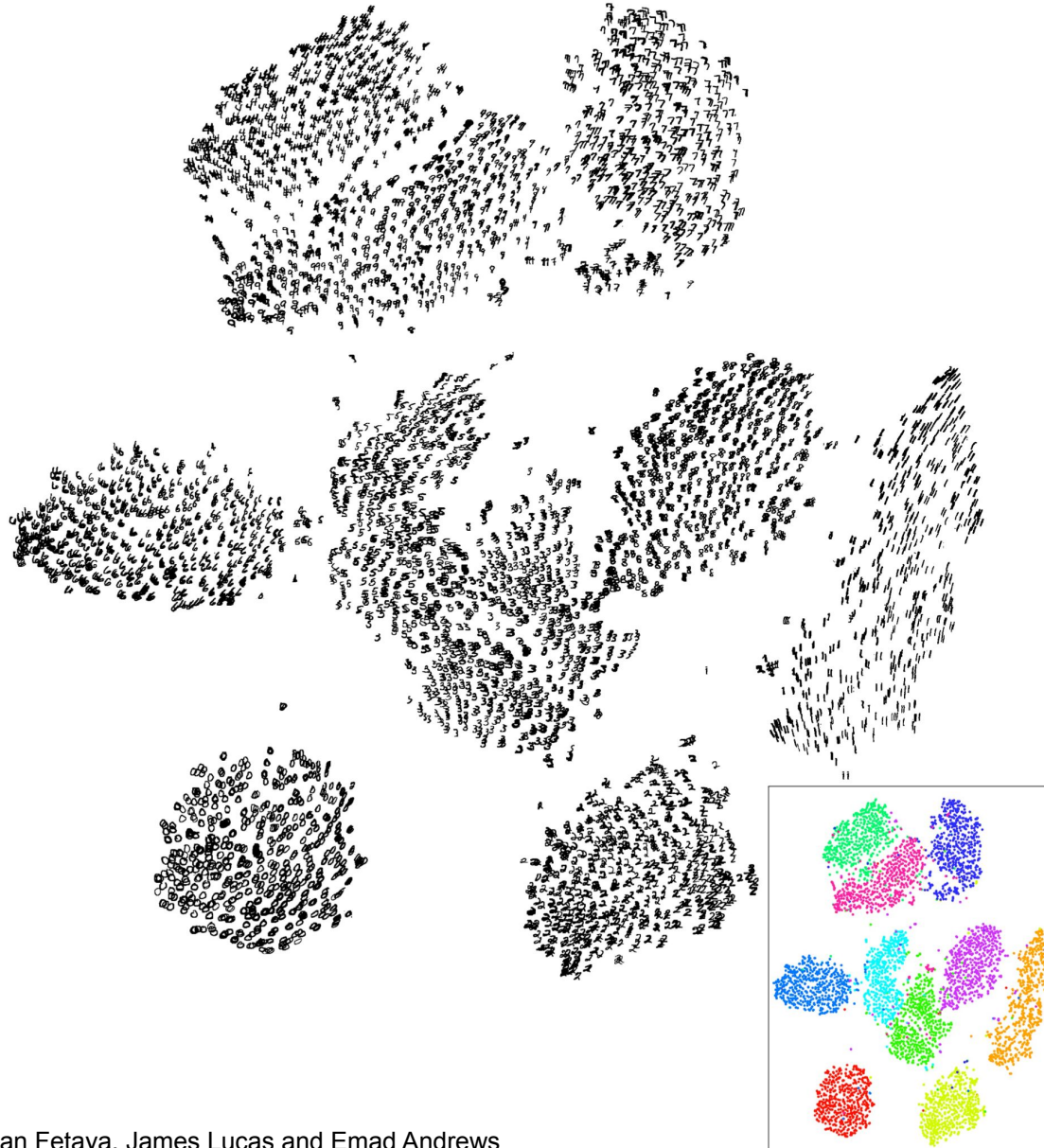
The swiss roll is a (locally linear) manifold



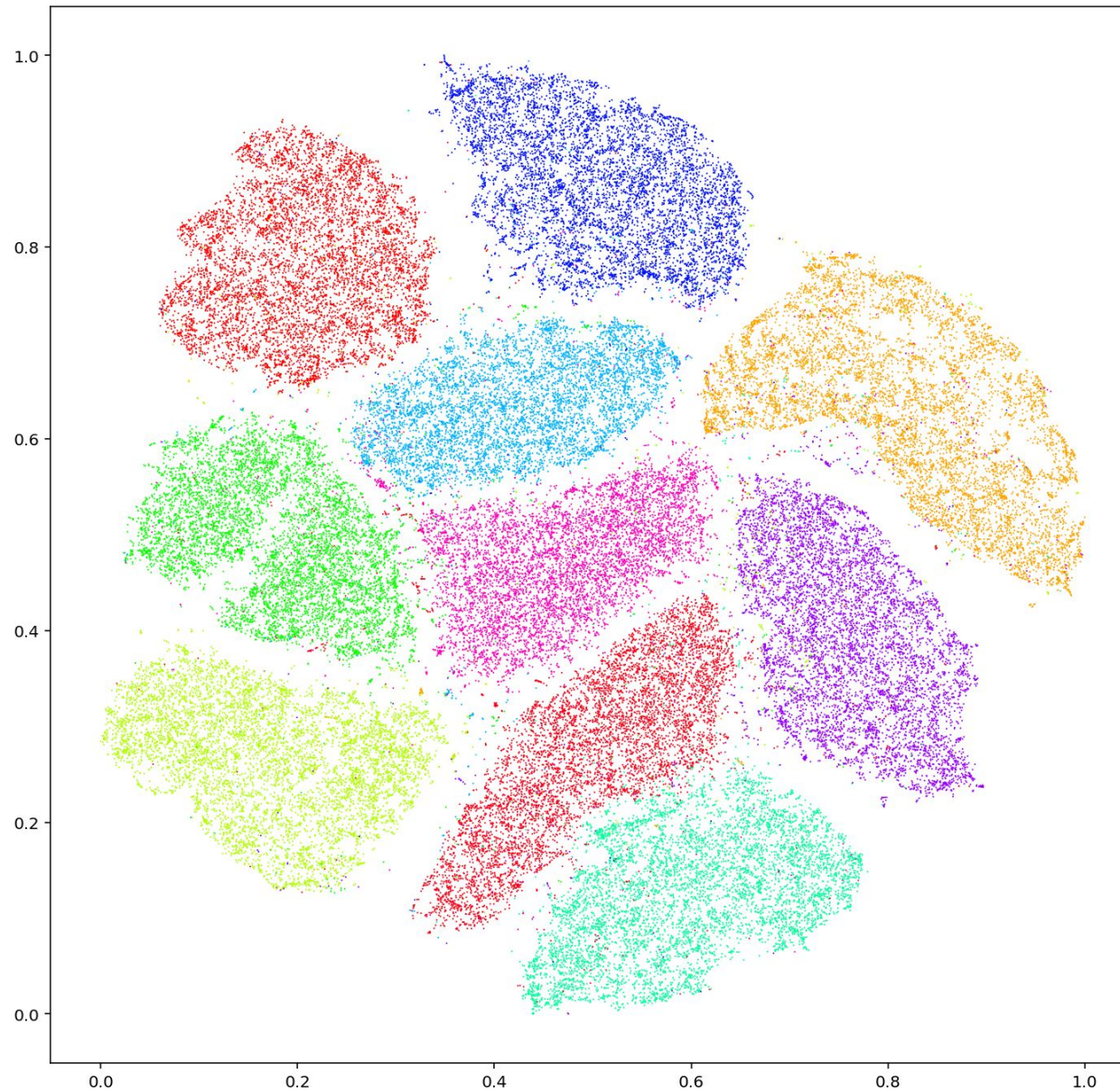
They will perform poorly on other structures

Let's use neighbors' adjacency to
unfold nonlinear structures

2D embedding of MNIST using tSNE



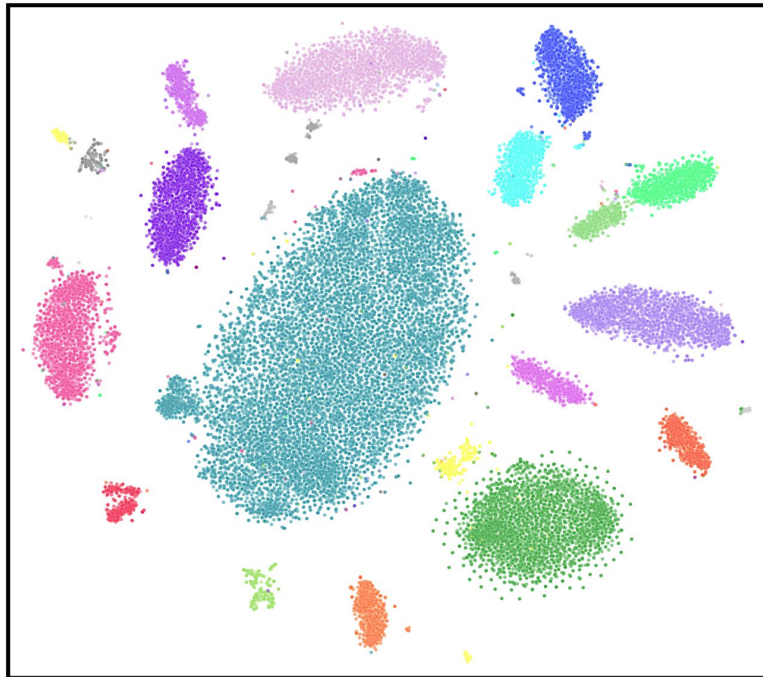
2D embedding of MNIST using tSNE



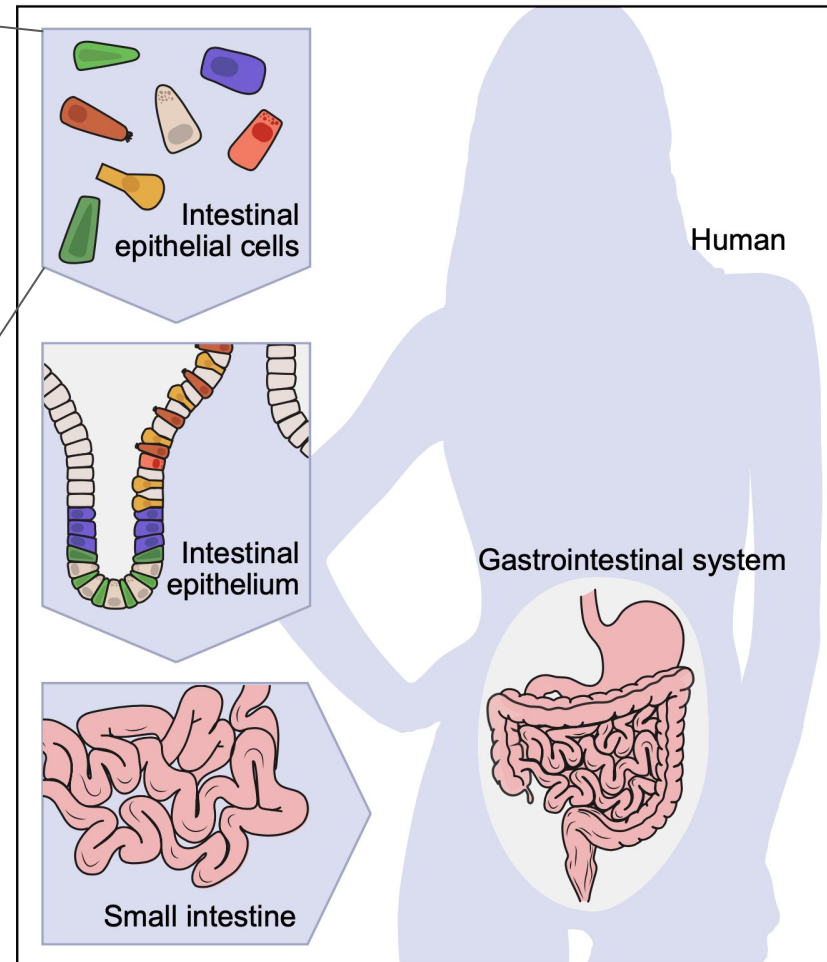
SCIENCE FORUM

The Human Cell Atlas

AVIV REGEV*, SARAH A TEICHMANN*, ERIC S LANDER*

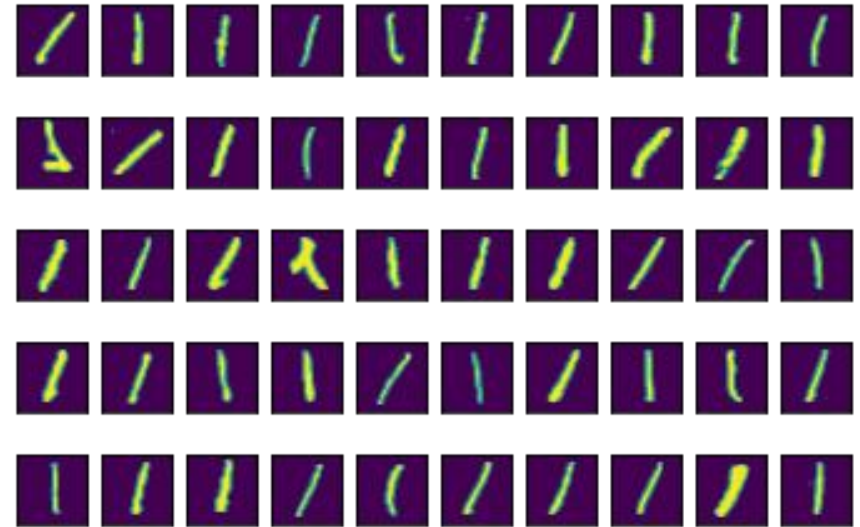
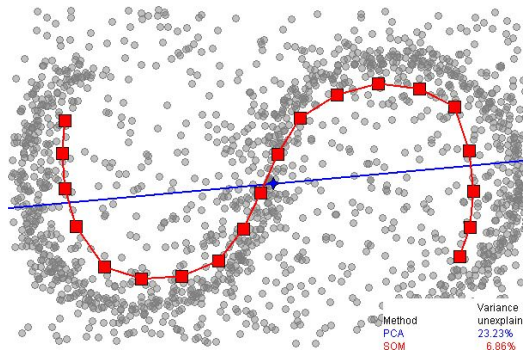


Clustered single-cell gene expression



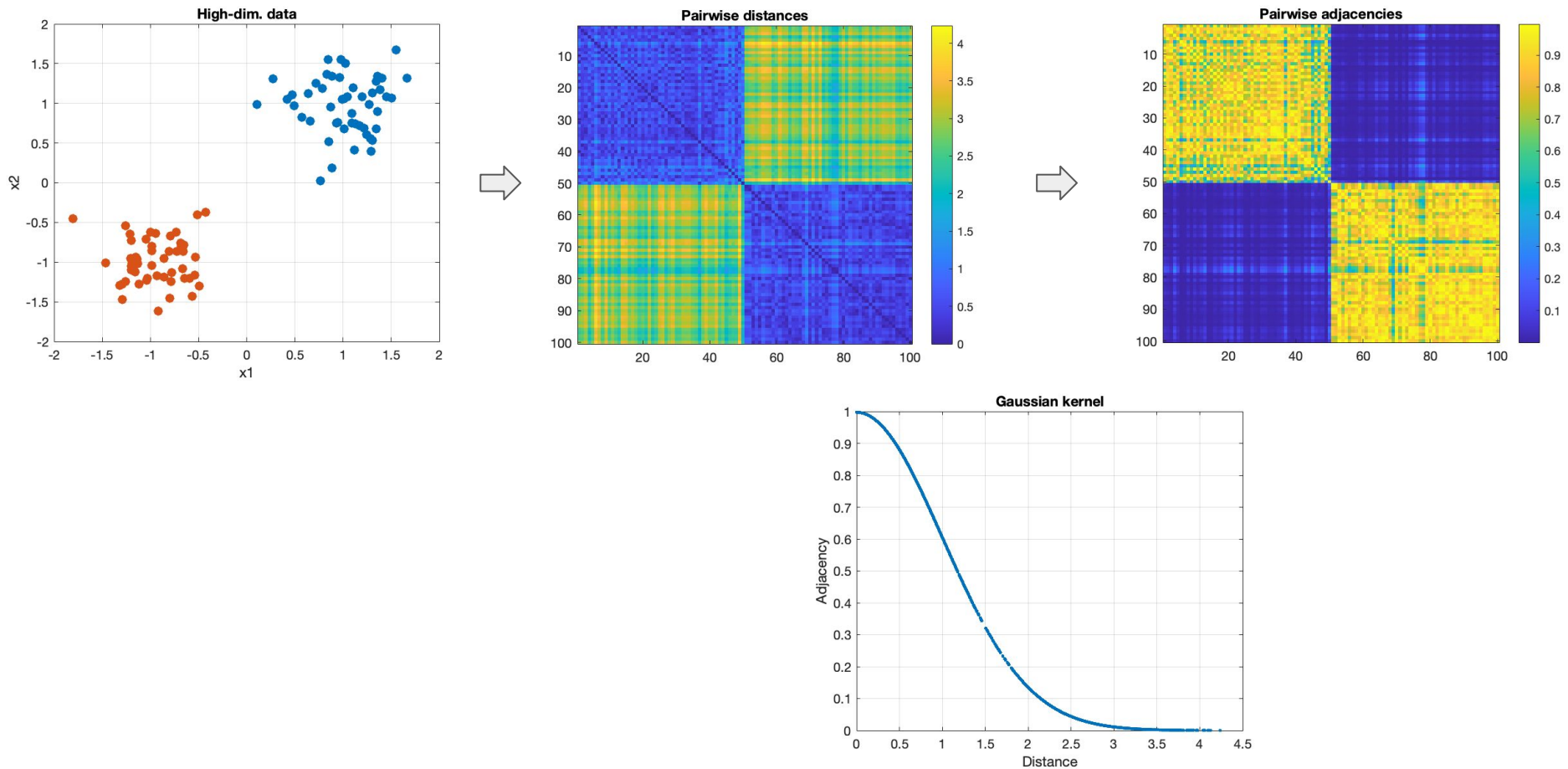
How does tSNE work?

Keep your neighbors close



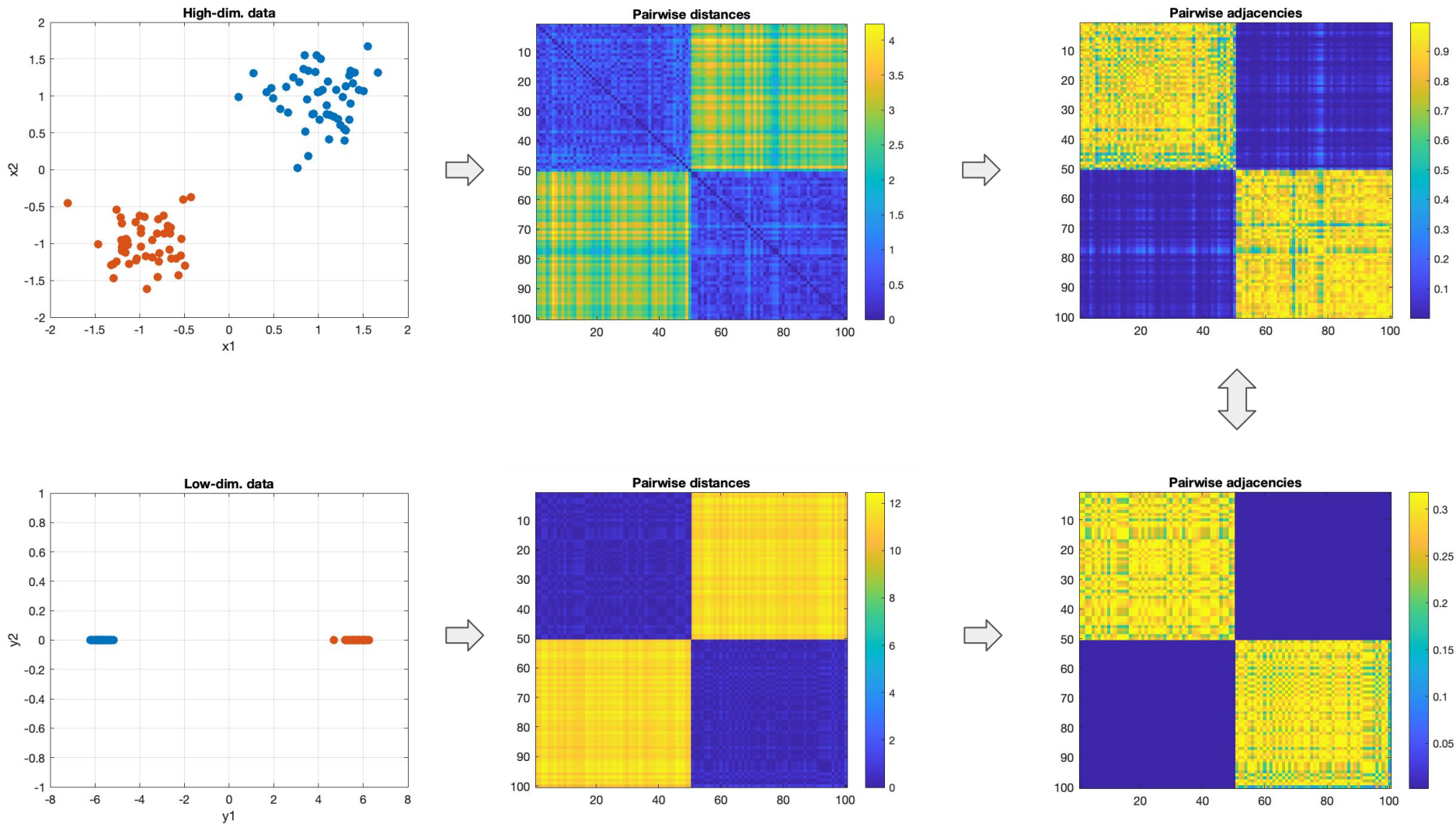
- tSNE **stochastically** projects data points to a lower dimension
- Iterates until converges to a local solution
- Close neighbors remain close
- Distant pairs could be distorted (less important)

How does tSNE work?

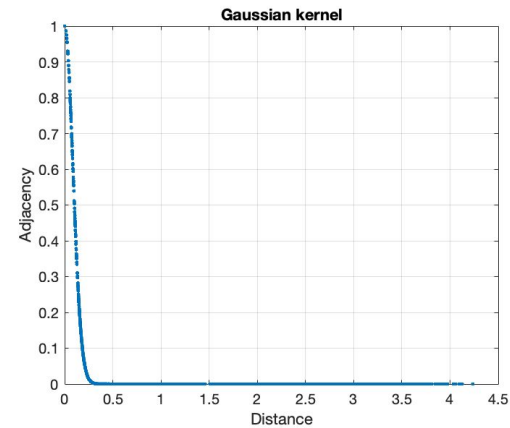
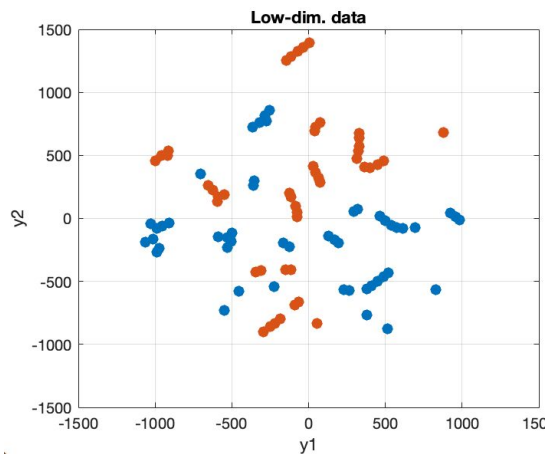
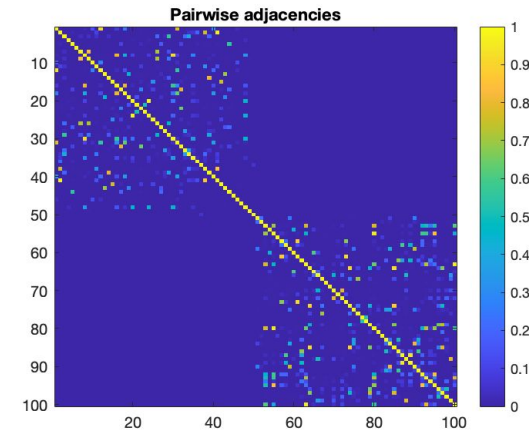
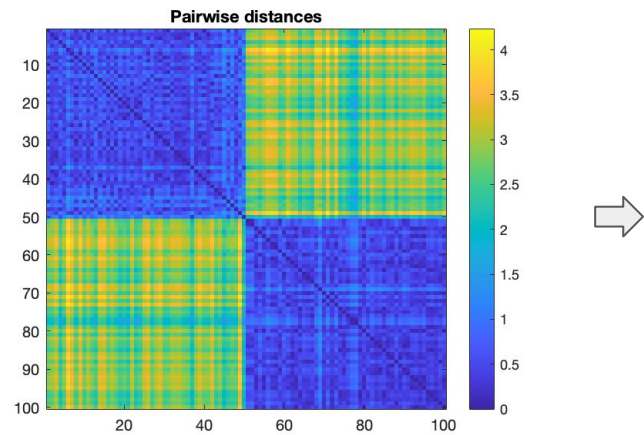
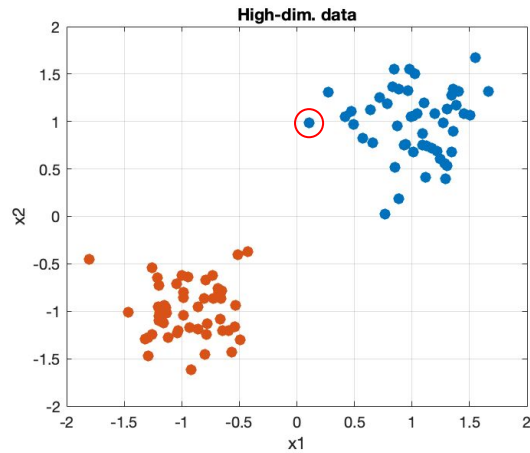


- Look for low-dimensional embedding with similar adjacencies

How does tSNE work?

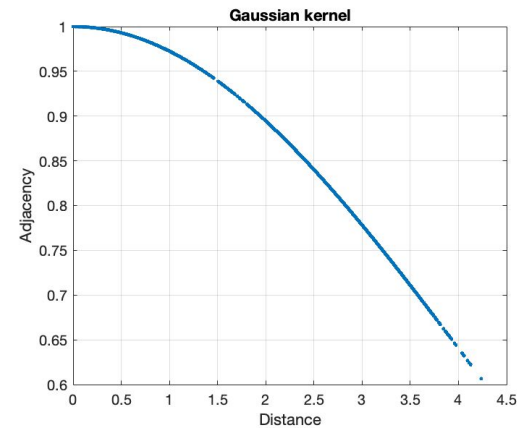
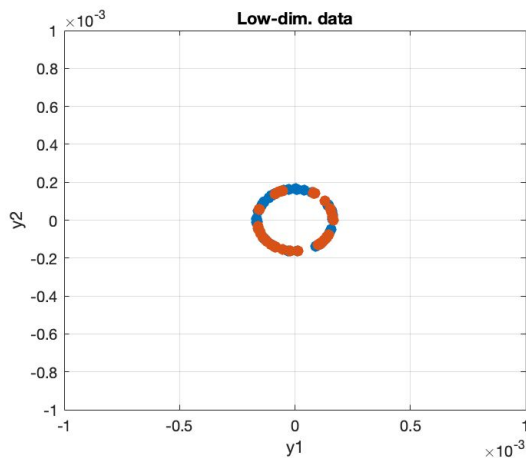
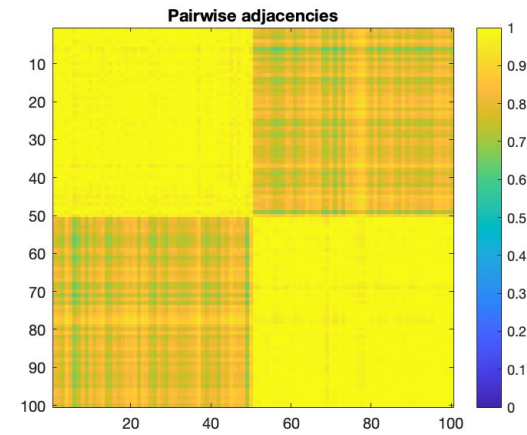
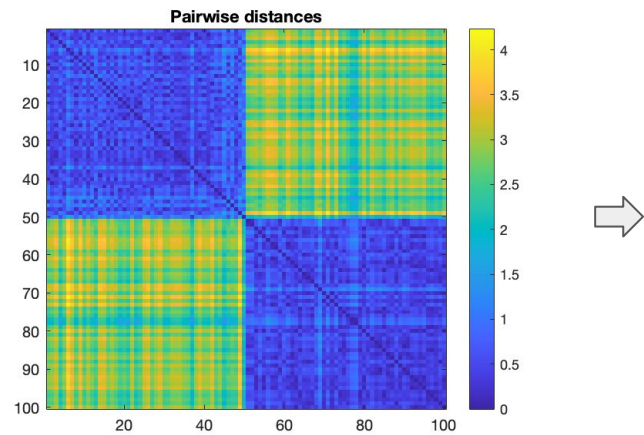
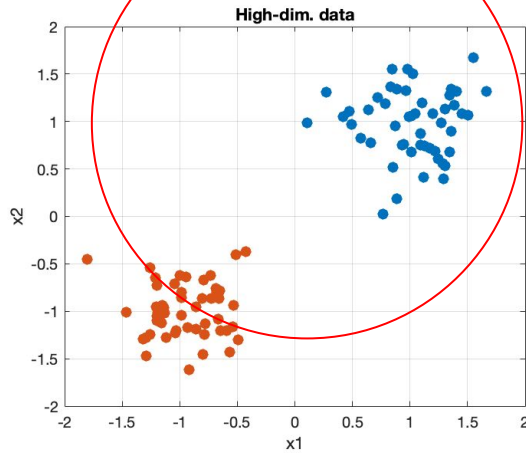


How does tSNE work?



Too few neighbors (tight neighborhood) and it'll fail

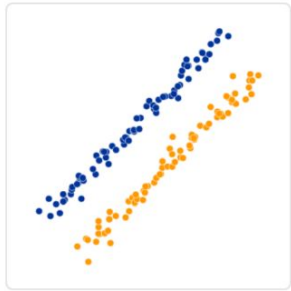
How does tSNE work?



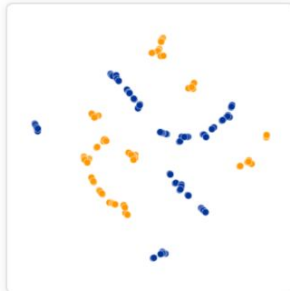
Too many neighbors (wide neighborhood) and it'll fail

How does tSNE work?

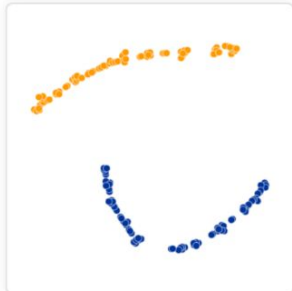
<https://distill.pub/2016/misread-tsne>



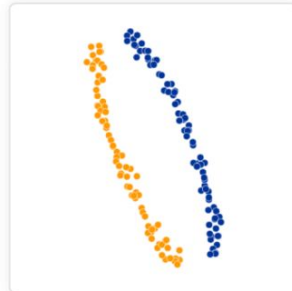
Original



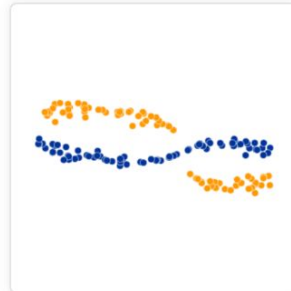
Perplexity: 2



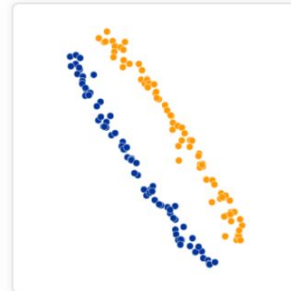
Perplexity: 5



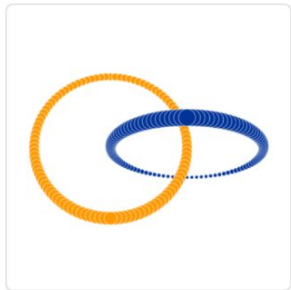
Perplexity: 30



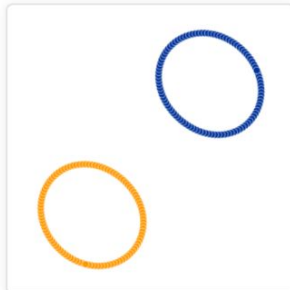
Perplexity: 50



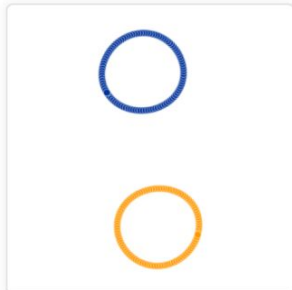
Perplexity: 100



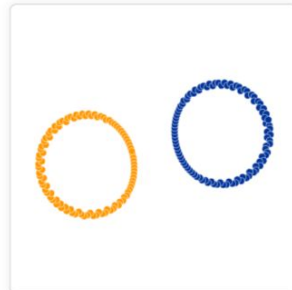
Original



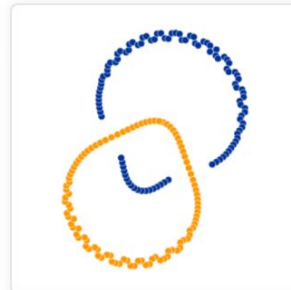
Perplexity: 2



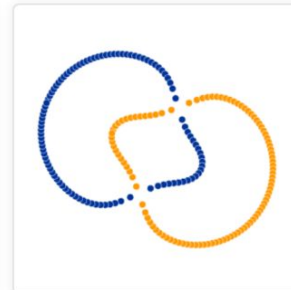
Perplexity: 5



Perplexity: 30



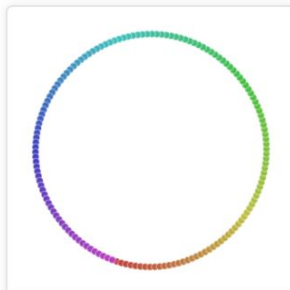
Perplexity: 50



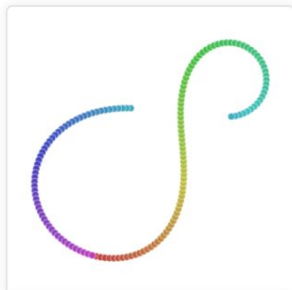
Perplexity: 100



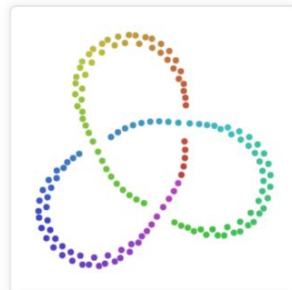
Original



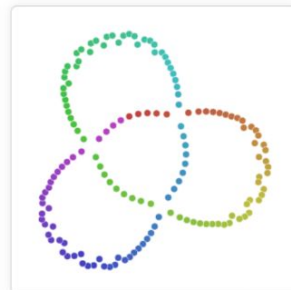
Perplexity: 2



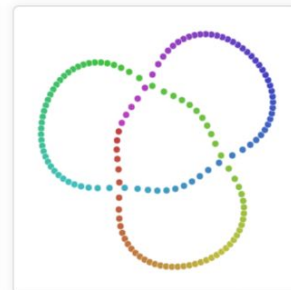
Perplexity: 5



Perplexity: 30



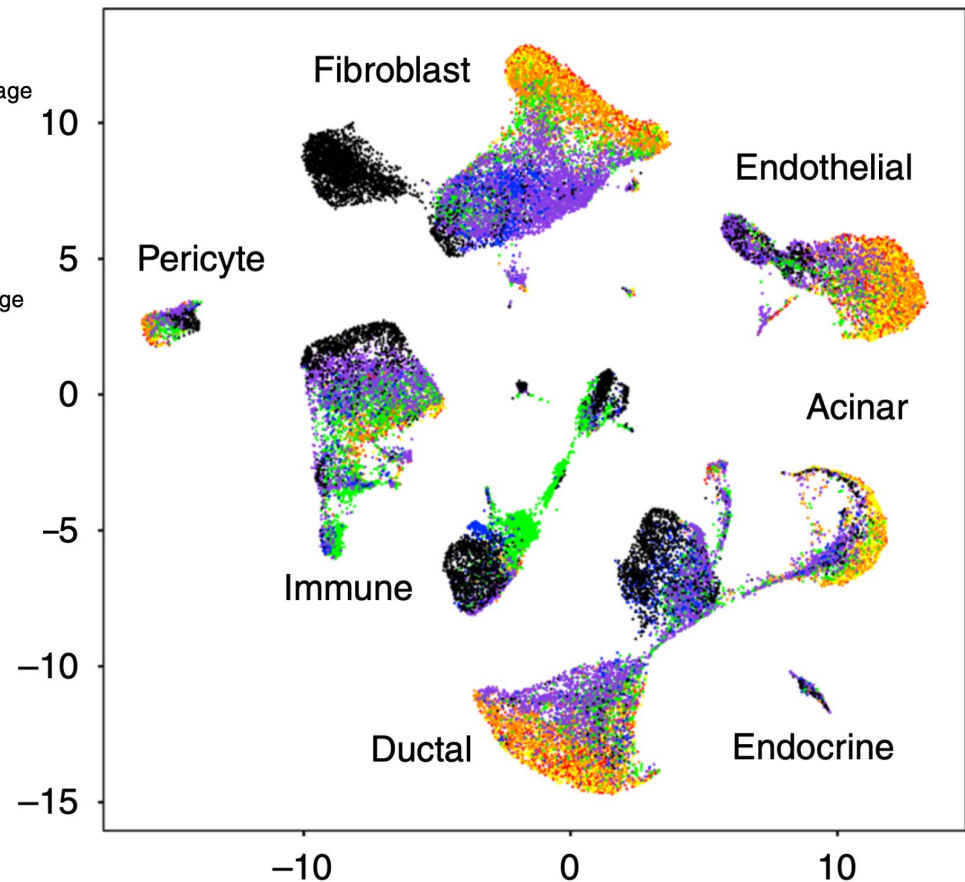
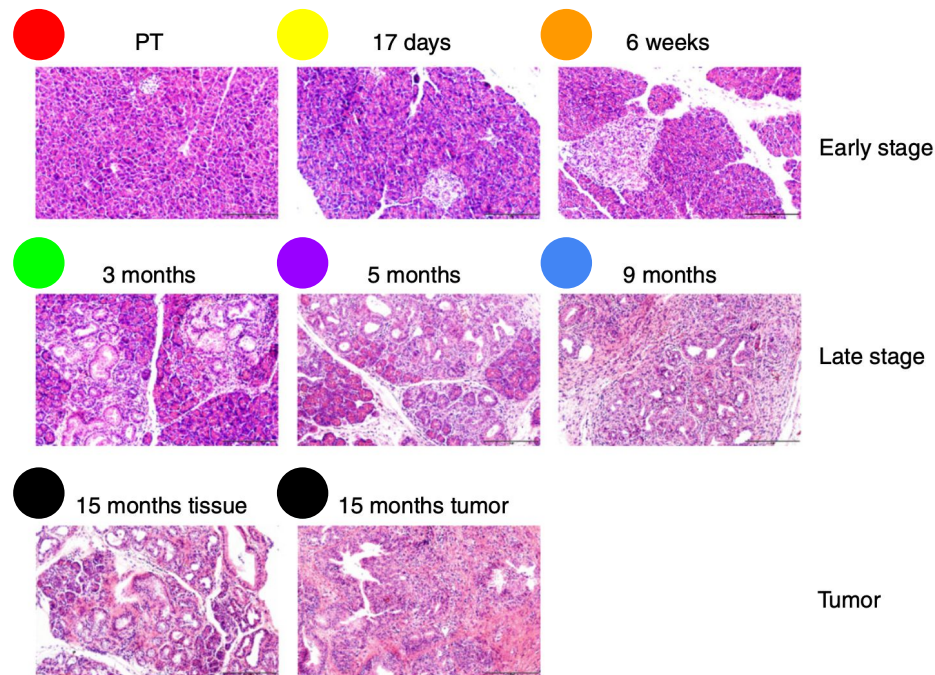
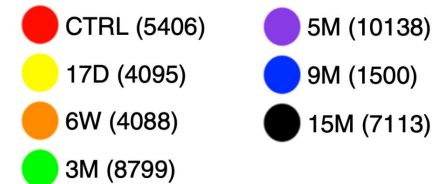
Perplexity: 50



Perplexity: 100

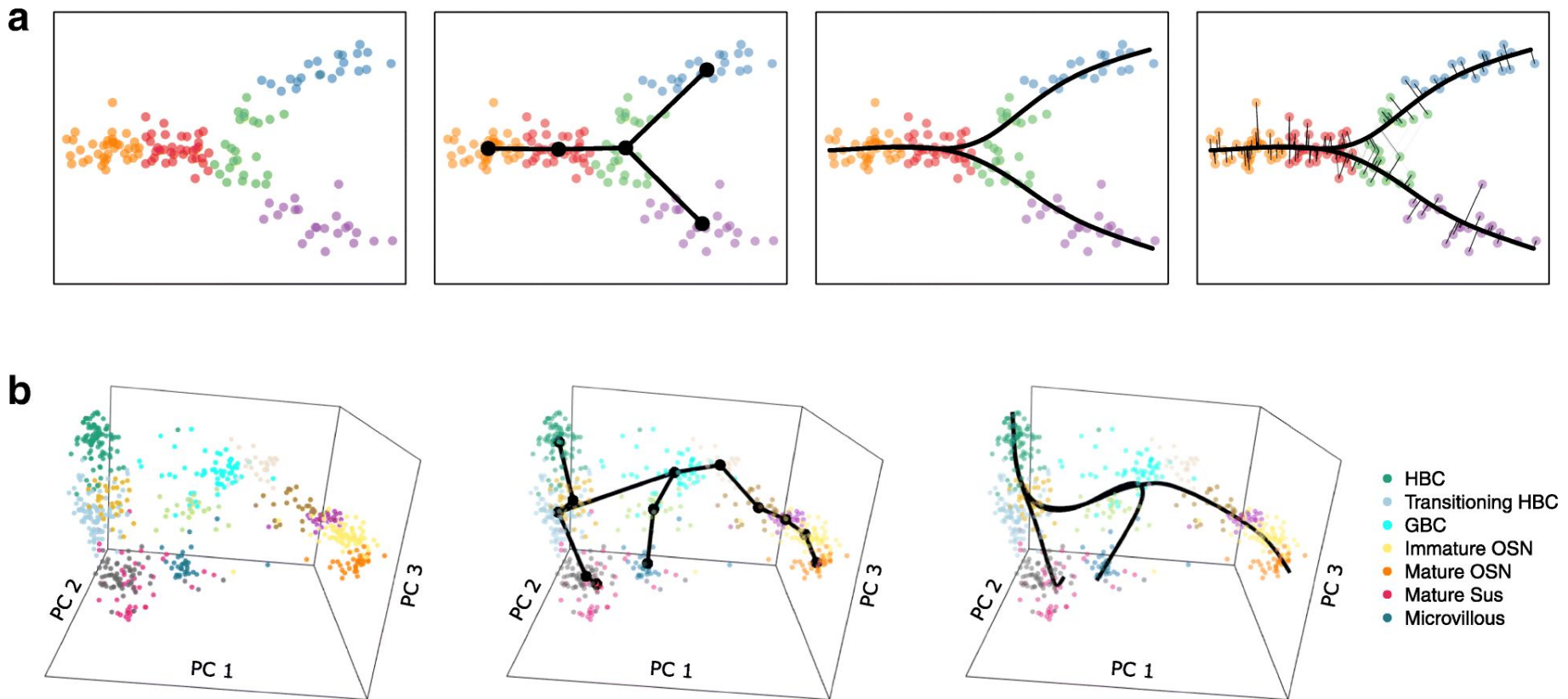
Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity

Yehuda Schlesinger^{1,7}, Oshri Yosefov-Levi^{1,7}, Dror Kolodkin-Gal^{2,3,7}, Roy Zvi Granit¹, Luriano Peters¹, Rachel Kalifa^{2,3}, Lei Xia¹, Abdelmajeed Nasereddin⁴, Idit Shiff⁴, Osher Amran¹, Yuval Nevo⁵, Sharona Elgavish⁵, Karine Atlan⁶, Gideon Zamir² & Oren Parnas¹



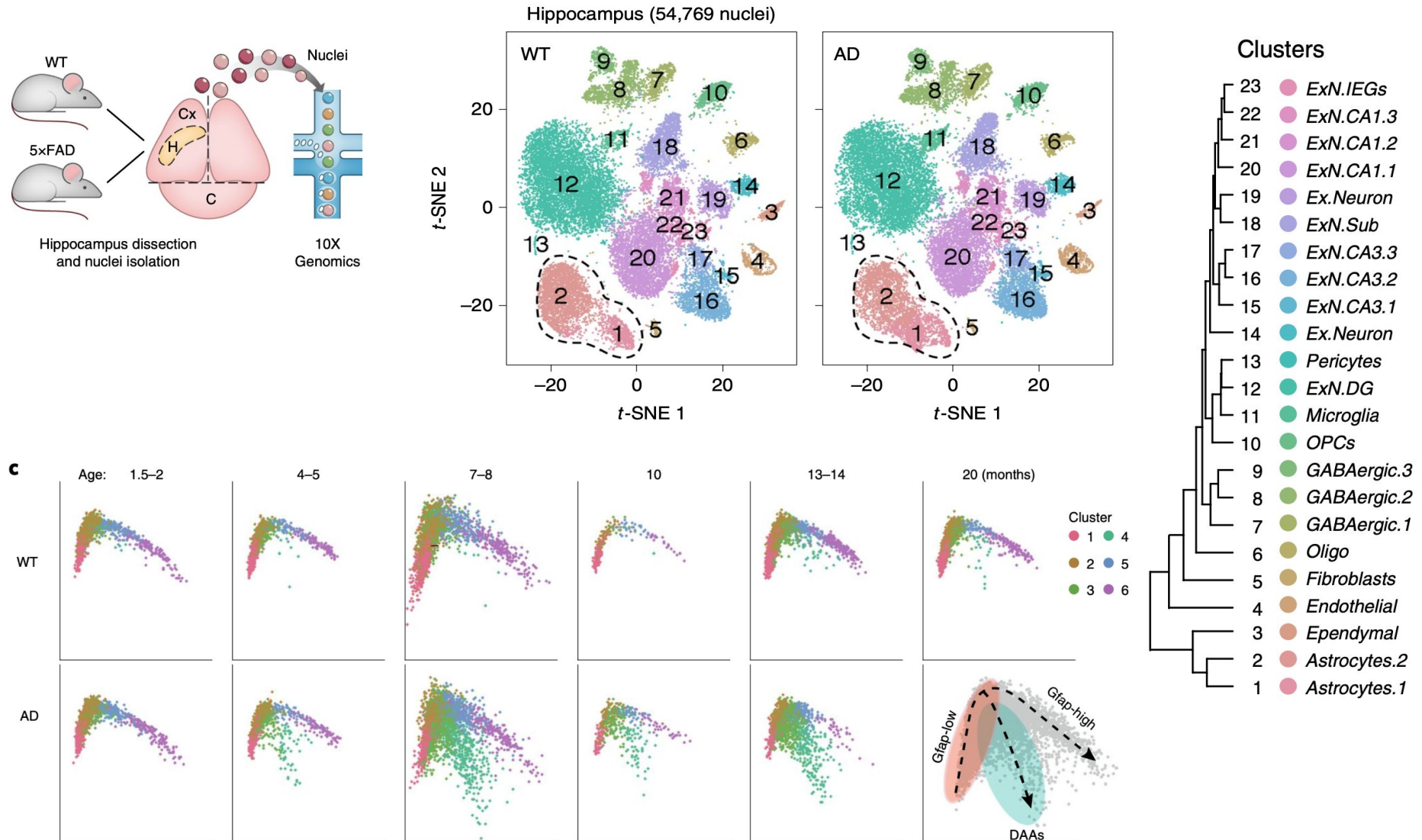
Trajectories and Pseudotime

- Many other alternatives were developed for analysis of single-cell RNA-seq data.
- Slingshot, Monocle and others track gradual changes in cells
- Most embed in a low-dim, cluster, and connect clusters



Disease-associated astrocytes in Alzheimer's disease and aging

Naomi Habib^{1,2}, Cristin McCabe^{2,8}, Sedi Medina^{3,8}, Miriam Varshavsky^{1,8}, Daniel Kitsberg¹, Raz Dvir-Szternfeld³, Gilad Green¹, Danielle Dionne², Lan Nguyen², Jamie L. Marshall², Fei Chen², Feng Zhang^{2,4,5}, Tommy Kaplan⁶, Aviv Regev^{2,7} and Michal Schwartz^{2,3}



תוכן

- למה להוריד מימד
 - להיפטר ממדידות לא רלוונטיות
 - לאחד מדידות קורלטיביות
 - להתגבר על רעש, אובר-פיטינג, curse of dimensionality
 - לאפשר אינטרפרטציה ותצוגה של הדוגמאות (קלאסטרिंग, רצף)
- הורדת מימד כרוכה בעיות - כמה מדוייק הייצוג החדש?
- מיזעור העיות הכולל - הורדת מימד לינארית PCA
- מיזעור העיות בין שכנים (קרובים) - הורדת מימד tSNE
- דוגמא כצירוף ארכיטיפים - זיהוי פנים, ניתוח מסמכים

What have we learned?

- Dimensionality reduction increases interpretability
- Projection onto important features
- PCA minimizes distortion, maximizes variance (information).
 - Each new coordinate is a linear function of old coords.
 - Each new axis is orthogonal to previous ones
 - Spans the maximal variance
- Non-linear embedding
 - tSNE converts pairwise distances to adjacencies
 - Stochastically seeks low-dim embedding that mimics original adjacencies. Converges to local optimum.
 - Other similar methods also used, allow inference of trajectories or pseudotime.