TRANSMAP: METRIC EMBEDDING OF GENE RELATIONAL STRUCTURES FOR IMAGE-NATIVE SINGLE-CELL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

In single-cell omics, gene expression profiles lack canonical spatial coordinates, yet exhibit relational structure that can be conserved across species. We propose an image-native representation scheme that incorporates this structure into the input geometry for single-cell learning. The TransMap layout maps gene relationships onto a two-dimensional grid using Gromov–Wasserstein (GW) transport, with fused GW incorporating soft ortholog correspondences across species. Each selected gene occupies its own pixel, so its expression value is preserved exactly, while gene–gene distances are preserved approximately. We develop TransMap, a masked self-supervised model with a shared encoder, to learn from these images across species. Across five single-species datasets, the layouts achieve competitive distance and neighborhood preservation, while analysis across five species shows corresponding spatial organization of conserved modules. In a layout ablation across five datasets and two model sizes, GW layouts achieve lower mean masked patch reconstruction error than random layouts in all ten comparisons. Using unified human–mouse layouts constructed from training data alone, we pretrain a shared model on approximately 732K training metacells. On two of three human benchmarks, zero-shot scores fall within the range of evaluated foundation models; the same encoder also yields cell embeddings on two mouse datasets. These results support image-native representation as a promising basis for cross-species foundation models and motivate further study at larger data and model scales.

1 INTRODUCTION

Single-cell foundation models seek reusable representations of cellular state from large expression collections (Theodoris et al., 2023; Cui et al., 2024). Learning across species requires accommodating different gene repertoires within a unified framework. Cross-species methods such as SATURN and UCE use protein representations to relate genes across organisms (Rosen et al., 2024; 2026). We propose an image-native representation scheme in which gene relationships define the spatial organization of expression measurements. This scheme provides a common input format across species and a basis for developing foundation models with image-based architectures.

Gene expression is typically structured as a table with permutable gene columns, lacking any canonical sequence analogous to word order in language. Expression profiles nevertheless exhibit relational structure: genes form co-expression modules, and subsets of these modules are conserved across species (Stuart et al., 2003). Our human and mouse corpora exhibit this modular organization and cross-species correspondence (Appendix A, Figure 5). Feature-to-image methods have previously arranged tabular variables onto grids for image-based learning (Sharma et al., 2019; Zhu et al., 2021; Islam & Xing, 2023). We extend this idea to multiple species: each species retains its own selected gene set, and all species use a common grid on which gene positions reflect within-species gene relationships and cross-species orthology.

We realize this scheme by constructing the TransMap layout through metric embedding. We quantify within-species gene dependencies as pairwise distances and use Gromov–Wasserstein (GW) transport (Mémoli, 2011; Peyré et al., 2016) to seek a two-dimensional grid layout whose pixel distances approximate these gene distances. Across species, we use fused GW (Vayer et al., 2020)

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

to combine this structural objective with soft ortholog correspondences, accommodating many-to-many homology relationships. A discrete injective assignment then maps each selected gene to a distinct pixel. Each cell can then be rendered as an expression image using its species' fixed layout.

To evaluate this representation as a route to foundation modeling, we develop TransMap, a masked self-supervised model with an encoder shared across species. TransMap implements SparK-style convolutional masking (Tian et al., 2023) to predict expression in masked image patches. Alongside patch reconstruction, a species-conditioned prediction head trains the latent cell representation itself to recover masked expression. The TransMap layout thus provides a biological prior through input geometry, while predictive expression relationships are learned from data.

Layout pilots on five single-species datasets support competitive distance and neighborhood preservation, and a five-species pilot provides qualitative evidence of corresponding spatial organization of conserved modules. In a layout ablation across five datasets and two model sizes, GW layouts achieve lower mean masked patch reconstruction error than random layouts in all ten comparisons. Constructing unified human–mouse layouts from training data only, we pretrain a shared model on approximately 732K training metacells. On two of three held-out human benchmarks, zero-shot aggregate scores fall within the range of evaluated foundation models; the same encoder also yields zero-shot cell embeddings on two held-out mouse datasets.

2 Related work

Gene expression as images. DeepInsight (Sharma et al., 2019) and IGTD (Zhu et al., 2021) arrange tabular features on two-dimensional grids for convolutional learning. GenoMap (Islam & Xing, 2023) applies this idea to gene expression through a Gromov–Wasserstein formulation. Its released implementation defines image distances by differences in pixel radii (Islam & Xing, 2024), so two pixels at the same distance from the image center have zero distance under this metric, even when they lie on opposite sides of the image. This geometry supports concentric patterns, but radial proximity need not correspond to convolutional or patch neighborhoods. The TransMap layout instead uses Euclidean pixel distances to embed gene relationships across the grid.

Single-cell representations across species. scVI (Lopez et al., 2018) learns cell representations through variational modeling, while Geneformer (Theodoris et al., 2023) and scGPT (Cui et al., 2024) use transformer pretraining on gene expression. SATURN (Rosen et al., 2024) learns a shared macrogene space informed by protein embeddings, and UCE (Rosen et al., 2026) uses protein language model embeddings to support a common cell encoder across species. We use fused Gromov–Wasserstein transport (Vayer et al., 2020) to incorporate within-species gene relationships and cross-species orthology into the image input geometry.

Masked image modeling. Masked autoencoders (He et al., 2022a) learn visual representations by reconstructing image patches from partially observed inputs. SparK (Tian et al., 2023) develops this approach for convolutional networks using sparse encoding and a hierarchical decoder. TransMap uses SparK-style convolutional masking for gene-expression images, together with a shared cell readout and species-conditioned expression prediction.

3 Methods

We developed TransMap to learn cell representations from gene-expression images through masked self-supervision. Its shared encoder processes species-specific images constructed with the TransMap layout.

3.1 TransMap layout: from gene relationships to image grids

Single-cell expression has no canonical gene sequence, but co-expression reveals dependencies within each species and conserved modules across species (Szklarczyk et al., 2023; Stuart et al., 2003). Our human–mouse analysis illustrates this relational organization (Appendix A, Figure 5). These observations motivate a representation that encodes within-species gene dependencies and cross-species correspondences alongside expression values.

The TransMap layout embeds these relational spaces into a common two-dimensional grid (Figure 1). Within each species, dependencies are quantified by pairwise gene–gene distances, such as transformed Pearson correlations. Across species, homology inferred from sequence alignments and gene trees provides correspondence (Vilella et al., 2009). For multiple species, we use fused Gromov–Wasserstein (FGW) optimal transport (Vayer et al., 2020) to combine these two sources of information: its structural term favors preservation of within-species distance geometry, while its correspondence term encourages homologous genes to occupy nearby grid positions. A weight balances the structural and correspondence terms. Each species retains its own selected gene set and its own assignment on the shared grid. For a single species, the construction reduces to Gromov–Wasserstein (GW) embedding (Mémoli, 2011; Peyré et al., 2016).

An injective gene-to-pixel assignment places each selected gene at a distinct pixel, directly preserving its supplied expression value. Appendices B and C detail layout construction.

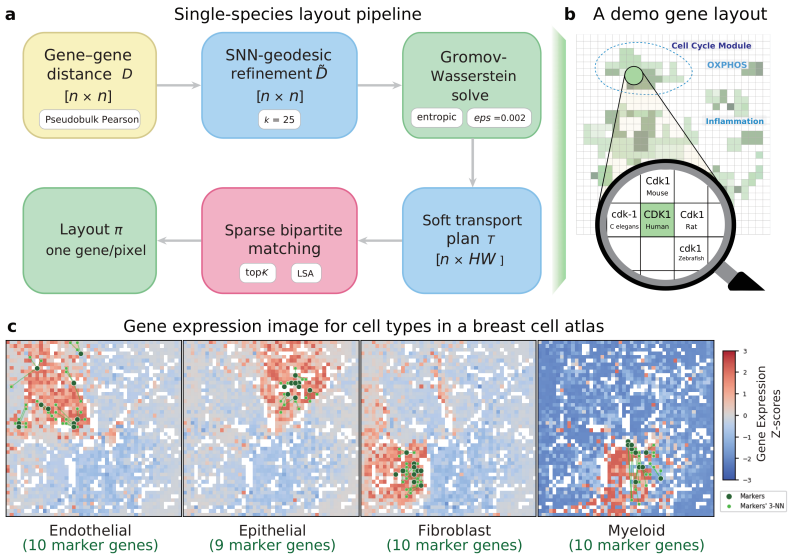

Figure 1: **Gene relationships guide the construction of expression images.** **(a)** Single-species layout construction: gene distances are refined using SNN geodesics, embedded through GW transport, and converted into a gene-to-pixel assignment by sparse bipartite matching. **(b)** An example gene layout with annotated functional regions and an ortholog detail. **(c)** Expression images for four cell types in a breast cell atlas. Colors indicate expression z-scores; green markers identify marker genes and their three nearest neighbors.

3.2 TransMap: Masked Expression Modeling

Design philosophy. TransMap uses SparK-style convolutional masking (Tian et al., 2023) for expression images (Figure 2). Reconstructing masked patches alone does not require the model to summarize a cell in a single vector. We therefore add a second head that predicts masked expression from a cell embedding and a species vector, encouraging the embedding to retain information needed for prediction.

Data and image representation. Let $\mathcal{S}$ be the species set and $\mathcal{G}_s$ the selected genes of species s. The dataset is $\mathcal{D} = \{(s_i, x_i, o_i)\}_{i=1}^{N}$, with species $s_i \in \mathcal{S}$, expression $x_i \in \mathbb{R}_{\geq 0}^{|\mathcal{G}_{s_i}|}$, and availability $o_i \in \{0,1\}^{|\mathcal{G}_{s_i}|}$. A gene is available if it appears in the source dataset's gene list, regardless of whether its count is zero. For an available gene, $x_{ig} = \log(1 + 10^4 c_{ig}/L_i)$, where c_{ig} is its raw count and $L_i > 0$ is the total count over all source genes before layout selection. Training profiles are metacells; the same representation accepts individual cells at inference.

Each fixed, injective layout $\pi_s : \mathcal{G}_s \to \mathcal{P}$ maps genes to the common grid $\mathcal{P} = \{1, \dots, H\} \times \{1, \dots, W\}$. We render $X_i \in \mathbb{R}^{2\times H\times W}$ as

$$X_i(:, p) = \begin{cases} (x_{ig}, 1)^\top, & p = \pi_{s_i}(g),\ o_{ig} = 1, \\ (0, 0)^\top, & \text{otherwise.} \end{cases} \tag{1}$$

The second channel distinguishes zero expression for an available gene from unavailable genes and unused pixels. We use $H = W = 144$ for human–mouse pretraining.

Masked encoding. The forward pass begins by partitioning the grid into 8×8 patches. A patch is occupied if it contains an available gene. Independently for each profile, we hide approximately 60% of occupied patches while retaining at least one visible and one hidden patch. Let $v_i(p) \in \{0, 1\}$ indicate membership in a visible occupied patch. Broadcasting this mask over both channels gives $\widetilde{X}_i = X_i \odot v_i$. A shared ConvNeXt-style convolutional encoder (Liu et al., 2022) produces

$$F_i = (F_{1,i}, F_{2,i}, F_{3,i}) = E_\theta(\widetilde{X}_i; v_i), \qquad F_{\ell,i} \in \mathbb{R}^{C_\ell \times (H/2^\ell) \times (W/2^\ell)}, \tag{2}$$

where $(C_1, C_2, C_3) = (128, 256, 512)$. The mask is resized and reapplied throughout the encoder, keeping nonvisible feature locations zero. Architecture and implementation details are in Appendix G.

Expression prediction and cell readout. Two branches consume the shared features. A hierarchical decoder $D_\phi(F_i; v_i) \in \mathbb{R}^{H\times W}$ fills nonvisible locations with learned mask tokens and combines features across scales. Its output gives local expression predictions $\hat{x}^{\text{recon}}_{ig} = [D_\phi(F_i; v_i)]_{\pi_{s_i}(g)}$. In parallel, a position-specific readout maps the deepest features to a cell representation:

$$z_i = \text{LayerNorm}(W_{\text{CLS}}\, \text{vec}(F_{3,i}) + b_{\text{CLS}}) \in \mathbb{R}^{512}. \tag{3}$$

This representation, denoted CLS in Figure 2, predicts expression through a second head conditioned on a learned species vector $e_{s_i} \in \mathbb{R}^{64}$:

$$\hat{x}^{\text{CLS}}_{ig} = q_g^\top [z_i; e_{s_i}] + b_g, \qquad q_g \in \mathbb{R}^{576}. \tag{4}$$

Here $[\cdot;\cdot]$ denotes concatenation and (q_g, b_g) are learned gene-specific parameters. Because this head is linear, the species vector contributes a gene-specific offset; the shared encoder and cell readout do not receive the species vector.

3.3 Training objective

For a minibatch $\mathcal{B}$, both branches predict the original normalized expression at the same masked, available gene positions,

$$\mathcal{H}_\mathcal{B} = \{(i, g) : i \in \mathcal{B},\ g \in \mathcal{G}_{s_i},\ o_{ig} = 1,\ v_i(\pi_{s_i}(g)) = 0\}. \tag{5}$$

The training objective is

$$\begin{aligned} \mathcal{L}_k &= \frac{1}{|\mathcal{H}_\mathcal{B}|} \sum_{(i,g)\in\mathcal{H}_\mathcal{B}} \left(\hat{x}^k_{ig} - x_{ig}\right)^2, \qquad k \in \{\text{recon}, \text{CLS}\}, \\ \mathcal{L} &= \mathcal{L}_{\text{recon}} + \lambda \mathcal{L}_{\text{CLS}}, \end{aligned} \tag{6}$$

with $\lambda = 1$. Each term averages over eligible sample–gene pairs, including zeros for available genes but excluding unavailable genes and unused pixels. The reconstruction loss trains the encoder and local decoder; the CLS loss directly trains the cell readout and propagates through it to the encoder. Human and mouse minibatches alternate during pretraining. At inference, all occupied patches are visible and the model returns z_i without either prediction head.

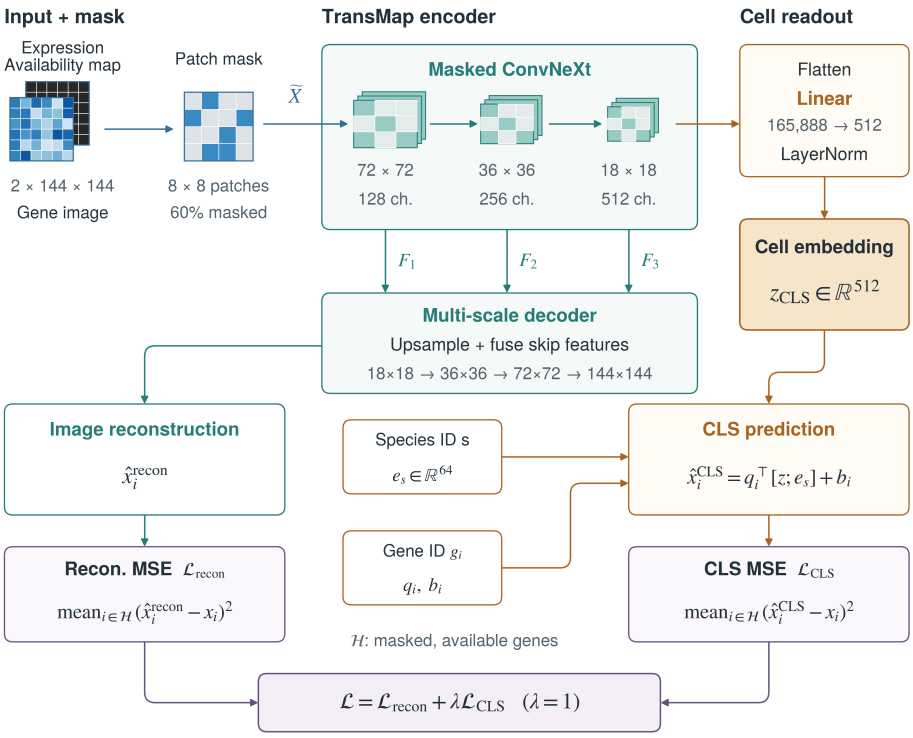

Figure 2: **Two masked-prediction heads train a shared cell encoder.** Expression and gene-availability images share a patch mask and a convolutional encoder. Multiscale features feed local expression reconstruction, while the deepest features produce a 512-dimensional CLS representation for species-conditioned expression prediction. Both losses are evaluated only at masked, available gene positions. Inference returns the CLS representation with all occupied patches visible.

4 Experiments

4.1 Pretraining corpus

We collect public human and mouse single-cell datasets spanning embryonic and adult tissues, diverse disease states, and stem-cell populations. Data resources include CELLxGENE (`https://cellxgene.cziscience.com`; CZI Cell Science Program et al., 2025) and hECA (`https://eca.xglab.tech`; Chen et al., 2022; Xi et al., 2026). We aggregate expression-similar cells within each sample and coarse lineage into metacells by summing raw counts. Pooling counts yields denser expression profiles, helping mitigate dropout and measurement noise, while reducing the number of profiles processed during pretraining (Bilous et al., 2022). The resulting compact pretraining corpus comprises 199,294 human and 591,919 mouse metacells (Figure 3a). Of these, 181,416 human and 550,601 mouse metacells (approximately 732K) form the training split (Appendix G). Dataset collection, preprocessing, and metacell construction are detailed in Appendix F.

4.2 Unified human–mouse gene layouts

We first conducted a single-species pilot on five benchmark datasets. The single-species TransMap layout achieved competitive preservation of gene-pair distance rankings and neighborhoods relative to existing layout methods, with both metrics substantially above random placement (Appendix D).

We next examined unified layouts for human (*Homo sapiens*), mouse (*Mus musculus*), zebrafish (*Danio rerio*), fruit fly (*Drosophila melanogaster*), and nematode (*Caenorhabditis elegans*) in a multi-species pilot. Even between human and *C. elegans*, whose lineages are estimated to have diverged more than 600 million years ago (Kumar et al., 2022), genes belonging to conserved func-

tional modules, including the cell cycle, proteasome, and tRNA charging, occupied corresponding grid regions (Appendix C, Figure 6).

Guided by these pilots, we built unified human–mouse layouts for pretraining (Figure 3b). We chose these two species because they are closely related and have extensive single-cell resources. Each species retains its own gene set on a shared 144×144 grid, with one gene per occupied pixel. Only training data were used to estimate gene relationships and construct these layouts; held-out expression data were excluded to avoid information leakage into layout construction. The full construction procedure and layout statistics are given in Appendix E.

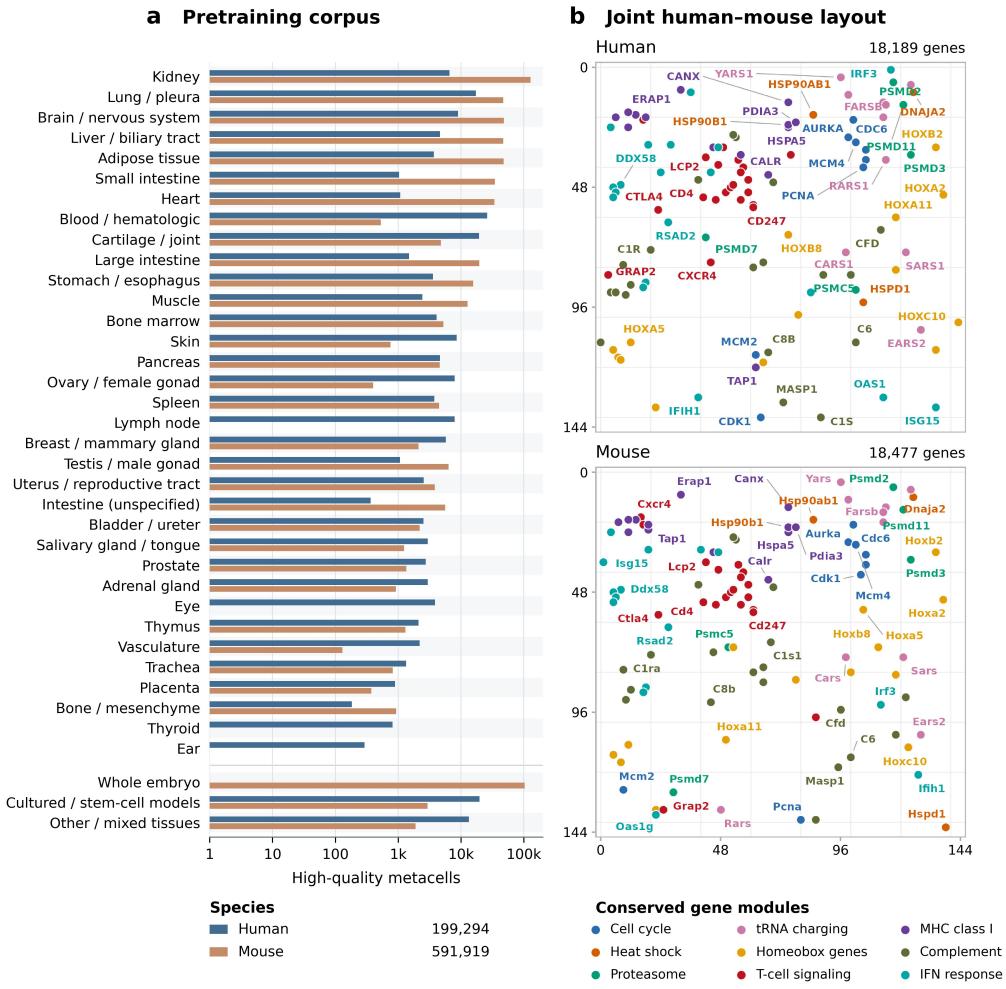

Figure 3: **A diverse human–mouse corpus supports unified gene layouts. a** High-quality metacell counts by source-organ or tissue group, with each metacell counted once (logarithmic horizontal axis; zero counts have no bar). Whole embryos and cultured or stem-cell samples are shown separately; mixed or unresolved sources are grouped as Other. These counts describe the collected corpus, not a training/validation split. **b** Human and mouse gene-to-pixel assignments on a shared 144×144 grid (18,189 and 18,477 genes, respectively). Colors identify nine conserved modules; the same 51 genes are labelled in both species. Layout construction uses training data only, combining within-species gene correlations with Ensembl Compara orthology. Corpus and layout details are in Appendices F and E.

4.3 LAYOUT ABLATION

We compare GW layouts with random layouts on five single-species datasets using approximately 458K- and 15M-parameter masked-prediction models. These smaller models make repeated layout ablations practical: we run five repeats per setting for the smaller model and three for the larger. This experiment evaluates single-species reconstruction; supplementary TransMapVI cell-representation results are in Appendix I.

We evaluate prediction at masked gene positions using the training-set mean expression of each gene as a baseline. For gene g, let $\mathcal{I}_g$ denote the test-cell instances in which the image patch containing that gene is masked. Let x_{ig} and $\hat{x}_{ig}$ denote the target and predicted expression, and let $\bar{x}_g^{\text{train}}$ be the training-set mean. The gene-level relative error is

$$r_g = \frac{\sum_{i \in \mathcal{I}_g} (x_{ig} - \hat{x}_{ig})^2}{\sum_{i \in \mathcal{I}_g} (x_{ig} - \bar{x}_g^{\text{train}})^2}. \tag{7}$$

For each run, we take the median of r_g across genes with non-negligible baseline variance; scores below one improve on the mean-expression baseline. Table 1 reports the mean and sample standard deviation across runs.

Table 1: Layout ablation: median gene-level relative MSE (mean $\pm$ sample standard deviation). GW layout denotes the single-species TransMap layout constructed with GW; Random layout denotes random gene placement. Lower is better; bold marks the better mean within each model-size pair. The mean-expression baseline has a score of one; n denotes repeated runs.

	458K parameters ($n = 5$)		15M parameters ($n = 3$)	
Dataset	GW layout	Random layout	GW layout	Random layout
Lung	$\mathbf{0.79 \pm 0.0028}$	0.87 ± 0.0044	$\mathbf{0.71 \pm 0.0013}$	0.75 ± 0.0039
Breast	$\mathbf{0.89 \pm 0.0030}$	0.95 ± 0.0041	$\mathbf{0.82 \pm 0.0028}$	0.84 ± 0.0008
CITE-seq PBMC	$\mathbf{0.98 \pm 0.0024}$	1.07 ± 0.0057	$\mathbf{0.94 \pm 0.0013}$	0.99 ± 0.0031
Gut Fetal	$\mathbf{0.85 \pm 0.0023}$	0.97 ± 0.0094	$\mathbf{0.78 \pm 0.0007}$	0.84 ± 0.0053
Lung Fetal	$\mathbf{0.72 \pm 0.0028}$	0.84 ± 0.0033	$\mathbf{0.64 \pm 0.0001}$	0.71 ± 0.0118

Across all ten dataset–model-size comparisons, GW layouts reduce mean relative reconstruction error by 2.4–14.3% compared with random placement. The reductions are smallest on Breast and largest on Lung Fetal at both model sizes.

Because masking operates on patches, changing the layout changes which genes are masked together. We have not evaluated prediction under an identical gene-level mask across layouts.

4.4 CELL EMBEDDING BENCHMARK ON OPENPROBLEMS

We evaluate the shared human–mouse encoder on five OpenProblems batch-integration datasets (Luecken et al., 2022; 2025), all absent from pretraining (Table 2). Tabula Sapiens (Tabula Sapiens Consortium, 2022) is excluded because it contributed pretraining metacells. We report scIB aggregate scores, combining biological conservation and batch correction with weights 0.6 and 0.4. Dataset sources and evaluation details are in Appendix H.

Protocol. Zero-shot evaluation uses the pretrained model without adaptation, returning the 512-dimensional CLS embedding with all occupied patches visible. TransMap receives all available layout genes; official foundation-model entries use 2,000 highly variable genes. We also evaluate per-dataset self-supervised fine-tuning with the pretraining objective, without cell-type or batch labels.

We take baseline scores from the official OpenProblems run of 17 August 2026, which reports SCimilarity (Heimberg et al., 2025), scGPT (Cui et al., 2024), and CellPLM (Wen et al., 2024) only on human datasets. PCA provides a no-integration baseline, and scVI (Lopez et al., 2018) a batch-aware reference trained separately on each dataset.

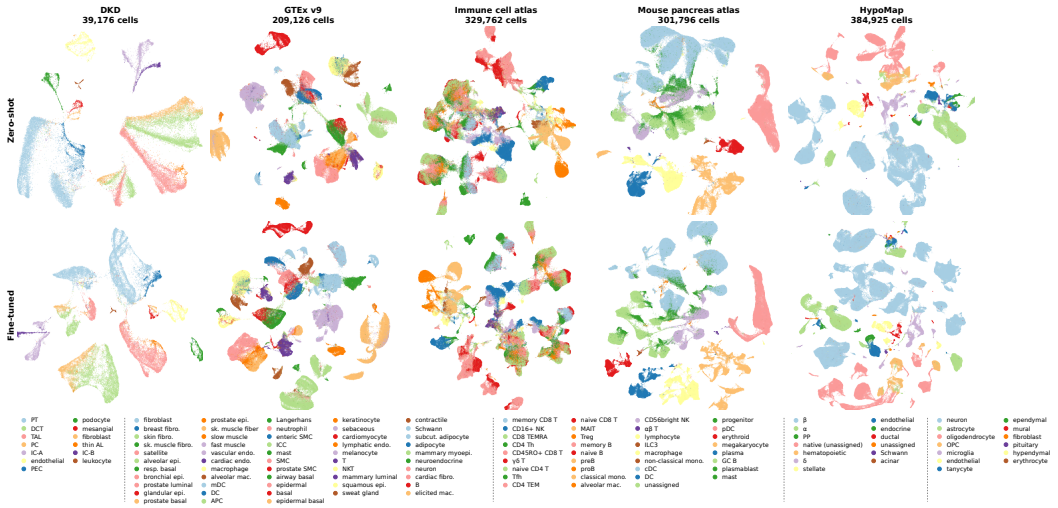

Figure 4: **Cell-type structure is visible in human and mouse embeddings.** UMAP of the 512-dimensional zero-shot (top) and fine-tuned (bottom) embeddings for the five evaluated datasets, colored by the official cell-type annotation; every annotated type is shown, with abbreviated legend names. UMAP is computed with 30 neighbors and a minimum distance of 0.3; colors are assigned per dataset from the zero-shot projection and reused for the fine-tuned row.

Table 2: OpenProblems batch-integration benchmark: scIB aggregate scores (higher is better). Rows list datasets with species and approximate numbers of individual cells. Each entry gives the overall score ($0.6 \cdot$bio $+ 0.4 \cdot$batch); bold marks the best reported zero-shot foundation-model score for each human dataset. Official scores are from the 17 August 2026 run; dashes denote unreported entries. TransMap uses full-gene input; official entries use 2,000 HVGs. scVI is trained per dataset with batch labels; PCA uses no integration. On HypoMap the official entries lack kBET, so their batch mean uses four metrics; TransMap uses all five.

	Zero-shot foundation models				Adapted	References	
Dataset	SCimilarity	scGPT	CellPLM	TransMap	TransMap fine-tuned	scVI (batch labels)	PCA (none)
DKD human, 39K cells	0.663	**0.674**	0.608	0.577	0.595	0.691	0.624
GTEx v9 human, 209K cells	0.541	**0.546**	0.545	0.545	0.538	0.592	0.560
Immune cell atlas human, 330K cells	**0.572**	0.526	0.513	0.537	0.487	0.604	0.500
Mouse pancreas atlas mouse, 302K cells	–	–	–	0.639	0.647	0.730	0.643
HypoMap mouse, 385K cells	–	–	–	0.490	0.469	0.499	0.470

Results. TransMap's zero-shot aggregate score falls within the range of the evaluated foundation models on GTEx v9 and Immune cell atlas, but below all three on DKD. On mouse pancreas it is close to PCA (0.639 versus 0.643); on HypoMap its reported aggregate score is higher than PCA's (0.490 versus 0.470), although the aggregate uses different metric sets across methods (Table 2). scVI has a higher overall score than every zero-shot entry on each dataset.

Self-supervised fine-tuning improves the aggregate score on DKD and mouse pancreas but lowers it on the other three datasets, with batch-correction scores decreasing throughout. The pretraining objective alone therefore does not consistently improve batch integration.

Figure 4 shows UMAP projections of the zero-shot and fine-tuned embeddings colored by the official cell-type labels. Without adaptation, the embeddings separate major lineages on every dataset, including the two mouse atlases, and resolve fine subdivisions such as the nephron segments of DKD and the islet endocrine types of the mouse pancreas atlas. Closely related states, such as the T-cell subsets of the Immune cell atlas and the epithelial subtypes of GTEx v9, remain intermixed, consistent with the aggregate scores. Fine-tuning yields more compact clusters without reorganizing the coarse structure.

5 DISCUSSION

This study demonstrates the feasibility of learning single-cell representations through an image-native format. A single pretrained TransMap encoder processes both human and mouse profiles.

With a 147M-parameter model trained on approximately 732K metacells, TransMap reaches the range of evaluated foundation models on two of three human benchmarks. The current evaluation does not establish state-of-the-art performance, and extending shared pretraining beyond human and mouse remains to be tested. Larger, more diverse corpora and increased model capacity will allow us to assess how well this approach scales.

AI USE STATEMENT

Claude Code (Anthropic), used as a coding agent under the authors' direction, assisted with data preparation and with implementing and running the analysis pipeline. Data preparation covered metadata harmonization, quality control, and cell-type annotation of the pretraining collections. Implementation covered metacell construction, single- and multi-species layout construction, TransMap pretraining, the layout ablation, the OpenProblems evaluation, and figure generation. OpenAI Codex/ChatGPT assisted with interpreting existing experimental results, refining the manuscript's conceptual framing and the scope of its claims, drafting and editing text, identifying relevant literature, formatting figures, and preparing the LaTeX manuscript and supporting scripts. Google Gemini 3.8 Flash, called through the Antigravity CLI, assisted with language editing of the abstract, introduction, layout motivation and construction, forward-pass, and training-loss descriptions. The authors directed these tasks and reviewed the resulting code, annotations, and outputs. AI-assisted text was checked against the figures, cited sources, and model implementation, and every reference was checked against its publisher, OpenReview, or arXiv record. The authors retain responsibility for the final manuscript, including AI-assisted code, text, claims, and other materials.

ETHICS STATEMENT

This work is a secondary computational analysis of published single-cell datasets. It involved no new data collection, participant recruitment, or animal experiments. All human data were obtained in de-identified form from public repositories. Consent and ethics approvals for the original data collection are reported in the source studies cited in Appendix F and Table A2. Arranging expression values on an image grid does not provide an additional privacy guarantee, so reuse of these data remains subject to the consent conditions and terms of use of the source datasets. The pretraining corpus covers species, tissues, and cell types unevenly, which can limit the applicability of the learned representations. The reported experiments do not establish clinical validity.

REPRODUCIBILITY STATEMENT

Appendices B and C give the distance definitions, optimization objectives, and algorithms for single- and multi-species layouts. Appendix E gives the settings of the human–mouse layout, including its fixed random seed, and Appendix A describes the co-expression module analysis. Appendix F summarizes the source collections of the pretraining corpus with their citations and describes metacell construction. Sections 3.2 and 3.3 specify the model's forward pass and training objective, and Appendix G gives the data split by collection, optimizer, learning-rate schedule, batch size, number of steps, and hardware of the pretraining run. Appendix I gives the gene selection, data splits, and optimization settings of the variational models. Section 4.4 identifies the official OpenProblems run that supplies all baseline scores, states the fine-tuning settings, and uses the official data files and metric implementations. Code, layout files, and configuration files will be released upon publication.

REFERENCES

Oren Ben-Kiki, Akhiad Bercovich, Aviezer Lifshitz, and Amos Tanay. Metacell-2: a divide-and-conquer metacell algorithm for scalable scRNA-seq analysis. *Genome Biology*, 23:100, 2022. doi: 10.1186/s13059-022-02667-1.

Mariia Bilous, Loc Tran, Chiara Cianciaruso, Aurélie Gabriel, Hugo Michel, Santiago J. Carmona, Mikael J. Pittet, and David Gfeller. Metacells untangle large and complex single-cell transcriptome networks. *BMC Bioinformatics*, 23:336, 2022. doi: 10.1186/s12859-022-04861-1.

Judith A. Blake, Richard Baldarelli, James A. Kadin, Joel E. Richardson, Cynthia L. Smith, Carol J. Bult, and the Mouse Genome Database Group. Mouse Genome Database (MGD): knowledgebase for mouse–human comparative biology. *Nucleic Acids Research*, 49(D1):D981–D987, 2021. doi: 10.1093/nar/gkaa1083.

Kristina Bubb, Tatjana Holzer, Janica L Nolte, et al. Mitochondrial respiratory chain function promotes extracellular matrix integrity in cartilage. *The Journal of biological chemistry*, 297(4): 101224, 2021. doi: 10.1016/j.jbc.2021.101224.

Sijie Chen, Yanting Luo, Haoxiang Gao, Fanhong Li, Yixin Chen, Jiaqi Li, Renke You, Minsheng Hao, Haiyang Bian, Xi Xi, Wenrui Li, Weiyu Li, Mingli Ye, Qiuchen Meng, Ziheng Zou, Chen Li, Haochen Li, Yangyuan Zhang, Yanfei Cui, Lei Wei, Fufeng Chen, Xiaowo Wang, Hairong Lv, Kui Hua, Rui Jiang, and Xuegong Zhang. hECA: The cell-centric assembly of a cell atlas. *iScience*, 25(5):104318, 2022. doi: 10.1016/j.isci.2022.104318.

Sade W. Clayton, Aimy Sebastian, Stephen P. Wilson, et al. GSE277892: Single cell rna sequencing reveals a shift in cell function and maturation of endogenous and infiltrating cell types in response to acute intervertebral disc injury. Gene Expression Omnibus dataset, 2024. URL https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE277892. Accessed 26 September 2026.

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*, 21(8):1470–1480, 2024. doi: 10.1038/s41592-024-02201-0.

CZI Cell Science Program, Shibla Abdulla, Brian Aevermann, Pedro Assis, Seve Badajoz, Sidney M. Bell, et al. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Research*, 53(D1):D886–D900, 2025. doi: 10.1093/nar/gkae1142.

C. Domínguez Conde, C. Xu, L. B. Jarvis, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594):eabl5197, May 2022. ISSN 1095-9203. doi: 10.1126/science.abl5197. URL http://dx.doi.org/10.1126/science.abl5197.

Janusz Dutkowski, Michael Kramer, Michal A. Surma, Rama Balakrishnan, J. Michael Cherry, Nevan J. Krogan, and Trey Ideker. A gene ontology inferred from molecular networks. *Nature Biotechnology*, 31:38–45, 2013. doi: 10.1038/nbt.2463.

Rasa Elmentaite, Natsuhiko Kumasaka, Kenny Roberts, et al. Cells of the human intestinal tract mapped across space and time. *Nature*, 597:250–255, 2021.

Gökcen Eraslan, Eugene Drokhlyansky, Shankara Anand, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science*, 376(6594):eabl4290, May 2022. ISSN 1095-9203. doi: 10.1126/science.abl4290. URL http://dx.doi.org/10.1126/science.abl4290.

Yue Fan, Xuzhao Bian, Xiaogao Meng, et al. Unveiling inflammatory and prehypertrophic cell populations as key contributors to knee cartilage degeneration in osteoarthritis using multi-omics data integration. *Annals of the rheumatic diseases*, 83(7):926–944, 2024. doi: 10.1136/ard-2023-224420.

Weili Fu, Sijie Chen, Runze Yang, et al. Cellular features of localized microenvironments in human meniscal degeneration: a single-cell transcriptomic study. *eLife*, 11, 2022. doi: 10.7554/eLife.79585.

Yibo Gan, Jian He, Jun Zhu, et al. Spatially defined single-cell transcriptional profiling characterizes diverse chondrocyte subtypes and nucleus pulposus progenitors in human intervertebral discs. *Bone research*, 9(1):37, 2021. doi: 10.1038/s41413-021-00163-z.

Avishai Gavish, Michael Tyler, Alissa C. Greenwald, et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature*, 618(7965):598–606, 2023. doi: 10.1038/s41586-023-06130-4.

Xiaoping Han, Renying Wang, Yincong Zhou, et al. Mapping the mouse cell atlas by microwell-seq. *Cell*, 172(5):1091–1107.e17, 2018. doi: 10.1016/j.cell.2018.02.001.

Xiaoping Han, Ziming Zhou, Lijiang Fei, et al. Construction of a human cell landscape at single-cell level. *Nature*, 581(7808):303–309, 2020. doi: 10.1038/s41586-020-2157-4.

Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184:3573–3587, 2021.

Peter W Harrison, M Ridwan Amode, Olanrewaju Austine-Orimoloye, et al. Ensembl 2024. *Nucleic Acids Research*, 52(D1):D891–D899, 2024. doi: 10.1093/nar/gkad1049.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022a. doi: 10.1109/CVPR52688.2022.01553.

Peng He, Kyungtae Lim, Dawei Sun, et al. A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell*, 185:4841–4860, 2022b.

Graham Heimberg, Tony Kuo, Daryle J. DePianto, Omar Salem, Tobias Heigl, Nathaniel Diamant, Gabriele Scalia, Tommaso Biancalani, Shannon J. Turley, Jason R. Rock, Héctor Corrada Bravo, Josh Kaminker, Jason A. Vander Heiden, and Aviv Regev. A cell atlas foundation model for scalable search of similar human cells. *Nature*, 638(8052):1085–1094, 2025. doi: 10.1038/s41586-024-08411-y.

Julian P. T. Higgins and Simon G. Thompson. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11):1539–1558, 2002. doi: 10.1002/sim.1186.

Karin Hrovatin, Aimée Bastidas-Ponce, Mostafa Bakhti, et al. Delineating mouse β-cell identity during lifetime and in diabetes with a single cell atlas. *Nature Metabolism*, 5(9):1615–1637, 2023. doi: 10.1038/s42255-023-00876-x.

Md Tauhidul Islam and Lei Xing. Cartography of genomic interactions enables deep analysis of single-cell expression data. *Nature Communications*, 14:679, 2023. doi: 10.1038/s41467-023-36383-6.

Md Tauhidul Islam and Lei Xing. GenoMap: released source code. GitHub repository, 2024. URL `https://github.com/xinglab-ai/genomap/blob/d09395a04c26c02b73c8efe9db7dbce24306305d/genomap/genomap.py`. Commit d09395a, `createMeshDistance`, lines 18–64. Accessed 26 September 2026.

Quanbo Ji, Yuxuan Zheng, Guoqiang Zhang, et al. Single-cell rna-seq analysis reveals the progression of human osteoarthritis. *Annals of the rheumatic diseases*, 78(1):100–110, 2019. doi: 10.1136/annrheumdis-2017-212863.

Shu Jia, Hongmei Liu, Tao Yang, et al. Single-cell sequencing reveals cellular heterogeneity of nucleus pulposus in intervertebral disc degeneration. *Scientific reports*, 14(1):27245, 2024. doi: 10.1038/s41598-024-78675-x.

Roy Jonker and Anton Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325–340, 1987. doi: 10.1007/BF02278710.

Tanguy Kerdoncuff, Rémi Emonet, and Marc Sebban. Sampled Gromov Wasserstein. *Machine Learning*, 2021. doi: 10.1007/s10994-021-06035-1.

Jacob C. Kimmel, Lolita Penland, Nimrod D. Rubinstein, David G. Hendrickson, David R. Kelley, and Adam Z. Rosenthal. Murine single-cell rna-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Research*, 29(12):2088–2103, 2019. doi: 10.1101/gr.253880.119.

Joe Kodama, Takeshi Oichi, Kevin J Wilkinson, et al. Apolipoprotein e is a marker of all chondrocytes in the growth plate resting zone. *Bone research*, 13(1):31, 2025. doi: 10.1038/s41413-025-00407-2.

Sudhir Kumar, Michael Suleski, Jack M. Craig, Adrienne E. Kasprowicz, Maxwell Sanderford, Michael Li, Glen Stecher, and S. Blair Hedges. TimeTree 5: an expanded resource for species divergence times. *Molecular Biology and Evolution*, 39(8):msac174, 2022. doi: 10.1093/molbev/msac174.

Tapsi Kumar, Kevin Nee, Runmin Wei, et al. A spatially resolved single-cell genomic atlas of the adult human breast. *Nature*, 620:181–191, 2023.

John E G Lawrence, Steven Woods, Kenny Roberts, et al. Single-cell transcriptomics identifies chondrocyte differentiation dynamics in vivo and in vitro. *Developmental cell*, 60(22):3066–3084.e8, 2025. doi: 10.1016/j.devcel.2025.06.031.

Pan Liu and Jingyi Jessica Li. mcRigor: a statistical method to enhance the rigor of metacell partitioning in single-cell data analysis. *Nature Communications*, 16, 2025. doi: 10.1038/s41467-025-63626-5.

Yuan Liu, Wacili Da, Ming-Jie Xu, et al. Single-cell transcriptomics reveals novel chondrocyte and osteoblast subtypes and their role in knee osteoarthritis pathogenesis. *Signal transduction and targeted therapy*, 10(1):40, 2025. doi: 10.1038/s41392-025-02136-8.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html.

Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15:1053–1058, 2018. doi: 10.1038/s41592-018-0229-2.

Malte D. Luecken, Maren Büttner, Kridsadakorn Chaichoompu, Anna Danese, Marta Interlandi, Michaela F. Müller, Daniel C. Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, and Fabian J. Theis. Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, 19:41–50, 2022.

Malte D. Luecken, Scott Gigante, Daniel B. Burkhardt, Robrecht Cannoodt, Daniel C. Strobl, Nikolay S. Markov, Luke Zappia, Giovanni Palla, Wesley Lewis, Daniel Dimitrov, et al. Defining and benchmarking open problems in single-cell analysis. *Nature Biotechnology*, 43(7):1035–1040, 2025. doi: 10.1038/s41587-025-02694-w. Batch integration task: https://openproblems.bio/benchmarks/batch_integration/.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011. doi: 10.1007/s10208-011-9093-5.

Mouse Cell Atlas. Mouse Cell Atlas: data resource. Dataset portal, n.d. URL https://bis.zju.edu.cn/MCA/. Accessed 26 September 2026.

Parse Biosciences. 5 million mouse single cell atlas from 7 tissues. Public dataset, 2025. URL https://www.parsebiosciences.com/datasets/5-million-mouse-single-cell-atlas-from-7-tissues/. Accessed 26 September 2026.

Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov–Wasserstein averaging of kernel and distance matrices. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2664–2672, 2016. URL https://proceedings.mlr.press/v48/peyre16.html.

Martin Proks, Nazmus Salehin, and Joshua M. Brickman. Deep learning-based models for preimplantation mouse and human embryos based on single-cell rna sequencing. *Nature Methods*, 22(1):207–216, 2025. doi: 10.1038/s41592-024-02511-3.

Chengxiang Qiu, Junyue Cao, Beth K. Martin, et al. Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nature Genetics*, 54(3):328–341, 2022. doi: 10.1038/s41588-022-01018-x.

Daniel Quiat, Seong Won Kim, Qi Zhang, et al. An ancient founder mutation located between robo1 and robo2 is responsible for increased microtia risk in amerindigenous populations. *Proceedings of the National Academy of Sciences of the United States of America*, 119(21):e2203928119, 2022. doi: 10.1073/pnas.2203928119.

Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):016110, 2006. doi: 10.1103/PhysRevE.74.016110.

Yanay Rosen, Maria Brbić, Yusuf Roohani, Kyle Swanson, Ziang Li, and Jure Leskovec. Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. *Nature Methods*, 21(8):1492–1500, 2024. doi: 10.1038/s41592-024-02191-z.

Yanay Rosen, Yusuf Roohani, Ayush Agrawal, Leon Samotorčan, Tabula Sapiens Consortium, Stephen R. Quake, and Jure Leskovec. Universal cell embedding provides a foundation model for cell biology. *Nature*, 656(8126):183–191, 2026. doi: 10.1038/s41586-026-10689-z.

Aimy Sebastian, Jillian L McCool, Nicholas R Hum, et al. Single-cell rna-seq reveals transcriptomic heterogeneity and post-traumatic osteoarthritis-associated early molecular changes in mouse articular chondrocytes. *Cells*, 10(6), 2021. doi: 10.3390/cells10061462.

Alok Sharma, Edwin Vans, Daichi Shigemizu, Keith A. Boroevich, and Tatsuhiko Tsunoda. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific Reports*, 9:11399, 2019. doi: 10.1038/s41598-019-47765-6.

Chenhao Shi, Yonggang Fan, Xiusheng Huang, et al. Single-cell rna sequencing reveals the differentiation and regulation of endplate cells in human intervertebral disc degeneration. *Scientific reports*, 14(1):21414, 2024. doi: 10.1038/s41598-024-71891-5.

Lisa Sikkema, Ciro Ramírez-Suástegui, Daniel C. Strobl, et al. An integrated cell atlas of the lung in health and disease. *Nature Medicine*, 29(6):1563–1577, 2023. doi: 10.1038/s41591-023-02327-2.

Lukas Steuernagel, Brian Y. H. Lam, Paul Klemm, et al. HypoMap—a unified single-cell gene expression atlas of the murine hypothalamus. *Nature Metabolism*, 4(10):1402–1419, October 2022. ISSN 2522-5812. doi: 10.1038/s42255-022-00657-y. URL http://dx.doi.org/10.1038/s42255-022-00657-y.

Joshua M. Stuart, Eran Segal, Daphne Koller, and Stuart K. Kim. A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302(5643):249–255, 2003. doi: 10.1126/science.1087447.

Hao Sun, Xingzhao Wen, Hongyi Li, et al. Single-cell rna-seq analysis identifies meniscus progenitors and reveals the progression of meniscus degeneration. *Annals of the rheumatic diseases*, 79 (3):408–417, 2020. doi: 10.1136/annrheumdis-2019-215926.

Zewen Sun, Mingyue Yan, Junjie Wang, et al. Single-cell rna sequencing reveals different chondrocyte states in femoral cartilage between osteoarthritis and healthy individuals. *Frontiers in immunology*, 15:1407679, 2024. doi: 10.3389/fimmu.2024.1407679.

Hannah Swahn, Kun Li, Tomas Duffy, et al. Senescent cell population with zeb1 transcription factor as its main regulator promotes osteoarthritis in cartilage and meniscus. *Annals of the rheumatic diseases*, 82(3):403–415, 2023. doi: 10.1136/ard-2022-223227.

Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L. Gable, Tao Fang, Nadezhda T. Doncheva, Sampo Pyysalo, Peer Bork, Lars J. Jensen, and Christian von Mering. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, 2023. doi: 10.1093/nar/gkac1000.

Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, 562(7727):367–372, 2018. doi: 10.1038/s41586-018-0590-4.

Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, 583(7817):590–595, 2020. doi: 10.1038/s41586-020-2496-1.

Tabula Sapiens Consortium. The tabula sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*, 376(6594):eabl4896, 2022. doi: 10.1126/science.abl4896.

Christina V. Theodoris, Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helene Mantineo, Elizabeth M. Brydon, Zexian Zeng, X. Shirley Liu, and Patrick T. Ellinor. Transfer learning enables predictions in network biology. *Nature*, 618:616–624, 2023. doi: 10.1038/s41586-023-06139-9.

Keyu Tian, Yi Jiang, Qishuai Diao, Chen Lin, Liwei Wang, and Zehuan Yuan. Designing BERT for convolutional networks: Sparse and hierarchical masked modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=NRxydtWup1S.

Vincent A. Traag, Ludo Waltman, and Nees Jan van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9:5233, 2019. doi: 10.1038/s41598-019-41695-z.

Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused Gromov–Wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020. doi: 10.3390/a13090212.

Albert J. Vilella, Jessica Severin, Abel Ureta-Vidal, Li Heng, Richard Durbin, and Ewan Birney. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, 19(2):327–335, 2009. doi: 10.1101/gr.073585.107.

Hanchen Wang, Jure Leskovec, and Aviv Regev. Limitations of cell embedding metrics assessed using drifting islands. *Nature Biotechnology*, 44:574–577, 2026. doi: 10.1038/s41587-025-02702-z.

Tianrui Wang, Junjie Wang, Zewen Sun, et al. Single-cell rna sequence presents atlas analysis for chondrocytes in the talus and reveals the potential mechanism in coping with mechanical stress. *Frontiers in cell and developmental biology*, 10:1047119, 2022. doi: 10.3389/fcell.2022.1047119.

Hongzhi Wen, Wenzhuo Tang, Xinnan Dai, Jiayuan Ding, Wei Jin, Yuying Xie, and Jiliang Tang. CellPLM: Pre-training of cell language model beyond single cells. In *International Conference on Learning Representations (ICLR)*, 2024.

Parker C. Wilson, Yoshiharu Muto, Haojia Wu, et al. Multimodal single cell sequencing implicates chromatin accessibility and genetic background in diabetic kidney disease progression. *Nature Communications*, 13(1):5253, September 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32972-z. URL http://dx.doi.org/10.1038/s41467-022-32972-z.

Xi Xi, Yixin Chen, Xinze Wu, Minsheng Hao, Jiaqi Li, Haiyang Bian, Qiuchen Meng, Fanhong Li, Chen Li, Chuxi Xiao, Xiaomin Dong, Renke You, Yifan Xiong, Peng Yang, Zijing Gao, Xuejian Cui, Yan Pan, Zhen Li, Wenrui Li, Zhuofeng Li, Xiaoyang Chen, Yanfei Cui, Hairong Lv, Rui Jiang, Lei Wei, and Xuegong Zhang. hECA v2.0: an AI-ready ensemble cell atlas of single-cell RNA and ATAC sequencing data. *Scientific Data*, 13:110, 2026. doi: 10.1038/s41597-025-06426-2.

Zihan Xu, Aileen Ugurbil, Joshua Kwan, et al. Single-cell multimodal profiling of pan-cancer cell lines uncovers gene regulatory principles underlying intrinsic cell states and environmental features. *Nature Communications*, 17(1):6975, 2026. doi: 10.1038/s41467-026-75360-7.

Mingyue Yan, Zewen Sun, Junjie Wang, et al. Single-cell rna sequencing reveals distinct chondrocyte states in femoral cartilage under weight-bearing load in rheumatoid arthritis. *Frontiers in immunology*, 14:1247355, 2023. doi: 10.3389/fimmu.2023.1247355.

Chenghao Zhang, Leilei Zhong, Yian Khai Lau, et al. Single cell rna sequencing reveals emergent notochord-derived cell subpopulations in the postnatal nucleus pulposus. *The FASEB Journal*, 38 (1):e23363, 2024. doi: 10.1096/fj.202301217R.

Zehao Zhang, Chloe Schaefer, Weirong Jiang, et al. A panoramic view of cell population dynamics in mammalian aging. *Science*, 387(6731):eadn3949, 2025. doi: 10.1126/science.adn3949.

Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049, 2017. doi: 10.1038/ncomms14049.

Yitan Zhu, Thomas Brettin, Fangfang Xia, Alexander Partin, Maulik Shukla, Hyunseung Yoo, Yvonne A. Evrard, James H. Doroshow, and Rick L. Stevens. Converting tabular data into images for deep learning with convolutional neural networks. *Scientific Reports*, 11:11325, 2021. doi: 10.1038/s41598-021-90923-y.

A GENE CO-EXPRESSION MODULES AND THEIR CROSS-SPECIES CORRESPONDENCE

A.1 BIOLOGICAL BACKGROUND AND INTERPRETATION

Genes participate in coordinated cellular programs, so their expression can covary across cell types and states. Co-expression provides evidence of functional association and is one of the information sources integrated by STRING (Szklarczyk et al., 2023). Biological organization also spans several scales: groups of genes contribute to protein complexes and pathways, which participate in broader cellular processes. In yeast, hierarchies inferred from molecular networks recover much of this organization (Dutkowski et al., 2013). These observations motivate representing gene relationships explicitly even though an expression table has no canonical gene sequence.

Some co-expression relationships persist across evolution. Comparative analysis of human, fly, worm, and yeast expression identified conserved modules associated with core cellular functions (Stuart et al., 2003). Conservation provides a prior for learning across species while allowing variation in gene membership and dependency strength. We use *approximate structural isomorphism* to describe this module-level correspondence: orthologous genes can participate in related programs without requiring identical expression patterns or identical gene–gene distance matrices.

Figure 5 illustrates these properties in our human and mouse corpora. The correlation blocks in panels **a** and **c** and the gene-neighbor embeddings in panels **b** and **d** reveal groups of coordinated genes. Their annotations include cell-cycle, motile-cilia, and immune-lineage programs. Panel **e** compares module membership through shared one-to-one orthologs, while panel **f** compares co-expression strength within matched gene groups. Together, these views connect within-species organization to cross-species correspondence. They motivate using conserved relationships to guide the TransMap layout while leaving species-specific expression patterns to be learned from data.

A.2 ANALYSIS OF HUMAN AND MOUSE GENE RELATIONSHIPS

Corpora. The co-expression analysis uses the metacell corpora described in Appendix F: 246 mouse and 227 human source units. Genes are restricted to per-species vocabularies of 19,131 mouse and 20,956 human genes. Features are joined on harmonized Ensembl identifiers, duplicate identifiers are summed, and genes absent from a unit's feature list are set to zero. Counts are normalized to 10^4 per metacell over the vocabulary genes and transformed as $\log(1 + x)$.

Pooled within-unit correlation. Let x_{gc} denote the normalized expression of gene g in metacell c, let $u(c)$ be the unit of c, and let $\bar{x}_{g,u}$ be the mean of x_{gc} over the metacells of unit u. The correlation between genes i and j is

$$r_{ij} = \frac{\sum_c (x_{ic} - \bar{x}_{i,u(c)})(x_{jc} - \bar{x}_{j,u(c)})}{\sqrt{\sum_c (x_{ic} - \bar{x}_{i,u(c)})^2}\,\sqrt{\sum_c (x_{jc} - \bar{x}_{j,u(c)})^2}}, \tag{8}$$

that is, a Pearson correlation over all metacells after centering every gene within its own unit. Equation (8) equals a weighted average of the per-unit correlations $r_{ij}^{(u)}$ with weights $n_u\, s_{i,u}\, s_{j,u}$, where n_u is the number of metacells and $s_{g,u}$ the standard deviation of gene g in unit u. Two properties follow. Offsets between units (sequencing platform, tissue identity, dataset) cannot correlate genes, because they are removed before the correlation; and a unit in which a gene does not vary contributes nothing to the pair, so a program restricted to a few tissues is not diluted by the many tissues in which it is silent. The matrix is computed exactly (chunked dense products in single precision, no approximation); a gene with zero pooled variance receives $r = 0$ with every other gene.

Modules. Each gene is linked to the $K = 30$ genes with which it has the largest positive correlation, edges are symmetrized and weighted by r_{ij}, and negative correlations are not used as edges. Modules are communities of this graph found by the Leiden algorithm (Traag et al., 2019) with the Reichardt–Bornholdt configuration null model (Reichardt & Bornholdt, 2006), resolution 1.0 and a fixed seed, which yields 21 mouse and 19 human modules without any size filtering. For display (Figure 5a, c), modules are ordered by average-linkage clustering of their mean correlation profiles and genes within a module by their summed within-module correlation; the human order is then permuted so that a module sits at the position of its mouse counterpart. Hub genes are the genes with the

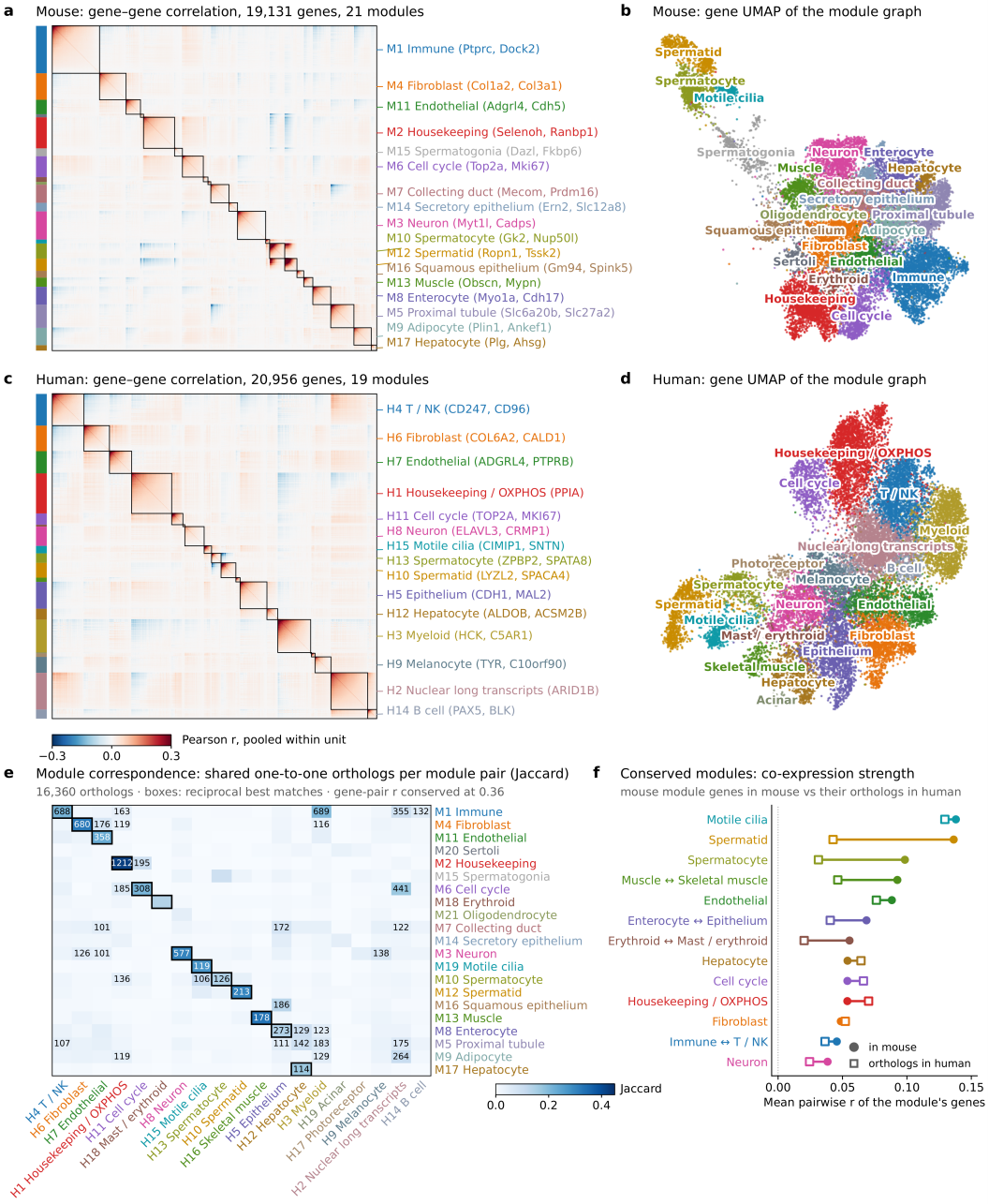

Figure 5: **Human and mouse co-expression modules show cross-species correspondence.** **(a, c)** Gene–gene Pearson correlations across metacells after centering expression within each dataset × unit. Genes are ordered by Leiden modules: 19,131 mouse genes in 21 modules and 20,956 human genes in 19 modules. **(b, d)** Gene-neighbor-graph UMAPs colored by module. **(e)** Module correspondence across 16,360 one-to-one orthologs; colors indicate Jaccard overlap, numbers give shared ortholog counts, and boxes mark reciprocal best matches. **(f)** Mean pairwise correlations for genes in 13 matched mouse modules, measured in mouse and for their orthologs in human.

largest mean correlation to the rest of their module. Module names were assigned by inspection of the hub genes and of module eigengenes (the mean per-unit z-score of the module genes, averaged within each coarse lineage). The gene UMAPs (McInnes et al., 2018) in Figure 5b, d embed the same K-nearest-neighbor graph that the modules were cut from, with connectivities r_{ij}, distances $1 - r_{ij}$, and $\min_\text{dist} = 0.8$, so the picture shows the neighborhoods the community detection saw.

Per-pair reliability. As an annotation of every gene pair we also compute, separately, how many units support it and how consistent they are. For each unit with at least 100 metacells (218 mouse and 197 human units, covering more than 99% of the metacells) we compute the unit's own Pearson matrix and a measurability mask: gene g is measurable in unit u when it is detected in at least 1% of the unit's metacells and $s_{g,u} \geq 0.1\, s_g$, with s_g the global standard deviation. A unit *votes* on pair (i, j) when both genes are measurable in it. Voting correlations are Fisher-transformed, $z_{ij}^{(u)} = \operatorname{artanh}(r_{ij}^{(u)})$, and combined with weights $w_u = n_u - 3$; Cochran's Q gives the heterogeneity $I^2 = \max\{0, (Q-(V-1))/Q\}$ for a pair with $V \geq 3$ votes (Higgins & Thompson, 2002). Votes and I^2 are not used to build the modules: an average over units would dilute tissue-restricted programs, which is exactly what Equation (8) avoids. They are reported alongside the modules as evidence counts (a pair measurable in 7 units and one measurable in 200 units carry different weight), and a high I^2 inside a module indicates a tissue-restricted program rather than an unreliable one.

Cross-species correspondence. Mouse and human genes are matched through the MGI mouse–human homology classes (Blake et al., 2021). We retain classes with exactly one gene per species, yielding 18,782 one-to-one pairs. Of these, 16,360 have both genes in the vocabularies, covering 86% of the mouse and 78% of the human vocabulary. We compare modules at both the module and gene-pair levels. For mouse module A and human module B restricted to shared orthologs, we compute

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{9}$$

Figure 5e shows $J(A, B)$ together with the shared ortholog count $|A \cap B|$. Modules A and B form a *reciprocal best match* when B maximizes $J(A, \cdot)$ and A maximizes $J(\cdot, B)$. We identify such matches for 13 of the 21 mouse modules. Across the 1.34×10^8 gene pairs formed by shared orthologs, corresponding mouse and human correlation coefficients correlate at $r = 0.36$. Of the 1% most correlated mouse pairs, 22% are also among the 1% most correlated human pairs (1% expected by chance). Figure 5f reports the mean pairwise correlation within each matched mouse module and among its orthologs in human. Colors are shared between reciprocally matched modules in every panel; species-specific modules are drawn in muted colors.

Relation to the layout construction. The pooled within-unit correlation can be converted into a dissimilarity with the same transform as Equation (11), $D_{ij} = (1 - r_{ij})/2$, and used as the input of Appendices B and C. The human–mouse layout of Figure 3 does so (Appendix E); Table A1 uses pseudobulk Pearson distances and the five-species layout uses STRING-derived distances. Figure 5 itself is evidence for module-level conservation of the relational structure used to construct the TransMap layout, not a measurement of layout fidelity.

B SINGLE-SPECIES LAYOUT CONSTRUCTION

Notation and overview. Let $G = \{g_1, \ldots, g_n\}$ be the selected gene set and $D \in \mathbb{R}_{\geq 0}^{n \times n}$ its symmetric dissimilarity matrix. Let $P = \{p_1, \ldots, p_m\}$ be an $H \times W$ pixel grid with $m = HW \geq n$, equipped with Euclidean distances $D_{uv}^{\text{grid}} = \|p_u - p_v\|_2$. We seek an injective assignment

$$\pi : G \longrightarrow P, \qquad \pi(g_i) \neq \pi(g_j) \quad \text{for } i \neq j, \tag{10}$$

that approximately preserves gene relationships in the grid. Algorithm 1 summarizes the construction.

Gene dissimilarities. We support data-derived and reference-network distances. For the data-derived construction, expression is aggregated into pseudobulks to obtain $X \in \mathbb{R}^{n \times b}$, where b is the

Algorithm 1 Single-species TransMap layout construction

Require: Gene dissimilarities D; grid (H, W) with $HW \geq n$; SNN neighborhood size k; entropy weight ε; initial candidate count K_0.
Ensure: Injective gene-to-pixel assignment π.
1: $\widetilde{D} \leftarrow \mathrm{SNNgeodesic}(D, k)$.
2: Compute Euclidean pixel distances D^{grid} on the $H \times W$ grid.
3: Obtain $T^\star$ by solving the entropic GW problem in Equation (13).
4: $\pi \leftarrow \mathrm{InjectiveAssign}(T^\star, K_0)$ using adaptive sparse matching.
5: **return** π.

number of pseudobulks. For gene pairs with defined Pearson correlations, we set

$$r_{ij} = \frac{(X_{i,:} - \bar{X}_i)(X_{j,:} - \bar{X}_j)^\top}{\|X_{i,:} - \bar{X}_i\|_2 \, \|X_{j,:} - \bar{X}_j\|_2}, \qquad D_{ij} = \frac{1 - r_{ij}}{2}, \qquad D_{ii} = 0. \tag{11}$$

This quantifies expression association; the raw correlation dissimilarity need not satisfy the triangle inequality. For the reference-network construction, a STRING co-expression score $s_{ij} \in [0, 1]$ gives $D_{ij} = 1 - s_{ij}$, with $D_{ij} = 1$ for unrecorded off-diagonal pairs (Szklarczyk et al., 2023). The single-species fidelity experiments use pseudobulk Pearson distances, whereas the five-species layout uses STRING-derived distances within each species.

SNN-geodesic refinement. Let $N_k(i)$ contain the k nearest genes to i under D, excluding i itself. On each directed neighbor edge (i, j), define $c_{ij} = |N_k(i) \cap N_k(j)|$ and assign length $1/c_{ij}$ when $c_{ij} > 0$, or length 2 otherwise. We symmetrize by retaining an undirected edge whenever either direction exists, taking the smaller length when both exist. All-pairs shortest paths give d^{geo}, which is normalized as

$$\widetilde{D}_{ij} = \frac{d^{\mathrm{geo}}_{ij}}{\max_{a,b} d^{\mathrm{geo}}_{ab}}, \qquad \widetilde{D}_{ii} = 0. \tag{12}$$

This construction incorporates paths through shared gene neighborhoods. The normalization requires finite shortest-path distances; SNN graph construction alone does not guarantee connectivity. The original layout procedure used $k = 25$; the fidelity benchmarks reported in Table A1 use $k = 7$.

GW optimization. Let $\mu_i = 1/n$, $\nu_u = 1/m$, and $\Pi(\mu, \nu) = \{T \geq 0 : T\mathbf{1} = \mu, \, T^\top \mathbf{1} = \nu\}$. Following the GW formulation (Mémoli, 2011) and its entropic regularization (Peyré et al., 2016), we obtain the soft gene-to-pixel coupling from

$$T^\star \in \operatorname*{arg\,min}_{T \in \Pi(\mu,\nu)} \mathcal{L}_{\mathrm{GW}}(T) - \varepsilon \mathcal{H}(T), \tag{13}$$

$$\mathcal{L}_{\mathrm{GW}}(T) = \sum_{i,j=1}^{n} \sum_{u,v=1}^{m} (\widetilde{D}_{ij} - D^{\mathrm{grid}}_{uv})^2 T_{iu} T_{jv}, \tag{14}$$

where $\mathcal{H}(T) = -\sum_{i,u} T_{iu}(\log T_{iu} - 1)$. The implementation uses sampled GW updates with Sinkhorn projections (Kerdoncuff et al., 2021), with $\varepsilon = 2 \times 10^{-3}$. These updates numerically optimize the nonconvex objective.

Injective assignment and rendering. The soft coupling is converted into a discrete layout by sparse bipartite matching (Jonker & Volgenant, 1987). For each gene, retain its top-K pixels under $T^\star$ and solve a minimum-cost full matching with edge costs $-T^\star_{iu}$. Start at $K = K_0$; if no full matching exists, expand $K \leftarrow \min(4K, m)$ and retry. At $K = m$, a full matching exists because $m \geq n$. The resulting π satisfies Equation (10). For a cell with expression vector x, define its image by $I_{\pi(g)} = x_g$, setting unused pixels to zero. Reading the assigned pixels recovers the selected expression values exactly. This property concerns rendering; the distance embedding and the rounding of the soft coupling can introduce geometric distortion.

C Cross-species layout construction

Species-specific structure and cross-species correspondence. For S species, retain a gene set G_s and refined distance matrix $\widetilde{D}_s$ for each species s. All species use one grid P with $m = HW \geq$

$\max_s |G_s|$, but have separate injective assignments $\pi_s : G_s \to P$. Cross-species information is supplied as weighted ortholog edges $E = \{(s, g, s', g', w) : s \neq s'\}$. We use Ensembl Compara homology calls, inferred through sequence comparison, multiple alignment, and gene-tree reconciliation (Vilella et al., 2009), retaining one-to-one, one-to-many, and many-to-many ortholog relationships.

Ortholog weights. Let $d_g^{s \to s'}$ denote the number of ortholog partners of $g \in G_s$ in species s'. An edge is weighted by

$$w(s, g, s', g') = \frac{1}{\sqrt{d_g^{s \to s'} d_{g'}^{s' \to s}}}. \tag{15}$$

This gives weight 1 for a one-to-one pair, $1/\sqrt{2}$ for each edge in a one-to-two relationship, and $1/2$ for each edge in a complete two-to-two group. It attenuates individual edges in expanded gene families without requiring a one-to-one ortholog restriction.

FGW update. Given current anchor coordinates $a_{s'}(g') \in P$ for other placed species, define the cost of placing gene $g \in G_s$ at pixel p_u as

$$C_s(g, u) = \sum_{(s,g,s',g',w) \in E:\, s' \in \mathcal{A} \setminus \{s\}} w \, \|p_u - a_{s'}(g')\|_2^2, \tag{16}$$

where $\mathcal{A}$ is the set of species with current anchors. The per-species fused GW update (Vayer et al., 2020) is

$$T_s \in \operatorname*{arg\,min}_{T \in \Pi(\mu_s, \nu)} \alpha \mathcal{L}_{\mathrm{GW},s}(T) + (1 - \alpha)\langle C_s, T \rangle - \varepsilon \mathcal{H}(T), \tag{17}$$

with $\mu_{s,g} = 1/|G_s|$, $\nu_u = 1/m$, and $\mathcal{L}_{\mathrm{GW},s}$ defined by Equation (14) using $\widetilde{D}_s$. We use $\alpha = 0.3$ and $\varepsilon = 2 \times 10^{-3}$. The structural and correspondence terms allow a trade-off between preserving each species' internal relationships and bringing orthologs together. Genes without an available ortholog anchor have zero correspondence cost and are placed through the structural term.

Algorithm 2 Cross-species TransMap layout construction

Require: Refined distances $\{\widetilde{D}_s\}_{s=1}^S$; ortholog edges E; shared grid (H, W); α, ε, K_0; maximum outer iterations $T_{\max}$.
Ensure: Per-species injective assignments $\{\pi_s\}_{s=1}^S$ on the shared grid.

```
Compute D^grid; choose the largest species s_0.
Solve pure GW for s_0 (α = 1); set a_{s_0}(g) ← p_{arg max_u T_{s_0}(g,u)}.
A ← {s_0}.
for t = 1, ..., T_max do
    for each species s do
        Build C_s from available anchors using Equation (16).
        Update T_s using Equation (17).
        a_s(g) ← p_{arg max_u T_s(g,u)} for each g; A ← A ∪ {s}.
    end for
    if the stopping criterion is satisfied then
        break
    end if
end for
π_s ← InjectiveAssign(T_s, K_0) for each species s.
return {π_s}_{s=1}^S.
```

Alternating optimization and final layouts. Algorithm 2 initializes the largest species with pure GW and then alternates species updates, rebuilding the correspondence costs from the latest anchors. Intermediate row-wise maxima provide provisional anchor coordinates and can assign multiple genes to the same pixel. Final sparse matching enforces injectivity separately within each species. The reported five-species construction uses eight outer iterations on a 141×141 grid. Ortholog proximity is a soft preference: corresponding genes can occupy different pixels, and species-specific relational structure remains part of every update.

Multi-species layout pilot. Figure 6 shows this construction for human, mouse, zebrafish, fruit fly, and nematode, with STRING-derived within-species distances (Szklarczyk et al., 2023) and Ensembl Compara orthologs. Genes of nine conserved modules are drawn at their assigned pixels in each species. The modules shared by all five species (cell cycle, heat shock, proteasome, tRNA charging) occupy corresponding regions in every panel, although the human and *C. elegans* lineages are estimated to have diverged more than 600 million years ago (Kumar et al., 2022); T-cell signaling, complement, and interferon-response genes have no orthologs in the two invertebrates and are absent from their layouts. Each species keeps its own gene set and assignment; the ortholog term of Equation (17) couples their positions.

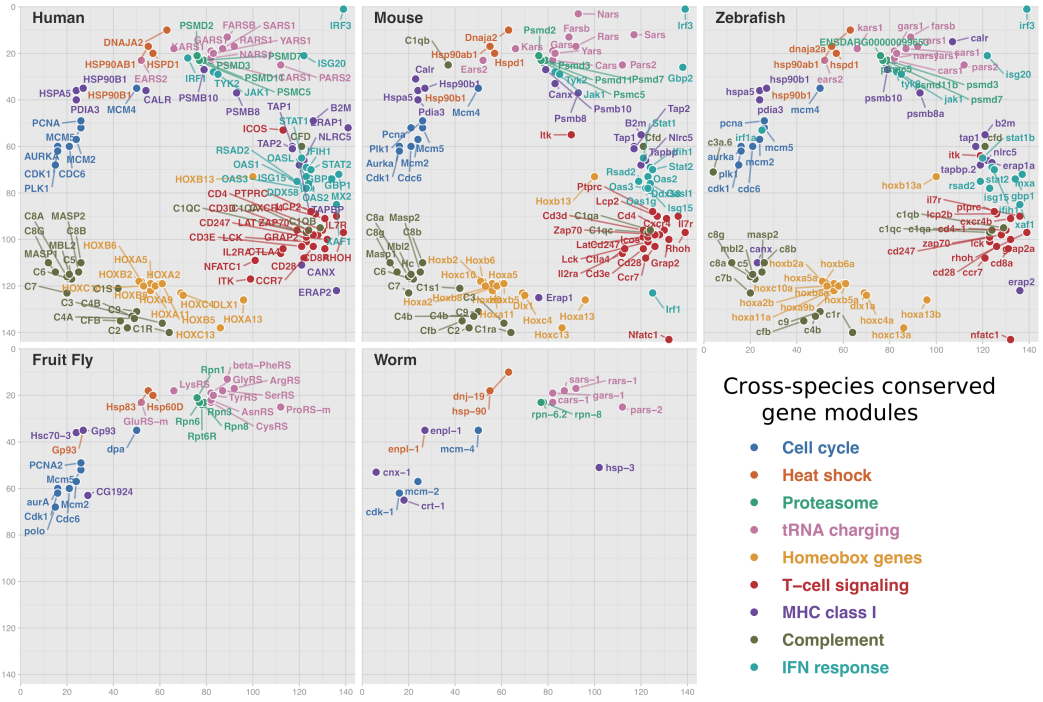

Figure 6: **Orthology-guided layouts place conserved modules in corresponding regions.** Selected genes from human, mouse, zebrafish, fruit fly, and nematode are shown at their assigned positions on the shared grid, colored by annotated functional group. Each species has its own gene-to-pixel assignment, coupled to other species through ortholog relationships; within-species distances are STRING-derived.

D SINGLE-SPECIES LAYOUT PILOT

D.1 FIDELITY TO SNN-GEODESIC DISTANCES

In the single-species pilot, we evaluate layouts on Lung, Breast, CITE-seq PBMC, Gut Fetal, and Lung Fetal, comparing the single-species TransMap layout with random placement, IGTD (Zhu et al., 2021), and DeepInsight (Sharma et al., 2019). We use SNN-geodesic distances with $k = 7$ as the common evaluation reference, since they define the gene relational structure targeted by the TransMap layout construction. For each layout, we compare these reference distances with the Euclidean distances between the assigned pixels. GenoMap's radial construction is discussed in Section 2 and is outside the scope of this Euclidean layout-fidelity comparison.

We assess two aspects of fidelity to the SNN-geodesic reference. Spearman correlation (ρ) measures agreement between the rankings of all gene-pair distances in the reference and the layout. It evaluates the preservation of overall distance ordering. R_{NX} AUC measures neighborhood preservation: for each gene, the overlap between its nearest-neighbor sets in the two spaces is corrected

for chance agreement and summarized across neighborhood sizes. All layouts are evaluated with the same implementation of this score.

Table A1: Gene-layout fidelity across five datasets. Entries report Spearman ρ / R_{NX} AUC against SNN-geodesic distances ($k = 7$). Ours denotes the GW-based TransMap layout. Higher is better; bold marks the best score for each metric within each dataset.

Dataset	Random	IGTD	DeepInsight	Ours
Lung	0.00 / 0.01	0.41 / 0.20	0.62 / **0.47**	**0.74** / 0.39
Breast	0.00 / 0.01	0.39 / 0.19	0.56 / **0.46**	**0.74** / 0.36
CITE-seq PBMC	0.00 / 0.01	0.48 / 0.23	0.53 / **0.45**	**0.73** / 0.38
Gut Fetal	0.00 / 0.01	0.45 / 0.22	0.55 / **0.43**	**0.72** / 0.38
Lung Fetal	0.00 / 0.01	0.38 / 0.20	0.56 / **0.48**	**0.72** / 0.42

Across all five datasets, the TransMap layout achieves the highest Spearman correlation (0.72–0.74) and outperforms IGTD on both metrics (Table A1). Its R_{NX} AUC (0.36–0.42) is lower than DeepInsight's (0.43–0.48). These results show competitive layout fidelity relative to the evaluated gene-layout methods, with an advantage in preserving overall SNN-geodesic distance ordering.

E Unified human–mouse layout construction

This appendix documents the two-species layout of Figure 3. Only training data are used to estimate the within-species gene relationships and construct the final layouts; held-out expression profiles do not enter this procedure. It applies the cross-species construction of Appendix C to human and mouse, with the pooled within-unit correlations of Appendix A as the within-species structure. It is independent of the five-species layout of Figure 6, which uses STRING-derived distances; the algorithm and its settings are the same.

Gene sets and distances. The gene sets are the vocabularies of Appendix A (20,956 human and 19,131 mouse genes) after removing genes whose Ensembl release 111 biotype (Harrison et al., 2024) is a long non-coding RNA or a pseudogene (2,767 human and 654 mouse genes, almost all lncRNA), leaving $|G_{\mathrm{hs}}| = 18{,}189$ and $|G_{\mathrm{mm}}| = 18{,}477$. The correlation matrix of Equation (8) is converted with $D_{ij} = (1 - r_{ij})/2$, the transform of Equation (11); every coefficient is defined, and a gene with zero pooled variance, which has $r = 0$, is at distance $1/2$ from every other gene. Each D_s is then refined by the SNN geodesic of Equation (12) with $k = 25$. Neighbor edges without a shared neighbor ($c_{ij} = 0$) receive length $2 + D_{ij}$ instead of the constant 2, so that structurally isolated genes stay ordered by their raw dissimilarity; all shortest paths were finite for both species, and each $\widetilde{D}_s$ is normalized to $[0, 1]$.

Ortholog edges. Cross-species edges are Ensembl Compara release 111 orthologs (Vilella et al., 2009; Harrison et al., 2024) flagged as high confidence, retaining one-to-one, one-to-many, and many-to-many relationships and weighted by Equation (15). The table holds 17,108 human–mouse ortholog edges; an edge whose genes are not both in the layouts contributes neither to the cost of Equation (16) nor to the evaluation below.

Optimization. Both species share a 144×144 grid ($m = 20{,}736 \geq \max_s |G_s|$). Following Algorithm 2, the larger gene set (mouse) is initialized by pure GW, and every outer iteration then updates human and mouse in that order with $\alpha = 0.3$ and $\varepsilon = 2 \times 10^{-3}$. Each update solves Equation (17) with the sampled-GW solver of Appendix B in its semi-relaxed form: the gene marginal μ_s is enforced exactly and the pixel marginal ν through a Kullback–Leibler penalty of weight 1, with at most 500 iterations, tolerance 10^{-5}, and single-precision inner Sinkhorn iterations. Provisional anchors are the row-wise maxima of T_s. The outer loop stops when fewer than $\tau = 10$ genes change pixel over a full iteration; the criterion was met at the fifth of at most five iterations, where no gene moved. The final injective assignment applies the sparse matching of Appendix B to the last coupling of each species with $K_0 = 12{,}800$ candidate pixels per gene, which admits a full matching without expansion. Every gene occupies its own pixel, and the two layouts fill 87.7% and 89.1% of the grid. A fixed random seed makes the run reproducible.

Canonical pose. The GW objective is invariant to the eight symmetries of the square grid (rotations and reflections), so independent runs may return equivalent layouts in different orientations. We fix the pose by the human positions of three anchor genes, ACTB, MKI67, and GAPDH: among the eight images of the layout we keep the one whose concatenated anchor coordinates are lexicographically smallest (here a vertical flip) and apply the same transformation to both species, which leaves all cross-species comparisons unchanged.

Conserved modules. Immune modules (MHC class I, T-cell signaling, interferon response) occupy corresponding regions in both layouts, as do proteasome, tRNA-charging, and heat-shock genes. Homeobox and complement genes are more broadly distributed over the grid in both species.

Ortholog alignment. For an ortholog pair (g, g') with both genes placed, let $\delta(g, g') = \|\pi_{\mathrm{hs}}(g) - \pi_{\mathrm{mm}}(g')\|_\infty$ be the Chebyshev distance between their pixels. The fraction of pairs with $\delta \le r$ is 0.495 for $r = 0$ (same pixel), 0.639 for $r \le 1$, 0.646 for $r \le 2$, 0.648 for $r \le 5$, and 0.652 for $r \le 8$. Because every pixel holds one gene, this rate cannot be inflated by several genes collapsing onto one pixel. Ortholog proximity remains a soft preference: about a third of the pairs are placed more than eight pixels apart, where the structural term of each species has prevailed.

Figure 3. The displayed genes belong to the nine conserved modules of Figure 6: cell cycle, heat shock, proteasome, tRNA charging, homeobox, T-cell signaling, MHC class I, complement, and interferon response. The selection comprises 115 human genes and their Ensembl mouse orthologs. All but MBL2, which is absent from the human vocabulary, are placed. We label the same genes in both panels. Among genes placed in both species, we select at most six per module, favoring those with the fewest other displayed genes within a Chebyshev radius of 10 pixels in either panel. This gives 51 labelled genes. All selected genes present in the layouts are plotted as points, with labels using species-native symbols.

F Atlas corpora and metacell construction

Table A2 provides an overview of publicly documented datasets in the atlas collection. A unit is a released tissue sample or collection from one dataset. The resulting metacells also supply the gene correlations of Appendix A and the human–mouse layouts in Figure 3.

Why metacells. Aggregating expression-similar cells produces denser profiles, mitigating dropout and measurement noise in individual cells (Bilous et al., 2022). It also reduces the number of profiles processed during pretraining, allowing us to retain broad biological coverage in a compact corpus.

Metacell construction. After quality control and cell-type annotation, we construct metacells by combining SuperCell-based grouping (Bilous et al., 2022), an outlier-filtering rule from Metacell-2 (Ben-Kiki et al., 2022), and heterogeneity checks based on mcRigor (Liu & Li, 2025). Raw counts are summed within groups restricted to the same sample and coarse cell lineage. The resulting corpus contains 199,294 human and 591,919 mouse metacells.

Figure 3a groups source units by organ, tissue, or sample source, counting each metacell once. Mixed or unresolved sources are assigned to Other; the plotted totals describe the collected corpus before the training/validation split.

Dataset sources. Table A2 cites the source atlases, collection papers, or public dataset releases. Mouse Cell Atlas version numbers refer to releases from the MCA portal (Mouse Cell Atlas, n.d.). The cartilage collection pools multiple studies: 13 human units (Ji et al., 2019; Sun et al., 2020; Fu et al., 2022; Wang et al., 2022; Swahn et al., 2023; Yan et al., 2023; Sun et al., 2024; Shi et al., 2024; Quiat et al., 2022; Gan et al., 2021; Jia et al., 2024; Lawrence et al., 2025; Fan et al., 2024) and six mouse units (Sebastian et al., 2021; Bubb et al., 2021; Kodama et al., 2025; Zhang et al., 2024; Clayton et al., 2024; Liu et al., 2025). For GSE277892, we cite the deposited dataset because its GEO record does not link a source paper.

Table A2: Overview of publicly documented datasets in the atlas collection. Units are released tissue samples or collections; cell and metacell counts are rounded to the nearest thousand (K), with counts below 1K shown as < 1. This table is a source overview, not an exhaustive inventory or a training/validation split. Dataset names follow the processing pipeline.

Species	Dataset	Source	Units	Cells (K)	Metacells (K)
Human	3CA	Gavish et al., 2023	105	1474	94
Human	Tabula Sapiens (droplet)	Tabula Sapiens Consortium, 2022	32	720	45
Human	Cartilage collection	Pooled studies in text	13	361	23
Human	Human Cell Landscape	Han et al., 2020	57	224	15
Human	EasySci-RNA	Xu et al., 2026	1	201	13
Human	PBMC 68k	Zheng et al., 2017	1	56	4
Human	Tabula Sapiens (Smart-seq2)	Tabula Sapiens Consortium, 2022	15	12	< 1
Mouse	PanSci	Zhang et al., 2025	67	5759	377
Mouse	TOME	Qiu et al., 2022	20	1395	94
Mouse	Mouse Cell Atlas 3.0	Han et al., 2018	29	590	39
Mouse	Mouse Cell Atlas 2.0	Han et al., 2018	11	296	19
Mouse	Parse 5M	Parse Biosciences, 2025	4	235	17
Mouse	Mouse Cell Atlas 1.1	Han et al., 2018	30	190	13
Mouse	Tabula Muris Senis (droplet)	Tabula Muris Consortium, 2020	16	168	11
Mouse	Cartilage collection	Pooled studies in text	6	78	5
Mouse	Tabula Muris Senis (FACS)	Tabula Muris Consortium, 2020	23	59	4
Mouse	3CA (mouse models)	Gavish et al., 2023	4	48	3
Mouse	Tabula Muris (droplet)	Tabula Muris Consortium, 2018	12	46	3
Mouse	Calico aging atlas	Kimmel et al., 2019	3	43	3
Mouse	Tabula Muris (FACS)	Tabula Muris Consortium, 2018	20	41	3
Mouse	Preimplantation atlas	Proks et al., 2025	1	2	< 1

G PRETRAINING DETAILS

The human–mouse model evaluated in Section 4.4 has 147,089,275 trainable parameters. Its encoder uses 3/3/18 ConvNeXt blocks per stage; the decoder has 64 channels. Input, masking, readout, and loss definitions follow Sections 3.2–3.3.

Encoder, decoder, and readout. The three encoder stages produce feature maps of 72×72, 36×36, and 18×18 pixels. The implementation uses dense convolutions with explicit masks, resized by nearest-neighbor interpolation and reapplied after every downsampling operation and ConvNeXt block. The local decoder projects each scale to 64 channels through a 1×1 convolution and replaces nonvisible features with a learned mask token specific to that scale. Starting at the coarsest scale, a decoder block is followed by bilinear upsampling and addition of the next finer scale's token-filled features before the next block. Final upsampling and a convolution produce the expression image. The CLS readout is a learned linear projection from the flattened 165,888-dimensional deepest feature map to 512 dimensions, followed by LayerNorm; CLS denotes its output, not an additional input token. The expression-prediction parameters (q_g, b_g) are shared only for identical gene identifiers across layouts; orthology alone does not tie these parameters.

Data and sampling. The corpus of Appendix F is split by collection: 181,416 human and 550,601 mouse metacells are used for training, 12,742 human and 20,990 mouse for validation, and 5,136 human and 20,328 mouse are held out as a test split that is not used in this paper. Whole collections, including related atlas releases, are assigned to one split; we do not claim donor independence across collections. Each optimizer step uses a minibatch from a single species, and the two species

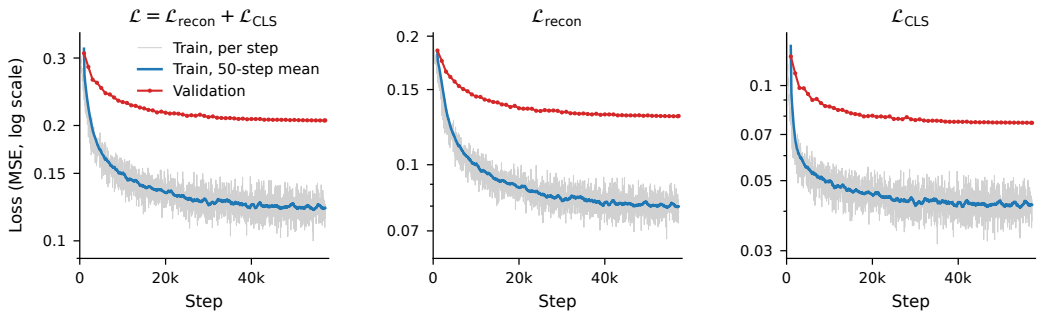

Figure 7: **Both masked-expression objectives improve during pretraining.** Total loss $\mathcal{L}$ (left) and its components $\mathcal{L}_{\text{recon}}$ (middle) and $\mathcal{L}_{\text{CLS}}$ (right) over the 57,200 training steps. Grey: per-step training loss of the current species minibatch; blue: 50-step moving average; red: validation loss every 1,000 steps, averaged over the human and mouse validation sets. Vertical axes are logarithmic.

alternate step by step, so both species receive the same number of updates regardless of corpus size. Within a species, a source file is drawn with probability proportional to its number of metacells, and a contiguous block of rows is read from it in random order. Metacell profiles are library-size normalized to 10^4 counts and $\log(1 + x)$ transformed before being placed on the layout.

Optimization. Training uses AdamW with peak learning rate 3×10^{-4}, weight decay 0.01, a linear warm-up over the first 1,000 steps, and cosine decay to 10^{-5}; gradients are clipped to a global norm of 5. The per-GPU batch size is 128 metacells and the run uses two GPUs with distributed data parallelism, giving a global batch of 256. One epoch is defined as 2,860 steps, the number of steps needed to sample the training metacells once at this batch size; the run lasts 20 epochs, or 57,200 steps. Forward and backward passes run under BF16 autocast with FP32 master weights and optimizer state. Validation loss is computed every 1,000 steps on 256 metacells per validation file with a fixed random mask, and the checkpoint with the lowest validation loss is retained (step 57,000; the final checkpoint at step 57,200 differs by less than 10^{-3}).

Hardware. Training used two NVIDIA H100 GPUs with 80 GB of memory each, connected via NVLink, with PyTorch 2.10 and CUDA 12.8.

Loss curves. Figure 7 shows the total loss and its two components. The validation total falls from 0.309 at step 1,000 to 0.206 at the end, with $\mathcal{L}_{\text{recon}}$ decreasing from 0.185 to 0.130 and $\mathcal{L}_{\text{CLS}}$ from 0.124 to 0.076, and both components are still decreasing slowly at the end of the cosine schedule. The training curve lies below the validation curve throughout. The two curves summarize losses differently: the training value is the minibatch loss of a single species, whereas validation averages the human and mouse validation sets. Their species-specific validation losses differ by about 0.03. Differences in collection composition may also contribute to the gap. No divergence between the two curves develops over the run.

H OpenProblems evaluation details

Datasets and reference run. The evaluation uses the official benchmark data files: DKD (39K cells, 13 cell types, 11 batches; Wilson et al., 2022), GTEx v9 (209K, 53, 16; Eraslan et al., 2022), Immune cell atlas (330K, 35, 12; Domínguez Conde et al., 2022), Mouse pancreas atlas (302K, 13, 56; Hrovatin et al., 2023), and HypoMap (385K, 13, 24; Steuernagel et al., 2022). Reference scores come from `run_2026-08-17_16-18-21`. Missing mouse foundation-model entries reflect the coverage of that run, not a restriction to human inputs.

Input processing and adaptation. Expression is normalized as log1p CP10K, as in pretraining, and placed on the species layout. Layout genes absent from a dataset are marked unavailable; 16,977–18,134 layout genes match each dataset. For fine-tuning, a separate copy of the pretrained

model is adapted to each dataset for 3,000–5,000 steps at learning rate 3×10^{-5}, using the reconstruction and CLS losses without cell-type or batch labels.

Metric aggregation. The biological-conservation score averages seven metrics: ASW label, NMI, ARI, isolated-label ASW and F1, cLISI, and cell-cycle conservation. The batch-correction score averages ASW batch, iLISI, kBET, graph connectivity, and PCR. The overall score is $0.6 \cdot \text{bio} + 0.4 \cdot \text{batch}$. We average raw values over the metrics reported by each method; the public website additionally rescales metrics using control methods. On HypoMap, official entries lack kBET and use four batch metrics, whereas TransMap uses all five. These reported aggregates therefore do not constitute a comparison over an identical metric set on HypoMap.

Fine-tuning outcomes. Fine-tuning increases biological conservation on DKD (0.674 to 0.718) and mouse pancreas (0.704 to 0.721), but reduces batch-correction scores throughout; isolated-label F1 on Immune cell atlas falls from 0.47 to 0.01.

I IMAGE REPRESENTATIONS FOR VARIATIONAL MODELING

We evaluate gene-expression images as inputs to a variational autoencoder, TransMapVI, trained separately on each dataset. For a fixed selected gene set, the TransMap layout retains each supplied expression value at a distinct pixel, $I_{\pi(g)} = x_g$. The experiments below assess the resulting cell representations.

I.1 MODEL AND TRAINING SETUP

TransMapVI encodes $\log(1 + x)$ expression images with a four-stage residual CNN, followed by correlation pooling and a variational head with a 30-dimensional latent space. The model uses a negative-binomial observation likelihood, observed library size, and batch conditioning. We compare it with our scVI baseline (Lopez et al., 2018), which uses a fully connected encoder and a standard MLP decoder. The models have approximately 15M and 14M parameters, respectively. Figure 8 illustrates the TransMapVI architecture.

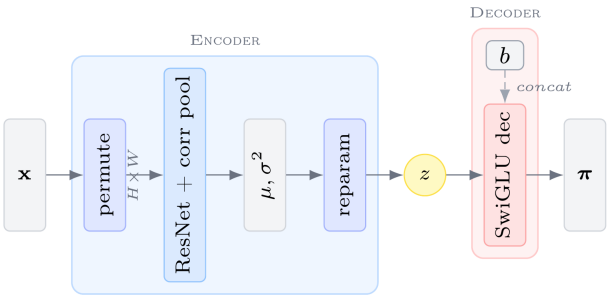

Figure 8: **Gene-expression images support variational cell modeling.** Expression values occupy a fixed gene-to-pixel grid. A residual CNN with correlation pooling estimates posterior parameters (μ, σ^2); reparameterization yields z. The decoder conditions on batch identity b to predict expression proportions $\boldsymbol{\pi}$ for the negative-binomial likelihood. Solid and dashed arrows indicate data flow and covariate conditioning, respectively.

The single-species protocol uses 2,000 selected highly variable genes and cell-type-stratified training, validation, and test splits of approximately 90%, 5%, and 5%. Gene selection uses the training split. Models are trained with Adam at a learning rate of 10^{-3} and batch size 512 for up to 400 epochs. The KL weight increases linearly to 1 over the first 30 epochs; validation-based early stopping then uses a patience of 30 epochs. The best validation checkpoint is evaluated on the test split.

I.2 CELL REPRESENTATION QUALITY

We compare posterior-mean cell embeddings on five human datasets: Lung (Sikkema et al., 2023), Breast (Kumar et al., 2023), CITE-seq PBMC (Hao et al., 2021), Gut Fetal (Elmentaite et al., 2021), and Lung Fetal (He et al., 2022b). Table A3 reports scIB scores, whose total weights biological conservation and batch integration by 0.6 and 0.4 (Luecken et al., 2022), and all three scGraph scores (Wang et al., 2026).

Relative to scVI, TransMapVI improves mean scIB total and biological conservation and all three mean scGraph scores on every dataset; total gains are 0.01–0.03. Batch integration improves on CITE-seq PBMC and both fetal datasets but declines on Lung and Breast. Figure 9 shows cell-type structure in individual fitted embeddings.

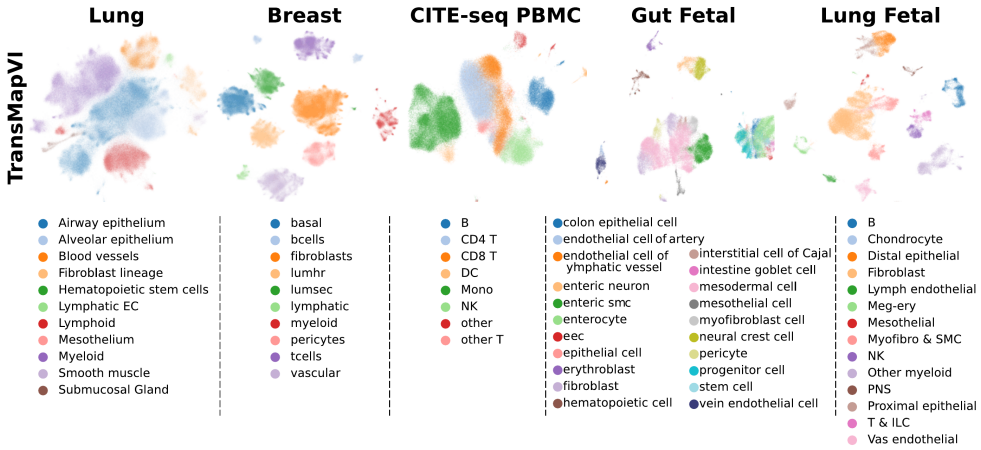

Figure 9: **TransMapVI embeddings retain cell-type structure across five datasets.** UMAPs of cell embeddings on the five single-species datasets, colored by cell type. Each panel shows one fit; Table A3 gives the repeated-run comparison with scVI.

Table A3: Cell-representation quality of scVI (approximately 14M parameters) and TransMapVI (15M). Entries report mean ± standard deviation over five repeated training/layout runs on the test split; TransMapVI uses the TransMap layout. Higher is better; bold marks the higher mean per dataset and metric.

scIB

Dataset	Total		Bio		Batch	
	scVI	TransMapVI	scVI	TransMapVI	scVI	TransMapVI
Lung	0.64 ± 0.0049	$\mathbf{0.67 \pm 0.0082}$	0.60 ± 0.0091	$\mathbf{0.67 \pm 0.0123}$	$\mathbf{0.69 \pm 0.0051}$	0.67 ± 0.0043
Breast	0.71 ± 0.0139	$\mathbf{0.72 \pm 0.0076}$	0.71 ± 0.0191	$\mathbf{0.74 \pm 0.0135}$	$\mathbf{0.70 \pm 0.0073}$	0.68 ± 0.0067
CITE-seq PBMC	0.68 ± 0.0069	$\mathbf{0.70 \pm 0.0094}$	0.65 ± 0.0094	$\mathbf{0.66 \pm 0.0177}$	0.73 ± 0.0137	$\mathbf{0.75 \pm 0.0034}$
Gut Fetal	0.66 ± 0.0072	$\mathbf{0.69 \pm 0.0023}$	0.62 ± 0.0031	$\mathbf{0.63 \pm 0.0027}$	0.72 ± 0.0172	$\mathbf{0.77 \pm 0.0036}$
Lung Fetal	0.69 ± 0.0116	$\mathbf{0.71 \pm 0.0117}$	0.64 ± 0.0180	$\mathbf{0.65 \pm 0.0174}$	0.77 ± 0.0071	$\mathbf{0.80 \pm 0.0073}$

scGraph

Dataset	r_w		r_{PCA}		$\mathrm{rank}_{\mathrm{PCA}}$	
	scVI	TransMapVI	scVI	TransMapVI	scVI	TransMapVI
Lung	0.44 ± 0.0220	$\mathbf{0.56 \pm 0.0263}$	0.50 ± 0.0187	$\mathbf{0.61 \pm 0.0180}$	0.32 ± 0.0224	$\mathbf{0.38 \pm 0.0270}$
Breast	0.52 ± 0.0138	$\mathbf{0.60 \pm 0.0326}$	0.56 ± 0.0060	$\mathbf{0.63 \pm 0.0309}$	0.40 ± 0.0157	$\mathbf{0.44 \pm 0.0395}$
CITE-seq PBMC	0.62 ± 0.0181	$\mathbf{0.69 \pm 0.0206}$	0.66 ± 0.0114	$\mathbf{0.72 \pm 0.0157}$	0.53 ± 0.0182	$\mathbf{0.60 \pm 0.0202}$
Gut Fetal	0.56 ± 0.0313	$\mathbf{0.67 \pm 0.0236}$	0.72 ± 0.0169	$\mathbf{0.78 \pm 0.0080}$	0.38 ± 0.0352	$\mathbf{0.47 \pm 0.0237}$
Lung Fetal	0.71 ± 0.0173	$\mathbf{0.79 \pm 0.0122}$	0.67 ± 0.0135	$\mathbf{0.75 \pm 0.0095}$	0.38 ± 0.0320	$\mathbf{0.45 \pm 0.0151}$