Decomposed Representations Mitigate the Alignment–Specificity Trade-off in Multi-Omics

Thao M. Dang[1], Feng Jiang[1], Hehuan Ma[1], Yuzhi Guo[1], Jingquan Yan[1], Haiqing Li[1], Saiyang Na[1], Zheng Zheng[1], Thuc Anh Tran[2], Jean Gao[1], Junzhou Huang[1,*]

[1]The University of Texas at Arlington, TX, USA [2]The University of Sydney, NSW, Australia

NeurIPS 2026

MOTIVATION

The alignment–specificity trade-off

One cell, three views of the same identity

RNA
expression

ATAC
chromatin

ADT
surface protein

Alignment-based

e.g., scPairing, scMODAL

- ✓ Good cross-modal retrieval
- ✗ Suppresses modality-specific signal

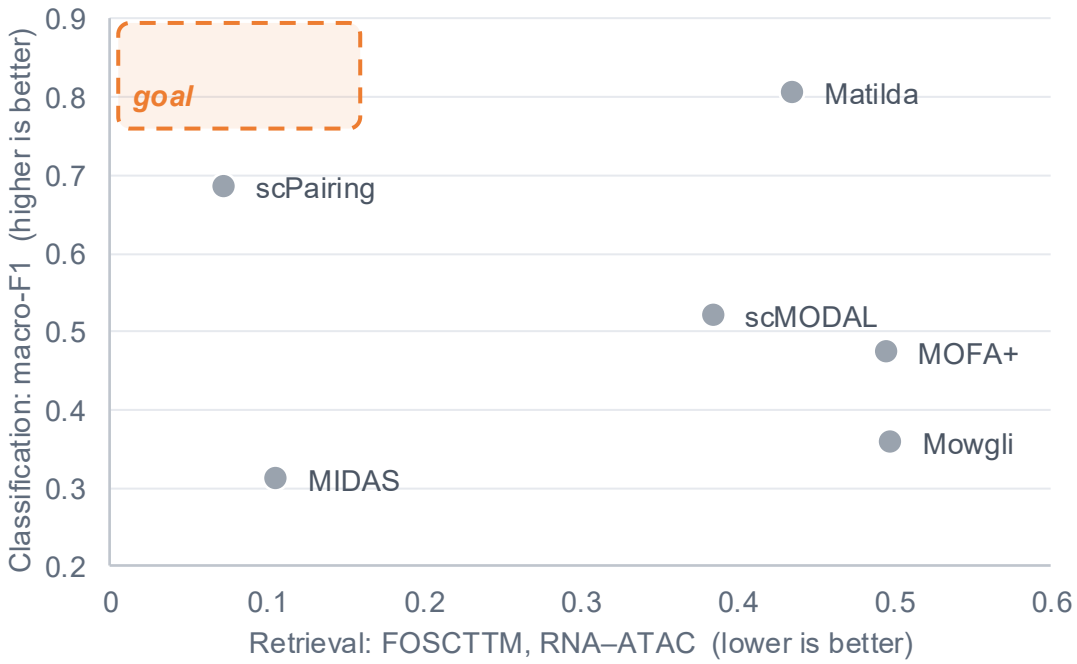
Fusion-based

e.g., Matilda, MIDAS

- ✓ Good cell-type classification
- ✗ Blurs cross-modal correspondence

The conflict is structural. The more one embedding agrees across modalities, the less non-shared, label-relevant signal it can keep (Props. 1–2).

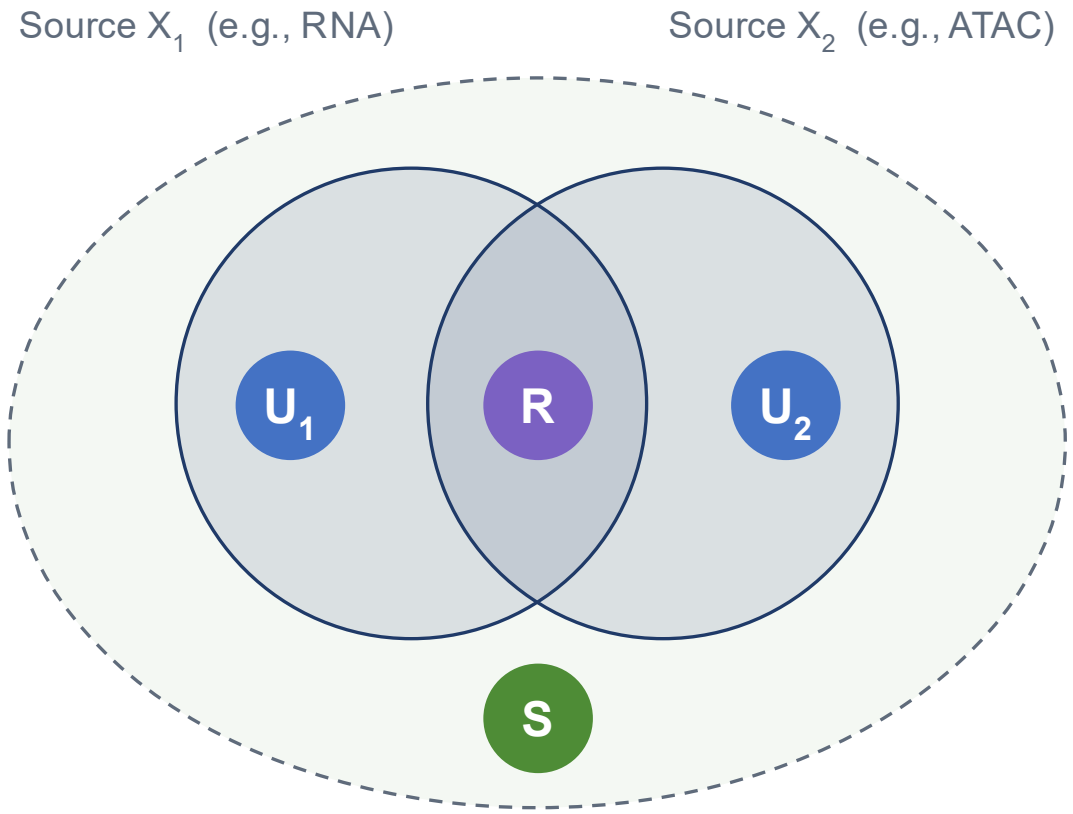
TEA-seq: no baseline does both



Baselines on TEA-seq (paper Tables 1–2). Macro-F1 from a linear probe.

KEY IDEA

Give each kind of signal its own branch



$$I(X_1, X_2; Y) = R + U_1 + U_2 + S$$

Partial information decomposition; Y = cell type

R **Redundancy** · shared by every modality

What alignment captures → cross-modal retrieval

U **Uniqueness** · in one modality only

What alignment discards → modality-specific, label-relevant detail

S **Synergy** · only in combination

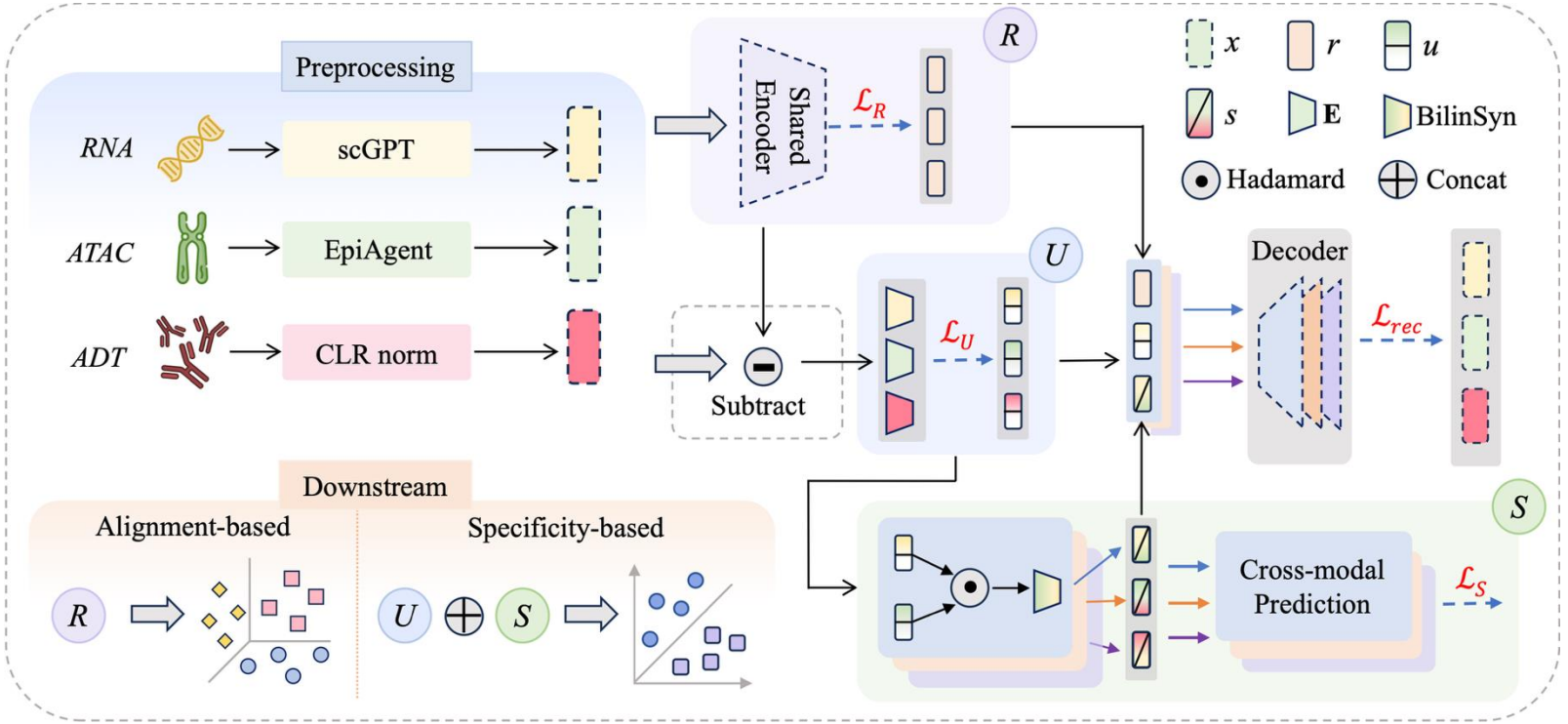
Interaction-dependent signal, e.g., regulatory relationships

scDecomp gives each component its own branch and an objective matched to its role.

PID is the design principle: R, U and S are role-specific components, not an exact trimodal PID estimate.

METHOD

scDecomp: three branches, three objectives



R **Redundancy**

Low-rank bottleneck shared by all modalities, trained with InfoNCE across modality pairs.

U **Uniqueness**

Residual (input − r) through a sparse encoder (L1 penalty); no alignment or label loss.

S **Synergy**

Bilinear (Hadamard) fusion of two U's predicts the held-out third modality.

TRAINING OBJECTIVE

$$L = L_R + \lambda_S L_S + \lambda_{rec} L_{rec} + \lambda_U L_U$$

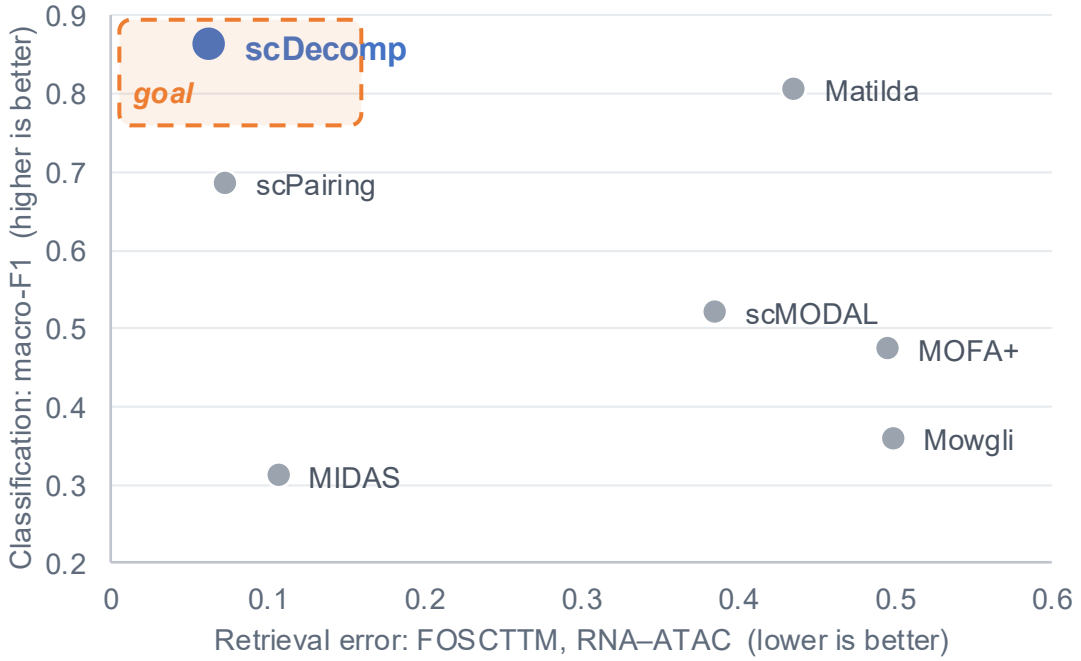
$\lambda_S = 1$, $\lambda_{rec} = 0.1$, $\lambda_U = 0.01$ — fixed for all datasets

DECODER

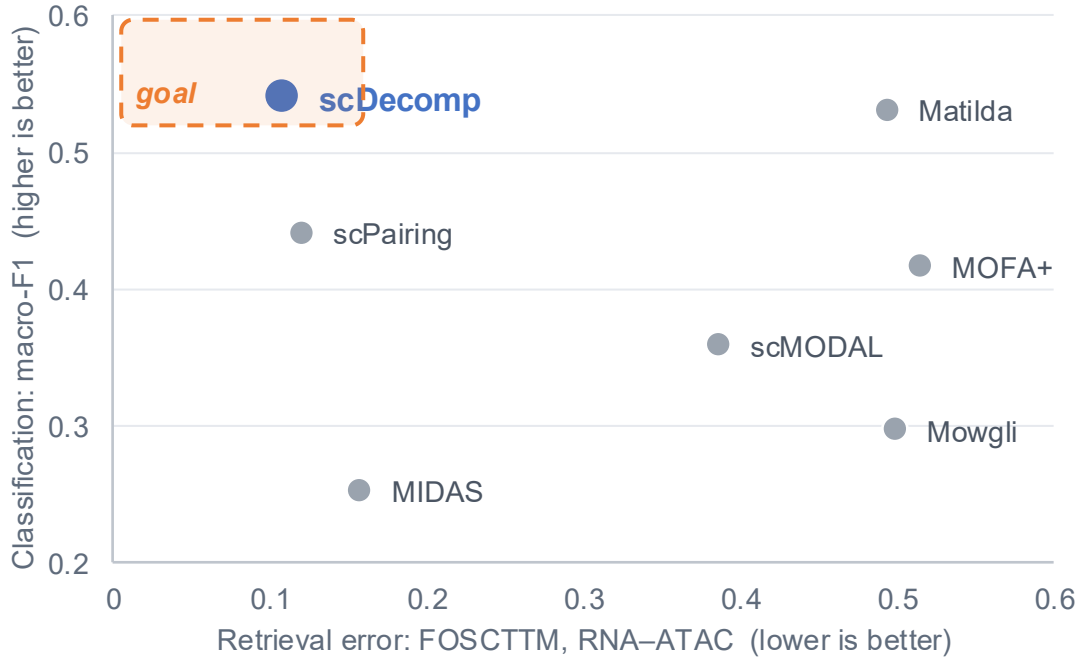
Reconstructs each modality from [r; u; s], so R, U and S jointly cover the input.

One model, best at both tasks

TEA-seq



DOGMA-seq



scDecomp: retrieval from R, classification from U + S of the same pretrained model. All methods use identical input embeddings; APOLLO (bimodal only) not shown.

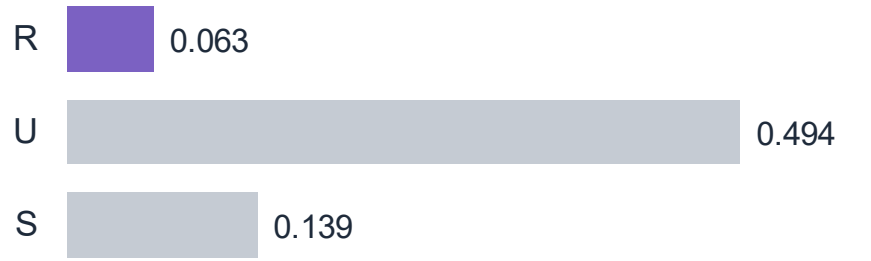
+6.9% macro-F1 on TEA-seq (0.861 vs. 0.805, Matilda)

−5.0% mean FOSCTTM vs. scPairing, all 6 modality pairs

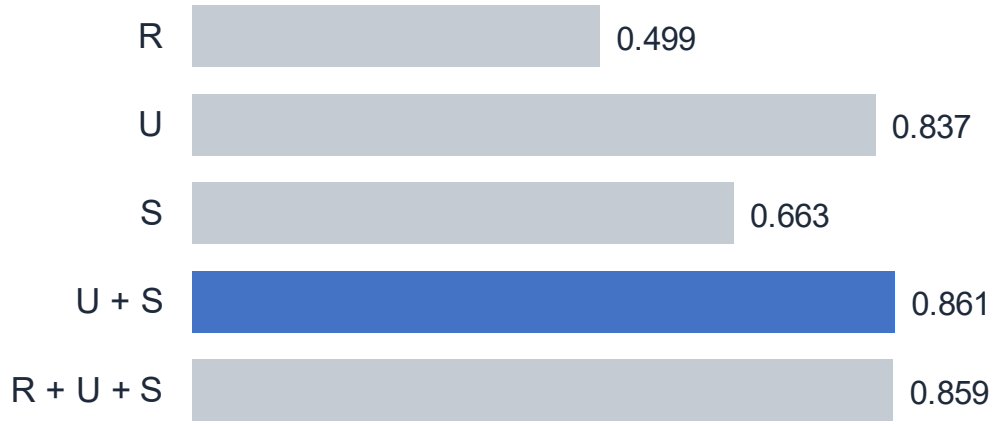
6 / 6 best Acc, wF1 and mF1 across both datasets

Each branch does its own job

Retrieval
FOSCTTM, RNA–ATAC · lower is better · 0.5 = random

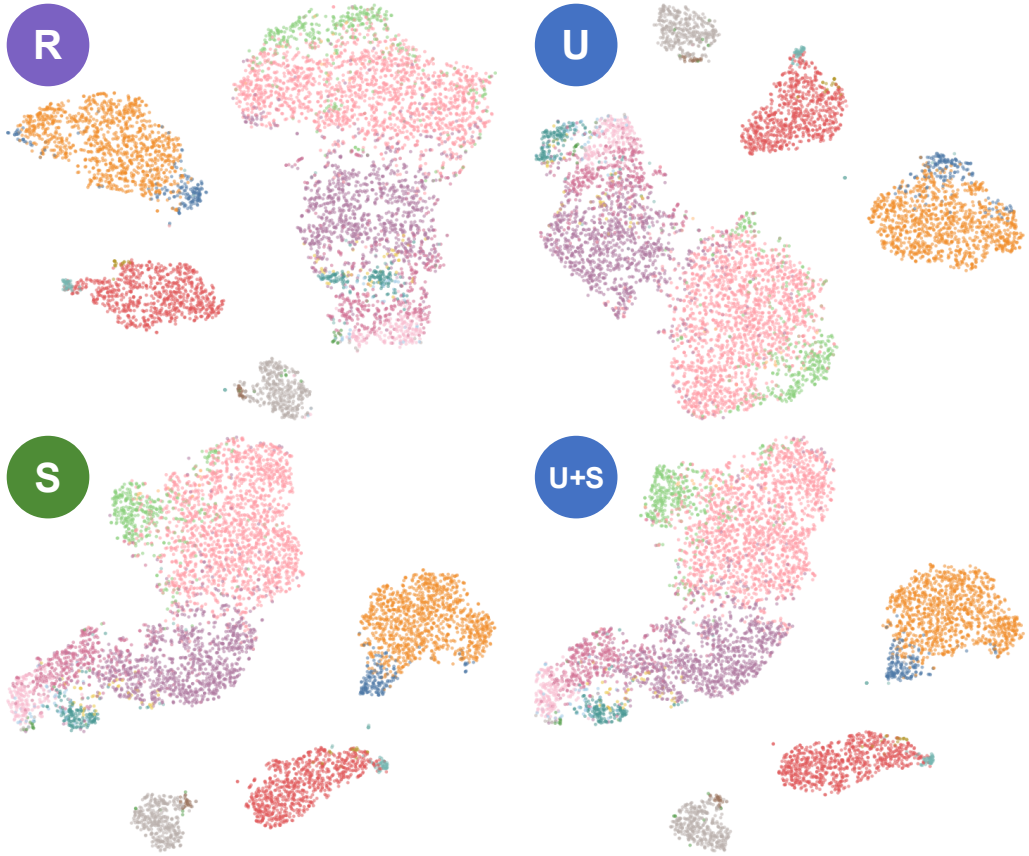


Classification
macro-F1 · higher is better



DOGMA-seq shows the same ordering (paper Tables 3–4).

What each branch sees
t-SNE, RNA modality



- MAIT cells mix into T cells in R, separate in U, and are clearest in U + S.

TAKE-HOME

Alignment or fusion? Neither alone

1

The trade-off is structural

Pushing one embedding toward cross-modal agreement discards non-shared, label-relevant signal.

2

Decompose instead of choosing

PID-inspired R, U and S branches, each with an objective matched to its role.

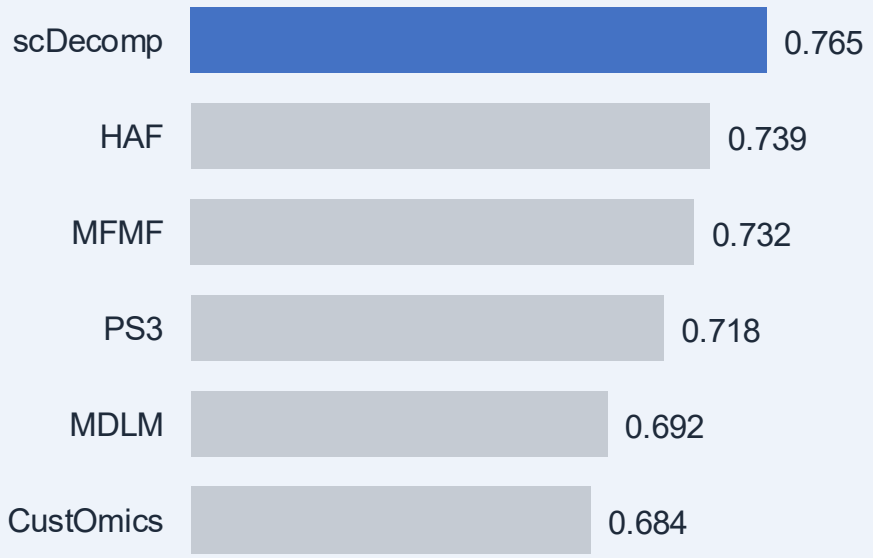
3

Each branch does its own job

R gives the best retrieval; U + S gives the best cell-type classification.

Beyond multi-omics

HANCOCK cancer cohort: survival status from 7 data types (histology, clinical, blood, ICD codes, …); 10-fold CV



AUC (bars start at 0.5 = chance); scDecomp +3.5% over HAF.

Thank you — questions welcome!

Scope: vertical integration of co-profiled cells. Mosaic or diagonal settings and batch-aware objectives are future work.

Email: tmd4090@mavs.uta.edu