Diversity in transcriptomics without cell types

Leyi Jiang[1,2], Katherine Benjamin[1,2,3], Jesse V Veenvliet[2], Emily Roff[4] and Heather A Harrington[1,2,3,5]

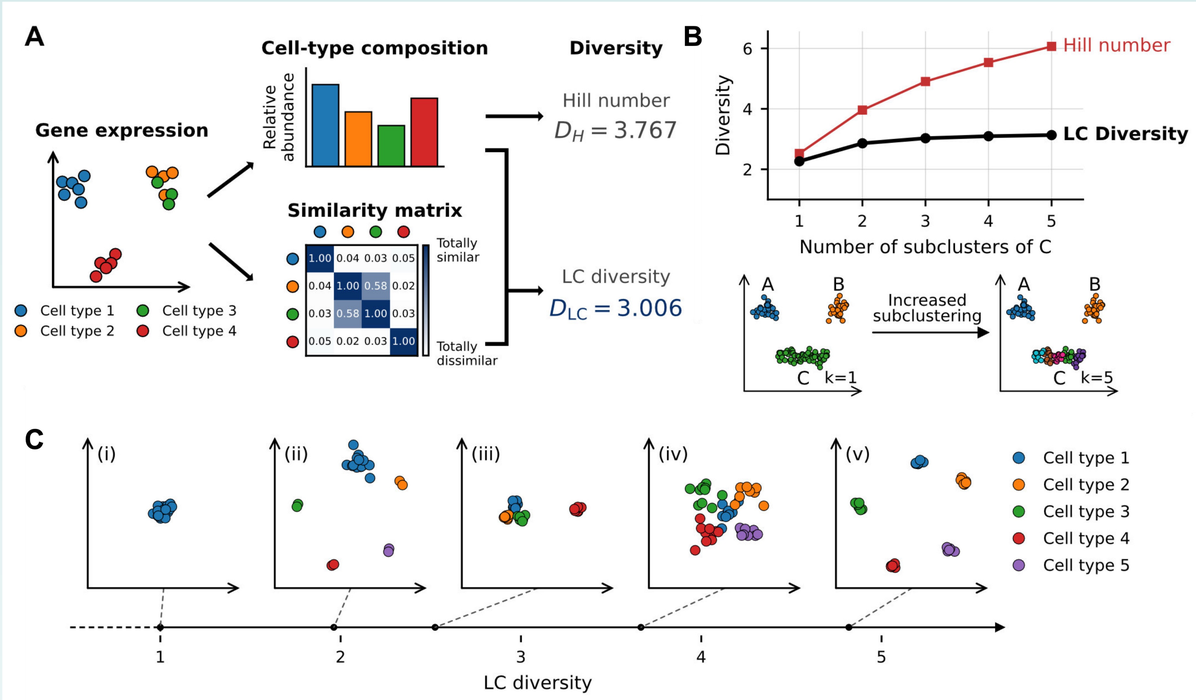
[1]CSBD, [2]MPI-CBG, [3]University of Oxford, [4]University of Edinburgh, [5]TU Dresden

The Challenge: The Non-Canonicity of Cell Types

Cellular diversity metrics usually rely on upstream cell-type clustering. However, cluster boundaries are highly non-canonical. Changing software versions (e.g., Seurat vs. Scanpy) or clustering resolutions changes downstream biological conclusions.

Our result: A similarity-sensitive measure of cellular diversity, inspired by ideas in ecological science[2], which is robust to the choice of cell type assignment and can even be used in the absence of any such assignment.

The Method: Leinster-Cobbold (LC) Diversity

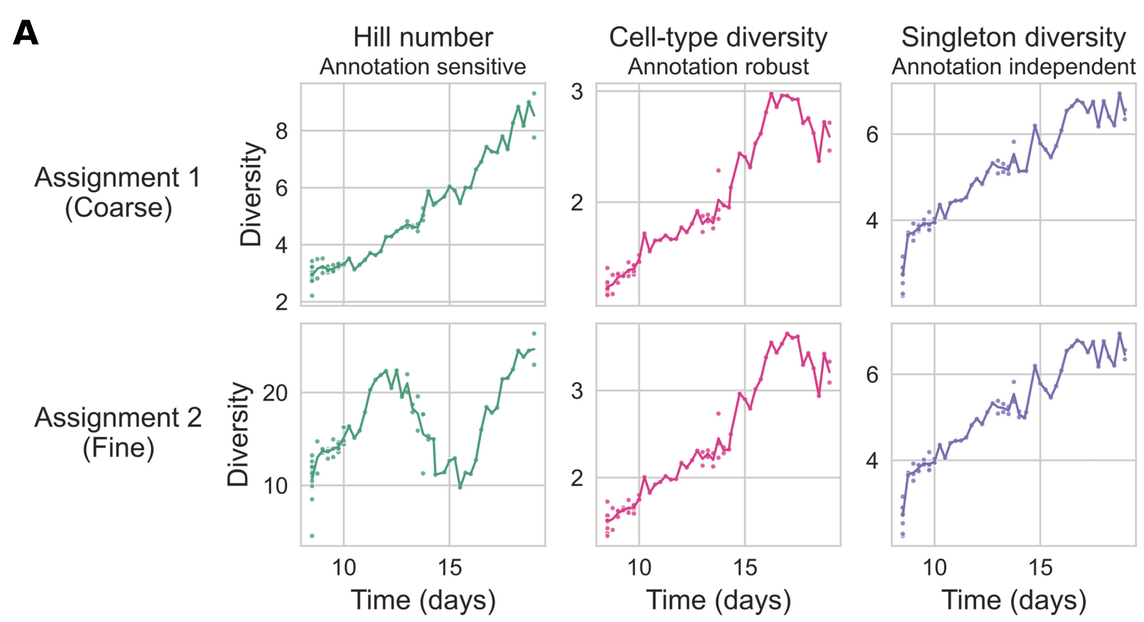


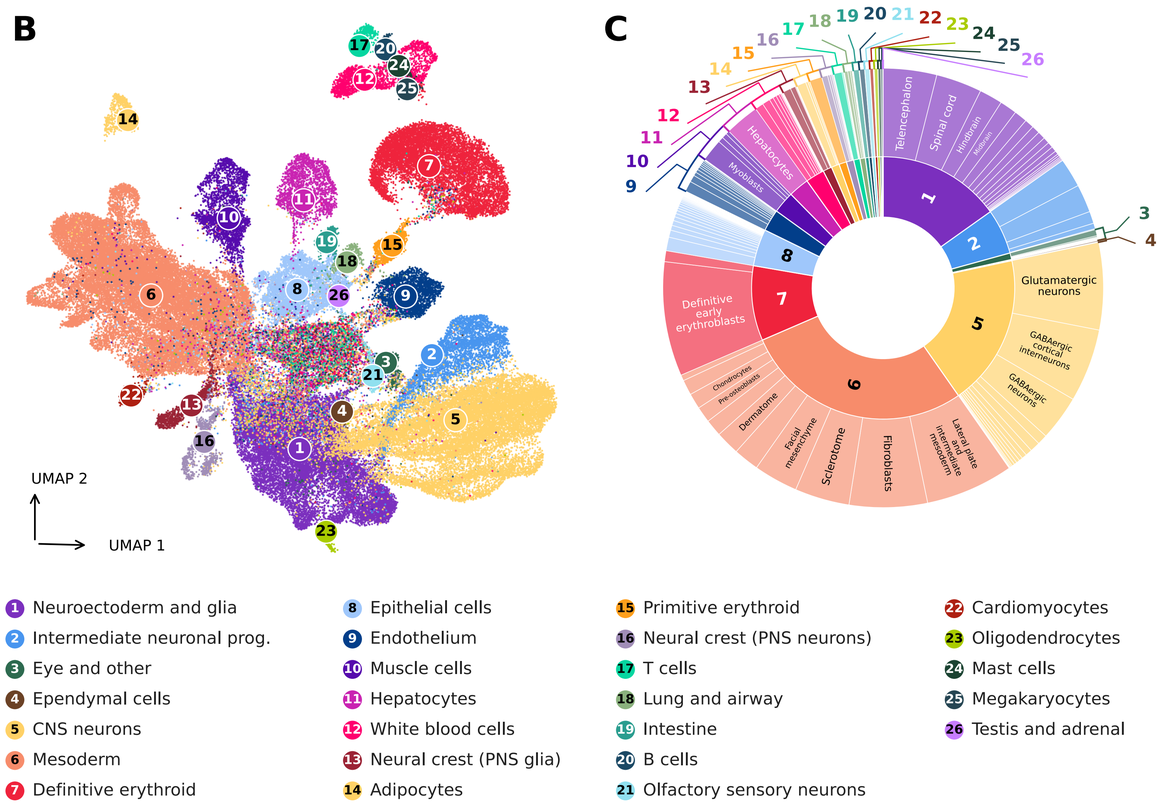
Instead of using abundance-only metrics like the Hill number or Shannon entropy, we calculate diversity by factoring in a pairwise cell similarity matrix $\mathbf{Z}$:

$$D_{\mathrm{LC}}(p;\, Z) = \frac{1}{\sum_{i=1}^{N} p_i (ZP)_i}$$

- p_i: Relative abundance of cell type i.
- Z_{ij}: Cosine similarity between average gene expression vectors.

Why it works: When a cluster is subdivided, the high similarity (Z_{ij} close to 1) between the subclusters prevents the diversity score from artificially inflating.





Application I - Time-Series Single-Cell Atlas

Mapping Mouse Embryogenesis

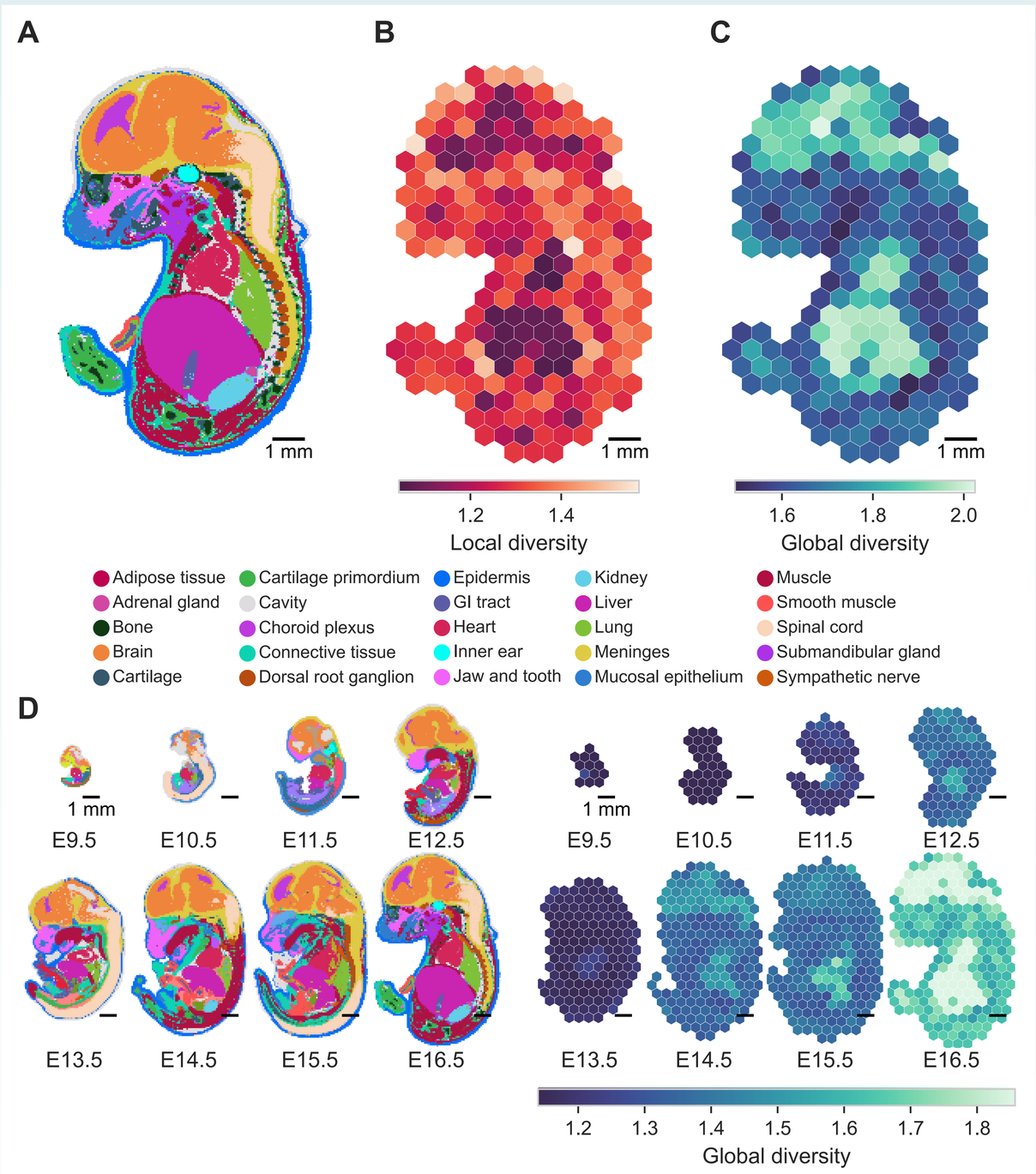
We applied our framework to a single-cell time-lapse atlas containing 11.4 million nuclei across 43 prenatal time points[3].

- Annotation Independence: Swapping from a coarse assignment (26 types) to a fine assignment (190 types) alters the trajectory shape of traditional Hill numbers. LC Diversity tracks perfectly across both.
- Singleton Mode: In singleton diversity, every single cell is treated as its own type, bypassing the clustering pipeline.
- Developmental Trend: Diversity rises near-linearly post-gastrulation (E8.5) and stabilises at E16.0, indicating that the primary cell state repertoire is established before tissue maturation.

Application II – Spatial Transcriptomics

For spatial transcriptomics, we're interested in the diversity of subregions of some larger spatial data set. Hence, we partitioned diversity into local and global components[4]:

- Local Diversity: Measures the heterogeneity of a distinct subregion. High scores highlight structural interfaces where distinct lineages mix.
- Global Diversity: Measures a region's unique contribution to the whole embryo.



Spatial Dynamics in Organogenesis

Using the MOSTA Stereo-seq dataset (E9.5 to E16.5)[1], we observed that the spatial diversity:

- The Developing Liver: Displays low local diversity (internally highly homogeneous) but exceptionally high global diversity. This diversity profile uniquely captures its phylogenetic divergence as an isolated endodermal branch.
- Spatiotemporal Onset: Global diversity maps remain uniform early (E9.5) but break into distinct, high-diversity organ patches by E12.5–E16.5, tracking terminal functionalisation.

Adopt scdiv in 3 Lines of Code

```
import scdiv

scdiv.tl.diversity(adata, order=1, cell_type_key="cell_type")
scdiv.spatial.diversity(
    adata, order=1,
    partition_kwargs={"method": "hex", "region_size": 100},
    cell_type_key="cell_type", mode="alpha",
)
```

Scan the QR code for the Python package

Citations:

[1] Chen A, Liao S, Cheng M et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell 2022;185(10):1777-1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.

[2] Leinster T, Cobbold CA. Measuring diversity: the importance of species similarity. Ecology 2012;93(3):477–89. https://doi.org/10.1890/10-2402.1.

[3] Qiu C, Martin BK, Welsh IC et al. A single-cell time-lapse of mouse prenatal development from gastrula to birth. Nature 2024;626(8001):1084–93. https://doi.org/10.1038/s41586-024-07069-w.

[4] Reeve R, Leinster T, Cobbold CA et al. How to partition diversity, arXiv:1404.6520. Preprint, arXiv, 8 Dec. 2016. https://doi.org/10.48550/arXiv.1404.6520.