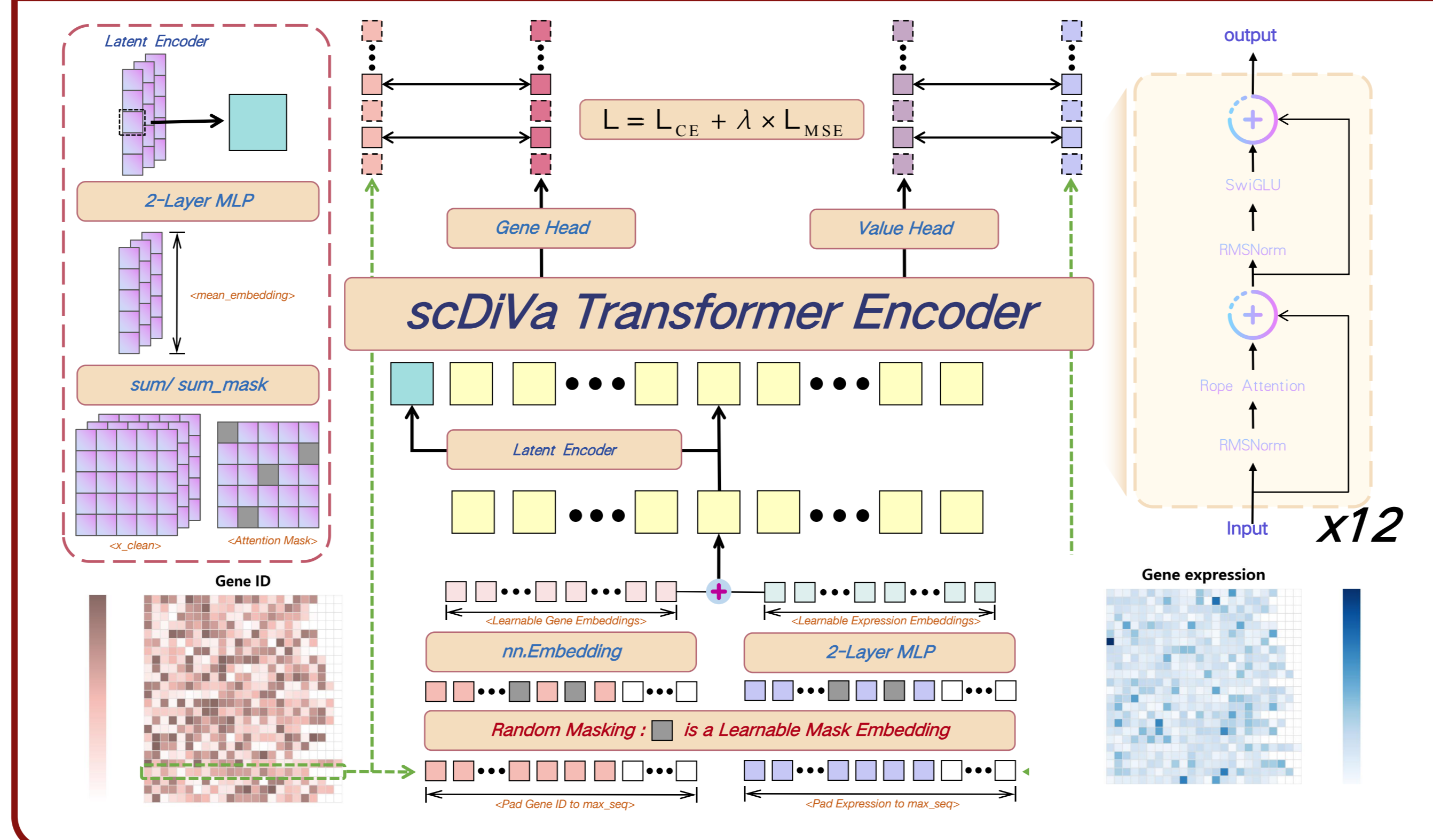


1. Introduction

Single-cell foundation models aim to learn transferable cellular representations from large-scale scRNA-seq data. However, unlike natural language, gene expression profiles do not follow a meaningful left-to-right order: genes within a cell form an unordered, sparse, and dropout-prone expression set. This makes standard autoregressive generation structurally mismatched with single-cell data. Instead of generating genes sequentially, scDiVa formulates single-cell modeling as dropout-aligned masked discrete diffusion, where corrupted gene tokens are recovered through bidirectional denoising.

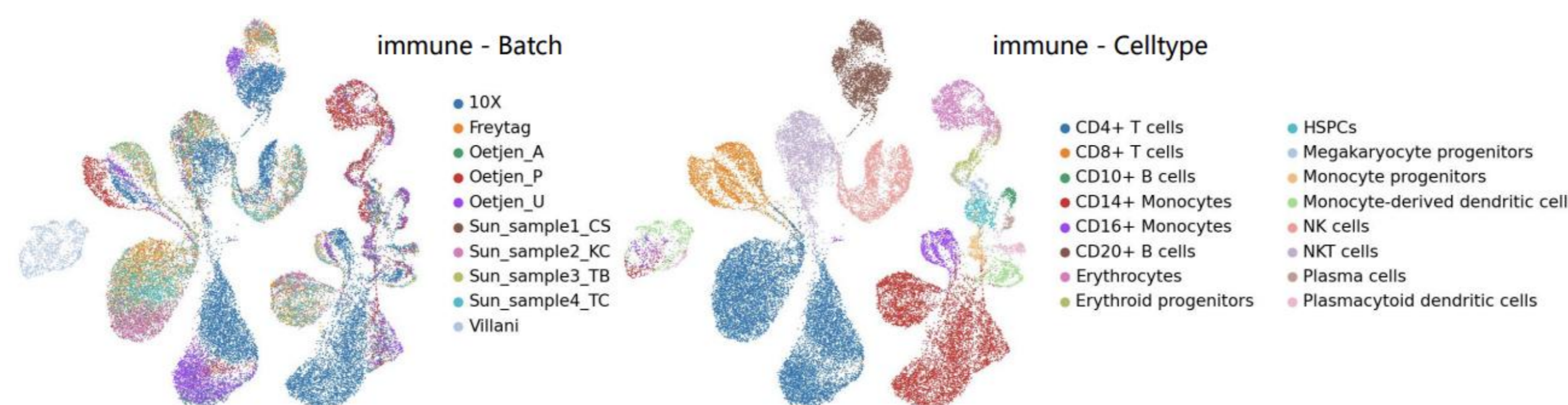
- **Single-cell RNA-seq profiles** are *sparse, unordered, and dropout-prone*, making them difficult to model with standard sequence-generation paradigms.
- **Autoregressive models** impose *artificial gene orders and accumulate errors*, while **Gaussian diffusion** *mismatches discrete gene-token corruption*.
- We propose **scDiVa**, a **masked discrete diffusion** model using *bidirectional denoising* to recover *gene identity and expression value* from *global cell context*.

2. Method: scDiVa

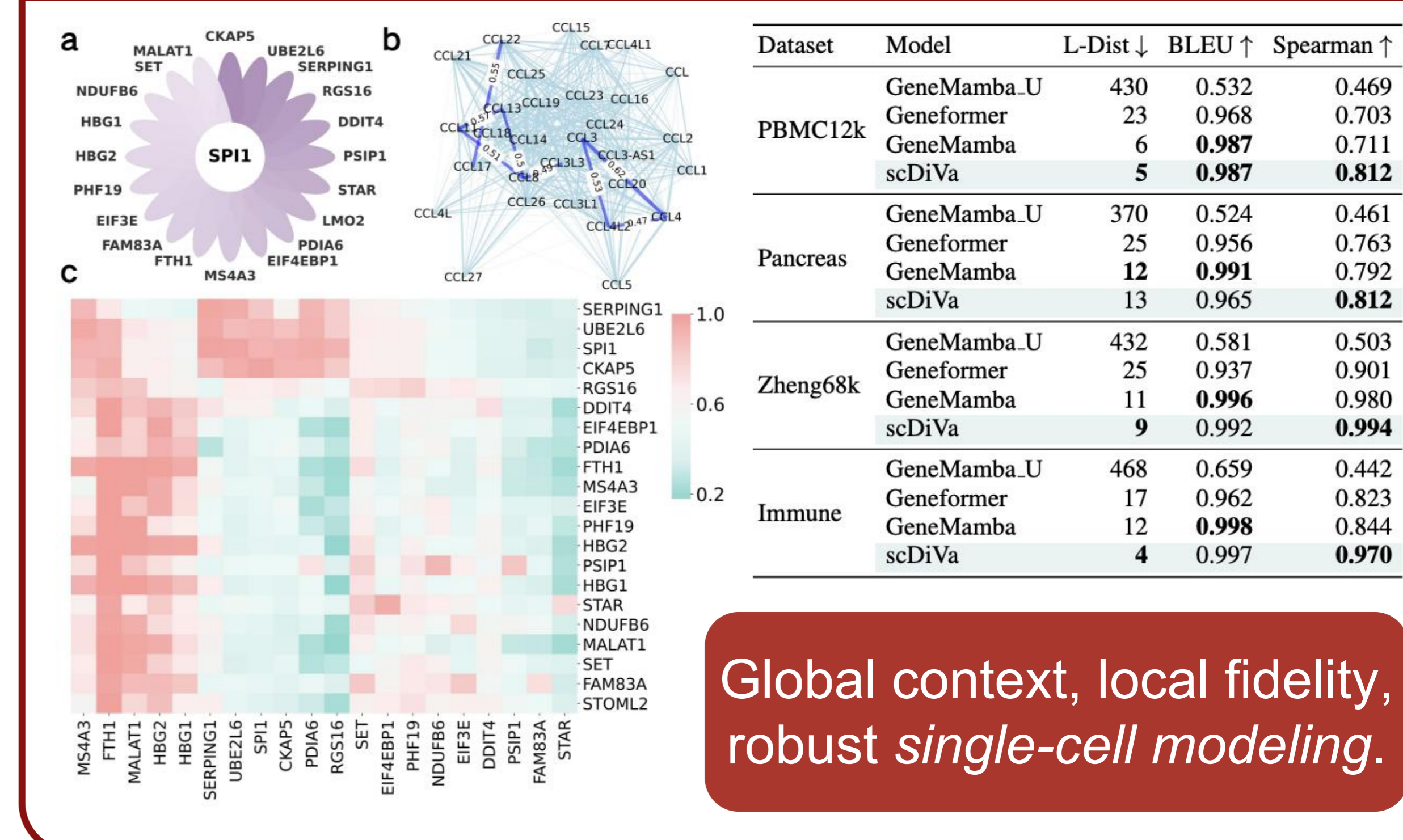
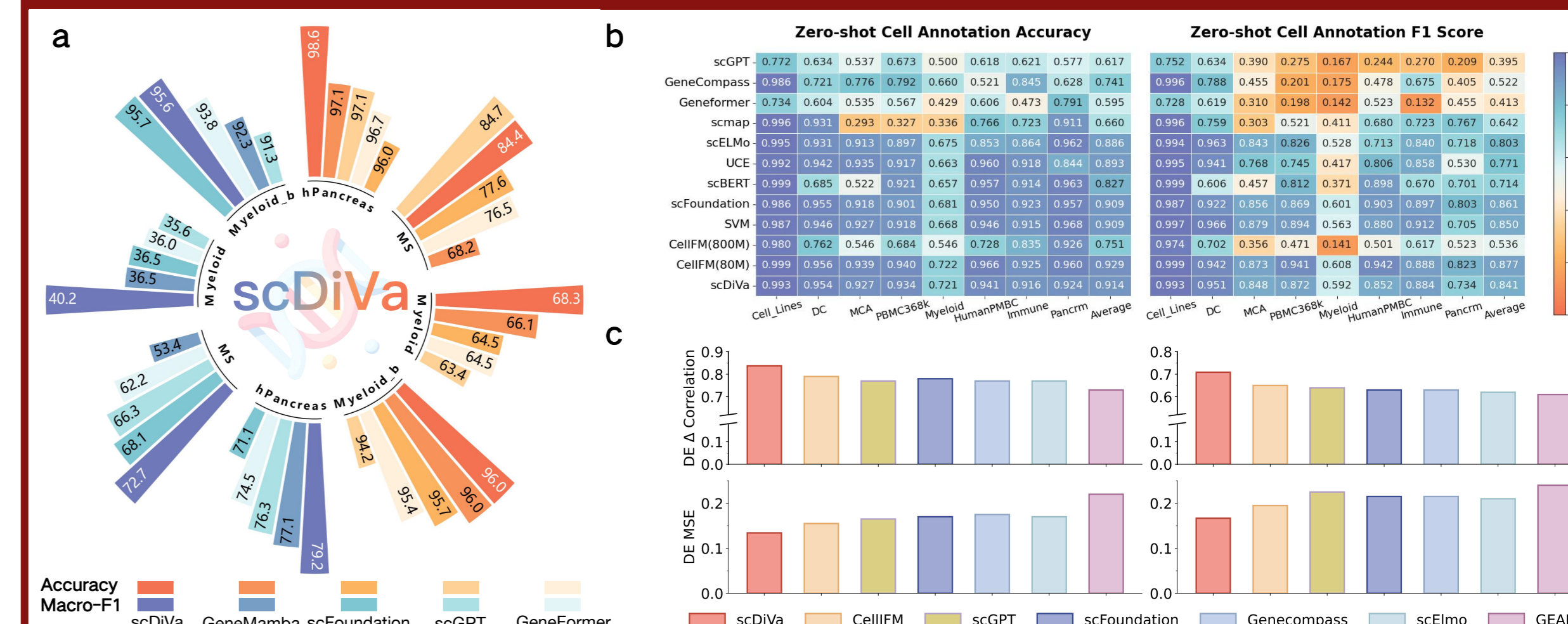


3. Main Results

Model	Immune		PBMC12k		BMBC		Perirhinal Cortex		COVID-19	
	Batch	Bio	Batch	Bio	Batch	Bio	Batch	Bio	Batch	Bio
Harmony	0.9514	0.6945	0.9341	0.7990	0.8999	0.6316	0.9442	0.8595	0.8781	0.4468
Geneformer	0.8153	0.6983	0.9545	0.7891	0.7720	0.6324	0.9127	0.8547	0.8240	0.5567
scGPT	0.9194	0.7879	0.9755	0.9018	0.8431	0.6576	0.9600	0.9552	0.8625	0.6476
scFoundation	0.8904	0.7337	0.9628	0.8662	0.7598	0.5250	0.9560	0.9606	0.8346	0.5468
GeneMamba	0.9536	0.8131	0.9604	0.8344	0.9157	0.7628	0.9673	0.9062	0.8742	0.5537
CellFM	0.9523	0.7934	0.9862	0.9741	0.9562	0.8014	0.9621	0.9692	0.9143	0.6415
UCE	0.9397	0.7482	0.9764	0.9323	0.9011	0.7225	0.9496	0.9281	0.8924	0.5916
scELMO	0.9342	0.7355	0.9726	0.9214	0.8893	0.7012	0.9445	0.9211	0.8842	0.5764
GeneCompass	0.9421	0.7684	0.9782	0.8963	0.9124	0.7581	0.9553	0.9442	0.9012	0.6093
scDiVa	0.9555	0.7785	0.9960	0.9566	0.9734	0.8712	0.9542	0.9895	0.9538	0.6689




3. Main Results



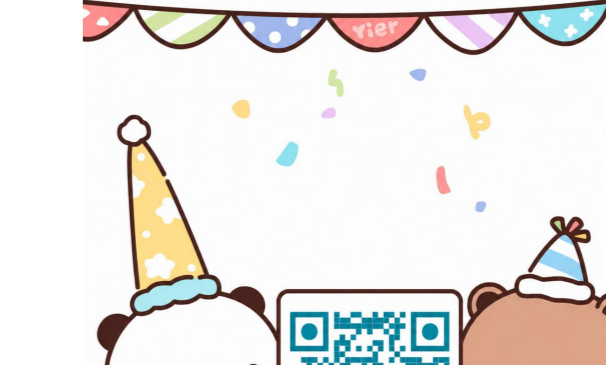
5. Resources

Code



please give me a star!

homepage



Paper

