Supplementary Note: Benchmarking Online iNMF Performance Across a Range of Conditions Using Real and Simulated Data

Benchmarking Online iNMF with Simulation Studies

We have demonstrated in the former sections the robust performance of online iNMF on multiple real datasets including human PBMC, human pancreas, and mouse cortex. To provide additional theoretical understanding, we performed extensive simulations using the Splatter scRNA-seq simulator. We investigated the effects of different dataset orderings, relative dataset sizes, and cell type compositions. To give a clearer view, these results are separated into five figures and organized in a similar fashion, including a schematic plot of the simulation design (**a, Fig. S9, S10, S11, S12, S13**), UMAP plots of a representative integration result (**b-d, Fig. S9, S10, S11, S12, S13**), and box or line plots of evaluation metrics (**e, Fig. S9, S10, S11, S12, S13**).

Overall, we found that imbalanced cell cluster proportions and dataset sizes have very little effect on the results of online iNMF, although scenario 2 is slightly more sensitive to imbalances in cell proportions than scenario 1. We computed the Spearman correlations between the Shannon entropy of batch and cluster sizes and the four evaluation metrics we employed (see **Online Methods** section for details). Most of the correlation p-values are much larger than 0.05, indicating that online iNMF performance is not significantly affected by imbalances in dataset size or cluster proportions. The one exception is that for online iNMF (scenario 2), there is a statistically significant correlation between cluster entropy and both adjusted rand index and alignment, indicating that scenario 2 is slightly more sensitive to imbalances in cell proportions than scenario 1. Furthermore, we used random dataset arrival orders in benchmarking scenario 2, and found that the relative order of small vs. large batches in scenario 2 makes little difference (**Fig. S9**).

With missing cell clusters, the proportion of the missing cell types also has little effect on the results of iNMF (scenarios 1 and 2) (**Fig. S11**). Under the same condition, we also performed simulations to test whether scenario 3 will force cells into the existing feature space. These results showed that online iNMF does not cause spurious alignment, even if one or more cell types in the dataset to be projected are missing from the reference dataset. In the case of multiple cell types missing from the reference dataset, all of the new cell types cluster with each other (but not with the reference cells). This causes a decrease in overall Purity and ARI (**Fig. S12e**, red line), but a much more gradual decrease in purity and ARI for the cell types shared between reference and query datasets (**Fig. S12e**, green line). This behavior makes sense, because the shared metagenes (W) learned from the reference dataset cannot be expected to distinguish among multiple unseen cell types, which explains the poor evaluation metrics for the cells in the missing cell types (**Fig. S12e**, blue line).

We also designed simulations in which no cell types occurred in every batch, but every pair of batches shared at least one cell type. This allowed us to test the performance of iNMF on data with complex biological compositions. Our results indicate that online iNMF is quite robust to

these situations and identifies every cluster clearly (**Fig. S10**). It seems that the most important factor determining the difficulty of identifying a particular cluster is the total number of cells in the cluster observed across all datasets, independent of how those cells are distributed across datasets.

Reading Mini-Batches from Disk Adds Minimal Overhead

One highlight of the proposed online iNMF algorithm is that it streams the mini-batches from the files on the disk without loading the entire data into the memory. Here we demonstrate that little overhead is added through this approach on the mouse frontal and posterior cortex scRNA-seq datasets (details are discussed in **Online Methods**). For a mini-batch size of 5,000 cells, reading each mini-batch from disk does not require significant overhead (an average of less than 0.56 seconds per iteration over 50 iterations) (**Fig. S4**).

Online iNMF Is Robust to Initialization and Input Ordering

The online iNMF algorithm starts with randomly initialized metagene factors (W and V^i). Therefore, we inspected the effect of random initialization on the analyses on the MOp datasets by assessing the agreement, as measured by ARI, between the resulting cell clusters and our annotations generated in scenario 2 (**Fig. 6d**). First, we performed online iNMF (scenario 1) on all eight MOP datasets with 10 different random initializations, using the same variable genes that we used for the scenario 2 analysis in **Fig. 6**. Based on the output cell clusters, the average ARI (vs. our annotations from scenario 2, shown in **Fig. 6d**) is 0.725. Similarly, we ran online iNMF (scenario 2) with 10 different initializations using the same set of genes (inputs ordered chronologically) and obtained an average ARI (vs annotations) of 0.744. These results indicate that online iNMF scenario 1 and 2 are both robust to the effects of random initialization.

To investigate the effects of different dataset orders on scenario 2 results, we repeated the scenario 2 analysis using each of the other five sc/snRNA-seq datasets as the initial dataset. As with the analysis shown in **Fig. 6**, we selected over 4,000 variable genes from the first dataset and sequentially incorporated all remaining MOp datasets. We found that initiating the analysis with any of the six sc/snRNA-seq datasets leads to clusters in good agreement with our annotation (average ARI = 0.759), indicating that the results are robust to choice of starting dataset. We can even select genes from the snATAC-seq dataset, and use it as the first input, with slightly lower agreement (ARI = 0.627). If we instead use the snATAC-seq dataset as the starting dataset but use the genes selected from the first RNA dataset (SMARTer cells), the ARI is 0.752. Because the distribution of methylation is so different from gene expression, the statistical model for variable gene selection reported zero variable genes, and thus we were not able to select genes from the methylation data. Additionally, the results from scenario 1 and 2 are quite congruent (ARI = 0.773).

Integration with RNA Data Detects More Clusters from Epigenome Data

In the original LIGER paper, we showed that integrating single-cell methylation data with scRNA-seq data resolved more methylation clusters than using methylation data alone. Here we confirmed that this still holds true for online iNMF on the mouse primary motor cortex (MOp) datasets: integrating methylation (snmC-seq) or chromatin accessibility data (snATAC-seq) with

RNA data (snRNA-seq) better separates clusters compared to the epigenome data alone (**Fig. S8**). In the first experiment (**Fig. S8a**), we started by factorizing the snATAC-seq data (n = 54,844) and obtained 9 clusters. After incorporating the snRNA-seq data (n = 101,647), the two datasets are well aligned. More importantly, we are able to observe 15 clusters, which implies the structure within the data is refined. Similarly, in the second experiment (**Fig. S8b**), the "resolution" of snmC-seq data (n = 9,366) is also increased after being jointly analyzed with the same snRNA-seq data, where 3 additional clusters are detected.

Online iNMF Identifies Rare Cell Types Present in Only a Subset of the Datasets

We also looked into the detection of rare cell types in MOp data, L5/6_NP and L6b, in separate analyses (**Fig. S5**). In the first experiment, we held out L5/6_NP and L6b cells from the first input (allen_10x_cells_v2, n = 117,382) in scenario 2. Next, we incorporated an snRNA-seq dataset (macosko_10x_nuclei_v3, n = 101,647) that includes L5/6_NP cells (3.3% of all cells). After Louvain clustering on the learned latent space, 96.4% of the L5/6_NP cells in the snRNA-seq dataset grouped together and formed a distinct cluster (highlighted with a red box). In the second experiment, we held out the L6b cells from the scRNA-seq dataset (n = 119,183) and subsequently incorporated the snRNA-seq dataset (n = 101,647), in which L6b cells make up 1.5% of all cells. L6b is rarer than L5/6_NP, which makes this task more challenging. Additionally, the L6b cluster is more continuous with the L6 CT cells, and the cluster boundary is somewhat unstable across different clustering runs. Nevertheless, 91.8% of the L6b cells formed a distinct cluster. Thus, these results indicate that online iNMF in scenario 2 can still detect rare cell types in late arriving datasets. We observed very similar results if the rare cell type was missing from the first dataset (99.3% of L5/6_NP cells formed a distinct cluster, and 94.3% of L6b cells formed a distinct cluster). Consistent with our simulation results, these analyses suggest that the order of dataset arrival is not strongly influential in whether rare cell types are detected.

Online iNMF Robustly Integrates Datasets with Non-Overlapping or Partially-Overlapping Cell Types

First, we examined the performance of online iNMF in integrating datasets of the same modality that do not share any common cell types. For this evaluation in scenario 2, we selected two datasets generated from MOp and only retained cells of dissimilar classes. The first input in scenario 2 (scRNA-seq, 10x v2) consists only of interneurons (n = 27,555), including medial ganglionic eminence (MGE)-derived cells and caudal ganglionic eminence (CGE)-derived cells. In contrast, the second input (snRNA-seq, 10x v3) only contains oligodendrocytes (n = 21,404). We also performed this analysis using online iNMF (scenario 1) and batch iNMF for comparison. The results are visualized in 2-dimensional UMAP coordinates (**Fig. S6a**). As expected, there is very little spurious alignment between the two cell classes when implementing online learning in scenario 2. The corresponding alignment scores for online iNMF (scenario 2), online iNMF (scenario 1) and batch iNMF are 0.106, 0.034 and 0.027 respectively, while the kBET acceptance rates are 0.050, 0.014 and 0.002. Thus, all three approaches are quite comparable in their ability to avoid spurious alignment of the non-overlapping cell types. Moreover, 30 metagenes effectively capture the structure within the interneurons. We were also interested in how online iNMF would perform in scenario 3 in a similar setting. In this experiment, we started

by creating a curated atlas of interneurons using scRNA-seq dataset (n = 27,555) and a snRNA-seq dataset (n = 15,255) through scenario 1. Then we projected a snATAC-seq dataset, which only consists of oligodendrocytes, into this atlas (n = 8,557). As **Fig. S6b** shows, the oligodendrocytes are clearly separated from interneurons, while the structure of interneurons are retained. This indicates that scenario 3 can still detect outliers even if the query sample has extra cell types, even across modalities.

Next, we investigated cases where the cell types in the input datasets partially overlap. As was discussed in the previous section, online iNMF (scenario 2) performs well at identifying the rare cell types in partially-overlapping datasets. We anticipate that scenario 3 is most useful for projecting small and specialized samples onto a large and comprehensive atlas, so we investigated performance when the reference dataset contains more cell types than the query (**Fig. S7**). We first integrated 6 sc/snRNA-seq datasets from the MOp (n = 344,675) using online iNMF in scenario 1. Afterwards, we held out the MGEs (i.e. Pvalb, Sst and Chodl cells), which are approximately 10.4% of all the cells, from the snATAC-seq dataset. Then we projected this processed ATAC dataset (n = 49,167) into the established atlas. The UMAP visualization annotated by our cell class labels is exhibited for reference (**Fig. S7a**) and it shows that different cell types are effectively identified. By coloring the cells by their data sources (**Fig. S7b**), it can be observed that very few cells from snATAC-seq data are spuriously aligned to the clusters corresponding to Pvalb, Sst and Chodl cells (highlighted in the red boxes).

Online iNMF Achieves Accurate Data Reconstruction

Here we demonstrate online iNMF's capability of data reconstruction by providing a supplementary figure showing only the reconstruction portion of the objective (**Fig. S14**). The resulting iNMF factors do indeed reconstruct the data comparably to PCA, batch iNMF, and regular NMF. In this experiment, we evaluated the performance of online iNMF on reconstructing the human PBMC dataset (n = 13,999) along with batch iNMF, regular NMF and PCA, using 2,001 variable genes. Next, we implemented the listed methods on the scaled data using the same setting (K = 30 for all methods and λ = 5 for iNMF-based methods). The metric for comparison is the mean squared error (MSE) between the scaled and the reconstructed gene expression matrices. We repeated the experiment 10 times for iNMF/NMF-based approaches to account for the effect of random initialization and reported the average MSE. As is displayed in the plot, the performances of online iNMF (mean MSE = 0.831), batch iNMF (mean MSE = 0.830) and batch NMF (mean MSE = 0.830) are quite similar, while PCA (mean MSE = 0.825) accomplishes this task slightly better.

Selection of Key Parameters (K and λ)

Selecting the dimensionality of the latent space is a perennial challenge in unsupervised data analysis. Due to the lack of ground truth, there is no way to pick the single best value for this parameter. In our previous paper, we described a heuristic for guiding the selection of K, by identifying an "elbow" in the plot of K vs. factor entropy. This is analogous to picking the number of eigenvectors for principal component analysis by inspecting a plot of the eigenvalue spectrum. In general, cell populations with a larger number of distinct cell types/states benefit from a larger K; for example, a sample of frontal cortex contains many more distinct subsets of

cells than a sample of peripheral blood mononuclear cells. In practice, any K value between 20 and 40 usually gives reasonable results. Here we added analyses to demonstrate that online iNMF performs well across different choices of K on simulated data (**Fig. S13**). The results show that, with any K in the range of 10 to 25, online iNMF (scenario 1 or 2) successfully aligns the datasets and recovers the 8 true cell clusters.

We also examined the effect of regularization parameter λ on data alignment (**Fig. S15**). To do so, we jointly analyzed the human PBMC datasets while varying λ and fixing K at 20. Similar to the original LIGER paper, an "elbow" shape was observed, which implies that the alignment quality remains robust for any $\lambda \geq 1$.

ANLS Outperforms HALS for Updating Cell Factor Loadings

We discovered an implementation detail that is crucial for achieving optimal online iNMF performance: using ANLS to calculate H (cell factor loading) updates. Although in principle either HALS (hierarchical alternating least squares) or ANLS can be used to update $H_M^{i(t)}$ for each minibatch, we found empirically that the ANLS updates converge much faster than HALS updates (**Fig. S7**). This may be because ANLS gives an optimal solution for all of the cell factor loadings (rows of H) simultaneously, while HALS updates are only optimal for one individual column of H at a time, requiring multiple iterations. We note also that Mairal et al. opted for least angle regression (LARS), which is directly analogous to our ANLS update, rather than a HALS-like update for H in their implementation.

Supplementary Note: Figures

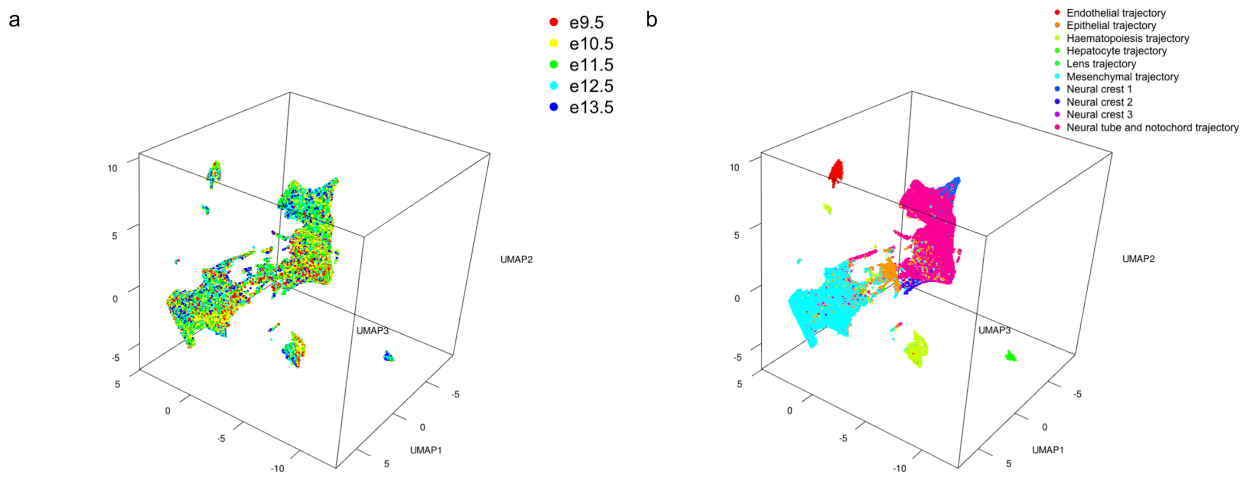
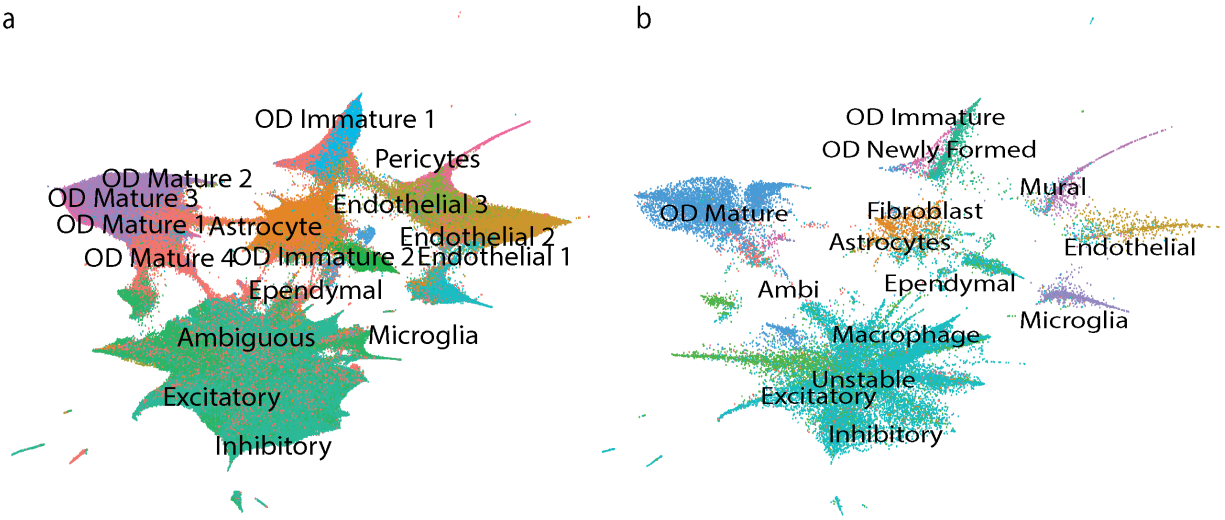


Figure S1. Online iNMF (scenario 1) efficiently factorizes the mouse organogenesis cell atlas (MOCA). The MOCA dataset consists of 1,363,063 cells from embryos between 9.5 to 13.5 days of gestation. The online iNMF analysis required 25 minutes and less than 2 GB of RAM on a MacBook Pro, compared to 98 minutes and 109 GB of RAM for Harmony, which could only be run on a large-memory server. **a-b**, 3D UMAP plot of the online iNMF results (n = 200,000 cells sampled for visualization), colored by dataset (**a**) and published developmental trajectory labels (**b**).



	Alignment	kBET	MERFISH purity	RNA Purity	MERFISH ARI	RNA ARI
Scenario 1	0.72	0.53	0.807	0.784	0.6	0.569
Scenario 3	0.74	0.56	0.802	0.781	0.64	0.57

Figure S2. Scenario 1 and scenario 3 achieve similar results on MERFISH data. The result of scenario 1 on the MERFISH dataset yielded similar MERFISH (**a**) and RNA (**b**) cluster placement as using scenario 3 (**Fig. 5**). The performance of the different approaches showed remarkable similarity, as demonstrated by the alignment, kBET, cluster purity, and ARI scores shown in the table.

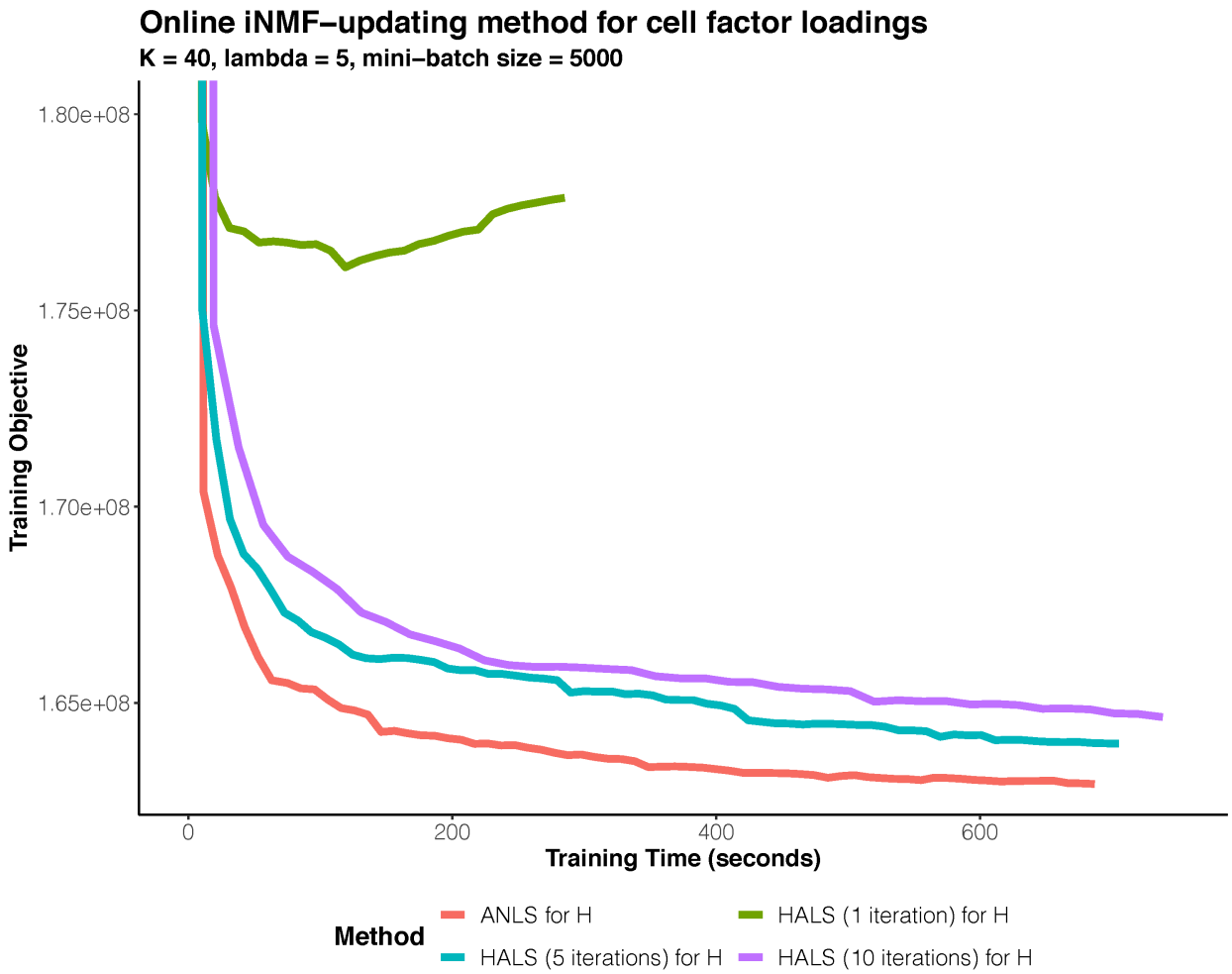


Figure S3. Comparison of methods for updating cell factor loadings (H). The training data are subsets (80%) of the adult mouse frontal (n = 124,934) and posterior cortex (n = 79,349) datasets. 1,111 were selected variable genes for this analysis. ANLS for H clearly outperforms the other in minimizing the objective.

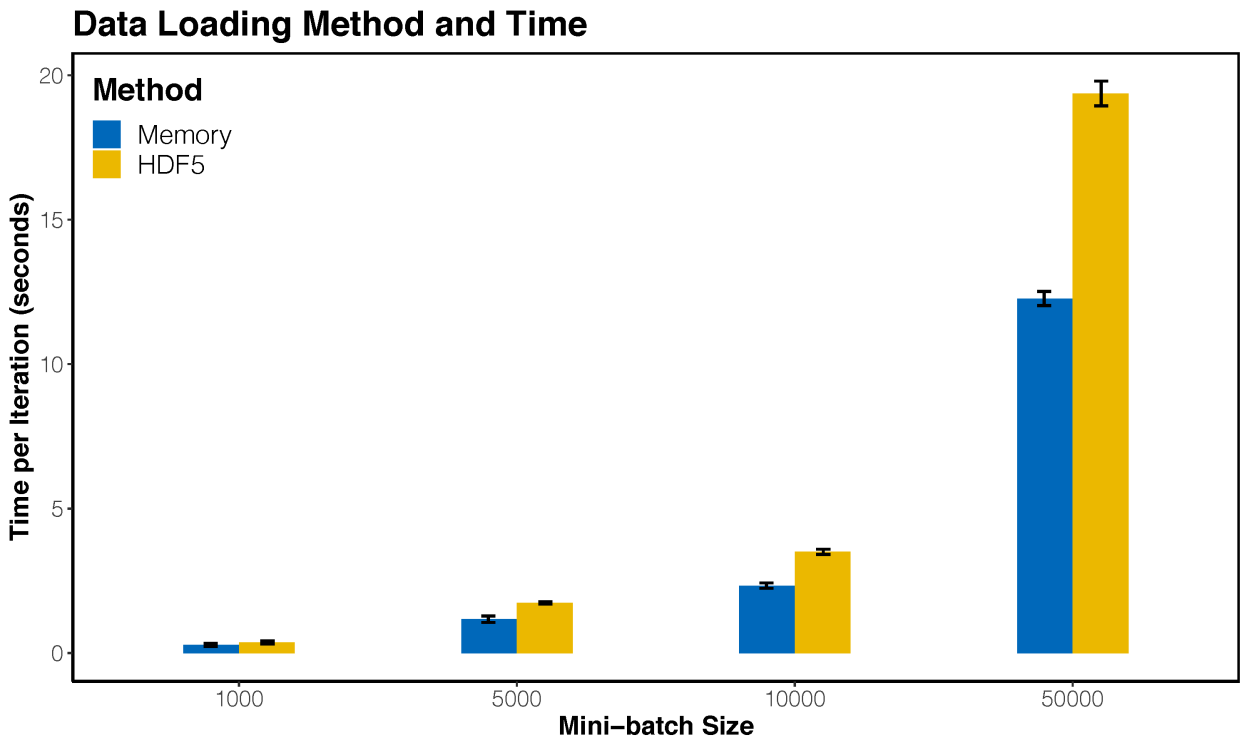


Figure S4. Reading mini-batches from disk adds minimal overhead. In this study, each chunk in HDF5 files stores 1,000 samples (cells). Pulling data from the disk does not add significant overhead compared to loading the data from memory, as long as the mini-batch size is close to the specified chunk size. Mean time per iteration (processing one mini-batch) (±SD) of 50 iterations in each setting is displayed.

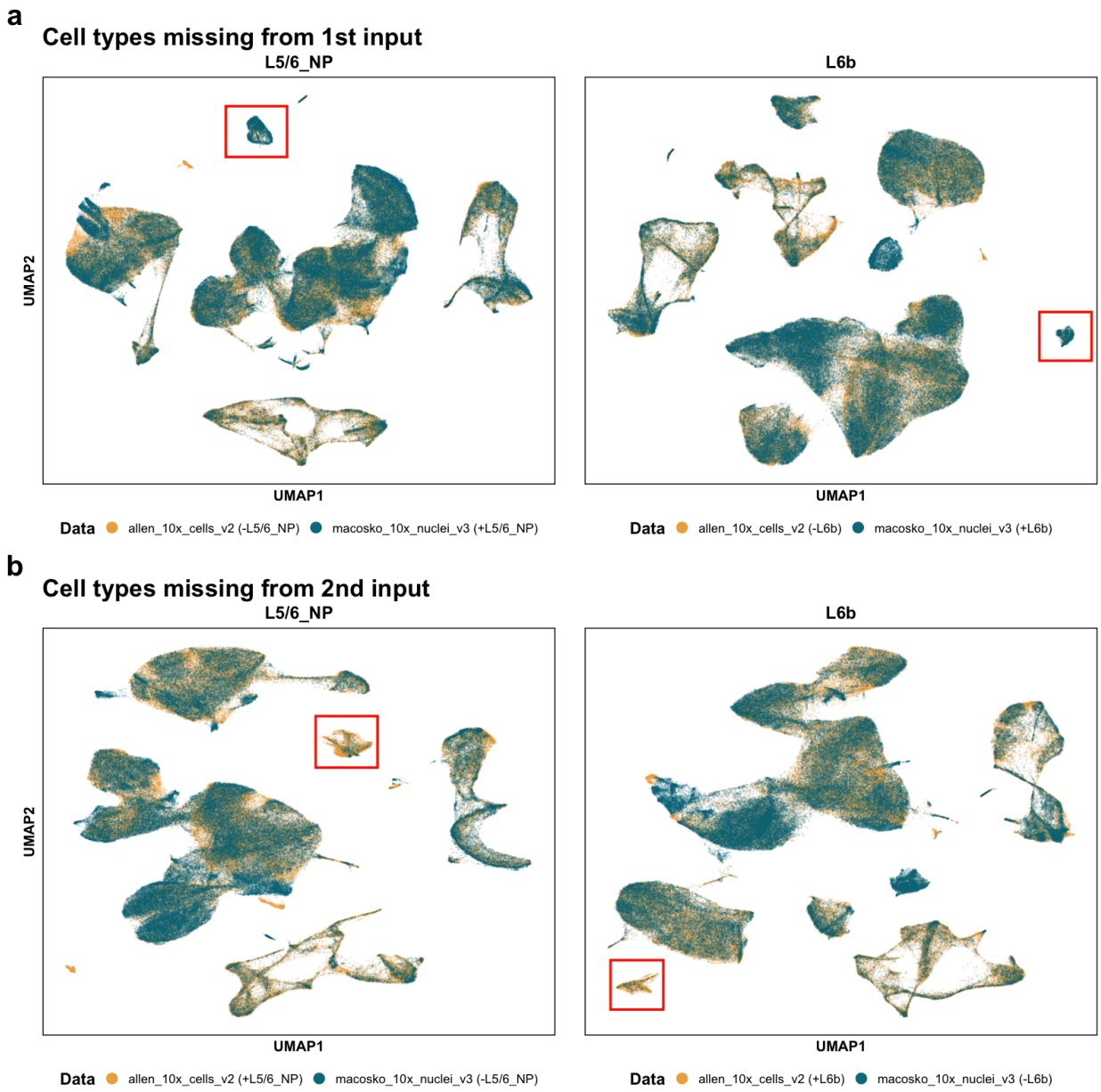


Figure S5. Performance of online iNMF (scenario 2) with missing rare cell clusters (real data). The L5/6_NP and L6b cells missing from early- or late- arriving datasets are successfully identified. **a**, The rare cell types were missing from the first input (allen_10x_cells_v2). **b**, The rare cell types were missing from the second input (macosko_10x_nuclei_v3).

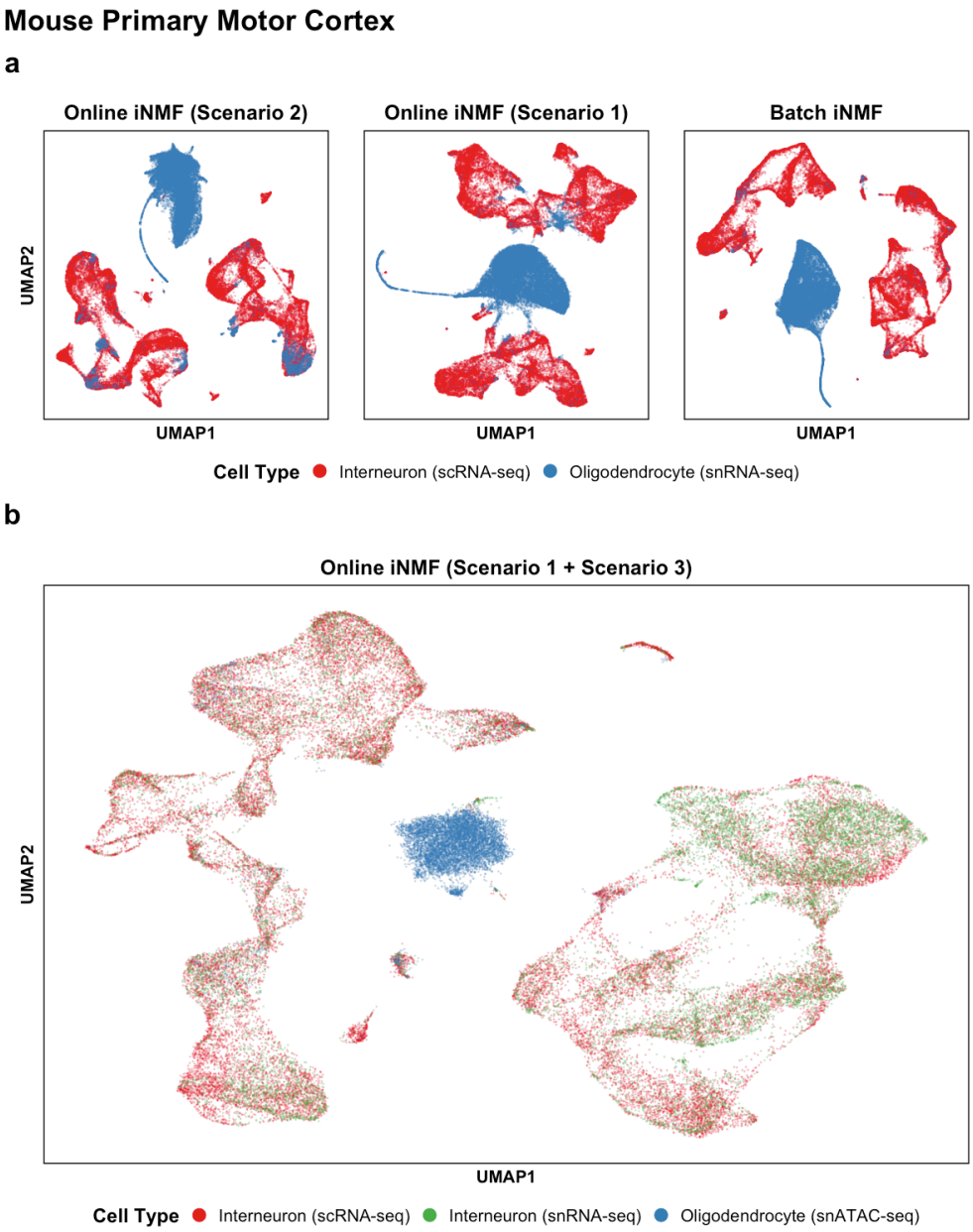


Figure S6. Online iNMF results in minimal spurious alignment for non-overlapping datasets (real data). a, Online iNMF (scenario 1 & 2) and batch iNMF are utilized to integrate one dataset containing only interneurons (scRNA-seq, n = 27,555) and another containing only oligodendrocytes (snRNA-seq, n = 21,404) using 30 metagenes. **b**, Projection of completely non-overlapping dataset into the existing latent space leads to minimal spurious alignment. An scRNA-seq dataset (n = 27,555) and a snRNA-seq dataset (n = 15,255) containing only interneurons are first integrated in scenario 1. Then an snATAC-seq dataset containing only oligodendrocytes (n = 8,557) is projected into this aligned latent space.

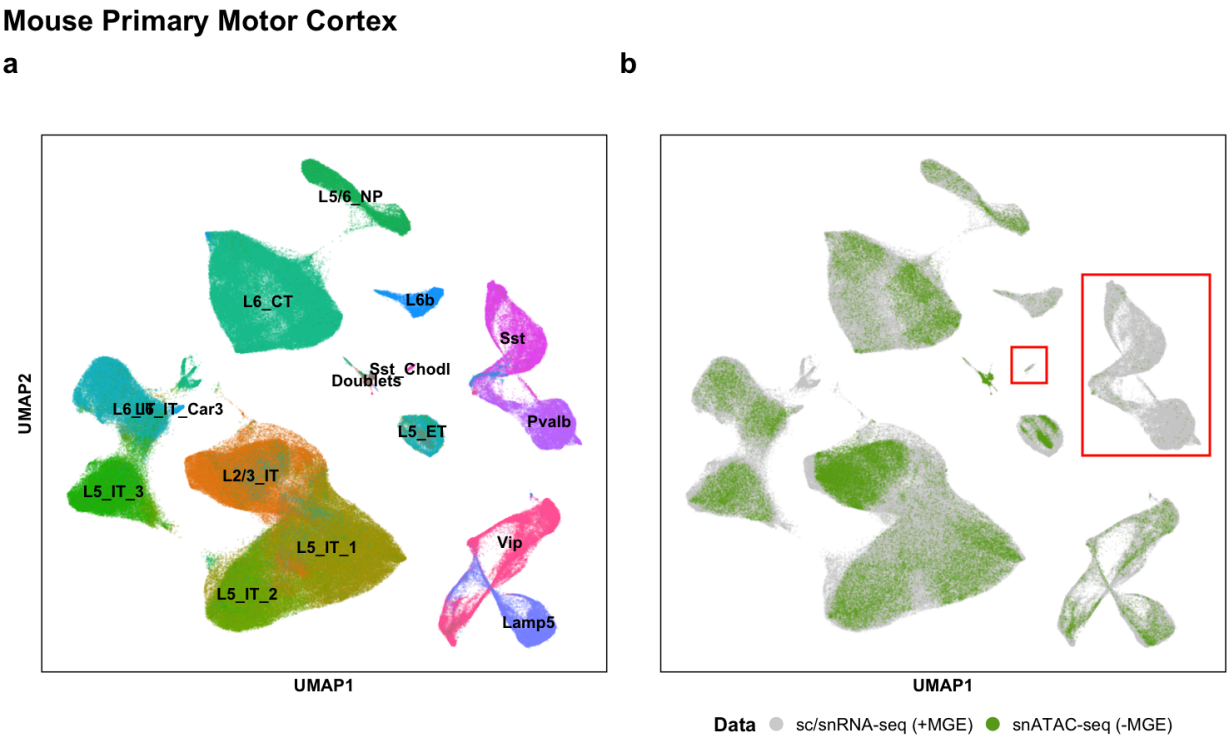


Figure S7. Online iNMF (scenario 3) leads to little spurious alignment when integrating partially-overlapping datasets. Six sc/snRNA-seq datasets from the MOp (n = 344,675) were integrated using online iNMF (scenario 1). Then an snATAC-seq dataset (n = 49,167) without MGEs (i.e. Pvalb, Sst and Chodl cells) was projected (scenario 3) into the atlas already built. **a**, The UMAP visualization annotated by our cell class labels. **b**, UMAP plot colored by dataset. Almost no cells from snATAC-seq data are observed in the clusters corresponding to Pvalb, Sst and Chodl cells (red boxes).

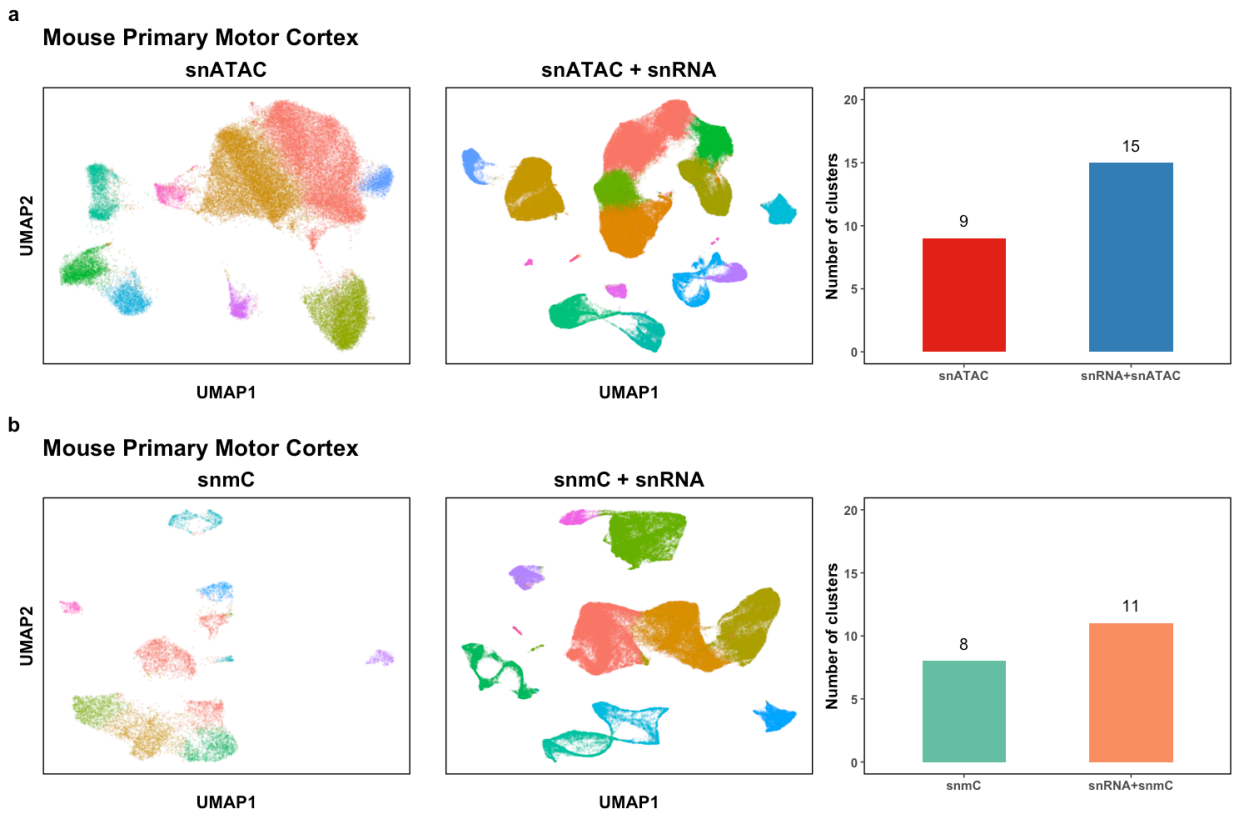


Figure S8. Integrating methylation or chromatin accessibility data with RNA data better separates clusters. a, 6 more clusters are observed after joint analysis of snATAC-seq data (n = 54,844) and snRNA-seq data (n = 101,647) than analysis of snATAC-seq data alone. **b**, 3 more clusters are obtained after incorporating the snRNA-seq data than investigating snmC-seq data (n = 9,366) alone.

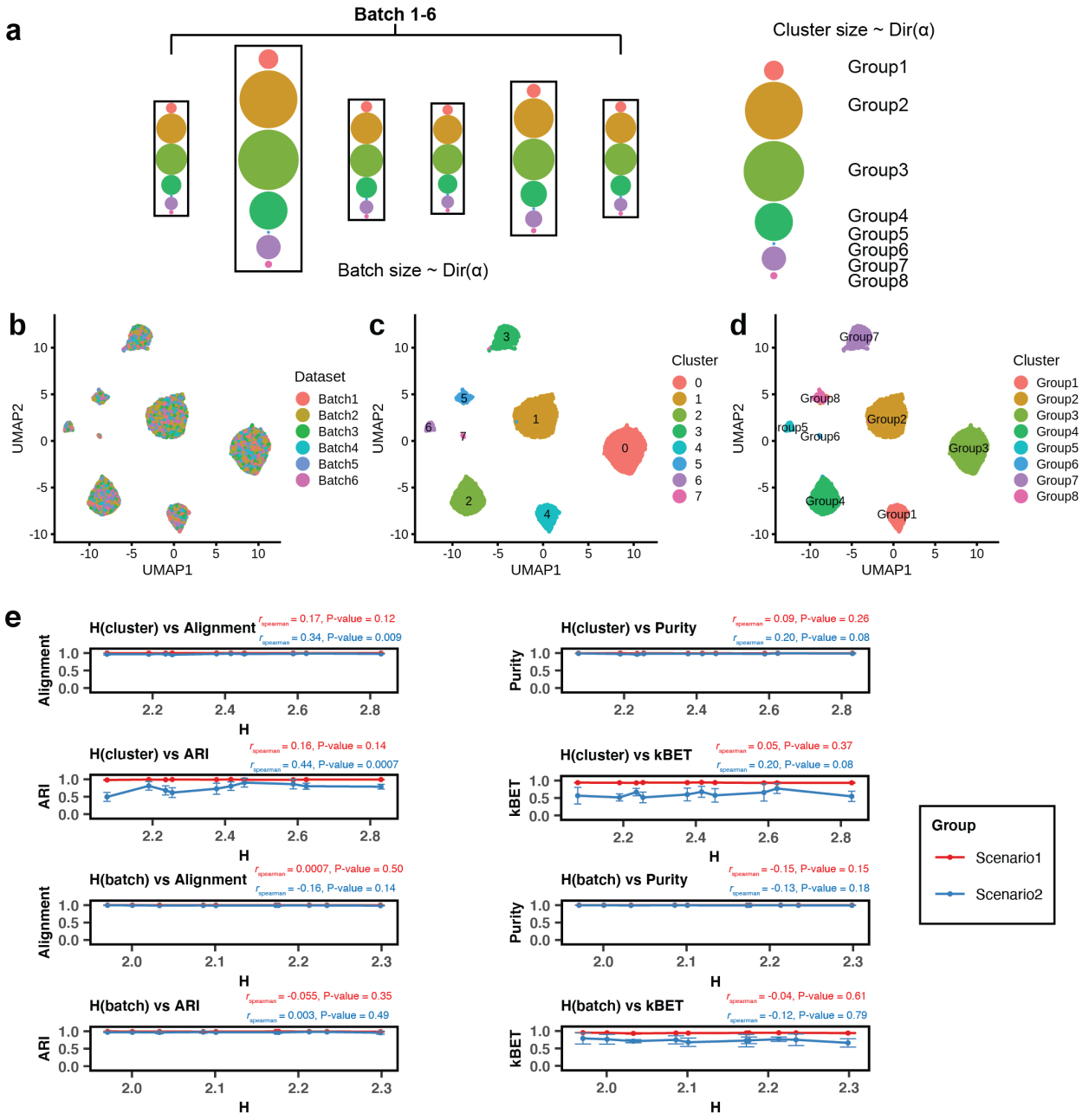


Figure S9. Performance of online iNMF under imbalanced cell clusters and dataset sizes (simulations). a, Schematic plot showing the composition of 8 clusters and 6 batches in each simulated dataset (with 10000 genes and 50000 cells). **b-d**) UMAP representations of an example integration result plotted using batch labels (**b**), LIGER cluster assignments (**c**), and ground truth cluster labels (**d**). **e**, Line plots of four evaluation metric scores for online iNMF (scenario 1 & 2) versus the Shannon entropy of cell type and batch size (larger H means more balanced composition). The data are presented as mean values $\pm$ standard deviation (5 random initializations for each simulated dataset, n = 50,000 cells in each simulated dataset). The p-value was obtained from one sided Spearman's rank correlation test without adjustment for multiple comparisons.

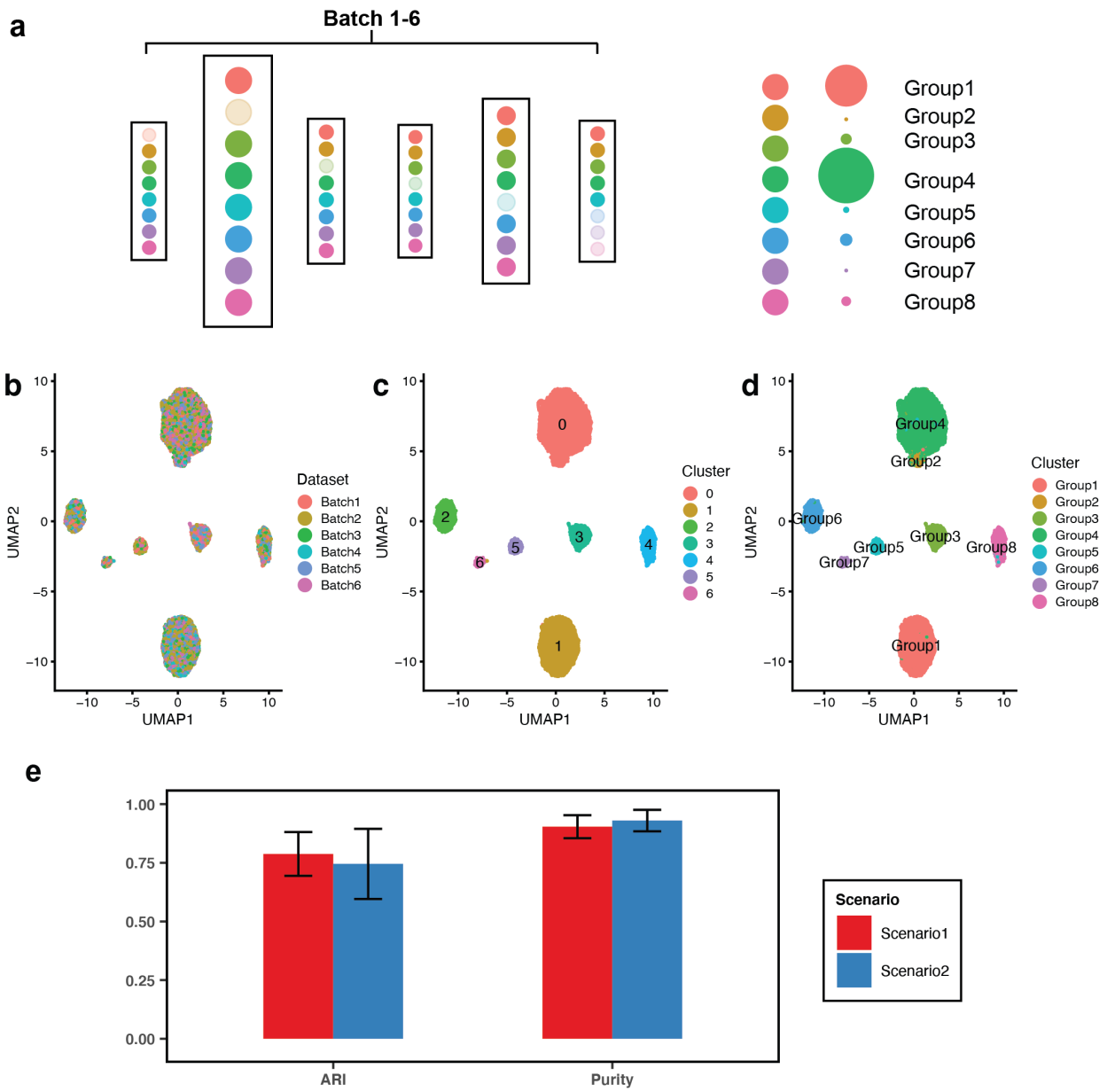


Figure S10: Performance of online iNMF with no cell types shared across all datasets (simulations). a, Schematic plot showing the composition of 8 clusters and 6 batches in each of ten simulated datasets; data were further filtered to make sure the intersection of all batches in each simulation is the empty set while the pairwise intersections of all batches are non-empty. **b-d**, UMAP representations of an integration example result under scenarios 1 & 2 plotted using batch labels (**b**), LIGER cluster assignments (**c**), and ground truth cluster labels (**d**). **e**, Bar plot of the two evaluation metric scores for online iNMF (scenario 1 & 2) simulated data. The data are presented as mean values $\pm$ standard deviation (50 runs in total for each metric, n = 50,000 cells in each simulated dataset).

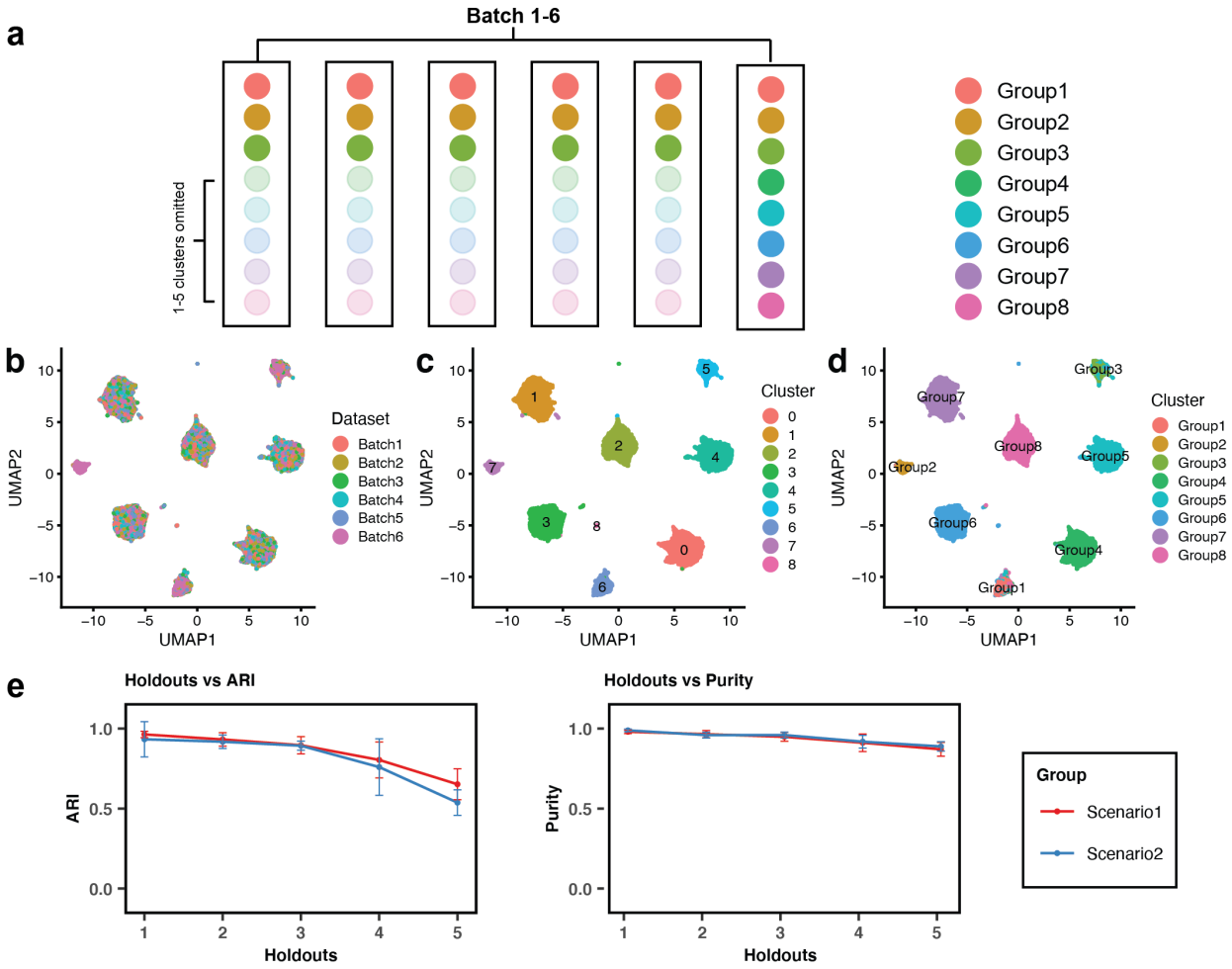


Figure S11: Performance of online iNMF (scenario 1 and 2) with missing cell clusters (simulations). a, Schematic plot showing the equal proportions of 8 clusters and 6 batches in each simulated dataset (with 10,000 genes and 50,000 cells) with 1-5 cell types excluded. **b-d**, UMAP representations of an example integration result from scenario 1 from a simulation with three held-out cell types. The plots are colored using batch labels (**b**), LIGER cluster assignments (**c**), and ground truth cluster labels (**d**). **e**, Line plots of two evaluation metrics for online iNMF (scenario 1 & 2) versus the number of cell types excluded. The data are presented as mean values $\pm$ standard deviation (10 random initializations for each simulated dataset, n = 50,000 cells in each simulated dataset before holding out any cell clusters).

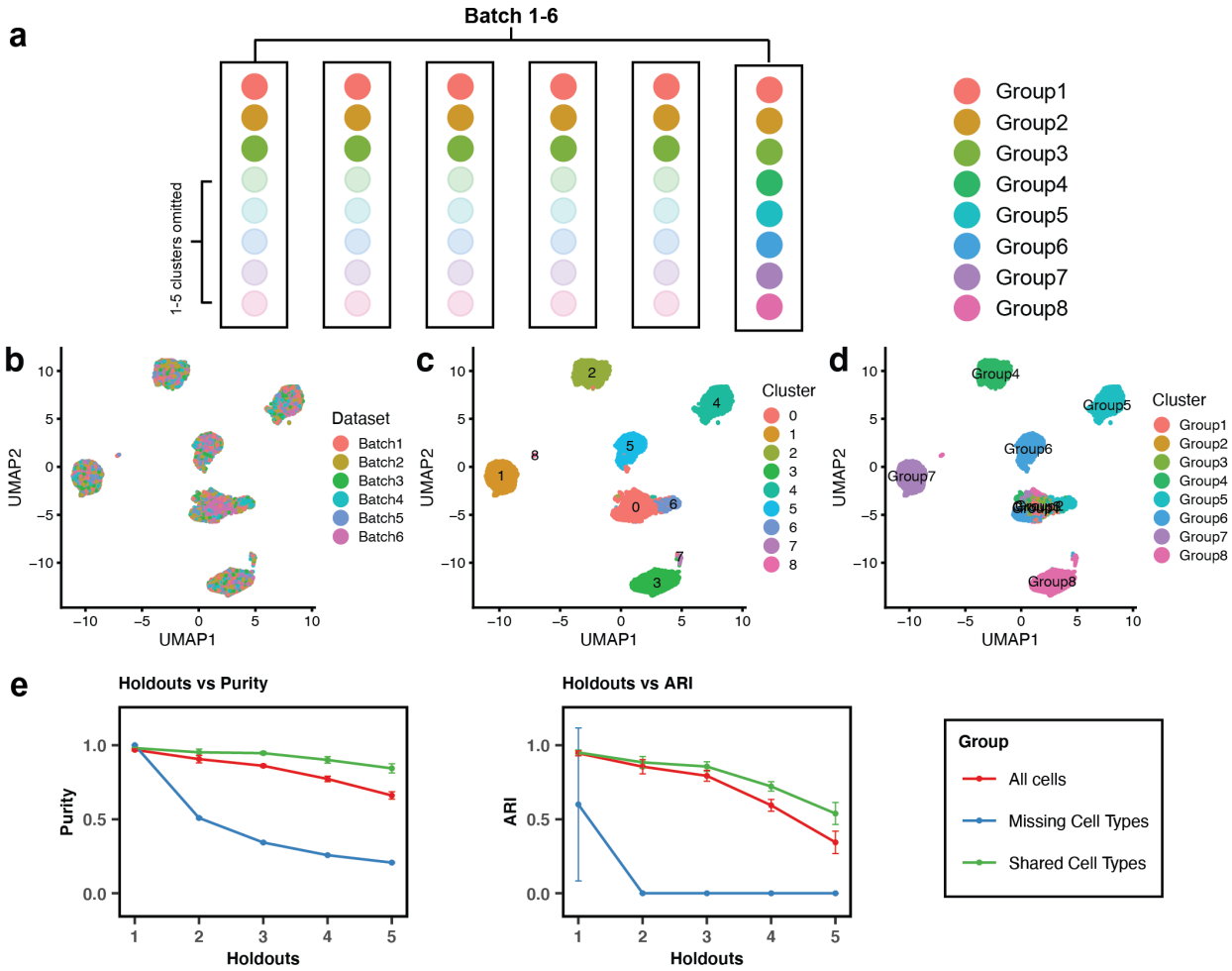


Figure S12: Performance of online iNMF (scenario 3) with missing cell clusters (simulations). a, Schematic plot showing the equal proportions of 8 clusters and 6 batches in each simulated dataset (with 10000 genes and 50000 cells) with 1-5 cell types excluded. **b-d**, UMAP representations of an example integration result from scenario 3 from a simulation with 3 held-out cell types. The plots are colored using batch labels (**b**), LIGER cluster assignments (**c**), and ground truth cluster labels (**d**). **e**, Line plots of two evaluation metric scores for online iNMF (scenario 3) on all cells (red line), cells in missing clusters (blue line), and cells in shared clusters (green line), versus the number of cell types excluded. The data are presented as mean values $\pm$ standard deviation (10 random initializations for each simulated dataset, n = 50,000 cells in each simulated dataset before holding out any cell clusters).

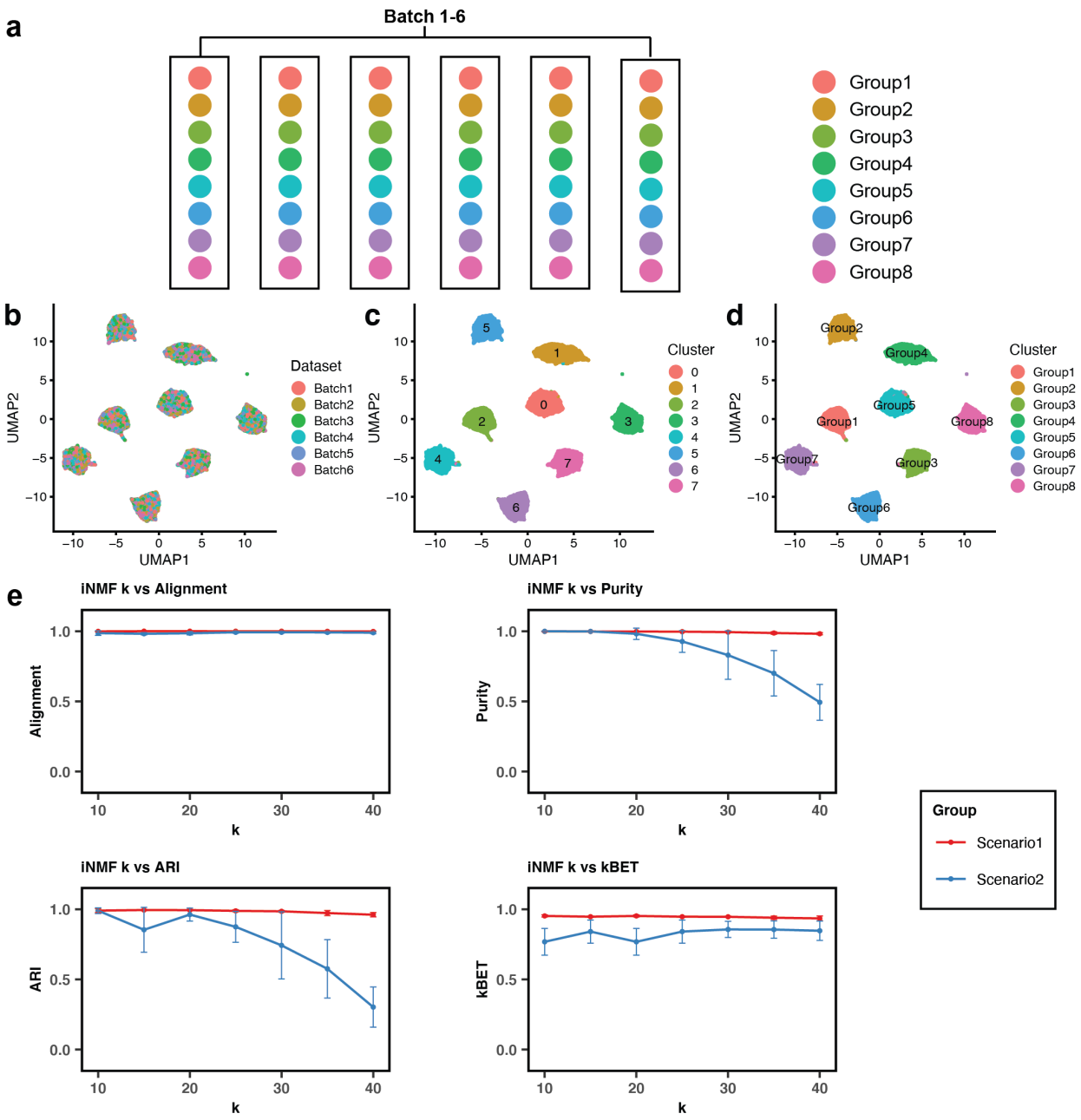


Figure S13. Performance of online iNMF (scenario 1 and 2) varying K (number of metagenes) (simulations). a, Schematic plot showing the composition of 8 clusters and 6 batches in each simulated dataset (with 10000 genes and 50000 cells). **b-d**, UMAP representations of an example result plotted using batch labels (b), LIGER cluster assignments (c), and ground truth cluster labels (d). **e**, Line plots of four evaluation metric scores for online iNMF (scenario 1 & 2) versus K varying from 10 to 40 incremented by 5. The data are presented as mean values $\pm$ standard deviation (10 random initializations for each K, n = 50,000 cells in each simulated dataset).

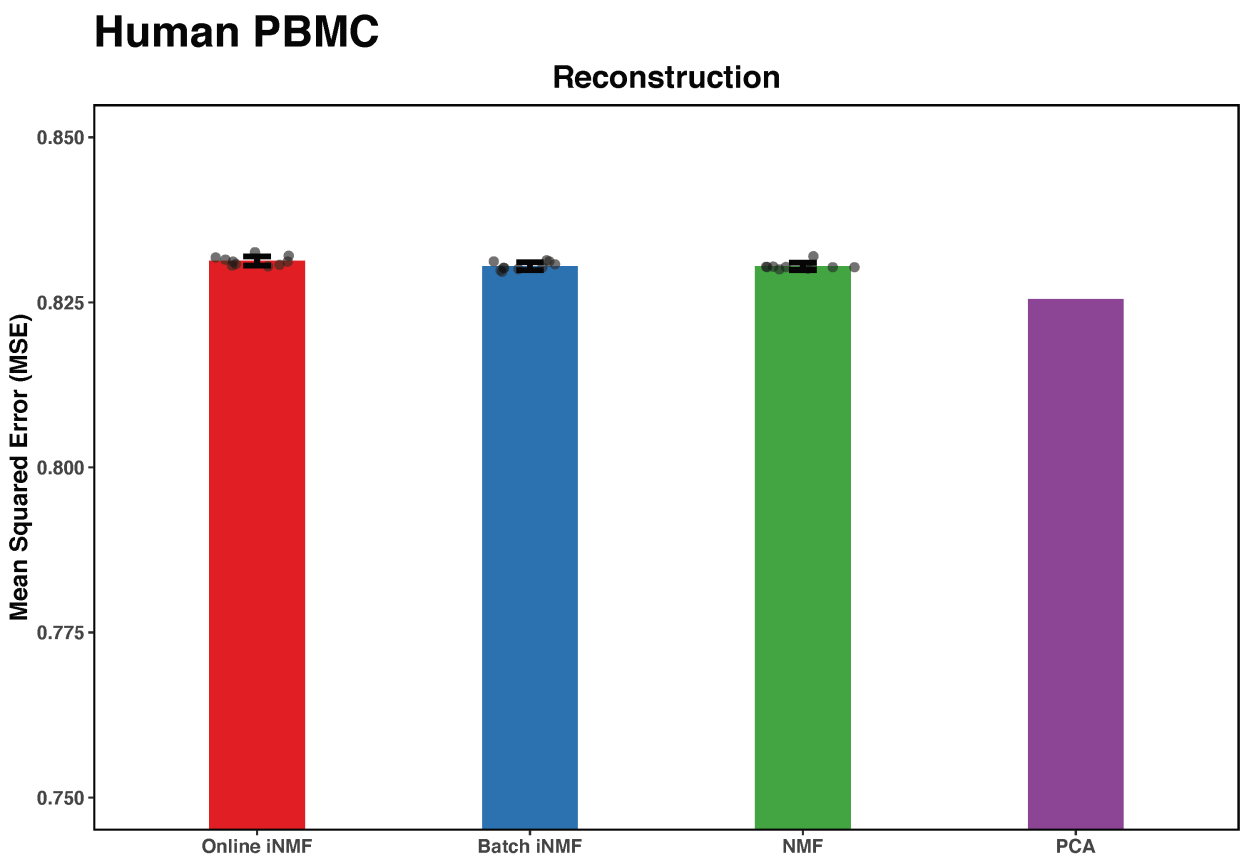


Figure S14. Comparison of data reconstruction among iNMF, NMF and PCA. Human PBMC dataset (n = 13,999) was used for this analysis. Average mean squared error (MSE) is shown and error bars indicate the standard deviation (10 random initializations for each method except for PCA, which is deterministic). Individual data points are shown for the NMF approaches.

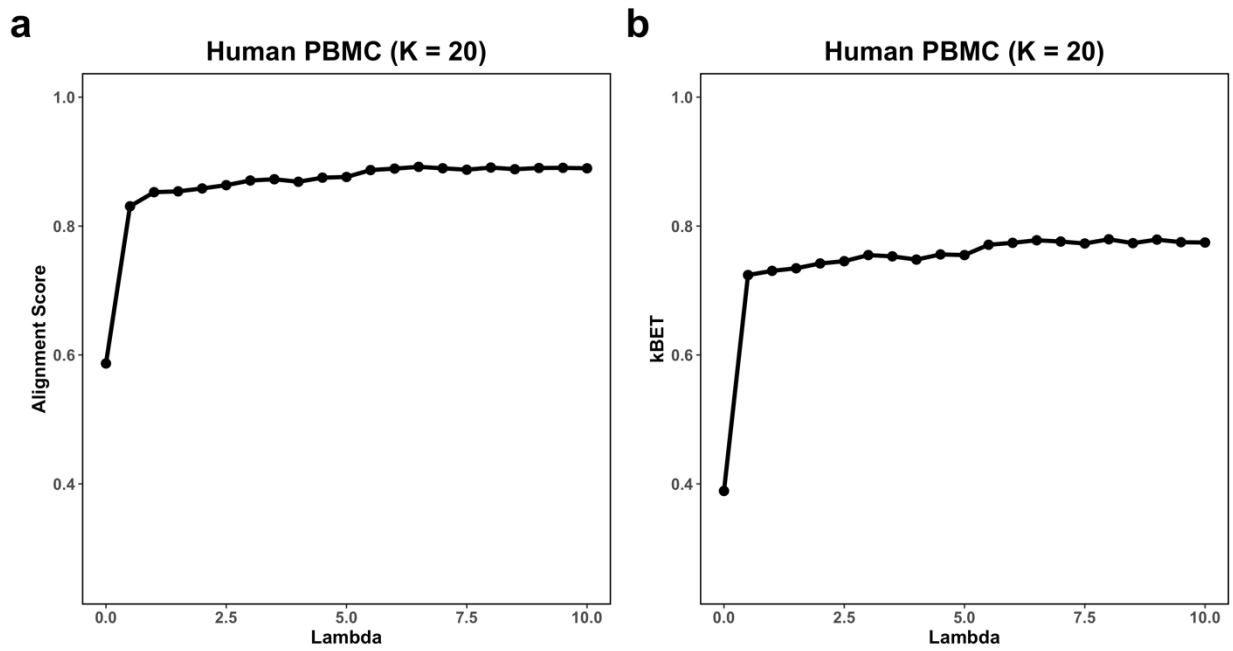


Figure S15. Selecting λ on human PBMC dataset. The human PBMC datasets (n = 13,999) were used to demonstrate the effect of λ on the data integration. Alignment score (**a**) and kBET (**b**) are reported to quantitatively assess dataset integration.

Supplementary Note: Tables

Corres. figure	Integration Strategy	K	λ	Mini-batch size	Variable Genes	# of iNMF initializations	# of simulations	# of total runs
Figure S13	Scenario 1	20	5	5000	~3000, selected from all batches	5	20	100
	Scenario 2	20	10	1000	3000, selected from the first batch	5	20	100
Figure S14	Scenario 1	20	5	5000	~3000, selected from all batches	5	10	50
	Scenario 2	20	10	1000	3000, selected from the first batch	5	10	50
Figure S15	Scenario 1	20	5	5000	~3000, selected from all batches	10	5	50
	Scenario 2	20	10	1000	3000, selected from the first batch	10	5	50
Figure S16	Scenario 3	20	5	5000	~3000, selected from batches with missing cell types	10	5	50
Figure S17	Scenario 1	10-40	5	5000	~3000, selected from all batches	1	70	70
	Scenario 2	10-40	10	1000	3000, selected from the first batch	1	70	70

Table S1. Key parameter settings for integrated analysis on simulated data

Supplementary Note: Derivation of Hierarchical Alternating Least Squares Updates for iNMF

1 iNMF objective

Given N input datasets $(X^1)_{m\times n_1}, \ldots, (X^N)_{m\times n_N}$, we want to jointly factorize them into $W_{m\times K}$, $V^i_{m\times K}$, and $H^i_{n_i\times K}$.

$$\min_{\substack{W,V^i,H^i\geq 0\\ i=1,\ldots,N}} \sum_{i=1}^{N} (\left\|X^i - (W+V^i)H^{i\top}\right\|_F^2 + \lambda \left\|V^i H^{i\top}\right\|_F^2) \tag{1}$$

2 Notation

$\boldsymbol{w}_j$: column vector, jth factor of W
$\boldsymbol{v}^i_j$: column vector, jth factor of V^i
$\boldsymbol{h}^i_j$: column vector, jth factor of H^i

3 W update

$$\begin{aligned} g(\boldsymbol{w}_j) &= \sum_i \left\|X^i - WH^{i\top} - V^i H^{i\top}\right\|_F^2 \\ &= \sum_i \left\|X^i - \sum_k^K \boldsymbol{w}_k \boldsymbol{h}_k^{i\top} - V^i H^{i\top}\right\|_F^2 \\ &= \sum_i \left\|\boldsymbol{w}_j \boldsymbol{h}_j^{i\top} - (X^i - V^i H^{i\top} - \sum_{k\neq j}^K \boldsymbol{w}_k \boldsymbol{h}_k^{i\top})\right\|_F^2 \end{aligned} \tag{2}$$

Let $Q^i_j = X^i - V^i H^{i\top} - \sum_{k\neq j}^K \boldsymbol{w}_k \boldsymbol{h}_k^{i\top}$; then we have

$$\begin{aligned} \boldsymbol{w}_j^* &= \arg\min_{\boldsymbol{w}_j\geq 0} g(\boldsymbol{w}_j) \\ &= \arg\min_{\boldsymbol{w}_j\geq 0} \sum_i \left\|\boldsymbol{w}_j^i \boldsymbol{h}_j^{i\top} - Q^i_j\right\|_F^2 \\ &= \arg\min_{\boldsymbol{w}_j\geq 0} \sum_i \mathbf{Tr}((\boldsymbol{w}_j \boldsymbol{h}_j^{i\top} - Q^i_j)^\top (\boldsymbol{w}_j \boldsymbol{h}_j^{i\top} - Q^i_j)) \\ &= \arg\min_{\boldsymbol{v}_j^i\geq 0} \sum_i \mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{w}_j^\top \boldsymbol{w}_j \boldsymbol{h}_j^{i\top} - \boldsymbol{h}_j^i \boldsymbol{w}_j^{i\top} Q^i_j - Q^{i\top}_j \boldsymbol{w}_j \boldsymbol{h}_j^{i\top} + Q^{i\top}_j Q^i_j) \end{aligned} \tag{3}$$

$Q_j^{i\top} Q_j^i$ is a constant with respect to $\boldsymbol{w}_j$, therefore, (3) is equivalent to the following

$$\begin{aligned}
\boldsymbol{w}_j^* &= \arg\min_{\boldsymbol{w}_j \geq 0} g^*(\boldsymbol{w}_j) \\
&= \arg\min_{\boldsymbol{v}_j^i \geq 0} \sum_i \mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{w}_j^\top \boldsymbol{w}_j \boldsymbol{h}_j^{i\top} - \boldsymbol{h}_j^i \boldsymbol{w}_j^{i\top} Q_j^i - Q_j^{i\top} \boldsymbol{w}_j \boldsymbol{h}_j^{i\top}) \\
&= \arg\min_{\boldsymbol{w}_j \geq 0} \sum_i \mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{w}_j^\top \boldsymbol{w}_j \boldsymbol{h}_j^{i\top}) - 2\,\mathbf{Tr}(\boldsymbol{w}_j \boldsymbol{h}_j^{i\top} Q_j^{i\top})
\end{aligned} \tag{4}$$

Based on the property of matrix trace, we have the following,
$\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{w}_j^\top \boldsymbol{w}_j \boldsymbol{h}_j^{i\top}) = \mathbf{Tr}(\boldsymbol{w}_j^\top \boldsymbol{w}_j \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) = \boldsymbol{w}_j^\top \boldsymbol{w}_j \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i$
$\mathbf{Tr}(\boldsymbol{w}_j \boldsymbol{h}_j^{i\top} Q_j^{i\top}) = (Q_j^i \boldsymbol{h}_j^i)^\top \boldsymbol{w}_j$
Therefore,

$$\boldsymbol{w}_j^* = \arg\min_{\boldsymbol{w}_j \geq 0} \sum_i (\boldsymbol{w}_j^\top \boldsymbol{w}_j)(\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) - 2(Q_j^i \boldsymbol{h}_j^i)^\top \boldsymbol{w}_j \tag{5}$$

To obtain $\boldsymbol{w}_j^*$, we set $\frac{\partial g^*(\boldsymbol{w}_j)}{\partial \boldsymbol{w}_j} = \mathbf{0}$.

$$\frac{\partial g^*(\boldsymbol{w}_j)}{\partial \boldsymbol{w}_j} = 2 \sum_i \boldsymbol{w}_j (\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) - Q_j^i \boldsymbol{h}_j^i = \mathbf{0} \tag{6}$$

$$\begin{aligned}
\boldsymbol{w}_j^* &= \left[\frac{\sum_i Q_j^i \boldsymbol{h}_j^i}{\sum_i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i} \right]_+ \\
&= \left[\frac{\sum_i (X^i - W H^{i\top} - V^i H^{i\top} + \boldsymbol{w}_j \boldsymbol{h}_j^{i\top}) \boldsymbol{h}_j^i}{\sum_i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i} \right]_+ \\
&= \left[\frac{\sum_i X^i \boldsymbol{h}_j^i - (W + V^i) H^{i\top} \boldsymbol{h}_j^i + \boldsymbol{w}_j \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i}{\sum_i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i} \right]_+ \\
&= \left[\boldsymbol{w}_j + \frac{\sum_i X^i \boldsymbol{h}_j^i - (W + V^i) H^{i\top} \boldsymbol{h}_j^i}{\sum_i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i} \right]_+ \\
W_{\cdot j}^* &= \left[W_{\cdot j} + \frac{\sum_i (X^i H^i)_{\cdot j} - (W + V^i)(H^{i\top} H^i)_{\cdot j}}{\sum_i (H^{i\top} H^i)_{jj}} \right]_+
\end{aligned} \tag{7}$$

Note that the nonnegativity constraint is enforced by thresholding, where $[a]_+ = \max(a, 0)$, assuming a is a scalar. As shown above, we have

$$\frac{\partial g^*(\boldsymbol{w}_j)}{\partial \boldsymbol{w}_j} = 2 \sum_i \boldsymbol{w}_j (\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) - Q_j^i \boldsymbol{h}_j^i$$

Then for mth element in $\boldsymbol{w}_j$

$$g^*((\boldsymbol{w}_j)_m) = \sum_i (\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i)(\boldsymbol{w}_j)_m^2 - 2(Q_j^i \boldsymbol{h}_j^i)_m (\boldsymbol{w}_j)_m$$

Here we assume $\boldsymbol{h}_j^i$ is not a zero-vector, therefore $\sum_i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i > 0$.

$$\frac{\partial g^*((\boldsymbol{w}_j)_m)}{\partial (\boldsymbol{w}_j)_m} = 2 \sum_i (\boldsymbol{w}_j)_m (\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) - (Q_j^i \boldsymbol{h}_j^i)_m$$

If $(Q_j^i \boldsymbol{h}_j^i)_m \geq 0$, then the minimal value of the function is achieved when $(\boldsymbol{w}_j)_m = \frac{(Q_j^i \boldsymbol{h}_j^i)_m}{\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i}$.
If $(Q_j^i \boldsymbol{h}_j^i)_m < 0$, then the value of function increases as $(\boldsymbol{w}_j)_m$ gets larger when $(\boldsymbol{w}_j)_m \geq 0$. Hence the minimal value of the function is achieved at $(\boldsymbol{w}_j)_m = 0$.
Combining both cases, $[\cdot]_+$ is used in expressing the resulting updates.

4 V update

For each V^i

$$\begin{aligned} h(\boldsymbol{v}_j^i) &= \left\|X^i - WH^{i\top} - V^i H^{i\top}\right\|_F^2 + \lambda \left\|V^i H^{i\top}\right\|_F^2 \\ &= \left\|X^i - WH^{i\top} - \sum_k^K \boldsymbol{v}_k^i \boldsymbol{h}_k^{i\top}\right\|_F^2 + \lambda \left\|\sum_k^K \boldsymbol{v}_k^i \boldsymbol{h}_k^{i\top}\right\|_F^2 \\ &= \left\|\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} - (X^i - WH^{i\top} - \sum_{k\neq j}^K \boldsymbol{v}_k^i \boldsymbol{h}_k^{i\top})\right\|_F^2 + \lambda \left\|\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + \sum_{k\neq j}^K \boldsymbol{v}_k^i \boldsymbol{h}_k^{i\top}\right\|_F^2 \end{aligned} \tag{8}$$

Let $S_j^i = \sum_{k\neq j}^K \boldsymbol{v}_k^i \boldsymbol{h}_k^{i\top} = V^i H^{i\top} - \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top}$ and $R_j^i = X^i - WH^{i\top} - S_j^i$; then we have

$$h(\boldsymbol{v}_j^i) = \left\|\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} - R_j^i\right\|_F^2 + \lambda \left\|\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + S_j^i\right\|_F^2 \tag{9}$$

Then

$$\begin{aligned} \boldsymbol{v}_j^{i*} &= \arg\min_{\boldsymbol{v}_j^i \geq 0} h(\boldsymbol{v}_j^i) \\ &= \arg\min_{\boldsymbol{v}_j^i \geq 0} \left\|\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} - R_j^i\right\|_F^2 + \lambda \left\|\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + S_j^i\right\|_F^2 \\ &= \arg\min_{\boldsymbol{v}_j^i \geq 0} \mathbf{Tr}((\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} - R_j^i)^\top (\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} - R_j^i)) + \lambda \,\mathbf{Tr}((\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + S_j^i)^\top (\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + S_j^i)) \\ &= \arg\min_{\boldsymbol{v}_j^i \geq 0} \mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} - \boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} R_j^i - R_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + R_j^{i\top} R_j^i) \\ &\quad + \lambda \,\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + \boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} S_j^i + S_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} + S_k^{i\top} S_j^i) \end{aligned} \tag{10}$$

Since $R_k^{i\top} R_j^i$ and $S_k^{i\top} S_j^i$ are constant, minimizing (10) is equivalent to minimizing the following

$$\begin{aligned} \boldsymbol{v}_j^{i*} &= \arg\min_{\boldsymbol{v}_j^i \geq 0} h^*(\boldsymbol{v}_j^i) \\ &= \arg\min_{\boldsymbol{v}_j^i \geq 0} \mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top}) - 2\,\mathbf{Tr}(\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} R_j^{i\top}) + \lambda \,\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top}) + 2\,\mathbf{Tr}(\lambda \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} S_j^{i\top}) \\ &= \arg\min_{\boldsymbol{v}_j^i \geq 0} (1+\lambda)\,\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top}) + 2\,\mathbf{Tr}(\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} (\lambda S_j^{i\top} - R_j^{i\top})) \end{aligned} \tag{11}$$

Based on the property of matrix trace, we have the following,
$\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top}) = \mathbf{Tr}(\boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) = \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i$
$\mathbf{Tr}(\boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top} (\lambda S_j^{i\top} - R_j^{i\top})) = ((\lambda S_j^i - R_j^i)\boldsymbol{h}_j^i)^\top \boldsymbol{v}_j^i$
Hence,

$$\boldsymbol{v}_j^{i*} = \arg\min_{\boldsymbol{v}_j^i \geq 0} (1+\lambda)(\boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i)(\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) + 2((\lambda S_j^i - R_j^i)\boldsymbol{h}_j^i)^\top \boldsymbol{v}_j^i \tag{12}$$

By setting $\frac{\partial h^*(\boldsymbol{v}_j^i)}{\partial \boldsymbol{v}_j^i} = \mathbf{0}$ we can solve for $\boldsymbol{v}_j^{i*}$, as shown below.

$$\frac{\partial h^*(\boldsymbol{v}_j^i)}{\partial \boldsymbol{v}_j^i} = 2(1+\lambda)\boldsymbol{v}_j^i (\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) + 2(\lambda S_j^i - R_j^i)\boldsymbol{h}_j^i = \mathbf{0} \tag{13}$$

$$
\begin{aligned}
\boldsymbol{v}_j^{i*} &= \left[\frac{(R_j^i - \lambda S_j^i)\boldsymbol{h}_j^i}{(1+\lambda)\boldsymbol{h}_j^{i\top}\boldsymbol{h}_j^i}\right]_+ \\
&= \left[\frac{(X^i - WH^{i\top} - (1+\lambda)(V^iH^{i\top} - \boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top}))\boldsymbol{h}_j^i}{(1+\lambda)\boldsymbol{h}_j^{i\top}\boldsymbol{h}_j^i}\right]_+ \\
&= \left[\frac{(X^i - WH^{i\top} - (1+\lambda)(V^iH^{i\top} - \boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top}))\boldsymbol{h}_j^i}{(1+\lambda)\boldsymbol{h}_j^{i\top}\boldsymbol{h}_j^i}\right]_+ \\
&= \left[\frac{X^i\boldsymbol{h}_j^i - (W + (1+\lambda)V^i)H^{i\top}\boldsymbol{h}_j^i + (1+\lambda)\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top}\boldsymbol{h}_j^i}{(1+\lambda)\boldsymbol{h}_j^{i\top}\boldsymbol{h}_j^i}\right]_+ \\
&= \left[\boldsymbol{v}_j^i + \frac{X^i\boldsymbol{h}_j^i - (W + (1+\lambda)V^i)H^{i\top}\boldsymbol{h}_j^i}{(1+\lambda)\boldsymbol{h}_j^{i\top}\boldsymbol{h}_j^i}\right]_+ \\
V_{\cdot j}^{i*} &= \left[V_{\cdot j}^i + \frac{(X^iH^i)_{\cdot j} - (W + (1+\lambda)V^i)(H^{i\top}H^i)_{\cdot j}}{(1+\lambda)(H^{i\top}H^i)_{jj}}\right]_+
\end{aligned}
\tag{14}
$$

5 H update

For each H^i

$$
\begin{aligned}
l(\boldsymbol{h}_j^i) &= \left\|X^i - (W+V^i)H^{i\top}\right\|_F^2 + \lambda\left\|V^iH^{i\top}\right\|_F^2 \\
&= \left\|X^i - \sum_k^K(\boldsymbol{w}_k + \boldsymbol{v}_k^i)\boldsymbol{h}_k^{i\top}\right\|_F^2 + \lambda\left\|\sum_k^K \boldsymbol{v}_k^i\boldsymbol{h}_k^{i\top}\right\|_F^2 \\
&= \left\|(\boldsymbol{w}_j + \boldsymbol{v}_j^i)\boldsymbol{h}_j^{i\top} - (X^i - \sum_{k\neq j}^K(\boldsymbol{w}_k + \boldsymbol{v}_k^i)\boldsymbol{h}_k^{i\top})\right\|_F^2 + \lambda\left\|\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top} + \sum_{k\neq j}^K \boldsymbol{v}_k^i\boldsymbol{h}_k^{i\top}\right\|_F^2
\end{aligned}
\tag{15}
$$

Let $\boldsymbol{z}_j^i = \boldsymbol{w}_j + \boldsymbol{v}_j^i$
$S_j^i = \sum_{k\neq j}^K \boldsymbol{v}_k^i\boldsymbol{h}_k^{i\top} = V^iH^{i\top} - \boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top}$
$U_j^i = X^i - \sum_{k\neq j}^K(\boldsymbol{w}_k + \boldsymbol{v}_k^i)\boldsymbol{h}_k^{i\top}$; then we have

$$
l(\boldsymbol{h}_j^i) = \left\|\boldsymbol{z}_j^i\boldsymbol{h}_j^{i\top} - U_j^i\right\|_F^2 + \lambda\left\|\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top} + S_j^i\right\|_F^2 \tag{16}
$$

Then

$$
\begin{aligned}
\boldsymbol{h}_j^{i*} &= \arg\min_{\boldsymbol{h}_j^i\geq 0} l(\boldsymbol{h}_j^i) \\
&= \arg\min_{\boldsymbol{h}_j^i\geq 0} \left\|\boldsymbol{z}_j^i\boldsymbol{h}_j^{i\top} - U_j^i\right\|_F^2 + \lambda\left\|\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top} + S_j^i\right\|_F^2 \\
&= \arg\min_{\boldsymbol{h}_j^i\geq 0} \mathbf{Tr}((\boldsymbol{z}_j^i\boldsymbol{h}_j^{i\top} - U_j^i)^\top(\boldsymbol{z}_j^i\boldsymbol{h}_j^{i\top} - U_j^i)) + \lambda\,\mathbf{Tr}((\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top} + S_j^i)^\top(\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top} + S_j^i)) \\
&= \arg\min_{\boldsymbol{h}_j^i\geq 0} \mathbf{Tr}(\boldsymbol{h}_j^i\boldsymbol{z}_j^{i\top}\boldsymbol{z}_j^i\boldsymbol{h}_j^{i\top} - \boldsymbol{h}_j^i\boldsymbol{z}_j^{i\top}U_j^i - U_j^{i\top}\boldsymbol{z}_j^i\boldsymbol{h}_j^{i\top} + U_j^{i\top}U_j^i) \\
&\quad + \lambda\,\mathbf{Tr}(\boldsymbol{h}_j^i\boldsymbol{v}_j^{i\top}\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top} + \boldsymbol{h}_j^i\boldsymbol{v}_j^{i\top}S_j^i + S_j^{i\top}\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top} + S_k^{i\top}S_j^i)
\end{aligned}
\tag{17}
$$

Because $U_k^{i\top}U_j^i$ and $S_k^{i\top}S_j^i$ are constant with respect to $\boldsymbol{h}_j^i$, minimizing (17) is equivalent to minimizing the following

$$
\begin{aligned}
\boldsymbol{h}_j^{i*} &= \arg\min_{\boldsymbol{h}_j^i\geq 0} l^*(\boldsymbol{h}_j^i) \\
&= \arg\min_{\boldsymbol{h}_j^i\geq 0} \mathbf{Tr}(\boldsymbol{h}_j^i\boldsymbol{z}_j^{i\top}\boldsymbol{z}_j^i\boldsymbol{h}_j^{i\top}) - 2\,\mathbf{Tr}(\boldsymbol{h}_j^i\boldsymbol{z}_j^{i\top}U_j^i) + \lambda\,\mathbf{Tr}(\boldsymbol{h}_j^i\boldsymbol{v}_j^{i\top}\boldsymbol{v}_j^i\boldsymbol{h}_j^{i\top}) + 2\lambda\,\mathbf{Tr}(\boldsymbol{h}_j^i\boldsymbol{v}_j^{i\top}S_j^i)
\end{aligned}
\tag{18}
$$

Based on the property of matrix trace, we have the following,
$\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{z}_j^{i\top} \boldsymbol{z}_j^i \boldsymbol{h}_j^{i\top}) = \mathbf{Tr}(\boldsymbol{z}_j^{i\top} \boldsymbol{z}_j^i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) = \boldsymbol{z}_j^{i\top} \boldsymbol{z}_j^i \boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i$
$\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{z}_j^{i\top} U_j^i) = \mathbf{Tr}(U_j^{i\top} \boldsymbol{z}_j^i \boldsymbol{h}_j^{i\top}) = (U_j^{i\top} \boldsymbol{z}_j^i)^\top \boldsymbol{h}_j^i$
$\mathbf{Tr}(\boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top} S_j^i) = \mathbf{Tr}(S_j^{i\top} \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top}) = (S_j^{i\top} \boldsymbol{v}_j^i)^\top \boldsymbol{h}_j^i$
Hence,

$$\boldsymbol{h}_j^{i*} = \underset{\boldsymbol{h}_j^i \geq 0}{\arg\min} (\boldsymbol{z}_j^{i\top} \boldsymbol{z}_j^i)(\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) + \lambda (\boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i)(\boldsymbol{h}_j^{i\top} \boldsymbol{h}_j^i) - 2(U_j^{i\top} \boldsymbol{z}_j^i)^\top \boldsymbol{h}_j^i + 2\lambda (S_j^{i\top} \boldsymbol{v}_j^i)^\top \boldsymbol{h}_j^i \tag{19}$$

By setting $\frac{\partial l^*(\boldsymbol{h}_j^i)}{\partial \boldsymbol{h}_j^i} = \mathbf{0}$ we can solve for $\boldsymbol{h}_j^{i*}$, as shown below.

$$\frac{\partial l^*(\boldsymbol{h}_j^i)}{\partial \boldsymbol{h}_j^i} = 2(\boldsymbol{z}_j^{i\top} \boldsymbol{z}_j^i)\boldsymbol{h}_j^i + 2\lambda (\boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i)\boldsymbol{h}_j^i - 2U_j^{i\top} \boldsymbol{z}_j^i + 2\lambda S_j^{i\top} \boldsymbol{v}_j^i = \mathbf{0} \tag{20}$$

$$\begin{aligned}
\boldsymbol{h}_j^{i*} &= \left[\frac{U_j^{i\top} \boldsymbol{z}_j^i - \lambda S_j^{i\top} \boldsymbol{v}_j^i}{\boldsymbol{z}_j^{i\top} \boldsymbol{z}_j^i + \lambda \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i} \right]_+ \\
&= \left[\frac{(X^{i\top} - H^i(W + V^i)^\top + \boldsymbol{h}_j^i (\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top)(\boldsymbol{w}_j + \boldsymbol{v}_j^i) - \lambda (H^i V^{i\top} - \boldsymbol{h}_j^i \boldsymbol{v}_j^{i\top}) \boldsymbol{v}_j^i}{(\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top (\boldsymbol{w}_j + \boldsymbol{v}_j^i) + \lambda \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i} \right]_+ \\
&= \left[\boldsymbol{h}_j^i + \frac{X^{i\top} (\boldsymbol{w}_j + \boldsymbol{v}_j^i) - H^i (W + V^i)^\top (\boldsymbol{w}_j + \boldsymbol{v}_j^i) - \lambda H^i V^{i\top} \boldsymbol{v}_j^i}{(\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top (\boldsymbol{w}_j + \boldsymbol{v}_j^i) + \lambda \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i} \right]_+ \\
H_{\cdot j}^{i*} &= \left[H_{\cdot j}^i + \frac{X^{i\top} (W_{\cdot j} + V_{\cdot j}^i) - H^i (W + V^i)^\top (W_{\cdot j} + V_{\cdot j}^i) - \lambda H^i V^{i\top} V_{\cdot j}^i}{[(W + V^i)^\top (W + V^i)]_{jj} + \lambda (V^{i\top} V^i)_{jj}} \right]_+
\end{aligned} \tag{21}$$

6 Appendix

Equivalently, we have

$$\begin{aligned}
\boldsymbol{h}_j^{i\top *} &= \left[\frac{\boldsymbol{z}_j^{i\top} U_j^i - \lambda \boldsymbol{v}_j^{i\top} S_j^i}{\boldsymbol{z}_j^{i\top} \boldsymbol{z}_j^i + \lambda \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i} \right]_+ \\
&= \left[\frac{(\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top (X^i - (W + V^i) H^{i\top} + (\boldsymbol{w}_j + \boldsymbol{v}_j^i) \boldsymbol{h}_j^{i\top}) - \lambda \boldsymbol{v}_j^{i\top} (V^i H^{i\top} - \boldsymbol{v}_j^i \boldsymbol{h}_j^{i\top})}{(\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top (\boldsymbol{w}_j + \boldsymbol{v}_j^i) + \lambda \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i} \right]_+ \\
&= \left[\boldsymbol{h}_j^{i\top} + \frac{(\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top X^i - (\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top (W + V^i) H^{i\top} - \lambda \boldsymbol{v}_j^{i\top} V^i H^{i\top}}{(\boldsymbol{w}_j + \boldsymbol{v}_j^i)^\top (\boldsymbol{w}_j + \boldsymbol{v}_j^i) + \lambda \boldsymbol{v}_j^{i\top} \boldsymbol{v}_j^i} \right]_+ \\
(H^{i*\top})_{j\cdot} &= \left[(H^{i\top})_{j\cdot} + \frac{(W_{\cdot j} + V_{\cdot j}^i)^\top X^i - (W_{\cdot j} + V_{\cdot j}^i)^\top (W + V^i) H^{i\top} - \lambda V_{\cdot j}^{i\top} V^i H^{i\top}}{[(W + V^i)^\top (W + V^i)]_{jj} + \lambda (V^{i\top} V^i)_{jj}} \right]_+
\end{aligned} \tag{22}$$