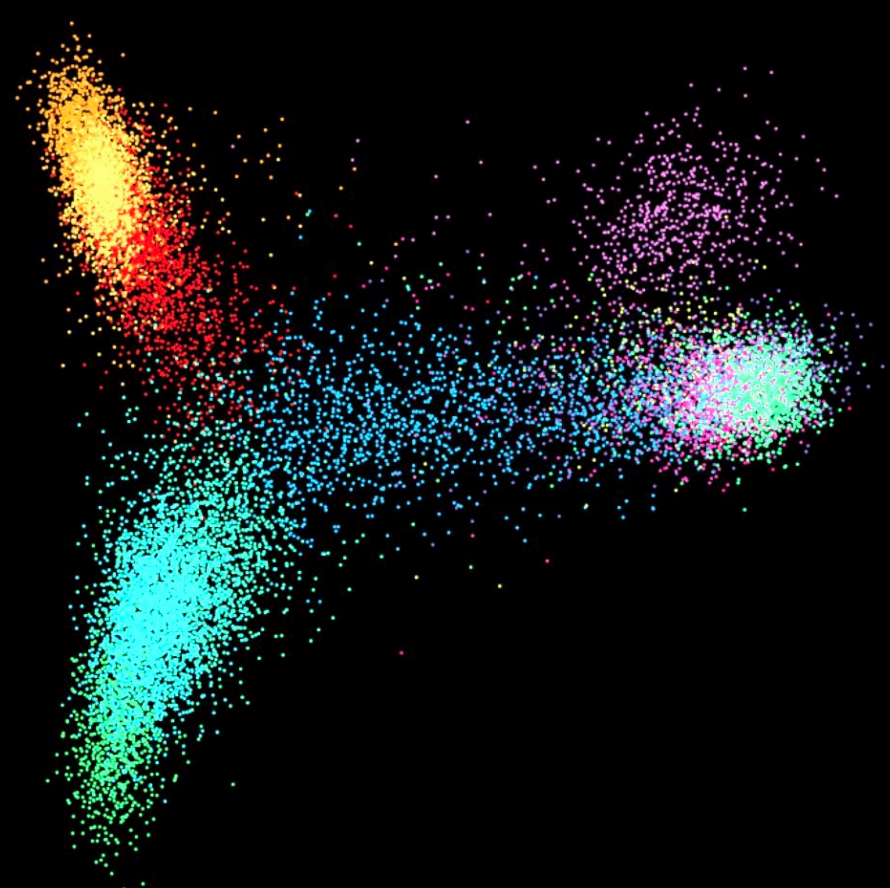
Day 3

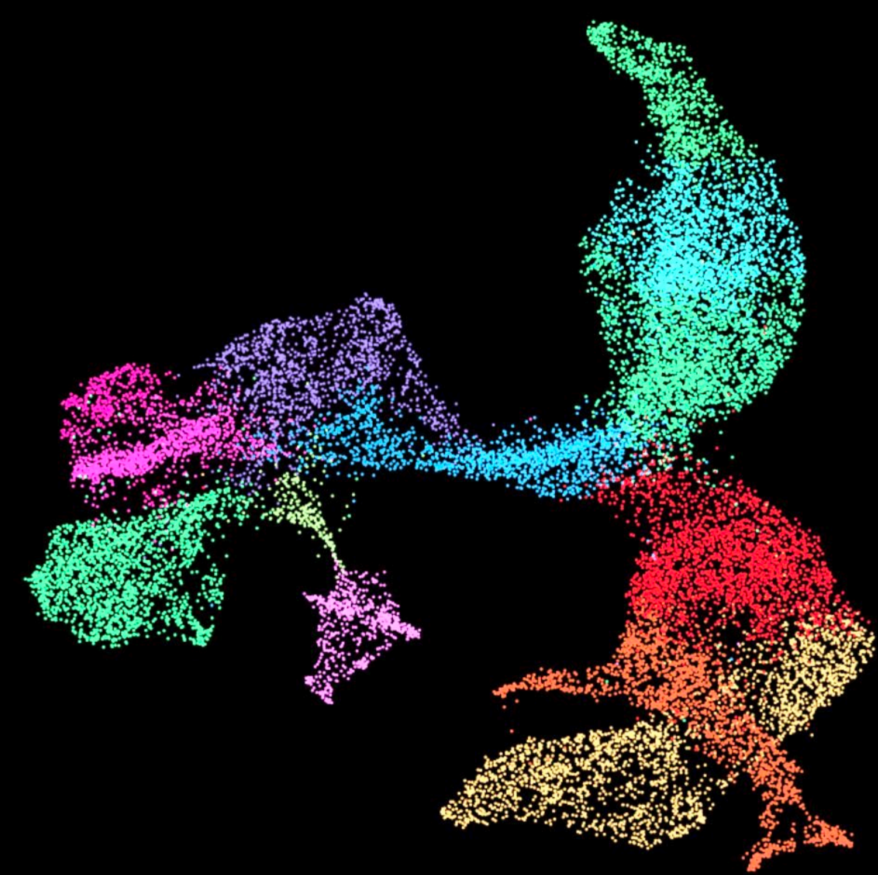
Linear representations
(PCA, ICA, …)

Easy comparison & integration across datasets



Nonlinear representations
(t-SNE, UMAP…)

Good representation of cell types / states

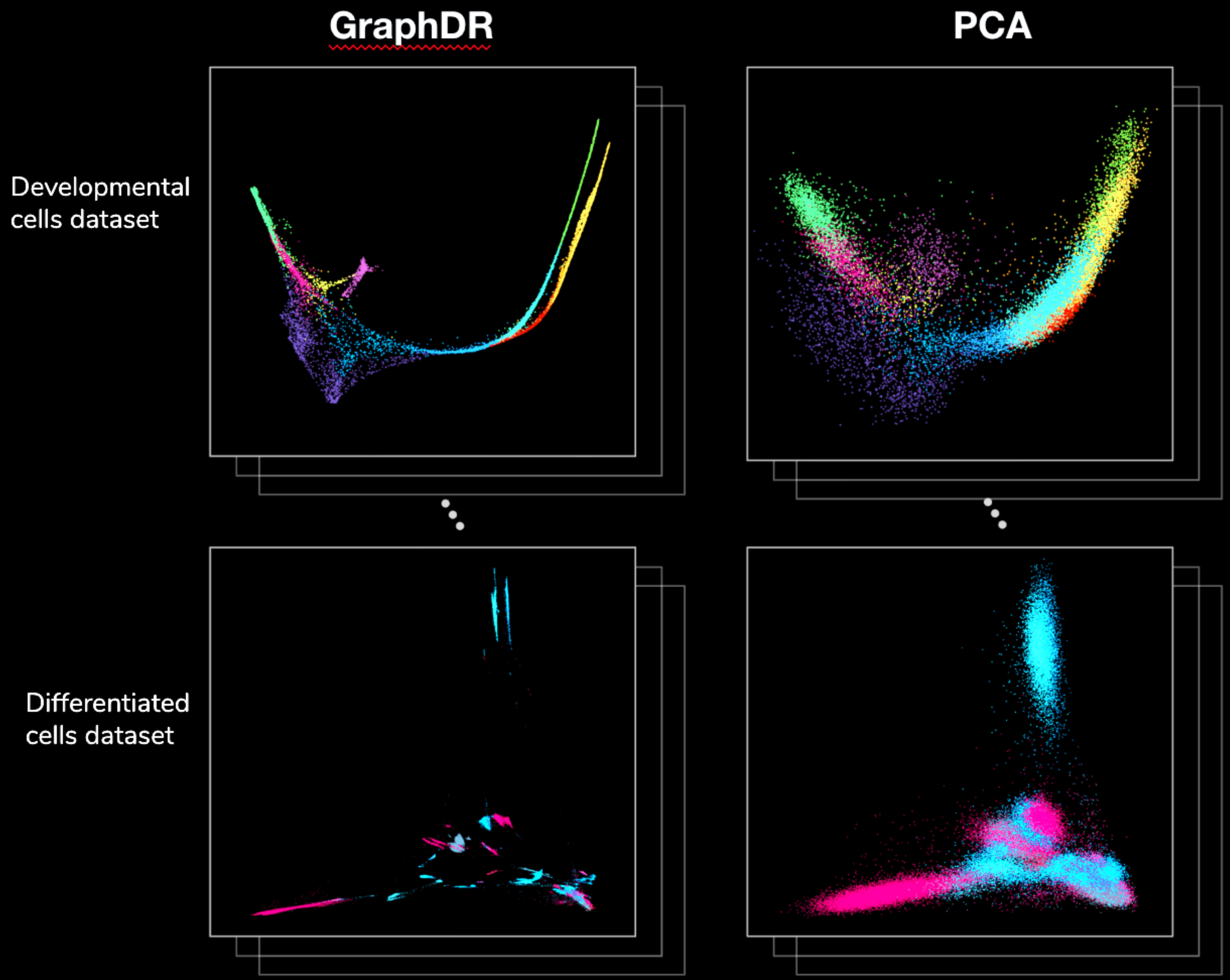


Can we combine advantages of linear and nonlinear representations?

Our solution: from linear to **quasilinear** representation

“Quasilinear” approaches:

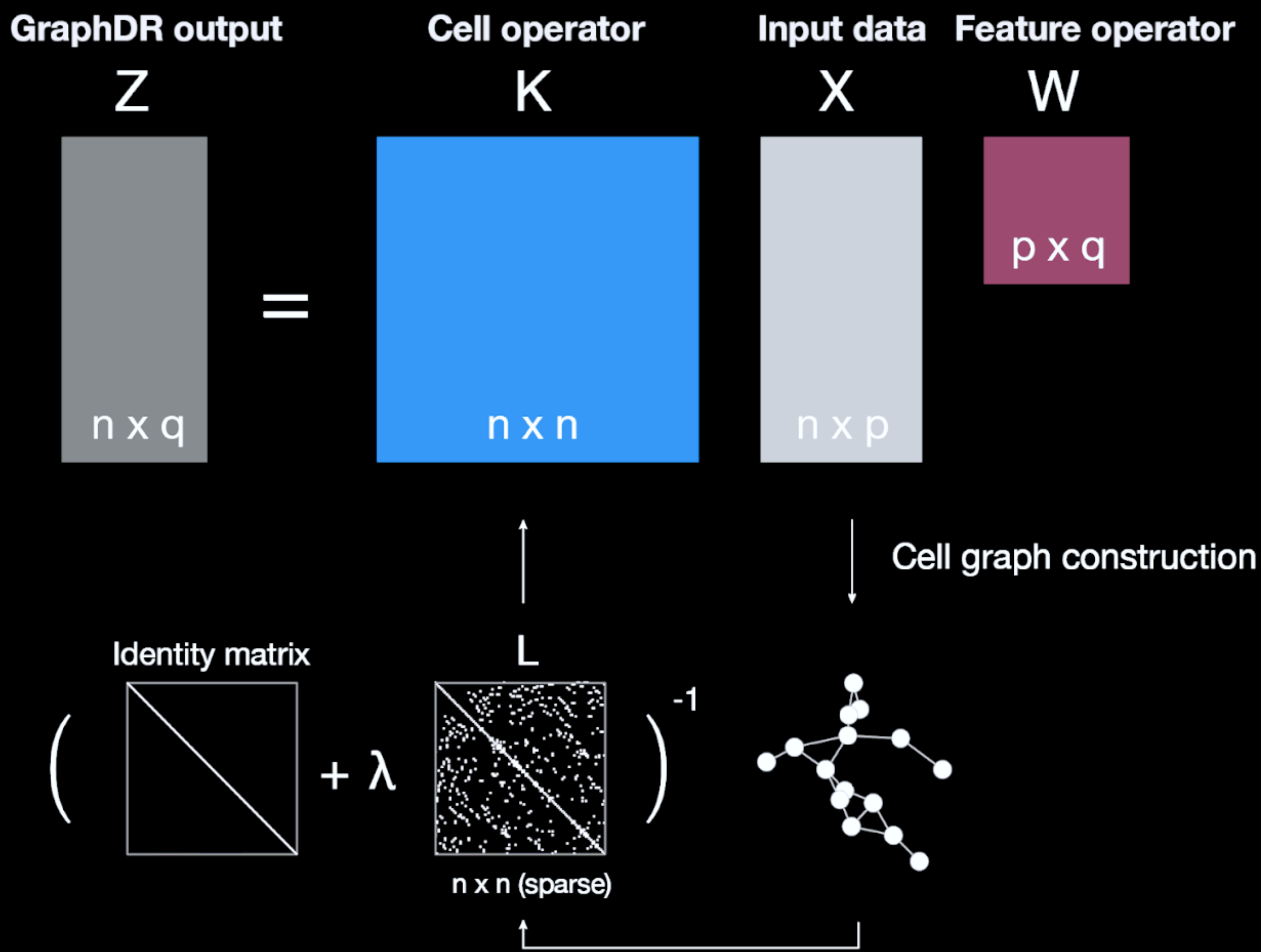
GraphDR - quasilinear visualization and general representation



Data from Hochgerner et al. 2018, Zeisel et al,. 2018

“Quasilinear” approaches:

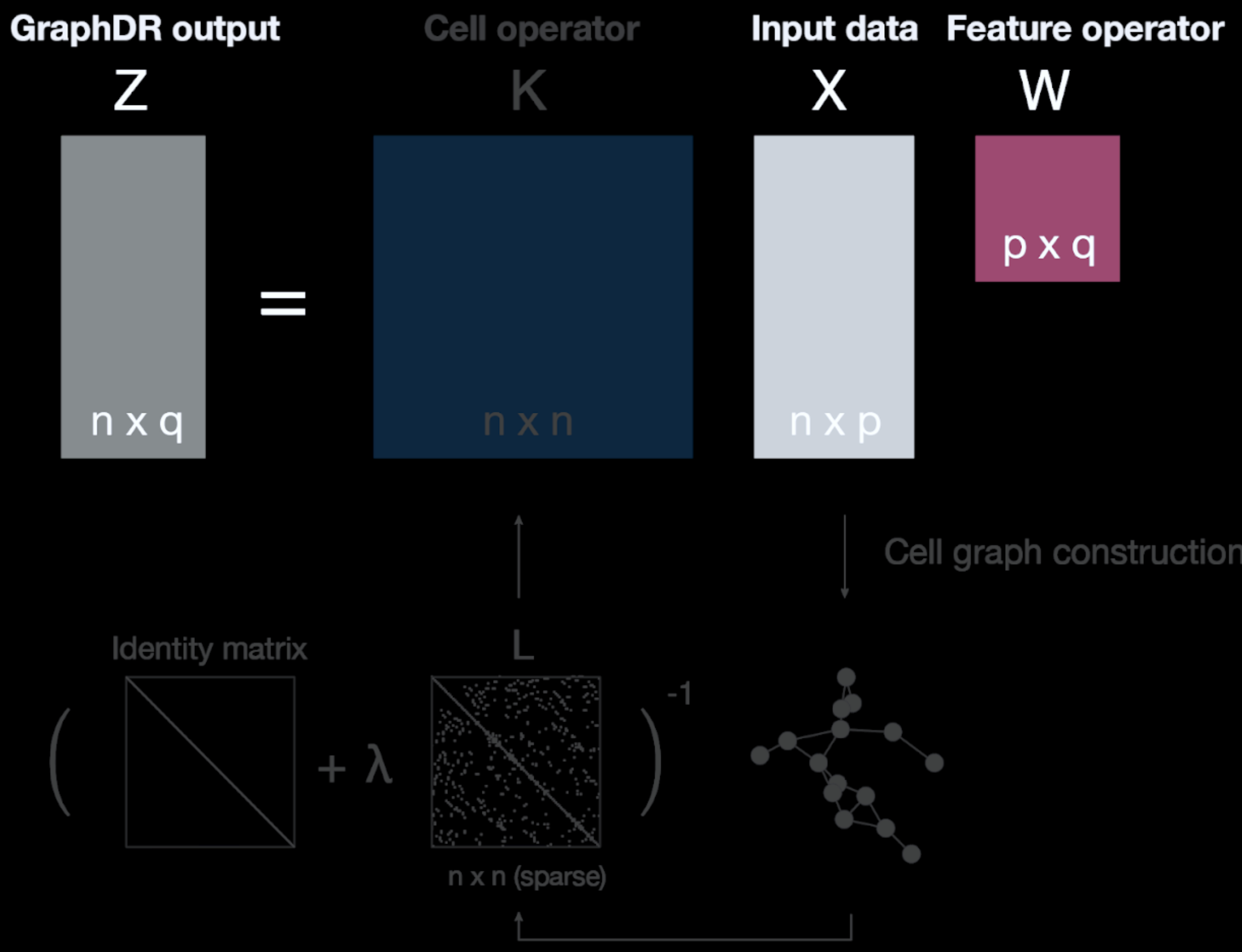
GraphDR - quasilinear visualization and general representation



GraphDR objective: $$\underset{W,Z}{\text{minimize}} \quad \|X - ZW^{\mathrm{T}}\|_2^2 + \lambda \sum_{\{i,j\}\in G} G_{ij}\left\|Z_i - Z_j\right\|_2^2, \qquad s.t.\ W^T W = I$$

“Quasilinear” approaches:

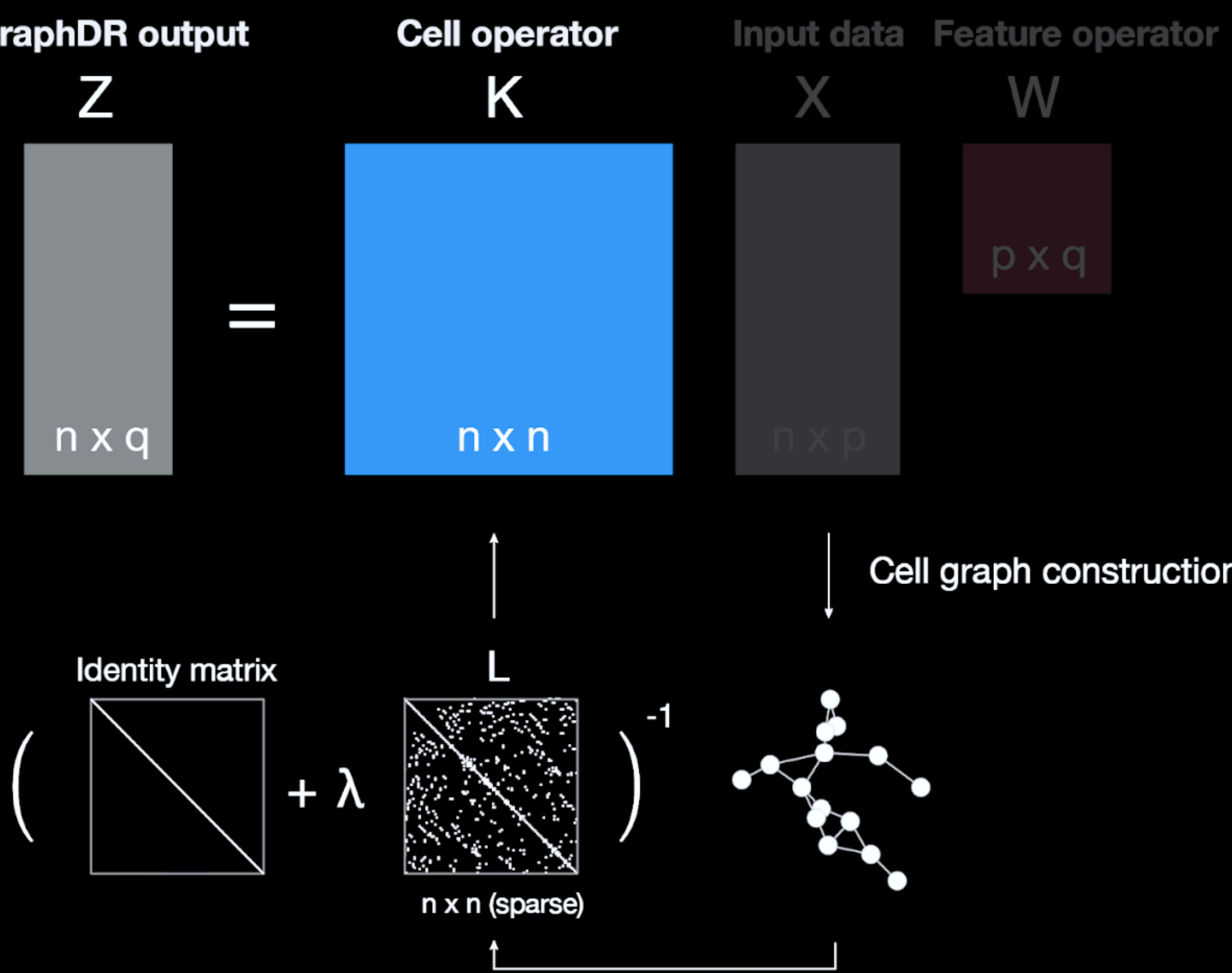
GraphDR - quasilinear visualization and general representation



GraphDR objective function: $\underset{W,Z}{minimize} \; \|XW - Z\|_2^2 + \lambda \sum_{\{i,j\}\in G} G_{ij}\|Z_i - Z_j\|_2^2, \quad s.t. \; W^T W = I$

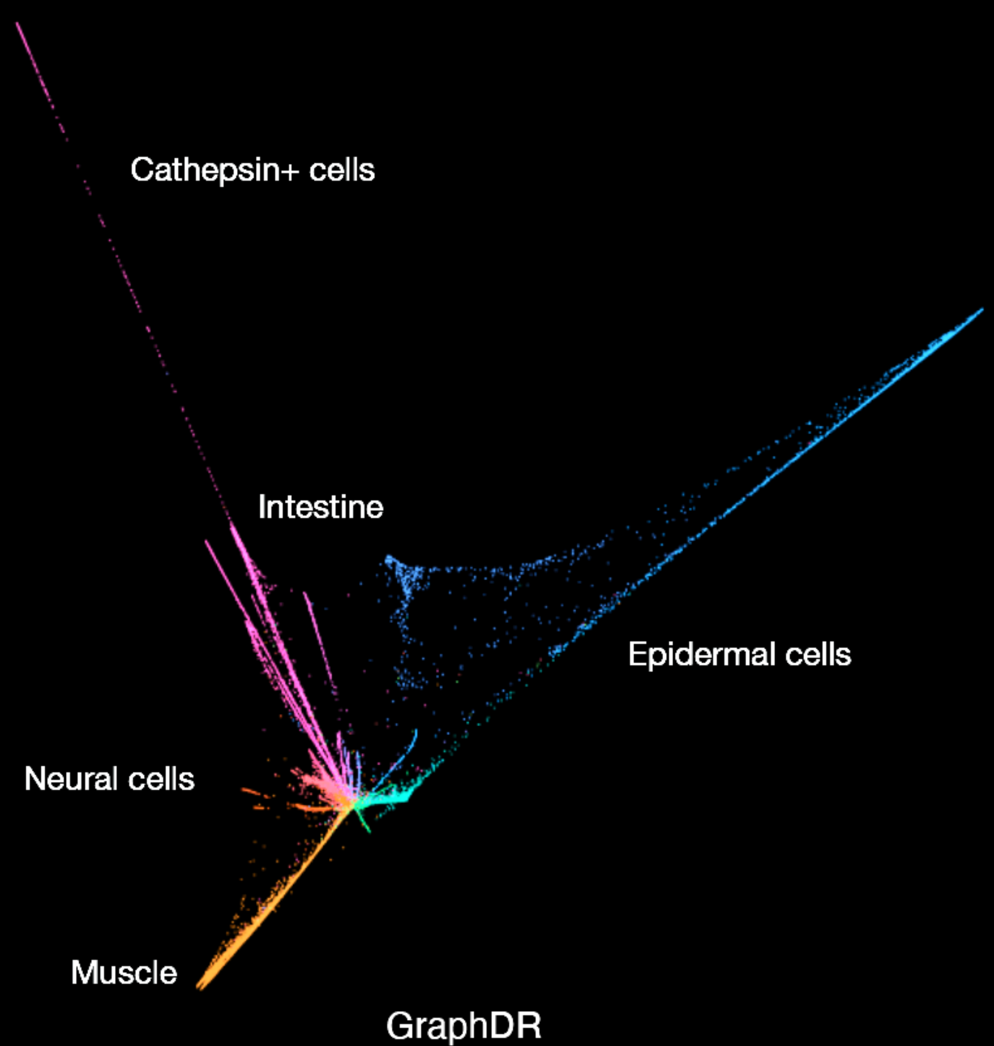
“Quasilinear” approaches:

GraphDR - quasilinear visualization and general representation

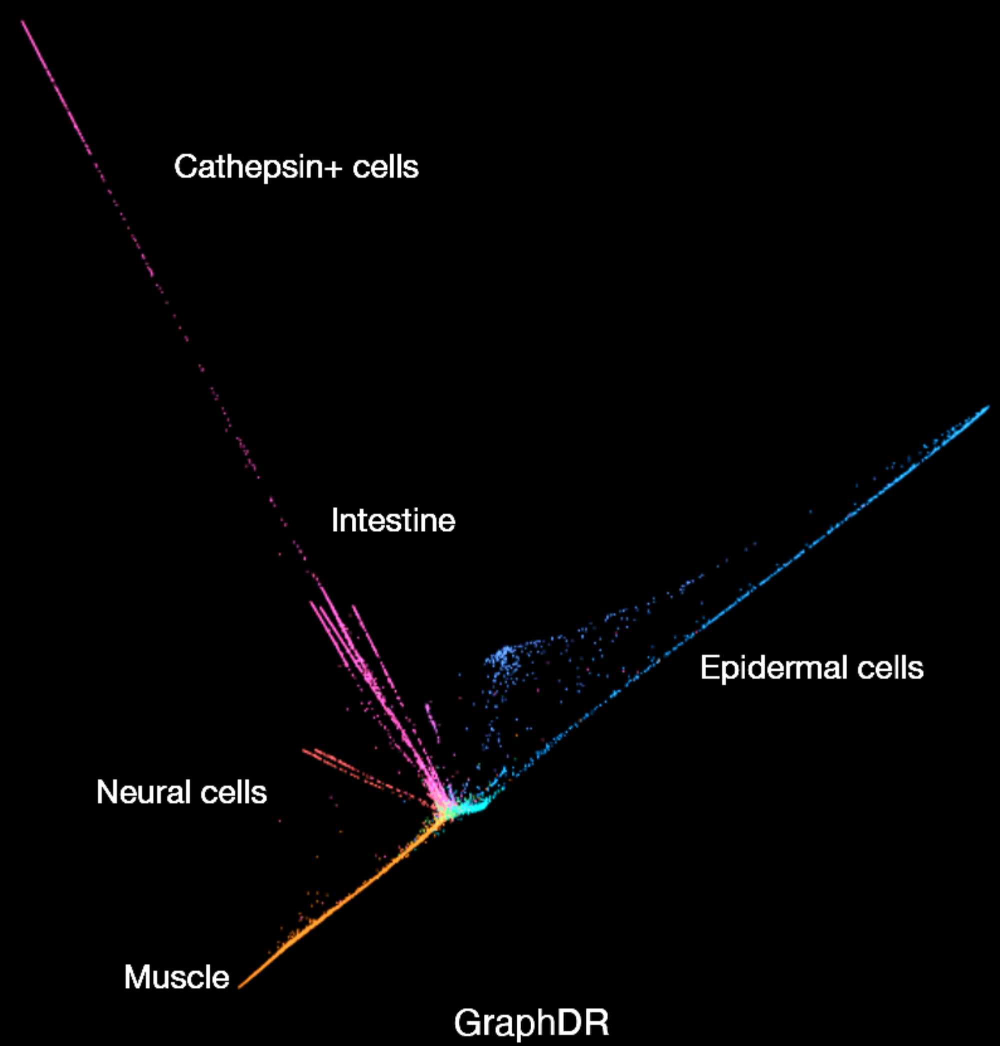


GraphDR objective: $$\underset{W,Z}{\text{minimize}} \quad \|X - ZW^{\mathrm{T}}\|_2^2 + \lambda \sum_{\{i,j\}\in G} G_{ij}\left\|Z_i - Z_j\right\|_2^2, \qquad s.t.\ W^TW = I$$

GraphDR representations allow direct comparison across datasets

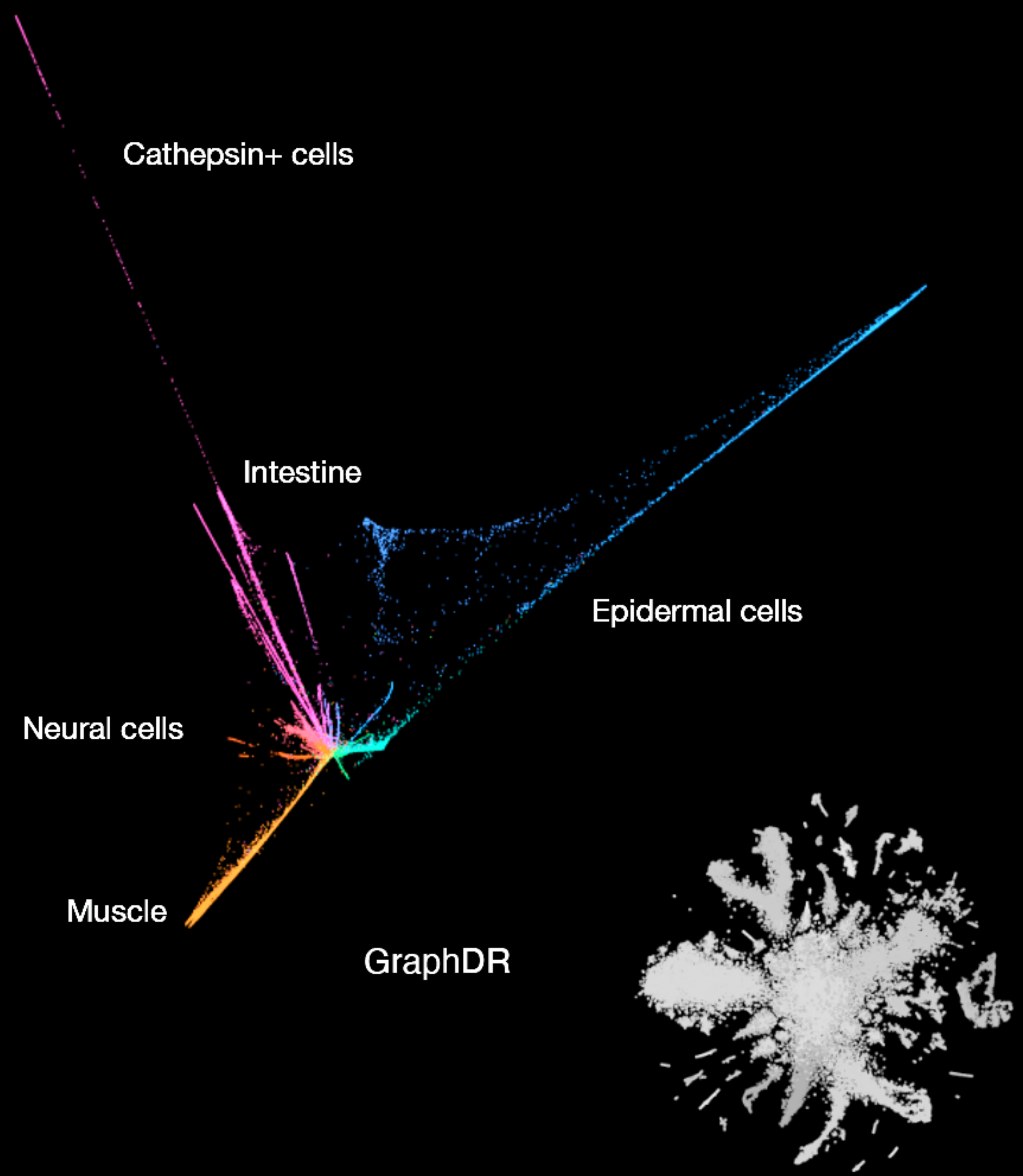


Planarian whole animal
(Fincher et. al.)

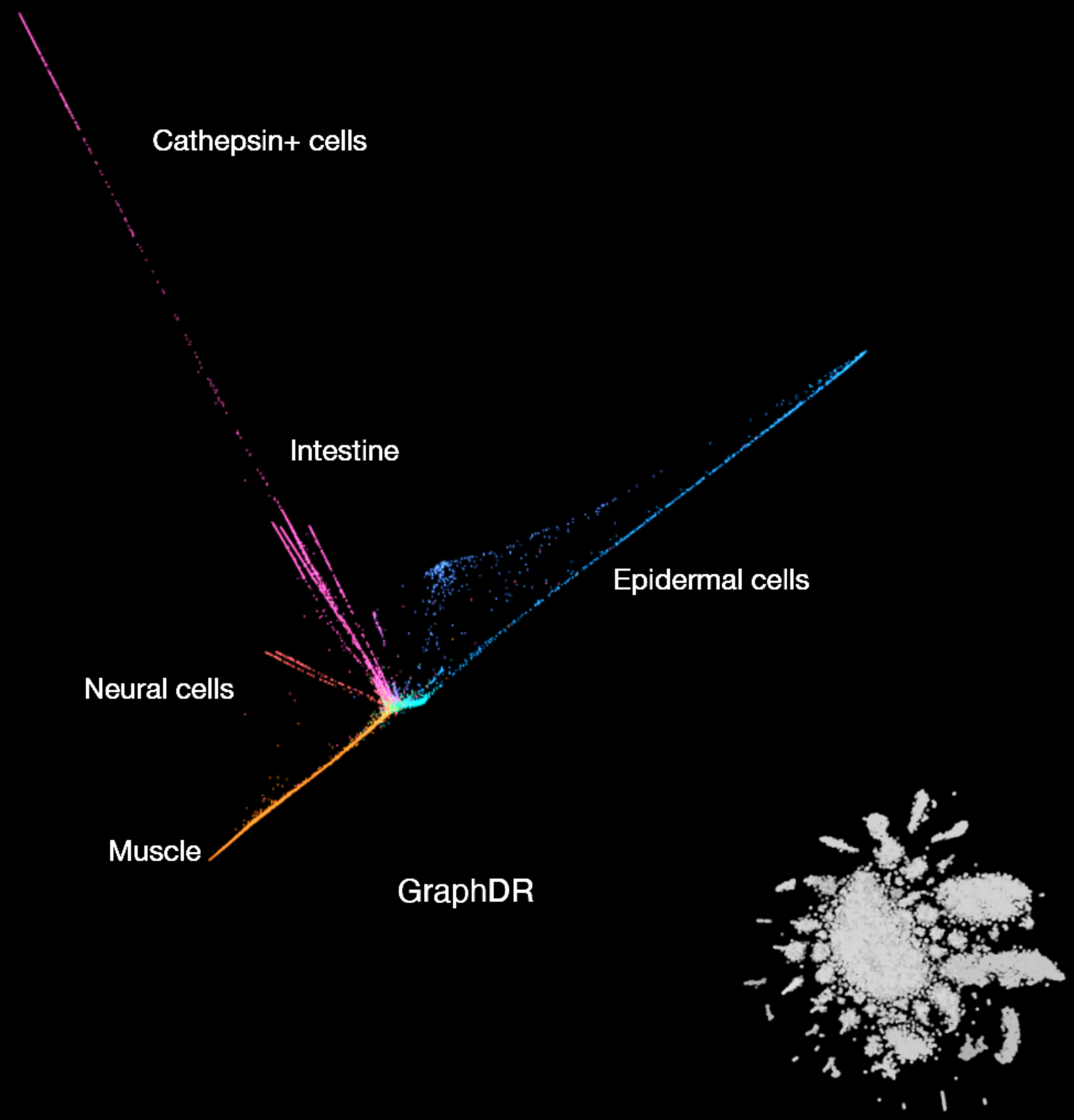


Planarian whole animal
(Plass et. al.)

GraphDR representations allow direct comparison across datasets

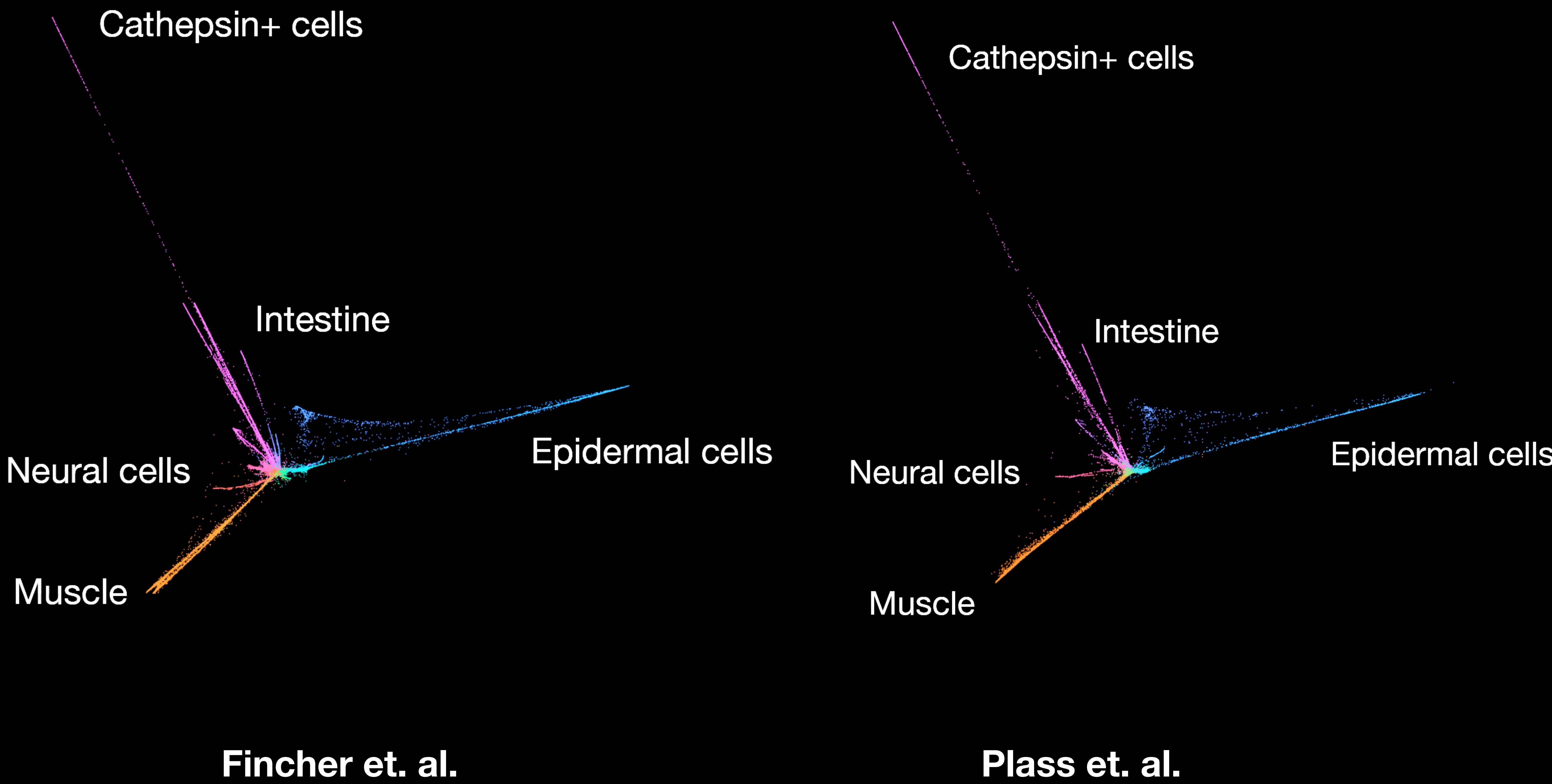


Planarian whole animal (Fincher et. al.)

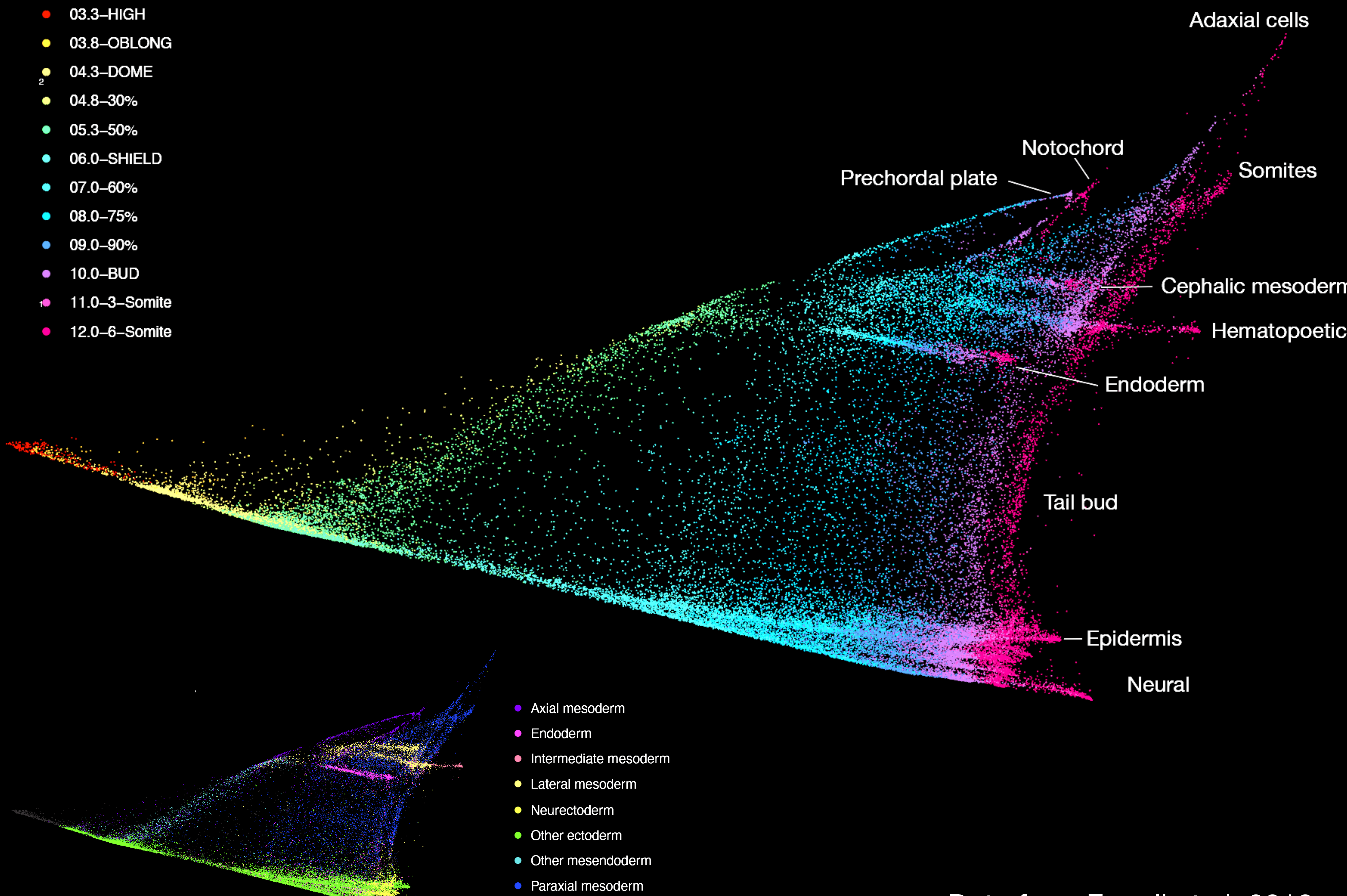


Planarian whole animal (Plass et. al.)

Improved comparison cross datasets with graph-based alignment

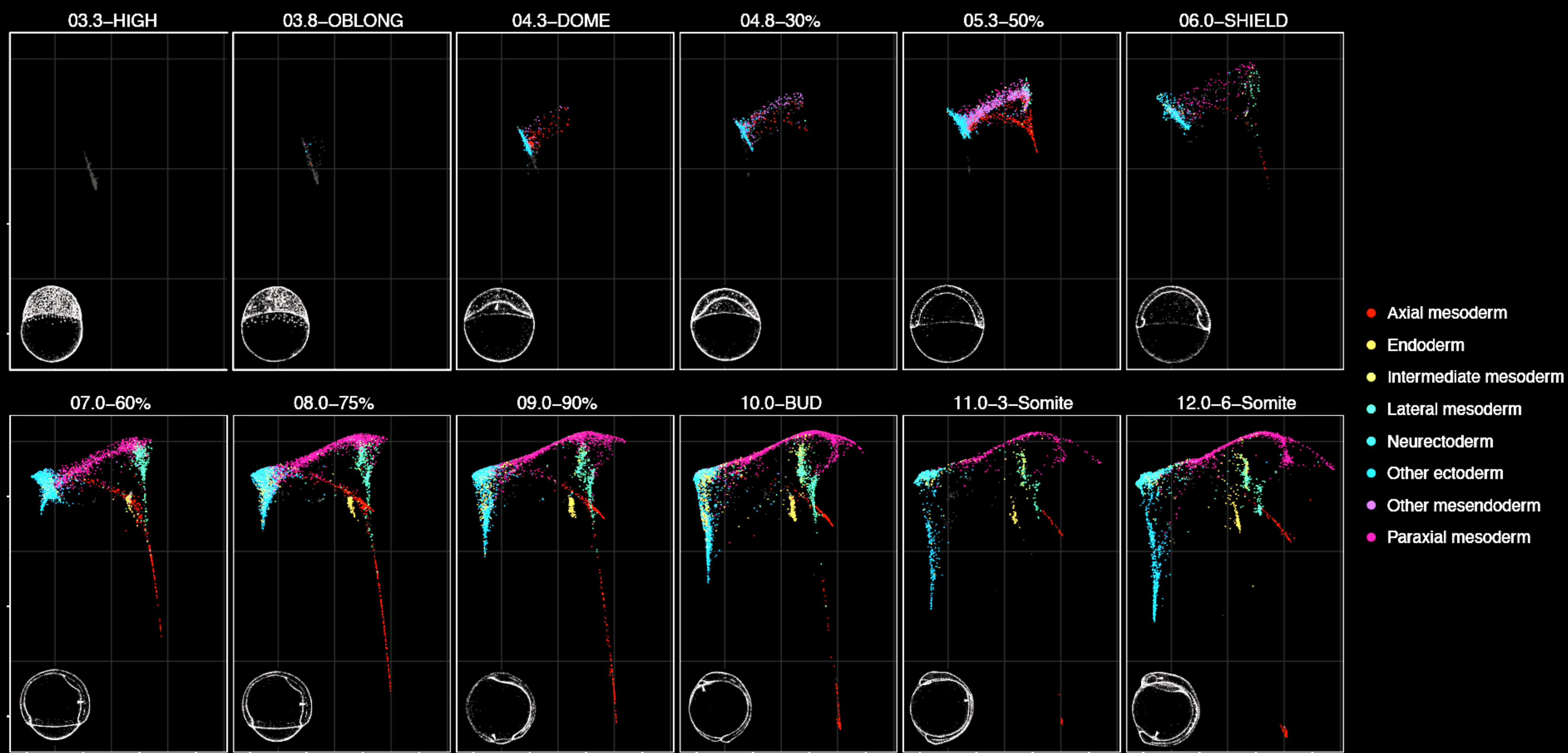


GraphDR visualization of zebrafish embryonic development

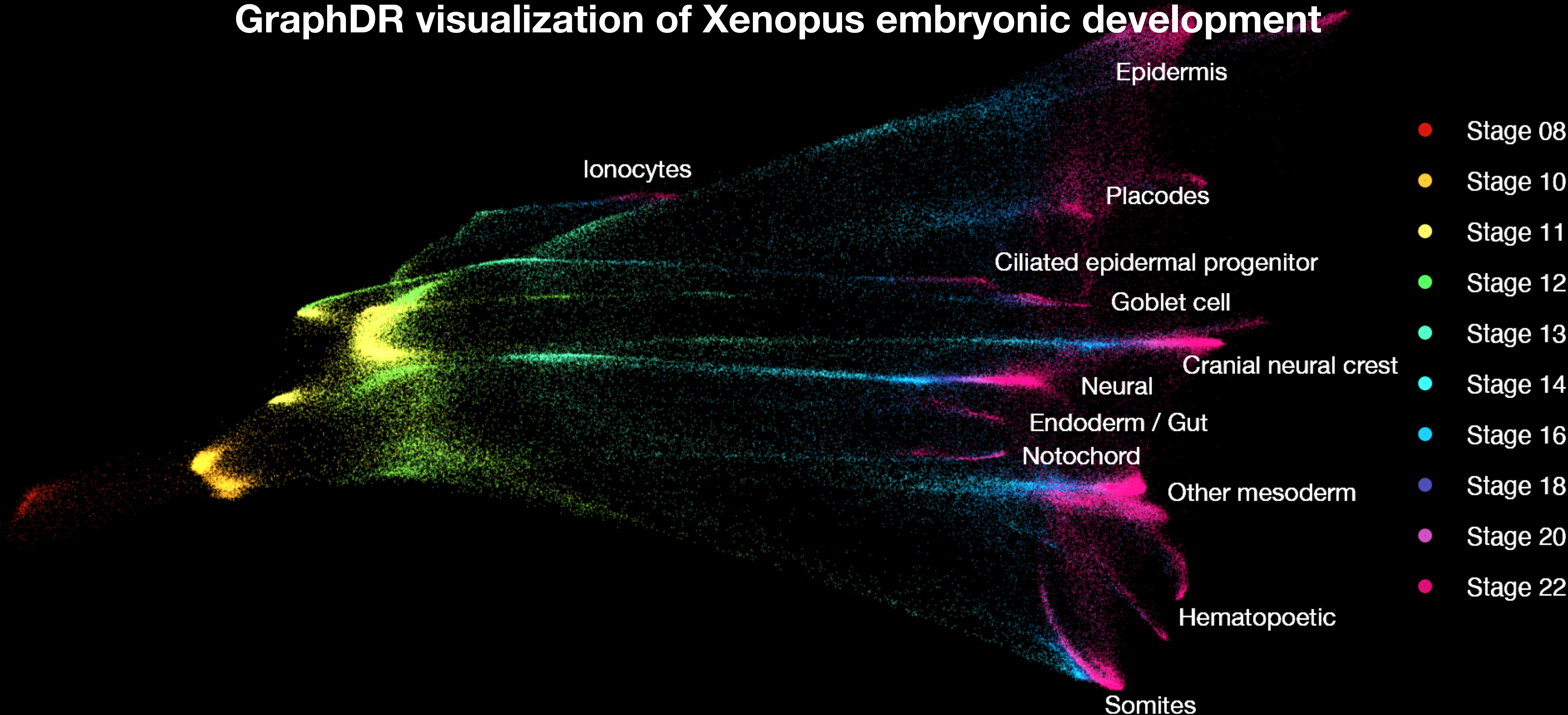


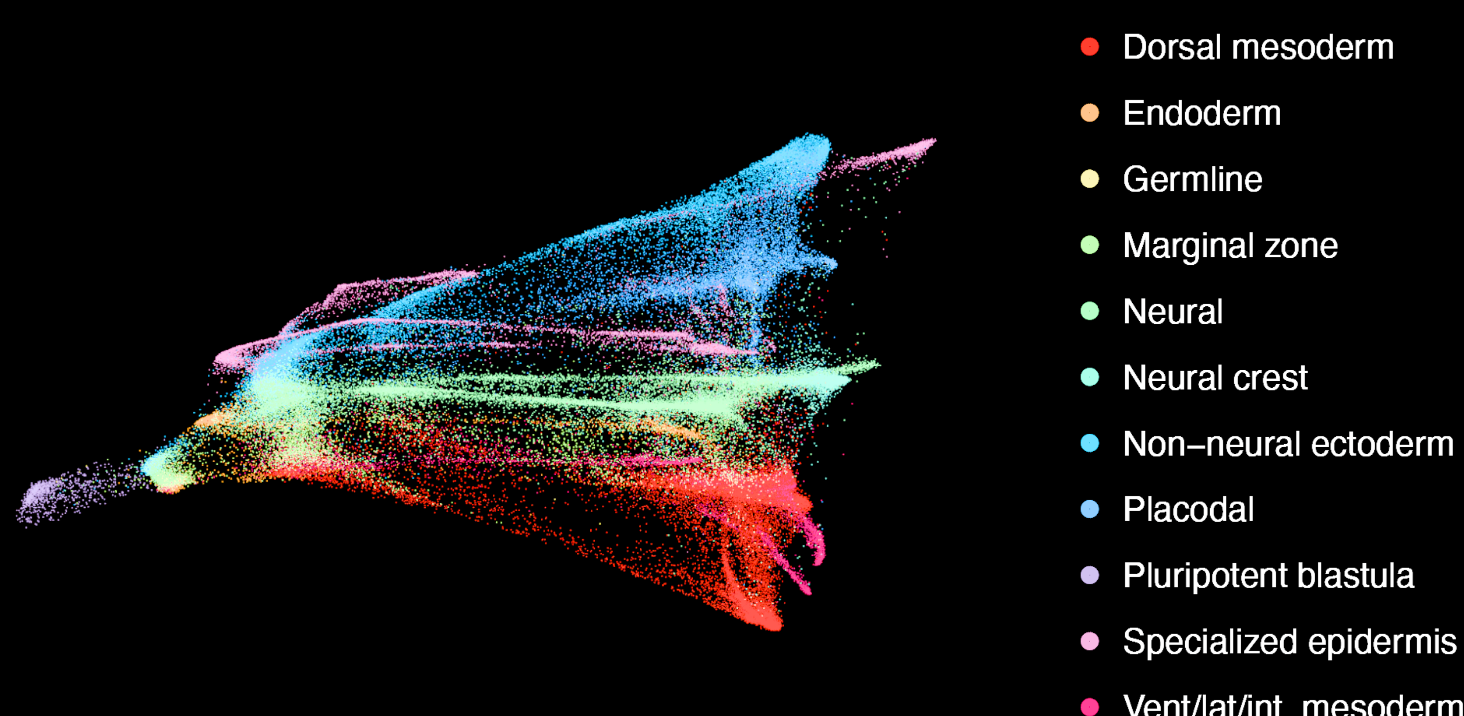
Data from Farrell et al. 2018

Uniform interpretability allows direct comparison across temporal “slices”

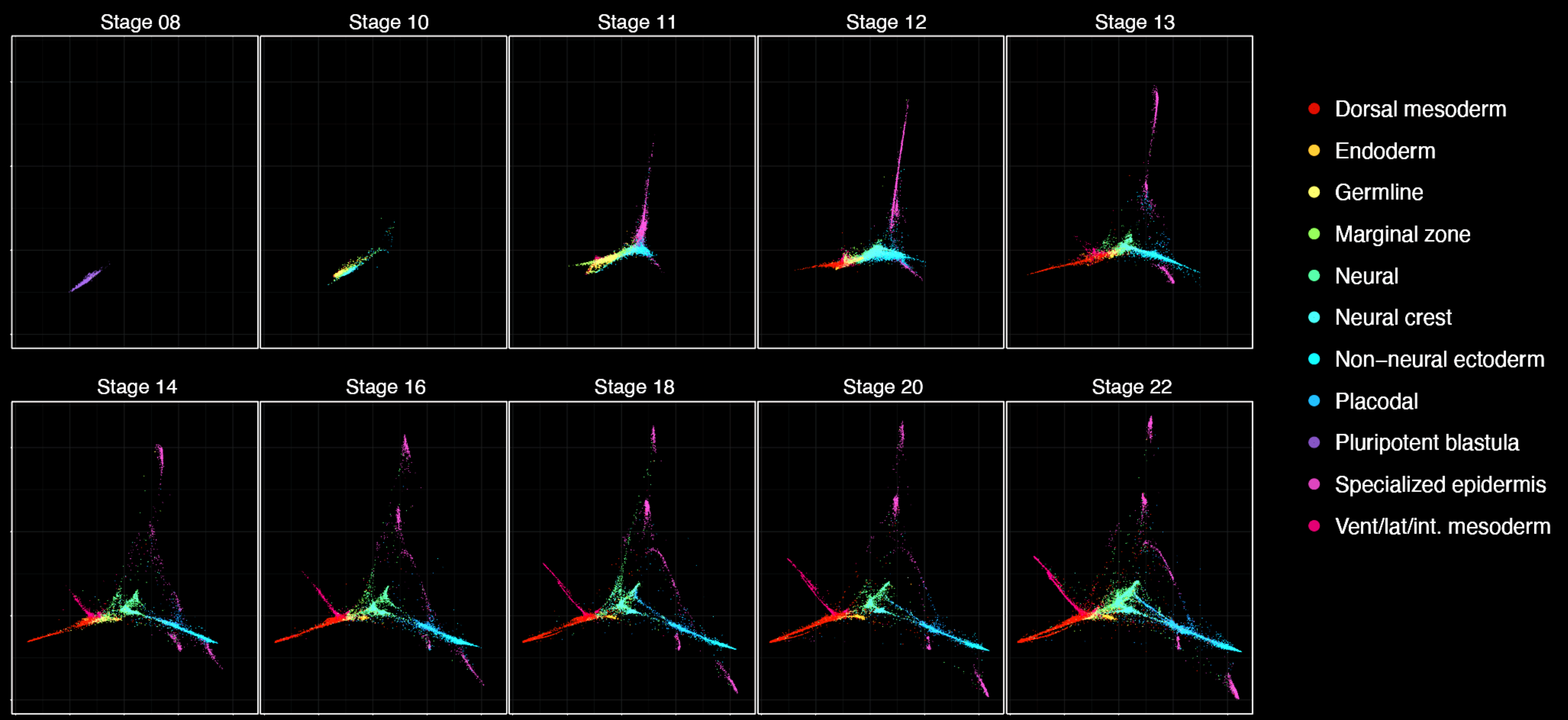


GraphDR visualization of Xenopus embryonic development





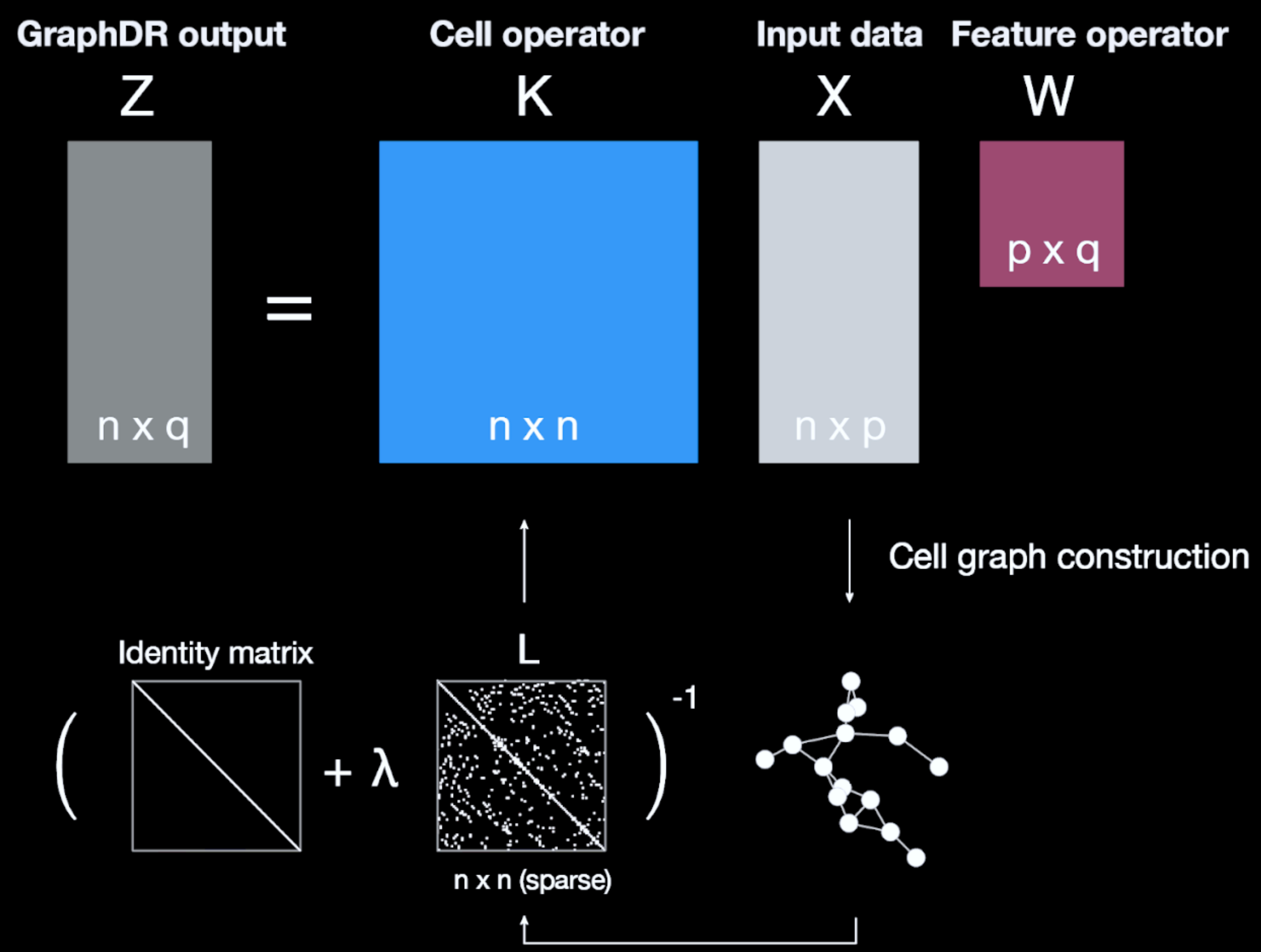
Data from Briggs et al. 2018



Stage 08
Stage 10
Stage 11
Stage 12
Stage 13
Stage 14
Stage 16
Stage 18
Stage 20
Stage 22
Dorsal mesoderm
Endoderm
Germline
Marginal zone
Neural
Neural crest
Non–neural ectoderm
Placodal
Pluripotent blastula
Specialized epidermis
Vent/lat/int. mesoderm

“Quasilinear” approaches:

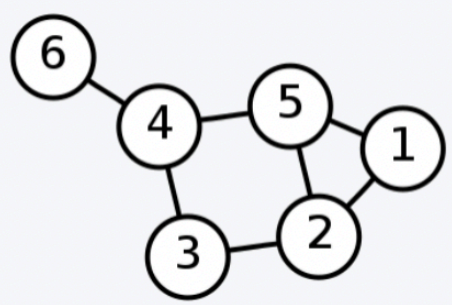
GraphDR - visualization and general-purpose representation



GraphDR objective: $$\underset{W,Z}{\mathit{minimize}} \quad \|X - ZW^{T}\|_2^2 + \lambda \sum_{\{i,j\}\in G} G_{ij}\|Z_i - Z_j\|_2^2, \qquad s.t.\ W^T W = I$$

New release! github.com/jzthree/quasildr

Graph Laplacian

Labelled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$

From https: // en. wikipedia. org/ wiki/ Laplacian_ matrix .

Why does it appear?

It automatically emerge when you compute sum of pairwise L2 distances

$$\text{trace}(X^T L X) = \sum_{(i,j)\in G} \|x_i - x_j\|^2$$

“Quasilinear” approaches:

Implement GraphDR (you can do it in ~10 lines of code!)

GraphDR objective: $$\underset{W,Z}{minimize} \quad \|X - ZW^{\mathrm{T}}\|_2^2 + \lambda \sum_{\{i,j\}\in G} G_{ij}\left\|Z_i - Z_j\right\|_2^2, \qquad s.t.\ W^T W = I$$

Analytical solution: $$Z = (I + \lambda L)^{-1} X W$$

W represents top d eigenvectors of $X^T (I + \lambda L)^{-1} X$

1. You may use: sklearn.neighbors.kneighbors_graph, scipy.sparse.csgraph.laplacian, scipy.sparse.eye, numpy.linalg.inv

2. Make skipping computing W an option (`no_rotation=True`):
   $Z' = (I + \lambda L)^{-1} X$ is the solution to the objective when we add the constraint $W = I$

3. Z does not have to reduce the number of dimensions in GraphDR

4. (Optional) make it scalable to > 1 million cells

Task for Day 3:

1. **Code review for Day 2 (Docstring for t-SNE)**
2. **Implement GraphDR (Code + Docstring)**
3. **Optional: 3D visualization (with plot.ly) or / and interactive visualization interface that allows adjusting regularization parameter (e.g. with Dash)**