Briefings in Bioinformatics, 22(5), 2021, 1–13

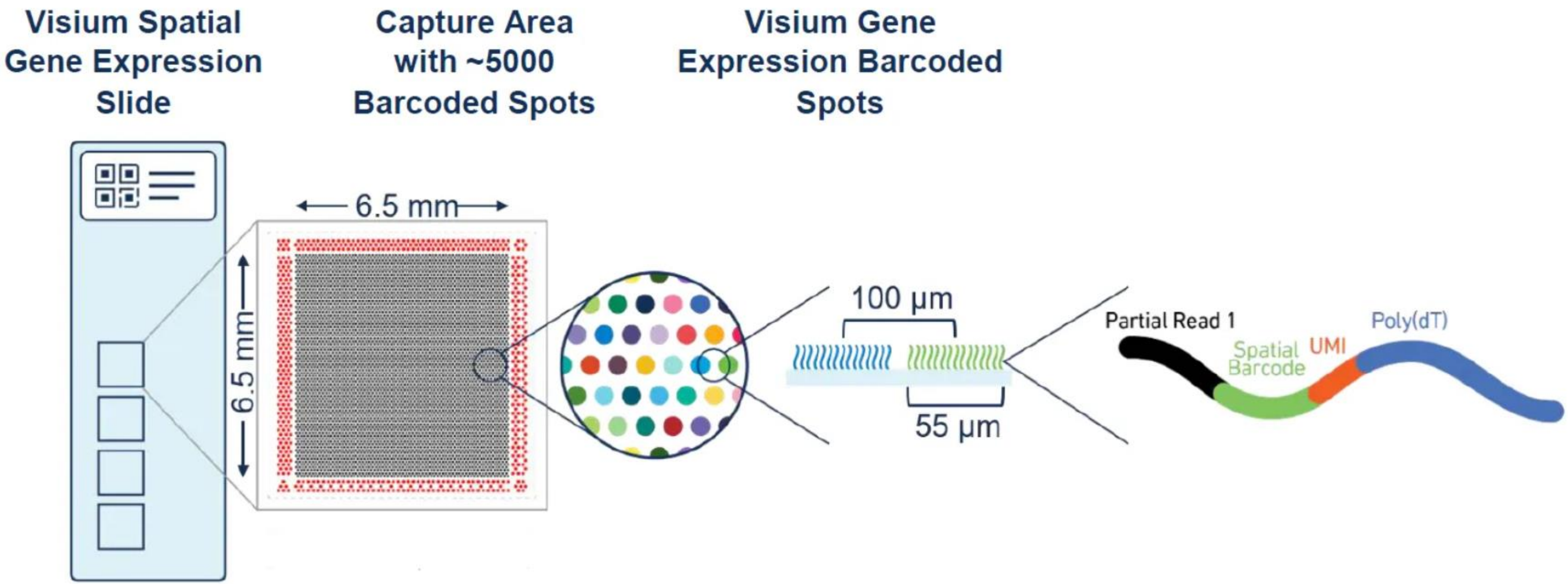
https://doi.org/10.1093/bib/bbaa414
Method Review

DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence

Qianqian Song and Jing Su

Corresponding authors: Jing Su, Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202, USA. Tel.: +1-678-642-7428. E-mail: su1@iu.edu; Qianqian Song, Department of Cancer Biology, Wake Forest School of Medicine, Winston Salem, NC 27157, USA. Tel.: +1-336-926-4972. E-mail: qsong@wakehealth.edu

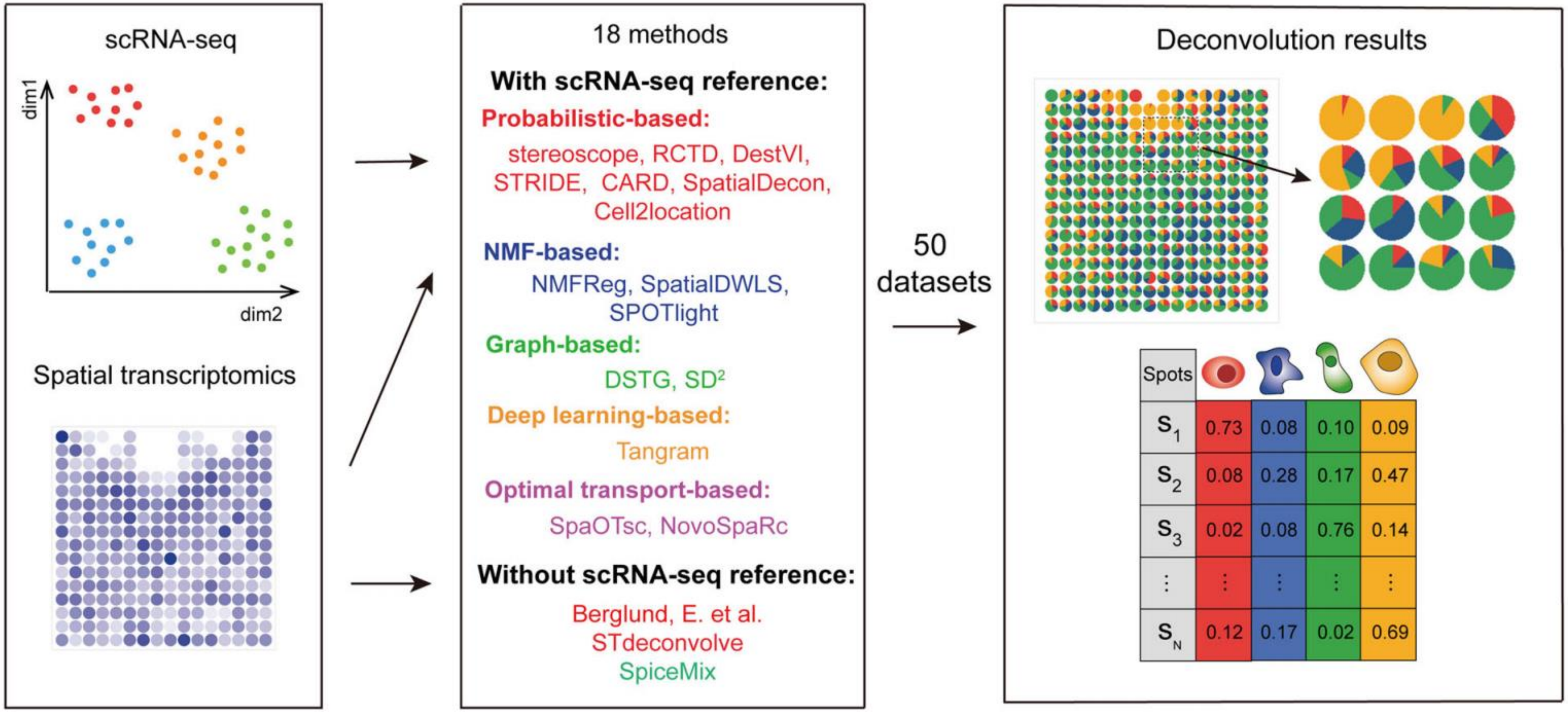
10X Visium


Visium Spatial Gene Expression Slide
Capture Area with ~5000 Barcoded Spots
Visium Gene Expression Barcoded Spots
6.5 mm
6.5 mm
100 µm
55 µm
Partial Read 1
Spatial Barcode
UMI
Poly(dT)

Cell Type Deconvolution

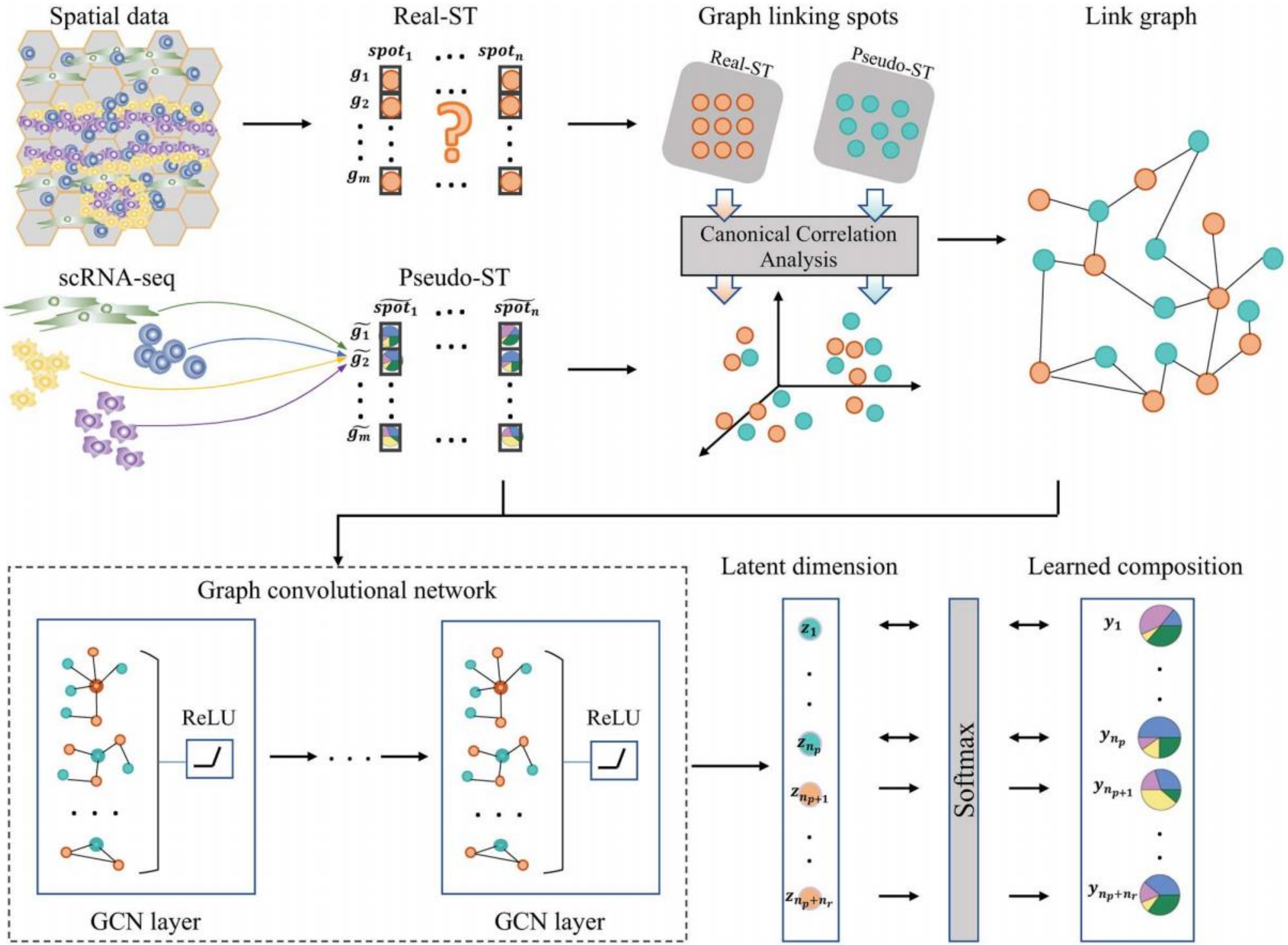
- A single spot usually contains multiple cells (2-15).
- The gene expression vector in a spot is extremely sparse and the applicability of linear regression methods is diminished.

Cellular Deconvolution Methods



DSTG

任务：节点预测



Spatial data
Real-ST
Graph linking spots
Link graph
Real-ST
Pseudo-ST
Canonical Correlation Analysis
scRNA-seq
Pseudo-ST
Graph convolutional network
ReLU
ReLU
GCN layer
GCN layer
Latent dimension
Softmax
Learned composition

Create Graph - Canonical Correlation Analysis

2 types of node

$$X = (x_1, \ldots, x_n)' \qquad Y = (y_1, \ldots, y_m)'$$

Mutual covariance matrix($n \times m$)

$$\Sigma_{XY} = \mathrm{cov}(X, Y) \qquad \mathrm{cov}(x_i, y_j)$$

Find a' and b'

$$\rho = \mathrm{corr}(a'X, b'Y)$$

Find another a'' and b'' which is orthogonal to a' and b'
$\min(n, m)$ times

canonical correlation analysis

$$\mu_s{}^T \left(X_{\text{pseudo}}{}^{m \times n_p} \right)^T X_{\text{real}}{}^{m \times n_r} \, \nu_s,$$

Subject to

$$\|\mu_s\|_2^2 \leq 1 \text{ and } \|\nu_s\|_2^2 \leq 1$$

Find 20 canonical correlation vector pairs using SVD

Create Graph - KNN

- KNN
  - Node i is the neighborhood of node j with KNN
- MNN
  - Node i is the neighborhood of node j with KNN
  - Node j is the neighborhood of node i with KNN

GCN

$$X = \left[X_{\text{pseudo}} \;\; X_{\text{real}}\right] \in R^{m \times N}$$

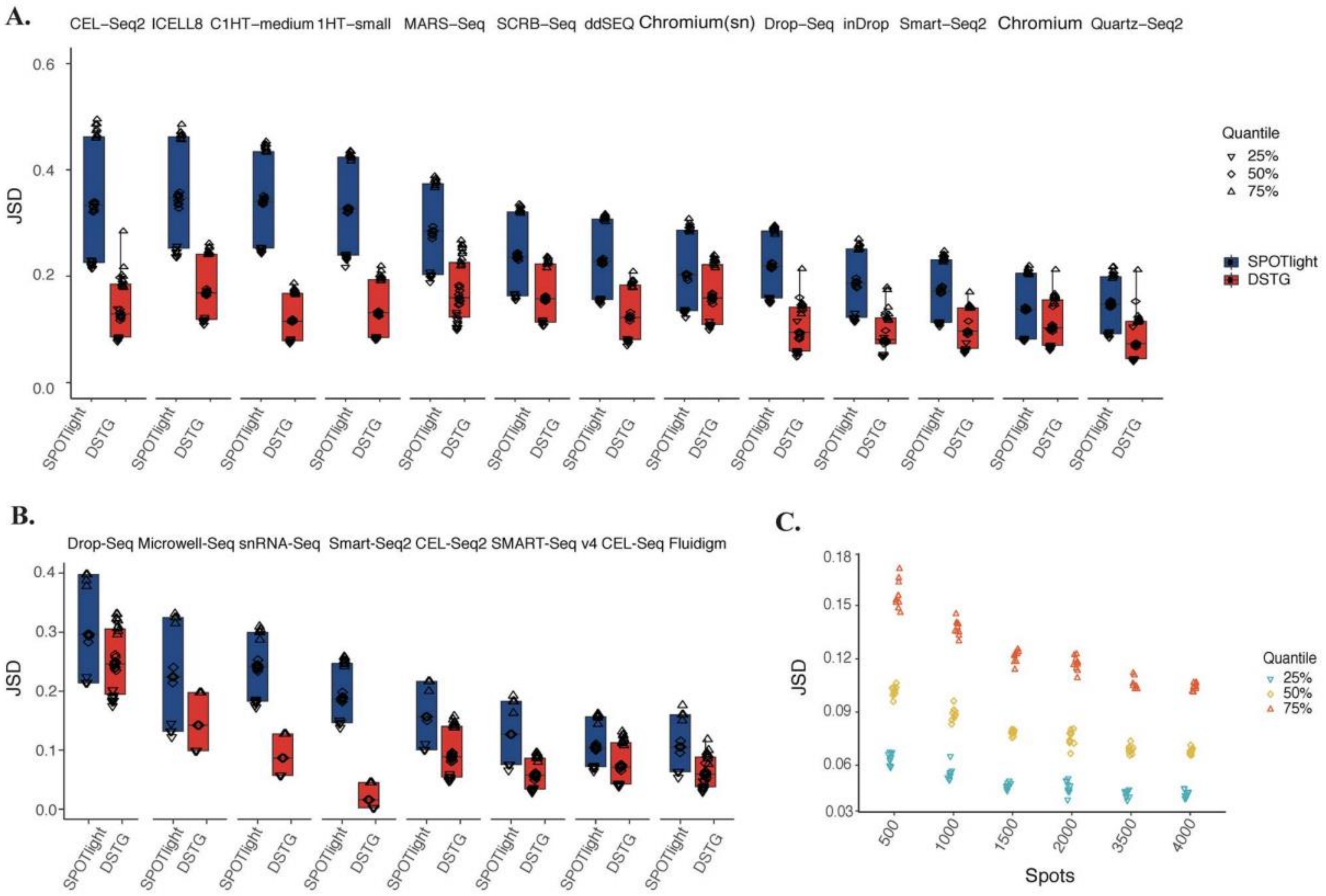
$$\widetilde{A} = \check{D}^{-1/2} \hat{A} \check{D}^{-1/2}$$

$$H^{(l+1)} = f\left(H^{(l)}, \widetilde{A}\right) = \sigma\left(\widetilde{A} H^{(l)} W^{(l)}\right) = \mathrm{ReLU}\left(\widetilde{A} H^{(l)} W^{(l)}\right)$$

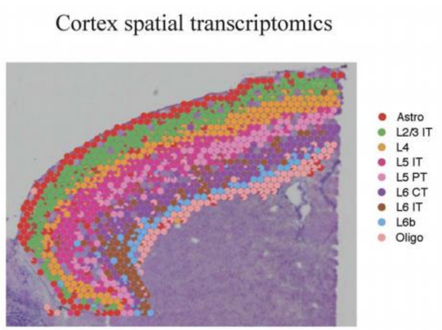
$$\hat{Y} = f\left(X, A^{H}\right) = \mathrm{softmax}(\widetilde{A}\ \mathrm{ReLU}(\widetilde{A} X^{T} W^{(0)}) W^{(1)})$$

Loss function $\mathcal{L} = -\sum_{i=1}^{n_p} \sum_{f=1}^{F} y_{i,f} \ln\left(\hat{y}_{i,f}\right)$

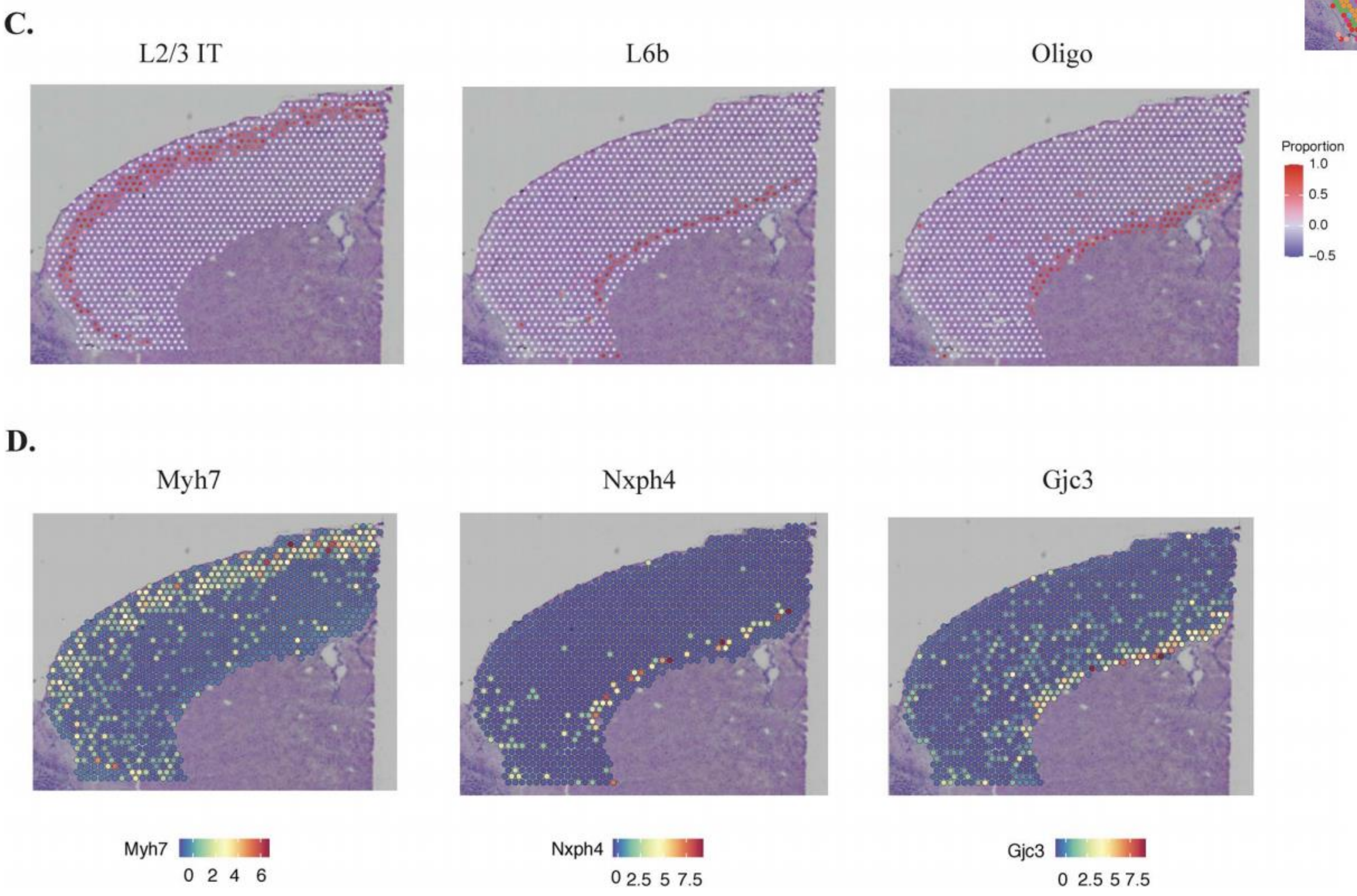
Results


A.
CEL–Seq2
ICELL8
C1HT–medium
1HT–small
MARS–Seq
SCRB–Seq
ddSEQ
Chromium(sn)
Drop–Seq
inDrop
Smart–Seq2
Chromium
Quartz–Seq2
JSD
0.0
0.2
0.4
0.6
SPOTlight
DSTG
Quantile
25%
50%
75%
SPOTlight
DSTG
B.
Drop-Seq
Microwell-Seq
snRNA-Seq
Smart-Seq2
CEL-Seq2
SMART-Seq v4
CEL-Seq
Fluidigm
JSD
0.0
0.1
0.2
0.3
0.4
SPOTlight
DSTG
C.
JSD
0.03
0.06
0.09
0.12
0.15
0.18
500
1000
1500
2000
3500
4000
Spots
Quantile
25%
50%
75%

Results



Cortex spatial transcriptomics
Astro
L2/3 IT
L4
L5 IT
L5 PT
L6 CT
L6 IT
L6b
Oligo



C.
L2/3 IT
L6b
Oligo
Proportion
1.0
0.5
0.0
−0.5
D.
Myh7
Nxph4
Gjc3
Myh7
0 2 4 6
Nxph4
0 2.5 5 7.5
Gjc3
0 2.5 5 7.5