

Assessing Normalization Methods in Spatial Transcriptomics

BEN ANDERSON

STAT 877

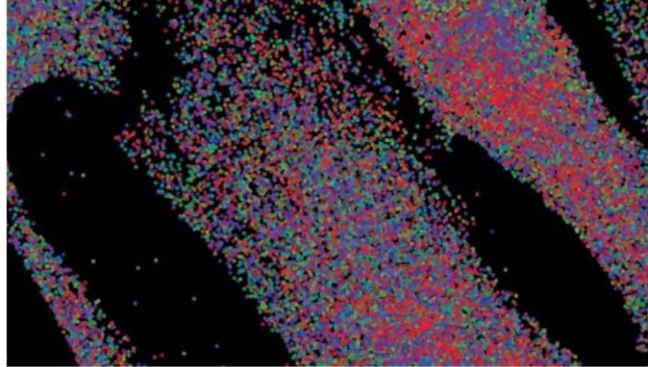
5/4/2022

Spatial Transcriptomics

FOCUS | 06 JANUARY 2021

Method of the Year 2020: spatially resolved transcriptomics

Spatially resolved transcriptomics is our Method of the Year 2020, for its ability to provide valuable insights into the biology of cells and tissues while retaining information about spatial context.

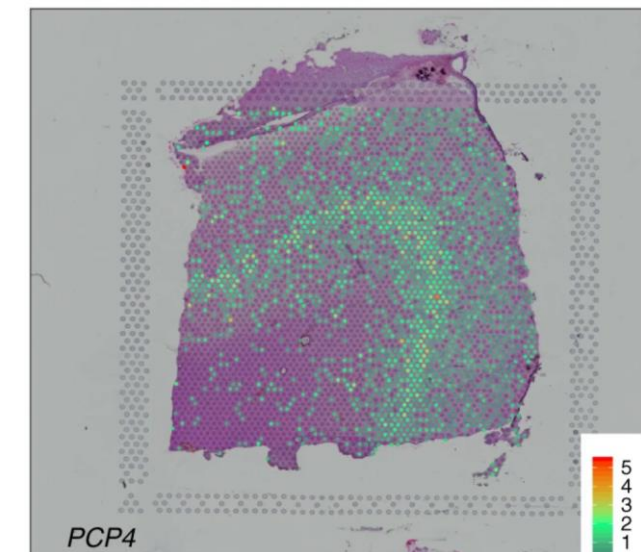
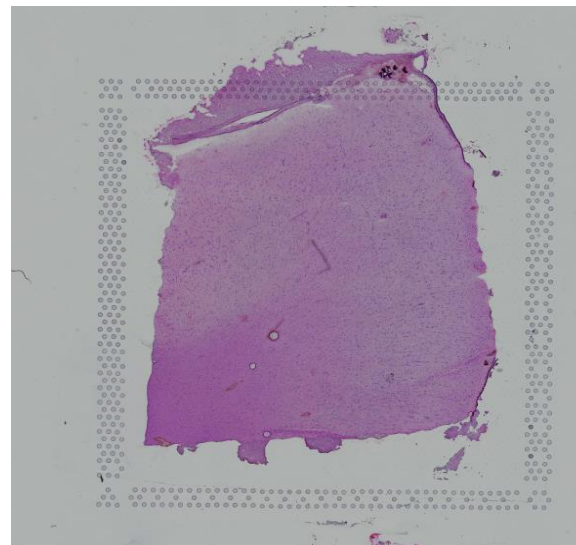
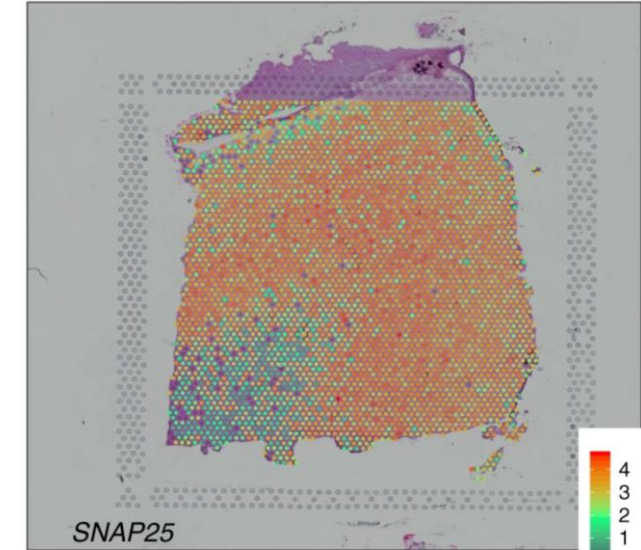


Major Benefit: Retains spatial information between samples

Spatial LIBD Paper

10x Genomics *Visium* method

Spatial analysis of brain tissue slices



Normalizing RNAseq data is Important

Key assumption for normalization

True biological gene expression is masked by technical noise

Within-Sample variance

- Length of Gene
Longer gives more reads
- GC-content
Affects amplification rate

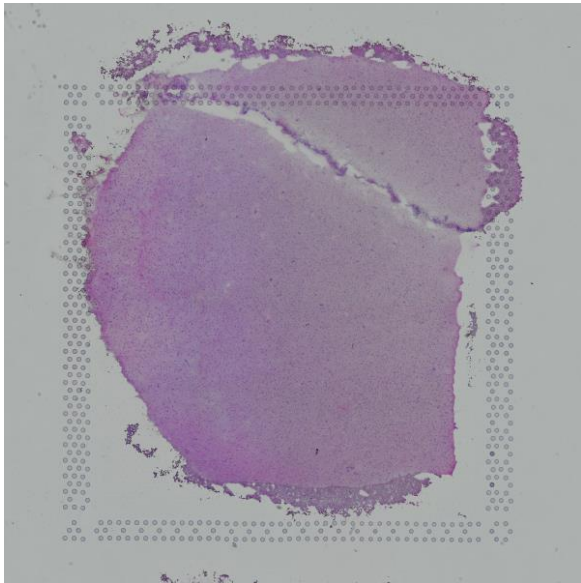
Between-Sample variance

- Sequencing depth
Biased against low exp. genes
- Sample preparation
How good is your pipetting?

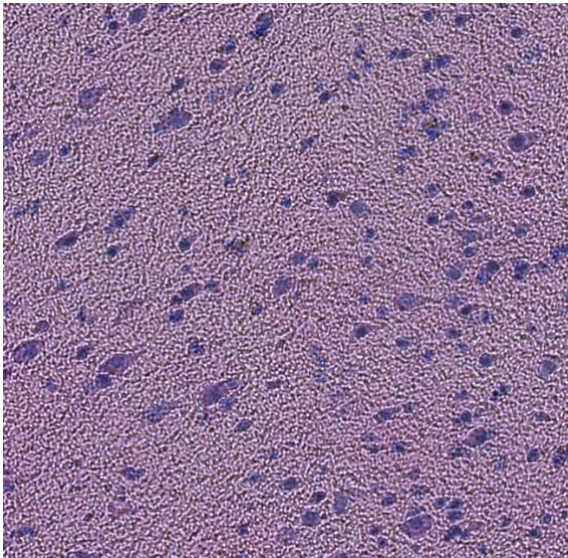
Challenges in Spatial Normalization

10x Genomics Visium samples all material in spot

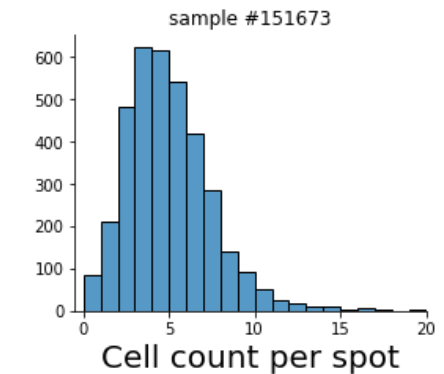
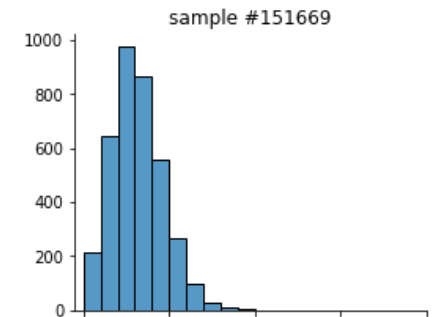
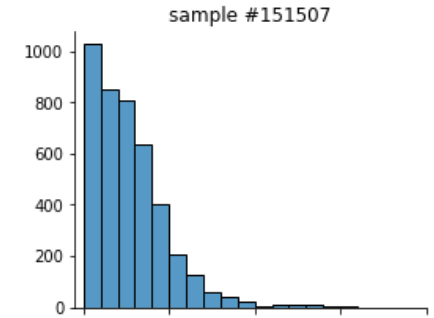
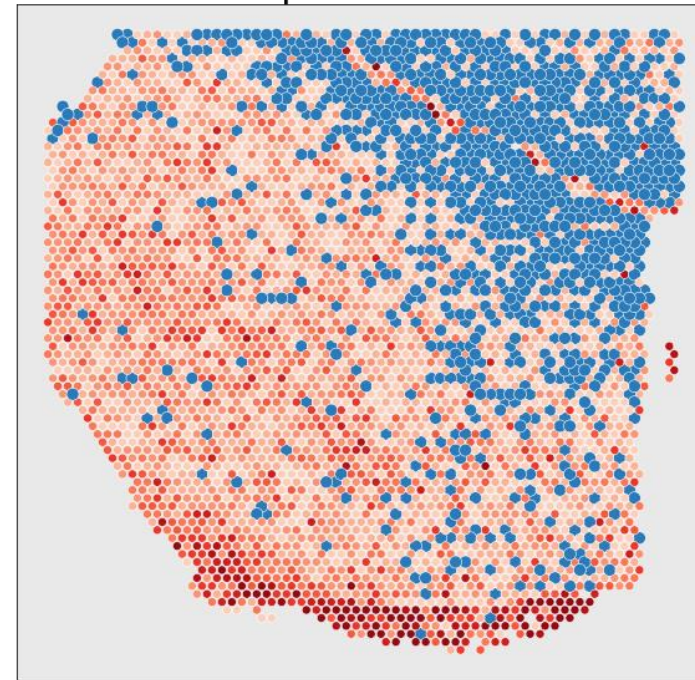
- How many cells?
- What size are the cells?



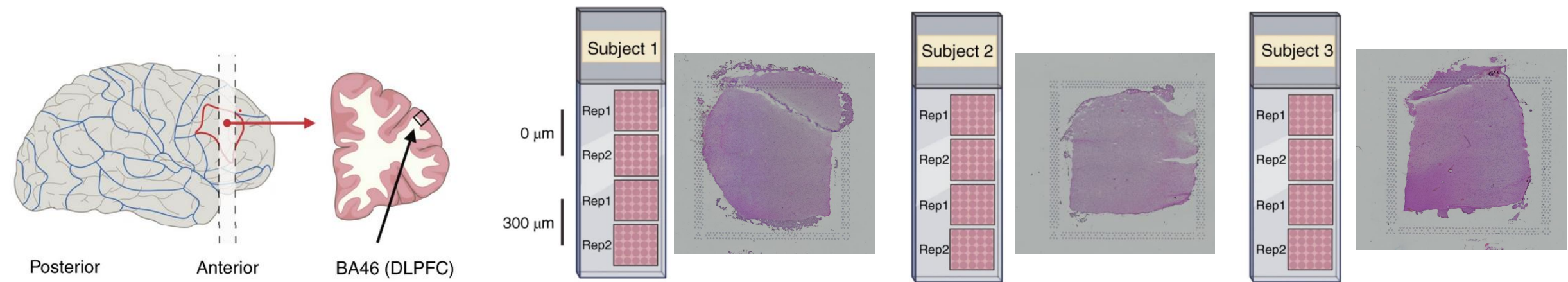
Zoomed-in view



Cell count per spot
sample #151507



Data Source



2831 genes
after filtering

~4000 spots
per sample

	Spot 1	Spot 2	...	Spot 3999	Spot 4000
Gene 1	0	0	1	0	1
Gene 2	0	3	0	0	0
...	4	7	0	11	3
Gene 2830	0	43	0	0	4
Gene 2831	0	0	1	1	0

x 6

Data Processing

Normalization Methods

1. Non-normalized (control)
2. Total count normalization
3. SCTransform
4. Dino (ZN & CK 2021)

Unsupervised clustering method

1. Filter genes and spots
2. PCA transform (50 PCs)
3. KMeans clustering on 50 PCs
num. clusters = 7 or 5

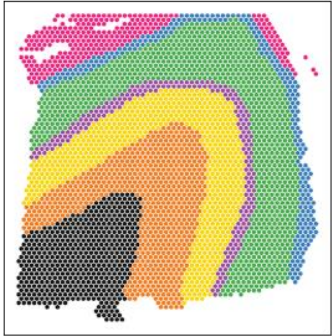
Scoring metric

Cluster assignment accuracy compared to Ground Truth

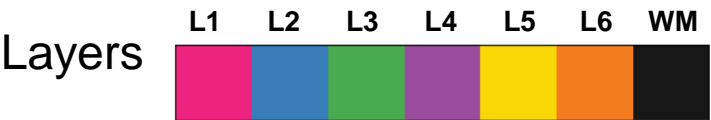
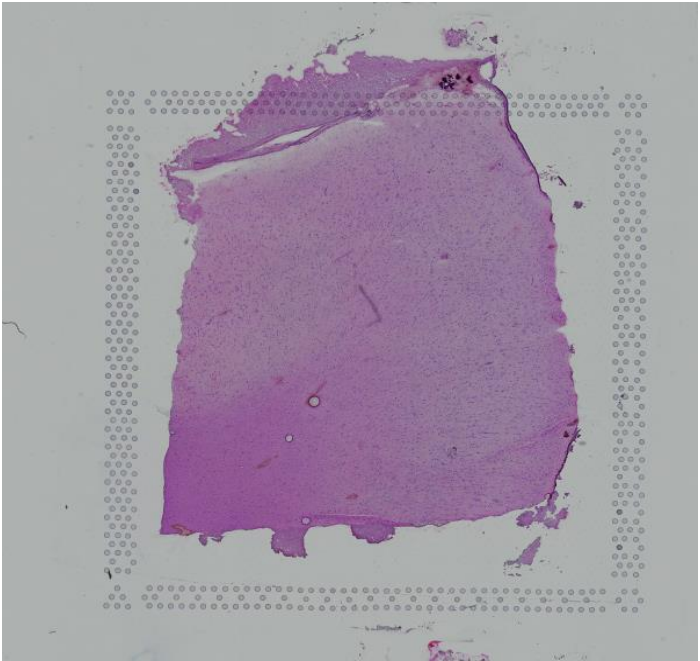
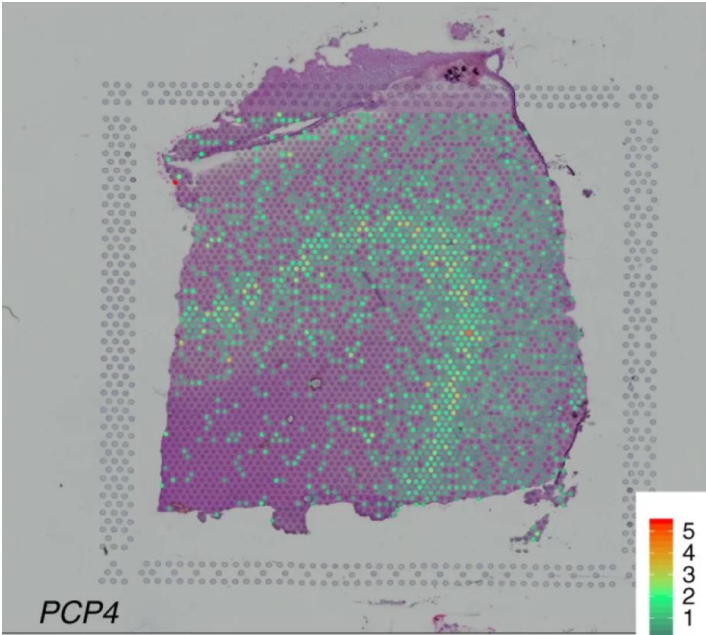
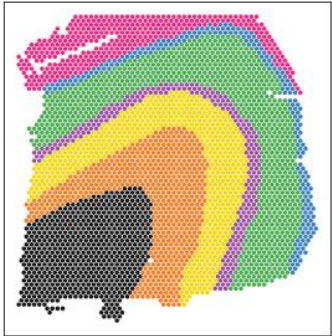
Results

Ground Truth

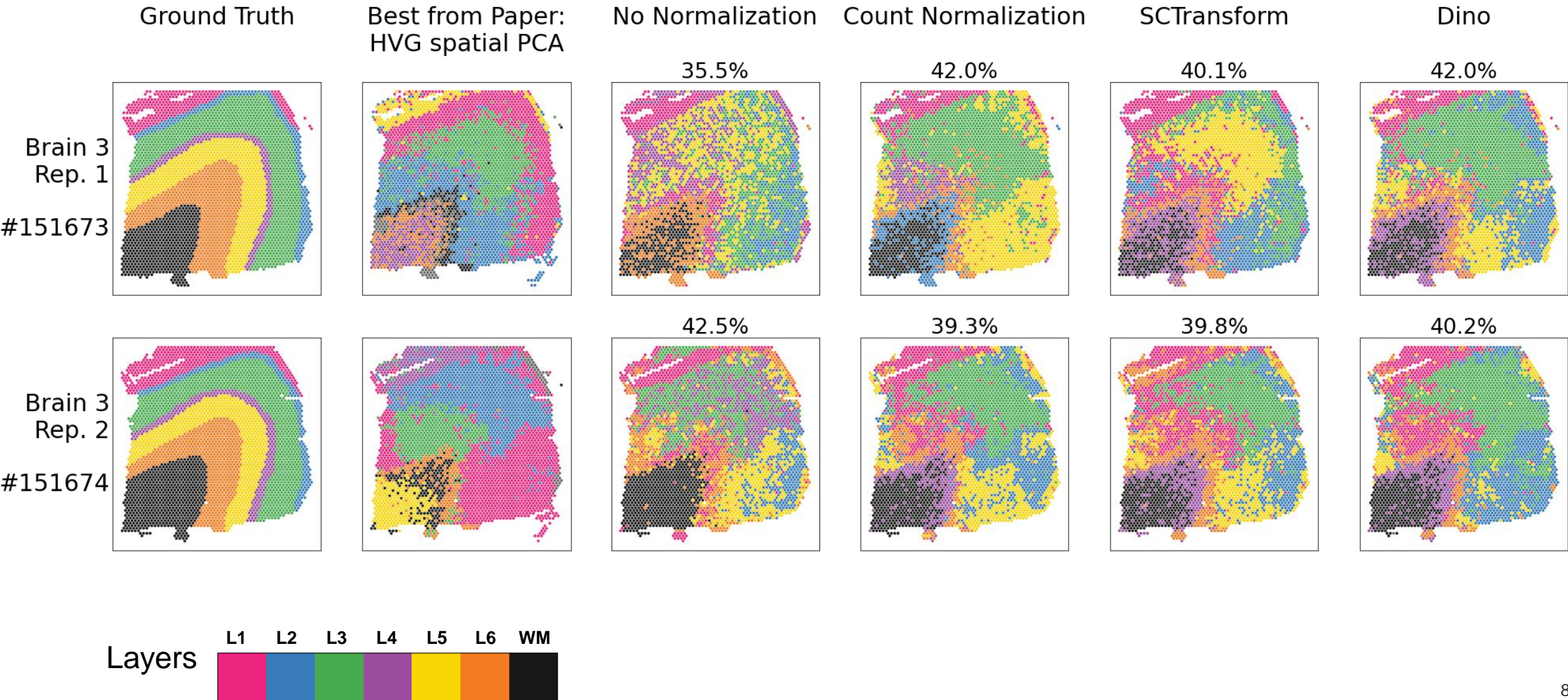
Brain 3
Rep. 1
#151673



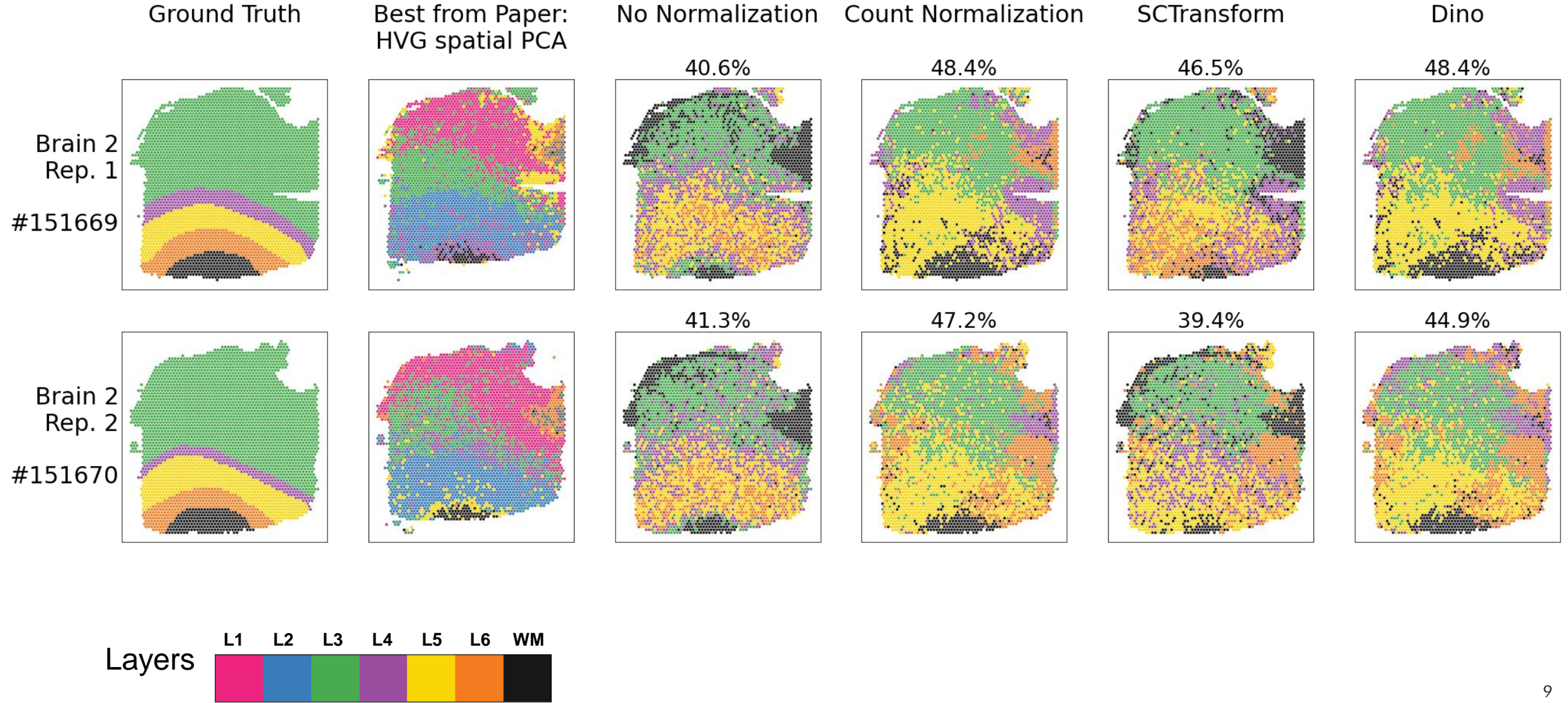
Brain 3
Rep. 2
#151674



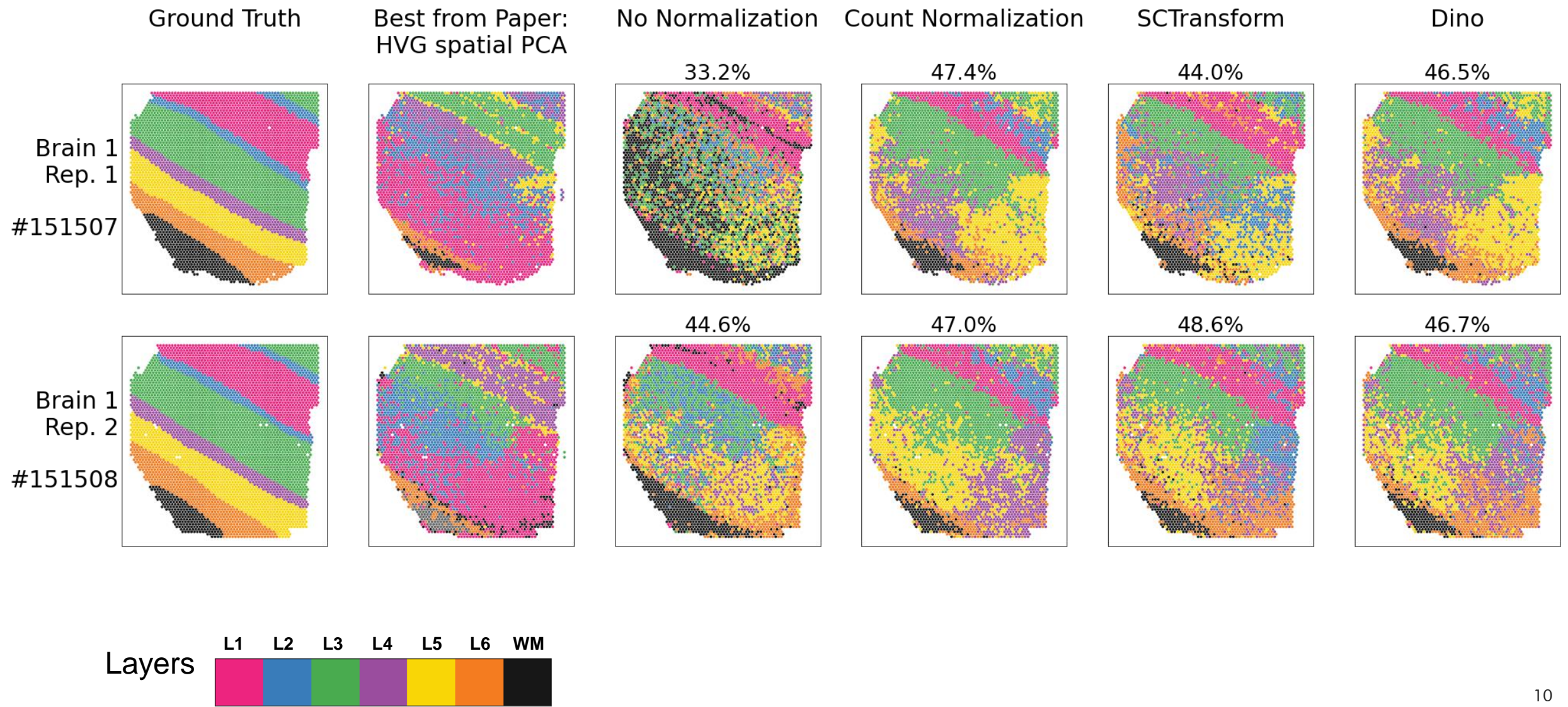
Results



Results



Results



Conclusion

Normalization tends to boost accuracy versus no normalization

% Accuracy

	Non-normalized	Count Norm.	SCTransform	Dino
Brain 1 rep 1	33.2	47.4	44.0	46.5
Brain 1 rep 2	44.6	47.0	48.6	46.7
Brain 2 rep 1	40.6	48.4	46.5	48.4
Brain 2 rep 2	41.3	47.2	39.4	44.9
Brain 3 rep 1	35.5	42.0	40.1	42.0
Brain 3 rep 2	42.5	39.3	39.8	40.2
Average	39.6	45.2	43	44.8

Normalization minimizes “spottiness” of cluster assignments

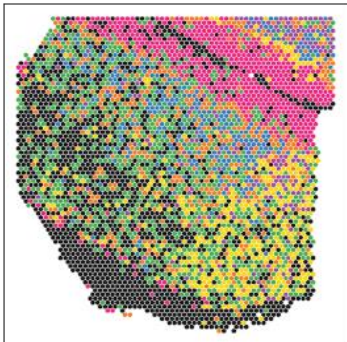
No Normalization

Count Normalization

SCTransform

Dino

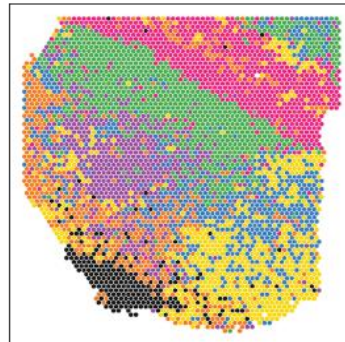
33.2%



47.4%



44.0%



46.5%



Analysis available at
github.com/benton-anderson/877