Supplementary Figures
Identification of spatial homogenous regions in tissues with concordex

Kayla C. Jackson[1,3], A. Sina Booeshaghi[2], Ángel Gálvez-Merchán[4], Lambda Moses[1], Tara Chari[1], Alexandra Kim[5], and Lior Pachter[1,6,*]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
[2]Department of Bioengineering, University of California, Berkeley, CA, USA
[3]Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[4]Cellarity, Boston, MA, USA
[5]Polytechnic High School, Pasadena, CA, USA
[6]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
[*]To whom correspondence should be addressed

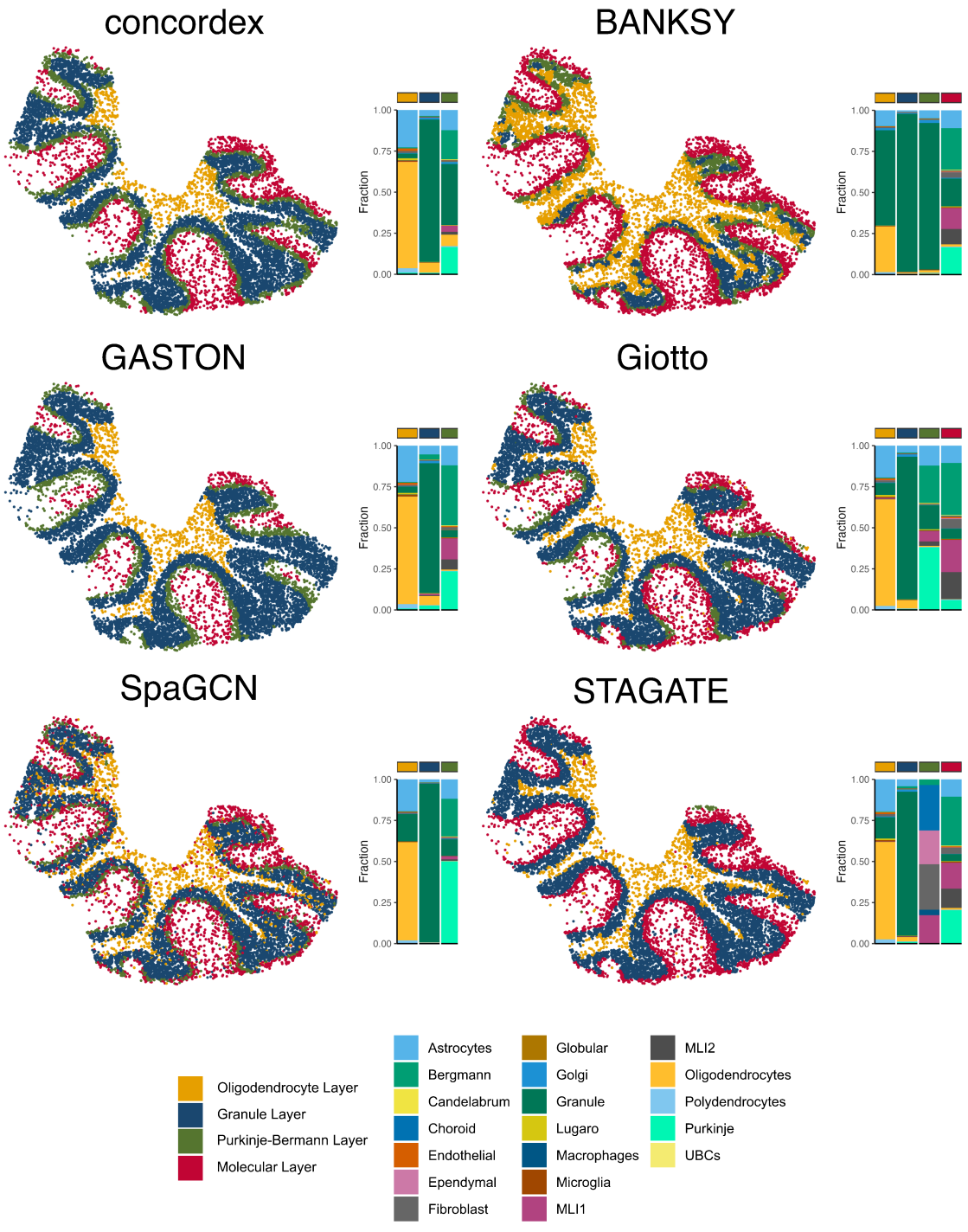

Fig. S1. Evaluation of concordex on mouse cerebellum. We compared A) concordex to B) BANKSY, C) GASTON, D) Giotto, E) SpaGCN, and F) STAGATE. The bar plots show the proportion of each cell type restricted to the SHRs identified by each method. While all methods broadly reconstruct the the cerebellar architecture, most methods failed to resolve the Purkinje-Bergmann Layer (B, D, E, F).

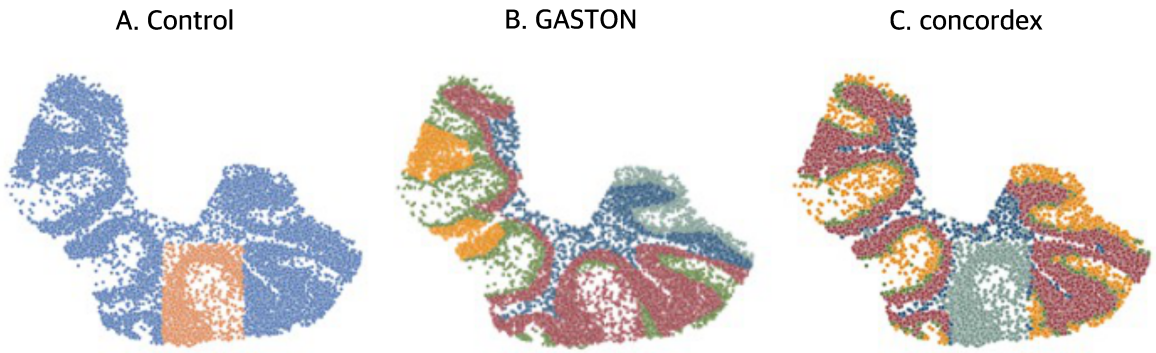

Fig. S2. Detection of anomalous region using GASTON. (A) shows the cells that were assigned a random expression profile. We tasked each method with identifying the disturbed region while and correctly assigning the remaining cells. (B) GASTON failed to identify the altered region and misassigned the remaining cells, while (C) concordex correctly identified the region and left the other assignments undisturbed.

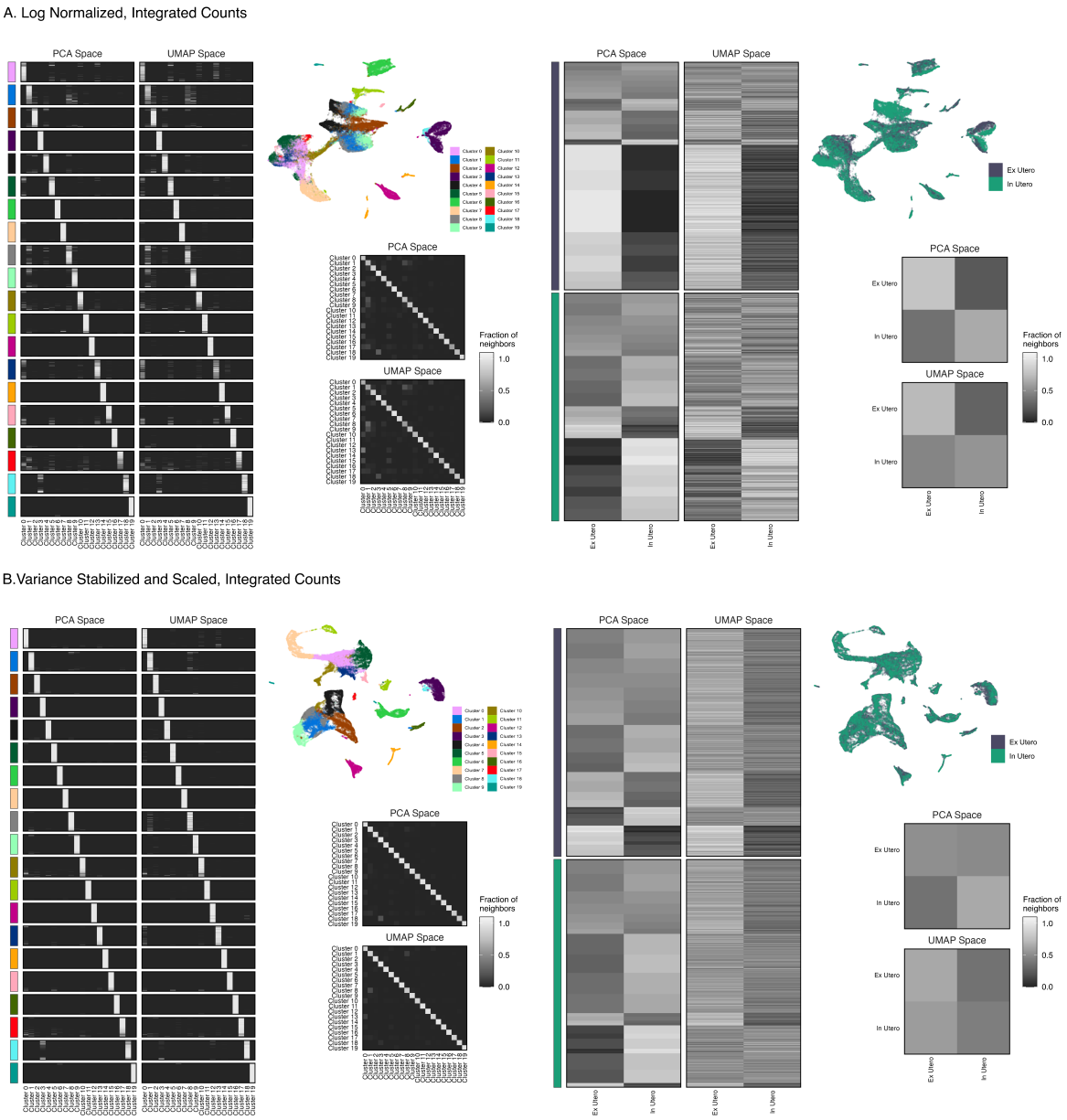

Fig. S3. Evaluation of concordex on mouse embryo. Left plots show the neighborhood consolidation matrices of the A) Log-normalized and integrated counts or B) Variance stabilized, scaled, and integrated counts with the UMAP embeddings and averaged similarity matrices next to it. In the neighborhood similarity matrix, each row is a cell and cells are grouped by their assigned cluster or batch. When the heatmaps are grouped by cluster assignment in the similarity matrices, the visualizations show that almost all clusters are well separated which contradicts visual inspection of the UMAP embedding. Similarly, visual inspection of the UMAP colored by growth condition suggests that the data are effectively integrated, however, the heatmaps show that most cells have neighborhoods that are exclusive of one growth condition.

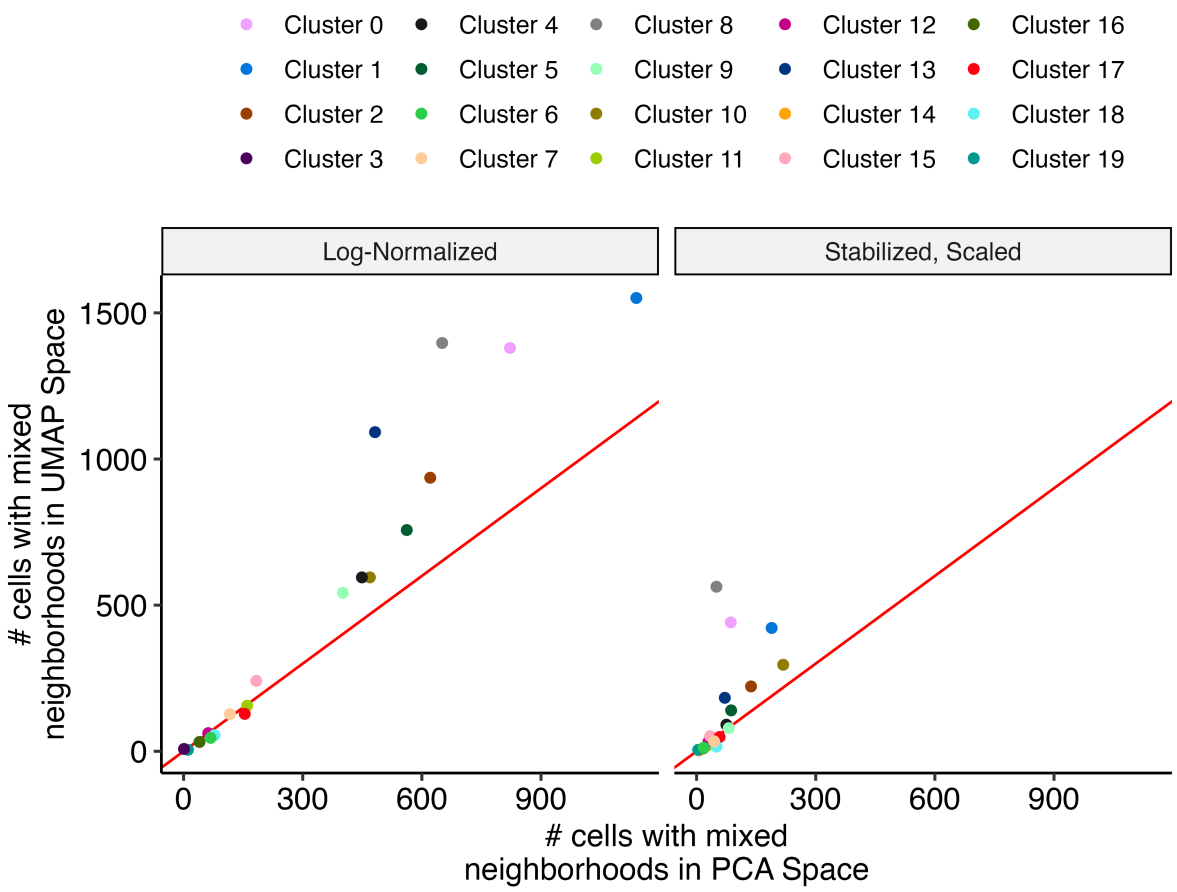

Fig. S4. Neighborhood mixing in reduced dimensions. Cells with majority mixed neighborhoods were identified for each cluster. A cell was called mixed if more than half of its nearest neighbors belong to a different cluster. For almost all clusters, the cells appeared to be more mixed in the UMAP space compared to the PCA space. The red line is the line y=x.