scMEDAL for the interpretable analysis of single-cell transcriptomics data with batch effect visualization using a deep mixed effects autoencoder

Aixa X. Andrade[1], Son Nguyen[1] and Albert Montillo[1]*
[1]Lyda Hill Department of Bioinformatics
University of Texas Southwestern Medical Center
Dallas, TX 75390, USA
*Corresponding author: Albert Montillo[1], Ph.D.; Email: albert.montillo@utsouthwestern.edu

Supplementary information

Table of contents

1. Supplementary materials

Table S1 shows the details of the datasets used to evaluate scMEDAL subnetworks as well as the baseline PCA and the AE and AEC models. Data processing steps are described in supplemental section 2.2

Table S1. Datasets used to evaluate scMEDAL models.

	Healthy Heart[1,2]	*ASD*[3]	*AML*[4]
Description	Heart cells from multiple tissues	ASD and TD (controls)	AML and healthy subjects
Data type	Single cell	Single nucleus	Single cell
Total cells (pre-filtering)	486,134	104, 559	41,090
Total cells (post-filtering)	486,134	104, 559	38,417
Total genes	33,538	36,501	27,899
Highly variable genes	3,000	2,916	2,916
Number of cell types	12 + Not Assigned cell type	17	6 malignant + 15 healthy
Batch effect type	Technical batch (z =147)	Biological donor (z = 31)	Technical batch (z=19)

2. Supplementary methods

2.1. Supplementary figures of the scMEDAL architecture

Random effects *Bayesian* layer block

Fig. S1 depicts the architecture of the Random Effects Dense (REDEN) Bayesian layer block, which is used to construct the Bayesian layers of the scMEDAL-RE subnetwork.

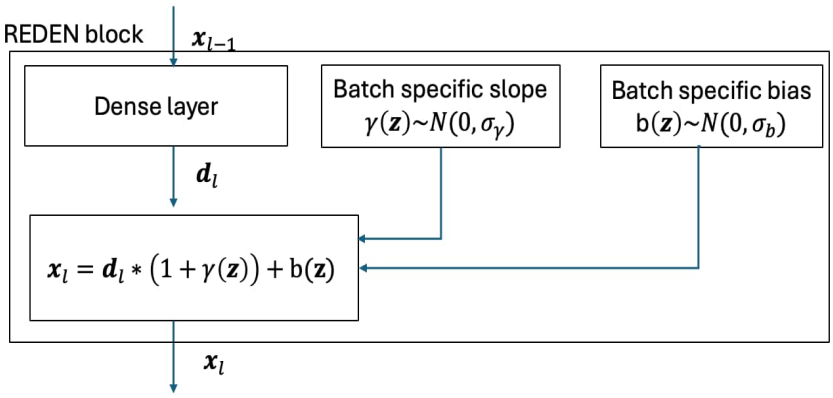

Fig. S1: Random effects dense (REDEN) block used in the random effects autoencoder (scMEDAL-RE). A dense layer with n_l neurons transforms an input representation from the preceding layer, x_{l-1}, into an intermediate output tensor, d_l. A batch-specific scale *γ(**z**)* and bias *b(**z**)* are then applied to each of the *f* feature maps. These batch-specific slopes and biases are regularized to follow normal distributions, with mean zero, and learned variances σ_γ and σ_b, respectively.

2.2. Data preprocessing

We used standard preprocessing steps from Yu et al. , 2023[5] which included: 1) filtering out cells exhibiting expression in fewer than ten genes and removal of genes detected in fewer than three cells, 2) normalizing total Unique Molecular Identifiers (UMI) counts per cell to 10,000, 3) log-transforming the data to stabilize variance log (X + 1), and 4) selecting the top highly variable genes (HVGs).

Since the ASD and Healthy Heart datasets had previously undergone quality control, no additional cells were removed during the filtering step. For the AML dataset, additional quality control measures were implemented. Cells with undefined or missing cell type annotations were excluded. Furthermore, samples identified by Dai et al., (2021) as having ambiguous annotations—specifically AML314, AML371, AML722B, and AML997[6]—were removed from the original 16 AML donors to ensure the dataset's integrity, resulting in 12 AML donors available for analysis. Including 5 healthy donors and 2 cell lines, the dataset comprised a total of 19 batches. The ASD dataset had previously undergone a log2 transformation as part of the original preprocessing by van Galen et al. (2019). To maintain consistency across all datasets, this transformation was reversed by applying an exponential function (base 2) to the data after adding 1 to each value, thereby returning the data to its original scale. Following this reversal, the standard preprocessing pipeline was applied.

The datasets were then partitioned through 5 fold cross-validation, stratified by batch and cell type in all three datasets. For each split, three folds were used for training, one for validation, and one for testing. The data were loaded for each subset and scaled using min-max scaling before model fitting. To reduce compute time for the ASW calculations, we used a random sample of 10,000 cells when datasets exceeded this size in training, validation, or testing subsets.

2.3. Hyperparameters used to build the AE, AEC, scMEDAL-FE, scMEDAL-FEC, and scMEDAL-RE models

Section 4.5 of the main paper describes the hyperparameter optimization process for the overall objective function in equation 9. In the following **Tables** S2-S4, we describe the values of the hyperparameter that are found to optimize the AE, AEC, scMEDAL-FE, scMEDAL-FEC, and scMEDAL-RE models in three datasets: Healthy Heart, the Autism Spectrum Disorder (ASD), and Acute Myeloid Leukemia (AML).

Table S2. Hyperparameters used for training PCA, AE and scMEDAL subnetworks in the **Healthy Heart dataset**.

Hyperparameter	AE	AEC	scMEDAL-FE	scMEDAL-FEC	scMEDAL-RE
$\lambda_{recon,F}$	1	81	5400	9450	-
$\lambda_{recon,R}$	-	-	-	-	110
λ_{y}	-	0.1	-	1	-
λ_{A}	-	-	1	1	-
λ_{L}	-				0.1
λ_{K}	-				1.00E-05

Table S3. Hyperparameters used for training PCA, AE and scMEDAL subnetworks in the **ASD dataset**.

Hyperparameter	AE	AEC	scMEDAL-FE	scMEDAL-FEC	scMEDAL-RE
$\lambda_{recon,F}$	1	1	1000	1000	-
$\lambda_{recon,R}$	-	-	-	-	110
λ_{y}	-	0.1	-	1	-
λ_{A}	-		1	1	-
λ_{L}	-				0.1
λ_{K}	-				1.00E-05

Table S4. Hyperparameters used for training PCA, AE and scMEDAL subnetworks **AML dataset**.

Hyperparameter	AE	AEC	scMEDAL-FE	scMEDAL-FEC	scMEDAL-RE
$\lambda_{recon,F}$	1	100	4000	1500	-
$\lambda_{recon,R}$	-	-	-	-	110
λ_y	-	0.1	-	1	-
λ_A	-	-	1	1	-
λ_L	-				0.1
λ_K	-				1.00E-05

2.4. UMAP visualization parameters

Table S5 describes the minimum distance and nearest neighbor's choices for UMAP visualizations presented in the main paper in sections 2.2, 2.3, 2.4, and 2.7.

Table S5. UMAP Visualization parameters.

Dataset	UMAP parameters	AE	AEC	scMEDAL-FE	scMEDAL-FE	scMEDAL-RE
Healthy Heart	Min distance	0.5				0
	Nearest Neighbors	15				1000
ASD	Min distance	0.5				
	Nearest Neighbors	15				
AML	Min distance	0.5				
	Nearest Neighbors	15				

2.5. Hardware and software used for model implementation and training

All deep learning models were developed and trained on Nvidia Tesla V100 GPUs with 32 GB of memory and Tesla P4 GPUs with 8 GB of memory. The software stack for model training included Python 3.8.5, TensorFlow 2.3[7], TensorFlow Probability 0.11.1[7-9], scikit-learn 0.23.2, and Scanpy 1.6.0. Performance measures, including ASW, DB, CH, accuracy, balanced accuracy, and chance accuracy, were computed using the Scikit-learn library[10]. The PCA model was also implemented through the Scikit-learn library. ScRNA-seq preprocessing and UMAP computations[11] were performed using Python 3.8.18 with the Scanpy 1.9.8 library[12]. Genomaps were generated using genomap 1.3.6 and the Mann-Whitney U test was implemented using the using the *stats* module from SciPy 1.10.1[13] and the linear mixed-effects models were implemented in the Statsmodels 0.10.0 library[14]. For more details on the software used for model training, plot creation, and generation of visualizations, please refer to the code repository described in the *Code and data availability* section 5.

2.6. Clustering metric definitions

We use the following metrics to quantify the clustering (separability) of the cells: the Average Silhouette Width (ASW)[15], Calinski-Harabasz (CH) index[16], and Davies-Bouldin (DB) index[17]. The *Silhouette Width* quantifies how well each cell fits within its assigned cluster compared to the other clusters. It is computed using two distances: the within-cluster distance (a_i) and the between-cluster distance (b_i). The within-cluster distance, a_i, represents the average distance between a cell i and all other cells in its cluster. This distance is defined as:

$$a_i = \frac{1}{|C_i| - 1} \sum_{\{j \in C_i, j \neq i\}} d(i,j)$$

where C_i is the set of cells in the cluster of cell i, $|C_i|$ is the total number of cells in this cluster, and $d(i,j)$ is the distance between cell i and cell j within the cluster. In contrast, the between-cluster distance, b_i, is the minimum of the average distance between cell i and cells from other clusters. It is defined as:

$$b_i = min_{\{k \neq i\}} \frac{1}{|C_k|} \sum_{\{j \in C_k\}} d(i,j)$$

where C_k is the set of cells assigned to cluster k which is different from the cluster of cell i, while $|C_k|$ is the total number of cells in cluster k. The silhouette coefficient for each cell i is calculated as:

$$s_i = \frac{b_i - a_i}{max(b_i, a_i)}$$

where a higher s_i value (closer to 1.0) indicates that the cell is well-clustered, being more similar to cells within its own cluster than to those in any other cluster. The *ASW* is defined as the average of s_i over all cells and provides an overall measure of clustering quality.

The *Davies–Bouldin* (DB) *index* quantifies the extent to which clusters overlap. For each cluster, we compare its average within-cluster dispersion to the distance between its centroid and the centroid of the most similar (closest) cluster. Consequently, a lower DB index value indicates less overlap and better separation between clusters. The DB index is computed as:

$$DB = \left(\frac{1}{K}\right) \sum_{\{k=1\}}^{K} D_k$$

where K is the number of clusters, and D_k for cluster k is the ratio, R_{kl} for the most similar cluster l to k, $(l \neq k)$. D_k is expressed as:

$$D_k = max_{\{l \neq k\}} R_{kl}$$

where the ratio R_{kl} measures how similar are the clusters k and l. R_{kl} is computed as:

$$R_{kl} = \frac{(q_k + q_l)}{M_{kl}}$$

where q_k is the is the average distance from each cell in cluster k to the centroid of cluster k, q_l is the corresponding average distance for cluster l, and M_{kl} is the distance between the centroids of clusters k and l. A larger M_{kl} indicates better separation between the two clusters.

The *Calinski-Harabasz* (CH) index measures the ratio of between-cluster variance to within-cluster variance, with higher values indicating more distinct and cohesive clusters. It is calculated as

$$CH = \frac{\sum_{\{k=1\}}^{K} n_k * ||\mu_k - \mu||^2}{\sum_{\{k=1\}}^{K} \sum_{\{i=1\}}^{n_k} ||p_i - \mu_k||^2} * \left[\frac{N-K}{K-1}\right]$$

where n_k is the number of points in the k^{th} cluster, μ_k is the centroid of the k^{th} cluster, μ is the centroid of the entire data (all the cells), p_i is the position (coordinates) of the cell i, N is the total number of cells and K is the number of clusters.

To align the interpretation of the DB index with the CH index and ASW, we used the reciprocal (1/DB), ensuring that higher scores consistently indicate better clustering quality. When measuring the clustering by cell type, higher scores signify improved separability in the latent space, enhancing the cell type signal—an objective of batch correction. Conversely, when measuring the clustering by batch label, lower scores indicate more effective batch correction, as it indicates that the batches become less distinguishable, which is another objective of batch correction. Additionally, higher batch scores in the batch-specific latent space of the random effects subnetwork indicates effective modeling of the batch effects and capture of the between-batch variance in gene expression data.

3. Supplementary results

3.1. Confounding within the AML dataset

In **Fig**. S2 the representation of donors across cell types in AML is shown graphically. We observe that not all donor samples contain cells from every cell type, making this dataset a particularly challenging and interesting one for batch correction due to potential confounding between donor identity and cell type.

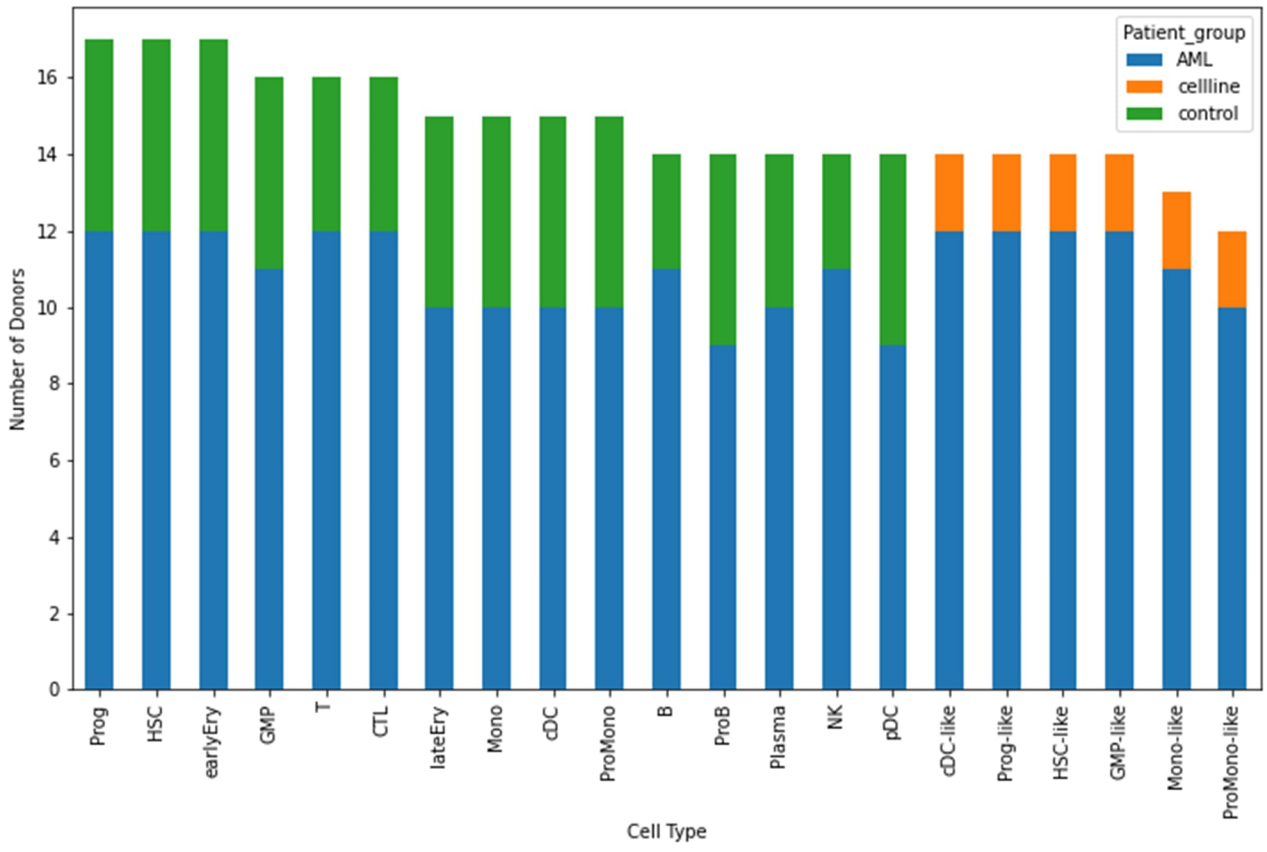

Fig. S2: AML dataset demonstrates intrinsec cell type-donor confounding. This stacked bar plot shows the number of donors for each cell type (y-axis), grouped by patient group, with cell types displayed along the x-axis. The colors indicate the donor patient group.

3.2. Hyperparameter Optimization (HPO) for AEC and scMEDAL-FEC models

Hyperparameter optimization (HPO) was conducted per dataset to tune the parameters using the held-out validation data (not used for model training). This tuning adjusts the weights of the reconstruction, adversarial, classification (cell type for scMEDAL-FE and batch for scMEDAL-RE), and Kullback–Leibler divergence loss function terms so that they are on a similar scale, allowing all terms to guide the model fitting, and optimizing performance on the validation data. The winning parameters are shown in **Tables** S2-S4. However, when cell type information is available from an external source (as described in Section 2.7 of the main paper) this can be used to further guide hyperparameter tuning. In this scenario, the ASW for cell type can be used to further refine the loss function weights. Both batch and cell-type ASW on the AEC and scMEDAL-FEC models can be evaluated during the training for 500 epochs with early stopping, prioritizing the cell-type ASW. **Table** S6 illustrates how the ASW scores for both batch and cell type separability can vary somewhat for different values of the weights assigned to the reconstruction loss (measured with mean squared error), adversarial loss, and classification loss of the cell type labels y, while the selected (winning) parameters are in boldface.

Table S6. Average Silhouette Width (ASW) scores for batch and cell type separability of the AML dataset (mean across 5 folds). Hyperparameter Optimization (HPO) selection of reconstruction loss λ_{recon} and classification loss λ_y weights using validation data. Adversarial loss weight λ_A = 1 for all models.

	λ_{recon}	λ_y	ASW (batch)			ASW (cell type)		
			mean	95% CI		mean	95% CI	
AEC	**1**	**0.1**	-0.43	-0.52	-0.34	-0.15	-0.23	-0.06
	10		-0.37	-0.44	-0.31	-0.11	-0.20	-0.02
	100		**-0.28**	**-0.34**	**-0.23**	**0.03**	**-0.04**	**0.10**
scMEDAL-FEC	500	1	-0.33	-0.35	-0.31	-0.10	-0.13	-0.06
	1500	**1**	**-0.30**	**-0.37**	**-0.24**	**-0.05**	**-0.08**	**-0.02**
	1500	2	-0.33	-0.40	-0.26	-0.07	-0.11	-0.04
	2000	2	-0.31	-0.33	-0.28	-0.07	-0.13	-0.01

3.3. Supplementary results for scMEDAL-FE and scMEDAL-RE models

3.3.1. CH and 1/DB scores

In **Tables** S7-S9, scores are shown from the Calinski-Harabasz (CH) clustering index and the reciprocal of the Davies-Bouldin (DB) clustering index (1/DB), as these complement the primary metric, ASW, used in the main paper. These scores confirm the ASW results. In particular, the scMEDAL-RE subnetwork consistently captures batch specific information as designed, demonstrating significantly more batch information (higher scores) than any of the other models, as measured with both the CH and 1/DB indices, across all three datasets (4th row, first 2 columns of **Tables** S7-S9). Meanwhile, the scMEDAL-FE subnetwork substantially suppresses batch specific information, demonstrating much lower CH scores (less batch contamination) than the baseline model (PCA) in all three datasets (3rd row, 2nd column of **Tables** S7-S9). In terms of preserving cell type information, across all datasets, the scMEDAL-FE subnetwork outperforms the baseline model (PCA) and scMEDAL-RE (by design) with higher scores in the 1/DB metric (rows 1,3,4 of 3rd column) and performs comparably well to the AE model in the CH metric (rows 2,3 of 4th column of **Tables** S7-S9).

Table S7. 1/DB and CH scores (Mean and 95% CI across 5 folds) for batch and cell type separability in the latent spaces of the **Healthy Heart dataset**, using PCA, AE, scMEDAL-FE, and scMEDAL-RE models.

	batch						cell type					
	1/DB			CH			1/DB			CH		
	mean	95% CI		mean	95% CI		mean	95% CI		mean	95% CI	
PCA (Baseline)	0.02	0.01	0.03	148.65	145.33	151.98	0.13	0.10	0.15	1502.74	1456.79	1548.69
AE	0.02	0.02	0.02	70.49	51.09	89.90	0.22	0.10	0.33	1712.76	1405.70	2019.83
scMEDAL-FE	0.02	0.01	0.02	61.38	38.74	84.03	0.19	0.13	0.24	1984.01	1078.08	2889.95
scMEDAL-RE	0.56	0.17	0.94	2555.27	544.02	4566.51	0.06	0.04	0.08	224.76	66.19	383.34

Table S8. 1/DB and CH scores (Mean and 95% CI across 5 folds) for batch and cell type separability in the latent spaces of the **ASD dataset**, using PCA, AE, scMEDAL-FE, and scMEDAL-RE models.

	batch						cell type					
	1/DB			CH			1/DB			CH		
	mean	95% CI		mean	95% CI		mean	95% CI		mean	95% CI	
PCA (Baseline)	0.03	0.02	0.03	32.67	32.06	33.28	0.18	0.15	0.21	9089.94	8906.86	9273.02
AE	0.02	0.02	0.03	23.72	20.74	26.69	0.78	0.71	0.85	9863.14	8673.06	11053.22
scMEDAL-FE	0.02	0.02	0.03	19.43	14.12	24.74	0.32	0.15	0.49	9631.75	8052.25	11211.25
scMEDAL-RE	0.16	-0.09	0.40	2423.99	-3510.80	8358.77	0.03	0.01	0.04	749.41	-261.64	1760.45

Table S9. 1/DB and CH scores (Mean and 95% CI across 5 folds) for batch and cell type separability in the latent spaces of the **AML dataset**, using PCA, AE, scMEDAL-FE, and scMEDAL-RE models.

	batch						cell type					
	1/DB			CH			1/DB			CH		
	mean	95% CI		mean	95% CI		mean	95% CI		mean	95% CI	
PCA (Baseline)	0.06	0.05	0.07	156.52	154.20	158.84	0.14	0.12	0.16	2560.96	2482.14	2639.79
AE	0.07	0.07	0.08	189.78	169.23	210.32	0.28	0.24	0.32	1956.48	1779.39	2133.58
scMEDAL-FE	0.07	0.06	0.09	80.63	69.88	91.37	0.19	0.09	0.29	1493.43	1353.85	1633.01
scMEDAL-RE	0.37	0.07	0.67	63228.21	9245.17	117211.25	0.07	0.05	0.09	114.73	69.53	159.94

3.3.2. Random Effects subnetwork (scMEDAL-RE) latent spaces

Figs. S3–S5 show the scMEDAL-RE latent spaces through UMAP visualizations. We observe that the batches (colored) are visibly apparent across all datasets. This further confirms the high degree of batch separability attainable by the scMEDAL-RE subnetwork, which is as designed, because its purpose is to model the batch effects.

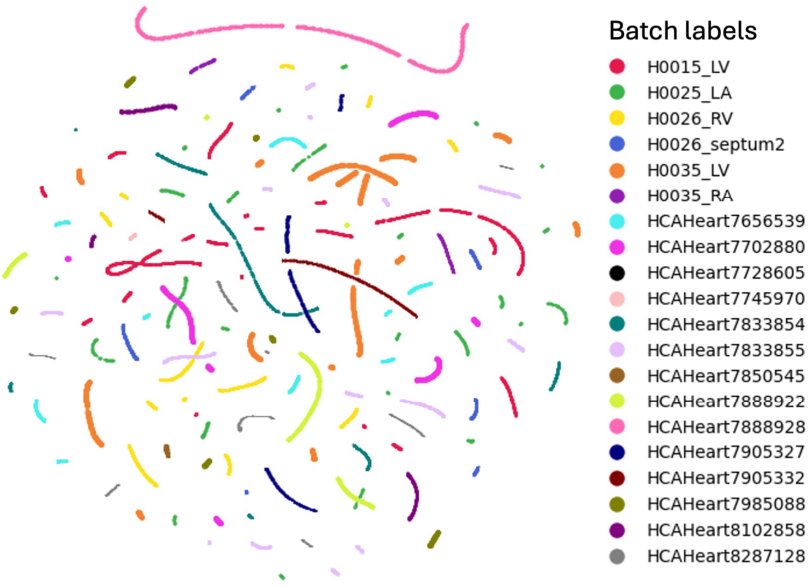

Fig. S3: scMEDAL-RE separates cells from different batches in the Healthy Heart dataset. UMAP visualization of scMEDAL-RE latent components, representing 44,987 cells from 20 selected batches (out of 147 total). Colors indicate different batches, demonstrating the separation of cells by batch in the scMEDAL-RE latent space.

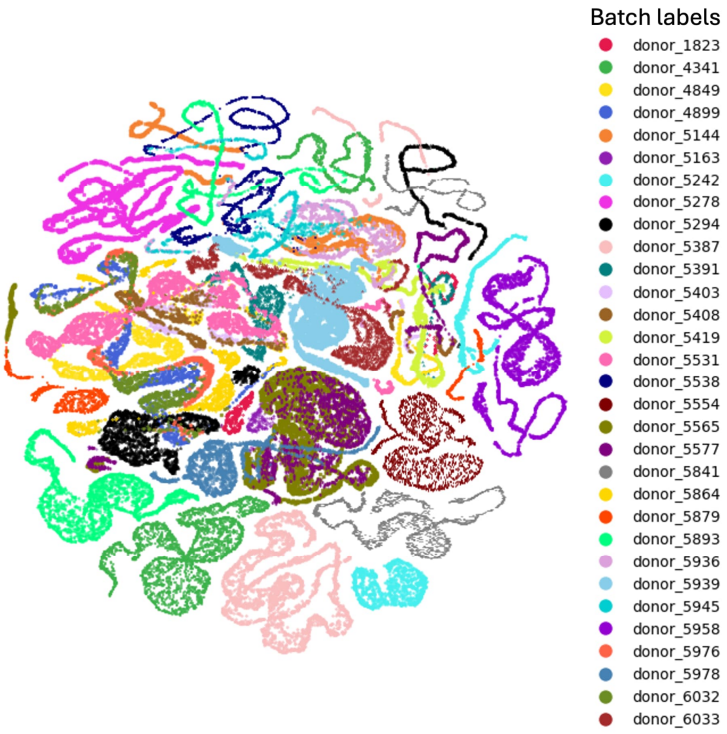

Fig. S4: scMEDAL-RE separates cells from different donors in the ASD dataset. UMAP visualization of scMEDAL-RE latent components, representing 62,735 cells from 31 donors. Colors indicate different donors, demonstrating the separation of cells by donors in the scMEDAL-RE latent space.

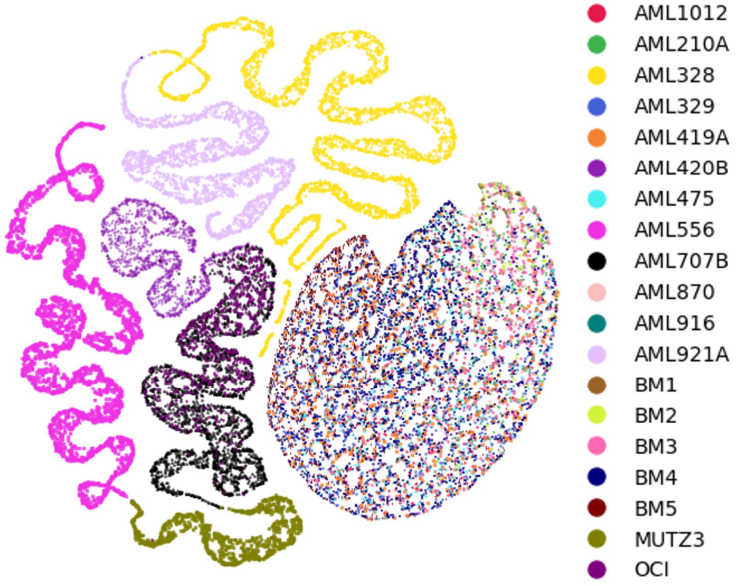

Fig. S5: scMEDAL-RE separates cells from cellines, AML and healthy donors in the AML dataset. Latent spaces from 23,050 cells from the AML dataset latent spaces obtained with scMEDAL-RE. UMAP applied to the scMEDAL-RE latent space.

3.3.3. Training and validation curves

Figs. S6-S8 present the training and validation curves, for the AE, scMEDAL-FE, and scMEDAL-RE models in all three datasets. For scMEDAL-FE, the total loss is calculated as the reconstruction loss minus the adversarial loss, with each component scaled by its respective weight. In the scMEDAL-RE curves, the left axis displays the reconstruction loss and total loss, while the right axis shows: the latent cluster loss classifying a cell's batch and the Kullback-Leibler divergence (KLD) loss. Also shown is the total loss, which is computed as the sum of the reconstruction loss, latent cluster loss, and KLD loss, each adjusted by their respective weights. We observe that through all models, no one term dominates. Instead the individual terms have values in similar ranges and can contribute effectively to the overall loss, thereby guiding the model fitting.

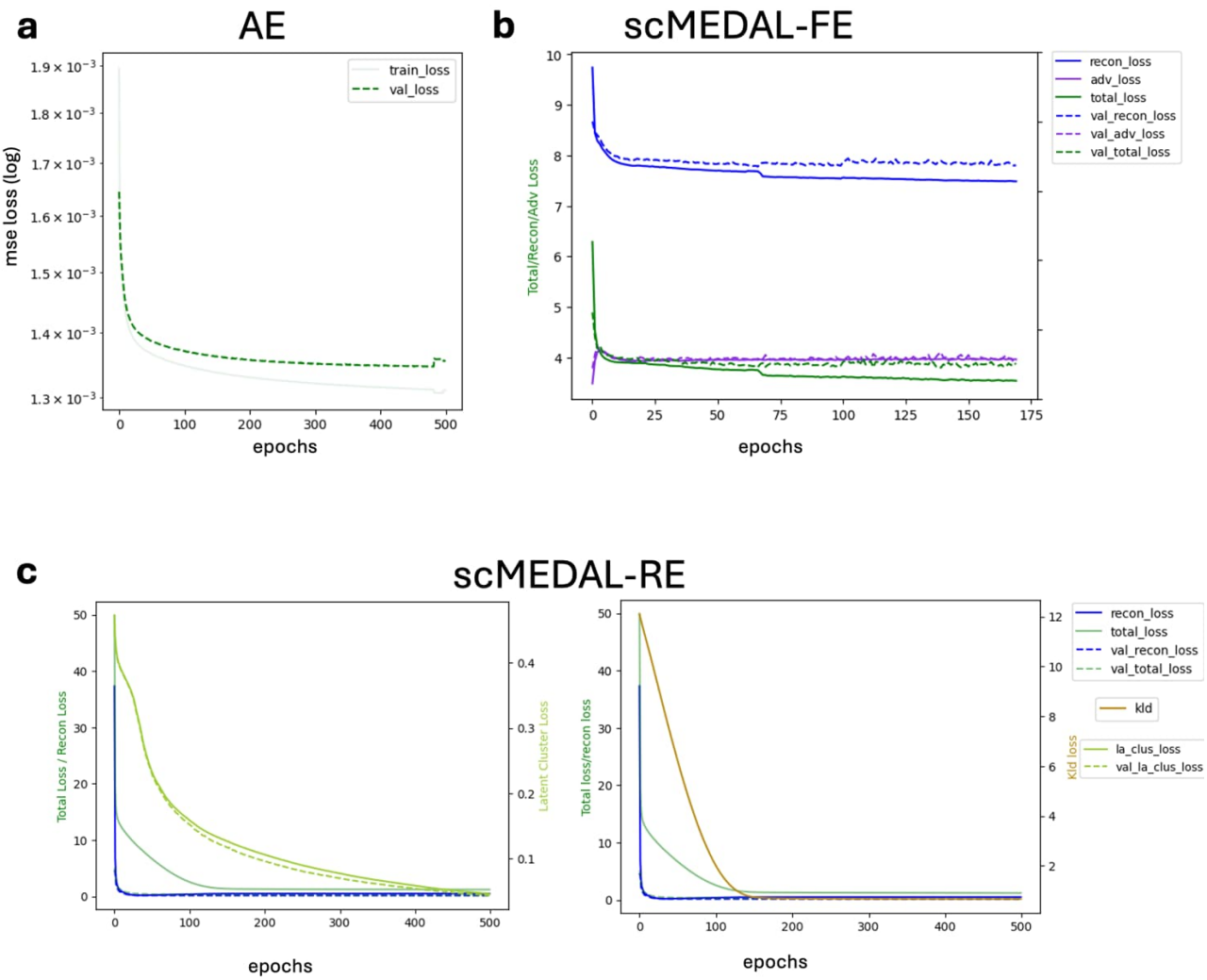

Fig. S6: Training and validation curves for the Healthy Heart dataset. Individual loss terms (see legend) maintain a similar scale throughout training. Training and validation curves adjusted by weights for **a** AE, **b** scMEDAL-FE (total loss = reconstruction loss - adversarial loss), and **c** scMEDAL-RE (total loss = reconstruction loss + KLD loss + latent cluster loss).

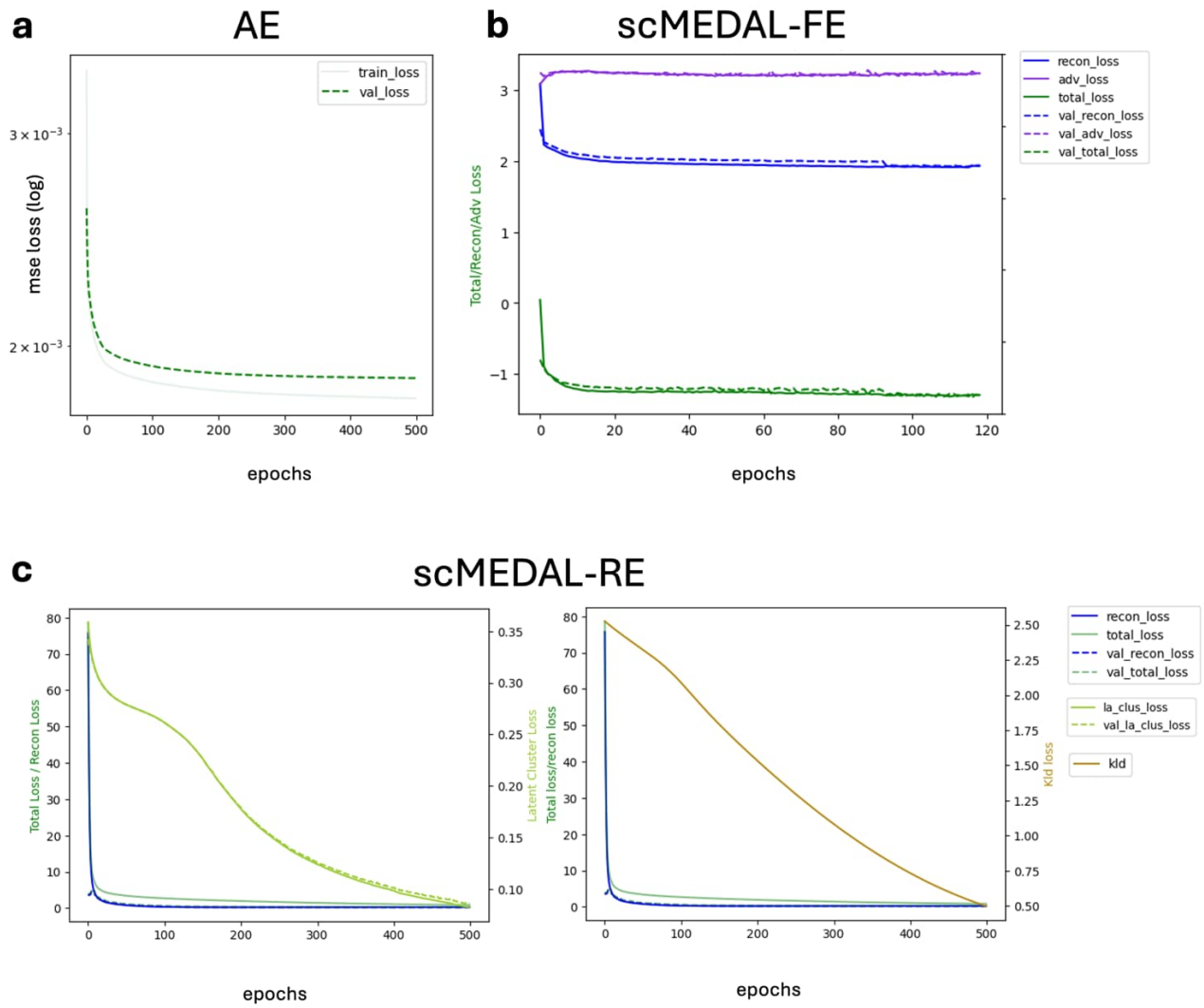

Fig. S7: Training and validation curves for the ASD dataset. Individual loss terms (see legend) maintain a similar scale throughout training. Training and validation curves adjusted by weights for **a** AE, **b** scMEDAL-FE (total loss = reconstruction loss - adversarial loss), and **c** scMEDAL-RE (total loss = reconstruction loss + KLD loss + latent cluster loss)

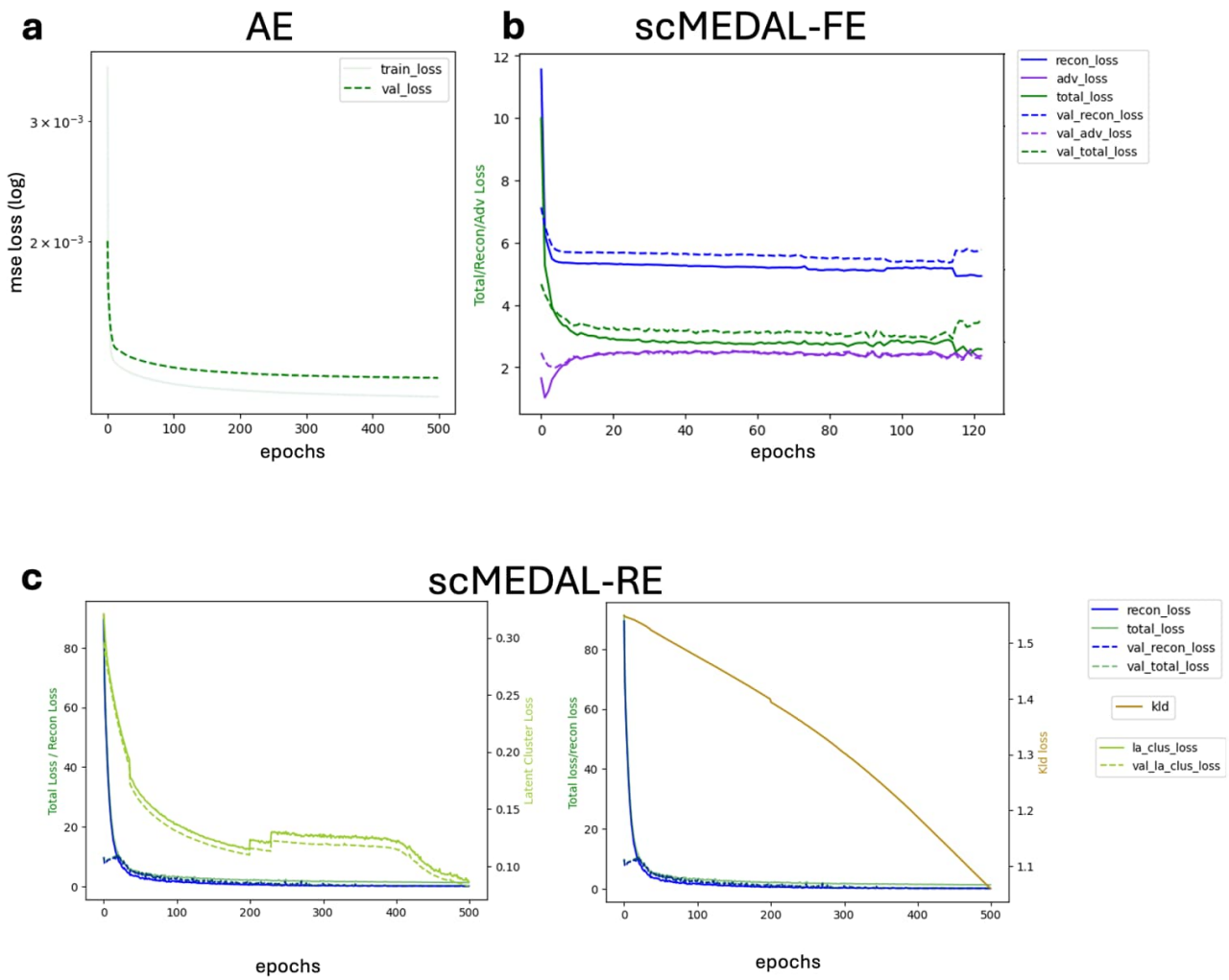

Fig. S8: Training and validation curves for the AML dataset. Individual loss terms (see legend) maintain a similar scale throughout training. Training and validation curves adjusted by weights for **a** AE, **b** scMEDAL-FE (total loss = reconstruction loss - adversarial loss), and **c** scMEDAL-RE (total loss = reconstruction loss + KLD loss + latent cluster loss)

3.4. Supplementary results for scMEDAL-FEC and AEC models

3.4.1. CH and 1/DB scores

In **Tables** S10-S12 scores that complement the primary metric, ASW, are shown, including the Calinski-Harabasz (CH) clustering index and the reciprocal of the Davies-Bouldin (DB) clustering index (1/DB). These results further corroborate the ASW results in Section 2.7 of the main paper. The scMEDAL-FEC subnetwork substantially suppresses batch specific information, demonstrating lower CH scores (less batch contamination) than the AEC in all three datasets (rows 1,2 of the 2nd column of **Tables** S10-S12). In terms of preserving cell type information, both models perform admirably and similarly, including in their CH scores for the ASD and AML datasets (4th column of **Tables** S11-S12) and the 1/DB score for the Healthy Heart dataset (3rd column of **Table** S10). As described in the main paper (Section 2.7) the UMAP latent space visualizations provide a more nuanced view including additional comparative performance benefits of scMEDAL-FEC over the AEC model across datasets.

Table S10. 1/DB and CH scores (Mean and 95% CI across 5 folds) for batch and cell type separability in the latent spaces of the **Healthy Heart dataset**, using AEC and scMEDAL-FEC models.

	batch						cell type					
	1/DB			CH			1/DB			CH		
	mean	95% CI		mean	95% CI		mean	95% CI		mean	95% CI	
AEC	0.02	0.01	0.03	48.29	15.26	81.32	0.23	0.06	0.41	5563.96	2833.93	8293.99
scMEDAL-FEC	0.02	0.01	0.02	41.39	27.01	55.77	0.21	0.04	0.38	3750.68	2318.16	5183.19

Table S11. 1/DB and CH scores (Mean and 95% CI across 5 folds) for batch and cell type separability in the latent spaces of the **ASD dataset**, using AEC and scMEDAL-FEC models.

	batch						cell type					
	1/DB			CH			1/DB			CH		
	mean	95% CI		mean	95% CI		mean	95% CI		mean	95% CI	
AEC	0.02	0.01	0.03	20.06	14.48	25.64	0.73	0.51	0.95	10631.01	5846.41	15415.61
scMEDAL-FEC	0.02	0.01	0.02	12.72	7.40	18.03	0.34	0.18	0.50	9911.04	8726.06	11096.02

Table S12. 1/DB and CH scores (Mean and 95% CI across 5 folds) for batch and cell type separability in the latent spaces of the **AML dataset**, using AEC and scMEDAL-FEC models.

	batch						cell type					
	1/DB			CH			1/DB			CH		
	mean	95% CI		mean	95% CI		mean	95% CI		mean	95% CI	
AEC	0.08	0.04	0.12	142.75	81.95	203.54	0.17	0.09	0.25	2159.15	1216.68	3101.61
scMEDAL-FEC	0.06	0.04	0.07	87.24	70.03	104.45	0.18	0.10	0.25	2065.85	1301.67	2830.04

3.4.2. Average training and validation curves for AEC and scMEDAL-FE across 5 folds in the Healthy Heart, ASD, and AML datasets

Figs. S9-S11 present the training and validation curves, for the AEC and scMEDAL-FEC models in the three datasets. For the AEC model, the total loss is calculated as the sum of the reconstruction loss and the cell type classification loss. For scMEDAL-FEC, the total loss is calculated as the sum of the reconstruction loss and the cell type classification loss minus the adversarial loss. We observe that in all cases the individual terms have similar value ranges and contribute effectively to the total loss.

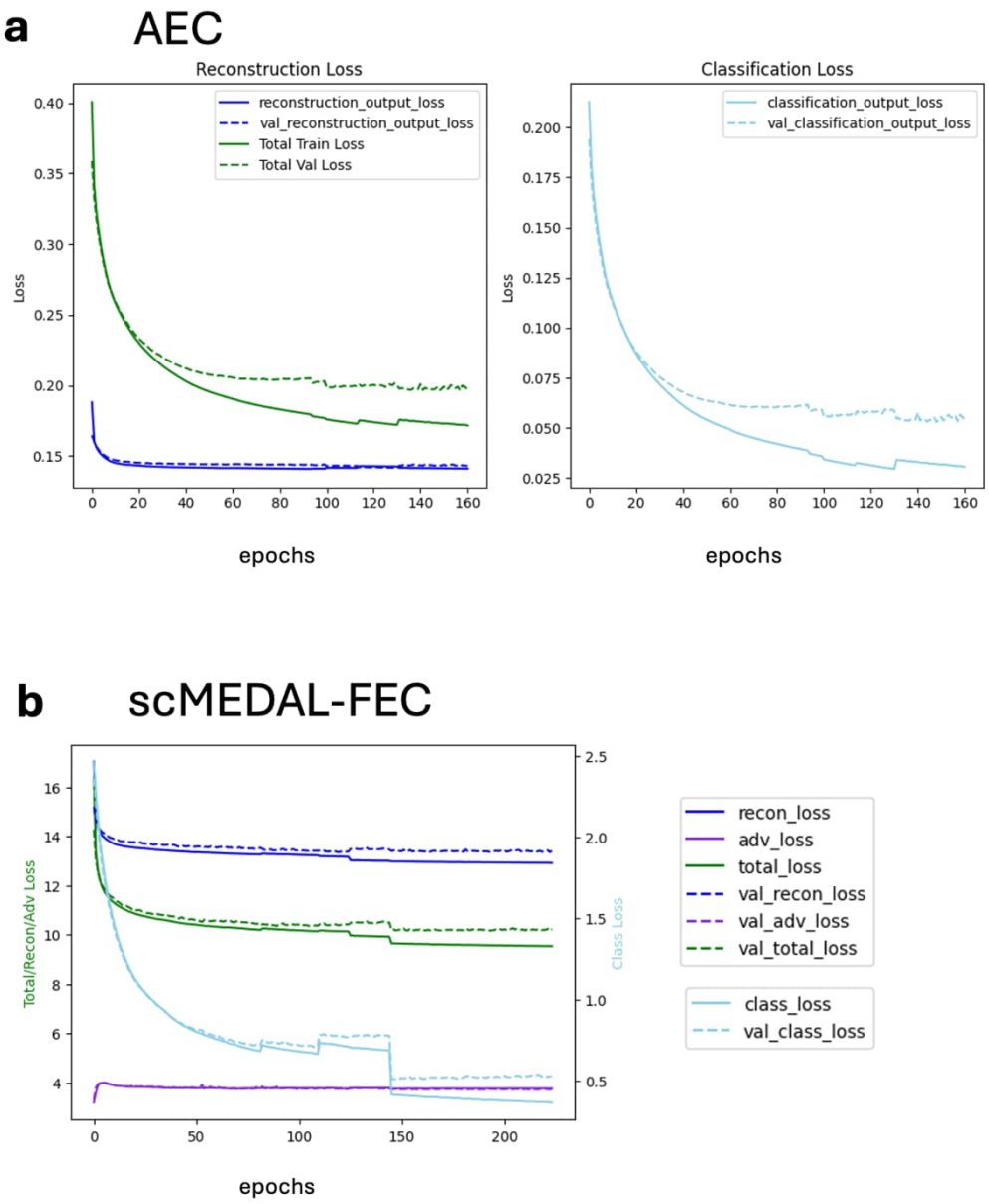

Fig. S9: The training and validation curves for the Healthy Heart dataset demonstrate that the weight-adjusted individual loss terms effectively balance their contributions to the total loss. Training and validation curves for the Healthy Heart dataset for **a** AEC and **b** scMEDAL-FEC models.

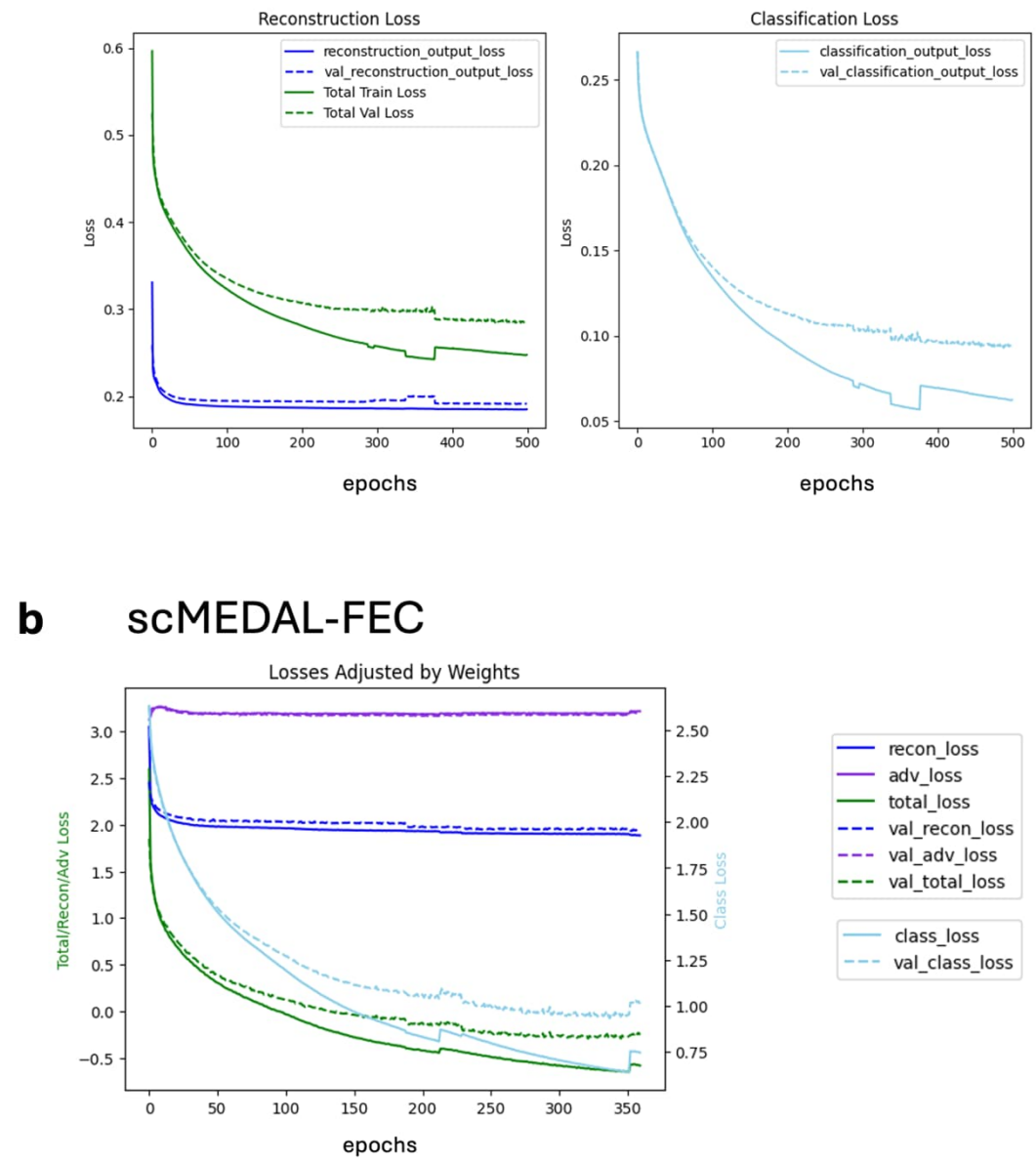

Fig. S10: The training and validation curves for the ASD dataset demonstrate that the weight-adjusted individual loss terms effectively balance their contributions to the total loss. Training and validation curves for the ASD dataset for **a** AEC and **b** scMEDAL-FEC models.

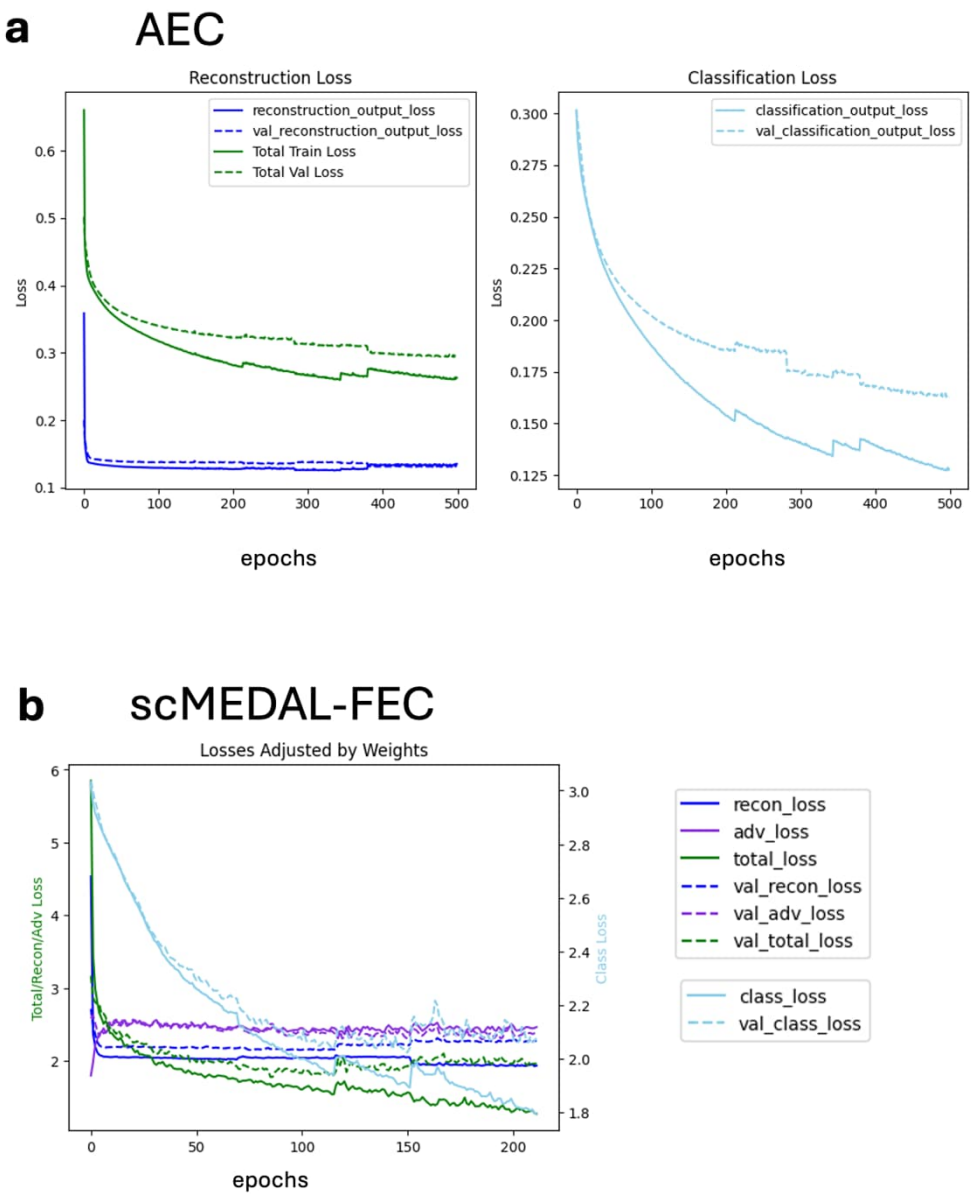

Fig. S11: The training and validation curves for the AML dataset demonstrate that the weight-adjusted individual loss terms effectively balance their contributions to the total loss. Training and validation curves for the AML dataset for **a** AEC and **b** scMEDAL-FEC models.

3.5. Significant genes from the genomaps for AML *versus* controls and ASD *versus* controls

Table S13 presents the statistics for significant genes related to ASD, identified using the scMEDAL-RE projection of the same 300 L2/3 cells across 15 ASD donors and then across 16 control (typically developing) donors. Statistical analysis of the gene expression between these groups was conducted using linear mixed-effects models implemented in Statsmodels[14]. A total of 230 genes were identified as

significantly associated with ASD ($p_{slope} < 0.05$). Statistics for these ASD-relevant genes are illustrated below. The full list of all ASD mapped genomap genes is available in Supplementary Data 1.

Table S13. Statistics for relevant genes associated with ASD. These genes are identified as significant through linear mixed-effects modeling comparing scMEDAL-RE projections of the same 300 L2/3 cells across 15 ASD and 16 control donors.

genes	pixel_i	pixel_j	intercept	p-val intercept	95% CI (intercept)		slope (control)	p-val slope	95% CI (slope)	
CNR1	1	1	0.3946	0.0048	0.1206	0.6686	-0.7758	0.0001	-1.1572	-0.3944
MYO1E	5	15	-0.2923	0.0752	-0.6144	0.0297	0.6037	0.0083	0.1554	1.052
CYP27A1	4	27	-0.2552	0.0801	-0.541	0.0306	0.5158	0.0111	0.118	0.9136
KCNJ10	48	15	-0.3308	0.1317	-0.761	0.0993	0.6454	0.0346	0.0466	1.2441
THBS1	17	34	-0.1595	0.0946	-0.3465	0.0275	0.2753	0.0382	0.015	0.5356
NOTCH3	5	34	-0.259	0.147	-0.6091	0.0911	0.5295	0.0332	0.0422	1.0167
IL1R1	40	15	-0.1641	0.1696	-0.3984	0.0701	0.3533	0.0337	0.0272	0.6793
COL4A1	17	3	-0.3247	0.0165	-0.5901	-0.0593	0.6232	0.0009	0.2538	0.9926
TGM2	10	11	-0.2724	0.1728	-0.664	0.1192	0.5951	0.0324	0.05	1.1402
FGF2	49	16	-0.3256	0.149	-0.7678	0.1167	0.6917	0.0277	0.0761	1.3072

Table S14 presents the statistics for significant genes related to AML. These are identified by a Mann-Whitney U test performed on the same set of 300 monocytes projected with scMEDAL-RE onto 12 AML donors and then onto 5 healthy control subjects. Averages from the AML subjects and the control (healthy) subjects were computed for each set of 300 monocytes, and *stats* module from SciPy[13]was used to conduct the Mann-Whitney U test to identify gene expression differences between these groups. In total, 358 genes were found to have significantly different expression levels ($p < 0.05$). The full list of all AML mapped genomap genes is available in Supplementary Data 1.

Table S14. Statistics for relevant genes associated with AML, identified as significant through a Mann-Whitney U test on the same 300 monocyte Genomap projections from 12 AML scMEDAL-RE analyses and 5 control scMEDAL-RE analyses.

genes	pixel_i	pixel_j	p-val
SH2B3	24	36	0.0023
HERPUD1	3	19	0.0194
PRDM1	2	23	0.0365
MKI67	11	27	0.0365
GATA2	45	46	0.0365
ABCB1	30	43	0.0485
SAMD9L	30	40	0.0194
SERPINA1	4	44	0.0013
SLC11A1	20	28	0.0094
VNN2	17	27	0.0365

4. References

1 Litvinukova, M., Talavera-Lopez, C., Maatz, H., Reichart, D., Worth, C. L., Lindberg, E. L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E. R., Roberts, K., Tuck, L., Fasouli, E. S., DeLaughter, D. M., McDonough, B., Wakimoto, H., Gorham, J. M., Samari, S., Mahbubani, K. T., Saeb-Parsy, K., Patone, G., Boyle, J. J., Zhang, H., Viveiros, A., Oudit, G. Y., Bayraktar, O. A., Seidman, J. G., Seidman, C. E., Noseda, M., Hubner, N. & Teichmann, S. A. Cells of the adult human heart. *Nature* **588**, 466-472 (2020).

2 Yu, X., Xu, X., Zhang, J. & Li, X. Batch alignment of single-cell transcriptomics data using deep metric learning. figshare https://doi.org/10.6084/m9.figshare.20499630.v2 (2023).

3 Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D. H. & Kriegstein, A. R. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).

4 van Galen, P., Hovestadt, V., Wadsworth Ii, M. H., Hughes, T. K., Griffin, G. K., Battaglia, S., Verga, J. A., Stephansky, J., Pastika, T. J., Lombardi Story, J., Pinkus, G. S., Pozdnyakova, O., Galinsky, I., Stone, R. M., Graubert, T. A., Shalek, A. K., Aster, J. C., Lane, A. A. & Bernstein, B. E. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281.e1224 (2019).

5 Yu, X., Xu, X., Zhang, J. & Li, X. Batch alignment of single-cell transcriptomics data using deep metric learning. *Nat. Commun.* **14**, 960 (2023).

6 Dai, C., Chen, M., Wang, C. & Hao, X. Deconvolution of Bulk Gene Expression Profiles with Single-Cell Transcriptomics to Develop a Cell Type Composition-Based Prognostic Model for Acute Myeloid Leukemia. *Front Cell Dev Biol* **9**, 762260 (2021).

7 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems. *Proc. of 12th USENIX Conference on Operating Systems Design and Implementation* (2016).

8 Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M. & Saurous, R. A. TensorFlow Distributions. arXiv:1711.10604, 2017.

9 TensorFlow. *TensorFlow Probability*, https://www.tensorflow.org/probability , last accessed 2/12/2025

10 Pedregosa, F., Eickenberg, M., Ciuciu, P., Thirion, B. & Gramfort, A. Data-driven HRF estimation for encoding and decoding models. *Neuroimage* **104**, 209-220 (2015).

11 McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://doi.org/10.48550/arXiv.1802.03426 (2018).

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

13 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, I., 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. & SciPy, C. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261-272 (2020).

14 Seabold, S. & Perktold, J. *Statsmodels: Econometric and Modeling with Python* in *Proc. 9th Python in Science Conference.* 57-61 (2010).

15 Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53-65 (1987).

16 Calinski, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat. Theory Methods* **3**, 1-27 (1974).

17 Davies, D. L. & Bouldin, D. W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 224-227 (1979).