

Integrating morphology and gene expression of neural cells in unpaired single-cell data using GeoAdvAE

Jinqui Turbo Du¹ Kevin Z Lin¹

University of Washington, Seattle WA 98107, USA, {turbodu,kzlin}@uw.edu,

Abstract. Cellular morphological transitions are widely observed in many diseases; however, the functional role of these morphologies remains unclear, as most technologies are unable to profile both form and function simultaneously. However, computationally linking single-cell morphology and transcriptomics is challenging due to a lack of feature correspondences. We present GeoAdvAE, a geometry-aware adversarial autoencoder for diagonal (unpaired) integration of single-cell morphology and single-cell RNA sequencing. GeoAdvAE combines modality-specific variational autoencoders with a Gromov-Wasserstein regularizer and an adversarial discriminator to embed unpaired morphologies and transcriptomes into a shared latent space, preserving both reconstruction fidelity and cross-modal geometry. To validate the correctness of integration, we leverage Patch-seq neurons with joint morphology-RNA measurements. Using these ground-truth pairings, GeoAdvAE achieves the best cross-modal cell-type matching accuracy compared to other diagonal integration methods, outperforming optimal transport, latent alignment, and adversarial baselines. We then apply GeoAdvAE to microglia from the 5xFAD mouse model, a model system of Alzheimer’s disease. We integrate 98 CAJAL-quantified morphologies, spanning amoeboid and ramified forms, with 31,948 single-cell RNA-seq profiles across homeostatic, proliferating, and disease-associated states to recover a one-dimensional axis that aligns the two modalities. We uncover novel biology by using integrated gradient attribution, where we highlight transcriptomic shifts (DNA repair in ramified; cell killing in amoeboid) and nominate gene markers (Ms4a6b; Ft1/Fth1) corresponding with morphological changes. Our integration also enables us to identify DAM signatures that do not correspond to morphological changes. GeoAdvAE provides a scalable and interpretable approach to connecting cellular ‘form’ and ‘function’ when joint profiling of morphology and transcriptomics is impractical. Our method is open-source, available at <https://github.com/turbodu222/GeoAdVAE>.

[[KZL: We should highlight neuron and microglia morphologies in title and abstract]] [[TD: Can we still change the title and abstract?]]

Keywords: Alzheimer’s disease, Cross-modal alignment, Deep learning, Diagonal integration, Patch-seq neurons, Microglia

1 Introduction

“Form follows function” is a powerful organizing principle in biology: a cell’s shape is often a visible readout of what it is doing. In the brain, for example, neurons in different layers exhibit distinct arborization patterns that enable specialized circuit roles. Microglia likewise remodel their processes as they surveil tissue, respond to damage, and interact with pathology. During Alzheimer’s disease (AD), these morphological programs shift alongside changes in immune function. Yet morphology is not a perfect proxy for cellular function. Similar shapes can conceal divergent molecular states, and conversely, functional reprogramming does not always manifest as a clear change in shape. This gap complicates efforts to understand how microglia respond to disease burden or therapy and limits our ability to translate visible phenotypes into mechanistic insight.

Molecular profiling provides a complementary view. High-throughput measurements of RNA, proteins, and chromatin capture internal programs and environmental responses at single-cell resolution. Recent advances in spatial transcriptomics and large-scale perturbation screens have begun to relate molecular state to tissue context and causal drivers. However, for many questions—especially in the nervous system where cell geometry is intricate and function is tightly coupled to morphology—simultaneous measurement of detailed morphology and transcriptomes remains rare. As a result, most available datasets are *unpaired*: large imaging collections with reconstructed shapes and, separately, large single-cell RNA-seq atlases. Unlocking the biology requires principled methods that can connect these modalities without joint measurements.

Integrating morphology with gene expression poses challenges that differ from conventional multi-omics alignment. Only a small fraction of genes directly influences morphological features, so information is intrinsically imbalanced across modalities. Unlike RNA–ATAC or RNA–protein, there is no straightforward feature-to-feature correspondence to anchor alignment. Moreover, morphology itself must first be converted into quantitative descriptors that respect geometric relationships, and naive embeddings can distort these relationships. Together, these issues make diagonal (unpaired) integration between morphology and transcriptomics both necessary and non-trivial.

We introduce **GeoAdvAE**, a geometry-aware adversarial autoencoder for diagonal integration of single-cell morphology and gene expression. GeoAdvAE learns a shared latent space from unpaired datasets using three complementary ingredients: modality-specific variational autoencoders to preserve reconstruction fidelity, an adversarial objective to align latent distributions across modalities, and a Gromov–Wasserstein regularizer to preserve within-modality geometric structure during alignment. The result is a joint representation that mixes cells across modalities while maintaining biologically meaningful neighborhoods.

We validate GeoAdvAE using Patch-seq neurons, where matched morphology–RNA measurements provide ground truth for cross-modal cell-type matching, and then deploy the framework to microglia in the 5xFAD mouse model of AD. In microglia, the integrated space reveals a one-dimensional continuum that aligns ramified-to-amoeboid morphologies with homeostatic-to-activated transcriptomic states. By probing the model with attribution analyses, we highlight gene programs associated with morphological transitions and identify transcriptomic signatures that are decoupled from visible shape changes—suggesting functional heterogeneity that morphology alone would miss.

1.1 Biological relevance

Alzheimer’s disease (AD) lacks a cure, motivating strategies that enhance resilience – the capacity of cells and circuits to preserve function despite pathology. Microglia are central to this idea: they adopt distinct morphologies that can encircle amyloid plaques and form barrier-like structures associated with tissue protection, even as dystrophic microglia accumulate with aging and may signal impaired responses. Much of this knowledge comes from microscopy, but direct links between these visible phenotypes and underlying molecular programs remain limited. To close that gap, we apply GeoAdvAE to the 5xFAD mouse, an amyloid-only transgenic model with reduced between-sample variability. This setting lets us cleanly interrogate RNA–morphology coupling: even if mouse models do not fully reproduce human resilience or dystrophic signatures, they robustly recapitulate a homeostatic-to-inflammatory microglial continuum (historically framed as M2–M1). By aligning morphology with transcriptomic state in this controlled context, GeoAdvAE provides a roadmap for dissecting putative resilience mechanisms and for prioritizing hypotheses to test in human tissue.

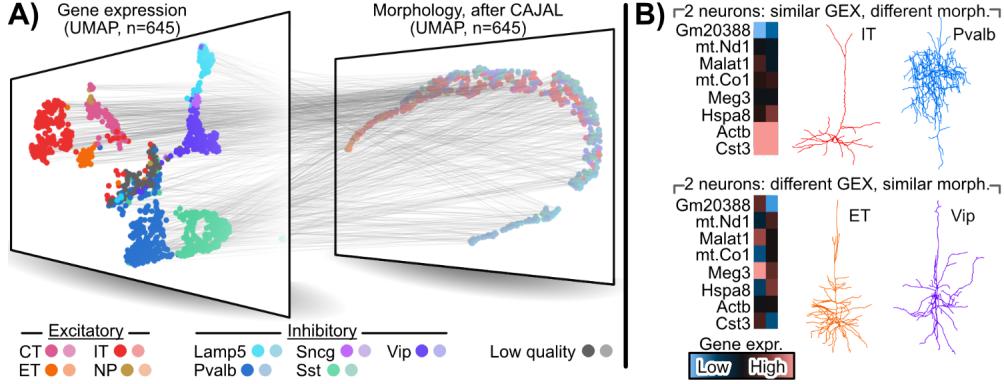


Fig. 1. Patch-seq data of cell GEX and morphology. A) Patch-seq data of 645 neurons [24], showing the complex relation between gene expression (GEX) and morphology (quantified by CAJAL [9]) among different subtypes of excitatory and inhibitory neurons. B) Exemplary genes and morphologies to demonstrate that despite neurons having similar gene expression, they could have vastly different morphologies (top) or vice-versa (bottom).

1.2 Related work and challenges in integrating morphology and gene expression

Our work addresses the key obstacle that integrating cellular morphology with gene expression is substantially different and arguably difficult challenge compared to conventional single-cell integration of two different such as RNA-ATAC or RNA-protein integration. We highlight two main reasons for this.

- **Imbalanced information and no clear correspondence:** Only a small subset of genes influences cell shape (ex: cytoskeleton, membrane dynamics), so morphology–GEX integration is intrinsically asymmetric and low signal-to-noise ratio. This is exemplified in Figure 1, which illustrates Patch-seq neurons whose GEX and morphology are measured simultaneously. Neurons with very similar GEX could have vastly different morphologies, and vice-versa. By contrast, GEX–ATAC and GEX–protein have natural anchors (i.e., regulatory links and protein readouts) supporting feature-level correspondence [5]. Methods built for such settings (ex: SCOT [8], cross-modal autoencoders [37]) tend to assume symmetric information and therefore underperform on morphology–GEX. Morphology emerges from many pathways rather than single genes, precluding one-to-one anchors and weakening co-occurrence/paired-feature strategies [5]. We note that the quantification of morphology itself is also nontrivial. Tools like CAJAL [10] and MorphOMICs [6] yield embeddings, but links to transcriptional programs remain limited.
- **Lack of datasets that measure both genes and morphology of brain cells:** Despite progress in multi-modal integration, linking single-cell morphology with gene expression remains difficult because paired datasets are rare. We note that there are plethora of works that study the relation between gene expression and morphology based on spatial transcriptomics [4,14]. However, these spatial transcriptomics platforms do not have the resolution to capture the intricate morphology of brain cells [23]. For example, while the latest versions of 10x Visium HD can capture spots of up to $2 \mu\text{m}$, this resolution is still not apt for capturing the intricate morphologies of microglia processes that are typically only $1 \mu\text{m}$ thick [35]. Patch-seq [9] is one of the few wet-bench protocols that enables accurate profiling of both gene expression and morphology of neural cells, but its low throughput contrasts with the scale of scRNA-seq and microscopy, which are typically generated separately. Consequently, new frameworks for effective inference from large single-cell unpaired gene expression and morphology datasets are needed.

2 Method

To integrate unpaired cellular morphology and gene expression data, we propose **GeoAdvAE** (geometry-aware adversarial autoencoder), a diagonal integration framework that learns joint representations of both modalities in a shared latent space (Fig. 2A). GeoAdvAE consists of two modality-specific Variational Autoencoders (VAEs) and one discriminator: the two encoders extract features from morphology and gene

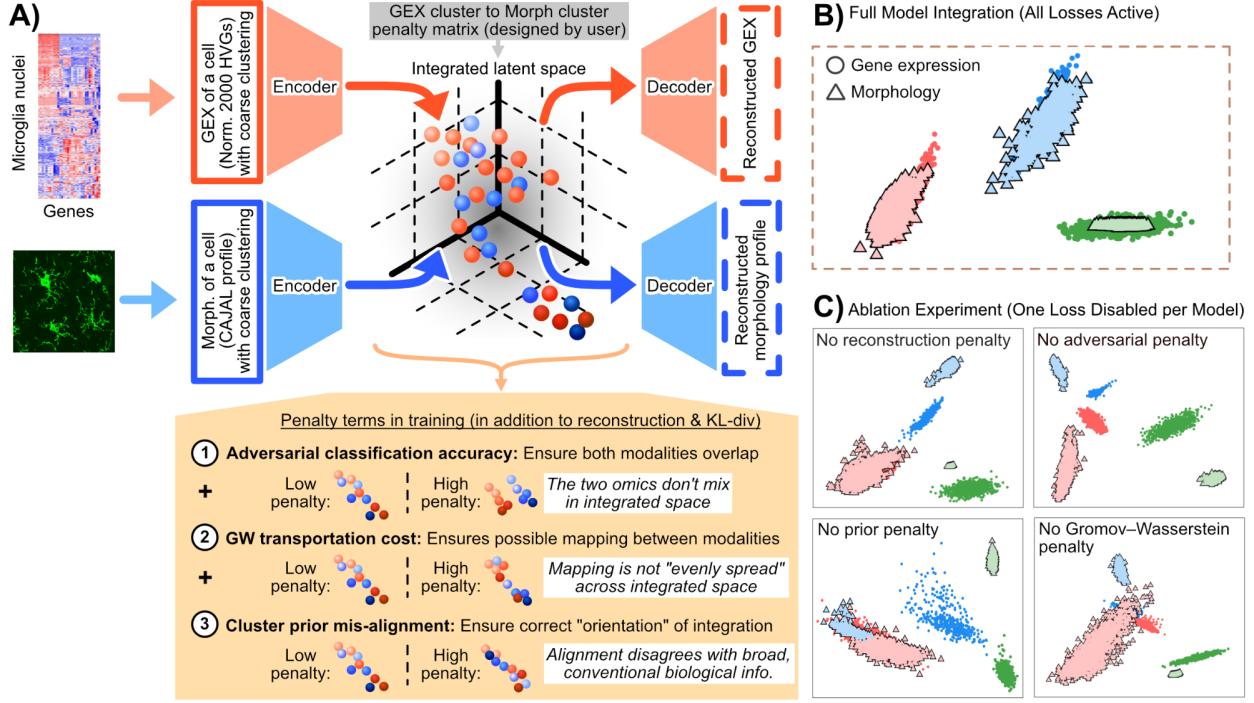


Fig. 2. Architecture and ablation study. A) Overview of the GeoAdvAE architecture integrating unpaired morphology and gene expression through two modality-specific encoders and decoders in a shared latent space. Three additional loss terms (adversarial alignment, Gromov–Wasserstein (GW) transport regularization, and cluster-prior guidance) jointly control modality overlap, geometric consistency, and biologically meaningful orientation, respectively. The white boxes denote qualitative statements on how to interpret potential integrations with high penalty terms (i.e., a not desirable integration). B) Integrated latent space from simulation data computed by GeoAdvAE. C) Ablation experiment (one penalty disabled for each fit), where removing any of the penalty term leads to distinct distortions in the integrated latent space.

expression, respectively, and project them into a common latent distribution. Unlike conventional VAEs that rely solely on reconstruction and KL-divergence losses, GeoAdvAE introduces three complementary penalty terms—variational reconstruction, adversarial alignment, and Gromov–Wasserstein-based structural regularization—that cooperatively guide the training process toward desirable cross-modal alignment. Together, these components enable the model to preserve modality-specific fidelity while enforcing semantic and geometric consistency across unpaired single-cell modalities, forming the foundation for integrating morphology and transcriptomic information in the subsequent stages of our framework.

In order for GeoAdvAE to integrate RNA and morphology, we first need to meaningfully represent a cell’s morphology as a vector of numbers. To do this, we use CAJAL, which represents each microglia’s morphology as a low-dimensional vector based on the Wasserstein distance between its and other cells’ morphologies. Compared to other methods to quantify morphology, we have empirically found CAJAL the most suitable for our proposed diagonal integration framework.

2.1 Model architecture

Encoder and decoder architecture. Each modality uses a VAE. The gene expression profiles ($\mathbf{x}_B \in \mathbb{R}^{2000}$ for the 2000 highly variable genes that are log-normalized) uses three layers. Both encoders output the mean and log-variance of a $d = 16$ integrated latent space. The morphology profiles ($\mathbf{x}_A \in \mathbb{R}^{30}$, quantified via CAJAL [9] as described in Section 3) pass through two hidden layers. Using the reparameterization trick, let \mathbf{z}_A and \mathbf{z}_B sampled from the posterior Gaussian distributions based on the gene expression and morphology encoders using \mathbf{x}_A and \mathbf{x}_B as inputs, respectively. Decoders mirror the encoders back to the input size with the same blocks; the final layer is linear to preserve real-valued reconstructions.

Discriminator architecture. To align the latent spaces across modalities, we add a discriminator $D : \mathbb{R}^d \rightarrow [0, 1]$ is a three-layer multi-layer perceptron. The encoders are updated adversarially to fool D , yielding a shared, modality-invariant latent.

2.2 Loss Function Formulation for Cross-Modal Alignment

We align gene expression (GEX, modality A) and CAJAL morphology embeddings (modality B) in a shared latent space through a weighted sum of five complementary objectives:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{recon}} \cdot \mathcal{L}_{\text{recon}} + \lambda_{\text{KL}} \cdot \mathcal{L}_{\text{KL}} + \lambda_{\text{GAN}} \cdot \mathcal{L}_{\text{GAN}} + \lambda_{\text{GW}} \cdot \mathcal{L}_{\text{GW}} + \lambda_{\text{prior}} \cdot \mathcal{L}_{\text{prior}}, \quad (1)$$

with hyperparameters $\lambda_{\text{KL}}, \lambda_{\text{recon}}, \lambda_{\text{prior}}, \lambda_{\text{GAN}}, \lambda_{\text{GW}} \geq 0$. Our default choices are $\lambda_{\text{KL}} = 0.1$, $\lambda_{\text{recon}} = 5$, $\lambda_{\text{prior}} = 1$, $\lambda_{\text{GAN}} = 4$, and $\lambda_{\text{GW}} = 1$. Training follows a curriculum: autoencoding ($\mathcal{L}_{\text{KL}} + \mathcal{L}_{\text{recon}}$) is active from the start; the adversarial term (\mathcal{L}_{GAN}) is enabled after an initial warmup so that encoders learn stable reconstructions first; the prior-guided semantic alignment ($\mathcal{L}_{\text{prior}}$) ramps up via a schedule; finally, the GW structure term (\mathcal{L}_{GW}) is enabled after the latent geometry stabilizes. As our results later in Section 4 will emphasize, integration that simply only focus on reconstruction are insufficient to integrate GEX and morphology. The additional of these other penalty terms encourage GeoAdvAE to prioritize other qualities that enable meaningful biological discovery beyond what existing methods optimize for.

Reconstruction loss and KL divergence. For each modality, the encoder parameterizes a diagonal-Gaussian posterior; we apply the Kullback–Leibler (KL) divergence penalty against a unit-normal prior to regularize the encoder. Each modality’s decoder reconstructs its original input; we use an L_1 loss (Manhattan distance) to measure the reconstruction criterion for robustness in high dimensions. Both the reconstruction and KL divergence penalties are summed across both modalities.

Adversarial classification based on the discriminator to ensure integration. Leveraging our discriminator, we include an adversarial classification term based on the discriminator to encourage the embeddings from both modalities to overlap. This ensures mixing between the two modalities. Specifically, let q_A and q_B denote the posterior distributions produced by E_A and E_B , respectively, and \mathbf{z}_A and \mathbf{z}_B be samples from these distributions. The discriminator $D : \mathbb{R}^d \rightarrow [0, 1]$ tries to identify the modality based on the samples \mathbf{z}_A and \mathbf{z}_B , while encoders try to produce posterior distributions that fool the discriminator.

To train, we alternate between training the discriminator and generator. The discriminator minimizes the following classification loss with a fixed hyperparameter $\epsilon = 0.1$ for stability, where a prediction closer to 0 or 1 means the discriminator predicts the cell to originate from modality A or B , respectively.

$$\text{Discriminator loss : } \mathcal{L}_D = \mathbb{E}_{\mathbf{z}_A \sim q_A} [(D(\mathbf{z}_A) - \epsilon)^2] + \mathbb{E}_{\mathbf{z}_B \sim q_B} [(D(\mathbf{z}_B) - (1 - \epsilon))^2]. \quad (2)$$

Training the discriminator D (but keeping the encoders fixed, and hence, the embeddings \mathbf{z}_A and \mathbf{z}_B fixed) to obtain a near-0 would reflect a clear separation between the cell morphology and GEX embeddings, which is not desirable for our integration goal (hence, adversarial).

The generator (i.e., encoders) has a penalty term to encourage better integration between GEX and morphology by updating the encoders (but holding the discriminator fixed),

$$\text{Penalty term in GeoAdvAE : } \mathcal{L}_{\text{GAN}} = \mathbb{E}_{\mathbf{z}_A \sim q_A} [(D(\mathbf{z}_A) - (1 - \epsilon))^2] + \mathbb{E}_{\mathbf{z}_B \sim q_B} [(D(\mathbf{z}_B) - \epsilon)^2]. \quad (3)$$

This particular formulation is inspired by Crossmodal-AE [37], where the updated encoder (and hence, updated embeddings \mathbf{z}_A and \mathbf{z}_B) are incentivized to cause the discriminator to wrongly predict whether the cell’s embedding originates from the GEX or morphology modality.

Gromov–Wasserstein loss to enable uniform alignment. We align intra-modality geometry by minimizing the Gromov–Wasserstein (GW) discrepancy between pairwise distances in the two latent spaces. This penalty term encourages the integration to spread out cells in each modality uniformly across the integration space so that as much of a one-to-one mapping between cells from each modality can be achieved. To define this, we first compute the intra-modality distances,

$$D_{ij}^{(A)} = \|\mathbf{z}_i^{(A)} - \mathbf{z}_j^{(A)}\|_2^2, \quad D_{kl}^{(B)} = \|\mathbf{z}_k^{(B)} - \mathbf{z}_l^{(B)}\|_2^2, \quad (4)$$

for all cell pairs $\{i, j\}$ and $\{k, l\}$ in modality A or B , respectively. Then, we match the geometry (among the cells in the minibatch) through an optimal transport plan to obtain a mapping T among all valid transportation plans Π , the set of all doubly stochastic matrices (i.e., rows and columns sum to 1).

$$\mathcal{L}_{\text{GW}} = \min_{T \in \Pi} \sum_{i,j,k,l} \left(D_{ij}^{(A)} - D_{kl}^{(B)} \right)^2 T_{ik} T_{jl}. \quad (5)$$

Prior-guided cluster alignment. To orient the shared latent space with coarse biology (ex: excitatory vs. inhibitory in Patch-seq), we impose a prior on broad cluster–cluster correspondences. Let $C_A : \{1, \dots, N_A\} \rightarrow \{1, \dots, K_A\}$ and $C_B : \{1, \dots, N_B\} \rightarrow \{1, \dots, K_B\}$ be precomputed broad cluster labels for gene expression (modality A) and morphology (modality B). For instance, in our Patch-seq analysis, our clustering is simply separating excitatory from inhibitory neurons. In this example, there are well established distinctions between these two categories of neurons from both gene expression and morphological data. Importantly, the clusterings we use here are *not* about more granular types of neurons since they are more difficult to define from one modality alone. It ensures that broad, literature-supported cell categories remain coherently aligned across modalities, providing semantic structure that complements the unsupervised adversarial and geometric objectives.

The user also provides a correspondence matrix $\mathbf{P} \in [0, 1]^{K_A \times K_B}$ whose entry P_{jk} encodes the expected association strength between GEX cluster j and morphology cluster k . This induces a target similarity

$$S_{ij}^{(\text{target})} = P_{C_A(i), C_B(j)}, \quad \text{for } i \in \{1, \dots, N_A\}, j \in \{1, \dots, N_B\}.$$

For a minibatch with latent codes $\mathbf{z}_A^{(i)}$ and $\mathbf{z}_B^{(j)}$, we define the predicted cross-modal similarity by cosine similarity of ℓ_2 -normalized embeddings,

$$S_{ij}^{(\text{pred})} = \left(\frac{\mathbf{z}_A^{(i)}}{\|\mathbf{z}_A^{(i)}\|_2} \right)^\top \left(\frac{\mathbf{z}_B^{(j)}}{\|\mathbf{z}_B^{(j)}\|_2} \right).$$

The prior loss matches these similarities via a temperature-scaled mean-squared error,

$$\mathcal{L}_{\text{prior}} = \frac{1}{N_A N_B} \cdot \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \left(\frac{S_{ij}^{(\text{pred})}}{\tau} - \frac{S_{ij}^{(\text{target})}}{\tau} \right)^2,$$

where $\tau > 0$ sharpens contrasts, emphasizing confident matches and down-weighting weak ones.

3 Data processing and evaluation

The model operates on two unpaired modalities: morphology and gene expression. The morphologies are “skeletons” of the cell in .swc format, recording the 3-dimensional location and connection between each dendrite/axon/process. These morphologies are processed using CAJAL [9], which encodes structural information in a continuous, numerical space suitable for GeoAdvAE’s deep-learning framework. The transcriptomics are the 2,000 highly variable genes, which we log-normalize to remove the impact of confounding sequencing depth.

Diagonal integration methods to compare against. We benchmark GeoAdvAE against representative cross-modal single-cell integrators spanning distinct paradigms. For instances, coupled/graph autoencoders or latent embedding methods learn shared spaces for unmatched modalities (ScDART [39]; STACI [38]; DCCA [1], scJoint [18]). Some methods also incorporate adversarial terms to encourage a better integration (CycleGAN [40], sciCAN [36], SCIM [26], Crossmodal-AE [37]). Implicit to such methods is the assumption that the axes of variation that best represent the modality are also the axes of variation that are shared between the two modalities. As we have seen in Patch-seq data (Fig. 1A), this assumption is not necessarily true for GEX-morphology alignment. In contrast, optimal-transport (i.e., geometry matching) methods align the global structures between the two modalities (SCOT [8], MMD-MA [19], UnionCom [2]). However, these methods implicitly assume that the global structure for each modality have similar structural properties, and they also lack the ability to distinguish the “proper orientation” of the integration. The Patch-seq data (Fig. 1A) demonstrates that both drawbacks are detrimental when aligning GEX-morphology of neural cells.

Evaluation criteria. We evaluate diagonal integration via cross-modal cell-type transfer accuracy. Importantly, these “true” cell-type labels are *not* used during the training of any of the proposed methods. For each query cell in one modality, we take its 1-nearest neighbor (in Manhattan distance) in the other modality within the latent integrated space and count a match if cell-type labels agree. Accuracy is defined as

$$\text{Acc} = \frac{1}{2} \cdot \left(\frac{\#\text{matches}_{A \rightarrow B}}{\#\text{queries}_{A \rightarrow B}} + \frac{\#\text{matches}_{B \rightarrow A}}{\#\text{queries}_{B \rightarrow A}} \right),$$

where higher values indicates a better cross-modal alignment.

Downstream gene investigation. The importance of GeoAdvAE’s diagonal integration lies in its ability to refine our biological understanding of which genes are associated with morphological changes. Since a cell’s morphology is one important readout of a cell’s function, understanding how the transcriptome relates to morphology provides another avenue to refine the functional relevance of certain gene pathways. Towards this end, we use integrated gradients [29], a computational framework to assess how impact perturbing a gene’s expression alters the resulting embedding, on the GEX encoder $E_A(\cdot)$. By interpreting each gene’s impact as its “importance scores,” we can identify pathways that concurrently have high importance scores, relative to all the other genes via a gene set enrichment analysis (GSEA, [28]). Importantly, this downstream analysis can also identify pathways that are known be differentially expressed across different cell populations that do *not* have a morphological component.

4 Results

4.1 Simulated data demonstrates GeoAdvAE’s advantage over other method and enables ablation studies

We build a simulator that generates synthetic gene-expression profiles and corresponding neuronal morphologies to evaluate GeoAdvAE under controlled conditions. This enables us to precisely diagnose how each loss component of GeoAdvAE contributes to its superior performance over other integration methods. Three canonical neuron types (pyramidal, multipolar, bipolar) are modeled by sampling GEX from three low-dimensional clusters; morphology is then generated by a process in which GEX controls polarity, branch density, and anisotropy, yielding three separable CAJAL morphology clusters (Supplemental Fig. S1; Appendix S2). For diagonal integration we use a 3×3 correspondence matrix \mathbf{P} (identity), which maps each GEX cluster to its morphology counterpart (Fig. 3B).

To illustrate the importance of each penalty term in GeoAdvAE, we perform an ablation study, where we remove one of the penalty terms and assess how the resulting integration suffered (Fig. 2C). Removing the adversarial classifier separates modalities in latent space. Removing the Gromov–Wasserstein penalty breaks local geometric coherence. Removing prior-guided cluster alignment misorients matched clusters. Removing reconstruction loss destabilizes embeddings and blurs cluster boundaries. Notably, although our correspondence matrix \mathbf{P} “reveals” how the two modalities should be integrated, we find that this biological information is *necessary* for GEX-morphology integration. In the presence of no paired samples and no matched features between the both modalities, the information represented by \mathbf{P} provides the necessary orientation signal that the data alone cannot recover.

We next compare GeoAdvAE with other competing methods using both quantitative and qualitative evaluations. GeoAdvAE achieves the highest alignment accuracy (Fig. 3A). Latent embedding methods (scJOINT and STACI) perform reasonably well but tended to under-align fine-grained structures, while optimal-transport methods (UnionComm, MMD-MA) struggle to uncover the correct integration orientation (Fig. 3B, Supplemental Fig. S2). Collectively, these results support combining adversarial alignment with cluster-prior regularization to achieve proper alignment between GEX and morphology.

4.2 Patch-seq neurons provide validation for GeoAdvAE’s integration

As mentioned in Section 1, Patch-sequencing is one of the few wet-bench protocols where the GEX and morphology of neural cells can be simultaneously measured. Such data provides an invaluable resources

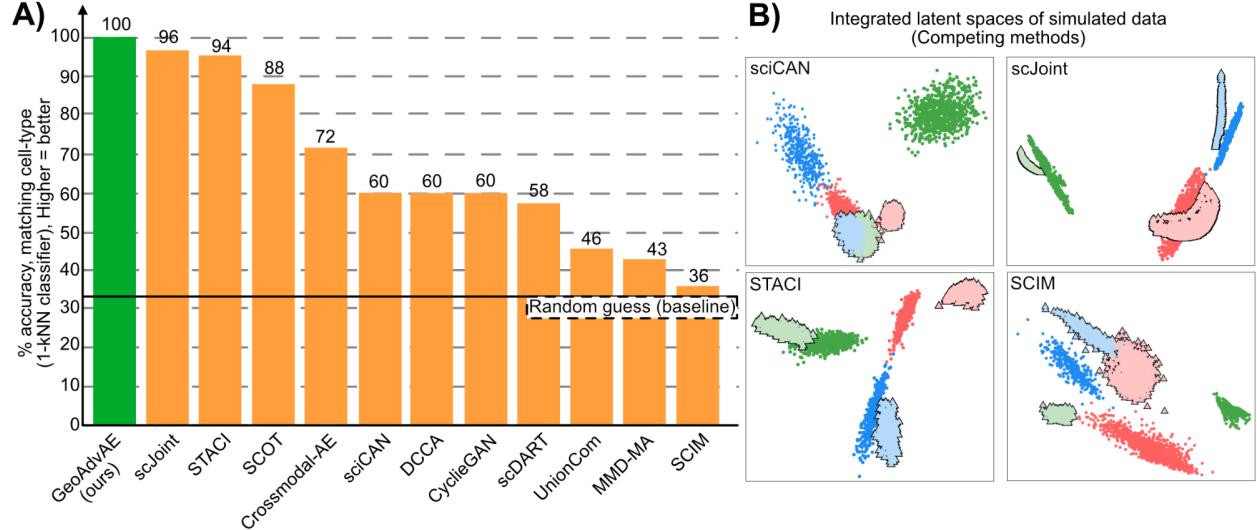


Fig. 3. Simulation study via synthetic brain cells. A) Comparison of cell-type matching accuracy on simulation data across GeoAdvAE and competing cross-modal integration methods. The solid horizontal line denotes the accuracy achieved by random guessing. B) Integrated latent spaces of simulated data produced by selected baseline methods (sciCAN [36], STACI [38], scJoint [18], SCIM [26]). The results by these competing methods can be compared against GeoAdvAE’s integration shown in Figure 2B.

to validate the performance of any GEX-morphology diagonal integration method. Hence, we evaluate our method on a Patch-seq dataset [24], which contains 645 neurons spanning a diverse set of excitatory and inhibitory neurons from the mouse motor cortex. To define the cluster prior matrix in the Patch-seq dataset, we *do not* use the provided cell annotations because the provided annotations were defined using by the original authors using many modalities. Instead, for GEX, we clustered the neurons into 4 clusters and annotated each clusters as either excitatory or inhibitory using canonical marker genes (ex: *Slc17a7* and *Gad1*). [[KZL: Citation]] For morphology, we embedded the neurons into a lower-dimensional space using CAJAL [9] and grouped them into 4 clusters. Based on our visually assessing 10 randomly chosen neurons in each cluster, we annotated the proportion of excitatory to inhibitory neurons based on whether or not the cell has a pyramidal shape. Based on these manual annotations, we construct the correspondence matrix \mathbf{P} mapping between the GEX and morphology clusters based on the how highly our coarse excitatory-vs.-inhibitory labels aligns between the two modalities. The details are described in Appendix S2.

After fitting GeoAdvAE, we use the true alignment between GEX and morphology to demonstrate that GeoAdvAE achieves more biologically meaningful and robust integration compared to competing methods. As shown in Fig. 4A, we evaluate all models by cell-type alignment accuracy, which measures the percentage of correctly matched cell types across modalities. GeoAdvAE attains the highest accuracy (34%), substantially outperforming all baselines. ScDART (28%) and STACI (27%) follow as the next best-performing methods, suggesting that approaches with graph-based priors are particularly effective at discovering cross-modal relations. In contrast, latent-space alignment methods (ex: scJoint, Crossmodal-AE) and adversarial models (ex: CycleGAN, sciCAN) exhibit degraded performance on Patch-seq data, likely due to difficulties in maintaining modality consistency under biological noise and nonlinear morphological variability. Similar to our simulation results (Fig. 3A), we find that optimal transport methods (ex: SCOT, UnionCom) perform poorly for GEX-morphology alignment. This is potentially due to fact that the GEX modality displays a much larger diversity of cell types when compared to the morphology modality (Fig. 1A). Hence, optimal transport methods that directly try to map the geometric relations between neuron’s GEX to those in morphology fail because the GEX modality captures many axes of variation not represented in the morphology modality. The UMAP visualizations in Fig. 4D further support this observation: compared with Crossmodal-AE, GeoAdvAE produces a smoother and more coherently mixed latent representation, where excitatory and inhibitory neurons form distinct yet biologically consistent distributions within the shared latent space.

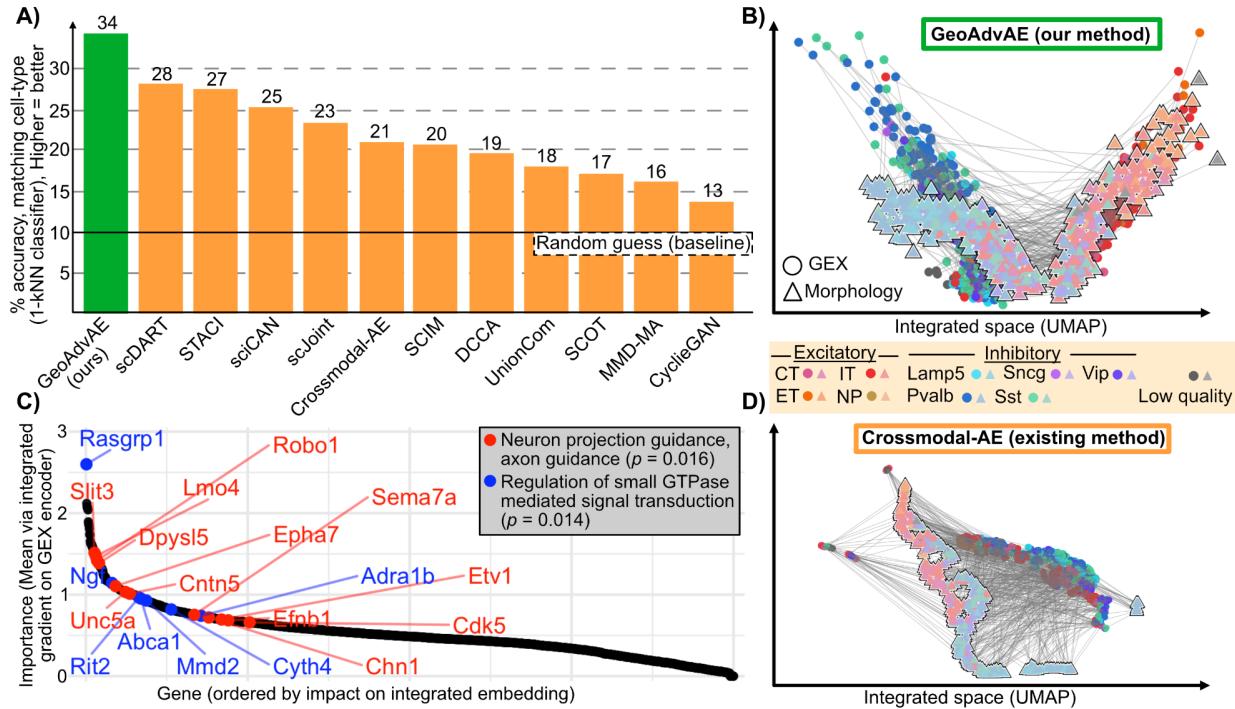


Fig. 4. Comparison and validation of integration methods using Patch-seq neurons. A) Comparison of cell-type matching accuracy on Patch-seq data across GeoAdvAE and competing cross-modal integration methods. GeoAdvAE achieves the highest alignment accuracy. B) Integrated latent space learned by GeoAdvAE on Patch-seq neurons. C) Integrated gradient importance scores of all the 2000 genes based on the GEX encoder. The genes in two critical pathways known to regulate neuron morphologies are highlighted in red and blue, both nominally significant. D) Integrated latent space generated by the existing Crossmodal-AE method [37], showing less continuous cross-modal alignment compared to GeoAdvAE. [[KZL (11/12 To Turbo): Fix the font sizes in B. Make it same as those in A]]

We also ensure that GeoAdvAE’s integration recapitulates known neuronal biology. After applying the integrated gradients [29] on the GEX encoder $E_A(\cdot)$, we see that GSEA uncovers many transcriptomic pathways known to impact a neuron’s morphology. We highlight two important pathways in Figure 4C. First, pathways for neuron projection guidance (GO:0097485) and axon guidance (GO:0007411) play a critical role in axon/neurite guidance and cytoskeletal regulators (example genes: *Slit3*, *Robo1*, *EphA7*, *Efnb1*, *Unc5a*; p-value: 0.016). Additionally, Rho-family small GTPases are master regulators of neuronal morphogenesis, coupling extracellular cues to actin remodeling that governs spine structure, filopodial dynamics, and growth-cone advance or collapse [20]. In diverse contexts, guidance and trophic signaling converge on these GTPases to control growth-cone behavior and neurite extension [34,7]. Consistent with this framework, we find that the relevant GO term (GO:0051056) is strongly associated with GEX–morphology, via genes such as *Rasgrp1*, *Robo1*, *Ngt*, *Rit2*, *Cyth4*, and *Chn1*.

4.3 GeoAdvAE uncovers novel biology of 5xFAD microglia

Our primary dataset of interest regard microglia from the 5xFAD mouse model which recapitulates amyloid- β progression, one of the hallmarks of Alzheimer’s disease [22]. [[KZL: I need to work on this transition]] For morphological data, we leverage 98 microglia skeletons analyzed by mice at 3 and 6 months of both genders [6]. After applying CAJAL and clustering, we observe 3 primary microglial clusters (Fig. 5A), where one microglial cluster (cyan) is enriched for larger microglia with a more ramified morphology primarily from male, younger mice. In contrast, another microglial cluster (pink) is enriched for smaller microglia with a more ameboid morphology primarily from female, older mice. For transcriptomic data, we leverage

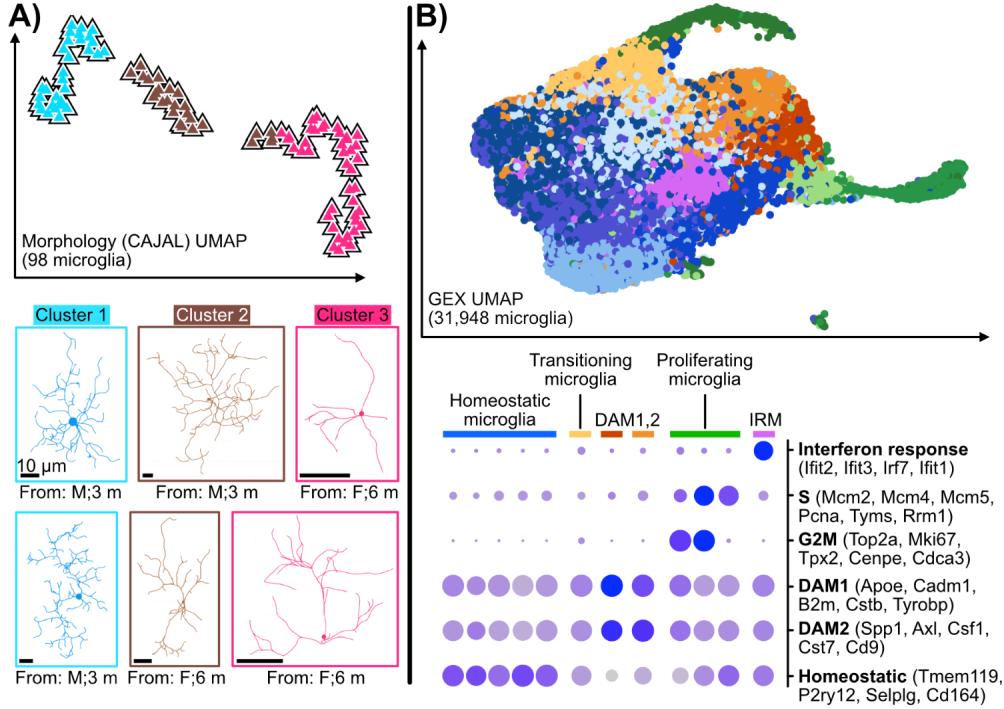


Fig. 5. 5xFAD microglia data. A) Microglia morphologies [6], quantified via CAJAL and clustered into 3 major morphological types. Exemplary microglia of each cluster are shown below. B) Microglia gene expressions [31], using the author's annotated microglial states. Marker genes for each microglia states are shown as a dot plot below.

31,948 microglial single-nuclei RNA-sequencing [31]. These microglia span many different states that the authors classified using marker genes, including homeostatic, proliferating, interferon-response (IRM), and disease-associated microglia (DAM) (Fig. 5B).

To form the correspondence matrix \mathbf{P} necessary for the prior-guided cluster alignment, we annotate broad correspondences between the two modalities. For GEX, we use the provided 12 clusters for GEX microglia and further quantify the proportion of homeostatic, proliferating, IRM, and DAM microglias within each cluster based on an enrichment analysis of marker genes. This enables us to capture the continuous state transitions among the microglia states, rather than discretely labeling each cluster as a “pure” state. For morphology, we cluster the CAJAL profiles into 3 clusters and labeled the cluster as ramified, intermediate, and amoeboid after visual inspection of the microglia in each cluster. Based on these annotations, we design \mathbf{P} that accounts for the within-cluster proportions and biological correspondence between microglial transcriptomic states and morphological states. See Appendix S2 for more details on this procedure.

When we applied GeoAdvAE on the microglia datasets, we observe an apparent 1-dimensional manifold that suggests novel biological insights into microglia. Figure 6A depicts the integrated embedding where the ramified morphologies (cyan triangles) integrate with the hoemostatic microglia (blue circles), while the ameoboid morphologies (pink triangles) integrate with the DAM microglia (orange circles). While this overall trend is not surprising due to GeoAdvAE’s prior-guided cluster alignment, we perform downstream analyses to investigate two notable biological questions: 1) which particular genes contributed to this alignment, and 2) what is the continuum of microglial states that spanned this 1-dimensional manifold?

We leverage our integrated gradient framework to investigate which genes are associated with microglial morphology. After ordering all the genes by their importance, we observe two significant pathways that have opposing directions (Fig. 6B). On one side, genes in the DNA repair pathway were strongly enriched among ramified microglia, while those in the cell killing pathway were strongly enriched among the ameoboid microglia. This finding recapitulated our known understanding of these different microglial morphologies [[Citation]], but direct transcriptomic evidence of the functional roles of these morphologies was not accessible prior to GeoAdvAE. [[Put in the blurb about ABeta clearance and C1Q]]

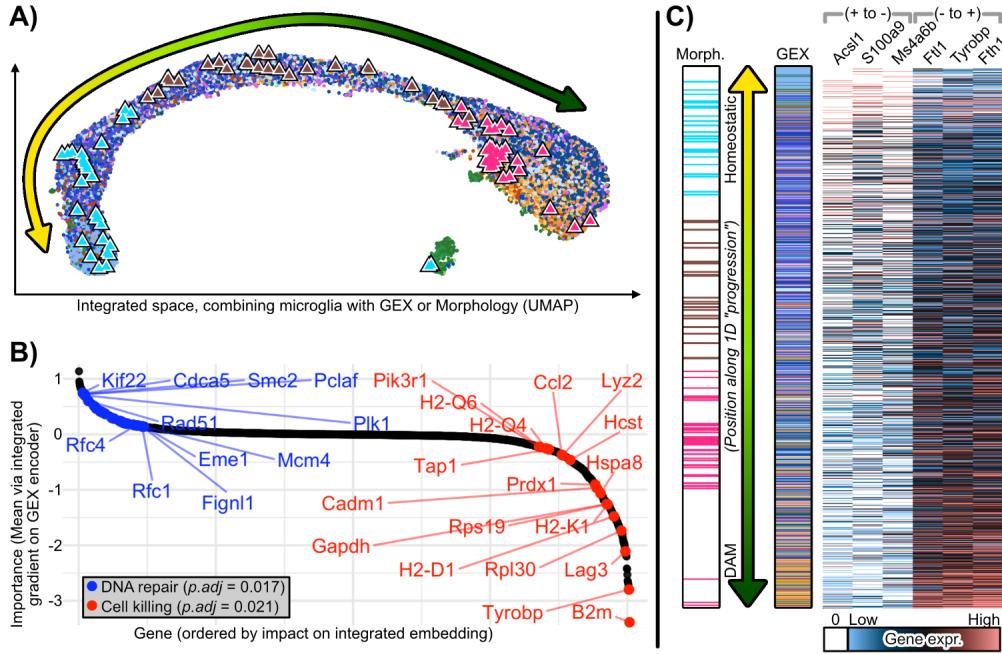


Fig. 6. Integration of GEX and morphology uncovers functional roles of microglia in 5xFAD. A) Integration of microglia GEX and morphology, uncovering a 1-dimensional manifold that correspond to a spectrum of microglial states. B) Genes ordered by their influence on a microglia's embedding in the integrated spaced as measured by integrated gradients [29], highlight two major pathways corresponding to the two different ends of the GEX-morphology integrated space. C) Ordering microglia along the 1-dimensional manifold, highlighting how microglia measured by morphology (left) or GEX (right) align along this manifold. Exemplary genes that also generally increase or decrease going from one end of the spectrum to the other are displayed as a heatmap.

Ordering the genes along the 1-dimensional manifold provided an alternative perspective on which genes are dynamic changing as microglia shift from one morphology to another. In Fig. 6C, we order the microglia (both morphology and transcriptomic) along the manifold using Slingshot [27]. Then, we computed the correlation of each gene against this ordering to understand which genes were associated with morphological changes. Surprisingly, we find *Ms4a6b* to be strongly associated with ramified morphologies [[KZL: I'll write the rationale later]]. On the other hand, we find *Ftl1* and *Fth1*, two genes critical for iron accumulation. This is a hallmark of dystrophic microglia [[Citation]], which our findings recapitulate. [[KZL: Include a few sentences about microglia that are DAM-associated but not correlated]]

5 Discussion

References

1. ANDREW, G., ARORA, R., BILMES, J., AND LIVESCU, K. Deep canonical correlation analysis. In *International conference on machine learning* (2013), PMLR, pp. 1247–1255.
2. CAO, K., BAI, X., HONG, Y., AND WAN, L. Unsupervised topological alignment for single-cell multi-omics integration. *Bioinformatics* 36, Supplement _1 (2020), i48–i56.
3. CAO, K., HONG, Y., AND WAN, L. Manifold alignment for heterogeneous single-cell multi-omics data integration using Pamona. *Bioinformatics* 38, 1 (2022), 211–219.
4. CHELEBIAN, E., AVENEL, C., AND WÄHLBY, C. What makes for good morphology representations for spatial omics? *arXiv preprint arXiv:2407.20660* (2024).
5. CHEN, S., ZHU, B., HUANG, S., HICKEY, J. W., LIN, K. Z., SNYDER, M., GREENLEAF, W. J., NOLAN, G. P., ZHANG, N. R., AND MA, Z. Integration of spatial and single-cell data across modalities with weakly linked features. *Nature Biotechnology* 42, 7 (2024), 1096–1106.
6. COLOMBO, G., CUBERO, R. J. A., KANARI, L., VENTURINO, A., SCHULZ, R., SCOLAMIERO, M., AGERBERG, J., MATHYS, H., TSAI, L.-H., CHACHÓLSKI, W., ET AL. A tool for mapping microglial morphology, morphomics, reveals brain-region and sex-dependent phenotypes. *Nature neuroscience* 25, 10 (2022), 1379–1393.
7. DANIELS, R. H., HALL, P. S., AND BOKOCH, G. M. Membrane targeting of p21-activated kinase 1 (pak1) induces neurite outgrowth from pc12 cells. *The EMBO journal* (1998).
8. DEMETCI, P., SANTORELLA, R., SANDSTEDE, B., NOBLE, W. S., AND SINGH, R. SCOT: Single-cell multi-omics alignment with optimal transport. *Journal of Computational Biology* 29, 1 (2022), 3–18.
9. GOUWENS, N. W., SORENSEN, S. A., BAFTIZADEH, F., BUDZILLO, A., LEE, B. R., JARSKY, T., ALFILER, L., BAKER, K., BARKAN, E., BERRY, K., ET AL. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell* 183, 4 (2020), 935–953.
10. GOVEK, K. W., NICODEMUS, P., LIN, Y., CRAWFORD, J., SATURNINO, A. B., CUI, H., ZOGA, K., HART, M. P., AND CAMARA, P. G. CAJAL enables analysis and integration of single-cell morphological data using metric geometry. *Nature Communications* 14, 1 (2023), 3672.
11. GREEN, T. R., AND ROWE, R. K. Quantifying microglial morphology: an insight into function. *Clinical and Experimental Immunology* 216, 3 (2024), 221–229.
12. GRUBMAN, A., CHOO, X. Y., CHEW, G., OUYANG, J. F., SUN, G., CROFT, N. P., ROSSELLO, F. J., SIMMONS, R., BUCKBERRY, S., LANDIN, D. V., ET AL. Transcriptional signature in microglia associated with A β plaque phagocytosis. *Nature Communications* 12, 1 (2021), 3015.
13. HAGHIGHI, M., CAICEDO, J. C., CIMINI, B. A., CARPENTER, A. E., AND SINGH, S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nature methods* 19, 12 (2022), 1550–1557.
14. HUANG, J., YUAN, C., JIANG, J., CHEN, J., BADVE, S. S., GOKMEN-POLAR, Y., SEGURA, R. L., YAN, X., LAZAR, A., GAO, J., ET AL. Bridging cell morphological behaviors and molecular dynamics in multi-modal spatial omics with morphlink. *Nature Communications* 16, 1 (2025), 5878.
15. JIANG, J., LIU, Y., QIN, J., CHEN, J., WU, J., PIZZI, M. P., LAZCANO, R., YAMASHITA, K., XU, Z., PEI, G., ET AL. METI: Deep profiling of tumor ecosystems by integrating cell morphology and spatial transcriptomics. *Nature Communications* 15, 1 (2024), 7312.
16. LENG, F., AND EDISON, P. Neuroinflammation and microglial activation in alzheimer disease: Where do we go from here? *Nature Reviews Neurology* 17, 3 (2021), 157–172.
17. LI, X., LI, Y., JIN, Y., ZHANG, Y., WU, J., XU, Z., HUANG, Y., CAI, L., GAO, S., LIU, T., ET AL. Transcriptional and epigenetic decoding of the microglial aging process. *Nature aging* 3, 10 (2023), 1288–1311.
18. LIN, Y., WU, T.-Y., WAN, S., YANG, J. Y., WONG, W. H., AND WANG, Y. R. scJoint integrates atlas-scale single-cell rna-seq and atac-seq data with transfer learning. *Nature biotechnology* 40, 5 (2022), 703–710.
19. LIU, J., HUANG, Y., SINGH, R., VERT, J.-P., AND NOBLE, W. S. Jointly embedding multiple single-cell omics measurements. In *Algorithms in bioinformatics:... International Workshop, WABI..., proceedings. WABI (Workshop)* (2019), vol. 143, p. 10.
20. LUO, L. Rho gtpases in neuronal morphogenesis. *Nature Reviews Neuroscience* 1, 3 (2000), 173–180.
21. MANCUSO, R., FATTORELLI, N., MARTINEZ-MURIANA, A., DAVIS, E., WOLFS, L., VAN DEN DAELE, J., GERIC, I., PREMEREUR, J., POLANCO, P., BIJNENS, B., ET AL. Xenografted human microglia display diverse transcriptomic states in response to Alzheimer's disease-related amyloid- β pathology. *Nature Neuroscience* 27, 5 (2024), 886–900.
22. OAKLEY, H., COLE, S. L., LOGAN, S., MAUS, E., SHAO, P., CRAFT, J., GUILLOZET-BONGAARTS, A., OHNO, M., DISTERHOFT, J., VAN ELDIK, L., ET AL. Intraneuronal β -amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial alzheimer's disease mutations: potential factors in amyloid plaque formation. *Journal of Neuroscience* 26, 40 (2006), 10129–10140.

23. QIAN, X., COLEMAN, K., JIANG, S., KRIZ, A. J., MARCIANO, J. H., LUO, C., CAI, C., MANAM, M. D., CAGLAYAN, E., LAI, A., ET AL. Spatial transcriptomics reveals human cortical layer and area specification. *Nature* (2025), 1–11.
24. SCALA, F., KOBAK, D., BERNABUCCI, M., BERNAERTS, Y., CADWELL, C. R., CASTRO, J. R., HARTMANIS, L., JIANG, X., LATURNUS, S., MIRANDA, E., ET AL. Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* 598, 7879 (2021), 144–150.
25. SINGH, R., DEMETCI, P., BONORA, G., RAMANI, V., LEE, C., FANG, H., DUAN, Z., DENG, X., SHENDURE, J., DISTECHÉ, C., ET AL. Unsupervised manifold alignment for single-cell multi-omics data. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (2020), pp. 1–10.
26. STARK, S. G., FICEK, J., LOCATELLO, F., BONILLA, X., CHEVRIER, S., SINGER, F., RÄTSCH, G., AND LEHMANN, K.-V. Scim: universal single-cell matching with unpaired feature sets. *Bioinformatics* 36, Supplement_2 (2020), i919–i927.
27. STREET, K., RISSO, D., FLETCHER, R. B., DAS, D., NGAI, J., YOSEF, N., PURDOM, E., AND DUDOIT, S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics* 19, 1 (2018), 477.
28. SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S., ET AL. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102, 43 (2005), 15545–15550.
29. SUNDARARAJAN, M., TALY, A., AND YAN, Q. Axiomatic attribution for deep networks. In *International conference on machine learning* (2017), PMLR, pp. 3319–3328.
30. VIDAL-ITRIAGO, A., RADFORD, R. A., ARAMIDEH, J. A., MAUREL, C., SCHERER, N. M., DON, E. K., LEE, A., CHUNG, R. S., GRAEBER, M. B., AND MORSCH, M. Microglia morphophysiological diversity and its implications for the CNS. *Frontiers in Immunology* 13 (2022), 997786.
31. WANG, S., MUSTAFA, M., YUEDE, C. M., SALAZAR, S. V., KONG, P., LONG, H., WARD, M., SIDDIQUI, O., PAUL, R., GILFILLAN, S., ET AL. Anti-human trem2 induces microglia proliferation and reduces pathology in an alzheimer’s disease model. *Journal of Experimental Medicine* 217, 9 (2020), e20200785.
32. WANG, X., FAN, Y., GUO, Y., FU, C., LEE, K., DALLAKYAN, K., LI, Y., YIN, Q., LI, Y., AND SONG, L. Prediction of cellular morphology changes under perturbations with a transcriptome-guided diffusion model. *Nature Communications* 16, 1 (2025), 8210.
33. WATSON, C., PAXINOS, G., AND PUELLES, L. *The mouse nervous system*. Academic Press, 2011.
34. WEGMEYER, H., EGEA, J., RABE, N., GEZELIUS, H., FILOSA, A., ENJIN, A., VAROQUEAUX, F., DEININGER, K., SCHNÜTGEN, F., BROSE, N., ET AL. EphA4-dependent axon guidance is mediated by the racgap $\alpha 2$ -chimaerin. *Neuron* 55, 5 (2007), 756–767.
35. WHITELAW, B. S., STOESSEL, M. B., AND MAJEWSKA, A. K. Movers and shakers: microglial dynamics and modulation of neural networks. *Glia* 71, 7 (2023), 1575–1591.
36. XU, Y., BEGOLI, E., AND MCCORD, R. P. scican: single-cell chromatin accessibility and gene expression data integration via cycle-consistent adversarial network. *NPJ Systems Biology and Applications* 8, 1 (2022), 33.
37. YANG, K. D., BELYAEVA, A., VENKATACHALAPATHY, S., DAMODARAN, K., KATCOFF, A., RADHAKRISHNAN, A., SHIVASHANKAR, G., AND UHLER, C. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nature Communications* 12, 1 (2021), 31.
38. ZHANG, X., WANG, X., SHIVASHANKAR, G., AND UHLER, C. Graph-based autoencoder integrates spatial transcriptomics with chromatin images and identifies joint biomarkers for alzheimer’s disease. *Nature Communications* 13, 1 (2022), 7480.
39. ZHANG, Z., YANG, C., AND ZHANG, X. scdart: integrating unmatched scRNA-seq and scatac-seq data and learning cross-modality relationship simultaneously. *Genome biology* 23, 1 (2022), 139.
40. ZHU, J.-Y., PARK, T., ISOLA, P., AND EFROS, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2223–2232.

Supplementary Materials

S1 Additional details on GeoAdvAE

S1.1 Additional details on architecture

Below, we provide additional implementation details of the GeoAdvAE architecture and training configuration. The model uses a latent-space dimension of $d=16$, with hidden-layer sizes of $\{256, 512, 1024\}$ depending on the modality. All networks were initialized using Kaiming normal initialization and optimized using the Adam optimizer with a batch size of 32 over the total iterations. Dropout was applied at a rate of 0.1 in the encoder and decoder networks and 0.3 in the discriminator to prevent overfitting. The prior-guided loss employed a temperature parameter of $\tau=0.07$, while label smoothing with $\epsilon=0.1$ was used in the adversarial discriminator to stabilize training. For the Gromov–Wasserstein (GW) alignment, we used the global alignment setting across all batches. When estimating cell-type proportions during evaluation, the neighborhood size was set to a maximum of $k=10$ for nearest-neighbor computations.

S1.2 Additional details on loss function

Below, we provide additional details about the various terms in GeoAdvAE’s loss function.

- **Reconstruction loss:** Each modality has its own decoder D_m and reconstruction $\hat{\mathbf{x}}_m = D_m(\mathbf{z}_m)$ with $\mathbf{z}_m \sim q_m(\mathbf{z}_m | \mathbf{x}_m)$. We use an L_1 reconstruction loss,

$$\mathcal{L}_{\text{recon}}^{(m)} = \frac{1}{N_m} \sum_{i=1}^{N_m} \left\| \mathbf{x}_m^{(i)} - \hat{\mathbf{x}}_m^{(i)} \right\|_1, \quad \mathcal{L}_{\text{recon}} = \mathcal{L}_{\text{recon}}^{(A)} + \mathcal{L}_{\text{recon}}^{(B)}. \quad (\text{S1})$$

- **KL divergence:** For each modality $m \in \{A, B\}$, the encoder E_m outputs parameters of a diagonal Gaussian $q_m(\mathbf{z}_m | \mathbf{x}_m) = \mathcal{N}(\boldsymbol{\mu}_m, \text{diag}(\boldsymbol{\sigma}_m^2))$, with latent dimension d . We penalize the KL divergence to the standard normal prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$:

$$\mathcal{L}_{\text{KL}}^{(m)} = \frac{1}{d} D_{\text{KL}}(q_m(\mathbf{z}_m | \mathbf{x}_m) \| p(\mathbf{z})) = -\frac{1}{2d} \sum_{i=1}^d (1 + \log \sigma_{m,i}^2 - \mu_{m,i}^2 - \sigma_{m,i}^2). \quad (\text{S2})$$

The total KL loss adds the two modalities:

$$\mathcal{L}_{\text{KL}} = \mathcal{L}_{\text{KL}}^{(A)} + \mathcal{L}_{\text{KL}}^{(B)}. \quad (\text{S3})$$

- **Adversarial classification:** For the alternating training between the generator and discriminator, we adopt an unbalanced alternating strategy in which the discriminator is updated twice for every generator update. This stabilizes adversarial learning and prevents model collapse. Each iteration processes one batch of 32 samples. For example, on the Patch-seq dataset (645 cells, 3000 iterations), this corresponds to roughly 150 epochs in total.
- **Prior-guided cluster alignment:** To stabilize early optimization, a linear warmup schedule gradually increases the influence of this loss during the first T_{warm} iterations:

$$\mathcal{L}_{\text{prior}}^{\text{final}} = \min \left(1, \frac{t}{T_{\text{warm}}} \right) \mathcal{L}_{\text{prior}}, \quad t = 1, \dots, T_{\text{warm}}. \quad (\text{S4})$$

S1.3 Additional details on training and timing

Training follows a three-stage warm-start activation schedule designed to stabilize optimization:

1. **Phase 1 (0–200 iterations):** Train both VAEs with \mathcal{L}_{KL} and $\mathcal{L}_{\text{recon}}$ active, establishing modality-faithful encoders. The prior loss $\mathcal{L}_{\text{prior}}$ is gradually warmed up to introduce coarse semantic alignment.

2. **Phase 2 (200–600 iterations):** Introduce the adversarial term \mathcal{L}_{GAN} to align latent marginal distributions while preserving reconstruction accuracy. The discriminator is updated twice per iteration ($n_D=2$) for every encoder-decoder update.
3. **Phase 3 (600–1000 iterations):** Introduce the structural Gromov–Wasserstein loss \mathcal{L}_{GW} (either global or groupwise mode) to refine geometric alignment of latent spaces.

Each encoder-decoder pair is optimized with the Adam optimizer ($\text{lr} = 2 \times 10^{-4}$, $\beta_1 = 0.5$, $\beta_2 = 0.999$) and a small weight decay (10^{-4}). The discriminator uses the same learning rate but maintains independent optimizer states. A step-decay policy (decay factor 0.8 every 800 iterations) is applied to the learning rate to ensure stable convergence across training phases.

Hyperparameters and default configuration. GeoAdvAE has multiple hyperparameters, as alluded to in the main text. Empirically, the most sensitive ones are:

- The prior loss weight λ_{prior} and temperature τ , which strongly influence semantic cluster alignment;
- The relative ratio between the reconstruction and GAN hyperparameters, λ_{recon} and λ_{GAN} , which controls the balance between information preservation and modality matching;
- The latent dimension d , which affects both alignment granularity and stability.

For the remaining parameters, moderate settings ($\lambda_{\text{prior}}=1.0$, $\tau=0.07$, $d=16$) provide the most stable and reproducible convergence.

Model efficiency and computational time. GeoAdvAE demonstrates high computational efficiency across datasets. Training on the synthetic simulation dataset completes in approximately 5 minutes for 2000 iterations. For the Patch-seq dataset, 3000 iterations require about 20 minutes. On the larger 5xFAD dataset, the model converges within 40 minutes over 9000 iterations, confirming the scalability and practicality of the framework for diverse single-cell integration tasks.

S2 Additional details on analysis workflow

S2.1 Additional details on simulating data

To systematically validate GeoAdvAE, we design a biologically interpretable simulation framework that jointly generates synthetic GEX and morphology data with known cross-modal relationships. The simulation is constructed to emulate three canonical neuron archetypes (pyramidal, multipolar, and bipolar neurons) each characterized by distinct transcriptional programs and morphological phenotypes.

The overall simulation process follows the chain:

$$\text{Gene expression } (g_1, g_2, g_3) \xrightarrow{\text{sigmoid mapping}} (P, D, A) \xrightarrow{\text{morphology rules}} \text{Neuronal morphology}.$$

This design ensures interpretability (each gene regulates one morphological axis), separability (distinct clusters in both modalities), and determinism (identical expression yields identical morphology), thereby enabling a biologically meaningful benchmark for evaluating GeoAdvAE’s integration performance.

Step 1. Generating separable gene expression clusters. We define three well-separated transcriptional clusters, each corresponding to a neuronal subtype, by sampling three synthetic genes (g_1, g_2, g_3) from Gaussian distributions with distinct means and small variance:

$$\begin{aligned} \text{Cluster 1 (Pyramidal)} : g_1 &\sim \mathcal{N}(2.0, 0.2^2), & g_2 &\sim \mathcal{N}(0.0, 0.2^2), & g_3 &\sim \mathcal{N}(1.2, 0.2^2), \\ \text{Cluster 2 (Multipolar)} : g_1 &\sim \mathcal{N}(-2.0, 0.2^2), & g_2 &\sim \mathcal{N}(2.0, 0.2^2), & g_3 &\sim \mathcal{N}(-1.0, 0.2^2), \\ \text{Cluster 3 (Bipolar)} : g_1 &\sim \mathcal{N}(0.0, 0.2^2), & g_2 &\sim \mathcal{N}(-1.0, 0.2^2), & g_3 &\sim \mathcal{N}(2.0, 0.2^2). \end{aligned}$$

The resulting clusters are linearly separable in 3D gene-expression space, providing an interpretable ground truth for downstream alignment analysis.

Step 2. Mapping GEX to morphological control parameters. Each gene expression vector (g_1, g_2, g_3) was transformed into three biologically interpretable morphological control parameters (polarity (P), proximal branching density (D), and anisotropy (A)) via a nonlinear sigmoid mapping:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad P = \sigma(1.5 \cdot g_1), \quad D = \sigma(1.5 \cdot g_2), \quad A = \sigma(1.5 \cdot g_3).$$

This mapping captures how transcriptional programs may regulate morphological features such as polarity, branching complexity, and spatial orientation. For instance, pyramidal neurons (high P , moderate D , high A) are simulated as having a dominant apical trunk with top-branch clusters; multipolar neurons (low P , high D , low A) exhibit dense isotropic branching; and bipolar neurons (intermediate P , low D , high A) form elongated bidirectional projections.

Step 3. Morphology generation via procedural rules. Given (P, D, A) , we procedurally generated synthetic neuronal trees reflecting their morphological archetypes:

- **Pyramidal** ($\bar{P}=0.95$, $\bar{D}=0.50$, $\bar{A}=0.87$): a single apical trunk extending along the y -axis, branching into three major apical clusters with 6 basal dendrites; the morphology .swc file for each neuron consists of $\sim 200\text{--}300$ points.
- **Multipolar** ($\bar{P}=0.05$, $\bar{D}=0.95$, $\bar{A}=0.18$): ~ 10 primary dendrites radiating isotropically with a branching probability $p_b=0.7e^{-d/2}$, forming compact bushy structures; the morphology .swc file for each neuron consists of $\sim 150\text{--}250$ nodes.
- **Bipolar** ($\bar{P}=0.50$, $\bar{D}=0.18$, $\bar{A}=0.95$): two symmetric trunks along $\pm y$ -axis, each forming distal trifurcations; minimal lateral branching; the morphology .swc file for each neuron consists of $\sim 100\text{--}200$ nodes.

The reported \bar{P} , \bar{D} and \bar{A} numbers denote the empirical mean value based on g_1 , g_2 and g_3 , respectively. These rule-based generative procedures ensure that morphological features are deterministic and interpretable given the underlying gene expression.

Step 4. Embedding and normalization. All generated morphologies were embedded into a fixed-length feature representation using the CAJAL framework [10], which computes pairwise Wasserstein distances between neuronal skeletons and derives low-dimensional morphology embeddings.

S2.2 Additional details on cluster alignment loss for Patch-seq data

To construct the cluster-level prior matrix \mathbf{P} used in the alignment loss for the Patch-seq dataset, we incorporated biological prior knowledge about excitatory–inhibitory (E/I) cell type correspondence between the gene expression (GEX) and morphology modalities. The rationale behind this design is that, although the two modalities are unpaired at the single-cell level, excitatory and inhibitory neurons exhibit consistent molecular and structural characteristics across modalities. Therefore, initializing cross-modal alignment with an E/I-consistent prior provides a biologically grounded correspondence.

We first performed clustering on the GEX modality and obtained four clusters. Each cluster was annotated as excitatory (E) or inhibitory (I) based on canonical marker genes (ex: [Cite]). This yielded the GEX-type proportion matrix:

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
		0	1	0	0
π_{GEX}	E	0	1	0	0
	I	1	0	1	1

Here, each element $\pi_{\text{GEX}}(t, i)$ represents the proportion of cells of type $t \in \{\text{E}, \text{I}\}$ within GEX cluster i .

For the morphology modality, we sampled 10 representative neuron morphologies from each cluster and manually classified them as excitatory or inhibitory based on dendritic and axonal branching patterns. This yielded the morphology-type proportion matrix:

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
		0.2	0.7	0.1	0.9
π_{Morpho}	E	0.2	0.7	0.1	0.9
	I	0.8	0.3	0.0	0.1

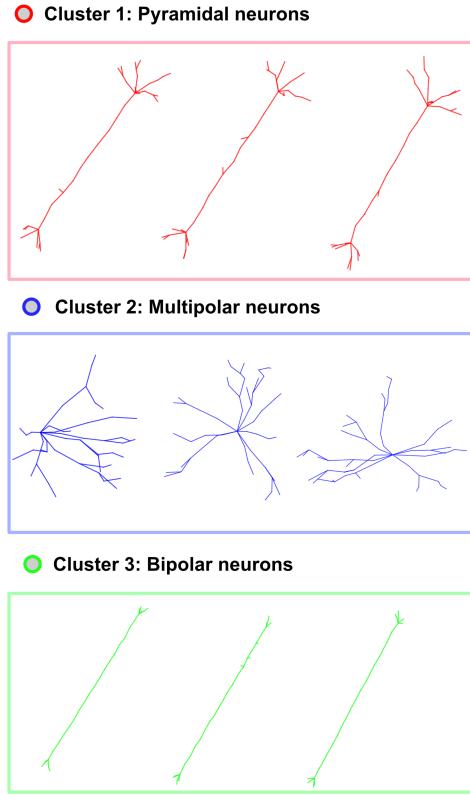


Fig. S1. Example simulated morphologies. Representative simulated neuron morphologies from three predefined clusters: Cluster 1 (red) — pyramidal neurons with elongated primary dendrites and sparse branching; Cluster 2 (blue) — multipolar neurons exhibiting complex and symmetric arborization; Cluster 3 (green) — bipolar neurons with simple, elongated structures. These synthetic morphologies were generated to mimic diverse neuronal geometries for simulation experiments.

Each element $\pi_{\text{Morpho}}(t, j)$ denotes the proportion of cells of type t in morphology cluster j , obtained from manual inspection.

Because excitatory and inhibitory neurons are expected to correspond across modalities—i.e., excitatory morphologies align to excitatory GEX profiles and inhibitory morphologies align to inhibitory ones—we defined a biological correspondence matrix:

$$R = \begin{array}{c|cc} E & I \\ \hline E & 1 & 0 \\ I & 0 & 1 \end{array}$$

The final cluster-level prior matrix \mathbf{P} was then computed as

$$\mathbf{P} = \pi_{\text{GEX}}^\top R \pi_{\text{Morpho}},$$

where each entry P_{ij} quantifies the biological correspondence between GEX cluster i and morphology cluster j based on their shared excitatory–inhibitory composition.

The resulting cluster-level prior matrix \mathbf{P} is shown below:

	Morpho Cluster 1	Morpho Cluster 2	Morpho Cluster 3	Morpho Cluster 4
GEX Cluster 1	0.80	0.30	0.90	0.10
GEX Cluster 2	0.15	0.70	0.10	1.00
GEX Cluster 3	0.80	0.30	0.90	0.10
GEX Cluster 4	0.75	0.25	0.85	0.05

Rows correspond to GEX clusters (i), and columns correspond to morphology clusters (j). Each element P_{ij} represents the prior correspondence strength between GEX cluster i and morphology cluster j , derived from the shared excitatory–inhibitory composition across modalities.

S2.3 Additional details on cluster alignment loss for 5xFAD data

To provide biologically grounded guidance for cross-modal alignment in 5xFAD microglia, we encode coarse correspondences between transcriptional states and morphology. Morphology clusters were first grouped into three canonical shapes based on visual inspection of CAJAL embeddings (cf. Fig. 5): *Ramified*, *Intermediate*, and *Amoeboid*. This yields a morphology-type proportion matrix that is an identity mapping:

$$\pi_{\text{Morpho}} = \begin{array}{c|ccc} & \text{Ramified} & \text{Intermediate} & \text{Amoeboid} \\ \hline \text{Ramified} & 1 & 0 & 0 \\ \text{Intermediate} & 0 & 1 & 0 \\ \text{Amoeboid} & 0 & 0 & 1 \end{array}.$$

Marker sets and enrichment procedure (GEX). We quantified the transcriptional composition of each GEX cluster using four microglial state marker sets: [[Cite!!!!]]

Homeostatic : $\{P2ry12, Tmem119, Cx3cr1\}$, Proliferating : $\{Mki67, Top2a, Pcn\}$,

IRM : $\{Ifit1, Ifit3, Isg15, Stat1\}$, DAM : $\{Apoe, Trem2, Axl, Cst7\}$.

Key steps. (i) Load the gene-by-cell expression matrix and the cluster labels; transpose so that genes are rows and cells are columns. (ii) For each GEX cluster, subset the corresponding cells. (iii) For each marker set, intersect with available genes, compute the mean expression across markers and cells to obtain a raw score per state. (iv) Normalize the four raw scores to percentages so that the four state proportions within each cluster sum to 100. This produces a 4×12 GEX-type proportion matrix π_{GEX} (rows: states; columns: GEX clusters 1–12).

GEX cluster compositions (percentage). The resulting π_{GEX} is:

$$\pi_{\text{GEX}} = \begin{array}{c|ccccccccccccc} & \text{H1} & \text{H2} & \text{H3} & \text{DAM 2} & \text{H4} & \text{H5} & \text{IRM} & \text{Transition} & \text{DAM 1} & \text{P1} & \text{P2} & \text{P3} \\ \hline \text{Homeostatic} & 87.98 & 93.05 & 80.85 & 35.82 & 65.37 & 79.50 & 61.79 & 76.99 & 15.62 & 46.50 & 84.38 & 48.51 \\ \text{Proliferating} & 0.97 & 0.78 & 0.79 & 0.63 & 0.87 & 0.51 & 0.73 & 1.12 & 0.48 & 19.46 & 1.97 & 36.68 \\ \text{IRM} & 0.64 & 0.64 & 0.73 & 0.60 & 1.07 & 0.77 & 11.01 & 1.62 & 0.65 & 2.62 & 1.03 & 1.64 \\ \text{DAM} & 10.41 & 5.53 & 17.63 & 62.96 & 32.69 & 19.21 & 26.47 & 20.27 & 83.25 & 31.42 & 12.62 & 13.17 \end{array}$$

Each element $\pi_{\text{GEX}}(t, i)$ denotes the percentage of state $t \in \{\text{Homeostatic, Proliferating, IRM, DAM}\}$ within GEX cluster i . In this matrix, each row represents the relative enrichment of one transcriptional state, and each column corresponds to one of the twelve GEX clusters ($H1-H5$, $DAM1-2$, $Transition$, IRM , $P1-P3$). Among these, $H1-H5$ denote five subtypes of homeostatic microglia, while $P1-P3$ represent proliferating sub-clusters. As expected, the homeostatic clusters ($H1-H5$) are predominantly enriched for homeostatic marker genes, whereas the proliferating clusters ($P1-P3$) show higher scores for proliferative markers. Similarly, the clusters $DAM1-2$ and IRM are characterized by strong enrichment of DAM- and IRM-associated signatures, respectively. Although all values in π_{GEX} are expressed as proportions—hence not large in magnitude—the enrichment pattern clearly reflects biologically meaningful state-specific compositions, indicating that the scoring procedure yields a reliable and interpretable π_{GEX} representation.

Biological mapping between transcriptional states and morphology. Consistent with prior knowledge—homeostatic microglia exhibit ramified morphology; DAM/IRM associate with amoeboid forms; proliferating microglia are typically intermediate—we encode a 4×3 state-to-morphology mapping matrix \mathbf{R} :

$$\mathbf{R} = \begin{array}{c|ccc} & \text{Ramified} & \text{Intermediate} & \text{Amoeboid} \\ \hline \text{Homeostatic} & 1 & 0.2 & 0 \\ \text{Proliferating} & 0.2 & 1 & 0.2 \\ \text{IRM} & 0 & 0.4 & 1 \\ \text{DAM} & 0 & 0.2 & 1 \end{array}.$$

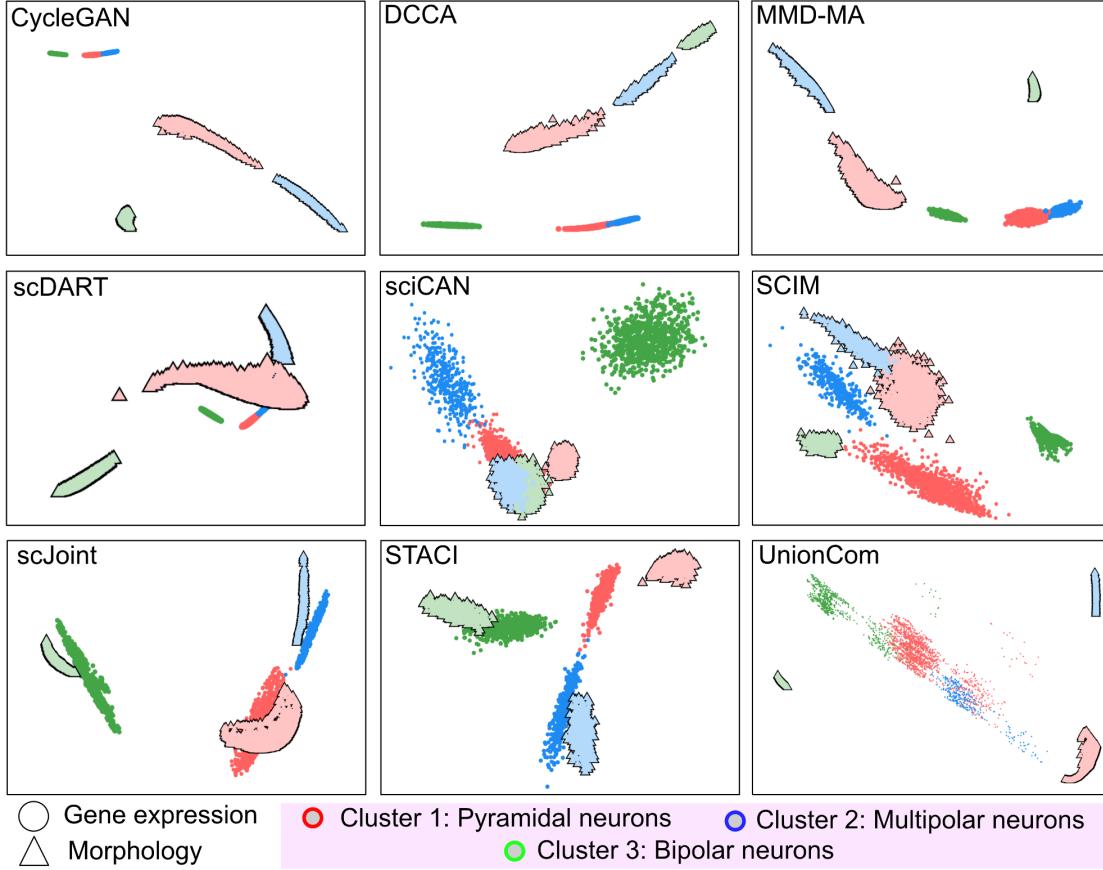


Fig. S2. Additional results of integration of simulated neurons. Integrated latent spaces obtained from simulation data using various competing cross-modal integration methods.

Prior construction. Given $\pi_{\text{GEX}} \in \mathbb{R}^{4 \times 12}$, $\mathbf{R} \in \mathbb{R}^{4 \times 3}$, and $\pi_{\text{Morpho}} \in \mathbb{R}^{3 \times 3}$, the cluster-level prior between GEX cluster i and morphology cluster j is computed by

$$\mathbf{P} = \pi_{\text{GEX}}^\top \mathbf{R} \pi_{\text{Morpho}} \in \mathbb{R}^{12 \times 3},$$

where rows correspond to GEX clusters (1–12) and columns to morphology clusters (Ramified, Intermediate, Amoeboid). Each entry P_{ij} reflects the expected correspondence strength derived from the *state-to-morphology* mapping encoded by \mathbf{R} (Homeostatic→Ramified, Proliferating→Intermediate, IRM/DAM→Amoeboid), independent of any E/I assumptions.

S3 Additional results

S3.1 Additional results on simulation

Figure S2 show the integrated latent spaces of simulated neurons produced by various baseline integration methods. Each point represents a single cell, with circles indicating gene expression embeddings and triangles indicating morphology embeddings; colors correspond to three simulated neuronal clusters. An ideal integration would mix the two modalities within each cluster while maintaining clear separation across clusters. However, most competing methods exhibit fragmented or modality-separated structures, indicating incomplete cross-modal alignment compared to the coherent integration achieved by GeoAdvAE.

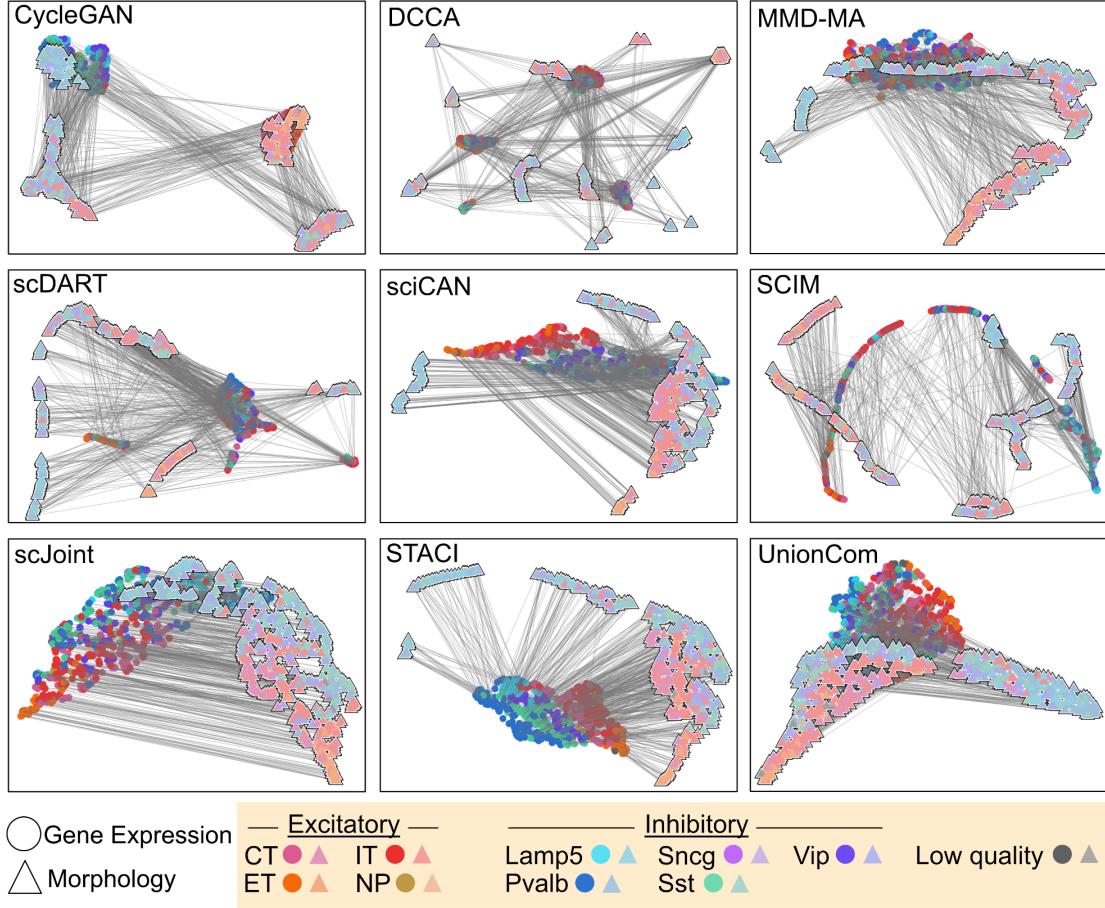


Fig. S3. Additional results of integration of Patch-seq neurons. Integrated latent spaces obtained from Patch-seq data using various competing cross-modal integration methods.

S3.2 Additional results on Patch-seq

Figure S3 show the integrated latent spaces of Patch-seq neurons produced by various baseline cross-modal integration methods. Each point represents a single cell, with circles denoting gene expression profiles and triangles denoting corresponding morphology representations. Ideally, cells of the same type should form overlapping, continuous clusters regardless of modality, indicating successful alignment. However, most competing methods yield fragmented or poorly mixed structures, highlighting the difficulty of achieving coherent cross-modal integration compared with GeoAdvAE.