arXiv:2602.04021v1 [cs.LG] 3 Feb 2026

弱配对多模态数据的组对比学习

Aditya Gorla[1,2,†], **Hugues Van Assel**[1,3], **Jan-Christian Huetter**[3], **Heming Yao**[3], **Kyunghyun Cho**[4,5], **Aviv Regev**[1,*], **Russell Littman**[1,*]

[1]Research and Early Development (gRED), Genentech [2]UCLA [3]Biology Research | AI Development (BRAID), Genentech [4]Genentech Computational Sciences [5]NYU

摘要

我们提出了 GROOVE，这是一种面向高含量扰动数据的半监督多模态表征学习方法。在这类数据中，不同模态的样本通过共享的扰动标签弱配对，但彼此之间没有直接对应关系。我们的主要贡献是 GroupCLIP，这是一种新颖的组级对比损失，弥合了面向配对跨模态数据的 CLIP 与用于单模态监督对比学习的 SupCon 之间的差距，从而解决了弱配对场景下对比学习的一项根本性缺失。我们将 GroupCLIP 与在线回译自编码器框架相结合，以促进跨模态纠缠表示，同时在共享潜在空间中保持群组级一致性。尤为重要的是，我们提出了一套全面的组合评估框架，系统地评估表征学习器在多个最优传输对齐器上的表现，以解决现有评估策略的关键局限。该框架还包括新颖的模拟，通过系统改变共享扰动效应与模态特异性扰动效应，来实现对方法鲁棒性的原则性评估。我们的组合基准测试表明，目前尚无一种对齐器能在所有设定或模态对中始终占优。在模拟实验和两个真实的单细胞遗传扰动数据集中，GROOVE 在下游跨模态匹配和插补任务上的表现与现有方法相当或更优。消融研究表明，GroupCLIP 是推动性能提升的关键组成部分。这些结果凸显了在仅有弱配对信息的场景中，利用组级约束进行有效多模态表征学习的重要性。

1 引言

近年来，扰动筛选因能够阐明因果基因调控网络（Dixit 等，2016年）、识别候选治疗靶点（Rood 等，2024年）并推动小分子药物再利用（Bhandari 等，2022年）而日益受到重视。每种给定的模态（例如 RNA 测序、ATAC测序或高内涵成像）只能观测系统底层生物学的一部分（Cui 等，2025年），因此，近期研究开始转向通过配对分析方法多模态地考察扰动效应（Frangieh 等，2021年；Martin-Rufino 等，2025年）。尽管前景可期，这种分析目前仍仅适用于特定的模态组合，例如基因表达与染色质可及性配对（Martin-Rufino 等，2025年），或基因表达与表面蛋白测量配对（Frangieh 等，2021年）。值得注意的是，目前尚无法同时获取细胞绘画实验中的扰动后显微图像（Feldman 等，2019年）和来自相同 单个细胞的扰动后基因表达谱，因为这两种测量本质上都是破坏性实验。在这种情况下，我们无法获取跨模态的配对样本，只能根据细胞的扰动（其他实验）标签对细胞进行大致的分组。

因此，多模态扰动筛选领域的近期工作已转向开发计算方法，以便在各模态之间对细胞进行事后“配对”（尽管真实配对实际上并不存在），或进行跨模态插补（见第 2 节）。这两个目标都依赖于仅利用群体层级信息，学习非配对多模态数据的有用联合表征。这种设定立即排除了现有用于联合跨模态推断的标准对比学习方法（见第 2 节）。跨模态对比学习方法

†在基因泰克实习期间完成的工作。*通讯作者： {regev.aviv, littman.russell}@gene.com

如 CLIP（Radford 等人，2021年）需要配对数据，而 SupCon（Khosla 等人，2020年）等单模态基于标签的对比方法本身无法兼容多模态数据。此外，现有的多模态单细胞深度学习方法，如非耦合自编码器（Samaran 等人，2024年；Ashuach 等人，2023年），要么依赖强人工定义的先验来建立假定的细胞对应关系，要么需要获取配对数据。目前尚无框架能够有效地从原生 弱成对多模态数据中学习，以获得充分混合的 多模态潜在表示（见第 2 节）。

贡献。 在本研究中，我们开发了**GROOVE** （面向弱配对多模态数据的组对比学习），以应对弱配对多模态数据中的上述挑战。我们的方法有三项主要贡献。首先，我们引入了组级半监督对比损失（**GroupCLIP**），弥合了 CLIP（Radford 等人，2021年）与 SupCon（Khosla 等人，2020年）在弱配对多模态数据上的差距。其次，GROOVE 将 GroupCLIP 与改编自神经机器翻译的即时 反向翻译自编码器（Artetxe 等人，2017年）相结合，构建了一个用于从弱配对单细胞数据中学习的统一架构。最后，我们开发了一个全面的评估框架，包括：（i）新颖的模拟实验，系统地改变共享信息与模态特异性信息的比例；（ii）组合基准测试，将不同的表示学习器与各种对齐算法配对，以评估匹配和跨模态插补性能。

符号说明。在标准的多模态学习设定中，所有样本都观测到全部数据模态，因此 $\mathcal{D} = \{(\boldsymbol{x}_i^{(1)}, \boldsymbol{x}_i^{(2)})\}_{i=1}^{N}$。然而，本文关注的是不存在此类配对数据的设定，即每个样本只能获取一种数据模态。相反，我们可以获取跨模态、跨样本共享的共同状态、环境、干预或扰动 标签 $t \in \mathcal{T} \subset \mathbb{Z}$。这项附加信息使我们的数据成为弱成对 数据，因此，不同模态中标签 t 不同的任何数据实例都严格无关。现在，我们可以将弱成对设定下的多模态学习重新表述为：数据 $\mathcal{D}^{(m)}$ 来自由 $m \in \mathbb{M} = \{1, 2\}$索引的两个不相交的数据模态。每个模态特异性数据集由 $N^{(m)}$ 个样本 $\mathcal{D}^{(m)} = \{(\boldsymbol{x}_i^{(m)}, t_i)\}_{i=1}^{N^{(m)}}$ 组成，其中每个 $\boldsymbol{x}_i^{(m)} \in \mathcal{X}_{(m)} \subseteq \mathbb{R}^{k^{(m)}}$ 是模态 m 的数据实例及其对应的标签 t_i。据此，我们的多模态表征学习问题是为每个样本学习一个嵌入 $\boldsymbol{z} \in \mathcal{Z} \subseteq \mathbb{R}^d$，使其位于共享的低维表示空间中，并满足 $d \ll \min(k^{(1)}, k^{(2)})$。令 $\mathcal{D}_z^{(m)} = \{(\boldsymbol{z}_i^{(m)}, t_i)\}_{i=1}^{N^{(m)}}$ 表示模态 m的潜在表示和标签的集合。我们用 $\bar{m}$表示另一种模态。

2 背景

在本节中，我们回顾更广泛的对比表征学习文献与单细胞领域中的相关工作，并指出我们的贡献旨在填补的空白。我们还简要回顾支撑我们架构的无监督神经翻译。

对比表征学习。 对比表征学习已成为一种强大的通用范式（Le-Khac 等人，2020年），既有理论基础（Alshammari 等人，2025年；Van Assel 等人，2025年；HaoChen 等人，2021年），也有广泛的实际实例。基础的 InfoNCE 损失（Oord 等人，2018年）在最大化正样本之间互信息的同时最小化负样本的互信息，并由此启发了众多有影响力的扩展（Chen 等人，2020年；He 等人，2020年；Grill 等人，2020年）。尽管这些方法取得了成功，它们仍存在一个关键局限：常常将同一类别的样本误判为负样本。有监督对比学习（Khosla 等人，2020年）通过利用标签为每个锚点形成多个正样本来缓解这一问题，但其设计仍是单模态的。类似地，Yao 等人（2024年）利用扰动标签跨批次配对样本，但作为一种单模态去噪方法在组级别上运作。对于多模态学习，CLIP（Radford 等人，2021年）和 CMC（Tian 等人，2020年）是最具影响力的方法，用于对齐图像–文本对，或将对比目标扩展到多种模态。其他工作，即 S-CLIP（Mo 等人，2023年）和 SemiCLIP（Gan 等人，2025年），探索了 CLIP 的有监督与半监督扩展，这些扩展利用有限的配对数据 以及额外的标签信息。然而，这些方法仍然需要模态之间存在一些实例级配对，无法在仅由组标签连接各模态的纯弱配对情形下运作。它们对严格配对数据的依赖仍然是一个关键瓶颈。因此，文献中的一个关键空白在于缺少能够利用弱配对的 CLIP 有监督扩展（图 1a）。尽管填补这一空白看似简单直接，但先前工作始终假设至少能够获得一些实例级配对，从而使纯弱配对情形未被涉及。

面向多模态单细胞数据的弱配对学习。 单细胞数据本质上是非配对的，因为测量过程具有破坏性，无法从同一个细胞获取多种模态。这一局限推动了非配对多模态整合（Yang 等人，2021年；Samaran 等人，2024年）和弱配对学习（Xi 等人，2024年；Ryu 等人，2025年）计算方法的发展，尤其是在涉及

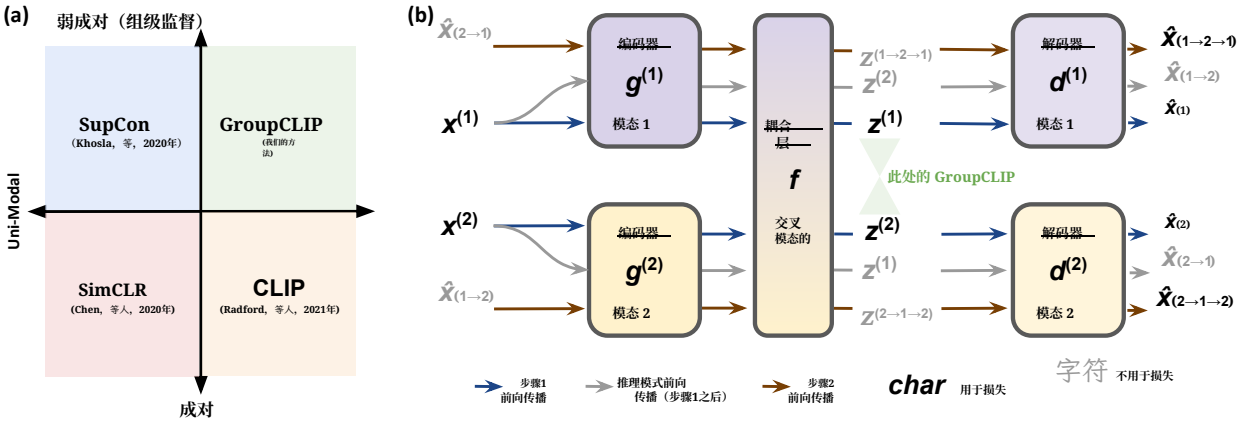


图1：**(a)** 在更广泛的对比学习背景下，GroupCLIP 可被视为 SupCon 的多模态泛化。**(b)** GROOVE 架构与训练步骤示意图。每次迭代包含两个步骤：（1）优化重建损失与 GroupCLIP 损失，然后（2）在推理模式下生成跨模态伪样本并优化回译损失。

扰动的场景中，这也是我们工作的重点。这些方法的一项尤其重要的下游应用是跨模态插补，即根据一种模态的观测结果预测另一种模态。接下来，我们将简要介绍解决这一问题的最相关工作。

首先，Xi 等人（2024）通过倾向评分匹配利用扰动标签。他们的方法训练独立分类器来预测每种模态中的扰动，然后使用所得的 logits 定义一个用于对齐的共同支持集，这遵循了 Rubin（1974年）的经典平衡得分。尽管这一方法在概念上颇具吸引力，但存在显著局限：它假定所有由扰动引起的变化都在各模态间完全共享，忽略了模态特异性效应；而且，学得的潜在表示（分类器 logits）仅捕捉可用于预测扰动的信息，舍弃了可能对下游任务至关重要、具有潜在价值的扰动内变异。其次，Ryu 等人（2025年）通过引入带标签约束的 Gromov–Wasserstein 最优传输（GW-OT）问题变体（M´emoli，2011年），解决了流程中的对齐阶段。与依赖样本级直接对应关系的方法不同，GW-OT 通过最小化两种模态所诱导度量空间之间的结构差异来对齐表示（Sebbouh 等人，2024；Van Assel 等人，2024）。这种形式能够对齐维度不同的模态特异性潜在空间，而这些空间均通过主成分分析获得。然而，该方法的有效性受到以下因素的限制：(i) 潜在表示的质量与稳健性；(ii) GW-OT 的三次计算复杂度（Peyr´e 等人，2016年）；以及 (iii) 其非凸性，使其容易陷入局部最优解（Vayer，2020年）。最后，Samaran 等人（2024）提出了唯一一种在统一框架中整合对齐与插补的方法。然而，他们的方法要求预先设定一组在各模态间共享的对齐特征——这是一个很强的假设，在实践中很少成立——并且没有利用能够显著提升对齐效果的扰动标签。

多语言神经机器翻译。 神经机器翻译已经表明，仅使用非成对单语数据，便可通过无监督方法训练出高质量的翻译系统；这些方法采用共享编码器架构、将重建与回译损失相结合的双重训练目标，以及跨语言初始化策略（Sennrich 等人，2015年；Artetxe 等人，2017年；Lample 等人，2018年；Conneau 等人，2018年）。这些方法的核心是一种即时回译策略，旨在促进纠缠的、跨模态[1] 信息丰富的潜在表示。在训练期间，来自一种模态的样本被编码到共享潜在空间，然后使用该表示在另一 模态中生成相应翻译。参数通过基于重建 （公式 1）和回译 （公式 2）损失的两步优化过程进行更新。重建（$\hat{\boldsymbol{x}}_i^{(m)}$）损失计算对同一模态中的 $\boldsymbol{x}_i^{(m)}$ 进行编码和解码时的误差，回译损失则计算将 $\boldsymbol{x}_i^{(m)}$ 编码并解码到另一模态（$\hat{\boldsymbol{x}}^{(m\to\bar{m})}{}_i$）时的误差，随后再对$\hat{\boldsymbol{x}}^{(m\to\bar{m}\to m)}{}_i$进行编码和解码，回到其原始模态。这种回译策略促使潜在表示在各模态之间充分混合。并且随着模型的改进，会生成更高质量的伪成对数据，从而形成正反馈循环，同时增强跨模态对齐与模态内表示。参见附录 A.1 以了解

[1]我们在此互换使用模态的和语言的。

有关此流程与架构的更多细节。据我们所知，来自多语言神经机器翻译的迭代回译尚未应用于弱成对单细胞多模态数据。

$$\mathcal{L}_{\text{reconstruction}} = \frac{1}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} \frac{1}{|\mathcal{D}^{(m)}|} \sum_{\boldsymbol{x}_i^{(m)} \in \mathcal{D}^{(m)}} \left\| \hat{\boldsymbol{x}}_i^{(m)} - \boldsymbol{x}_i^{(m)} \right\|_2^2 \tag{1}$$

$$\mathcal{L}_{\text{backtranslation}} = \frac{1}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} \frac{1}{|\mathcal{D}^{(m)}|} \sum_{\boldsymbol{x}_i^{(m)} \in \mathcal{D}^{(m)}} \left\| \hat{\boldsymbol{x}}_i^{(m \to \bar{m} \to m)} - \boldsymbol{x}_i^{(m)} \right\|_2^2 \tag{2}$$

3 提出的方法

我们使用非耦合自编码器来实例化基础架构，这种架构在单细胞文献中很常见（Samaran 等人，2024年；Lopez 等人，2018年），并在不同模态之间添加了一个共享的线性（耦合）投影层。我们采用无监督机器翻译中的两步优化流程进行训练（更多详情见附录 A）。然而，这个基础模型并未利用样本关联（扰动）标签 $\mathcal{T}$ 中包含的额外监督信号。我们可以利用这些额外信息，将*on-the-fly*自编码器的监督程度从无监督提升到半监督。标签信息使我们能够根据以下要求更新潜在表示：（1）不同模态中具有相同标签的样本应彼此靠近；（2）没有相同标签的样本应尽可能远离（相互排斥）。这种设定自然地引出了对比学习这一解决方案。因此，我们着手为弱配对多模态数据设计一种新颖的对比损失。我们的方法利用弱配对结构，将来自不同模态的样本视为

与正样本对具有相同标签。对于来自模态 m 的潜在表示 $\boldsymbol{z}_i^{(m)}$，我们将来自另一模态 $\bar{m}$ 且具有相同标签的所有潜在表示定义为其吸引子。

形式上，我们将 $\mathcal{P}_i^{(m)} = \{ \boldsymbol{z}_j^{(\bar{m})} \in \mathcal{D}_z^{(\bar{m})} : t_j = t_i \}$ 定义为锚点 $\boldsymbol{z}_i^{(m)}$（标签相同、模态相反）的正样本 集合，并将 $\mathcal{A}_i^{(m)} = \mathcal{D}_z^{(\bar{m})}$ 定义为来自另一模态的所有候选样本 集合（包括正样本和负样本）。遵循 CLIP （ Radford 等人，2021年），我们在来自另一模态的所有 候选样本（包括正样本）上进行归一化。给定锚点 $\boldsymbol{z}_i^{(m)}$，损失为：

$$\ell_i^{(m)} \;=\; -\log \frac{\displaystyle\sum_{\boldsymbol{z}_p \in \mathcal{P}_i^{(m)}} \exp\!\big(\mathrm{sim}(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_p)/\tau\big)}{\displaystyle\sum_{\boldsymbol{z}_a \in \mathcal{A}_i^{(m)}} \exp\!\big(\mathrm{sim}(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_a)/\tau\big)} \tag{3}$$

其中，$\mathrm{sim}(\boldsymbol{u},\ \boldsymbol{v})$ 是两个嵌入之间的相似度函数（例如余弦相似度），而 $\tau > 0$ 是用于缩放相似度的温度参数。请注意，除了标准余弦相似度外，我们还尝试了基于*t*分布的相似性核，参见附录 A.4。对所有锚点和模态取平均后得到：

$$\mathcal{L}_{\text{GroupCLIP}} = \frac{1}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} \frac{1}{|\mathcal{D}_z^{(m)}|} \sum_{\boldsymbol{z}_i^{(m)} \in \mathcal{D}_z^{(m)}} \ell_i^{(m)} \tag{4}$$

所得的 GroupCLIP 损失鼓励相同标签的潜在表示在不同模态之间聚集在一起，同时将其他所有标签的表示推离。由于对比学习对批次组成较为敏感，我们采用平衡欠采样策略，使每个小批次中各标签的样本数量保持相等，从而避免类别不平衡，而无需对稀有标签进行过采样，详见附录 A.5。

我们的方法将有监督对比学习拓展至跨模态场景（图 1a），填补了对比学习领域的一项关键空白。正如 CLIP（Radford 等人，2021年）是 SimCLR（Chen 等人，2020年）无监督对比框架的典型跨模态扩展，GroupCLIP 也是 SupCon（Khosla 等人，2020年）有监督对比方法的自然跨模态扩展。CLIP 利用自然配对数据，无需明确的标签监督；而 GroupCLIP 利用标签监督，无需模态之间存在自然配对。这一区别对生物学应用至关重要，因为真实的跨模态配对往往难以获得，而扰动标签却能提供丰富的监督信号。通过将有监督对比学习引入多模态领域，GroupCLIP 弥合了单模态监督方法与跨模态无监督方法之间的差距，为配对关系较弱但标签信息丰富的场景提供了一个有原则的框架。

最后，我们将新颖的 GroupCLIPloss（公式 4）与即时 回译自编码器框架（公式 1 和 2）相结合，构建了 GROOVE 架构——一种新颖的方法，能够通过以下损失函数从弱成对多模态数据中学习统一表示：

$$\mathcal{L}_{\text{step-1}} = \alpha \cdot \mathcal{L}_{\text{GroupCLIP}} + \beta \cdot \mathcal{L}_{\text{reconstruction}} \quad (5) \qquad \mathcal{L}_{\text{step-2}} = \beta \cdot \mathcal{L}_{\text{backtranslation}} \quad (6)$$

α，β 是用于平衡各损失分量的超参数。附录 A.6 中的算法 1 概述了 GROOVE 的整体训练循环，图 1b 则直观展示了这一架构。

4 评估框架

我们通过两种方法评估所学习潜在表示的质量。首先，我们评估它们在基于最优传输的跨模态样本匹配中的效用。这是一种标准评估方式（Ryu 等人，2025年；Xi 等人，2024年），因为高质量的潜在表示应能够恢复有意义的传输计划，从而正确识别并匹配不同模态中的相似样本。随后，我们评估一项利用该传输计划的关键下游任务的表现：单向插补，即根据另一种模态的样本预测一种模态。

4.1 组合评估

我们发现，先前评估存在一个关键局限：它们要么固定表征学习器，并比较各种最优传输方法（Ryu 等人，2025年）；要么固定最优传输方法，并比较不同的表示学习器（Xi 等人，2024年）。这种评估范式可能产生系统性偏差，因为表征学习方法的表现与下游对齐算法的选择密不可分。这种关联源于不同最优传输公式所依据的几何假设存在根本差异。
例如，熵正则最优传输（EOT）隐式假设两种模态都可以嵌入到共享特征空间中，因此，使用相同度量比较表征时，跨模态对应关系才有意义。换言之，预期学习到的嵌入能在共同坐标系中一致地保留不同模态之间的距离。相比之下，熵正则 Gromov–Wasserstein 最优传输（EGWOT）不假设存在共享空间；相反，它将不同模态视为可能各异的度量空间，并通过比较各自内部的距离结构，而非绝对坐标来进行对齐。这些不同的假设表明，针对某一种几何框架优化的表征学习器，在其他最优传输方法下可能无法发挥最佳性能。为了解决这一局限，我们提出了一种组合评估框架，系统地测试每种表征学习器与所有可用最优传输对齐器的组合。该方法使我们能够评估表征学习器与对齐算法之间的相互作用，从而得出更稳健的性能评估，并将下游选择的敏感性纳入考量。

4.2 基线方法

我们的评估聚焦于能够直接处理弱成对数据、无需预先指定特征对应关系或配对样本的方法。其中包括以下表征学习基线方法：倾向评分（PS）匹配（Xi 等人，2024年），以及带标签适配的领域对抗变分自编码器（DAVAE）。后者是我们在 Ashuach 等人（2023年）的方法基础上进行的定制修改，详见 Ryu 等人（2025年）。PS 是最新的弱成对跨模态匹配方法，也是现有方法中唯一以原生形式直接解决这一问题的方法。DAVAE 在概念上与我们提出的方法最为接近，但它采用模态识别对抗损失，并使用线性探针监督进行潜在空间正则化。所有方法均采用相同的架构和一致的潜在嵌入维度；我们评估了每种表征学习器与标准最优传输方法（EOT、EGWOT）及标签约束最优传输方法（LabeledEOT、LabeledEGWOT、labeledCOOT）的组合。详情见附录 B。

对于单向插补，我们训练了一个由 ReLU 激活交错构成的 2 层多层感知机，并添加了一个最终线性投影层。该模型使用 OT 对齐器返回的传输计划进行训练，该计划定义了训练采样策略。具体而言，在使用模态 2 预测模态 1 时，给定传输计划$\boldsymbol{T} \in [0,1]^{N^{(1)} \times N^{(2)}}$，对于每个来自模态 2 的样本索引-$i$，我们根据$j \sim$多项分布$\left(\boldsymbol{T}_{:,i} / \sum \boldsymbol{T}_{:,i}\right)$从模态 1 中采样索引-$j$，每个小批次均如此。

4.3 数据集

模拟。 以往评估的一个关键局限在于假设扰动引起的变异在不同模态间完全共享，导致模态特异性变异与标签完全无关；对于生物学数据而言，这一假设并不现实（Argelaguet 等人, 2020年；Lin 和 Zhang, 2023年；Gorla 等人, 2025a）。为解决这一问题，我们开发了一个概率模拟框架，用于捕捉受扰动影响的共享潜在变异和模态特异性潜在变异。该框架通过共享潜在因子（对两种模态产生相同影响）和模态特异性因子（各自仅存在于一个分子层中）对每个细胞进行建模，并设置两种扰动类型：同时影响两种模态的共享扰动，以及分别作用于单个模态的模态特异性扰动。我们通过改变共享维度与模态特异性维度的比例，在不同耦合水平下系统评估方法性能，并测试共享变异占比为 100%、80% 和 50% 的情形。完整的模拟详情见附录 C。对于每种情形，我们模拟 10 个重复样本，并采用扰动平衡的 80-20 训练-测试划分进行评估。

Perturb-Multiome 我们使用 Martin-Rufino 等人（2025年）的成对基因表达与染色质可及性（多组学）数据集，其中包含基于转录因子的扰动，以进行更贴近实际的评估。原始数据经预处理并抽取子集后，得到覆盖 20 种扰动的 2560 个细胞，每种扰动 128 个细胞，以及 512 个基因和 256 个基因可及性评分。详见附录 D。该数据集在两种框架下进行分析：（1）扰动平衡的五折划分；（2）留一扰动法（LOPO）。该数据集的所有插补结果均聚焦于由可及性评分预测基因表达。

Perturb-CITE-seq. 我们还使用 Frangieh 等（2021年）的 Perturb-CITE-seq 数据，其中包含遗传扰动以及成对的 RNA 测序与表面蛋白测量，以进行更贴近实际的评估。原始数据经预处理并抽取子集后，得到覆盖 19 种遗传扰动的 3689 个细胞（每种条件下的细胞数：中位数 [，最小值、最大值] = 201[72, 270]），以及 20 种蛋白质和 500 个基因，并直接用作输入。详见附录 E。该数据集同样在平衡五折划分和留一扰动交叉验证下进行分析。该数据集的所有插补结果均聚焦于由蛋白质水平测量值预测基因表达。

4.4 指标

由于我们的数据集包含真实的细胞配对，我们使用两项基于真实值的指标评估最优传输匹配准确率。归一化传输计划的迹 衡量被完美匹配的真实配对所占比例（Xi 等人，2024年），例如，当全部权重都分配给真实配对时该值为1.0，而对于无信息的均匀分配则为 $1/N$。对称的比真实匹配更近的细胞重心分数（*Bary. FOSCTTM*）衡量传输计划相对于真实配对错误分配给错误配对的权重程度（Demetci 等人，2022年；Liu 等人，2019年），其中0.0表示完美匹配，0.5表示随机均匀分配。我们通过对两种模态的性能取平均来报告对称Bary. FOSCTTM。接下来，我们使用6项指标量化插补准确率：均方误差、1-瓦瑟斯坦距离（加权距离）、余弦相似度（余弦相似度）、*KNN* 召回率、KNN 平均精确率-召回率（*KNN PR*）和KNN 受试者工作特征曲线下面积（*KNNROC*）。均方误差和加权距离是插补文献中的标准指标（Gorla 等人，2025年b），其中前者评估全局误差，后者量化分布

表 1：模拟中各共享比例设置下前五种方法组合的匹配性能指标。标准误遵循 ±；最佳结果以粗体标示，次佳结果加下划线。

共享比例	方法	Mean Rank	迹	重心 FOSCTTM
100	GROOVE（余弦）+ LabeledCOOT	1.5	**0.856±0.027**	0.027±0.006
	GROOVE（余弦）+ LabeledEOT	3.0	0.466±0.020	**0.026±0.004**
	DAVAE+ LabeledCOOT	3.5	0.669±0.034	0.066±0.009
	DAVAE+ LabeledEOT	4.5	0.453±0.023	0.042±0.006
	倾向评分+LabeledEOT	5.5	0.366±0.006	0.054±0.006
80	DAVAE+ LabeledEOT	2.5	0.165±0.008	**0.146±0.014**
	GROOVE（余弦）+ LabeledCOOT	3.5	**0.237±0.015**	0.180±0.012
	GROOVE（余弦）+ LabeledEOT	4.5	0.148±0.008	0.150±0.013
	倾向评分+LabeledEOT	4.5	0.156±0.005	0.161±0.011
	DAVAE+ LabeledCOOT	6.0	0.202±0.014	0.196±0.011
50	DAVAE+ LabeledEOT	2.5	0.125±0.007	**0.162±0.006**
	GROOVE（余弦）+ LabeledEOT	4.0	0.118±0.007	**0.162±0.009**
	GROOVE（余弦）+ LabeledCOOT	5.0	**0.184±0.015**	0.206±0.008
	GROOVE (tdist)+ LabeledEOT	5.5	0.105±0.006	0.168±0.009
	DAVAE+ LabeledCOOT	6.0	0.164±0.011	0.214±0.007

对齐。余弦相似度是扰动预测任务中的常用指标（Littman 等，2025年；Adduri 等，2025年）。基于KNN的指标受到（Littman 等，2025年）近期工作的启发，该工作表明，KNN指标能够稳健评估扰动效应预测任务中的局部相似性保持。KNN指标通过评估邻域保持而非全局预测误差，避免了单个基因失效带来的偏差。所有指标均在留出数据（测试集或交叉验证折）上报告，并给出标准误（SE；跨折或重复样本）。标准误四舍五入至小数点后三位；报告为0.000的值表示SE<0.0005。有关指标的更多细节，见附录F。

5 结果

5.1 模拟数据的总体性能

表 1 展示了模拟数据中不同共享比例设置下排名前五的方法的匹配性能。在标准的 100% 共享扰动相关变异设置下，GROOVE（与 LabeledCOOT 结合）相比所有其他方法组合均取得了更优的性能。在 80% 和 50% 共享比例下，GROOVE 平均排名第二；然而，在所有共享比例条件下，它都保持着最高的基于 Trace 的匹配性能。由于匹配性能是实现下游目标的中间步骤，表 2 在相同的模拟数据上评估了下游插补性能。在 100% 和 80% 共享条件下，基于 GROOVE 的方法再次取得最高排名，而在 50% 共享条件下，基于 PS 的组合排名略高。进一步考察各项指标可以发现，基于 GROOVE 的组合在几乎所有单项指标上都持续取得最佳性能，但 KNN PR 除外；在该指标上，它们比 PS+LabeledEOT 低 0.001（远小于 SE）。比较表 1 和表 2 中的方法排名，我们注意到，更好的匹配性能并不能保证最优的下游任务性能。例如，在共享比例为 $<100\%$ 时，DAVAE 展现出更好的平均匹配性能，但在相同条件下，基于 PS 的组合取得了更好的插补性能。尽管存在这些差异，基于 GROOVE 的组合在 100% 共享变异下取得了顶尖性能，并展现出稳健且一致的表现，在所有评估场景中均稳居至少前两名。

5.2 多模态单细胞性能

现在，我们在更贴近实际的 Perturb-CITE-seq 数据集上评估性能。表 3 展示了五折交叉验证下排名前十的方法组合的插补性能。与模拟数据的结果相比，基于 GROOVE 的方法占据了前五名，而基于 PS 的方法位列倒数两位。这一模式支持了基于 GROOVE 的组合在不同实验条件下的稳健性。接下来，表 6（附录 H）进一步显示，基于 GROOVE 的方法在匹配性能排名中表现最佳。值得注意的是，一种基于 PS 的方法在最佳匹配性能方面与基于 GROOVE 的方法并列；然而，与我们的模拟结果一致，更优的匹配性能并不一定能转化为更好的下游插补结果。我们还报告了留一扰动交叉验证下排名前十的方法的性能，匹配性能和插补性能结果分别见表 7 和表 8（附录 H）。即使在这一设置下，

表 2：模拟中各共享比例设置下前五种方法组合的插补性能指标。标准误遵循 ±；最佳结果以粗体标示，次佳结果加下划线。

共享比例	方法	Mean Rank	MSE	余弦相似度	KNN 召回率	KNN PR	KNN ROC	WD
100	GROOVE（余弦）+LabeledCOOT	1.00	**0.099±0.012**	**0.949±0.005**	**0.444±0.029**	**0.272±0.024**	**0.706±0.015**	**0.065±0.003**
	DAVAE+LabeledCOOT	2.00	0.121±0.013	0.931±0.006	0.412±0.027	0.244±0.021	0.689±0.014	0.092±0.005
	倾向评分+LabeledCOOT	3.50	0.148±0.012	0.920±0.006	0.382±0.023	0.221±0.017	0.673±0.012	0.107±0.004
	GROOVE（tdist）+LabeledCOOT	4.00	0.146±0.014	0.931±0.004	0.372±0.020	0.216±0.014	0.668±0.011	0.097±0.005
	DAVAE+LabeledEOT	4.67	0.156±0.010	0.916±0.005	0.379±0.021	0.220±0.015	0.672±0.011	0.142±0.004
80	GROOVE（余弦）+LabeledEOT	2.83	0.623±0.119	**0.837±0.011**	**0.241±0.014**	**0.137±0.008**	**0.598±0.008**	0.199±0.027
	倾向评分+LabeledEOT	3.00	0.593±0.115	0.836±0.010	0.239±0.013	0.135±0.007	0.597±0.007	**0.198±0.027**
	GROOVE（余弦）+LabeledCOOT	5.00	0.740±0.197	0.832±0.016	0.240±0.017	0.134±0.008	**0.598±0.009**	0.126±0.010
	倾向评分+LabeledEGWOT	5.00	0.614±0.113	0.834±0.009	0.238±0.014	0.135±0.007	0.597±0.007	0.256±0.028
	倾向评分+EOT	5.33	**0.601±0.111**	0.832±0.010	0.236±0.014	0.134±0.007	0.596±0.007	0.187±0.022
50	倾向评分+LabeledEOT	3.83	0.516±0.040	0.830±0.011	**0.215±0.011**	**0.123±0.006**	**0.585±0.006**	0.183±0.011
	GROOVE（余弦）+EOT	4.00	0.519±0.039	**0.838±0.009**	**0.215±0.009**	0.122±0.006	**0.585±0.005**	0.208±0.012
	GROOVE（余弦）+LabeledEOT	4.00	**0.513±0.040**	0.835±0.011	0.212±0.010	0.121±0.006	0.583±0.005	0.194±0.012
	DAVAE+LabeledEOT	5.33	0.517±0.041	0.833±0.009	0.211±0.012	0.121±0.007	0.582±0.006	0.189±0.009
	GROOVE（余弦）+LabeledCOOT	6.00	0.559±0.049	0.832±0.012	0.213±0.010	0.121±0.006	0.584±0.005	**0.134±0.009**

表 3：Perturb-CITE-seq 数据集中前 10 种方法组合在五折评估下的插补性能指标。标准误位于 ± 之后；最优值加粗，次优值加下划线，同质指标不作标注。

方法	Mean Rank	MSE	余弦相似度	KNN 召回率	KNN PR	KNN ROC	WD
GROOVE（余弦）+LabeledEGWOT	6.83	**0.261±0.001**	0.049±0.002	0.020±0.001	0.017±0.000	0.503±0.000	0.353±0.001
GROOVE（余弦）+LabeledCOOT	7.33	0.282±0.000	0.016±0.002	**0.021±0.001**	0.017±0.000	0.503±0.000	0.297±0.000
GROOVE（tdist）+LabeledCOOT	7.50	0.282±0.001	0.018±0.001	0.020±0.001	0.017±0.000	0.503±0.000	0.297±0.000
GROOVE（tdist）+LabeledEGWOT	7.67	**0.261±0.000**	0.046±0.002	0.019±0.001	0.017±0.000	0.502±0.000	0.353±0.001
GROOVE (tdist)+LabeledEOT	7.67	0.262±0.000	0.042±0.002	0.020±0.001	0.017±0.000	0.503±0.000	0.347±0.002
DAVAE+ LabeledEGWOT	8.00	**0.261±0.000**	**0.059±0.002**	0.018±0.001	0.017±0.000	0.502±0.000	0.352±0.000
DAVAE+ LabeledCOOT	8.33	0.295±0.000	0.019±0.001	0.020±0.001	0.017±0.000	0.503±0.001	**0.281±0.000**
GROOVE（余弦）+LabeledEOT	9.00	0.262±0.001	0.044±0.002	0.019±0.001	0.017±0.000	0.502±0.000	0.350±0.001
倾向评分+ LabeledCOOT	9.67	0.295±0.001	0.018±0.002	0.019±0.001	0.017±0.000	0.502±0.000	**0.281±0.001**
倾向评分+ LabeledEOT	10.00	0.262±0.001	0.040±0.001	0.019±0.001	0.017±0.000	0.502±0.000	0.344±0.001

表 4：在 Perturb-Multiome 数据集上采用留一扰动评估时，排名前十的方法组合的插补性能指标。标准误遵循 ±；最佳结果以粗体表示，第二佳结果加下划线。

方法	Mean Rank	MSE	余弦相似度	KNN 召回率	KNN PR	KNN ROC	WD
GROOVE (tdist)+ LabeledEOT	2.17	**0.306±0.003**	**0.140±0.015**	0.192±0.007	0.132±0.004	0.558±0.004	0.425±0.003
GROOVE（余弦）+LabeledEOT	4.83	0.310±0.002	0.078±0.008	**0.198±0.008**	**0.133±0.004**	**0.561±0.004**	0.432±0.001
GROOVE（tdist）+ EOT	5.00	0.309±0.002	0.100±0.012	0.180±0.008	0.126±0.004	0.551±0.005	0.431±0.002
倾向评分+ LabeledEOT	6.67	0.310±0.002	0.074±0.008	0.177±0.007	0.122±0.003	0.550±0.004	0.431±0.002
GROOVE（余弦）+ EOT	7.17	0.311±0.002	0.073±0.006	0.183±0.006	0.127±0.003	0.553±0.004	0.433±0.001
GROOVE(余弦)+LabeledCOOT	7.67	0.357±0.014	0.072±0.030	0.170±0.008	0.122±0.003	0.546±0.004	0.326±0.004
DAVAE+ LabeledEOT	8.33	0.310±0.003	0.080±0.004	0.117±0.006	0.101±0.002	0.517±0.003	0.428±0.002
倾向评分+ EOT	9.50	0.310±0.002	0.077±0.008	0.129±0.006	0.104±0.002	0.524±0.003	0.432±0.002
GROOVE (tdist)+ LabeledCOOT	10.33	0.363±0.013	0.050±0.030	0.161±0.008	0.119±0.004	0.541±0.004	**0.324±0.005**
DAVAE+ LabeledCOOT	10.50	0.344±0.010	0.062±0.023	0.148±0.007	0.113±0.003	0.534±0.004	0.348±0.004

基于 GROOVE 的方法在插补方面取得了前两名。我们还在 Trace 和 Bary. FOSCTTM 指标上分别取得了唯一的最高成绩。

接下来，我们在 Perturb-Multiome 数据集上评估性能。表 12 和表 10（附录 I）分别总结了五折交叉验证下排名前 10 的方法在匹配和插补任务上的结果。与此前的数据集一致，GROOVE 在两项任务中的总体排名均最高。结合模拟实验来看，这些结果表明，余弦相似度核通常是不错的默认选择，但在某些情况下，t分布核也会有所帮助。我们还再次观察到，较高的匹配性能并不意味着良好的插补性能：DAVAE 的匹配性能优于基于 PS 的方法，但其插补性能未能进入前 10 名。类似地，在 LOPO 评估下，基于 GROOVE 的方法在匹配性能（附录 I，表 11）和插补性能方面均位居前三。有趣的是，在固定底层表征学习器时，EOT 变体在 Perturb-Multiome 数据集上的插补性能更优；而在 Perturb-CITE-seq 数据集的插补任务中，EGWOT 变体（COOT 是其中之一）更受青睐。这一规律表明，最优传输对齐器的选择取决于数据集特征，包括具体的模态对，而与所选表征学习器无关。

表 5：在 80% 共享变异模拟和 Perturb-Multiome 数据集下，GROOVE 消融分析的性能指标。标准误遵循 ±；最佳结果以粗体显示，次佳结果加下划线。

数据集	消融类型	重心 FOSCTTM	MSE	余弦相似度	KNN 召回率	KNN PR	KNN ROC	WD
扰动-多组学	GROOVE（余弦）	**0.489±0.007**	**0.308±0.003**	**0.102±0.028**	**0.044±0.007**	**0.028±0.002**	**0.512±0.004**	**0.428±0.005**
	NoGroupCLIP	0.497±0.001	0.310±0.001	0.076±0.007	0.036±0.002	0.025±0.000	0.507±0.001	0.432±0.002
	仅自编码器	0.500±0.000	0.311±0.001	0.069±0.002	0.029±0.002	0.025±0.000	0.504±0.001	0.433±0.001
模拟	GROOVE（余弦）	**0.143±0.012**	**0.622±0.122**	**0.836±0.011**	**0.239±0.015**	**0.137±0.008**	**0.597±0.008**	**0.195±0.021**
	No GroupCLIP	0.321±0.013	0.965±0.225	0.778±0.016	0.187±0.014	0.106±0.007	0.570±0.008	0.226±0.007
	仅自编码器	0.280±0.013	0.846±0.175	0.793±0.013	0.195±0.012	0.111±0.006	0.574±0.006	0.240±0.012

5.3 消融分析

我们进行了消融分析，以评估 GROOVE 方法中各关键组件的相对重要性。这些分析固定使用余弦相似度核和对齐器。我们考察了两种消融配置：（1）“No GroupCLIP”：采用即时 回译，并使用重建损失和回译损失，但移除 GroupCLIP 损失；（2）“仅自编码器”：采用标准重构损失，不使用 GroupCLIP 或回译。我们在 80% 共享比例模拟设置下（使用 LabeledEOT）以及真实的 Perturb-Multiome 数据集上（使用 EOT）评估了这两种配置。表 9 展示了消融结果。在模拟数据和真实数据上，移除 GroupCLIP 损失都会导致所有指标中最大的性能下降。尽管这两种消融配置相较于完整的 GROOVE 均出现了显著退化，但标准自编码器与不使用 GroupCLIP 的回译方法之间的差异相对较小，这凸显了 GroupCLIP 是性能提升的主要驱动因素。GroupCLIP 提供的额外监督为潜在表示引入了有意义的软约束。这支持了我们的假设：仅靠回译不足以促进群组级一致性和跨群组判别。

在 Perturb-Multiome 数据集上（采用五折交叉验证），我们发现，在标准自编码器框架中加入即时回译后，大多数指标都有显著提升。然而，在模拟中，这两种方法之间的差异微乎其微。对此，我们认为有两种可能的解释，但二者并不互斥。首先，我们的模拟仍未能捕捉真实多模态单细胞数据的全部复杂性，而回译可以利用这种复杂性。其次，原始的即时回译框架（Artetxe 等人，2017年）使用了一个共享的预训练多模态编码器，而本研究未能使用该编码器。这些因素可能共同或单独导致我们低估回译的效用，以及 GROOVE 在模拟中的整体潜力。需要指出的是，我们尝试在 Perturb-CITE-seq 数据集上进行消融分析（附录 H），但指标的动态范围不足以支持有意义的结论。

此外，我们还进行了超参数扫描，以定量评估模型对关键参数的敏感性。图 2 展示了在共享比例为 100%、80% 和 50% 的设置下，平均匹配性能（迹和 Bary. FOSCTTM）随温度（τ）以及重建/回译权重（β）变化的等高线图。结果表明，GROOVE 对 τ 相对稳健，但对 β 更为敏感。与所有深度学习方法一样，我们预计超参数敏感性图谱会因数据集和目标而显著不同。我们建议用户针对具体数据集进行超参数优化，以获得最佳结果。

6 讨论

本研究提出了一种用于弱成对数据的多模态半监督表示学习方法，在一种即时回译自编码器中采用了新颖的群组对比损失（GroupCLIP）。我们还引入了组合式基准测试（针对最优传输对齐器），用于多模态单细胞扰动对齐和跨模态插补。实证评估表明，GROOVE 能够学习到更有用的潜在表示，以用于细胞级匹配和下游插补。消融实验表明，GroupCLIP 是对齐多模态表示并取得良好对齐性能的关键组成部分。我们认为，GroupCLIP 的用途不只限于单细胞数据，因为弱配对多模态数据在许多领域都很常见（Yang 等，2020；Mei 等，2024；Sun 等，2024）。还需注意的是，GroupCLIP 和 GROOVE 并未作出将其限制于双模态数据的强假设；不过，我们将该方法在 > 2 模态上的应用留待未来研究。

虽然我们的方法在各种数据集上均展现出稳定且有意义的性能，但我们也承认其存在一些局限，并指出了未来的研究方向。首先，在模拟与真实数据集评估中，相对排名存在不一致。这种差异凸显了学界开发更真实的模拟框架的必要性，以便更好地指导未来的模型开发。至关重要的是，这些模拟框架应纳入不同比例的共享变异，因为真实数据集所属的实际情形事先并不明确。其次，学界需要就评估优先级达成共识：应将样本匹配还是插补等下游任务作为主要性能标准。我们的结果表明，匹配性能更优并不意味着下游插补效果更好，这说明这些目标可能需要不同的方法学方法。优先级的确定至关重要，因为适用于匹配任务的最优方案，很可能不同于在下游应用中表现出色的方案。第三，我们需要更灵敏的插补性能指标（动态范围更大），以便稳健地评估扰动预测。开发这些指标需要与生物学家密切合作，以明确相关研究目标，从而构建能够准确衡量模型在具有科学意义的任务上表现的评估框架。

致谢与经费披露

H.V.A、J-C.H、K.C.、A.R. 和 R.L. 是罗氏的员工。H.Y.、J-C.H、K.C. 和 A.R. 持有罗氏股权。相关工作是在基因泰克实习期间完成的（AG）。A.R. 是 Celsius Therapeutics 的联合创始人及股权持有人，也是 Immunitas 的股权持有人；在 2020年7月31日之前，曾担任 ThermoFisher Scientific、SyrosPharmaceuti- cals、Neogene Therapeutics 和 Asimov 的科学顾问委员会成员。自 2020年8月1日 起，A.R. 成为基因泰克员工，并持有罗氏股权。

参考文献

Atray Dixit、Oren Parnas、Biyu Li、Jenny Chen、Charles P Fulco、Livnat Jerby-Arnon、Nemanja D Marjanovic、Danielle Dionne、Tyler Burks、Raktima Raychndhury、Britt Adamson、Thomas M Norman、Eric S Lander、Jonathan S Weissman、Nir Friedman 和 Aviv Regev。《Perturb-seq：通过对混合遗传筛选进行可扩展的单细胞RNA分析，剖析分子回路》。细胞，2016年。

Jennifer E. Rood、Anna Hupalowska和Aviv Regev。迈向基于扰动细胞与组织图谱的细胞与组织因果生物学基础模型。《*Cell*》，2024年。Ashish Bhandari、Rebecca Sarto Basso、Daria Beshnova、Aurora Blucher、Carl Brooks、Lu Chen、China Chien、Jacob Cooper、Michael Cuccarese、Hayley Donnella、Bryan Ellis、Marie Evangelista、Kevin Fales、Imran Haque、Aimee Iberg、Weston Judd、Kiran Nadella、Chase Neumann、Janet Paulsen、Paul Rearden、Shane Rowley、Jenny Rudnick、Ashraf Saeed、Bahar Shamloo和Lee Zamparo。在基因组与化学扰动的统一表征空间中识别并优化免疫检查点耐药的新型小分子调节剂。收录于美国癌症研究协会，2022年。Haotian Cui、Alejandro Tejada-Lapuerta、Maria Brbi´c、Julio Saez-Rodriguez、Simona Cristea、Hani Goodarzi、Mohammad Lotfollahi、Fabian J. Theis和Bo Wang。迈向分子细胞生物学中的多模态基础模型。《自然》，2025年。Chris J Frangieh、Johannes C Melms、Pratiksha I Thakore、Kathryn R Geiger-Schuller、Patricia Ho、Adrienne M Luoma、Brian Cleary、Livnat Jerby-Arnon、Shruti Malu、Michael S Cuoco等。患者模型中的多模态混合扰动-cite-seq筛选揭示癌症免疫逃逸机制。《自然·遗传学》，53(3):332–341，2021年。Jorge Diego Martin-Rufino、Alexis Caulier、Seayoung Lee、Nicole Castano、Emily King、Samantha Joubran、Marcus Jones、Seth R Goldman、Uma P Arora、Lara Wahlster等。转录因子网络对血细胞表型遗传力的富集程度不成比例。《科学》，388(6742):52–59，2025年。David Feldman、Avtar Singh、Jonathan L Schmid-Burgk、Rebecca J Carlson、Anja Mezger、Anthony J Garrity、Feng Zhang和Paul C Blainey。人类细胞中的光学混合筛选。《*Cell*》，2019年。Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark等。从自然语言监督学习可迁移的视觉模型。收录于国际机器学习会议，第8748–8763页。PmLR，2021年。Prannay Khosla、Piotr Teterwak、Chen Wang、Aaron Sarna、Yonglong Tian、Phillip Isola、Aaron Maschinot、Ce Liu和Dilip Krishnan。有监督对比学习。神经信息处理系统进展，33: 18661–18673，2020年。Jules Samaran、Gabriel Peyr´e和Laura Cantini。scConfluence：基于弱连接特征的正则化逆最优传输实现单细胞对角整合。《自然·通讯》，15(1):7762，2024年。Tal Ashuach、Mariano I Gabitto、Rohan V Koodli、Giuseppe-Antonio Saldi、Michael I Jordan和Nir Yosef。MultiVI：用于整合多模态数据的深度生成模型。《自然·方法》，20(8):1222–1231，2023年。Mikel Artetxe、Gorka Labaka、Eneko Agirre和Kyunghyun Cho。无监督神经机器翻译。*arXiv* 预印本 *arXiv:1710.11041*，2017年。Phuc H Le-Khac、Graham Healy和Alan F Smeaton。对比表征学习：框架与综述。《*IEEE Access*》，8:193907–193934，2020年。Shaden Alshammari、John Hershey、Axel Feldmann、William T. Freeman和Mark Hamilton。I-Con：表征学习的统一框架，2025年。URL https://arxiv.org/abs/2504.16929。Hugues Van Assel、Mark Ibrahim、Tommaso Biancalani、Aviv Regev和Randall Balestriero。联合嵌入与重建：潜在空间预测对自监督学习的可证明益处。神经信息处理系统进展，39，2025年。Jeff Z HaoChen、Colin Wei、Adrien Gaidon和Tengyu Ma。基于谱对比损失的自监督深度学习的可证明保证。神经信息处理系统进展，34:5000–5011，2021年。

Aaron van den Oord、Yazhe Li和Oriol Vinyals。基于对比预测编码的表征学习。*arXiv* 预印本 *arXiv:1807.03748*，2018年。Ting Chen、Simon Kornblith、Mohammad Norouzi和Geoffrey Hinton。视觉表征对比学习的简单框架。收录于 国际机器学习会议，第1597–1607页。PmLR，2020年。Kaiming He、Haoqi Fan、Yuxin Wu、Saining Xie和Ross Girshick。用于无监督视觉表征学习的动量对比。收录于 *IEEE/CVF*计算机视觉与模式识别会议论文集，第9729–9738页，2020年。Jean-Bastien Grill、Florian Strub、Florent Altch´e、Corentin Tallec、Pierre Richemond、Elena Buchatskaya、Carl Doersch、Bernardo Avila Pires、Zhaohan Guo、Mohammad Gheshlaghi Azar等。自举潜在表征——一种自监督学习新方法。神经信息处理系统进展，33:21271–21284，2020年。Heming Yao、Phil Hanslovsky、Jan-Christian Huetter、Burkhard Hoeckendorf和David Richmond。弱监督集合一致性学习改进单细胞图像的形态学表征分析。收录于 *IEEE/CVF*计算机视觉与模式识别会议论文集，第6978–6987页，2024年。Yonglong Tian、Dilip Krishnan和Phillip Isola。对比多视图编码。收录于 欧洲计算机视觉会议，第776–794页。Springer，2020年。Sangwoo Mo、Minkyu Kim、Kyungmin Lee和Jinwoo Shin。S-CLIP：使用少量专家标注文本的半监督视觉语言学习。收录于 第*37*届神经信息处理系统国际会议论文集，NIPS '23，美国纽约州雷德胡克，2023年。Curran Associates Inc.。Kai Gan、Bo Ye、Min-Ling Zhang和Tong Wei。通过强制语义一致性与梯形一致性实现半监督 CLIP 适配。见Y. Yue、A. Garg、N. Peng、F. Sha和R. Yu编，国际表征学习会议，第2025卷，第12142–12163页，2025年。Karren Dai Yang、Anastasiya Belyaeva、Saradha Venkatachalapathy、Karthik Damodaran、Abigail Katcoff、Adityanarayanan Radhakrishnan、GV Shivashankar和Caroline Uhler。使用自编码器在单细胞成像与测序数据之间进行多域转换。《自然·通讯》，12(1):31，2021年。Johnny Xi、Jana Osea、Zuheng Xu和Jason Hartford。未配对多模态数据的倾向评分对齐，2024年。Jayoung Ryu、Charlotte Bunne、Luca Pinello、Aviv Regev和Romain Lopez。利用带标签的 Gromov-Wasserstein 最优传输进行跨模态匹配与扰动响应预测，2025年。Donald B Rubin。估计随机研究与非随机研究中治疗的因果效应。《教育心理学杂志》，66(5):688，1974年。Facundo M´emoli。Gromov–Wasserstein 距离及对象匹配的度量方法。《计算数学基础》，11(4):417–487，2011年。Othmane Sebbouh、Marco Cuturi和Gabriel Peyr´e。基于代价正则化最优传输的跨空间结构化变换。收录于 国际人工智能与统计会议，第586–594页。PMLR，2024年。Hugues Van Assel、C´edric Vincent-Cuaz、Nicolas Courty、R´emi Flamary、Pascal Frossard和Titouan Vayer。分布约简：利用Gromov-Wasserstein统一降维与聚类。*arXiv* 预印本 *arXiv:2402.02239*，2024年。Gabriel Peyr´e、Marco Cuturi和Justin Solomon。核矩阵与距离矩阵的Gromov-Wasserstein 平均。收录于国际机器学习会议，第2664–2672页。PMLR，2016年。Titouan Vayer。关于不可比空间上最优传输的一项贡献。*arXiv* 预印本 *arXiv:2011.04447*，2020年。Rico Sennrich、Barry Haddow和Alexandra Birch。利用单语数据改进神经机器翻译模型。*arXiv* 预印本 *arXiv:1511.06709*，2015年。Guillaume Lample、Myle Ott、Alexis Conneau、Ludovic Denoyer和Marc'Aurelio Ranzato。基于短语与神经的无监督机器翻译。*arXiv* 预印本 *arXiv:1804.07755*，2018年。Alexis Conneau、German Kruszewski、Guillaume Lample、Lo¨ıc Barrault和Marco Baroni。单个向量能装下什么：探测句子嵌入的语言学属性。*arXiv* 预印本 *arXiv:1805.01070*，2018年。Romain Lopez、Jeffrey Regier、Michael B Cole、Michael I Jordan和Nir Yosef。面向单细胞转录组学的深度生成建模。《自然·方法》，15(12):1053–1058，2018年。Ricard Argelaguet、Damien Arnol、Danila Bredikhin、Yonatan Deloro、Britta Velten、John C Marioni和Oliver Stegle。Mofa+：多模态单细胞数据全面整合的统计框架。《基因组生物学》，21 (1):111，2020年。Kevin Z Lin和Nancy R Zhang。利用倾斜典型相关分析量化单细胞多模态数据中的共有与差异信息。《美国国家科学院院刊》，120(32):e2303647120，2023年。

Aditya Gorla、Jonathan Isaac Witonsky、Zeyuan Johnson Chen、Jennifer R Elhawary、Joel Andrew Mefford、Javier Perez-Garcia、Anne-Marie Madore、Scott Huntsman、Donglei Hu、Celeste Eng 等。《表观遗传患者分层揭示了具有改变的B细胞反应的2型哮喘亚内型》。*medRxiv*，第 2025–08 页，2025a。Pinar Demetci、Rebecca Santorella、Bj¨orn Sandstede、William Stafford Noble 和 Ritambhara Singh。Scot：利用最优传输进行单细胞多组学对齐。计算生物学杂志，29(1):3–18，2022。Jie Liu、Yuanhao Huang、Ritambhara Singh、Jean-Philippe Vert 和 William Stafford Noble。联合嵌入多种单细胞组学测量。载于生物信息学算法：……国际研讨会，*WABI*……论文集。*WABI*（研讨会），第 143 卷，第 10 页，2019。Aditya Gorla、Ryan Wang、Zhengtong Liu、Ulzee An 和 Sriram Sankararaman。CACTI：利用副本掩蔽和上下文信息改进表格数据插补。载于第*42*届国际机器学习会议论文集，机器学习研究论文集第 267 卷，第 20187–20225 页。PMLR，2025年7月13–19日。Russell Littman、Jacob Levine、Sepideh Maleki、Yongju Lee、Vladimir Ermakov、Lin Qiu、Alexander Wu、Kexin Huang、Romain Lopez、Gabriele Scalia 等。利用 presage 进行基因嵌入预测和扰动表达响应的功能评估。*bioRxiv*，第 2025–06 页，2025年。Abhinav K Adduri、Dhruv Gautam、Beatrice Bevilacqua、Alishba Imran、Rohan Shah、Mohsen Naghipourfar、Noam Teyssier、Rajesh Ilango、Sanjay Nagaraj、Mingze Dong 等。利用 state 预测多种情境下的细胞扰动响应。*bioRxiv*，第 2025–06 页，2025年。Zhenguo Yang、Zehang Lin、Lingni Guo、Qing Li 和 Wenyin Liu。Mmed：一个多领域多模态事件数据集。信息处理与管理，57(6):102315，2020年。Xinhao Mei、Chutong Meng、Haohe Liu、Qiuqiang Kong、Tom Ko、Chengqi Zhao、Mark D Plumbley、Yuexian Zou 和 Wenwu Wang。Wavcaps：用于音频语言多模态研究的 ChatGPT辅助弱标注音频描述数据集。*IEEE/ACM* 音频、语音与语言处理汇刊，32:3339–3354，2024年。Xintian Sun、Benji Peng、Charles Zhang、Fei Jin、Qian Niu、Junyu Liu、Keyu Chen、Ming Li、Pohsun Feng、Ziqian Bi 等。从像素到散文：推进遥感多模态语言模型。*arXiv* 预印本 *arXiv:2411.05826*，2024年。Shengjia Zhao、Jiaming Song 和 Stefano Ermon。Infovae：平衡变分自编码器中的学习与推理。载于*AAAI* 人工智能会议论文集，第 33 卷，第 5885–5892 页，2019年。Diederik P Kingma 和 Max Welling。自编码变分贝叶斯。*arXiv* 预印本 *arXiv:1312.6114*，2013年。Jan Niklas B¨ohm、Philipp Berens 和 Dmitry Kobak。使用对比学习无监督可视化图像数据集。*arXiv* 预印本 *arXiv:2210.09879*，2022年。Tianyang Hu、Zhili Liu、Fengwei Zhou、Wenjia Wang 和 Weiran Huang。你的对比学习实际上在暗中进行随机邻域嵌入。*arXiv* 预印本 *arXiv:2205.14814*，2022年。Laurens Van Der Maaten 和 Geoffrey Hinton。使用 t-SNE 可视化数据。机器学习研究杂志，9（11月）：2579–2605，2008年。Leonid V Kantorovich。组织和规划生产的数学方法。管理科学，6(4)：366–422，1960年。Vayer Titouan、Ievgen Redko、R´emi Flamary 和 Nicolas Courty。协同最优传输。神经信息处理系统进展，33:17559–17570，2020年。

GROOVE 扩展详情

A.1 在线回译

每个模态特异性编码器 $g_\theta^{(m)}$ 由一个带有批归一化和 ReLU 激活的 3 层多层感知机构成，而每个解码器 $d_\theta^{(m)}$ 则使用一个带有额外线性输出层的 2 层多层感知机（见下一小节）。一个共享线性投影（耦合层） f_θ 将所有模态特异性编码器连接到各自的解码器（即，所有模态特异性嵌入均经过该投影），从而得到最终的编码器组合 $f_\theta^{(m)} = f_\theta \circ g_\theta^{(m)}$。
在弱配对的单细胞数据中，每种模态都可以视为一种具有共享语义信息的不同语言。我们通过自回译策略，确保潜在表示能够实现有意义的跨模态翻译，从而促进跨模态纠缠表示。令 $\bar{m}$ 表示另一种模态。给定来自模态 m 的样本潜在表示： $\boldsymbol{z}_i^{(m)} = f_\theta \circ g_\theta^{(m)}(\boldsymbol{x}_i^{(m)})$，我们分三个步骤进行跨模态翻译：

1. **跨模态生成**：使用模态 $\bar{m}$ 的解码器在推理模式下，生成 $\boldsymbol{x}^{(m\to\bar{m})}{}_i = d_\theta^{(\bar{m})}(\boldsymbol{z}_i^{(m)})$
2. **重新编码**：切换到训练模式并对生成的样本进行编码：$\boldsymbol{z}_i^{(m\to\bar{m})} = f_\theta \circ g_\theta^{(\bar{m})}(\boldsymbol{x}^{(m\to\bar{m})}{}_i)$
3. **回译**：重建原始模态： $\boldsymbol{x}_i^{(m\to\bar{m}\to m)} = d_\theta^{(m)}(\boldsymbol{z}_i^{(m\to\bar{m})})$

这一即时 过程会在每个小批次中生成合成伪配对样本，前提是潜在表示包含跨模态信息。

A.2 编码器架构与正则化

编码器的第一层将每种数据模态从其原生维度 $k^{(m)}$ 映射到最终嵌入维度的两倍（$2 \times d$）。这一设计遵循 Samaran 等人（2024年）的建议，他们采用类变分编码器架构：对于每个 d 维嵌入，编码器都会输出均值和对数方差参数。

然而，与标准变分自编码器不同，我们不优化变分证据下界（ELBO），该下界将重构损失与 Kullback-Leibler（KL）散度结合起来。相反，我们仅最小化标准重构损失（与普通自编码器类似）。这一设计选择受到先前研究（Zhao 等人，2019年）的启发；该研究表明，KL散度可能与重建目标相冲突，并降低下游推理性能——我们在内部分析中也观察到了这一现象。
我们的方法利用编码器输出的参数定义一个具有对角协方差的高斯后验分布。具体而言，我们将编码器输出的 $2d$ 维向量解释为 d 维高斯分布的均值 $\boldsymbol{\mu}$ 和对数方差 $\log \boldsymbol{\sigma}^2$ 参数。训练期间，我们使用重参数化技巧从该分布中采样（Kingma and Welling，2013）。

为确保数值稳定性并提供轻度正则化，我们仅在训练期间向对角协方差矩阵添加一个较小的固定常数（10^{-4}）。这种随机采样机制有三个重要作用：（1）在训练期间使编码器具有非确定性；（2）通过鼓励采用分布而非点估计，防止样本在潜在空间中过度拥挤；（3）为解码器提供隐式正则化，要求其根据潜在编码的分布而非确定性点进行重构，从而有助于防止过拟合。

推理期间，编码器仅使用均值参数 $\boldsymbol{\mu}$ 进行确定性运算。

A.3 解码器

解码器接收来自共享投影层 f_θ 的 d 维输出，并将其输入一个包含 1D 批归一化和 ReLU 激活函数的两层多层感知机，随后通过一个最终线性层，将嵌入从 d 维投影回原生的 $k^{(m)}$ 特征

A.4 相似性核

相似度核的选择会显著影响学习到的表征的质量和特性。虽然在大多数对比学习方法中，余弦相似度仍是标准选择，通常在超球面上学习表征（Radford 等人，2021；Chen 等人，2020），但近期研究表明，由参数化重尾t分布可在欧几里得空间中产生表达能力更强的表征（B¨ohmet al.,2022; Hu 等人，2022）。

余弦相似度。标准（余弦）形式如下：

$$\mathrm{sim}_{\cos}(\boldsymbol{a},\boldsymbol{b}) := \frac{\langle \boldsymbol{a},\boldsymbol{b}\rangle}{\|\boldsymbol{a}\|_2 \cdot \|\boldsymbol{b}\|_2} \tag{7}$$

公式 3 中的逐锚点损失可直接应用：

$$\ell_i^{(m)} = -\log \frac{\sum\limits_{\boldsymbol{z}_p \in \mathcal{P}_i^{(m)}} \exp\big(\mathrm{sim}_{\cos}(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_p)/\tau\big)}{\sum\limits_{\boldsymbol{z}_a \in \mathcal{A}_i^{(m)}} \exp\big(\mathrm{sim}_{\cos}(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_a)/\tau\big)} \tag{8}$$

其中 τ 控制温度。

***t* 分布相似度**。t 分布核使用重尾衰减来衡量欧几里得距离：

$$\mathrm{sim}_t(\boldsymbol{a},\boldsymbol{b}) := \left(1 + \frac{\|\boldsymbol{a}-\boldsymbol{b}\|_2^2}{\tau \cdot \eta}\right)^{-\frac{\eta+1}{2}} \tag{9}$$

该核有两个作用不同的参数。自由度 η 控制分布的形状，具体来说是尾部的厚重程度。设置为 $\eta = 1$时，得到用于 t-SNE（Van Der Maaten 和 Hinton，2008）的柯西核；这是本研究的默认设置 η。温度 τ 控制相似度函数的尺度，决定相似度随距离增加而衰减的速度。

由于公式 9 已包含温度缩放，并映射到 $(0, 1]$,，这些性质使得指数变换不再必要。逐锚点损失（公式 3）可简化为：

$$\ell_i^{(m)} = -\log \frac{\sum\limits_{\boldsymbol{z}_p \in \mathcal{P}_i^{(m)}} \mathrm{sim}_t(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_p)}{\sum\limits_{\boldsymbol{z}_a \in \mathcal{A}_i^{(m)}} \mathrm{sim}_t(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_a)} \tag{10}$$

A.5 小批次平衡欠采样策略

为解决 GroupCLIP 框架中可能存在的类别不平衡问题，我们采用平衡欠采样策略，确保每个批次中的所有类别均衡呈现。算法首先计算第一模态中每个标签 l 对应的实例数n_l，并假设两个模态的标签分布相同。随后，我们找出实例数最少的少数类： $n_{\min} = \min_l (n_l)$。

给定初始批次大小 B 和标签数量 L，我们使用以下公式计算有效批次大小： $B_{\text{eff}} = B -$（$B \mod L$），其中每个标签在每个批次中恰好贡献 B_{eff}/L 个样本。这确保批次大小能被标签数量整除。重要的是，我们的采样策略遵循少数类标签约束，在整个训练过程中从任何标签（即 $B_{\text{eff}}/L \le n_{\min}$）采样的实例数从不超过 $n_{\min}$，从而避免对少数类标签进行过采样。

对于每个批次，我们从两种模态中独立地为每个标签随机采样 B_{eff}/L 个实例，在保持数据弱配对特性的同时确保表示均衡。该方法可防止主导标签压倒对比学习信号，并确保所有标签对学习到的表征贡献均等，这在处理不平衡的多模态数据集时尤为重要。

A.6 算法概要

算法 1GROOVE 训练流程

要求： 模态特定数据集 $\mathcal{D}^{(1)}$, $\mathcal{D}^{(2)}$
要求： 编码器 $g_\theta^{(1)}$, $g_\theta^{(2)}$、解码器 $d_\theta^{(1)}$, $d_\theta^{(2)}$、共享投影 f_θ
要求： 超参数 α (GroupCLIP), β (重建)
对于当尚未达到最大训练迭代次数时**执行**
  采样平衡小批量 $(\boldsymbol{x}^{(1)}, t^{(1)}) \sim \mathcal{D}^{(1)}$, $(\boldsymbol{x}^{(2)}, t^{(2)}) \sim \mathcal{D}^{(2)}$ ▷ 参见第A.5节
  步骤1：模态内重建与对比对齐
  $\boldsymbol{z}^{(1)} \leftarrow f_\theta(g_\theta^{(1)}(\boldsymbol{x}^{(1)}))$
  $\boldsymbol{z}^{(2)} \leftarrow f_\theta(g_\theta^{(2)}(\boldsymbol{x}^{(2)}))$
  $\hat{\boldsymbol{x}}^{(1)} \leftarrow d_\theta^{(1)}(\boldsymbol{z}^{(1)})$
  $\hat{\boldsymbol{x}}^{(2)} \leftarrow d_\theta^{(2)}(\boldsymbol{z}^{(2)})$
  $\mathcal{L}_{\text{recon}} \leftarrow \sum_{m\in\{1,2\}} \text{均方误差}(\boldsymbol{x}^{(m)}, \hat{\boldsymbol{x}}^{(m)})$
  $\mathcal{L}_{\text{GroupCLIP}} \leftarrow \frac{1}{2|\mathcal{D}_z^{(1)}|}\sum_{\boldsymbol{z}_i^{(1)}\in\mathcal{D}_z{}^{(1)}} \ell_i^{(1)} + \frac{1}{2|\mathcal{D}_z^{(2)}|}\sum_{\boldsymbol{z}_i^{(2)}\in\mathcal{D}_z^{(2)}} \ell_i^{(2)}$ ▷参见公式 4
  $\mathcal{L}^{\text{step1}} \leftarrow \beta\cdot\mathcal{L}_{\text{recon}} + \alpha\cdot\mathcal{L}_{\text{GroupCLIP}}$
  使用来自 $\mathcal{L}^{\text{step1}}$ 的梯度更新参数 θ
  步骤 2：在线回译
  将 $g_\theta^{(1)}$, $g_\theta^{(2)}$, $d_\theta^{(1)}$, $d_\theta^{(2)}$, f_θ 设为评估 ▷ 在不计算梯度的情况下生成跨模态样本
  $\boldsymbol{x}^{(1\to2)} \leftarrow d_\theta^{(2)} \circ f_\theta \circ g_\theta^{(1)}(\boldsymbol{x}^{(1)})$
  $\boldsymbol{x}^{(2\to1)} \leftarrow d_\theta^{(1)} \circ f_\theta \circ g_\theta^{(2)}(\boldsymbol{x}^{(2)})$
  设置 $g_\theta^{(1)}$ $g_\theta^{(2)}$ $d_\theta^{(1)}$, $d_\theta^{(2)}$, f_θ 以进行训练
  $\boldsymbol{z}^{(1\to2)} \leftarrow f_\theta(g_\theta^{(2)}(\boldsymbol{x}^{(1\to2)}))$ ▷ 重新编码翻译样本
  $\boldsymbol{z}^{(2\to1)} \leftarrow f_\theta(g_\theta^{(1)}(\boldsymbol{x}^{(2\to1)}))$
  $\hat{\boldsymbol{x}}^{(1\to2\to1)} \leftarrow d_\theta^{(1)}(\boldsymbol{z}^{(1\to2)})$ ▷ 重构原始模态
  $\hat{\boldsymbol{x}}^{(2\to1\to2)} \leftarrow d_\theta^{(2)}(\boldsymbol{z}^{(2\to1)})$
  $\mathcal{L}_{\text{bt}} \leftarrow \frac{1}{2}[\text{均方误差}(\boldsymbol{x}^{(1)}, \hat{\boldsymbol{x}}^{(1\to2\to1)}) + \text{均方误差}(\boldsymbol{x}^{(2)}, \hat{\boldsymbol{x}}^{(2\to1\to2)})]$
  $\mathcal{L}^{\text{step2}} \leftarrow \beta\cdot\mathcal{L}_{\text{bt}}$
  更新参数 θ 使用来自 $\mathcal{L}^{\text{step2}}$的梯度
返回 训练后的编码器 $f_{(1)\theta}$, $f_\theta^{(2)}$

B 基线详情

基线实现细节 所有方法均使用相同的编码器和解码器架构[2]，并在所有实验中采用一致的潜在嵌入维度 $d = 128$。对于 DAVAE，我们采用此前报告的最优超参数设置（Ryu 等人，2025年）。我们使用余弦相似度核和 t 分布（tdist）相似度核对 GROOVE 进行评估，在所有实验中设置 $\alpha = 1.0,\ \tau = 0.2,$ ，并针对单细胞数据训练设置 $\beta = 0.1$。需要注意的是，我们未进行严格的超参数探索，因此不声称所报告的结果代表 GROOVE 在任何受评估数据集上的最佳性能。
我们结合五种最优传输方法对所有表示学习器进行了评估：两种标准方法和三种受标签约束的变体。标准最优传输方法包括 EOT 和 EGWOT（Peyr´e 等人，2016年；Kantorovich，1960年）。受标签约束的方法包括：labeledEOT、labeledEGWOT 和 labeledCOOT（协同最优传输）（Ryu 等人，2025年；Titouan 等人，2020年）。在所有分析中，我们使用 Perturb-OT 软件包中的默认熵正则项设置[3]。

C 模拟详情

为模拟复杂的多模态细胞数据，我们定义了一个概率模型，用以捕捉共享和模态特异性的潜在结构。该模型首先定义一个共享潜在空间和两个非共享的模态特异性潜在空间，然后将它们组合起来。使用以下扰动类型：

1. 共享扰动：通过共享潜在空间协同影响两种模态，且效应大小和细胞特异性外显率值相同 2. 模态特异性扰动：分别针对每种模态的独有维度，并采用不同的效应大小和外显率参数

我们使用以下设置初始化模拟：

- 潜在维度：10
- 共享变异比例：– 100% 共享：10 个共享维度，0 个独有维度– 80% 共享：8 个共享维度，每种模态各有 2 个独有维度– 50% 共享：5 个共享维度，每种模态各有 5 个独有维度

- 实验设计：9种扰动 + 1 对照条件
- 样本量：每种条件下每个模态包含100个细胞
- 特征数量：模态1和模态2分别包含1000个和500个观测特征

C.1 生成模型

我们首先定义潜变量。对于 n总细胞数和 d_s 共享维度，共享潜变量 Z 的每个系数均从缩放标准正态分布中采样：

$$Z_{ij} \sim \mathcal{N}(0, \text{scale}^2) \quad \text{for } i = 1, \ldots, n,\ j = 1, \ldots, d_s$$

其中，scale表示全局潜在信号强度（默认为0.1）。
类似地，我们定义两个非共享潜变量 U_X 和 U_Y，它们具有 d_u 个非共享维度：

$$U_{X,ij} \sim \mathcal{N}(0, \text{scale}^2), \quad U_{Y,ij} \sim \mathcal{N}(0, \text{scale}^2).$$

将它们拼接起来，形成每种模态的完整潜在表示：

$$V_X = [Z \,\|\, U_X], \quad V_Y = [Z \,\|\, U_Y],$$

总维度为 $d = d_s + d_u$。

[2]PS 无需解码器 [3]
https://github.com/Genentech/Perturb-OT

每种模态都对应一个变换矩阵。对于具有 p_X 个特征的模态 X，$A_X \in \mathbb{R}^{d \times p_X}$ 的系数被重新采样为

$$(A_X)_{jk} \sim \mathcal{N}(0,1).$$

同样，对于具有 p_Y 特征的模态 Y：

$$(A_Y)_{jk} \sim \mathcal{N}(0,1).$$

偏置向量 $b_X \in \mathbb{R}^{p_X}$ 和 $b_Y \in \mathbb{R}^{p_Y}$ 具有系数

$$(b_X)_j \sim \mathcal{N}(0,1), \quad (b_Y)_j \sim \mathcal{N}(0,1).$$

缩放因子 $s_X \in \mathbb{R}^{p_X}$ 和 $s_Y \in \mathbb{R}^{p_Y}$ 控制特征变异性：

$$(s_X)_j \sim \Gamma(1,1), \quad (s_Y)_j \sim \Gamma(1,1).$$

为考虑模态特异性噪声，扰动参数按系数定义 :

$$\begin{aligned} \mu_{X,j} \sim \mathcal{N}(0,1), \quad & \mu_{Y,j} \sim \mathcal{N}(0,1), \\ \text{offsetsd}_{X,j} \sim \mathcal{N}(0,1), \quad & \text{offsetsd}_{Y,j} \sim \mathcal{N}(0,1), \\ \sigma_{X,j} = \exp(-3.0 + \text{offsetsd}_{X,j}), \quad & \sigma_{Y,j} = \exp(-3.0 + \text{offsetsd}_{Y,j}). \end{aligned}$$

随后，每个细胞 i和特征 j 的噪声为：

$$\xi_{X,ij} \sim \mathcal{N}(\mu_{X,j}, \sigma_{X,j}^2) \cdot \frac{\text{scale}}{\text{snr}}, \quad \xi_{Y,ij} \sim \mathcal{N}(\mu_{Y,j}, \sigma_{Y,j}^2) \cdot \frac{\text{scale}}{\text{snr}},$$

其中 snr 表示信噪比（默认值为 0.2）。

接下来，我们在 $L+1$ 个条件（包括一个对照）下引入 L 个扰动。扰动以循环方式作用于共享或非共享潜在维度。

对于**共享扰动**，目标维度 n is

$$t_s(I) = ((I-1) \bmod d_s).$$

效应大小采样如下

$$|e_s(I)| \sim \max(3, \Gamma(1,1)), \quad \text{sign}_s(I) \sim 2 \cdot \text{Bernoulli}(0.5) - 1,$$
$$e_s(I) = \text{sign}_s(I) \cdot |e_s(I)|.$$

每个细胞 i 的外显率为

$$q_{s,i} \sim \text{Beta}(1,10).$$

对于处于扰动 I 下的细胞，目标潜在维度在两种模态中均发生偏移：

$$v'_{X,it_s(I)} = v_{X,it_s(I)} + e_s(I) q_{s,i}, \quad v'_{Y,it_s(I)} = v_{Y,it_s(I)} + e_s(I) q_{s,i}.$$

对于**模态特异性扰动**，目标维度为

$$t_u(I) = ((I-1) \bmod d_u) + d_s.$$

每种模态的效应大小分别定义为：

$$|e_{uX}(I)| \sim \max(3, \Gamma(1,1)), \quad \text{sign}_{uX}(I) \sim 2 \cdot \text{Bernoulli}(0.5) - 1, \quad e_{uX}(I) = \text{sign}_{uX}(I)|e_{uX}(I)|,$$
$$|e_{uY}(I)| \sim \max(3, \Gamma(1,1)), \quad \text{sign}_{uY}(I) \sim 2 \cdot \text{伯努利分布}(0.5) - 1, \quad e_{uY}(I) = \text{符号}_{uY}(I)|e_{uY}(I)|.$$

在具有外显率的情况下

$$q_{uX,i} \sim \text{Beta}(1,10), \quad q_{uY,i} \sim \text{Beta}(1,10),$$

潜变量彼此独立地受到扰动：

$$v''_{X,it_u(I)} = v'_{X,it_u(I)} + e_{uX}(I) q_{uX,i}, \quad v''_{Y,it_u(I)} = v'_{Y,it_u(I)} + e_{uY}(I) q_{uY,i}.$$

令 V_X^{pert} 和 $V_{\text{扰动}Y}$表示最终的潜在空间。观测数据的生成方式如下

$$X_{ij} = \Big(((V_X^{\text{pert}} + \xi_X) A_X + b_X) \odot s_X \Big)_{ij}, \quad Y_{ij} = \Big(((V_Y^{\text{pert}} + \xi_Y) A_Y + b_Y) \odot s_Y \Big)_{ij},$$

其中 $\odot$表示逐元素乘法。

该模型使用 Pyro 实现，并使用 Pyro 生成模拟数据。

D Perturb-Multiome 预处理

我们从（Martin-Rufino 等，2025年）下载了 Perturb-Multiome 数据。我们聚焦于第14天的数据，因为这一天的细胞类型异质性最低。为降低 ATAC-seq 数据的维度和稀疏性，我们筛选出至少在 1% 的细胞中测得的 ATAC-seq 峰，并使用 Gencode M38 注释文件将峰映射到其 100 kb 范围内最近的基因。我们使用逆文档频率和奇异值分解，将 ATAC 测序数据降至 256 个成分。对于 RNA 测序数据，我们将计数归一化至每个细胞 10,000，并进行 log1p 变换。我们筛除了在少于 1% 的细胞中表达的基因。随后，我们使用 Seurat 方法（通过 scanpy 实现）选取排名前 512 的高变基因。我们从每种扰动中选取 128 个细胞，并将数据拆分为 RNA 测序和 ATAC 测序数据，用于跨模态表征和预测。我们测试了各方法利用 ATAC-seq 数据对 RNA 测序数据进行插补的准确性。

E Perturb-CITE-seq 预处理

我们从 Franghei 等（Frangieh 等，2021年）下载了 Perturb-CITE-seq 数据。我们聚焦于 IFNy 条件，并将数据子集限定为该条件。数据被拆分为 RNA 测序和 CITE-seq，以便分别进行预处理。对于 RNA 测序，我们将每个细胞的总计数归一化至 10,000，并应用了 log1p 变换。对于两种模态，我们减去了具有对照扰动的细胞的平均表达，因此基因或蛋白质表达矩阵中的所有值均为相对于平均对照的相对变化。为减少扰动数量，我们计算了每个目标基因相对于对照的能量距离，并筛选出至少包含 50 个细胞且能量距离为 0.05 的扰动。由此得到 18 个用于下游任务的扰动。最后，我们使用 scanpy 中实现的 Seurat 方法，将基因表达子集化为前 500 个高变基因。随后，我们测试了各方法利用 CITE-seq 数据向 RNA 测序数据中进行插补的准确性。

F 指标扩展细节

我们评估跨模态。 匹配和预测性能，这些指标能够捕捉不同的 t
对齐质量的方面以及 分布相似性。

迹指标 假设样本索引对应真实匹配，我们计算正确匹配上的平均权重，即传输计划 $\boldsymbol{T}$ 的归一化迹：

$$\text{Trace}(\boldsymbol{T}) = \frac{1}{n}\text{Tr}(\boldsymbol{T}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{T}_{ii} \tag{11}$$

传输计划 $\boldsymbol{T}$ 首先按行归一化，使得对于所有 i，$\sum_j \boldsymbol{T}_{ij} = 1$。均匀随机匹配为每个细胞分配$\boldsymbol{T}_{ij} = 1/n$，得到 Trace($\boldsymbol{T}$) $= 1/n$。完美匹配得到 Trace($\boldsymbol{T}$) $= 1$。

重心 FOSCTTM 我们使用重心投影计算比真实匹配更近的样本比例。给定匹配矩阵 $\boldsymbol{T}$和目标数据 $\boldsymbol{X}^{(1)}$，我们进行投影，得到 $\hat{}\boldsymbol{X}^{(1)} = \boldsymbol{T}\boldsymbol{X}^{(1)}$。对于每个投影样本 $\hat{}\mathbf{x}_i^{(1)}$，我们计算它与$\mathbf{X}^{(1)}$ 中的所有样本之间的欧几里得距离，并计算比真实匹配更近的样本比例：

$$\text{FOSCTTM}(\boldsymbol{T}, \boldsymbol{X}^{(1)}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n-1}\sum_{j\neq i}\mathbf{1}\{d(\hat{\boldsymbol{x}}_i^{(1)}, \boldsymbol{x}_j^{(1)}) < d(\hat{\boldsymbol{x}}_i^{(1)}, \boldsymbol{x}_i^{(1)})\}, \tag{12}$$

其中，$d(\cdot,\cdot)$ 表示欧几里得距离。最终报告的对称重心 FOSCTTM 是两种模态上的平均值：$0.5 \times$ (FOSCTTM($\boldsymbol{T}$, $\boldsymbol{X}^{(1)}$) $+$FOSCTTM($\boldsymbol{T}^\top, \boldsymbol{X}^{(2)}$))
数值越低表示匹配质量越好，随机匹配的预期值为 0.5 .

均方误差（MSE）对于直接预测评估，我们计算真实样本$\mathbf{X}^{(1)}$ 与预测样本 $\hat{}\mathbf{X}^{(1)}$之间的 MSE：

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(1)}\|_2^2. \tag{13}$$

我们报告所有特征的均值均方误差。

1-瓦瑟斯坦距离（WD）为评估分布相似性，我们计算真实样本与预测样本之间的1-瓦瑟斯坦距离，并对所有特征取平均:

$$\text{WD} = \frac{1}{d}\sum_{j=1}^{d} W_1(\boldsymbol{X}^{(1)}_{:,j}, \hat{\boldsymbol{X}}^{(1)}_{:,j}), \tag{14}$$

其中， W_1 表示单变量分布之间的1-瓦瑟斯坦距离，$\boldsymbol{X}_{:,\ j}$ 表示第 j个特征列。

余弦相似度 我们　　计算对应的真实特征向量与预测特征向量之间的平均余弦相似度　　s:

$$\text{Cosine} = \frac{1}{d}\sum_{j=1}^{d} \frac{\boldsymbol{X}^{(1)}_{:,j} \cdot \hat{\boldsymbol{X}}^{(1)}_{:,j}}{\|\boldsymbol{X}^{(1)}_{:,j}\|_2 \|\hat{\boldsymbol{X}}^{(1)}_{:,j}\|_2}, \tag{15}$$

其中，点积和范数均针对每个特征跨样本计算。

基于 KNN 的指标 为评估邻域保持，我们使用余弦相似度为真实数据和预测数据分别构建 k近邻图。设 $\boldsymbol{G}^{\text{true}}$ 和 $\boldsymbol{G}^{\text{pred}}$ 分别表示对应 KNN 图的二元邻接矩阵（其中 $k = 10$）。对于每个样本 i，我们将$\boldsymbol{G}^{\text{true}}_{i:,}$ 视为真实标签，并将$\boldsymbol{G}^{\text{pred}}_{i\ ,\ :}$ 视为预测结果，然后计算:

KNN 召回率：正确识别出的真实邻居所占的比例:

$$\text{KNN Recall} = \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_j \boldsymbol{G}^{\text{true}}_{i,j}\boldsymbol{G}^{\text{pred}}_{i,j}}{\sum_j \boldsymbol{G}^{\text{true}}_{i,j}}. \tag{16}$$

KNN 平均值 精确率（KNN PR）： 每个样本邻域预测的平均精度分数:

$$\text{KNN PR} = \frac{1}{n}\sum_{i=1}^{n} \text{AP}(\boldsymbol{G}^{\text{true}}_{i,:}, \boldsymbol{G}^{\text{pred}}_{i,:}), \tag{17}$$

其中 AP 表示平均精度分数。

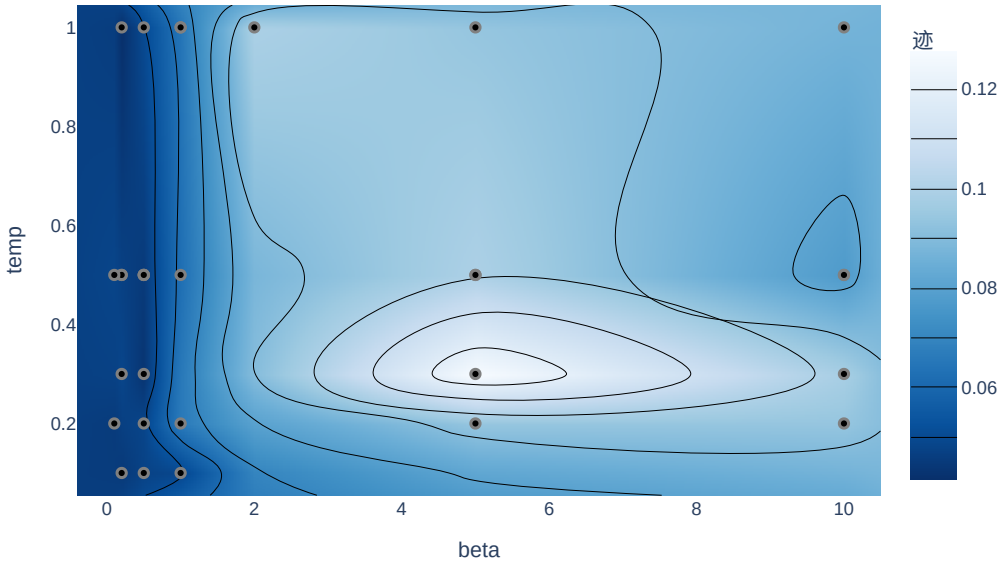
KNN ROC-AUC（KNN ROC）： 邻域预测的 ROC 曲线下面积　　:

$$\text{KNN ROC} = \frac{1}{n}\sum_{i=1}^{n} \text{AUC}(\boldsymbol{G}^{\text{true}}_{i,:}, \boldsymbol{G}^{\text{pred}}_{i,:}), \tag{18}$$

其中 AUC 表示受试者工作特征曲线下面积。

G 模拟扩展结果

(a)



(b)

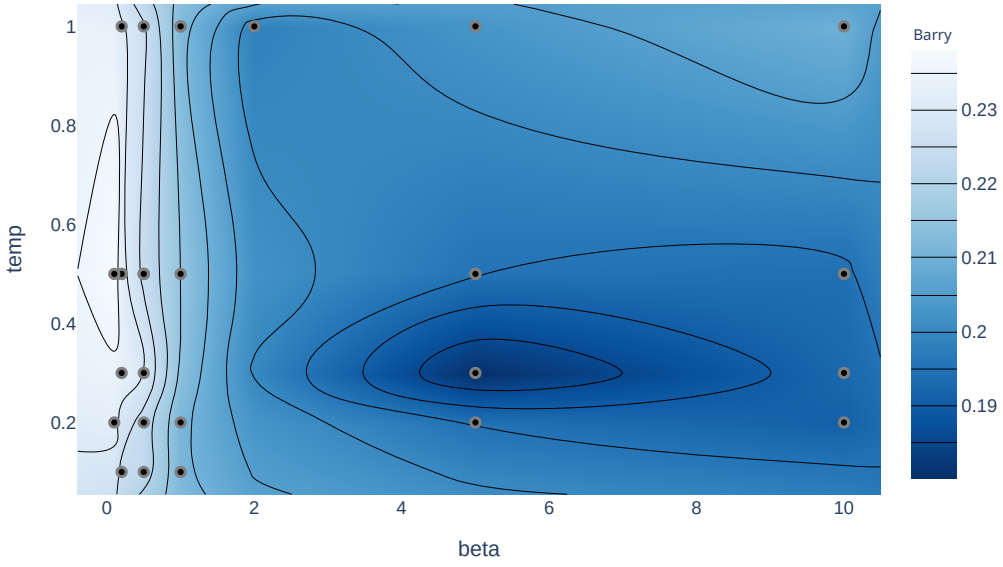


图 2：匹配性能的超参数敏感性图谱。等高线图显示了在每种共享变异设置（100%、80% 和 50%）下，β（x 轴）与 τ（y 轴）各组合的平均性能；使用基于 Optuna 的超参数搜索进行分析。 $\alpha = 1$ 用于本研究的所有分析。**(a)**基于 Trace 的匹配性能（越高越好）。**(b)** 重心 FOSCTTM（越低越好）。

H Perturb-CITE-seq 扩展结果

表 6：采用五折评估的 Perturb-CITE-seq 数据集中排名前 10 的方法组合的匹配性能指标。标准误见 ±；最佳结果以粗体标出，次佳结果加下划线。

方法	Mean Rank	迹	重心 FOSCTTM
GROOVE（余弦）+LabeledEOT	3.0	0.039±0.002	**0.381±0.004**
倾向评分+ LabeledEOT	3.0	0.039±0.002	0.383±0.002
GROOVE (tdist)+ LabeledEGWOT	3.5	0.033±0.001	0.365±0.001
GROOVE（余弦）+LabeledCOOT	5.0	**0.040±0.003**	0.451±0.004
GROOVE（余弦）+LabeledEGWOT	5.0	0.031±0.001	0.369±0.001
GROOVE（tdist）+ LabeledEOT	5.5	0.036±0.001	0.394±0.001
DAVAE+ LabeledEGWOT	7.5	0.024±0.000	0.383±0.003
DAVAE+ LabeledEOT	7.5	0.032±0.001	0.397±0.011
GROOVE（tdist）+ LabeledCOOT	7.5	0.036±0.002	0.452±0.004
倾向评分+ LabeledEGWOT	9.0	0.023±0.000	0.389±0.001

表 7：Perturb-CITE-seq 数据集上留一扰动外评估中排名前 10 的方法组合的匹配性能指标。标准误见 ±；最佳结果以粗体标示。

方法	Mean Rank	迹	重心 FOSCTTM
PS+ LabeledEOT	2.5	0.008±0.001	**0.487±0.004**
GROOVE（tdist）+ LabeledEOT	4.5	0.008±0.002	0.490±0.003
GROOVE（余弦）+ EGW	6.0	0.008±0.001	0.491±0.003
GROOVE（tdist）+ EOT	6.0	0.007±0.001	0.485±0.003
GROOVE（余弦）+ EOT	6.5	0.007±0.001	0.489±0.003
GROOVE（余弦）+ LabeledCOOT	6.5	**0.010±0.002**	0.495±0.003
倾向评分+ EOT	6.5	0.007±0.001	0.488±0.003
GROOVE（余弦）+ LabeledEOT	7.0	0.007±0.001	**0.487±0.003**
DAVAE+ EOT	7.0	0.008±0.002	0.491±0.003
DAVAE+ LabeledCOOT	7.5	0.008±0.002	0.495±0.002

表8：在Perturb-CITE-seq 数据集上采用留一扰动评估时，排名前10的方法组合的插补性能指标。标准误紧随 ±；最优以粗体标示，次优加下划线，同质指标不作标注。

方法	Mean Rank	MSE	余弦相似度	KNN 召回率	KNN PR	KNN ROC	WD
GROOVE（余弦）+ LabeledEOT	7.00	0.262±0.001	0.057±0.007	0.075±0.008	0.072±0.008	0.505±0.001	0.352±0.001
GROOVE (tdist)+ LabeledEGWOT	7.33	0.262±0.001	0.065±0.008	0.074±0.008	0.072±0.008	0.505±0.000	0.355±0.001
DAVAE+ LabeledEOT	7.83	0.262±0.001	0.082±0.008	0.074±0.008	0.073±0.008	0.504±0.001	0.354±0.000
GROOVE（tdist）+ LabeledCOOT	8.00	0.282±0.001	0.021±0.003	0.077±0.008	0.072±0.008	0.506±0.001	0.300±0.001
PS+ LabeledEOT	8.17	0.263±0.001	0.070±0.008	0.074±0.008	0.072±0.008	0.505±0.001	0.348±0.001
GROOVE（余弦）+ EGW	8.67	0.262±0.001	0.085±0.010	0.074±0.008	0.072±0.008	0.504±0.001	0.363±0.000
DAVAE+ LabeledCOOT	9.17	0.290±0.002	0.031±0.003	0.074±0.007	0.073±0.008	0.504±0.001	**0.289±0.001**
GROOVE（余弦）+ LabeledEGWOT	9.67	0.262±0.001	0.066±0.008	0.073±0.008	0.072±0.008	0.504±0.001	0.355±0.000
GROOVE（tdist）+ LabeledEOT	9.67	0.262±0.001	0.054±0.007	0.074±0.008	0.072±0.008	0.505±0.001	0.352±0.001
DAVAE+ LabeledEGWOT	10.33	0.262±0.001	**0.088±0.009**	0.071±0.008	0.073±0.008	0.502±0.001	0.354±0.001

表 9：Perturb-CITE-seq 数据集中 GROOVE 消融分析的性能指标。标准误紧随 ±；最佳值以粗体标示，次佳值以下划线标示，同质度量不作标注。

消融类型	重心 FOSCTTM	MSE	余弦相似度	KNN 召回率	KNN PR	KNN ROC	WD
GROOVE（余弦）	**0.368±0.005**	0.261±0.001	0.047±0.002	0.019±0.001	0.017±0.000	0.502±0.000	0.353±0.000
No GroupCLIP	0.380±0.003	0.261±0.001	0.048±0.001	0.019±0.000	0.017±0.000	0.502±0.000	0.353±0.001
仅自编码器	0.381±0.002	0.261±0.001	0.048±0.004	0.018±0.000	0.017±0.000	0.502±0.000	0.353±0.001

I Perturb-Multiome 扩展结果

表 10：采用五折评估的 Perturb-Multiome 数据集中排名前 10 的方法组合的插补性能指标。标准误见 ±；最佳结果以粗体标出，次佳结果加下划线。

方法	Mean Rank	MSE	余弦相似度	KNN 召回率	KNN PR	KNN ROC	WD
GROOVE（余弦）+ EOT	4.50	**0.308±0.003**	0.098±0.028	0.052±0.005	0.029±0.001	0.515±0.003	0.428±0.005
GROOVE（余弦）+ LabeledEOT	5.17	0.311±0.001	0.075±0.004	**0.066±0.003**	**0.032±0.001**	**0.523±0.002**	0.433±0.001
倾向评分+ EOT	6.17	**0.308±0.003**	**0.113±0.029**	0.043±0.004	0.027±0.001	0.511±0.002	0.426±0.004
GROOVE（tdist）+ LabeledEOT	6.50	0.311±0.001	0.071±0.002	**0.066±0.003**	**0.032±0.001**	**0.523±0.002**	0.433±0.001
GROOVE（tdist）+ EOT	7.33	0.310±0.002	0.090±0.022	0.047±0.007	0.028±0.002	0.513±0.003	0.430±0.003
GROOVE（余弦）+ LabeledCOOT	7.50	0.355±0.023	0.091±0.054	0.045±0.005	0.028±0.001	0.512±0.002	0.323±0.008
倾向评分+ LabeledEOT	7.50	0.311±0.001	0.070±0.002	0.058±0.003	0.030±0.001	0.518±0.001	0.433±0.001
GROOVE（tdist）+LabeledCOOT	9.33	0.364±0.029	0.047±0.079	0.048±0.005	0.028±0.001	0.514±0.003	**0.317±0.011**
倾向评分+ LabeledCOOT	9.50	0.334±0.006	0.041±0.025	0.048±0.005	0.028±0.001	0.514±0.002	0.379±0.001
倾向评分+ LabeledEG WOT	10.00	0.311±0.001	0.075±0.003	0.037±0.006	0.026±0.001	0.508±0.003	0.430±0.001

表 11：Perturb-Multiome 数据集中的最佳方法组合匹配

扰动输出评估

Perturb-Multiome 数据集中前 10 个方法组合的性能指标，并开启 leave e 评估。标准误遵循 ±；最佳结果以粗体显示，次佳结果加下划线。

方法	Mean Rank	迹	重心 FOSCTTM
GROOVE（tdist）+ LabeledCOOT	1.5	**0.014±0.003**	0.481±0.019
GROOVE（tdist）+ LabeledEOT	3.0	0.008±0.000	**0.480±0.004**
GROOVE（余弦）+ EOT	4.5	0.008±0.000	0.487±0.004
DAVAE+ LabeledCOOT	4.5	0.010±0.002	0.488±0.009
GROOVE（余弦）+ LabeledEOT	5.0	0.008±0.000	0.486±0.003
GROOVE（tdist）+ EOT	7.0	0.008±0.000	0.487±0.003
GROOVE（tdist）+ LabeledEGWOT	7.0	0.008±0.000	0.496±0.005
GROOVE（余弦）+ LabeledCOOT	7.5	0.011±0.003	0.500±0.018
倾向评分+ EOT	9.5	0.008±0.000	0.497±0.002
DAVAE+ LabeledEOT	10.5	0.008±0.000	0.494±0.002

表 12：Perturb-Multiome 数据集中采用五折评估的前 10 种方法组合的匹配性能指标。标准误遵循 ±；最佳结果以粗体显示，次优结果加下划线。

方法	Mean Rank	迹	重心 FOSCTTM
GROOVE（余弦）+LabeledCOOT	2.0	0.048±0.009	0.431±0.028
GROOVE（余弦）+LabeledEOT	2.5	0.041±0.001	**0.427±0.008**
GROOVE（tdist）+ LabeledEOT	4.5	0.040±0.000	0.441±0.003
DAVAE+ LabeledCOOT	6.0	**0.052±0.010**	0.470±0.025
GROOVE（tdist）+ LabeledEGWOT	6.0	0.040±0.000	0.448±0.003
GROOVE（余弦）+LabeledEGWOT	6.5	0.039±0.001	0.446±0.003
GROOVE（tdist）+ LabeledCOOT	6.5	0.047±0.017	0.457±0.050
PS+ LabeledEOT	7.0	0.039±0.000	0.445±0.003
DAVAE+ LabeledEGWOT	8.5	0.039±0.001	0.448±0.003
倾向评分+ LabeledEG WOT	8.5	0.039±0.001	0.447±0.003