Multi-omic Latent Interaction Modelling at Single-Cell Resolution

Extending Latent Interaction Variational Inference (LIVI) Model with Protein Modality

Author: Jakub Fręchowicz Responsible Professor: Marcel Reinders Supervisors: Inez den Hond, Kirti Biharie

Delft University of Technology, The Netherlands

Glossary

- **scRNA-seq** (single-cell RNA sequencing) – method to obtain gene expression measurements at a single-cell resolution.
- **eQTL** (expression Quantitative Trait Locus) – a single nucleotide polymorphism that affects gene expression; trans-eQTL affect genes further away from the polymorphism.
- **VAE** (Variational Autoencoder) – a neural network architecture able to compress and reconstruct data, where the encoder and decoder are both neural networks.
- **CITE-seq** – a method that measures cell surface proteins and mRNA in a single cell.

1. Introduction

Background: The analysis of data that comes from scRNA-seq is promising in terms of new insights about how organisms work at a lower level. Studying how to perform such an analysis is important because given human data, we can potentially gain new knowledge about unknown biological mechanisms in our physiology.

Problem: scRNA-seq data is highly dimensional and sparse. Previous methods were lacking in terms of their scalability, interpretability, or ability to find trans eQTL. One of the recent methods, LIVI [1], proves promising because it uses VAE architecture for scalability with linear decoders for better interpretability. It extends VAE by separating cell-state and donor variation, also modelling their interaction. What is missing in the LIVI model, and what is also mentioned as its possible extension, is enriching the cell-state latent space with data from other modalities.

Research Question: How can LIVI be extended to jointly model RNA and protein data, and does this improve the biological interpretability of its latent representations?

2. Multimodal LIVI Models

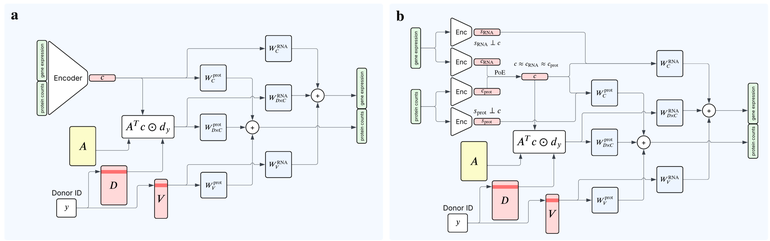

Figure 3: **Diagrams of the architecture of**: (**a**) **MultiSLIVI** (**b**) **DMLIVI**. In MultiSLIVI, gene expression and protein counts are encoded together into a cell-state embedding $\mathbf{c}$, whereas in DMLIVI each modality is encoded into a modality-specific part $\mathbf{s}$ and a shared part $\mathbf{c}$. In DMLIVI the shared parts are aligned using product of experts (PoE), and they are kept dissimilar to the modality-specific embeddings. In both, the cell-donor interaction is modelled with the term $\mathbf{A}^T\mathbf{c} \odot \mathbf{d}_y$, where $\mathbf{d}_y$ is a donor embedding, and $\mathbf{A}$ is the assignment matrix. Each modality can be reconstructed by multiplying the cell-state embedding, the cell-donor interaction, and the persistent donor effects ($\mathbf{V}$) with appropriate modality-specific linear decoders ($\mathbf{W}$) and adding the results.

2.1. MultiSLIVI

MultiSLIVI (**Multi**modal **S**hared-space **L**atent **I**nteraction **V**ariational **I**nference) extends the architecture of the original LIVI with a few changes:

1. Encoder: Same architecture (multilayer perceptron), however its input is concatenated RNA and protein measurements.
2. Decoder: All decoders for each of the latent spaces are now split into two, one for each modality.

Most importantly, there are no new latent variables, and the interaction term stays unchanged. To reconstruct the gene expression vector for a cell i, we compute:

$$\hat{\mathbf{x}}_i^{\text{RNA}} = \mathbf{W}_C^{\text{RNA}}\mathbf{c}_i + \mathbf{W}_{D\times C}^{\text{RNA}}\mathbf{z}_i^{D\times C} + \mathbf{W}_V^{\text{RNA}}\mathbf{v}_y$$

where $\mathbf{z}_i^{D\times C} = (\mathbf{A}^T\texttt{Softmax}(\mathbf{c}_i)) \odot \mathbf{d}_y$. Similarly, to reconstruct protein counts:

$$\hat{\mathbf{x}}_i^{\text{prot}} = \mathbf{W}_C^{\text{prot}}\mathbf{c}_i + \mathbf{W}_{D\times C}^{\text{prot}}\mathbf{z}_i^{D\times C} + \mathbf{W}_V^{\text{prot}}\mathbf{v}_y$$

2.2. DMLIVI

DMLIVI (**D**isentangled **M**ultimodal **L**atent **I**nteraction **V**ariational **I**nference) combines the architecture of the original LIVI with the disentanglement principles from Disentangled Multimodal Variational Autoencoders (DMVAE) [2]. For each modality we learn a shared (c) and a modality-specific embedding (s). To separate the signals, we employ the following measures:

1. Alignment of the shared embeddings: minimize the MSE between the shared components and compute the final shared embedding using product of experts.

$$\frac{1}{K_{\text{shared}}}\left\|\mu_{\text{RNA}}^{(c)} - \mu_{\text{protein}}^{(c)}\right\|_2^2 \qquad p(\mathbf{c} \mid \mathbf{x}_{\text{RNA}}, \mathbf{x}_{\text{prot}}) \propto \frac{1}{p(\mathbf{c})}p(\mathbf{c} \mid \mathbf{x}_{\text{RNA}})\,p(\mathbf{c} \mid \mathbf{x}_{\text{prot}})$$

2. Disentanglement of the modality-specific embeddings: use a simple (dot-product) correlation penalty between the shared embedding and the modality-specific ones.

$$\sum_{m\in\text{RNA,protein}} \left(\mathbf{c}^\top \mathbf{s}_m\right)^2$$

Then, to reconstruct the input vector for a cell i, we need both the shared and modality-specific embeddings and compute:

$$\hat{\mathbf{x}}_i^{\text{RNA}} = \mathbf{W}_C^{\text{RNA}}\begin{bmatrix}\mathbf{c}_i \\ \mathbf{s}_i^{\text{RNA}}\end{bmatrix} + \mathbf{W}_{D\times C}^{\text{RNA}}\mathbf{z}_i^{D\times C} + \mathbf{W}_V^{\text{RNA}}\mathbf{v}_y,$$

$$\hat{\mathbf{x}}_i^{\text{prot}} = \mathbf{W}_C^{\text{prot}}\begin{bmatrix}\mathbf{c}_i \\ \mathbf{s}_i^{\text{prot}}\end{bmatrix} + \mathbf{W}_{D\times C}^{\text{prot}}\mathbf{z}_i^{D\times C} + \mathbf{W}_V^{\text{prot}}\mathbf{v}_y.$$

3. Results

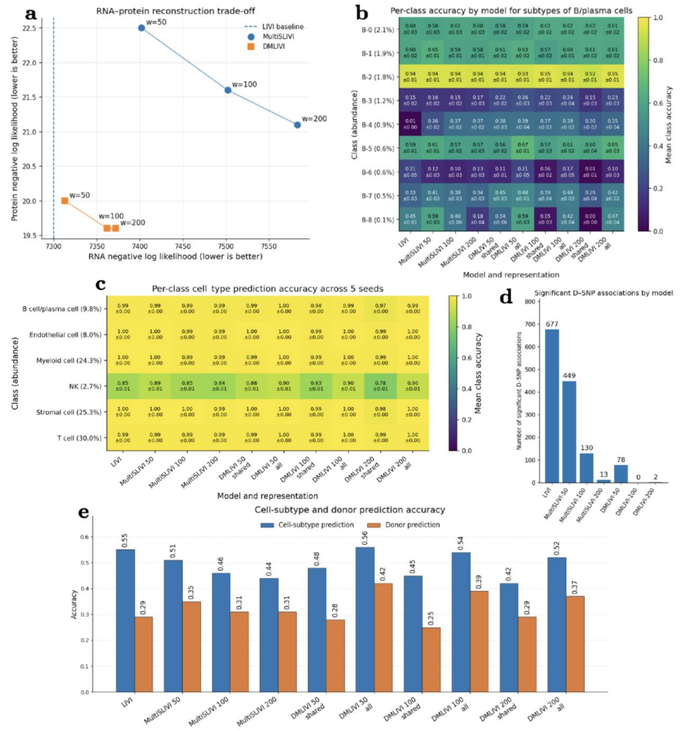

Figure 4: **Evaluation of MultiSLIVI and DMLIVI.** (**a**) RNA–protein reconstruction trade-off for MultiSLIVI and DMLIVI across protein reconstruction weights ($\lambda_{\text{prot}} = \text{w}$) compared to the LIVI baseline. (**b**) Per-class prediction accuracy for B/plasma-cell subtypes across models. (**c**) Per-class prediction accuracy for major cell types across models. (**d**) Number of significant D–SNP associations discovered by each model. (**e**) Cell-subtype and donor-identity prediction accuracy from the learnt cell-state latent representations. For panels (**b**–**e**) the number after a model signifies the protein reconstruction weight. For DMLIVI, "shared" means the shared-embedding-only version, whereas "all" also includes modality-specific embeddings.

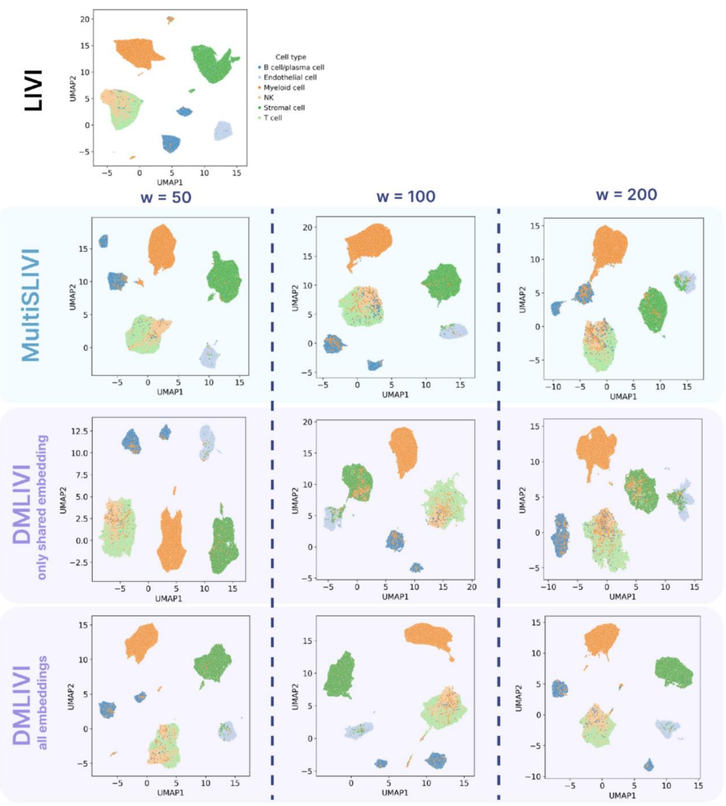

Figure 5: **UMAP visualisation of cell-state latent representations learnt by LIVI, MultiSLIVI, and DMLIVI.** The top panel shows the LIVI baseline. The remaining panels compare MultiSLIVI, DMLIVI using only the shared embedding, and DMLIVI using all embeddings across protein reconstruction weights ($\lambda_{\text{prot}} = w = 50, 100, 200$).

4. Discussion

- MultiSLIVI and DMLIVI extend LIVI to include RNA and protein data.
- Both models reconstructed protein counts, but this reduced RNA reconstruction quality.
- DMLIVI gave a better RNA-protein reconstruction balance than MultiSLIVI.
- Cell-type information was mostly preserved in the latent space.
- Higher protein weighting made cell-type clusters less clear in MultiSLIVI.
- Multimodal models showed more donor information in the cell-state embeddings.
- LIVI still found the most significant SNP-D factor associations among tested models.
- Modelling additional protein data this way did not improve downstream eQTL discovery.
- Overall, LIVI remained the strongest model for genetic association testing.
- Future models should more carefully separate shared, modality-specific, and donor-related variation.

5. References

[1] Danai Vagiaki et al. “Mapping trans-eQTLs at single-cell resolution using Latent Interaction Variational Inference”. In: bioRxiv (2026). doi: 10.64898/2026.02.04.703363.

[2] Imant Daunhawer et al. “Self-supervised Disentanglement of Modality-Specific and Shared Factors Improves Multimodal Generative Models”. In: Pattern Recognition. Ed. by Zeynep Akata, Andreas Geiger, and Torsten Sattler. Cham: Springer International Publishing, 2021, pp. 459–473.

Contact Information

Jakub Fręchowicz, J.S.Frechowicz@student.tudelft.nl