

BEYOND GLOBAL EMBEDDINGS: A NEW APPROACH TO SINGLE-CELL REPRESENTATIONS

Anonymous authors

Paper under double-blind review

ABSTRACT

Uncovering the latent structure of high-dimensional data is a fundamental challenge in single-cell analysis. While many methods seek to structure single-cell data, most rely on a single global embedding space, which can obscure fine-grained variation. Here, we introduce Connectorama, a locally adaptive framework that constructs neighborhoods by aggregating information across overlapping local patches, allowing similarity metrics to adapt to local covariance structures. Applying this approach to large single-cell RNA sequencing datasets, we recover biologically meaningful subpopulations that global methods fail to resolve, including distinct immune cell subsets and hepatocyte populations with specialized gene expression signatures. By reframing single-cell representation as an ensemble of local views rather than a single projection, Connectorama offers a powerful framework for studying cellular diversity at scale.

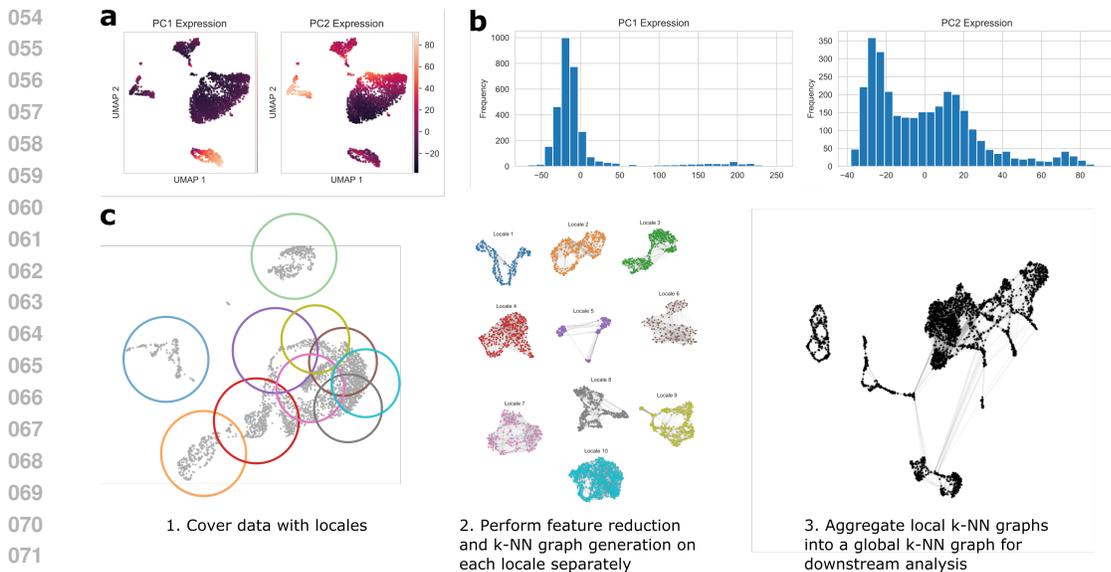
1 INTRODUCTION

Single-cell RNA sequencing (scRNA-seq) enables high-resolution characterization of cellular heterogeneity, driving discoveries across immunology, cancer, and complex diseases (Xiao et al., 2019; Ren et al., 2021; Azizi et al., 2018; Potter, 2018). A fundamental challenge in analyzing these datasets is defining cell-cell similarity, which underpins clustering, trajectory inference, and other downstream tasks. Most pipelines rely on a k -nearest-neighbor (k -NN) graph constructed after dimensionality reduction, but standard approaches assume that a single global embedding can faithfully capture biological variation.

To mitigate noise, methods such as PCA (Pearson, 1901), ICA (Comon, 1992), and SCA (DeMeo & Berger, 2023) reduce dimensionality before computing k -NN graphs. More recent deep learning-based approaches, including variational autoencoders (Lopez et al., 2018), contrastive learning (Lotfollahi et al., 2022), and transformer-based models (Cui et al., 2024), learn latent representations optimized for specific tasks. While these methods extract biologically meaningful features, they still impose a single global structure, which can obscure fine-grained subpopulations and context-dependent gene relationships.

We introduce Connectorama, a locally adaptive approach that constructs more accurate k -NN graphs by aggregating *local* dimensionality reductions. Instead of enforcing a single embedding, Connectorama partitions the dataset into overlapping *locales*, within which dimensionality reduction is performed separately. These locale-specific graphs are then stitched together using a principled topological aggregation method, preserving fine-grained relationships while mitigating distortions introduced by global projections.

By leveraging local structure, Connectorama provides greater resolution than global embeddings alone. We demonstrate its advantages in both synthetic and real datasets, showing improved detection of rare but biologically meaningful populations, including microglia, hepatocyte subtypes, and specialized immune cells. Because Connectorama refines the k -NN graph itself, it integrates seamlessly into existing analysis pipelines, providing a robust alternative for single-cell representation learning.



2 RESULTS

2.1 OVERVIEW OF THE METHOD

086
087
088
089
090

Connectorama consists of three main components: (1) the dataset is covered by a set of overlapping contiguous subsets called *locales*, ensuring each cell is included in at least C locales; (2) dimensionality reduction is performed *within* each locale, followed by the construction of locale-specific k -nearest-neighbor (k -NN) graphs; and (3) these local graphs are merged into a global k -NN graph through a principled topological aggregation strategy. We detail each step in Figure 1c and below.

091
092
093
094

Generating Locales: The key requirement in defining locales is that *every* cell is well-covered by at least one locale where its neighbors exhibit similar expression profiles. A well-constructed locale framework must also ensure robust representation of *rare* cell types, preventing their omission due to under-sampling.

095
096
097
098
099
100
101

To achieve this, we adapt the Hopper algorithm (DeMeo & Berger, 2020) to create a representative *sketch* of the dataset. This sketch consists of a set of subsampled points that collectively preserve the global structure of the data, including rare cell types. Locales are then constructed around these sketch points, with the constraint that each cell is included in at least C locales for some user-specified C . The result is an overlapping cover of the dataset that ensures all cells, including rare subpopulations, are included within a well-defined local structure. We describe our locale construction approach in detail in section section 4.

102
103
104
105
106
107

Local Dimensionality Reduction: Within each locale, we seek to identify the most informative features for distinguishing between local cells. This is achieved by applying dimensionality reduction *restricted* to the locale. Our framework is agnostic to the choice of dimensionality reduction technique, allowing flexibility based on dataset characteristics. However, linear projections are a natural choice for their simplicity, interpretability and widespread use. In our experiments, we evaluate Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Surprisal Component Analysis (SCA).

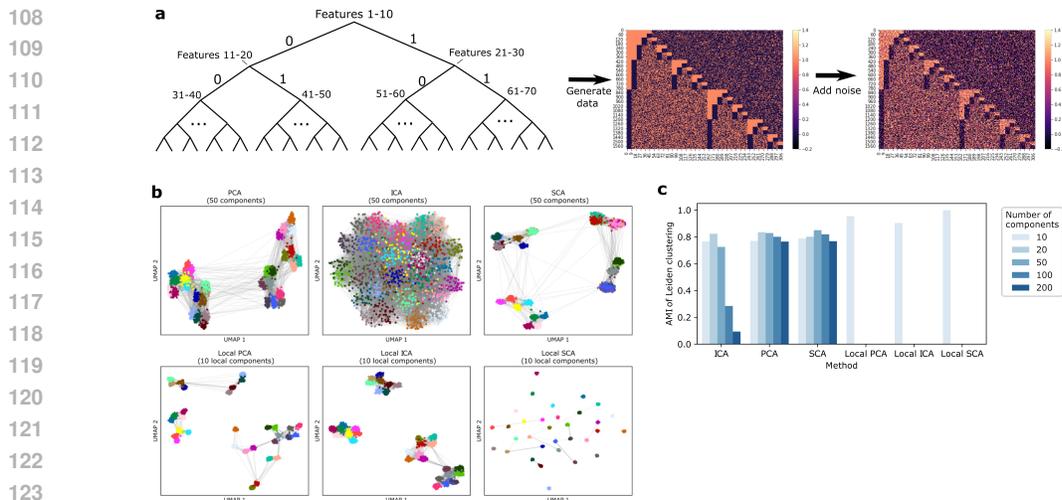


Figure 2: Results on synthetic, hierarchical data. **a**: Construction of the synthetic dataset. Clusters are related in a binary tree structure, with different sets of features defining each split. We then add noise to the resulting dataset to simulate dropouts and variable capture rate. **b**: UMAP plots with k -nearest-neighbor graphs computed downstream of global PCA, ICA, and SCA (top), and local version (bottom), colored according to cluster. Connectorama consistently produces better separation between clusters. **c**: Adjusted mutual information (AMI) of Leiden clusterings downstream of each method with the true cluster labels. Connectorama gives higher AMI than global dimensionality reductions, even with as many as 200 global components.

A key advantage of our localized approach is that even simple linear methods (e.g., PCA) perform well. Since each locale captures only a subset of cellular variation, the axes of variation within a locale are more biologically meaningful than those derived from a single global decomposition. In contrast, traditional global reductions force a single set of latent axes across the dataset, potentially discarding signals relevant to rare or context-dependent cell states. By analyzing locales separately, Connectorama preserves local gene relationships that may be masked in a global embedding space.

Once reduced representations are computed, we construct a k -NN graph within each locale using Euclidean distance in the reduced space. This ensures that similarity metrics are tailored to the local structure rather than being dictated by global characteristics.

Topological Stitching: After constructing locale-specific k -NN graphs, we integrate them into a single global k -NN graph. Since locales overlap, a given cell may have different neighbor sets across the locales to which it belongs. The challenge is to reconcile these local graphs into a cohesive structure that accurately reflects cellular relationships.

We achieve this through a *mean random walk with restart* (RWR) probability approach (section 4). Given two points i and j that co-occur in multiple locales, their similarity is computed as the mean RWR probability of reaching j from i across all shared locales. This approach ensures that points that are consistently well-connected in multiple local embeddings have high similarity in the final global graph.

By leveraging RWR-based aggregation rather than naive averaging of local distances, Connectorama avoids discontinuities and better preserves hierarchical structure. This allows for improved separation of fine-grained subpopulations, overcoming the limitations of global distance metrics that fail to capture local variation.

The final output of Connectorama is a refined global k -NN graph that integrates local structure while mitigating the distortions of a single global embedding space. This improved graph structure enhances the accuracy of downstream tasks, such as clustering, trajectory inference, and differential expression analysis.

2.2 CONNECTORAMA RESOLVES CLASSES IN SYNTHETIC DATA

To test Connectorama’s ability to resolve fine-grained structure, we created a synthetic dataset in which clusters are hierarchically arranged in a binary tree (Figure 2a). At each level, different sets of features define divisions, mimicking real biological heterogeneity. Gaussian noise and random dropout events were introduced to simulate realistic single-cell data.

Traditional global reductions struggle with this dataset because broad-classifying features introduce noise at deeper levels. We applied Connectorama with 20 locales (coverage of 3 per cell), using PCA, ICA, or SCA for local embedding, and RWR for aggregation.

Compared to global methods, Connectorama significantly improved cluster separability, as visualized in UMAP embeddings (Figure 2b). Adjusted Mutual Information (AMI) scores showed that Connectorama better preserved the hierarchy, achieving 0.953, 0.903, and 0.999 with local PCA, ICA, and SCA, respectively, compared to a maximum of 0.816 for global methods (Figure 2c). These results highlight how local structure enhances fine-grained resolution.

2.3 APPLICATION TO THE TABULA MURIS DATASET

The Tabula Muris Consortium compiled a comprehensive mouse single-cell dataset covering over 100,000 cells from 20 organs (Consortium et al., 2018). Their analysis applied PCA followed by Louvain clustering separately for each tissue, manually refining clusters and assigning cell types based on differentially expressed genes. Given the dataset’s diversity, we hypothesized that applying Connectorama jointly to all tissues would recover known populations in a single pass while potentially revealing novel subpopulations.

Data Processing and Locale Construction: We obtained raw counts from GEO accession GSE109774 and followed standard preprocessing, including transcript-per-million normalization and log transformation. To remove confounding effects of donor sex, we excluded five sex-linked genes (*XIST*, *TSIX*, *DDX3Y*, *EIF2S3Y*, and *UTY*).

Locales were constructed using Hopper on a 50-dimensional PCA embedding, ensuring each cell was assigned to at least three locales. Dimensionality reduction was performed within locales using PCA, ICA, or SCA, followed by 15-nearest-neighbor graphs constructed using Euclidean distance. These local graphs were aggregated into a global 15-nearest neighbor graph using mean random walk with restart (RWR) similarity with a restart probability of 1% (Methods). For comparison, we generated k -NN graphs using global PCA, ICA, or SCA embeddings with 50, 100, or 200 components. Leiden clustering (resolution 3.0) and UMAP visualization were applied to assess clustering resolution.

Improved Resolution of Rare and Specialized Cell Types: UMAP plots suggest that Connectorama provides a more granular representation of cellular identity, yielding stronger separation of annotated cell types compared to global approaches (Figure 3a). To quantify this, we evaluated whether Leiden clusters derived from different k -NN graphs successfully recovered known cell types. We computed F1 scores by identifying the Leiden clusters with maximal overlap with known cell annotations (Methods).

Connectorama improved the resolution of multiple rare cell types, including:

- Bergmann glial cells, a specialized class of astrocytes involved in cerebellar function.
- Myofibroblasts, contractile cells crucial for wound healing and fibrosis.
- Natural killer (NK) cells, innate lymphocytes with cytotoxic and immunoregulatory roles.
- Pancreatic D cells, a minor islet cell type secreting somatostatin, a regulator of insulin/glucagon secretion.

These populations were poorly resolved by global methods but emerged distinctly under Connectorama (Figure 3f), highlighting the power of local representations for enhancing rare cell detection.

Discovery of Microglial and Hepatocyte Subpopulations Beyond rare cell detection, Connectorama revealed finer-grained subpopulations within known lineages.

For example, microglia exhibited a previously unrecognized subset marked by *CLEC7A*, a receptor promoting immune activation, and *GAS2L3*, a cytoskeletal regulator implicated in mitotic stability

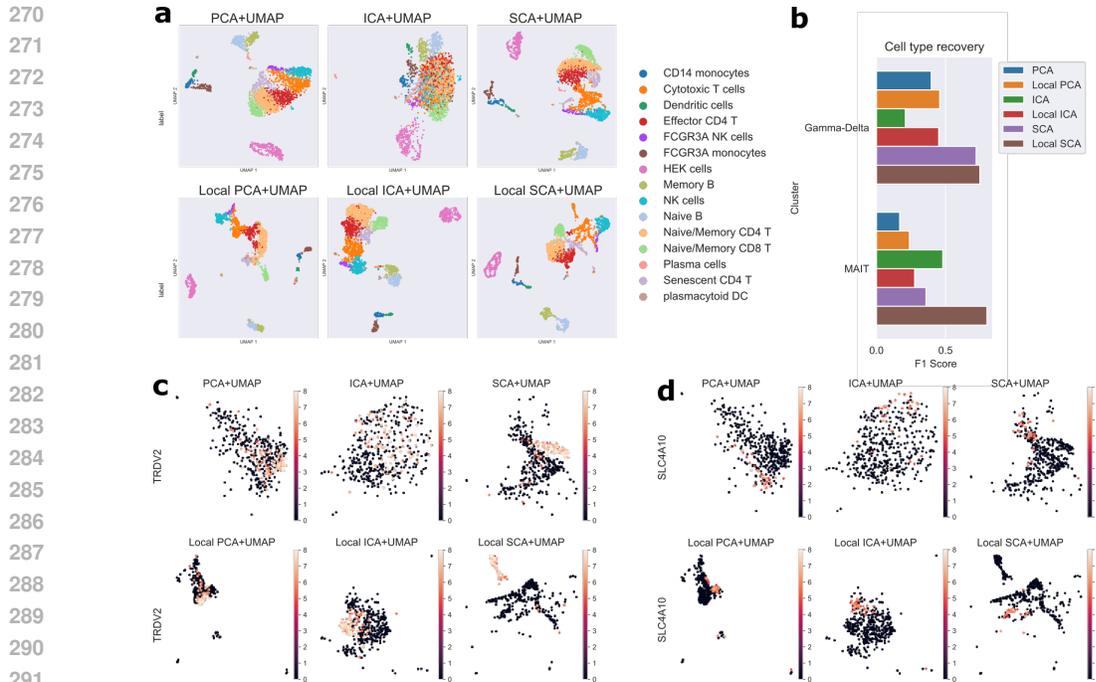


Figure 4: Running Connectorama on cellular populations from (Hagemann-Jensen et al., 2020). **a**: UMAP plots downstream of PCA, ICA, SCA, and their local versions computed with Connectorama. **b**: F1 scores for recovery of the gamma-delta and MAIT populations using Leiden clusterings downstream of each method. **c**: UMAP plots of the cytotoxic T subset of the data, colored by TRDV2, a marker gene for gamma-delta T-cells. **d**: UMAP plots of the cytotoxic T subset colored by SLC4A10, a MAIT marker gene.

(Figure 3b). These markers suggest an actively proliferating microglial state, possibly engaged in immune surveillance or neuroinflammation.

Within hepatocytes, Connectorama resolved distinct functional groups (Figure 3e):

- Cluster 116 lacked *CD302* but highly expressed excision repair genes (*ERCCs*), suggesting a DNA-damage repair state (McWhir et al., 1993).
- Cluster 119 exhibited strong cytochrome P450 (*CYP2B9*, *CYP2B13*) expression, indicating specialized roles in xenobiotic metabolism (Villeneuve & Pichette, 2004).
- Clusters 124 and 129 expressed cytochrome genes distinct from clusters 116 and 119, implying further functional heterogeneity (Figure 3c).

These hepatocyte and microglial subpopulations were not explicitly characterized in the original study, demonstrating how Connectorama enhances biological resolution.

Implications for Large-Scale Single-Cell Analysis: By analyzing all tissues simultaneously, Connectorama identified cross-tissue relationships that were missed when treating organs independently. For example, activated pancreatic stellate cells emerged within the broader myofibroblast niche, marked by *ACTA2* and *MYL9* expression (Apte et al., 1999). These cells, sparse in pancreatic samples alone, were clearly distinguishable in a multi-tissue setting, suggesting that joint analysis improves detection of shared functional states.

Overall, Connectorama refines cell-type classification by preserving local relationships lost in global embeddings, making it a powerful tool for large-scale, heterogeneous single-cell datasets.

2.4 APPLICATION TO HUMAN IMMUNE CELL PROFILING

The human immune system consists of highly specialized cell types that coordinate responses to pathogens and disease (Cillo et al., 2020; Ren et al., 2021). Immunotherapies increasingly target immune subtypes to treat conditions ranging from cancer to autoimmune disorders (Azizi et al., 2018; Ren et al., 2021; Xiao et al., 2019). However, single-cell techniques often struggle to resolve subtle distinctions among immune subtypes, particularly in large datasets (Potter, 2018). We hypothesized that Connectorama could improve immune cell discrimination by adapting to local transcriptional variation rather than imposing a single global embedding.

Enhanced Resolution of Immune Subpopulations: As before, we computed global PCA, ICA, and SCA embeddings with 10–200 components, constructing k -NN graphs using Euclidean distance. In parallel, we applied Connectorama using 30 locales, ensuring each cell belonged to at least three locales. Local 15-nearest-neighbor graphs were computed in 10-dimensional PCA, ICA, or SCA space and aggregated via mean random walk with restart (RWR) similarity (restart probability 0.1). UMAP plots downstream of each network-construction strategy are shown in Figure 4a.

Compared to global embeddings, Connectorama improved separation between immune subtypes, particularly within the cytotoxic T-cell compartment. Further examination revealed that these substructures corresponded to gamma-delta T-cells and mucosal-associated invariant T (MAIT) cells, functionally distinct but closely related populations. Gamma-delta T-cells are an evolutionarily conserved T-cell lineage involved in rapid responses to stressed or infected cells. These cells were distinguished by expression of the delta T-cell receptor gene *TRDV2* (Figure 4c). **MAIT cells** specialize in antimicrobial immunity by detecting microbial metabolites and play key roles in mucosal defense (Le Bourhis et al., 2010). These were identified by strong expression of *SLC4A10* (Figure 4d).

These subsets were visually indistinct in global embeddings but emerged clearly under Connectorama, indicating that local graph construction better preserves fine-grained immune variation.

Quantitative Assessment of Immune Subtype Recovery: To quantify improvements, we applied Leiden clustering (resolution 2.0) and evaluated F1 scores for recovering gamma-delta and MAIT cells. We annotated cells expressing at least two of *TRDV2*, *TRGV9*, and *TRDC* as gamma-delta T-cells; and cells expressing either *SLC4A10* or *LTK* as MAIT cells.

Connectorama combined with SCA yielded the highest F1 scores, significantly outperforming global methods (Figure 4b). These results reinforce that Connectorama effectively preserves immune heterogeneity, making it a valuable tool for refining immune cell atlases.

3 DISCUSSION

Extracting meaningful biological signal from high-dimensional data is a central challenge in modern biological analysis. k -nearest-neighbor (k -NN) graphs provide a concise summary of cellular relationships and serve as the foundation for clustering, integration, and trajectory inference. However, standard methods construct these graphs using global embeddings, which impose rigid geometric assumptions that can obscure fine-grained biological variation.

Connectorama reframes this problem by leveraging local embeddings, sidestepping the need to embed an entire dataset into a single embedding space. This approach aligns with the fundamental principle that while global embeddings may distort underlying relationships, local patches remain well-approximated in Euclidean space. By aggregating local structure, Connectorama achieves both fine-grained resolution and global coherence, offering a robust alternative for structuring high-dimensional biological data.

Beyond its practical advantages, Connectorama bridges two perspectives in high-dimensional data analysis: the *geometric* view, which assumes an underlying continuous manifold, and the *topological* view, which emphasizes point-wise connectivity. Traditional methods define k -NN graphs directly from a single global embedding, whereas Connectorama constructs them as an ensemble of locally adaptive representations. This removes the constraints imposed by global coordinate systems and better reflects the natural complexity of biological data.

Because k -NN graphs underpin a wide range of downstream bioinformatics algorithms—including clustering (Traag et al., 2019), integration (Hie et al., 2019), and trajectory inference (Moon et al.,

2019; Haghverdi et al., 2016)—improving their fidelity can have far-reaching implications. As single-cell datasets continue to grow in scale and complexity, locally adaptive representations such as Connectorama may provide a critical foundation for high-resolution biological discovery.

4 METHODS

Mathematical Setup: We represent an scRNA-seq dataset as an $N \times G$ matrix X , where each cell i is a G -dimensional transcript count vector $x_i \in \mathbb{R}^G$. A k -nearest neighbor (k -NN) graph $G_X^{(k)}$ is constructed by linking each cell to its k closest neighbors according to a chosen distance function $d_X(i, j)$. Our goal is to construct a high-fidelity k -NN graph by learning an improved distance metric that better preserves local structure.

Locale Construction: A key challenge in defining similarity in high-dimensional space is that gene co-expression patterns are context-dependent. Instead of applying a single global metric, Connectorama partitions the dataset into overlapping *locales*, where each cell belongs to multiple neighborhoods. Unlike clusters, which impose hard partitions, locales provide redundant coverage, ensuring each cell is well-represented in at least one appropriate context.

To construct locales, we employ a modified version of the Hopper algorithm (DeMeo & Berger, 2020) adapted for graph distances (Graph-Hopper). Hopper selects a representative subsample (*sketch*) of M points that minimizes the Hausdorff distance to the dataset, ensuring rare cell types are well-covered. Each cell is then assigned to its C closest locales based on a random walk with restart (RWR) similarity measure over a preliminary PCA-based k -NN graph. This strategy ensures that locales reflect the dataset’s intrinsic structure rather than being arbitrarily defined.

Local Dimensionality Reduction: Within each locale, we perform dimensionality reduction to identify the most informative features for distinguishing local cell states. By restricting the decomposition to local neighborhoods, we capture context-specific variation that global methods overlook.

We evaluate Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Surprisal Component Analysis (SCA) (DeMeo & Berger, 2023) for this step. While PCA is widely used for noise reduction, it prioritizes variance over biological relevance, which can obscure rare subpopulations. SCA instead optimizes an information-theoretic objective to emphasize marker genes over noise. In the PCA-based version of Connectorama, the distance between cells i and j in locale m is defined as:

$$d_m(i, j) = \|P_m x_i - P_m x_j\|,$$

where P_m is the learned projection matrix for the locale. Each locale-specific distance metric induces a k -NN graph, which we denote G_m .

Stitching Local Graphs into a Global Graph: The final step is integrating local k -NN graphs into a single global structure. Since each cell belongs to multiple locales, it may have different neighbor sets across graphs. To reconcile these differences, we use mean random walk with restart (RWR) similarity, which measures connectivity between nodes by simulating a biased diffusion process.

For cells i and j appearing in c locales with corresponding graphs G_{m_1}, \dots, G_{m_c} , the similarity score $s(i, j)$ is computed as:

$$s(i, j) := \frac{1}{c} \sum_{k=1}^c p_{RWR}(i, j; G_{m_k}), \quad (1)$$

where $p_{RWR}(i, j; G_{m_k})$ is the steady-state probability of reaching j from i in locale graph G_{m_k} . The global k -NN graph is then constructed by connecting each cell to its k most similar neighbors according to $s(i, j)$:

$$N_i^{(k)} = \{x_j : \text{rank}(s(i, j)) \leq k\}. \quad (2)$$

This final k -NN graph forms the output of Connectorama, preserving both local structure and global coherence without assuming a single global embedding space.

432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485

MEANINGFULNESS STATEMENT

In current practice, a “meaningful representation” of data often means a low-dimensional embedding of the data that captures its salient features. Here, we show that this *geometric* approach is inherently limited, and introduce Connectorama, a novel method that constructs a global *topological* representation by stitching together local embeddings. By adapting to context-dependent gene expression, Connectorama better captures rare cell types and uncovers novel biology from atlas-scale scRNA-seq data, moving beyond static embeddings to a more faithful representation of cellular identity and function.

REFERENCES

- MV Apte, PS Haber, SJ Darby, SC Rodgers, GW McCaughan, MA Korsten, RC Pirola, and JS Wilson. Pancreatic stellate cells are activated by proinflammatory cytokines: implications for pancreatic fibrogenesis. *Gut*, 44(4):534–541, 1999.
- Elham Azizi, Ambrose J Carr, George Plitas, Andrew E Cornish, Catherine Konopacki, Sandhya Prabhakaran, Juozas Nainys, Kenmin Wu, Vaidotas Kisieliovas, Manu Setty, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell*, 174(5):1293–1308, 2018.
- Anthony R Cillo, Cornelius HL Kürten, Tracy Tabib, Zengbiao Qi, Sayali Onkar, Ting Wang, Angen Liu, Umamaheswar Duvvuri, Seungwon Kim, Ryan J Soose, et al. Immune landscape of viral-and carcinogen-driven head and neck cancer. *Immunity*, 52(1):183–199, 2020.
- Pierre Comon. Independent component analysis, 1992.
- Tabula Muris Consortium et al. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, 562(7727):367–372, 2018.
- Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, pp. 1–11, 2024.
- Benjamin DeMeo and Bonnie Berger. Hopper: a mathematically optimal algorithm for sketching biological data. *Bioinformatics*, 36(Supplement_1):i236–i241, 2020.
- Benjamin DeMeo and Bonnie Berger. Sca: recovering single-cell heterogeneity through information-based dimensionality reduction. *Genome Biology*, 24(1):195, 2023.
- Michael Hagemann-Jensen, Christoph Ziegenhain, Ping Chen, Daniel Ramsköld, Gert-Jan Hendriks, Anton JM Larsson, Omid R Faridani, and Rickard Sandberg. Single-cell rna counting at allele and isoform resolution using smart-seq3. *Nature Biotechnology*, 38(6):708–714, 2020.
- Laleh Haghverdi, Maren Büttner, F Alexander Wolf, Florian Buettner, and Fabian J Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods*, 13(10):845, 2016.
- Brian Hie, Bryan Bryson, and Bonnie Berger. Efficient integration of heterogeneous single-cell transcriptomes using scanorama. *Nature biotechnology*, 37(6):685–691, 2019.
- Lionel Le Bourhis, Emmanuel Martin, Isabelle Péguillet, Amélie Guihot, Nathalie Froux, Maxime Coré, Eva Lévy, Mathilde Dusseaux, Vanina Meyssonier, Virginie Premel, et al. Antimicrobial activity of mucosal-associated invariant t cells. *Nature immunology*, 11(8):701–708, 2010.
- Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.
- Mohammad Lotfollahi, Mohsen Naghipourfar, Malte D Luecken, Matin Khajavi, Maren Büttner, Marco Wagenstetter, Žiga Avsec, Adam Gayoso, Nir Yosef, Marta Interlandi, et al. Mapping single-cell data to reference atlases by transfer learning. *Nature biotechnology*, 40(1):121–130, 2022.

486 Jim McWhir, Jim Selfridge, David J Harrison, Shoshana Squires, and David W Melton. Mice with
487 dna repair gene (ercc-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die
488 before weaning. *Nature genetics*, 5(3):217–224, 1993.

489
490 Kevin R Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B Burkhardt, William S Chen,
491 Kristina Yim, Antonia van den Elzen, Matthew J Hirn, Ronald R Coifman, et al. Visualizing
492 structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):
493 1482–1492, 2019.

494 Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London,
495 Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

496
497 S Steven Potter. Single-cell rna sequencing for the study of development, physiology and disease.
498 *Nature Reviews Nephrology*, 14(8):479–492, 2018.

499 Xianwen Ren, Lei Zhang, Yuanyuan Zhang, Ziyi Li, Nathan Siemers, and Zemin Zhang. Insights
500 gained from single-cell analysis of immune cells in the tumor microenvironment. *Annual Review
501 of Immunology*, 39:583–609, 2021.

502
503 Nicholas Schaum, Jim Karkanias, Norma F Neff, Andrew P May, Stephen R Quake, Tony Wyss-
504 Coray, Spyros Darmanis, Joshua Batson, Olga Botvinnik, Michelle B Chen, et al. Single-cell
505 transcriptomics of 20 mouse organs creates a tabula muris: The tabula muris consortium. *Nature*,
506 562(7727):367, 2018.

507 Vincent A Traag, Ludo Waltman, and Nees Jan van Eck. From louvain to leiden: guaranteeing
508 well-connected communities. *Scientific reports*, 9(1):1–12, 2019.

509
510 J-P Villeneuve and V Pichette. Cytochrome p450 and liver diseases. *Current drug metabolism*, 5(3):
511 273–282, 2004.

512 Zhengtao Xiao, Ziwei Dai, and Jason W Locasale. Metabolic landscape of the tumor microenviron-
513 ment at single cell resolution. *Nature communications*, 10(1):1–12, 2019.

514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539