Single-cell multiomics data integration and generation with scPairing

Jeffrey Niu[1], Jiarui Ding[1*]

[1]*Department of Computer Science, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4*

**Corresponding author jiarui.ding@ubc.ca*

Single-cell multiomics technologies generate paired or multiple measurements of different cellular properties (modalities), such as gene expression and chromatin accessibility. However, multiomics technologies are more expensive than their unimodal counterparts, resulting in smaller and fewer available multiomics datasets. Here, we present scPairing, a variational autoencoder model inspired by Contrastive Language-Image Pre-Training, which embeds different modalities from the same single cells onto a common embedding space. We leverage the common embedding space to generate novel multiomics data following bridge integration, a method that uses an existing multiomics bridge to link unimodal data. Through extensive benchmarking, we show that scPairing constructs an embedding space that fully captures both coarse and fine biological structures. We then use scPairing to generate new multiomics data from retina and immune cells. Furthermore, we extend to co-embed three modalities and generate a new trimodal dataset of bone marrow immune cells. Researchers can use these generated multiomics datasets to discover new biological relationships across modalities or confirm existing hypotheses without the need for costly multiomics technologies.

Keywords: single-cell multiomics, contrastive learning, deep generative models, variational autoencoders, scRNA-seq, scATAC-seq

1 Introduction

Recent advances in single-cell sequencing technologies have enabled the joint profiling of multiple modalities within individual cells.[1–4] These multiomics technologies jointly profile different aspects of cell state, such as the transcriptome,[5,6] chromatin accessibility,[7] and surface epitopes.[4] Single-cell multiomics facilitate a deeper characterization of cell state diversity and the interplay between different modalities,[8,9] potentially revealing the complex mechanisms that drive cellular processes.

Despite the increasing availability of single-cell multiomics technologies, large, high-quality multiomics datasets remain scarce due to cost. A typical joint single-cell transcriptomics and epigenomics experiment usually produces less than 100,000 high-quality cells, whereas unimodal sequencing using single-cell RNA sequencing (scRNA-seq) or single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) has produced much larger data, including cell atlases with over a million cells.[10,11] Multiomics data also have lower data quality compared to single-modality data, potentially obscuring downstream analyses.[12]

A wide range of single-cell multiomics integration tools perform various analyses on multiomics data.[13–22] Both statistical[20] and deep learning[13,15] methods are used to model multiomics data to learn data representations suitable for downstream analyses to identify cell populations or differentiation trajectories. Because of the scalability to handle large datasets and the effectiveness in capturing the structure in complex high-dimensional single-cell data, deep learning methods, especially deep generative models, have been used to co-embed multiomics data to a shared low-dimensional latent space.[13–18] These models integrate the separate information sources present in multiomics data to uncover patterns that are hard to detect from a single modality alone, such as gene regulatory relationships. However, although these models have decoders to impute the missing data modalities given the shared cell-specific embeddings, these imputed data may miss fine-grained cell state information or have distorted characteristics compared to experimentally generated data. Consequently, these imputed data are seldom used for downstream analysis for biological discoveries, such as gene regulatory network inference.[19] Furthermore, these models are typically more time-consuming to run compared to those for unimodal data analysis and may rely on heuristics to combine the modality-specific embeddings to obtain joint cell-specific embeddings,[13,16,17,23] or not natural and have difficulties in embedding data in the presence of missing modalities.[14,19,22] Lastly, many models do not easily adapt to valuable multiomics data with more than two modalities.[24,25]

Here, we introduce scPairing, an efficient and flexible deep generative model for multiomics data integration and generation by pairing independently generated experimental unimodal data. scPairing has a modality-specific encoder to embed each modality data to a shared latent space. Furthermore, scPairing uses contrastive learning to encourage the modality-specific embeddings of the same cell to be closeby, while pushing away embeddings of different cells. scPairing uses dropout, a technique to randomly draw a subset of latent dimensions from each modality embedding, to merge the modality-specific embeddings, which has been shown to effectively merge modality-specific embeddings from cardiovascular data.[26–28] This

formulation enables both multiomics data integration and generation, where scPairing can produce joint embeddings suitable for biological investigation and facilitate the generation of artificial multiomics data from pairing independent experimental unimodal datasets, using a multiomics dataset as a bridge.[29] scPairing is flexible to accept both raw counts data or low-dimensional embeddings for each modality as inputs, and can be very efficient in processing atlas-level data when using low-dimensional embeddings as input. We extensively test scPairing on nine human and mouse datasets, demonstrating scPairing's superior performance on multimodal integration and pairing unimodal data. First, we demonstrate scPairing's integration capabilities on common multiomics integration tasks. Second, we show how we can use the aligned embedding space learned by scPairing to perform bridge integration, resulting in artificial multiomics data by generating new multiomics human retina data, which are used to study the correlations between chromatin accessibility and gene expression. Third, we use a specialized bridge dataset to generate multiomics profiles of a rare immune cell type to study the shared chromatin profiles across tissues. Lastly, we extend scPairing to integrate and generate single-cell trimodal sequencing data. Given its flexibility and scalability in processing large datasets, effectiveness in multiomics data integration, and its ability to pair large-scale unimodal data to generate multiomics data for biological discoveries, we envision that scPairing will become a valuable tool for single-cell multiomics data analysis. scPairing is a publicly available software tool, available at https://github.com/Ding-Group/scPairing.

2 Results

scPairing jointly embeds cells onto a common hyperspherical space. scPairing model uses a variational autoencoder (VAE), a deep generative model augmented with a contrastive loss, to embed single-cell multiomics data onto a common hyperspherical space (**Fig. 1A**). scPairing takes in low-dimensional representations of each modality computed from domain-specific methods such as principal component analysis (PCA), VAEs designed for transcriptomics,[30,31] or single-cell large language models (LLMs)[32,33] for scRNA-seq data, and latent semantic indexing (LSI) or PeakVI[34] for scATAC-seq data. scPairing's encoders (feed-forward neural networks, **Methods**) then transform these low-dimensional representations onto a common hyperspherical latent space. Following the approach of Contrastive Language Image Pre-training (CLIP),[26] scPairing employs a contrastive objective that encourages transformed embeddings of different modalities from the same cell to be proximal, while embeddings from different cells are pushed further apart. Our model adds further alignment objectives to ensure modality alignment (**Methods**).

Following the motivations of CLIP, we designed scPairing to take representations from modality-specific models by default. Using modality-specific representations provides flexibility in the final representations learned by scPairing, such as accounting for the trade-off between batch correction and biological conservation.[35] Modality alignment becomes easier to learn with low-dimensional representations, as fewer dimensions make finding the alignment simpler than if the inputs were high-dimensional raw counts. Using low-dimensional representations also makes the application of scPairing computationally inexpensive, as the model has fewer parameters than a model that directly takes counts as input (**Supplementary Fig. S1**). The embeddings learned by scPairing can be used for downstream analyses, such as clustering and visualization.

Following a bridge integration procedure,[29] scPairing can generate new multiomics data from separate unimodal measurements by artificially pairing these cells (**Fig. 1B**). For simplicity, we will first discuss the case with two data modalities and discuss cases with more than two modalities in later sections. First, the individual modalities from the multiomics data are integrated with the unimodal data. We train scPairing on only the bridge data to learn a common embedding space and then pass the unimodal data through scPairing encoders to project the two unimodal datasets onto the same space. We assign pairings between cells of the two modalities by computing the assignment of cells from one modality to the other modality such that the total similarity among all assignments is maximized (**Methods**).

scPairing preserves biological structure while aligning modalities. We first applied scPairing to a joint scRNA-seq and scATAC-seq benchmarking dataset of bone marrow mononuclear cells (BMMCs) sequenced with 10X Multiome, produced by the Multimodal Single-Cell Data Integration Challenge at NeurIPS 2021.[36] For the low-dimensional representations of the scRNA-seq data, we applied embeddings from Harmony-corrected PCA,[37] scVI,[30] scPhere,[31] scGPT,[32] and CellPLM,[33] while for the low-dimensional representations of the scATAC-seq data, we applied embeddings from Harmony-corrected LSI and

PeakVI[34] (**Fig. 2A and 2B, Supplementary Fig. S2**).

To quantify the preservation of biological structure, we computed cross-batch k-nearest neighbor cell type classification accuracies with scPairing embeddings. We held out cells from one batch and predicted their types using the cells from the other batches. Our results show scPairing preserves the structure just as well as other multiomics methods that learn the structure from scratch, and better than unimodal methods (**Fig. 2C, Supplementary Fig. S3A**). We also performed ten-fold cross-validation of k-nearest neighbors accuracies, where we saw the same trends as the cross-batch evaluation (**Supplementary Fig. S3B and S3C**). Next, we performed cross-site k-nearest neighbor cell type classification, where we trained each model on cells from three of the four data sequencing sites and transferred the cells from the remaining site onto the training data. Our model is competitive with other methods in this data transfer experiment (**Supplementary Fig. S4**).

Next, we evaluated scPairing's ability to align the two modalities in the common embedding space using the Fraction of Samples Closer than True Match (FOSCTTM) metric[38] (**Methods**). We observed that scPairing embeddings derived from scVI with PeakVI or Harmony-corrected LSI outperformed two methods designed for modality alignment, CLUE[39] and Cobolt[15] (**Fig. 2D, Supplementary Fig. S5A**). We also performed cross-site validation, where scPairing with scVI or CellPLM and PeakVI embeddings outperformed CLUE and Cobolt at aligning the held-out site (**Supplementary Fig. S5B**).

We also demonstrated scPairing's alignment of modalities using a dataset of mouse skin cells sequenced by SHARE-seq,[2] using Palantir[40] to order the cells in pseudotime. In the mouse skin, the scRNA-seq data do not show the correct differentiation trajectory, while the scATAC-seq data do show the correct trajectory (**Supplementary Fig. S6A and S6B**). After applying scPairing, the transformed scRNA-seq embeddings followed the correct differentiation trajectory, while the scATAC-seq embeddings retained the correct differentiation trajectory (**Supplementary Fig. S6C and S6D**). This experiment demonstrates that scPairing representations infused the accurate information from both modalities.

We also investigated modality mixing in scPairing's representations, as aligning the modalities does not preclude the possibility that the modalities are separated in the common embedding space. The uniform manifold approximation and projection (UMAP) visualizations of the two modalities in the benchmarking dataset indicate strong mixing (**Supplementary Fig. S7A–C**). For a quantitative evaluation, we calculated the Local Inverse Simpson's Index (LISI), which measures the diversity within the nearest neighborhoods in a k-nearest neighbor graph. We see that scPairing embeds modalities with greater mixing compared to CLUE and Cobolt (**Fig. 2e, Supplementary Fig. S7D**). Along each dimension of scPairing's common hyperspherical embedding space, we observe that each dimension has near-identical overlap of the two modalities (**Supplementary Fig. S7E**).

Since both decoders in scPairing decode from the common latent space, scPairing can perform cross-modality prediction. We performed the same cross-site split of the NeurIPS 2021 benchmarking dataset and used the scATAC-seq embeddings to impute scRNA-seq profiles and vice-versa. Compared to MultiVI, scPairing with pre-trained or learned decoders

achieved slightly better imputations based on the Spearman correlation between true and imputed gene expression (**Fig. 2F**), and the imputations accurately reflected the expression of marker genes (**Supplementary Fig. S8**). The binary cross-entropy between the true binarized and imputed peak accessibility was lower in three of four sites when comparing scPairing with pre-trained decoders and MultiVI (**Fig. 2G**). When using learned decoders, the binary cross-entropy drastically decreased.

Bridge integration produces artificial cell pairings that approximate true multiomics data. We evaluated scPairing's ability to produce realistic artificial multiomics data by pairing gene expression and chromatin accessibility profiles from human retinal cells, using another paired multiomics dataset as a bridge. The unpaired Liang *et al.* dataset contains over 250,000 cells sequenced with single-nuclei RNA sequencing (snRNA-seq) and 137,000 cells sequenced with single-nuclei ATAC-seq (snATAC-seq).[41] The paired Wang *et al.* bridge dataset contains over 50,000 cells sequenced with 10X Genomics Single Cell Multiome[42] (**Supplementary Fig. S9A–C**).

After harmonizing the unpaired and bridge datasets and applying scPairing, we obtained 105,613 artificially paired cells (**Fig. 3A**). Visualization of our artificial pairings shows that scPairing correctly separated the discrete major cell types in the retina. To determine the quality of our pairings, we focused on 24,446 retinal bipolar cells. Liang *et al.* annotated 14 retinal bipolar cell subtypes, which our pairings also successfully captured (**Fig. 3B**).

We further assessed whether our artificial pairings are concordant with true paired data by calculating the correlation between gene expression and chromatin accessibility of peaks within 250kb of the gene transcription start site (TSS) across 12 abundant retinal bipolar cell subtypes. We compared our pairings against imputations from canonical correlation analysis (CCA) between gene activities and the scRNA-seq data,[43] and imputations solely derived from gene activities.[44] Sparsity presents a challenge in computing faithful gene-peak correlations. Thus, we applied the SEACells algorithm[45] to aggregate retinal bipolar cell subtypes into robust metacells (**Supplementary Note S1**). For each gene-peak pair, we computed the mean and standard deviation of the Pearson correlations across the 12 subtypes (**Fig. 3C**).

We found that the gene-peak correlations of the artificial pairings most closely match the correlations from the paired multiomics data (**Fig. 3D**). The multiomics data suggest that the correlation between most genes and peaks close to the TSS have little correlation with little variance between subtypes. Artificial pairings produced by scPairing have a distribution of means and standard deviations closer to the paired data. The correlations produced by gene activities and gene activities followed by CCA both overestimate the gene-peak correlations and variance of correlations between subtypes. While the addition of CCA to gene activities moves the distribution closer to the correlations of the paired data, it still overestimates the correlations. This overestimation may arise from the direct transformation of gene activities performed to impute gene expression, neglecting the variability present within cells.

We then looked specifically at the relationship between *GNB3* expression and the peak at

chr12:6850801-6853197 (**Fig. 3E, Supplementary Fig. S9D**). Using aggregated meta-cells, we found that this gene-peak pair has strong positive correlation across all subtypes (**Fig. 3F**). This result aligns with the correlations seen in the paired multiomics data, which also showed positive correlations (**Supplementary Fig. S9E**), while aggregating cells by donors might detect heterogeneous correlations because cells from different donors were well mixed (**Supplementary Note S1, Supplementary Fig. S10 and Fig. S11**).

Using a specialized bridge dataset pairs a rare immune cell type. Ulezko Antonova *et al.* recently described a new rare cell type with characteristics of both dendritic cells (DCs) and type 3 innate lymphoid cells (ILC3s), which they termed R-DC-like cells.[46,47] In their study, they applied a sorting strategy to yield a joint scRNA-seq and scATAC-seq dataset of human tonsil cells enriched for R-DC-like cells. To test the versatility of scPairing, we attempted to pair scRNA-seq and scATAC-seq data from two other datasets where very few R-DC-like cells have been captured, using the Ulezko Antonova *et al.* dataset as a bridge. To evaluate the quality of our pairings, we identified cells that express *PRDM16*, *RORC*, *PIGR*, and *FOXP2* as candidate R-DC-like cells. If we successfully paired these cells, the artificially paired cells should show a chromatin accessibility profile similar to the R-DC-like cells in the Ulezko Antonova *et al.* dataset (**Fig. 4**). We highlight four regions examined by Ulezko Antonova *et al.* at the *RORC*, *AIRE*, and *CD4* loci. They found that R-DC-like cells have an accessible *RORγt* promoter region but not at the *RORγ* promoter region, an inaccessible *AIRE* conserved non-coding sequence, and an accessible *CD4* promoter region.

We first artificially paired data from an atlas of human tonsil cells.[48] This data contains both paired and unpaired scRNA-seq and scATAC-seq cells. Starting with the paired data, we verified that scPairing correctly re-pairs the R-DC-like cells. Since the two modalities were already paired, we checked whether an scRNA-seq modality of an R-DC-like cell paired with an scATAC-seq modality of an R-DC-like cell. Out of 27 R-DC-like scRNA-seq profiles, 24 paired with a scATAC-seq profile from a R-DC-like cell, including 12 perfect matches (**Supplementary Table S1**). Comparing the chromatin accessibility profiles, we found that the profiles of our re-pairings matched the profiles from the original paired data and the Ulezko Antonova *et al.* data at all four key regions (**Fig. 4**).

Given the success of scPairing at re-pairing rare R-DC-like cells, we then applied scPairing to the unpaired cells. Examining the chromatin accessibility profiles paired to R-DC-like scRNA-seq profiles, we see that they mirror both the Ulezko Antonova *et al.* dataset and the original paired tonsil atlas dataset at the four regions of interest (**Fig. 4**).

Next, we applied scPairing to paired human intestine cells[49] to show that our pairing procedure works across tissues. The chromatin accessibility profiles of the original paired data and our re-pairings are consistent with each other (**Fig. 4, Supplementary Table S2**). The *RORC* and *CD4* loci are consistent with the Ulezko Antonova *et al.* dataset, but the *AIRE* conserved non-coding sequence has accessibility close to the region in both the paired and re-paired data. Thus, the chromatin accessibility profile of R-DC-like cells in the intestine may be different from the R-DC-like cells in the tonsil.

scPairing extends to trimodal sequencing data. To demonstrate that scPairing can integrate more than two data modalities, we applied our model to a TEA-seq dataset, which contains joint measurements of gene expression, chromatin accessibility, and cell epitopes.[25]

The representations learned by scPairing captures major cell types and recapitulates known benefits of detecting cell epitopes, such as $CD56^{bright}$ and $CD56^{dim}$ natural killer cell subtypes[4] (**Fig. 5A**). The representations for each individual modality also showed separation of cell types (**Fig. 5B**) while mixing the three modalities in the common embedding space (**Fig. 5C, Supplementary Fig. S12**). Compared to CLUE and Cobolt, scPairing embeds each modality closer to its true match across all pairs of the three modalities (**Fig. 5D**).

Next, we applied scPairing to re-pair DOGMA-seq[24] data using the TEA-seq data as a bridge. The DOGMA-seq embeddings produced by scPairing well preserve the matching triplets, as the true matching ranks near the top in terms of embedding similarity among all possible matchings (**Fig. 5E**). The true matching triplets also exhibited high pairwise cosine similarity and low FOSCTTM, indicating that they are co-embed closely (**Supplementary Tables S3, S4**). The scRNA-seq and scATAC-seq modality pair showed better alignment metrics, suggesting that finding the true pairing with epitope data is harder. This difficulty may stem from fewer overlapping features between datasets, with only 39 proteins shared by the TEA-seq and DOGMA-seq data, while both the scRNA-seq and scATAC-seq modalities have much greater overlap.

Then, we used a custom randomized greedy algorithm to generate a re-pairing of the DOGMA-seq data, as no efficient optimal algorithm exists for this problem (**Methods**). Next, we evaluated whether our re-pairing reflects the original data. Without ground-truth cell type labels, we compared the similarity of clusters discovered in the original data versus the re-paired data. The cells in the re-pairing contain three modalities combined to form one cell, each of which may originate from different cells. Thus, we computed the similarity with respect to each modality. We observed that the clusterings show strong similarities, with substantial overlap regardless of modality (**Supplementary Fig. S13**), with an average adjusted Rand index (ARI) of 0.522 and normalized mutual information (NMI) of 0.623.

Finally, we synthesized a new trimodal dataset using the joint scRNA-seq and scATAC-seq BMMC benchmarking dataset and the epitopes from the CITE-seq dataset from the 2021 NeurIPS challenge,[36] using the TEA-seq data as a bridge. After projecting all three modalities onto the bridge, we assigned epitope profiles to the Multiome cells. Out of 69,249 Multiome cells, we successfully paired 56,569 cells to high-similarity epitope profiles (**Fig. 5F**). Despite the TEA-seq bridge missing cell types present in the BMMC data, we could still find pairings for the unobserved erythroid cell types, which suggests that scPairing could learn an embedding space that matches out-of-distribution cells. We also observed that the matched epitope profiles are consistent with the cell identities through the expression of marker epitopes, including the separation between CD4 and CD8 expressing cells, and CD45RA and CD45RO expressing cells (**Fig. 5G**).

3 Discussion

In summary, scPairing builds upon the powerful variational autoencoder architecture by adding contrastive learning to align representations of multiomics data together onto a common embedding space using dropout. By learning the alignment between paired multiomics data, scPairing enables the alignment of unpaired data using its trained encoders, from which new artificial multiomics data can be generated by matching cells together.

Generating artificial pairings with scPairing can be viewed as a reformulation of bridge integration. As originally defined, bridge integration expresses unimodal cells in terms of multiomics cells (the bridge) to harmonize the features for cells from separate modalities using dictionary learning.[29] While both our method and the original bridge integration first harmonize the bridge with the unimodal data, our method differs by using the bridge data to find an alignment between the modalities. This procedure is more efficient than dictionary learning, which expresses unimodal cells in terms of bridge cells.

Unlike many other multiomics integration tools taking raw counts as input, scPairing can take low-dimensional representations of each modality as inputs. This strategy reduces scPairing's runtime, although computing the modality-specific representations can be computationally expensive for complex deep learning models. However, the availability of pre-trained models has increased, including pre-trained scVI models[30] or single-cell LLMs,[32,33,50] each trained on millions of cells. These pre-trained models can compute embeddings either after fine-tuning on a specific dataset or directly in a zero-shot manner without any additional training, drastically reducing the time required to obtain input representations for scPairing. Our benchmarking with scGPT[32] and CellPLM[33] embeddings demonstrates that these pre-trained models can be used with scPairing. As large pre-trained models designed for chromatin accessibility become available, such as EpiGePT,[51] we foresee scPairing interfacing directly with pre-trained models for any modality, allowing for fast multiomics data integration.

The flexibility in choosing modality-specific integration methods as input for scPairing opens the door to cross-species multiomics data generation. Fewer non-human paired multiomics data exist due to the high cost of multiomics sequencing. Given the success in pairing R-DC-like cells using bridge and unimodal data from different tissues, one can potentially also apply scPairing to bridge and unimodal data from different species. Specialized tools developed for cross-species integration can assist in integrating the bridge and unimodal data.[52–54]

We believe that scPairing presents an opportunity to leverage the wealth of unimodal single-cell data to generate high-quality artificial multiomics data, including difficult to obtain trimodal data. These generated datasets can be analyzed through a multiomics lens using other powerful multiomics tools to identify novel relationships between modalities.

4 Methods

The scPairing framework. Though scPairing applies to any combination of data modalities, we describe the two-modality version of scPairing in terms of joint scATAC-seq and scRNA-seq data. The scPairing model takes in a multimodal dataset with gene expression matrix $X^{\text{RNA}} = [\mathbf{x}_1^{\text{RNA}}, ..., \mathbf{x}_C^{\text{RNA}}] \in \mathbb{N}^{C \times d_1}$, and chromatin accessibility matrix $X^{\text{ATAC}} = [\mathbf{x}_1^{\text{ATAC}}, ..., \mathbf{y}_C^{\text{ATAC}}] \in \mathbb{N}^{C \times d_2}$, with C cells, d_1 genes, and d_2 peaks. scPairing also takes in a batch-corrected low-dimensional representation of each modality, $Y^{\text{RNA}} = [\mathbf{y}_1^{\text{RNA}}, ..., \mathbf{y}_C^{\text{RNA}}] \in \mathbb{R}^{C \times r_1}$ and $Y^{\text{ATAC}} = [\mathbf{y}_1^{\text{ATAC}}, ..., \mathbf{y}_C^{\text{ATAC}}] \in \mathbb{R}^{C \times r_2}$, with r_1 and r_2 being the dimensions of the representations.

These representations can be obtained from approaches such as PCA for gene expression data and LSI for chromatin accessibility data. Alternatively, these representations can also be obtained from deep learning models such as scVI[30] for gene expression data and PeakVI[34] or cisTopic[55] for chromatin accessibility data, or from pre-trained models based on transformers.[32,33,56]

In scPairing, we incorporate the Contrastive Language Image Pre-training (CLIP) model described by Radford *et al.*[26] designed for joint embedding of image and text representations into a variational autoencoder model. For each cell i, we assume that the representations $\mathbf{y}_i^{\text{RNA}}$ and $\mathbf{y}_i^{\text{ATAC}}$ are described by a low-dimensional latent vector $\mathbf{z}_i \in \mathbb{R}^d$ that lies on a $(d-1)$-sphere $\mathbb{S}^{d-1}$,

$$\begin{aligned} p_{\boldsymbol{\theta}_i}(\mathbf{y}_i^{\text{RNA}}, \mathbf{y}_i^{\text{ATAC}}, \mathbf{z}_i) &= p(\mathbf{z}_i) \cdot p_{\boldsymbol{\theta}_i}(\mathbf{y}_i^{\text{RNA}}, \mathbf{y}_i^{\text{ATAC}} \mid \mathbf{z}_i) \\ &= p(\mathbf{z}_i) \cdot p_{\boldsymbol{\theta}_i}(\mathbf{y}_i^{\text{RNA}} \mid \mathbf{z}_i) \cdot p_{\boldsymbol{\theta}_i}(\mathbf{y}_i^{\text{ATAC}} \mid \mathbf{z}_i) \end{aligned}$$

where $\boldsymbol{\theta}_i$ is the parameter set for each distribution. The prior distribution $p(\mathbf{z}_i)$ is assumed to be the uniform distribution on the $(d-)$-sphere. Both $p_{\boldsymbol{\theta}_i}(\mathbf{y}_i^{\text{RNA}} \mid \mathbf{z}_i)$ and $p_{\boldsymbol{\theta}_i}(\mathbf{y}_i^{\text{ATAC}} \mid \mathbf{z}_i)$ are assumed to be normally distributed, with parameters specified by neural networks (decoders).

We use variational inference to approximate the posterior distribution for $\mathbf{z}_i$. The variational distribution $q_{\boldsymbol{\phi}_i}(\mathbf{z}_i \mid \mathbf{y}_i^{\text{RNA}}, \mathbf{y}_i^{\text{ATAC}})$ is assumed to be a Power Spherical distribution[57] on a $(d-1)$-sphere. The Power Spherical distribution is an alternative to the von Mises-Fisher (vMF) distribution, which is considered the normal distribution on the hypersphere. We chose the Power Spherical distribution because of its sampling efficiency compared to the vMF distribution. The Power Spherical distribution has two parameters, the direction $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ and the concentration $\kappa \in \mathbb{R}_{>0}$.

This variational distribution is parameterized by the three neural network encoders f_{RNA}, f_{ATAC} and f_κ, which compute $\boldsymbol{\mu}_{\text{RNA}}$, $\boldsymbol{\mu}_{\text{ATAC}}$, and κ, respectively. Since f_{RNA} and f_{ATAC} both produce a d-dimensional direction, we merge the directions from both encoders using a method derived from dropout.[27,28] For each index $i \in \{1, 2, \cdots, d\}$, we take the value corresponding to the i-th index in $\boldsymbol{\mu}_{\text{RNA}}$ with 0.5 probability. Otherwise, we take the value corresponding to the i-th index in $\boldsymbol{\mu}_{\text{ATAC}}$. This approach allows the model to better generalize to varying emphases on $\boldsymbol{\mu}_{\text{RNA}}$ or $\boldsymbol{\mu}_{\text{ATAC}}$.

The evidence lower bound (ELBO) objective for training both encoder and decoder neural

networks become

$$\begin{aligned}\mathcal{L}_{\text{ELBO}} = & -D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{y}^{\text{RNA}}, \mathbf{y}^{\text{ATAC}}) \parallel p(\mathbf{z})) \\ & + \mathbb{E}_{\mathbf{z}\sim q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y}^{\text{RNA}},\mathbf{y}^{\text{ATAC}})}[\log p_{\boldsymbol{\theta}}(\mathbf{y}^{\text{RNA}}, \mathbf{y}^{\text{ATAC}} \mid \mathbf{z})].\end{aligned}$$

Here $D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{y}^{\text{RNA}}, \mathbf{y}^{\text{ATAC}}) \parallel p(\mathbf{z}))$ represents the Kullback-Leibler (KL) divergence between $q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{y}^{\text{RNA}}, \mathbf{y}^{\text{ATAC}})$ and $p(\mathbf{z})$. We also remove the cell index i here for brevity, assuming just one given cell. Given N cells, the ELBO is the sum of the ELBO over the N cells.

Given a Power Spherical distribution P and a uniform distribution on the sphere $Q = \mathcal{U}(\mathcal{S}^{d-1})$, the KL divergence has a closed form solution,[57]

$$KL(P \parallel Q) = -H(P) + H(Q),$$

where H is the entropy of the distribution. Since we assumed the prior distribution $p(\mathbf{z}_i)$ to be a uniform distribution on the hypersphere, we can compute the closed-form solution for the KL divergence between the variational approximation and the prior distribution.

To enforce that the two embeddings of a cell are close on the hypersphere, the encoders are learned to minimize the CLIP contrastive loss. In each batch of N cells, scPairing maximizes the cosine similarity of the embeddings for the N matching measurements, while minimizing the cosine similarity of the embeddings for the $N^2 - N$ non-matching measurements. For each RNA direction $\boldsymbol{\mu}_i^{\text{RNA}}$, we compute the cross-entropy loss and average over all cells,

$$\mathcal{L}_{\text{RNA}} = -\frac{1}{N}\sum_i \log \frac{\exp({\boldsymbol{\mu}_i^{\text{RNA}}}^{\text{T}} \boldsymbol{\mu}_i^{\text{ATAC}}/\tau)}{\sum_{k=1}^N \exp({\boldsymbol{\mu}_i^{\text{RNA}}}^{\text{T}} \boldsymbol{\mu}_k^{\text{ATAC}}/\tau)},$$

where τ is the learned temperature parameter. For each ATAC direction $\boldsymbol{\mu}_j^{\text{ATAC}}$, we compute the symmetric version of the previous loss and average over all cells,

$$\mathcal{L}_{\text{ATAC}} = -\frac{1}{N}\sum_j \log \frac{\exp({\boldsymbol{\mu}_j^{\text{RNA}}}^{\text{T}} \boldsymbol{\mu}_j^{\text{ATAC}}/\tau)}{\sum_{k=1}^N \exp({\boldsymbol{\mu}_k^{\text{RNA}}}^{\text{T}} \boldsymbol{\mu}_j^{\text{ATAC}}/\tau)}.$$

In addition to the contrastive objective, we add an adversarial objective by including a discriminator neural network $h : \mathbb{S}^{d-1} \to [0, 1]$ that takes each direction, $\boldsymbol{\mu}_i^{\text{RNA}}$ or $\boldsymbol{\mu}_i^{\text{ATAC}}$, as input and outputs the probability that it originates from an RNA measurement. The RNA and ATAC encoders are trained to fool the discriminator such that the discriminator cannot determine the difference between the RNA embeddings and ATAC embeddings. The adversarial loss is given by

$$\mathcal{L}_D = \frac{1}{N}\sum_i^N \log(h(\boldsymbol{\mu}_i^{\text{RNA}})) + \frac{1}{N}\sum_j^N \log(1 - h(\boldsymbol{\mu}_j^{\text{ATAC}})).$$

Finally, we add an alignment objective to further encourage modalities from the same cell to be similar,

$$\mathcal{L}_A = -\frac{1}{N}\sum_i^N \text{CosineSimilarity}(\boldsymbol{\mu}_i^{\text{RNA}}, \boldsymbol{\mu}_i^{\text{ATAC}})$$

In datasets where batch effects are prominent and cannot be sufficiently corrected by the modality-specific encoders, the CLIP contrastive loss may over-separate cells based on their batch. To counteract this effect, we add another adversarial objective, except this discriminator predicts the batch each embedding came from. Given a dataset with B batches, the discriminator neural network $b : \mathbb{S}^{d-1} \to [0,1]^B$ takes in $\boldsymbol{\mu}_i^{\text{RNA}}$ or $\boldsymbol{\mu}_i^{\text{ATAC}}$ and produces the probabilities of cell i coming from each batch. The RNA and ATAC encoders are trained to fool this discriminator. The batch adversarial loss is a B-class cross-entropy loss,

$$\mathcal{L}_B = \frac{1}{N}\sum_i^N \sum_{t=1}^B y_{i,t}\log(b(\boldsymbol{\mu}_i^{\text{RNA}})_t) + \frac{1}{N}\sum_i^N \sum_{t=1}^B y_{i,t}\log(b(\boldsymbol{\mu}_i^{\text{ATAC}})_t),$$

where $y_{i,t}$ is a binary indicator for cell i belonging to batch t.

The overall objective of scPairing is:

$$\begin{cases} \max\limits_{h,b} \mathcal{L}_D + \mathcal{L}_B \\ \min\limits_{f_{\text{RNA}}, f_{\text{ATAC}}, f_\kappa, g_{\text{RNA}}, g_{\text{ATAC}}} -\mathcal{L}_{\text{ELBO}} + \mathcal{L}_{\text{RNA}} + \mathcal{L}_{\text{ATAC}} + \mathcal{L}_D + \mathcal{L}_A + \mathcal{L}_B. \end{cases}$$

Extending the model to allow for full data reconstruction. While the model described so far works for producing aligned embeddings usable in downstream analyses, the model does not generate raw count data, unlike other methods.[13,15,39] To enable data imputation, we augment the model with two additional decoders that take the reconstructed representations $\hat{\mathbf{y}}_i^{\text{RNA}}$ and $\hat{\mathbf{y}}_i^{\text{ATAC}}$ and reconstruct the count data $\hat{\mathbf{x}}^{\text{RNA}}$ and $\hat{\mathbf{x}}^{\text{ATAC}}$. We also define probabilistic models of scRNA-seq and scATAC-seq count data to modify the scPairing ELBO.

We model the scRNA-seq counts with a negative binomial distribution. For each observed count $x_{i,g}^{\text{RNA}}$ of gene g in cell i the scRNA-seq data matrix $X^{\text{RNA}} \in \mathbb{N}^{C \times d_1}$, we model $x_{i,g}^{\text{RNA}}$ as

$$x_{i,g}^{\text{RNA}} \sim \text{NB}(\mu_{i,g}, \sigma_g),$$

where the mean $\mu_{i,g}$ is specified by a neural network decoder $\hat{g}_{\text{RNA}}$, while the dispersion parameter σ_g is a gene-specific non-negative number shared across cells.

We model the scATAC-seq counts with a Bernoulli distribution. For each observed count $x_{i,r}^{\text{ATAC}}$ of region r in cell i, we binarize the observation into either zero or one. This binarized count is $x_{i,r}^{\prime\text{ATAC}} = 0$ if $x_{i,r}^{\text{ATAC}} = 0$ and $x_{i,r}^{\prime\text{ATAC}} = 1$ if $x_{i,r}^{\text{ATAC}} \geq 1$. Then, we model this observation as

$$x_{i,r}^{\prime\text{ATAC}} \sim \text{Ber}(p_{i,r}),$$

where the probability parameter $p_{i,r}$ is specified by a different neural network decoder $\hat{g}_{\text{ATAC}}$.

The new ELBO formulation becomes

$$\begin{aligned}\mathcal{L}_{\text{ELBO}} = &- KL(q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{y}^{\text{RNA}}, \mathbf{y}^{\text{ATAC}}) \parallel p(\mathbf{z})) \\ &+ \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y}^{\text{RNA}},\mathbf{y}^{\text{ATAC}})}[\log p_{\boldsymbol{\theta}}(\mathbf{y}^{\text{RNA}}, \mathbf{y}^{\text{ATAC}} \mid \mathbf{z}) + \\ &\qquad \log p_{\boldsymbol{\theta}}(\mathbf{x}^{\text{ATAC}} \mid \mathbf{y}^{\text{ATAC}}) + \\ &\qquad \log p_{\boldsymbol{\theta}}(\mathbf{x}^{\text{RNA}} \mid \mathbf{y}^{\text{RNA}})].\end{aligned}$$

Leveraging pre-trained decoders for full data reconstruction. Instead of using a decoder that is trained from scratch during the training process, if the low-dimensional representations Y^{RNA} or Y^{ATAC} were computed using a method that allows for reconstruction, we can leverage those decoders for faster training. Methods such as PCA and variational autoencoders already contain a reconstructor, either by multiplying the representations by the principal axes in PCA, or passing the latent representations through the VAE decoder. Instead of learning new decoders to reconstruct the counts, we can pass $\hat{Y}^{\text{RNA}}$ and $\hat{Y}^{\text{ATAC}}$ into the reconstructor from the method originally used to compute Y^{RNA} or Y^{ATAC}.

Extension to datasets with more than two modalities. To extend scPairing to M modalities with count matrices $X^1, X^2, ..., X^M$ and batch-corrected low-dimensional representations $Y^1, Y^2, ..., Y^M$, we add additional encoders and decoders for the extra modalities, and compute the contrastive loss between all pairs of modalities. There are encoders $f_1, ..., f_M$, where f_m takes in Y^m as input and produces d-dimension directions, $\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_M$, and decoders $g_1, ..., g_M$, where g_m takes in the merged d-dimension embedding and reconstructs X^m. The process to merge the M d-dimension directions from each encoder uses the same dropout method, except each encoding has a $1/M$ chance of being selected at each index. The pairwise CLIP loss between modality m_1 and m_2 is given by

$$\mathcal{L}_{m_1,m_2} = -\frac{1}{N}\sum_i \log \frac{\exp(\boldsymbol{\mu}_i^{m_1\mathrm{T}}\boldsymbol{\mu}_i^{m_2}/\tau)}{\sum_{k=1}^N \exp(\boldsymbol{\mu}_i^{m_1\mathrm{T}}\boldsymbol{\mu}_k^{m_2}/\tau)},$$

where this loss is calculated across all modalities m_1, m_2, where $m_1 \neq m_2$. The discriminator h changes from a function $h : \mathbb{R}^d \to [0, 1]$ to $h : \mathbb{R}^d \to [0, 1]^M$, a function that outputs the probabilities that the embedding came from the m-th modality for $m \in \{1, ..., M\}$. The discriminative loss becomes an M-class cross-entropy loss,

$$\mathcal{L}_D = \frac{1}{N}\sum_i^N \log(h(\boldsymbol{\mu}_i^{m_1})_1) + ... + \frac{1}{N}\sum_i^N \log(h(\boldsymbol{\mu}_i^{m_M})_M).$$

The alignment objective becomes a pairwise cosine similarity between all modalities,

$$\mathcal{L}_A = -\sum_{i=1}^M \sum_{j=i+1}^M \frac{1}{N}\sum_k^N \text{CosineSimilarity}(\boldsymbol{\mu}_k^{m_i}, \boldsymbol{\mu}_k^{m_j}).$$

The batch discriminator objective is modified to calculate the prediction across all modality embeddings,

$$\mathcal{L}_B = \frac{1}{N}\sum_{i}^{N}\sum_{t=1}^{B} y_{i,t}\log(b(\boldsymbol{\mu}_i^{m_1})_t) + ... + \frac{1}{N}\sum_{i}^{N}\sum_{t=1}^{B} y_{i,t}\log(b(\boldsymbol{\mu}_i^{m_M})_t).$$

The new objective becomes:

$$\begin{cases} \max\limits_{h,b} \mathcal{L}_D + \mathcal{L}_B \\ \min\limits_{f_\kappa, f_i, g_i, i\in\{1,\ldots,M\}} -\mathcal{L}_{\text{ELBO}} + \sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M}\mathcal{L}_{i,j} + \mathcal{L}_D + \mathcal{L}_A + \mathcal{L}_B \end{cases}$$

Model structure and training. All of the neural network components of our model use a single-layer neural network with 128 units, layer normalization, and exponential linear unit activation.[58] Training is a two-step process due to the two objectives, training the discriminators to maximize the adversarial objectives, and then training the encoders and decoders to minimize the augmented VAE objective. We use the Adam optimizer[59] during training, with an initial learning rate of 0.005 and learning rate decay of 0.00006. We use a minibatch size of either 2000 or 5000 since larger minibatch sizes are better for representation learning tasks.[60]

Artificially pairing unpaired multiomics data. Given a joint scRNA-seq and scATAC-seq bridge dataset R with matrices X_R^{RNA} and X_R^{ATAC}, scPairing can generate pairings between unpaired scRNA-seq dataset P and scATAC-seq dataset Q with matrices X_P^{RNA} and X_Q^{ATAC}. As introduced by Hao *et al.*,[29] we first compute a common batch-corrected low-dimensional representation for each modality across datasets. For the scRNA-seq data, X_R^{RNA} and X_P^{RNA} are processed into Y_R^{RNA} and Y_P^{RNA}, where batch effects between R and P are resolved. Likewise, X_R^{ATAC} and X_Q^{ATAC} are processed into batch-corrected representations Y_R^{ATAC} and Y_Q^{ATAC}. Then, we train scPairing on the bridge dataset Y_R^{RNA} and Y_R^{ATAC}. With trained RNA and ATAC encoders, we pass Y_P^{RNA} through the RNA encoder and Y_Q^{ATAC} through the ATAC encoder to obtain embeddings $\boldsymbol{\mu}_P^{\text{RNA}}$ and $\boldsymbol{\mu}_Q^{\text{ATAC}}$ on the common hyperspherical space.

The problem of pairing these cells is treated as a maximum weight matching problem in bipartite graphs, which aims to find a pairing between vertices from one partition to the other such that the sum of edge weights is maximized. We first generate a bipartite graph between cells i from dataset P and j from dataset Q, with the weight of each edge between cells i and j being

$$w(i,j) = \begin{cases} \text{CosineSimilarity}(\boldsymbol{\mu}_{P_i}^{\text{RNA}}, \boldsymbol{\mu}_{Q_j}^{\text{ATAC}}) & \text{if cosine similarity} \geq 1-\varepsilon \\ 0 & \text{otherwise,} \end{cases}$$

for some cutoff adjustment $\varepsilon > 0$. We prune edges between cells with similarity less than $1-\varepsilon$ to prevent dissimilar cells from being paired together during matching. The ε parameter

can be manually tuned in the context of the data. In data with many similar cell types, choosing a smaller ε will prevent similar but distinct cell states from pairing together, such as cells present along a differentiation trajectory.

To solve the maximum weight matching problem, we use the linear sum assignment function from SciPy,[61] which uses a modification of the Jonker-Volgenant algorithm.[62] In large datasets, computing the maximum weight matching is infeasible due to time and memory constraints. In such cases, we find an approximate solution by dividing the cells in both datasets into random disjoint subsets and computing the maximum weight matching on each subset.

Trimodal artificial pairing. Given a joint scRNA-seq, scATAC-seq and epitope dataset bridge dataset R with matrices X_R^{RNA}, X_R^{ATAC}, and X_R^{epitope}, and corresponding unpaired data from datasets P, Q, and S with matrices X_P^{RNA}, X_Q^{ATAC}, and X_S^{epitope}, we can apply scPairing to embed all three modalities onto the common hyperspherical space using the trimodal extension of scPairing. This procedure follows from pairing two modalities, except a tripartite graph is produced, with one partition each for the scRNA-seq, scATAC-seq, and epitope modality cells.

The problem of finding artificial pairings for three modalities is intractable as finding a maximum tripartite graph is NP-hard. In lieu of finding an optimal pairing, we propose a randomized greedy algorithm to generate a respectable pairing. Our algorithm randomly selects a cell i from the modality with the fewest cells (without loss of generality, assume the scRNA-seq dataset has the fewest cells) and finds the cell j from the scATAC-seq dataset and cell k from the epitope dataset that maximizes the pairwise cosine similarity between $\boldsymbol{\mu}_i^{\text{RNA}}$, $\boldsymbol{\mu}_j^{\text{ATAC}}$, and $\boldsymbol{\mu}_k^{\text{epitope}}$. Cells i, j, and k are removed from consideration and the algorithm repeats until all cells are paired or the pairwise cosine similarity drops below a threshold. The full algorithm pseudocode is provided in **Supplementary Note S2**.

Alternatively, if we have a joint scRNA-seq and scATAC-seq dataset P with a separate epitope dataset Q, we can also generate trimodal data. In this case, all three modalities are still embedded on the common hyperspherical space learned by scPairing with bridge dataset R, except we can make a bipartite graph with one partition being the paired scRNA-seq and scATAC-seq cells, and the other partition being the epitope cells. The edge weights between cell i and cell j is

$$w(i,j) = \begin{cases} \text{CosineSimilarity}(\boldsymbol{\mu}_{P_i}^{\text{RNA}}, \boldsymbol{\mu}_{Q_j}^{\text{epitope}}) & \text{if total similarity} \geq 2-\varepsilon \\ + \text{CosineSimilarity}(\boldsymbol{\mu}_{P_i}^{\text{ATAC}}, \boldsymbol{\mu}_{Q_j}^{\text{epitope}}) & \\ 0 & \text{otherwise,} \end{cases}$$

for some cutoff adjustment ε. With this formulation, we can find optimal pairings using the linear sum assignment algorithm.[61] This pairing method can also be applied if we have a joint scRNA-seq and epitope dataset with a separate scATAC-seq dataset.

Benchmarking evaluation. We assessed multiomics data integration using different metrics for data integration and alignment on the NeurIPS 2021 dataset.[36] The evaluation

metrics are described below.

Conservation of biological structure using kNN accuracies. To compare the quality of embeddings produced by each model, we performed both cross-batch validation and ten-fold cross-validation of cell type prediction. Cross-batch cell type prediction takes batch correction into account, as batches must be integrated together for accurate classification. Conversely, ten-fold validation does not account for batch effects. We conducted the k-nearest neighbor predictions with $k = 5, 17, 29, 41, 53$, and 65.

Cell matching. To evaluate whether the model can correctly embed matching measurements from the same cell, we calculated the FOSCTTM score.[38] The FOSCTTM metric indicates the proportion of cells that are closer to the embedding of one modality of a cell than the embedding of the other modality for that same cell. Since FOSCTTM is not symmetric, the final metric is the average between the two FOSCTTMs. A lower FOSCTTM indicates better cell matching.

Modality mixing. We check that the embeddings for each modality in the multiomics dataset mixes together using the Local Inverse Simpson's Index (LISI) metric. We use the graph iLISI implementation from scIB.[35]

Cross-modality imputation. We evaluate the ability of each model to impute across modalities by training the model on three out of four sites and imputing using the held-out site. We quantified the scRNA-seq imputations using the Pearson correlation between the true and imputed expression. We quantified the scATAC-seq imputations using the binary cross-entropy between the binarized true accessibility and the imputed accessibility.

Comparison methods. We compare our method to five other multiomics tools: MultiVI,[13] MoETM,[17] MIRA,[19] Cobolt,[15] and CLUE.[39] Of these, Cobolt and CLUE were designed for modality alignment. We ran the methods based on their publicly available tutorials, or if no tutorials exist, based on their default settings.

Evaluating the quality of retina pairings. With unpaired data, no ground truth pairing exists to evaluate the quality of the computed pairings. One common analysis performed on joint chromatin accessibility and gene expression data is inferring transcription factor binding by identifying correlated genes and peaks. Thus, we compare the preservation of gene-peak correlations found in a true multiomics dataset to indicate realistic pairings.

We compare our pairings against two other methods. The first is gene activities computed with Signac,[44] which directly transforms scATAC-seq fragments into gene counts by counting fragments overlapping each gene plus 2kb upstream. The second method uses canonical correlation analysis implemented in the Seurat[43] `FindTransferAnchors` function to integrate the gene activities onto the scRNA-seq data. Then, we use `TransferData` to impute the gene expression of the scATAC-seq cells using the scRNA-seq data as a reference. This second method is the strategy applied by Liang *et al.* in their analysis of gene-peak correlations. We ran both these methods on the same set of scATAC-seq cells that were present in our artificial pairings.

We focus on the retinal bipolar cells as they were also analyzed for gene-peak correlations

in Liang *et al.*[41] In particular, Liang *et al.* analyzed the correlation of gene *GNB3* expression and peak chr12:6959945-6962406 accessibility across all retina bipolar cell types. We compare the means and standard deviations of gene-peak correlations in twelve retina bipolar cell subtypes. We choose to compute correlations for all genes with peaks within 250kb of the gene TSS. Since not all genes are imputed by gene activities, we only compute gene-peak correlations for genes present in the multiomics data, the artificially paired data, and gene activities.

We use metacells, specifically the SEACells algorithm, to compute gene-peak correlations as it has been shown that metacells more faithfully capture relationships between the two modalities.[45] To obtain a sufficient number of metacells to compute meaningful correlations, we run SEACells using a target of ten cells per metacell. We modify the maximum iterations of SEACells to 400, but use otherwise default parameters.

We quantify the difference between the gene-peak means and standard deviation distributions using optimal transport, which measures the distance between two distributions. We calculate both the Earth Mover distance and Sinkhorn distance using the Python Optimal Transport library.[63] As the optimal transport is not scalable to the number of gene-peak pairs, we randomly sample 2% of the gene-peak pairs to perform optimal transport, and repeat for 100 trials.

Trimodal pairing evaluation. We evaluate the similarity of the re-pairings to the original pairings using three methods: FOSCTTM, the percentile rank of the true pairing among all possible pairings, and the similarity between clusters in the original data and re-paired data. The FOSCTTM is calculated as the average FOSCTTM of the three pairs of unique modalities.

The percentile rank of the true pairing calculates the cosine similarity rank of the true pairing, scRNA-seq, scATAC-seq, and epitope data all from cell i, among all possible pairings. Let $\mathbf{z}_i^{\text{RNA}}$, $\mathbf{z}_i^{\text{ATAC}}$, $\mathbf{z}_i^{\text{epitope}}$ be the embeddings for cell i for each of the three modalities. We define the triplet similarity for cells i, j, and k to be

$$\text{Tri-SIM}(i,j,k) = d(\mathbf{z}_i^{\text{RNA}}, \mathbf{z}_j^{\text{ATAC}}) + d(\mathbf{z}_i^{\text{RNA}}, \mathbf{z}_k^{\text{epitope}}) + d(\mathbf{z}_j^{\text{ATAC}}, \mathbf{z}_k^{\text{epitope}}),$$

where d is a similarity measure (in this case, cosine similarity), as the embeddings are located on the unit hypersphere. Then, for cell i, we calculate the percentile rank of $\text{Tri-SIM}(i,i,i)$ in

$$\{\text{Tri-SIM}(i,j,k) \mid j,k \in \{1,...,N_{\text{cells}}\}\}.$$

A high percentile rank triplet indicates that the true triplet match is very close to the best possible match, meaning the embeddings closely follow the true pairings.

Cluster similarity is measured using the ARI and NMI metrics. High ARI and NMI indicate that clusters formed in the re-pairings and pairings share commonalities, and the structure of the original data is preserved in the re-pairings. The ARI and NMI are calculated for each modality in the re-pairing as one artificially paired cell is composed of three cells, one for each modality.

Data preprocessing. Our computational experiments start with raw gene expression, chromatin accessibility, and epitope matrices. The chromatin accessibility matrices are loaded directly from the source, or computed from fragment files. The data are then transformed into low-dimensional representations, which are used as input to scPairing. Full details regarding preprocessing methods are available in **Supplementary Note S3**.

Data availability. The 10X Multiome BMMCs from the Multimodal Single-Cell Data Integration Challenge at NeurIPS 2021 is available via GEO accession GSE194122.[36] The SHARE-seq mouse skin dataset is available via GEO accession GSE140203.[2] The TEA-seq dataset is available via GEO accession GSE158013.[25] The DOGMA-seq dataset is available via GEO accession GSE156478.[24]

The pre-trained model checkpoints for CellPLM and scGPT were downloaded from the authors' sources on Dropbox[64] and Google Drive,[65] respectively.

The scRNA-seq raw counts data from the Wang *et al.* paired retina dataset is available via the UCSC Cell Browser[66] dataset ID chang-retina-atlas and the scATAC-seq fragment files are available via GEO accession GSE196235.[42] The snRNA-seq raw counts data and snATAC-seq fragment files from the Liang *et al.* unpaired retina dataset is available via the UCSC Cell Browser[66] dataset ID retina-atac.[41]

The paired Ulezko Antonova *et al.* bridge dataset of human tonsil cells is available via GEO accession GSE247692.[46] The paired and unpaired human tonsil cell atlas is available via Zenodo.[48,67] The paired human intestine dataset is available via HuBMAP ID HBM692.JRZB.356, while the data matrices and fragment files are available through Data Dryad via DOI 10.5061/dryad.0zpc8672f.[49]

Code availability. The scPairing model is available on GitHub (https://github.com/Ding-Group/scPairing). The scripts and computational notebooks used to produce and analyze the results are deposited in Zenodo (https://doi.org/10.5281/zenodo.13910219).

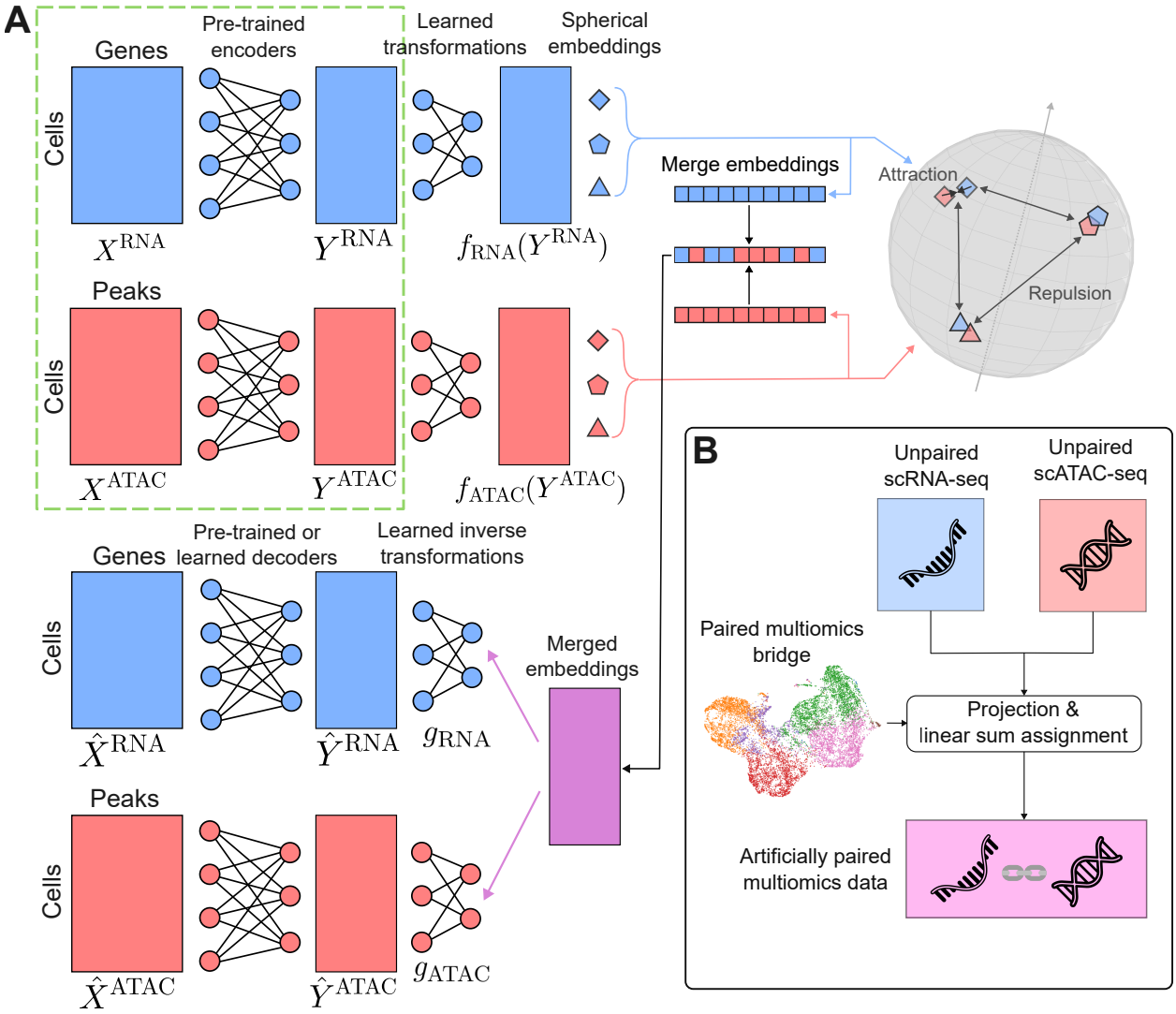

Figure 1: scPairing model overview. (A) Schematic of scPairing. Pre-trained encoders compute low-dimension embeddings Y^{RNA} and Y^{ATAC} that scPairing transforms onto a common hyperspherical embedding space. The dashed-green box indicates that the low-dimensional representations are assumed to be computed prior to running scPairing. scPairing applies a contrastive loss that encourages embeddings of two different modalities to embed close together, while repelling different cells' embeddings. The two embeddings are merged and decoded back to their low-dimensional representations. Either a pre-trained decoder or a learned decoder reconstructs the count matrices for each modality. (B) Training scPairing on a multiomics bridge enables artificial pairing of independent scRNA-seq and scATAC-seq datasets. The scPairing encoders co-embed the two modalities in the same space. Solving the linear sum assignment problem with the similarities between scRNA-seq and scATAC-seq embeddings produces artificially paired data.

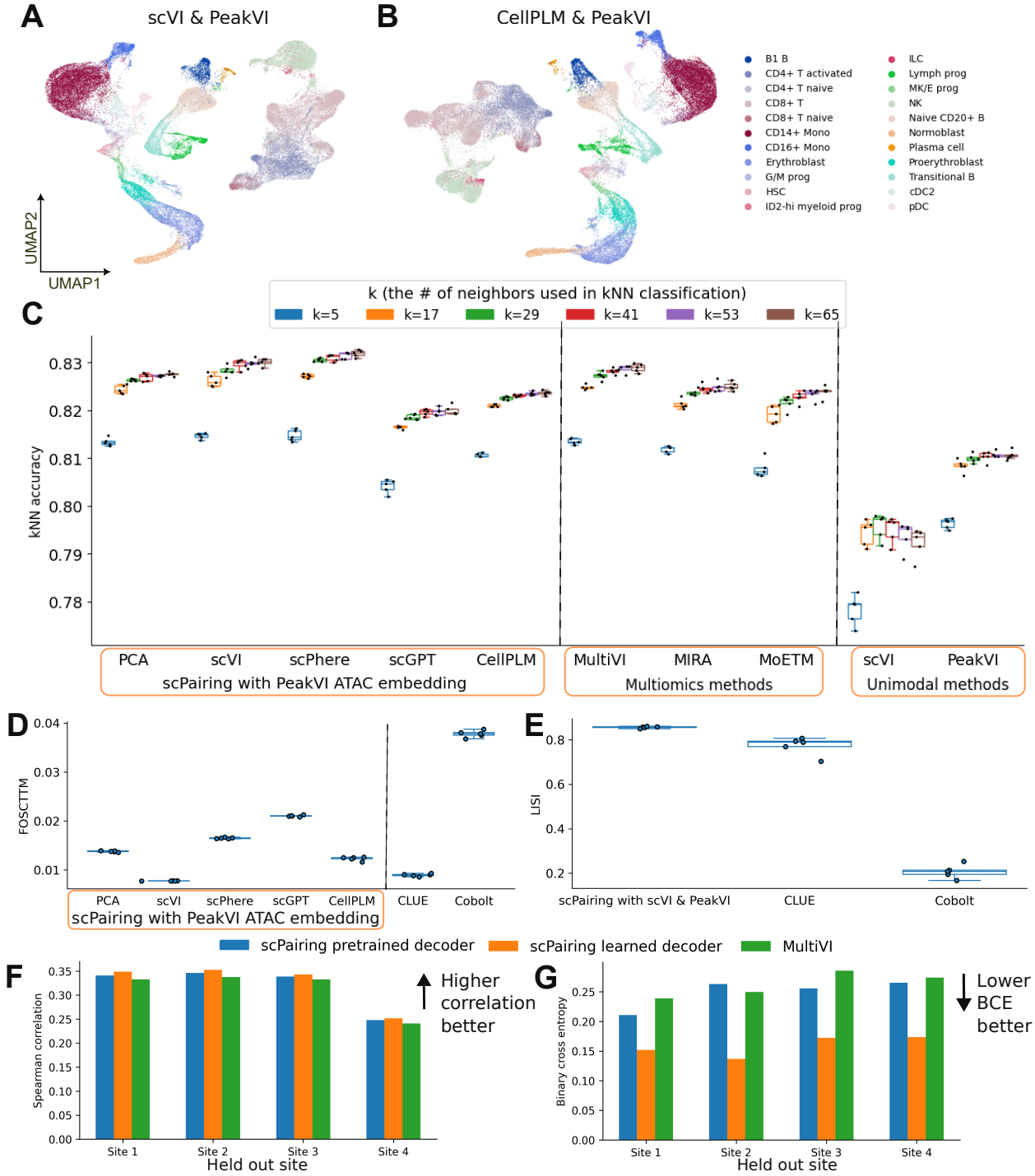
A
scVI & PeakVI
B
CellPLM & PeakVI
UMAP2
UMAP1
B1 B
CD4+ T activated
CD4+ T naive
CD8+ T
CD8+ T naive
CD14+ Mono
CD16+ Mono
Erythroblast
G/M prog
HSC
ID2-hi myeloid prog
ILC
Lymph prog
MK/E prog
NK
Naive CD20+ B
Normoblast
Plasma cell
Proerythroblast
Transitional B
cDC2
pDC
C
k (the # of neighbors used in kNN classification)
k=5
k=17
k=29
k=41
k=53
k=65
kNN accuracy
PCA
scVI
scPhere
scGPT
CellPLM
scPairing with PeakVI ATAC embedding
MultiVI
MIRA
MoETM
Multiomics methods
scVI
PeakVI
Unimodal methods
D
FOSCTTM
PCA
scVI
scPhere
scGPT
CellPLM
scPairing with PeakVI ATAC embedding
CLUE
Cobolt
E
LISI
scPairing with scVI & PeakVI
CLUE
Cobolt
scPairing pretrained decoder
scPairing learned decoder
MultiVI
F
Spearman correlation
Higher correlation better
Site 1
Site 2
Site 3
Site 4
Held out site
G
Binary cross entropy
Lower BCE better
Site 1
Site 2
Site 3
Site 4
Held out site

Figure 2: Integration of joint scRNA-seq and scATAC-seq data. (A and B) UMAP visualization of scPairing merged representations from scVI (A) or CellPLM (B), and PeakVI of a human BMMC joint scRNA-seq and scATAC-seq dataset. (C) Comparison of biological structure preservation quantified by k-nearest neighbor cell type classification accuracies, where we predicted cell types for one batch given the cells from the remaining batches. Each dot represents a trial of each method at a particular k. The boxplots denote the medians and interquartile ranges (IQRs), while the whiskers indicate the lowest and highest data point within 1.5 IQR of the lower and upper quartile, respectively. (D) Comparison of modality alignment with FOSCTTM computed across all cells. Each dot denotes an average FOSCTTM of one trial. (E) Comparison of modality mixing with LISI computed across all cells. Each dot denotes the LISI of one trial. (F) Comparison of cross-site scATAC-seq to scRNA-seq imputation with Spearman correlations between imputed and true gene counts. (G) Comparison of cross-site scRNA-seq to scATAC-seq imputation with binary cross-entropy between the imputed and true binarized peaks.

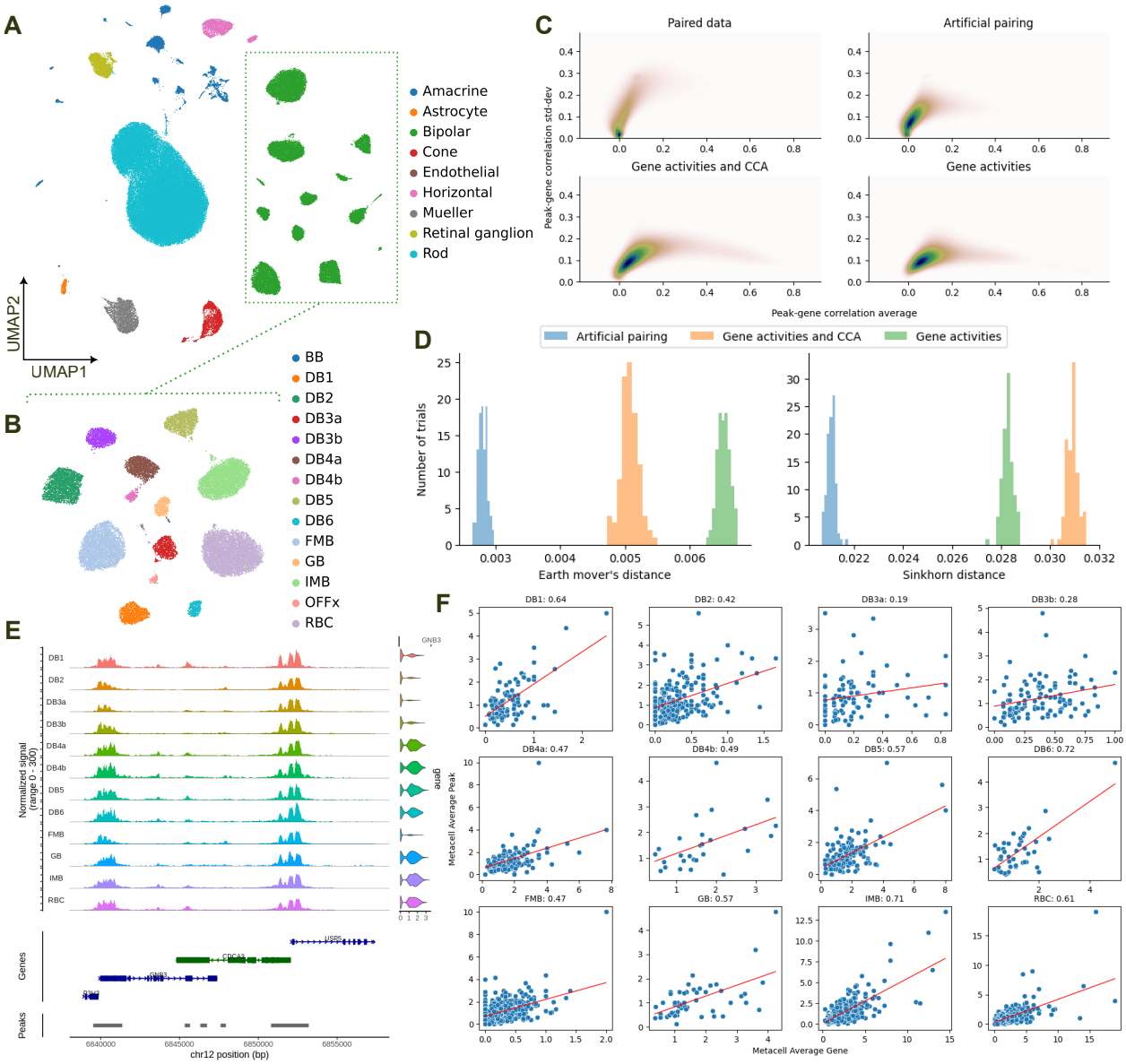

Figure 3: Artificial pairing of human retinal data. (A) UMAP visualization of the artificially paired human retinal cells from Liang *et al.*[41], colored by major cell type. (B) UMAP visualization of the retinal bipolar subset from the artificial pairing, colored by cell subtype. (C) Comparison of density plots of peak-gene correlation means and standard deviations across the retinal bipolar subtypes. We computed the means and standard deviations on the Wang *et al.* multiomics data[42] (top-left), scPairing artificial pairings of the Liang *et al.* unpaired data[41] (top-right), imputed gene expression of the Liang *et al.* scATAC-seq data using gene activities followed by CCA with scRNA-seq data (bottom-left), and imputed gene expression of the Liang *et al.* scATAC-seq data using only gene activities (bottom-right). Correlations were computed using metacells produced by SEACells[45]. (D) Distances between the distribution of gene-peak correlation means and standard deviations for the Liang *et al.* dataset computed from the artificial pairings, imputation using gene activities and CCA, and imputation using only gene activities, compared to the distribution from the Wang *et al.* dataset. Distances were calculated using optimal transport, considering both the Earth Mover distance and Sinkhorn distance. (E) Coverage plot of retinal bipolar subtypes at the *GNB3* TSS. (F) Correlation of average *GNB3* gene expression and chr12:6850801-6853197 accessibility across 12 retinal bipolar subtypes aggregated into metacells.

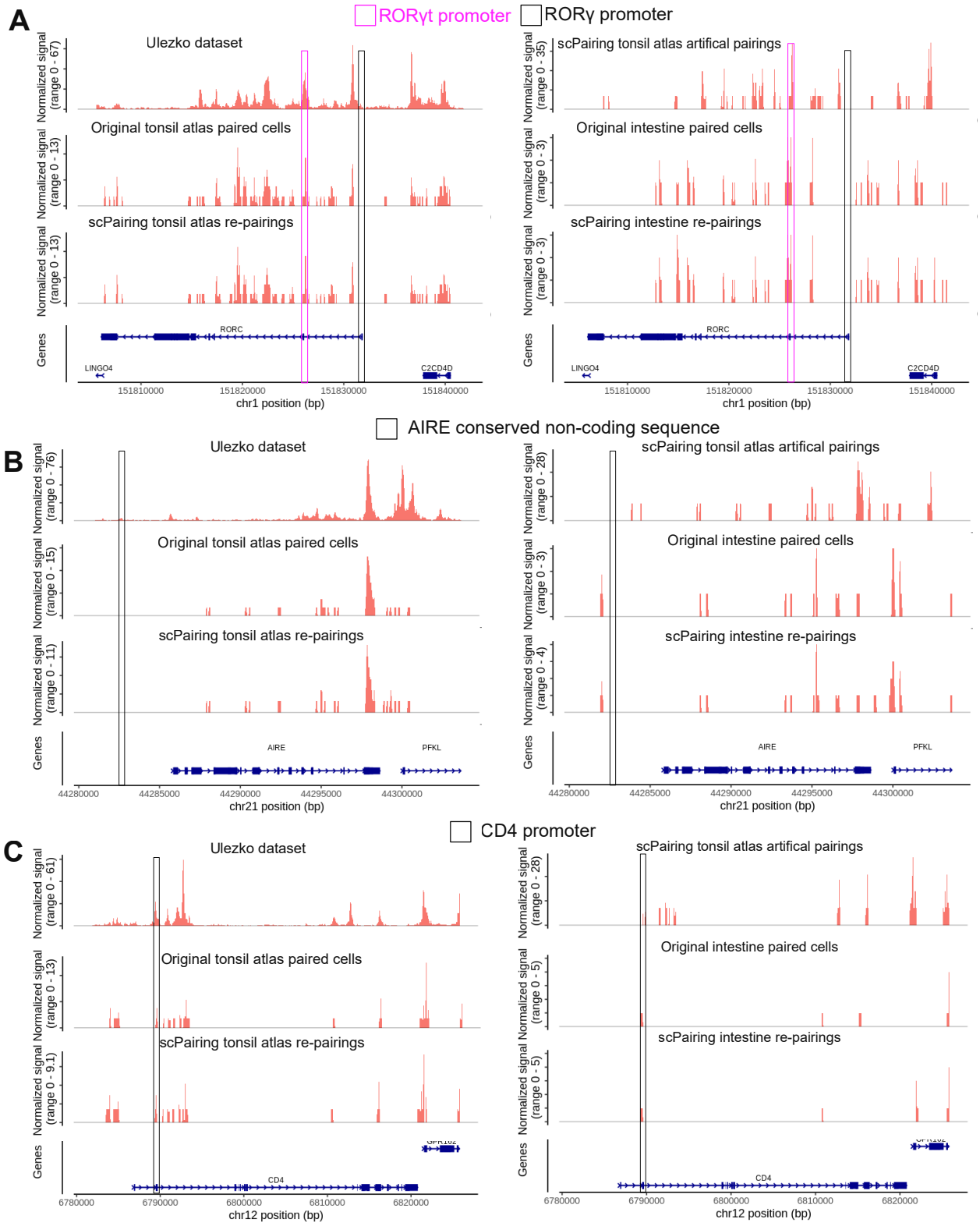
A
RORγt promoter
RORγ promoter
Ulezko dataset
Original tonsil atlas paired cells
scPairing tonsil atlas re-pairings
scPairing tonsil atlas artifical pairings
Original intestine paired cells
scPairing intestine re-pairings
Normalized signal (range 0 - 67)
Normalized signal (range 0 - 13)
Normalized signal (range 0 - 13)
Normalized signal (range 0 - 35)
Normalized signal (range 0 - 3)
Normalized signal (range 0 - 3)
Genes
RORC
LINGO4
C2CD4D
151810000
151820000
151830000
151840000
chr1 position (bp)
AIRE conserved non-coding sequence
B
Ulezko dataset
Original tonsil atlas paired cells
scPairing tonsil atlas re-pairings
scPairing tonsil atlas artifical pairings
Original intestine paired cells
scPairing intestine re-pairings
Normalized signal (range 0 - 76)
Normalized signal (range 0 - 15)
Normalized signal (range 0 - 11)
Normalized signal (range 0 - 28)
Normalized signal (range 0 - 3)
Normalized signal (range 0 - 4)
Genes
AIRE
PFKL
44280000
44285000
44290000
44295000
44300000
chr21 position (bp)
CD4 promoter
C
Ulezko dataset
Original tonsil atlas paired cells
scPairing tonsil atlas re-pairings
scPairing tonsil atlas artifical pairings
Original intestine paired cells
scPairing intestine re-pairings
Normalized signal (range 0 - 61)
Normalized signal (range 0 - 13)
Normalized signal (range 0 - 9.1)
Normalized signal (range 0 - 28)
Normalized signal (range 0 - 5)
Normalized signal (range 0 - 5)
Genes
CD4
6780000
6790000
6800000
6810000
6820000
chr12 position (bp)

Figure 4: Chromatin accessibility at the *RORC*, *AIRE*, and *CD4* loci in R-DC-like cells. (A–C) Accessibility from the *RORC* locus (A), *AIRE* locus (B), and *CD4* locus (C) from the paired Ulezko Antonova *et al.* dataset that originally described these R-DC-like cells[46] (top-left), paired tonsil cell atlas data[48] (middle-left), scPairing's re-pairing of the paired tonsil cell atlas (bottom-left), scPairing's artificial pairing of the unpaired tonsil cell atlas (top-right), paired intestine data[49] (middle-right), and scPairing's re-pairing of the paired intestine data (bottom-right). In (A), the pink box indicates the *ROR*γt promoter region and the black box indicates the *ROR*γ promoter. In (B), the black box indicates a conserved non-coding sequence of *AIRE*. In (C), the black box indicates a *CD4* promoter region.

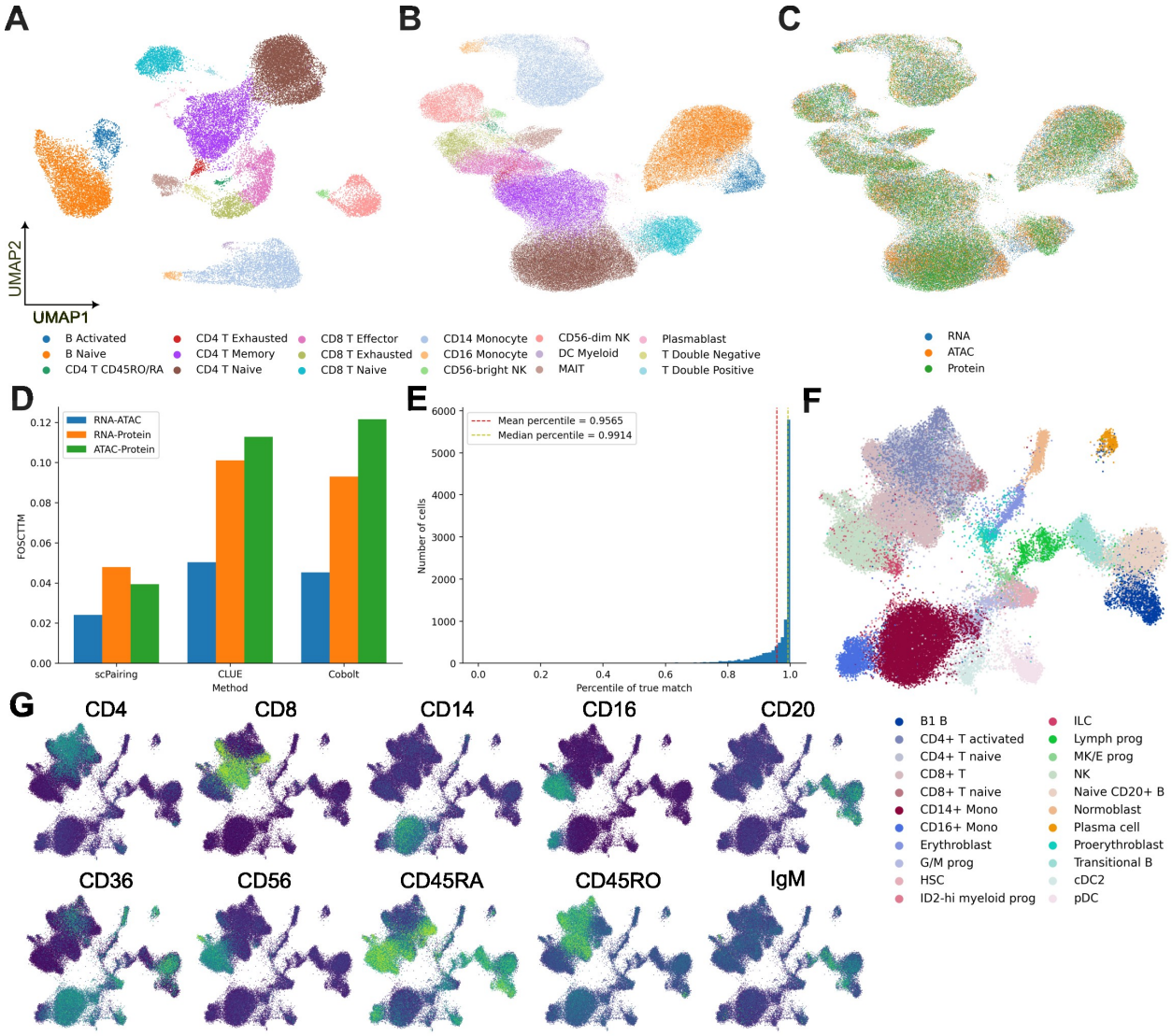

Figure 5: Integration of joint scRNA-seq, scATAC-seq and cell epitope data. (A) UMAP visualization of scPairing representations of the TEA-seq data, colored by manually annotated cell types. Each dot represents a cell. (B and C) UMAP visualization of scPairing representations of all three modalities combined together, colored by manual cell type annotation (B) and modality (C). Each dot represents a cell from a specific modality. (D) Comparison of pairwise-modality FOSCTTM on TEA-seq embeddings. (E) Histogram of percentile rank of true match among all possible pairings for DOGMA-seq cells projected onto the TEA-seq bridge data. The yellow dashed line denotes the median percentile rank, while the red dashed line denotes the mean percentile rank. (F) UMAP visualization of the scRNA-seq modality of the artificially generated trimodal BMMCs, colored by cell type. (G) Marker epitope expression in our artificially generated trimodal BMMCs.

References

1. Chen, S., Lake, B.B., and Zhang, K. (2019). High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nature Biotechnology *37*, 1452–1457.

2. Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T. et al. (2020). Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell *183*, 1103–1116.

3. Belhocine, K., DeMare, L., and Habern, O. (2021). Single-cell multiomics: Simultaneous epigenetic and transcriptional profiling: 10x Genomics shares experimental planning and sample preparation tips for the Chromium Single Cell Multiome ATAC+ Gene Expression system. Genetic Engineering & Biotechnology News *41*, 66–68.

4. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nature Methods *14*, 865–868.

5. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J. et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications *8*, 14049.

6. Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature Protocols *9*, 171–181.

7. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

8. Badia-i Mompel, P., Wessels, L., Müller-Dott, S., Trimbour, R., Ramirez Flores, R.O., Argelaguet, R., and Saez-Rodriguez, J. (2023). Gene regulatory network inference in the era of single-cell multi-omics. Nature Reviews Genetics *24*, 739–754.

9. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nature Reviews Molecular Cell Biology *24*, 695–713.

10. Zhang, K., Hocker, J.D., Miller, M., Hou, X., Chiou, J., Poirion, O.B., Qiu, Y., Li, Y.E., Gaulton, K.J., Wang, A. et al. (2021). A single-cell atlas of chromatin accessibility in the human genome. Cell *184*, 5985–6001.

11. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P. et al. (2017). The human cell atlas. eLife *6*, e27041.

12. Booeshaghi, A.S., Gao, F., and Pachter, L. Assessing the multimodal tradeoff (2021). doi: 10.1101/2021.12.08.471788 preprint at *bioRxiv* https://doi.org/10.1101/2021.12.08.471788.

13. Ashuach, T., Gabitto, M.I., Koodli, R.V., Saldi, G.A., Jordan, M.I., and Yosef, N. (2023). MultiVI: deep generative model for the integration of multimodal data. Nature Methods *20*, 1222–1231.

14. Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K.L., Streets, A., and Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nature Methods *18*, 272–282.

15. Gong, B., Zhou, Y., and Purdom, E. (2021). Cobolt: integrative analysis of multimodal single-cell sequencing data. Genome Biology *22*, 351.

16. Minoura, K., Abe, K., Nam, H., Nishikawa, H., and Shimamura, T. (2021). scMM: Mixture-of-experts multimodal deep generative model for single-cell multiomics data analysis. Cell Reports Methods *1*, 100071.

17. Zhou, M., Zhang, H., Bai, Z., Mann-Krzisnik, D., Wang, F., and Li, Y. (2023). Single-cell multi-omics topic embedding reveals cell-type-specific and COVID-19 severity-related immune signatures. Cell Reports Methods *3*, 100563.

18. Li, G., Fu, S., Wang, S., Zhu, C., Duan, B., Tang, C., Chen, X., Chuai, G., Wang, P., and Liu, Q. (2022). A deep generative model for multi-view profiling of single-cell RNA-seq and ATAC-seq data. Genome Biology *23*, 20.

19. Lynch, A.W., Theodoris, C.V., Long, H.W., Brown, M., Liu, X.S., and Meyer, C.A. (2022). MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. Nature Methods *19*, 1097–1108.

20. Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J.C., and Stegle, O. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biology *21*, 111.

21. Tang, X., Zhang, J., He, Y., Zhang, X., Lin, Z., Partarrieu, S., Hanna, E.B., Ren, Z., Shen, H., Yang, Y. et al. (2023). Explainable multi-task learning for multi-modality biological data analysis. Nature Communications *14*, 2546.

22. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M. et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.

23. Xiong, L., Chen, T., and Kellis, M. (2023). scCLIP: Multi-modal single-cell contrastive learning integration pre-training. In NeurIPS 2023 AI for Science Workshop. URL: https://openreview.net/forum?id=KMtM5ZHxct.

24. Mimitou, E.P., Lareau, C.A., Chen, K.Y., Zorzetto-Fernandes, A.L., Hao, Y., Takeshima, Y., Luo, W., Huang, T.S., Yeung, B.Z., Papalexi, E. et al. (2021). Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nature Biotechnology *39*, 1246–1258.

25. Swanson, E., Lord, C., Reading, J., Heubeck, A.T., Genge, P.C., Thomson, Z., Weiss, M.D., Li, X.j., Savage, A.K., Green, R.R. et al. (2021). Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. eLife *10*, e63632.

26. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. et al. (2021). Learning transferable visual models from natural language supervision. In International Conference on Machine Learning. PMLR pp. 8748–8763.

27. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research *15*, 1929–1958.

28. Radhakrishnan, A., Friedman, S.F., Khurshid, S., Ng, K., Batra, P., Lubitz, S.A., Philippakis, A.A., and Uhler, C. (2023). Cross-modal autoencoder framework learns holistic representations of cardiovascular state. Nature Communications *14*, 2436.

29. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C. et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology *42*, 293–304.

30. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nature Methods *15*, 1053–1058.

31. Ding, J., and Regev, A. (2021). Deep generative model embedding of single-cell RNA-seq profiles on hyperspheres and hyperbolic spaces. Nature Communications *12*, 2554.

32. Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods *21*, 1470–1480.

33. Wen, H., Tang, W., Dai, X., Ding, J., Jin, W., Xie, Y., and Tang, J. (2024). CellPLM: Pre-training of cell language model beyond single cells. In The Twelfth International Conference on Learning Representations.

34. Ashuach, T., Reidenbach, D.A., Gayoso, A., and Yosef, N. (2022). PeakVI: A deep generative model for single-cell chromatin accessibility analysis. Cell Reports Methods *2*, 100182.

35. Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Müller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M. et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. Nature Methods *19*, 41–50.

36. Luecken, M.D., Burkhardt, D.B., Cannoodt, R., Lance, C., Agrawal, A., Aliee, H., Chen, A.T., Deconinck, L., Detweiler, A.M., Granados, A.A. et al. (2021). A sandbox for prediction and integration of DNA, RNA, and proteins in single cells. In Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2).

37. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.r., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods *16*, 1289–1296.

38. Liu, J., Huang, Y., Singh, R., Vert, J.P., and Noble, W.S. (2019). Jointly embedding multiple single-cell omics measurements. In 19th International Workshop on Algorithms in Bioinformatics vol. 143. NIH Public Access.

39. Tu, X., Cao, Z.J., Mostafavi, S., and Gao, G. (2022). Cross-linked unified embedding for cross-modality representation learning. Advances in Neural Information Processing Systems *35*, 15942–15955.

40. Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'Er, D. (2019). Characterization of cell fate probabilities in single-cell data with palantir. Nature Biotechnology *37*, 451–460.

41. Liang, Q., Cheng, X., Wang, J., Owen, L., Shakoor, A., Lillvis, J.L., Zhang, C., Farkas, M., Kim, I.K., Li, Y. et al. (2023). A multi-omics atlas of the human retina at single-cell resolution. Cell Genomics *3*, 100298.

42. Wang, S.K., Nair, S., Li, R., Kraft, K., Pampari, A., Patel, A., Kang, J.B., Luong, C., Kundaje, A., and Chang, H.Y. (2022). Single-cell multiome of the human retina and deep learning nominate causal variants in complex eye diseases. Cell Genomics *2*, 100164.

43. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology *36*, 411–420.

44. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nature Methods *18*, 1333–1341.

45. Persad, S., Choo, Z.N., Dien, C., Sohail, N., Masilionis, I., Chaligné, R., Nawy, T., Brown, C.C., Sharma, R., Pe'er, I. et al. (2023). SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data. Nature Biotechnology *41*, 1746–1757.

46. Ulezko Antonova, A., Lonardi, S., Monti, M., Missale, F., Fan, C., Coates, M.L., Bugatti, M., Jaeger, N., Fernandes Rodrigues, P., Brioschi, S. et al. (2023). A distinct human cell type expressing MHCII and RORγt with dual characteristics of dendritic cells and type 3 innate lymphoid cells. Proceedings of the National Academy of Sciences *120*, e2318710120.

47. Ding, J., Garber, J.J., Uchida, A., Lefkovith, A., Carter, G.T., Vimalathas, P., Canha, L., Dougan, M., Staller, K., Yarze, J. et al. (2024). An esophagus cell atlas reveals dynamic rewiring during active eosinophilic esophagitis and remission. Nature Communications *15*, 3344.

48. Massoni-Badosa, R., Aguilar-Fernández, S., Nieto, J.C., Soler-Vila, P., Elosua-Bayes, M., Marchese, D., Kulis, M., Vilas-Zornoza, A., Bühler, M.M., Rashmi, S. et al. (2024). An atlas of cells in the human tonsil. Immunity *57*, 379–399.

49. Hickey, J.W., Becker, W.R., Nevins, S.A., Horning, A., Perez, A.E., Zhu, C., Zhu, B., Wei, B., Chiu, R., Chen, D.C. et al. (2023). Organization of the human intestine at single-cell resolution. Nature *619*, 572–584.

50. Rosen, Y., Roohani, Y., Agrawal, A., Samotorcan, L., Consortium, T.S., Quake, S.R., and Leskovec, J. Universal cell embeddings: A foundation model for cell biology (2023). doi: 10.1101/2023.11.28.568918 preprint at *bioRxiv* https://doi.org/10.1101/2023.11.28.568918.

51. Gao, Z., Liu, Q., Zeng, W., Jiang, R., and Wong, W.H. (2023). EpiGePT: a Pretrained Transformer model for epigenomics. bioRxiv. doi: 10.1101/2023.07.15.549134.

52. Rosen, Y., Brbić, M., Roohani, Y., Swanson, K., Li, Z., and Leskovec, J. (2024). Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. Nature Methods pp. 1492–1500.

53. Hrovatin, K., Moinfar, A.A., Zappia, L., Lapuerta, A.T., Lengerich, B., Kellis, M., and Theis, F.J. Integrating single-cell RNA-seq datasets with substantial batch effects (2023). Preprint at *bioRxiv* https://doi.org/10.1101/2023.11.03.565463.

54. Fan, Y., Li, Y., Ding, J., and Li, Y. (2024). GFETM: Genome Foundation-based Embedded Topic Model for scATAC-seq Modeling. In International Conference on Research in Computational Molecular Biology. Springer pp. 314–319.

55. Bravo González-Blas, C., Minnoye, L., Papasokrati, D., Aibar, S., Hulselmans, G., Christiaens, V., Davie, K., Wouters, J., and Aerts, S. (2019). cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. Nature Methods *16*, 397–400.

56. Yang, F., Wang, W., Wang, F., Fang, Y., Tang, D., Huang, J., Lu, H., and Yao, J. (2022). scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. Nature Machine Intelligence *4*, 852–866.

57. De Cao, N., and Aziz, W. (2020). The Power Spherical distribution. In Second workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models.

58. Clevert, D., Unterthiner, T., and Hochreiter, S. (2016). Fast and accurate deep network learning by exponential linear units (ELUs). In 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings.

59. Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

60. Chen, C., Zhang, J., Xu, Y., Chen, L., Duan, J., Chen, Y., Tran, S., Zeng, B., and Chilimbi, T. (2022). Why do we need large batchsizes in contrastive learning? A gradient-bias perspective. Advances in Neural Information Processing Systems *35*, 33860–33875.

61. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods *17*, 261–272.

62. Crouse, D.F. (2016). On implementing 2D rectangular assignment algorithms. IEEE Transactions on Aerospace and Electronic Systems *52*, 1679–1696.

63. Flamary, R., Courty, N., Gramfort, A., Alaya, M.Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N. et al. (2021). POT: Python Optimal Transport. Journal of Machine Learning Research *22*, 1–8.

64. Ding, J. (2023). CellPLM - Dropbox. . URL: https://www.dropbox.com/scl/fo/i5rmxgtqzg7iykt2e9uqm/h?dl=0.

65. Pang, K. (2023). scGPT_CP. . URL: https://drive.google.com/drive/folders/1_GROJTzXiAV8HB4imruOTk6PEGuNOcgB.

66. Speir, M.L., Bhaduri, A., Markov, N.S., Moreno, P., Nowakowski, T.J., Papatheodorou, I., Pollen, A.A., Raney, B.J., Seninge, L., Kent, W.J., and Haeussler, M. (2021). UCSC Cell Browser: visualize your single-cell data. Bioinformatics *37*, 4578–4580.

67. Massoni-Badosa, R. (2023). Seurat objects associated with the tonsil cell atlas. Zenodo. URL: https://doi.org/10.5281/zenodo.8373756. doi: 10.5281/zenodo.8373756.

68. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F. et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Research *34*, D590–D598.

69. Chari, T., and Pachter, L. (2023). The specious art of single-cell genomics. PLOS Computational Biology *19*, e1011288.

70. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biology *19*, 15.

71. Bredikhin, D., Kats, I., and Stegle, O. (2022). MUON: multimodal omics analysis framework. Genome Biology *23*, 42.

72. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biology *9*, R137.

Acknowledgements

We thank Minuk Ma for his comments on the manuscript.

This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada, and a department startup fund from the University of British Columbia (to J.D.). J.D. is a Canada Research Chair and is supported by the Canadian Institutes of Health Research through the Canada Research Chair Program. The computational resource is partially supported by the Canada Foundation for Innovation & John. R. Evans Leader Fund (to J.D.). This research was supported in part through the computational resources and services provided by Advanced Research Computing at the University of British Columbia. J.N. is supported by a Master's scholarship from NSERC and a graduate scholarship from the province of British Columbia.

Author's contributions

J.N. and J.D. conceived the study. J.N. designed and wrote the code for scPairing, ran computational experiments, and interpreted the results. J.D. supervised the project. J.N. and J.D. approved the final manuscript.

Competing financial interests

The authors declare no competing interests.

Supplementary Figures

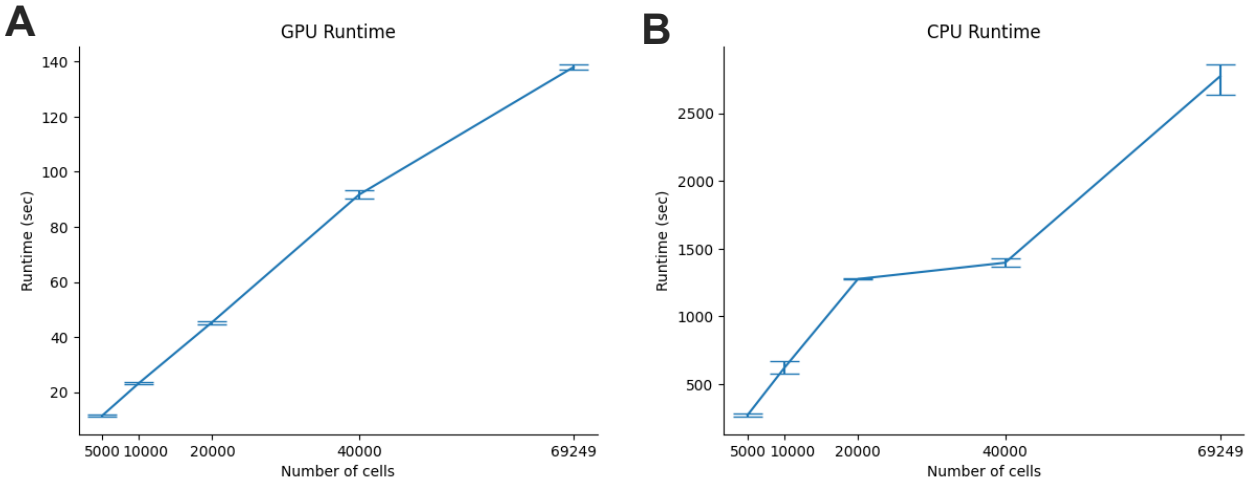

Figure S1: Runtime of scPairing. (A and B) Runtime using a single NVIDIA Tesla V100 GPU (A) or a single Intel Xeon Gold 6130 CPU (B) in seconds across five dataset sizes. Each dataset size was run for three trials, with the average runtime plotted. The error bars indicate the maximum and minimum times across the three trials.

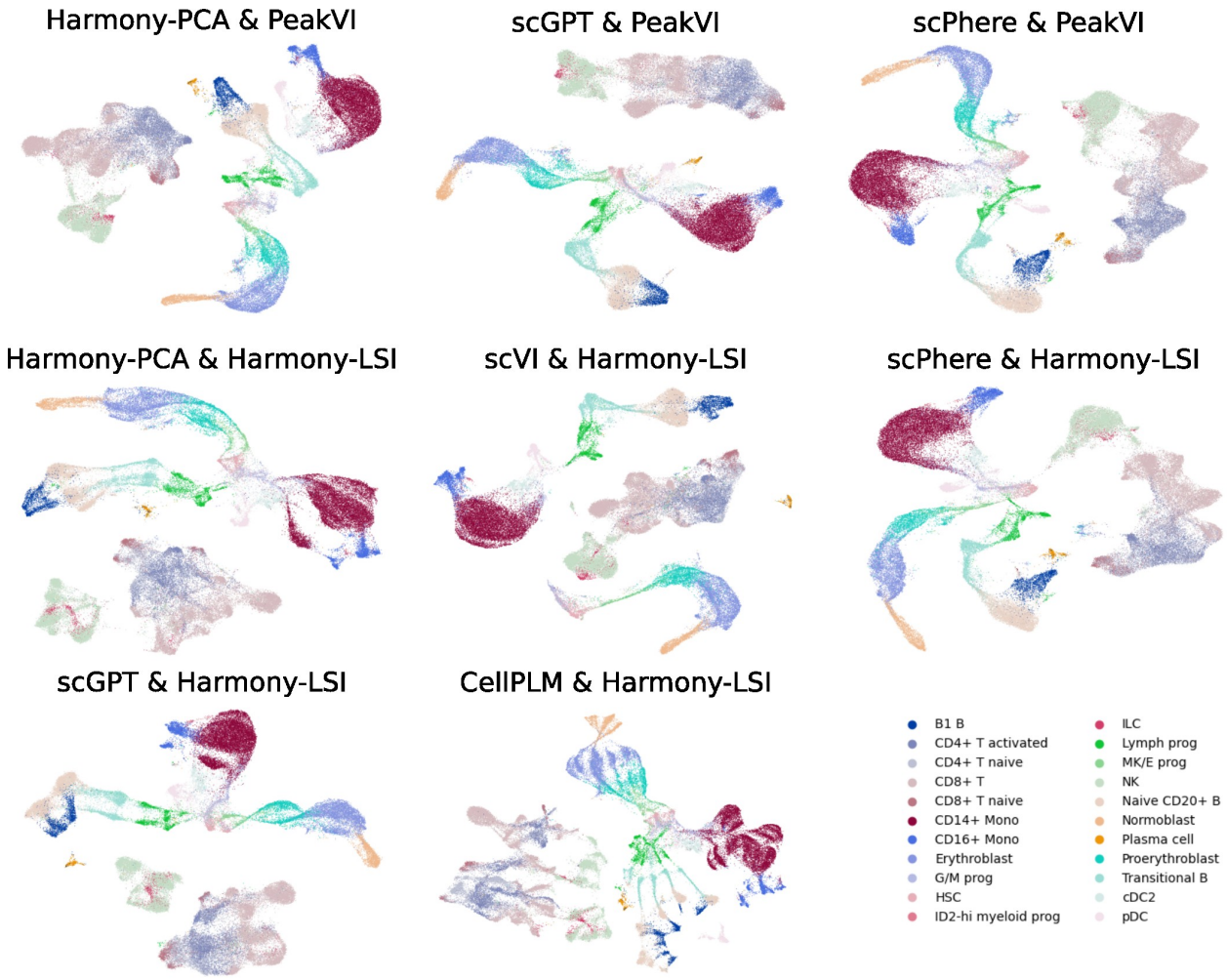

Figure S2: UMAP visualizations of scPairing embeddings. UMAP visualizations of scPairing embeddings with different scRNA-seq and scATAC-seq representations as inputs.

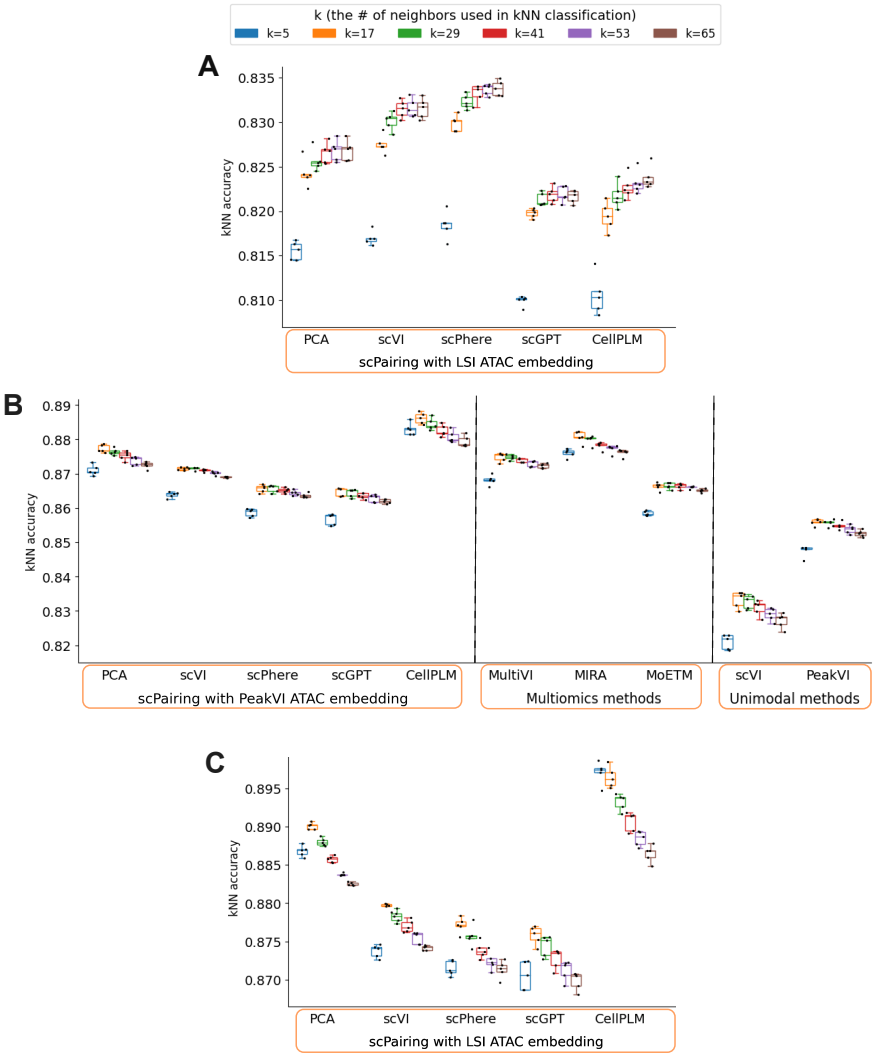

Figure S3: Benchmarking scPairing on the BMMC multiomics dataset. (A) Same as Figure 2C, except with a Harmony-corrected LSI ATAC embedding instead of PeakVI. (B) Comparison of biological structure preservation as quantified by 10-fold k-nearest neighbor cell type classification accuracies. We compared scPairing using five different scRNA-seq representations and a PeakVI scATAC-seq representation. Each dot represents a trial at a particular k. (C) Same as (A) except with a Harmony-corrected LSI ATAC embedding instead of PeakVI. The boxplots denote the medians and interquartile ranges, while the whiskers indicate the lowest and highest data point within 1.5 IQR of the lower and upper quartile, respectively.

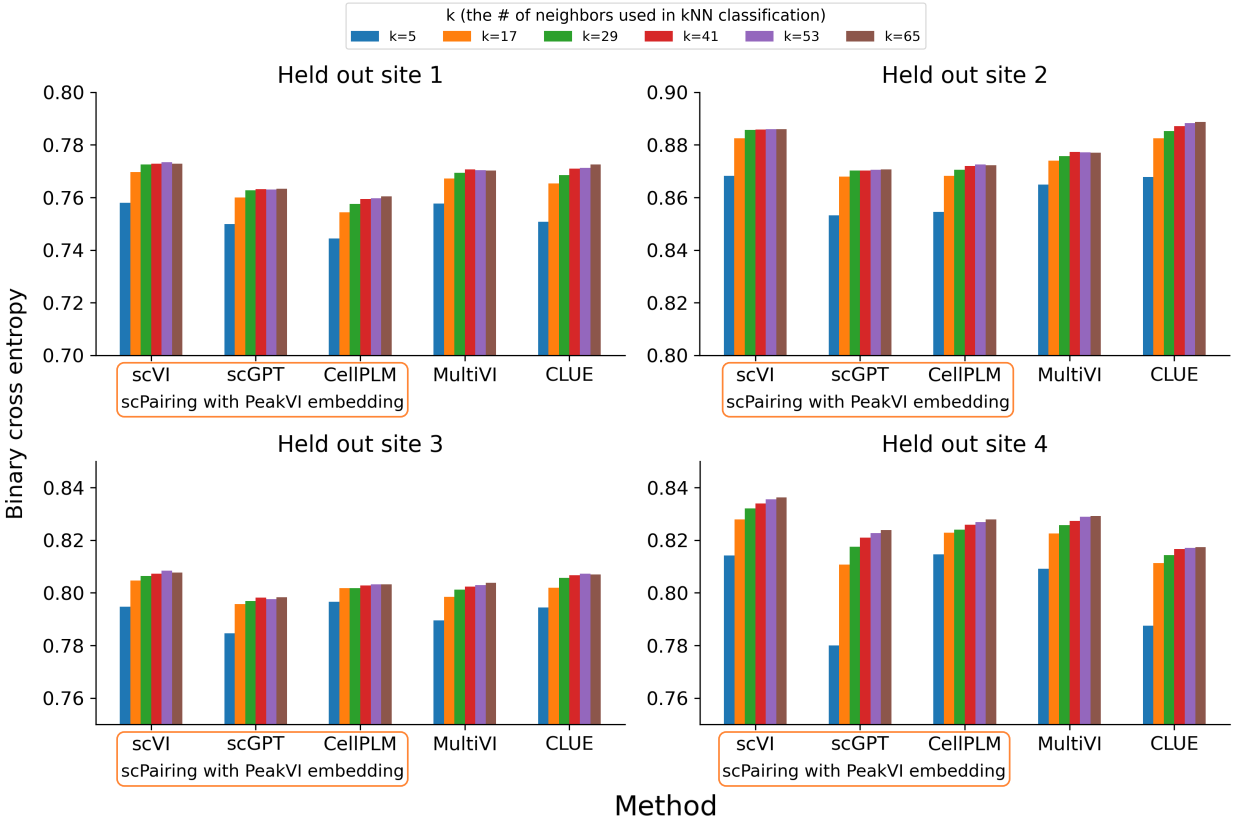

Figure S4: Evaluating data transfer using cross-site cell type classification. Each model was trained using cells from three out of four sites. We projected the held-out site onto the training data for cell type classification with k-nearest neighbors.

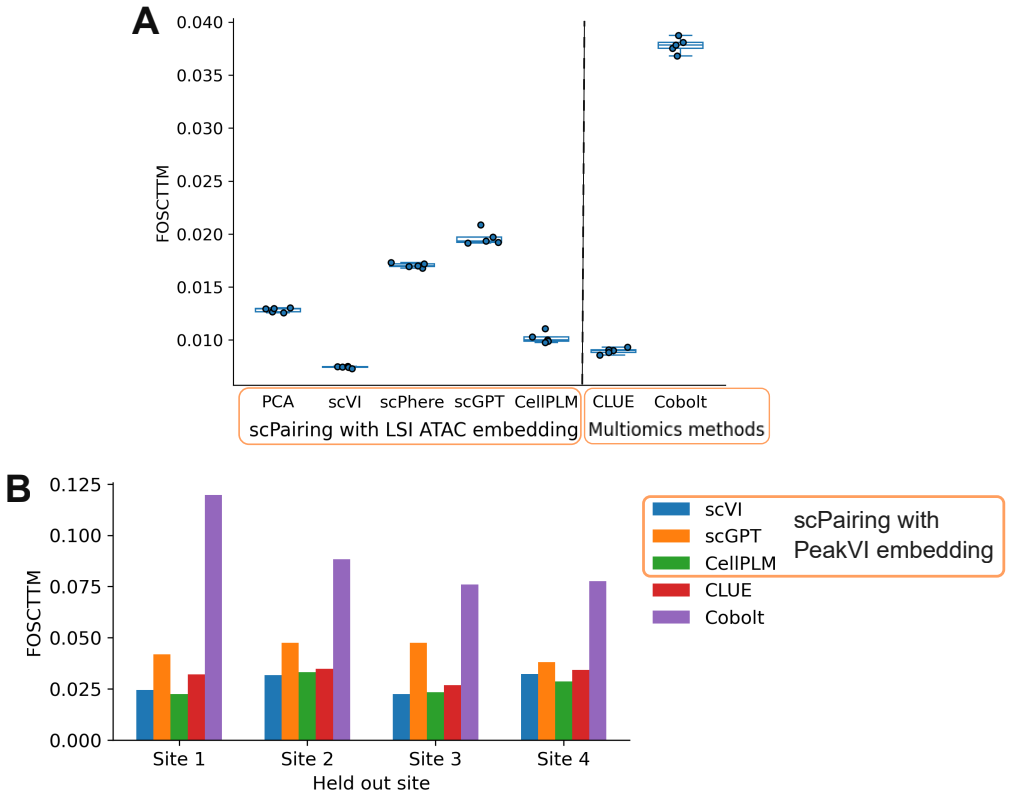

Figure S5: Evaluating modality alignment with FOSCTTM. (A) Same as Fig. 2D, except with a Harmony-corrected LSI ATAC embedding instead of PeakVI. (B) Cross-site validation, where each model was trained using cells from three out of four sites. We projected the held-out site onto the training data and computed the FOSCTTM of the held-out cells.

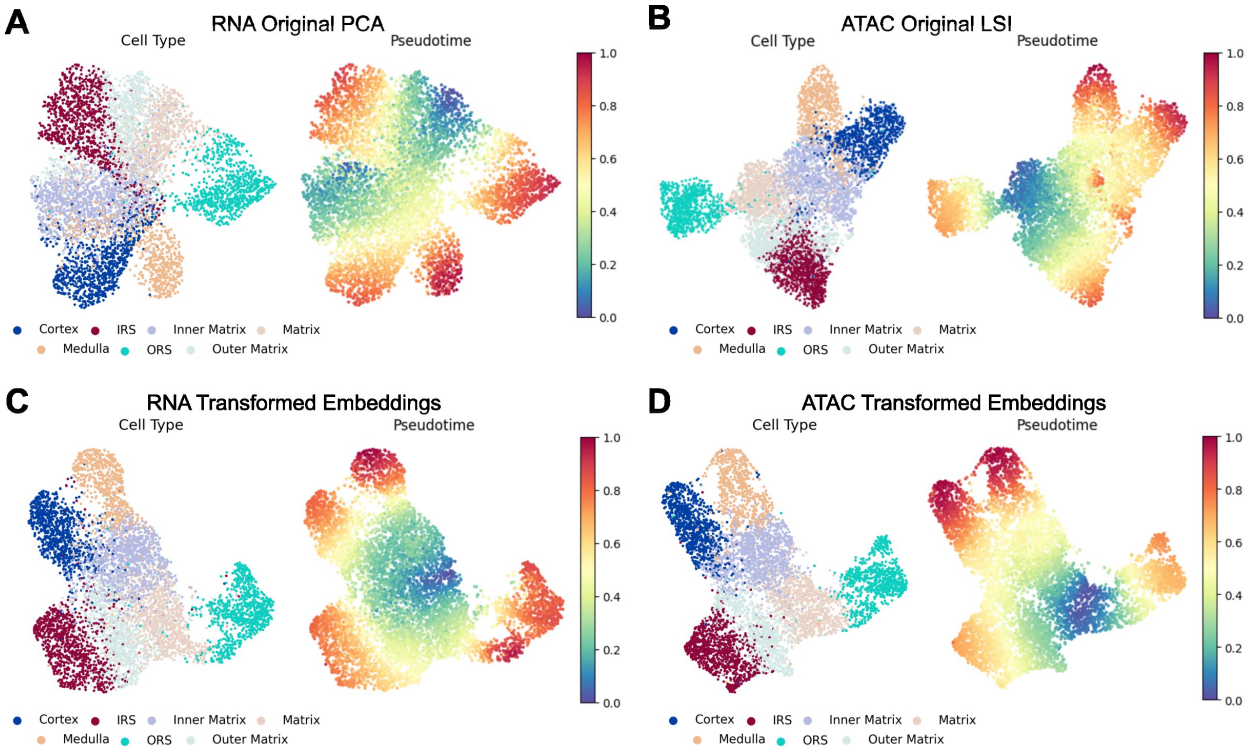

Figure S6: Alignment of PCA embeddings of the scRNA-seq modality in mouse skin cells and LSI embeddings of the scATAC-seq modality. (A) UMAP visualization of cell type and pseudotime using PCA embeddings of the scRNA-seq modality. (B) UMAP visualization of cell type and pseudotime using LSI embeddings of the scATAC-seq modality. (C) UMAP visualization of cell type and pseudotime using scPairing's transformation of the PCA embeddings. (D) UMAP visualization of cell type and pseudotime using scPairing's transformation of the LSI embeddings.

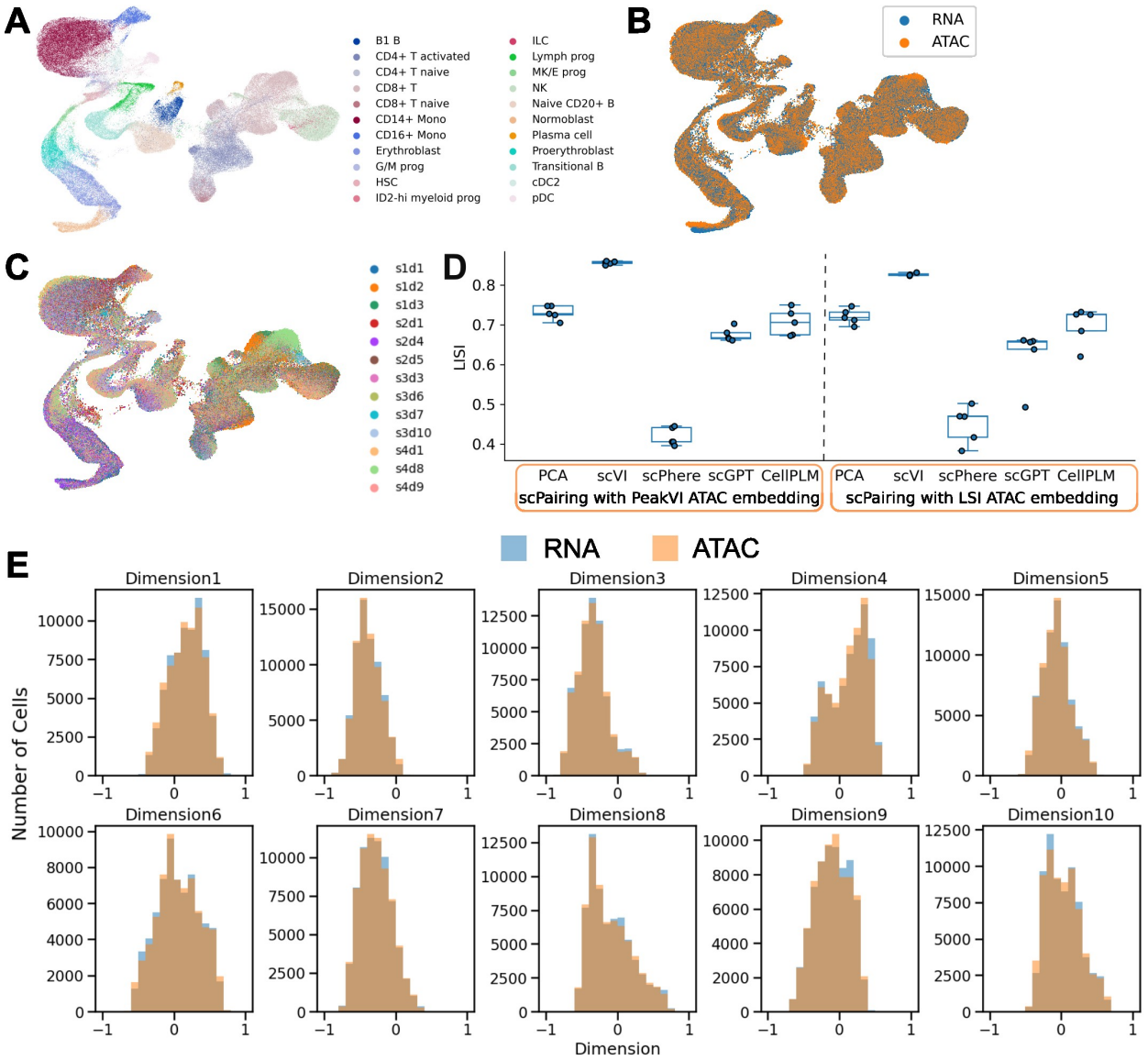

Figure S7: Modality mixing within scPairing representations in the BMMC dataset. (A–C) UMAP visualizations for both scRNA-seq and scATAC-seq modalities combined together, colored by cell type (A), modality (B), and batch (C), when scPairing is run with scVI and PeakVI input embeddings. (D) LISI of scPairing run on different combinations of input embeddings. Each dot represents a trial of each method. The boxplots denote the medians and IQRs, while the whiskers indicate the lowest and highest data point within 1.5 IQR of the lower and upper quartile, respectively. (E), Histograms of modality density along each dimension of the common hyperspherical embedding space, when scPairing is run with scVI and PeakVI input embeddings.

Figure S8: Imputation of gene expression data from chromatin accessibility data using pre-trained decoders. (A) UMAP visualization of the scRNA-seq BMMC benchmarking dataset using scVI embeddings. (B) Visualization of scPairing marker gene expression imputations. We applied scVI and PeakVI pre-trained decoders to impute the counts.

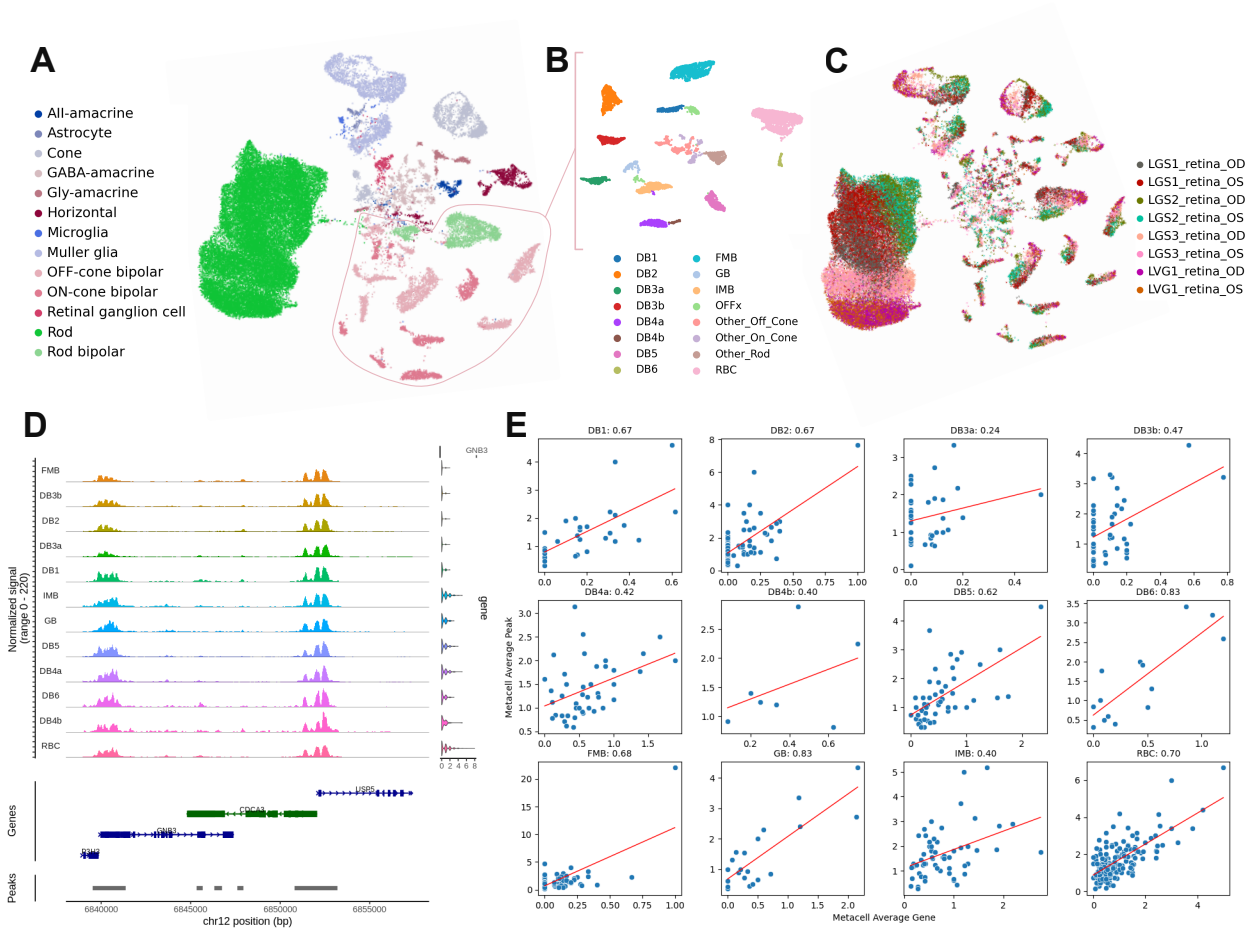

Figure S9: Human retina bridge dataset. (A) UMAP visualization of the bridge dataset, colored by major cell type. (B) UMAP visualization of bipolar retina cells in the bridge dataset, colored by cell subtype. Cell identities not present in the unpaired dataset were annotated as other cell types from the major cell type. (C) UMAP visualization of the bridge dataset, colored by batch. (D) Coverage plot of the retina bipolar cell subtypes at the *GNB3* TSS. (E) Correlation of average *GNB3* gene expression and chr12:6850801-6853197 accessibility across 12 bipolar subtypes aggregated into metacells.

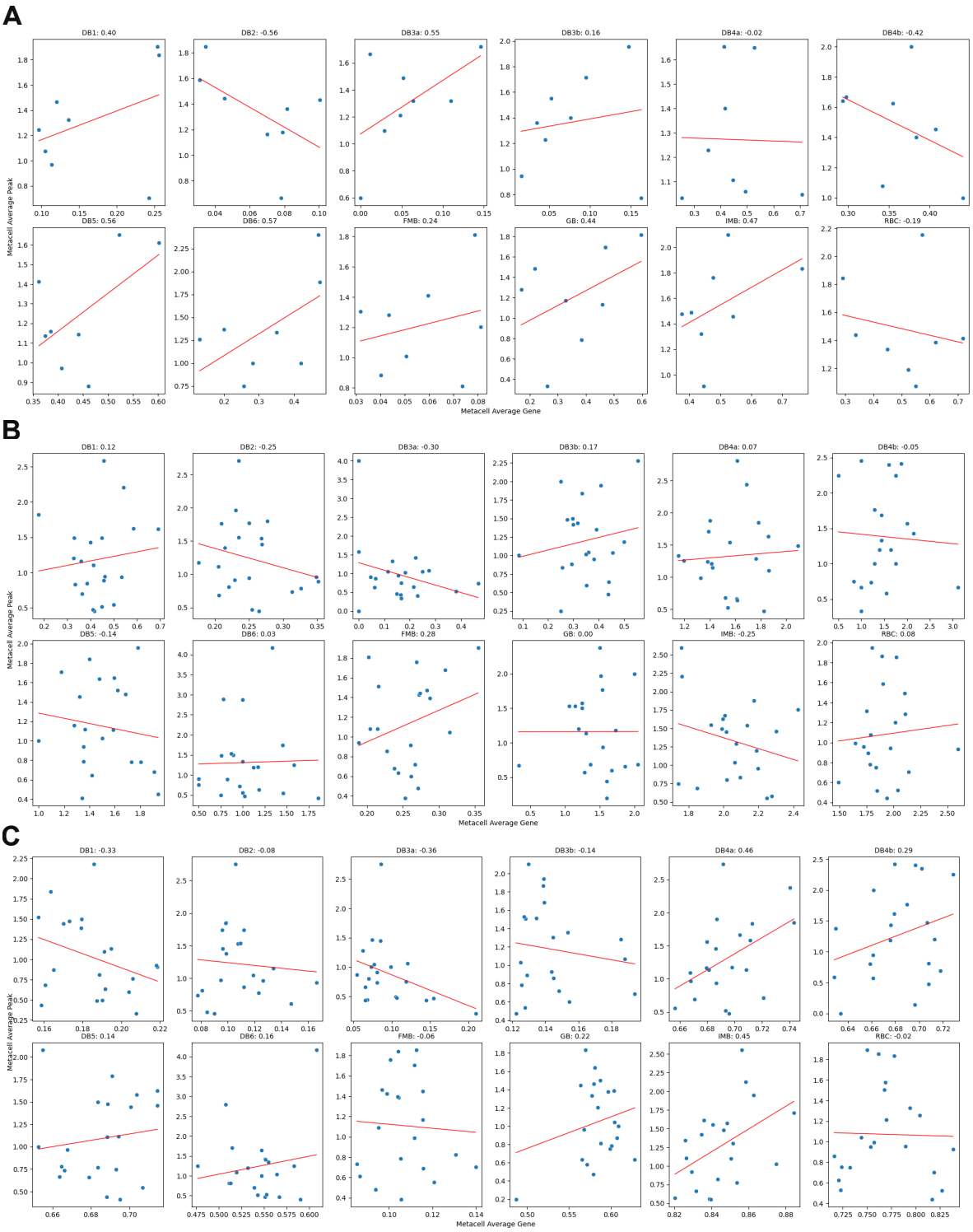

Figure S10: Correlation between *GNB3* expression and chr12:6850801-6853197 accessibility using metacells aggregated by batch. (A) Correlations computed on the Wang *et al.* paired multiomics dataset. The data contained four donors with two retina each, for eight batches in total. (B) Correlations computed on the artificial pairings. The data contained 21 donors. (C) Correlations computed on the gene activities and CCA imputations. This is the same data as in (B), with 21 donors.

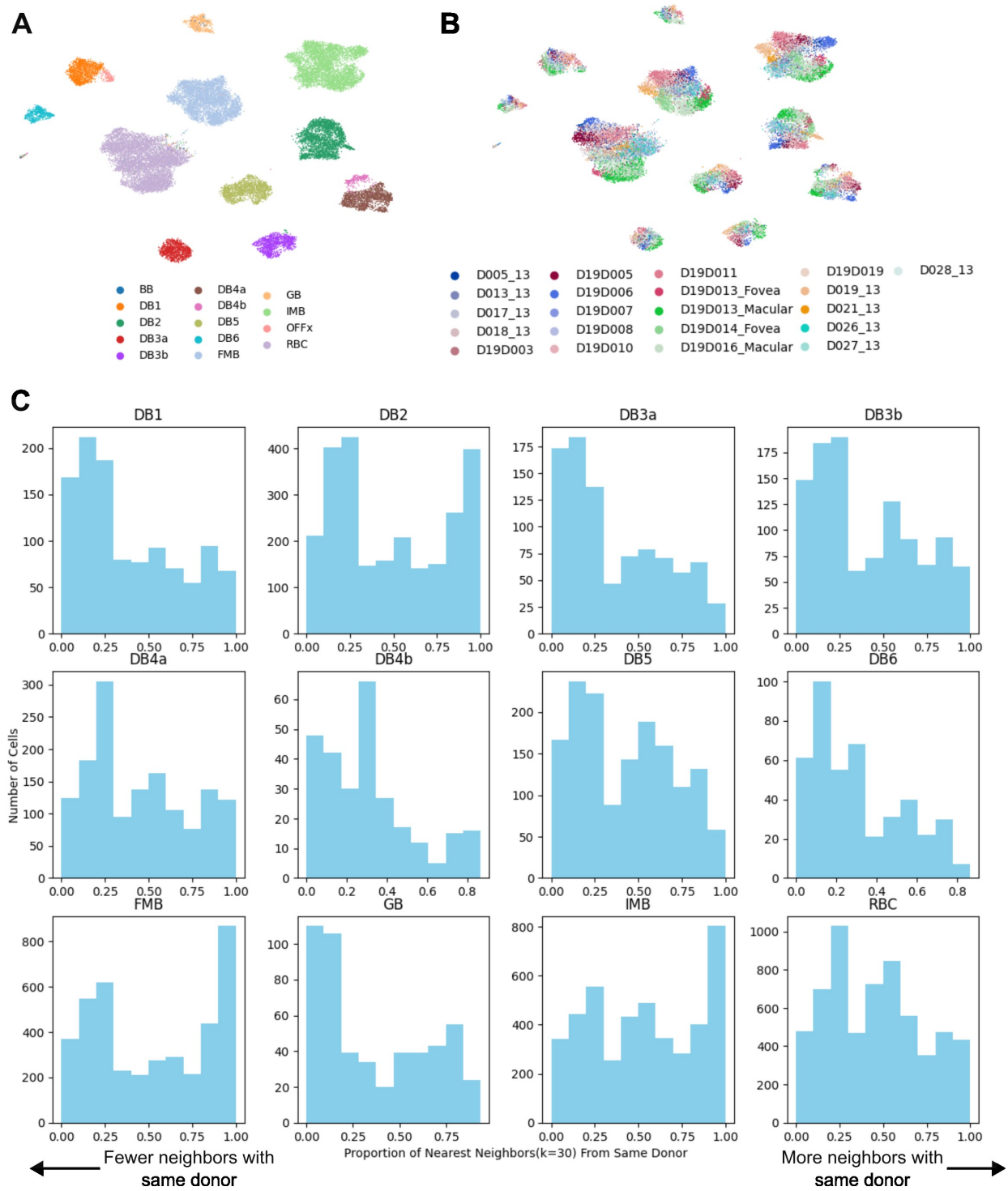

Figure S11: Mixing of donors in scATAC-seq data. (A) UMAP visualization of bipolar retina cells using LSI dimensionality reduction without batch correction, colored by cell subtype. (B) Same UMAP visualization, colored by donor. (C) Proportion of nearest neighbors ($k = 30$) having the same donor for each cell subtype. The 12 subtypes shown are the same as in Figure 3F.

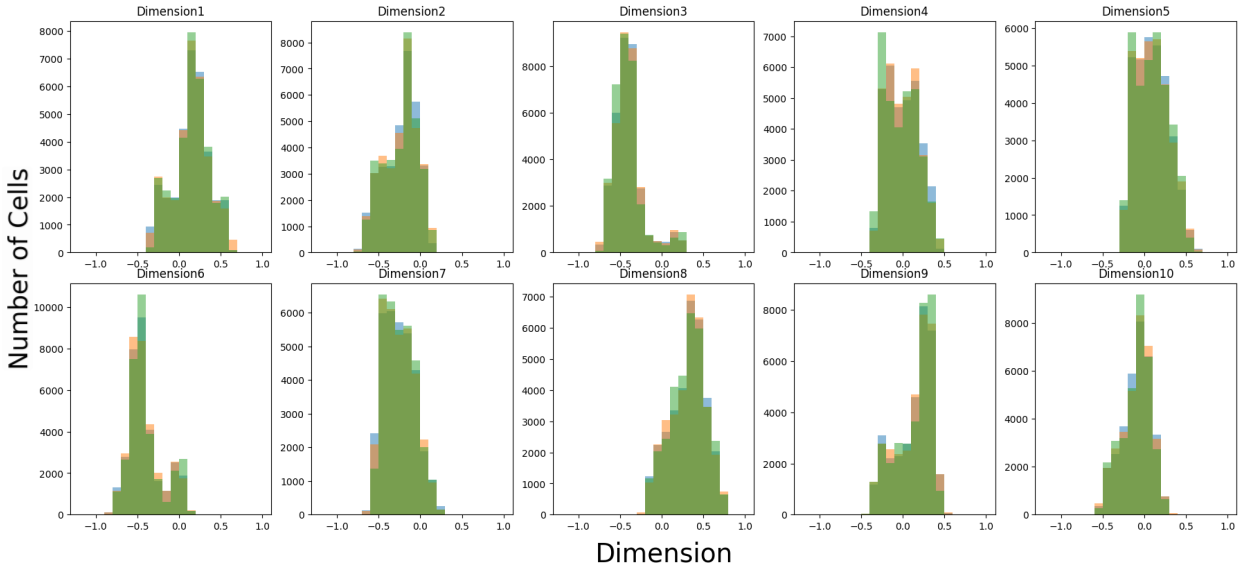

Figure S12: Modality mixing within scPairing representations in the TEA-seq dataset. Histograms of modality density along each dimension of the common hyperspherical latent space.

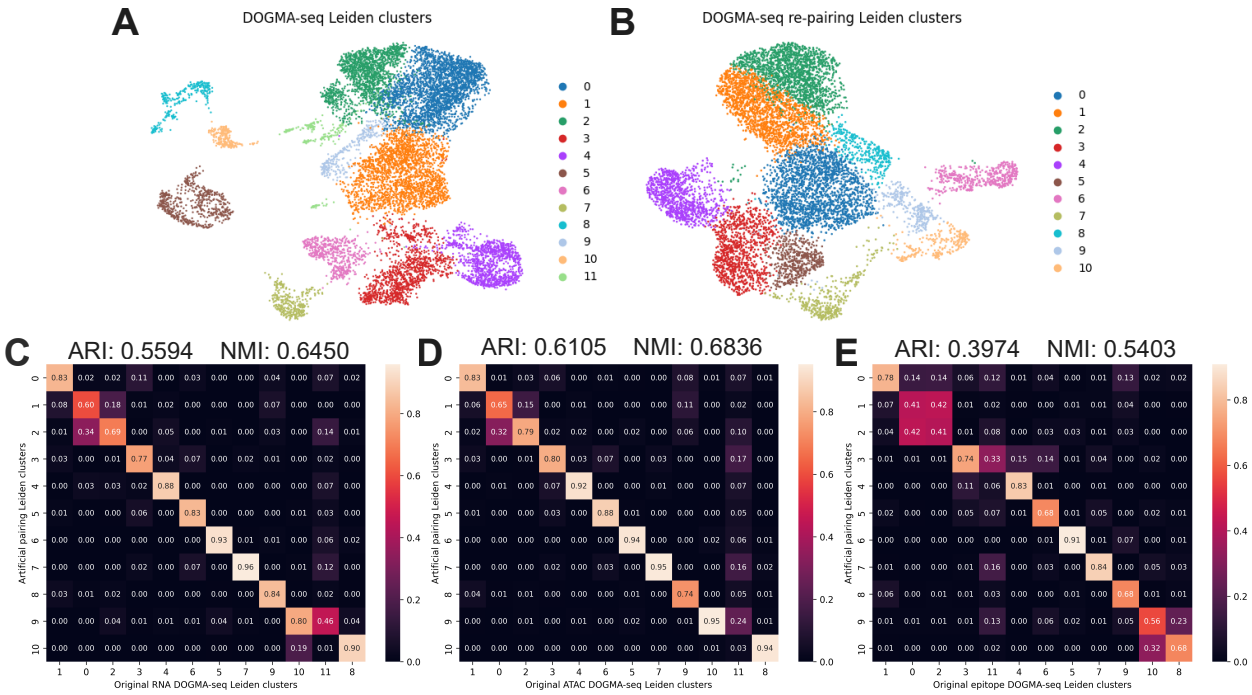

Figure S13: Similarity between original DOGMA-seq clusters and scPairing re-pairings. (A and B) UMAP visualization of the original (A) and re-paired (B) DOGMA-seq data, colored by Leiden clusters. (C–E) Proportion of cells from each original DOGMA-seq Leiden cluster in each re-pairing Leiden cluster. The matrix is normalized by column; each column sums to one. Since the re-pairings constitute three different sets of cells, one for each modality, we computed the similarity for the selected scRNA-seq cells (C), scATAC-seq cells (D), and epitope cells (E). The ARI and NMI between the clusters are given for each modality.

Supplementary notes

S1 Computing gene-peak correlations using SEACells

Persad *et al.* introduced an algorithm for computing metacells, which are aggregations of single cells into robust cells that represent distinct cell states.[45] The authors demonstrate that the correlation between gene expression and chromatin accessibility in metacells are substantially more clear than correlations in single-cells. This clarity enables the discovery of more gene-peak associations.

To combat data sparsity, another approach used by Liang *et al.* is to aggregate the cells by donor.[41] This approach is only applicable when the data contains a substantial number of different donors to have a sufficient number of aggregated cells to compute a meaningful correlation. Using this approach, Liang *et al.* identified the expression of *GNB3* had a heterogeneous range of correlations with the peak chr12:6850801-6853197 across bipolar retina cell subtypes. The original peak in Liang *et al.* is labeled as chr12:6959945-6962406 using the hg19 genome. We used LiftOver[68] to lift the fragments to the hg38 genome and re-call peaks to obtain the corresponding peak chr12:6850801-6853197.

Here, we argue that using metacells also provides clearer correlations than donor-aggregated cells. When aggregating cells by donor into metacells, we make the assumption that the population of cells from each donor represents a different state compared to the other donors. While this assumption may hold for data with high intra-donor heterogeneity, other data have relatively small differences between donors.

To illustrate the lack of heterogeneity between donors in the Liang *et al.* scATAC-seq retina data,[41] we analyzed the mixing of donors after dimensionality reduction using LSI, but without batch correction (**Supplementary Fig. S11A and S11B**). We calculate the proportion of nearest neighbors for each cell that belong to the same donor in the LSI embedding space[69] (**Supplementary Fig. S11C**). We observe that for most subtypes, the bulk of cells have a diverse set of neighbors, with almost all subtypes having the proportion's mode near zero. Only the FMB and IMB subtypes show low donor mixing.

Our results indicate that there is little variability between the donors. Thus, when aggregating cells into metacells by donor, the difference between each metacell is likely noise. The correlations obtained through these metacells show positive, neutral, or negative correlation not through biological variation, but through the randomness between donors. Conversely, since we used SEACells on integrated embeddings of the data, the SEACells-derived metacells are more likely indicative of finer cell states that bring more accurate correlations.

S2 Trimodal pairing algorithm

Algorithm 1 Trimodal pairing

Require: $A \in [0,1]^{n \times m}, B \in [0,1]^{n \times l}, C \in [0,1]^{m \times l}$, A, B, C are pairwise cosine similarity matrices between RNA and ATAC, RNA and epitope, and ATAC and epitope, respectively. There are n RNA profiles, m ATAC profiles, and l epitope profiles.

Require: cutoff $\in [0,1]$

Ensure: A maximal trimodal pairing

```
Without loss of generality, assume that RNA has the fewest number of profiles
Pairings ← {}
All RNA, ATAC, and epitope profiles are in consideration
while ∃ RNA profiles for consideration do
    i ← random RNA profile
    j, k ← ATAC and epitope profiles that maximize average similarity between i, j, k
    if average similarity of i, j, k > cutoff then
        Pairings ← Pairings ∪ {(i, j, k)}
        Remove i, j, k from consideration
    else
        Remove i from consideration
    end if
end while
return Pairings
```

S3 Data preprocessing and pairing

In this supplementary note, we detail the data preprocessing steps taken to convert the raw data into inputs for the model, and the steps taken to perform artificial pairing for each dataset.

NeurIPS 2021 benchmarking dataset. This dataset was produced as part of the Multimodal Single-Cell Data Integration Challenge at NeurIPS 2021.[36] The dataset contains 69,249 cells from 10 donors across four sites, sequenced using the 10X Multiome Gene Expression and Chromatin Accessibility kit. Due to repeat samples from donors at different sites, there are 13 batches present in the data.

NeurIPS 2021 benchmarking dataset scRNA-seq pre-processing. For the scRNA-seq data, we computed five representations using PCA, scVI,[30] scPhere,[31] scGPT,[32] and CellPLM.[33] The PCA embedding was computed with 50 PCs on counts after normalization and `log1p` transformation using Scanpy.[70] Then, we applied Harmony[37] to the PCA embeddings and kept the first 20 dimensions. The scVI representation was computed on all genes using 50 latent dimensions, but otherwise default parameters. The scPhere representation was computed on all genes using 10 dimensions, 100 epochs of training, a minibatch size of 2000, and a learning rate of 0.001. We chose fewer dimensions for scPhere due to declining surface area in high-dimensional hyperspheres. The scGPT representation was computed following their reference mapping tutorial using the continual pre-trained model checkpoint for zero-shot cell embedding. The CellPLM representation was computed following their cell embedding tutorial using checkpoint `20230926_85M`. Both scGPT and CellPLM produce 512-dimension embeddings.

NeurIPS 2021 benchmarking dataset scATAC-seq pre-processing. For scATAC-seq data, we computed two representations using LSI and PeakVI.[34] The LSI embedding was computed on peaks expressed in at least 1% of cells and TF-IDF transformed counts. We used the Muon[71] LSI implementation with 50 dimensions. The embeddings were then corrected using Harmony[37] and then kept the first 20 dimensions. The PeakVI embedding was also computed using peaks expressed in at least 1% of cells with 50 dimensions, but otherwise default parameters.

Mouse skin dataset. This dataset contains 28,429 cells sequenced by SHARE-seq from Ma *et al.*[2] Specifically, we used a pre-processed and re-annotated version of the data produced by Lynch *et al.*[19] We computed the scPairing input representations using PCA and LSI with default parameters.

Human retina datasets. We used two human retina datasets. The first dataset is a set of over 50,000 joint scRNA-seq and scATAC-seq human retina cells sequenced using 10X Multiome by Wang *et al.*[42] The samples included both left and right retinas from four different donors. The second dataset is a set of over 250,000 cells sequenced using snRNA-seq and another 137,000 cells sequenced using snATAC-seq by Liang *et al.*[41] The scRNA-seq samples came from six donors, of which we only used the four donors with the majority of cells (211,757 cells). The scATAC-seq samples came from 21 donors, all of which

we used.

We filtered the scRNA-seq data by keeping cells that had between 300 and 25,000 counts and at least 10 unique genes expressed. The scATAC-seq data was filtered by retaining cells that had total counts between 3,000 and 30,000 and a nucleosome signal below four. Peaks that were expressed in less than 0.5% of cells were also removed to reduce the data size.

We prepared the two datasets by integrating the individual modalities in the paired and unpaired datasets using scVI[30] and PeakVI,[34] using the batch and dataset as categorical covariates. The peaks from the two datasets were merged using Signac[44] by calling peaks with MACS2[72] on each dataset individually and combining the two peak sets using `reduce`. The two scRNA-seq modalities were merged by taking the common genes sequenced. During pairing, we used similarity cutoff $\varepsilon = 0.02$.

R-DC-like datasets We used three datasets containing R-DC-like cells. The first dataset is a set of over 10,000 paired human tonsil cells that we use as the bridge.[46] The dataset authors applied a sorting strategy to enrich for DC1, DC2, $CD34^+$, and $Lin^-HLA\text{-}DR^+CD4^-CD1c^-$ cells. The second dataset consists of both paired and unpaired scRNA-seq and scATAC-seq human tonsil cells. In both the paired and unpaired data, we restricted our attention to DCs and ILC3s as other cell types are unlikely to contain R-DC-like cells. After applying this filter, we obtained 501 paired cells, 1,536 scATAC-seq cells, and 7,098 scRNA-seq cells. The third dataset is a set of paired and unpaired scRNA-seq and scATAC-seq cells from the human intestine. Like the second dataset, we also restricted our analysis to cell types that may be R-DC-like. Only the cell type annotations in the paired data indicated the potential presence of R-DC-like cells. Thus, we only used the paired data, consisting of 280 cells.

We filtered the first dataset for cells that had between 300 and 20,000 total counts and filtered for genes expressed in at least three cells, leaving 10,523 cells. The other two datasets were not filtered. To merge the scRNA-seq data, we took the common genes. For the scATAC-seq data, we merged the peaks by calling peaks separately and reducing the peaks together to create a common peak set. For both the unpaired and paired data from the second dataset, we computed low-dimensional representations for scPairing input using scVI for the scRNA-seq data and Harmony-corrected LSI for the scATAC-seq data. During pairing, we used similarity cutoff $\varepsilon = 0.5$.

Trimodal data. We used two trimodal datasets, TEA-seq[25] and DOGMA-seq,[24] with both sequencing peripheral blood mononuclear cells (PBMCs). The TEA-seq data contained 37,124 cells sequenced in five batches. The DOGMA-seq data consisted of four batches, of which we only used one batch, the control digitonin-permeabilized cells, containing 11,868 cells.

We first processed the scATAC-seq fragments from the TEA-seq dataset[25] using Signac,[44] calling peaks on each fragment file separately, then merging the peaks. The scRNA-seq and epitope modalities were merged by keeping the common genes and proteins. We kept cells that had between 500 and 2,750 unique genes, at least 1000 unique peaks, and were not part of a cluster that expressed the mouse control protein. Peaks or genes expressed in fewer than

15 cells were removed.

We compute the low-dimensional representations for the TEA-seq scRNA-seq data by Harmony-corrected PCA and the scATAC-seq data by Harmony-corrected LSI.[37] For the epitope data, we did not reduce the dimension as only 46 proteins were measured. Instead, we used centered log ratio (CLR) transformed[4,71] counts as input.

We performed peak calling on the DOGMA-seq scATAC-seq fragments and merged the peaks with the peaks from the processed TEA-seq data. The scRNA-seq and epitope data were merged by taking common genes or proteins. We kept cells with at least 500 peak counts, between 200 and 30,000 gene counts, between 20 and 10,000 epitope counts, and less than 50 rat control epitopes. In our experiment to re-pair DOGMA-seq data using the TEA-seq data as a bridge, we integrated each modality separately using scVI[30] for scRNA-seq data, PeakVI[34] for scATAC-seq data, and Harmony[37] on scaled CLR counts for epitope data. During pairing, we used similarity cutoff $\varepsilon = 0.1$.

The joint scRNA-seq and scATAC-seq dataset used to generate a new trimodal dataset is the NeurIPS 2021 benchmarking data. The epitope data comes from the same NeurIPS 2021 benchmarking data,[36] except from the CITE-seq data. We discarded the scRNA-seq data from the CITE-seq data to retain only the epitope sequencing. This data contained 90,261 cells, with 134 proteins sequenced. We performed no additional pre-processing on the data as it was already processed by the authors.

For low-dimensional representations, we used scVI and PeakVI for the scRNA-seq and scATAC-seq data, respectively. For the epitopes, we used Harmony-corrected scaled CLR transformed counts. During pairing, we used similarity cutoff $\varepsilon = 0.1$.

Supplementary tables

R-DC-like scRNA-seq cell barcode	Artificially paired scATAC-seq cell barcode	Successful pairing	True match
TGCCGGTAGGCATTAC-1	TGCCGGTAGGCATTAC-1	✓	✓
TAGTGTGGTACGGTAC-1	AACCCGCAGCAACATC-1	✓	✗
AAGCCTTAGTGAGAGC-1	GTTACCTGTTAGCCAA-1	✗	✗
CTTACTAGTGCTCCGT-1	GTATTGCAGCTTCTCA-1	✓	✗
GTACCGGGTCCTTCAG-1	TCCATCATCCCGAAGC-1	✓	✗
TCCTTTACAGCAACAG-1	TCCTTTACAGCAACAG-1	✓	✓
ACTTTGTTCGCTAGAT-1	ACTTTGTTCGCTAGAT-1	✓	✓
TCCATCATCCCGAAGC-1	AAGCCTTAGTGAGAGC-1	✓	✗
GATTGCAGTCCAAGAC-1	GATTGCAGTCCAAGAC-1	✓	✓
GCCAATAGTGATGAAA-1	ACGGGAAGTATTGGTG-1	✓	✗
GTATTGCAGCTTCTCA-1	GTACCGGGTCCTTCAG-1	✓	✗
TCTTTGTAGTTCCCAC-1	TCACTGACATAAGCAA-1	✗	✗
AACCCGCAGCAACATC-1	TAGTGTGGTACGGTAC-1	✓	✗
ATGCAGGCATTGTTGG-1	ATGCAGGCATTGTTGG-1	✓	✓
CGCTTAACACCAAAGG-1	CGCTTAACACCAAAGG-1	✓	✓
TTCCTTGAGTAATCCA-1	GTCTTTAGTTGCTGGG-1	✗	✗
CCAGGATGTTTATCTG-1	CCAGGATGTTTATCTG-1	✓	✓
ACCAGGACATAATCAC-1	ACCAGGACATAATCAC-1	✓	✓
ATATAGGCATTAGCCA-1	CGTGACATCAGGGCCT-1	✓	✗
CCGGTAGGTAGCAGCT-1	CCGGTAGGTAGCAGCT-1	✓	✓
CTACAACAGGAAACTG-1	TCTTTGTAGTTCCCAC-1	✓	✗
ACGGGAAGTATTGGTG-1	ATATAGGCATTAGCCA-1	✓	✗
CACTTAAAGGAACACA-1	CACTTAAAGGAACACA-1	✓	✓
CGTGACATCAGGGCCT-1	CTACAACAGGAAACTG-1	✓	✗
GGTCCTGCAACCTAAT-1	GGTCCTGCAACCTAAT-1	✓	✓
GATGCGACAGGATAAC-1	CTTACTAGTGCTCCGT-1	✓	✗
GCTGAGGAGATAAAGC-1	GCTGAGGAGATAAAGC-1	✓	✓

Table S1: Re-pairings of human tonsil R-DC-like cells. The cells in the first column were identified as likely R-DC-like cells from their gene expression profiles. The second column indicates the scATAC-seq profile that was paired with each R-DC-like gene expression profile. The third column indicates whether the scRNA-seq and scATAC-seq profiles both belonged to R-DC-like cells. The fourth column indicates whether the scRNA-seq and scATAC-seq profiles came from the same cell, meaning scPairing recapitulated the true pairing of the two modalities. The cell barcodes listed have had their prefixes stripped off, but no duplicate barcodes are present in the R-DC-like cells.

R-DC-like scRNA-seq cell barcode	Artificially paired scATAC-seq cell barcode	Successful pairing	True match
AGCCTGGGTCCGTAGG-1	CATCCTCAGTGCTGTG-1	✓	✗
ACGAATCTCCGCAAGC-1	AGCCTGGGTCCGTAGG-1	✓	✗
AGTGAGTCAGCTTAAT-1	AGTGAGTCAGCTTAAT-1	✓	✓
GATGCAGCAGTCTATG-1	TGTAACTCAGAGGGAG-1	✓	✗
GTGCACGGTTAAGCCA-1	TACTAAGTCGCTTGCT-1	✗	✗
GGTGCTGGTCAAGTAT-1	TGACTTAAGTATTGCA-1	✓	✗
TGTAAGTGTTTAGTCC-1	GATGCAGCAGTCTATG-1	✓	✗
GATGAGGGTAAGTCGC-1	GGTGCTGGTCAAGTAT-1	✓	✗
TGACTTAAGTATTGCA-1	TGAGTTTCACCATATG-1	✗	✗
GTCTAATCATGTTGCA-1	GTCTAATCATGTTGCA-1	✓	✓
TGTAACTCAGAGGGAG-1	TAAGCCAGTGCGCAAT-1	✗	✗
CATCCTCAGTGCTGTG-1	ACGAATCTCCGCAAGC-1	✓	✗
TAGCGGACAGACAAAC-1	TAGCGGACAGACAAAC-1	✓	✓

Table S2: Re-pairings of intestine R-DC-like cells. The cells in the first column were identified as likely R-DC-like cells from their gene expression profile. The second column indicates the scATAC-seq profile that was paired with each R-DC-like gene expression profile. The third column indicates whether the scRNA-seq and scATAC-seq profiles both belonged to R-DC-like cells. The fourth column indicates whether the scRNA-seq and scATAC-seq profiles came from the same cell, meaning scPairing recapitulated the true pairing of the two modalities. The cell barcodes listed have had their prefixes stripped off, but no duplicate barcodes are present in the R-DC-like cells.

	RNA-ATAC	RNA-epitope	ATAC-epitope
Mean Cosine Similarity	0.9579	0.9060	0.9253
Median Cosine Similarity	0.9640	0.9183	0.9366

Table S3: Mean and median pairwise modality cosine similarities between the true matching triplets from DOGMA-seq. The cosine similarities are calculated using the embeddings produced by applying scPairing with TEA-seq as the bridge data. The similarity is computed between each pair of modalities.

	RNA-ATAC	RNA-epitope	ATAC-epitope
FOSCTTM	0.05378	0.1620	0.1294

Table S4: Pairwise FOSCTTM between the true matching triplets from DOGMA-seq. The FOSCTTM scores are calculated using the embeddings produced by applying scPairing with TEA-seq as the bridge data. The FOSCTTM score is computed between each pair of modalities.