Variational Inference for Single-Cell Transcriptomics

Valentin Badea **Etienne Lefranc**
ENS Paris-Saclay | Université Paris-Cité | CentraleSupélec

Key message

We re-implement scVI from scratch and validate it on CORTEX. On CITE-seq PBMC, we extend scVI to **jointly model RNA and ADT** with a **shared cell latent state**, improving alignment with protein markers (CD proteins).

Context / Motivation

Single-cell RNA sequencing (scRNA-seq) measures gene expression at cellular resolution, enabling the discovery of cell types and states in complex tissues. However, scRNA-seq data are **high-dimensional**, **sparse**, and strongly affected by **technical confounders** such as library size variation, batch effects, and dropouts.

Goal. Learn a **low-dimensional latent representation** of cells that preserves biologically meaningful variation while accounting for technical noise, and extend this representation to **multimodal** measurements when available (CITE-seq: RNA + surface proteins).

Approach.

- **CORTEX (RNA-only):** validate RNA latent space via t-SNE and cell-type separation.
- **CITE-seq PBMC (RNA-only):** check consistency with protein markers (CD proteins).
- **CITE-seq PBMC (RNA+ADT):** learn a joint RNA–protein representation to improve alignment with protein-defined populations.

Experimental setup (implementation)

- **From-scratch PyTorch** implementation of scVI (ZINB) and citeVI (RNA+ADT), trained end-to-end.
- **Latent space:** $p = 10$ for scVI, p = 15 for the other one; posterior means used as embeddings.
- **Training:** 20 epochs on PBMC; loss stabilizes after ∼10 epochs.
- **Visualization:** t-SNE on CORTEX and on PBMC (CD proteins overlays).

Methods: scVI (RNA generative model + amortized VI)

What scVI models. For each cell n, scVI separates **biological variation** from **technical effects** using:

- latent state $z_n \in \mathbb{R}^p$ (cell identity / state),
- library-size variable l_n (sequencing depth / capture efficiency),
- batch label s_n (batch-effect correction via embeddings).

Generative process (counts).

$$z_n \sim \mathcal{N}(0, I), \qquad \log l_n \sim \mathcal{N}(l_\mu^{(s_n)}, (l_\sigma^{(s_n)})^2),$$
$$x_{ng} \mid z_n, l_n, s_n \sim \mathrm{ZINB}\big(\mu_{ng}(z_n, s_n, l_n),\ \theta_g,\ \pi_{ng}(z_n, s_n)\big),$$

where θ_g is gene-specific dispersion and π_{ng} captures dropout/zero inflation.

Amortized variational inference.

$$q_\phi(z_n, \log l_n \mid x_n, s_n) = q_\phi(z_n \mid x_n, s_n)\, q_\phi(\log l_n \mid x_n, s_n).$$

Training objective (ELBO).

$$\mathcal{L}_{\text{scVI}}(x_n) = \mathbb{E}_{q_\phi}[\log p_\theta(x_n \mid z_n, l_n, s_n)] - \mathrm{KL}(q_\phi(z_n \mid x_n, s_n) \| p(z_n)) - \mathrm{KL}(q_\phi(\log l_n \mid x_n, s_n) \| p(\log l_n \mid s_n))$$

Output embedding. Posterior mean $\mu_z(x_n, s_n)$ as RNA representation; visualize with t-SNE.

Datasets

CORTEX (scRNA-seq). Curated mouse cortex scRNA-seq with annotated neuronal and glial types.

CITE-seq PBMC (RNA+ADT). Human PBMCs with RNA counts and surface protein abundances (ADT). Protein markers provide an independent signal to assess embedding structure and multimodal integration.

Preprocessing (high level).

- RNA: raw UMI counts as scVI input; batch labels used when available.
- ADT: log-transform $\log(1 + a)$ for the protein module input.

Graphical model (scVI)

Factorization.

$$p(x_n, z_n, l_n \mid s_n) = p(z_n)\, p(l_n \mid s_n)\, p(x_n \mid z_n, l_n, s_n).$$

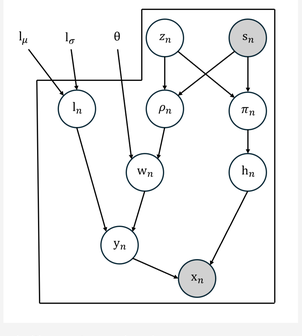
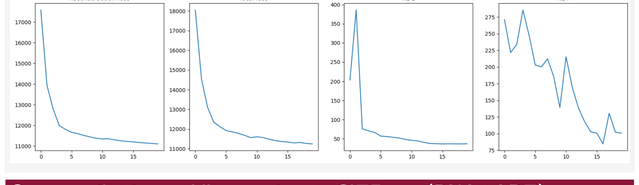


Plate: z_n captures biology, l_n library size, s_n batch.

Training diagnostics (scVI)

Optimization behaviour. Reconstruction decreases then plateaus; KL stabilizes ⇒ well-behaved optimization.



Our contribution: scVI extension to CITE-seq (RNA+ADT)

Idea. Extend scVI to CITE-seq with a **shared cell latent state** z_n and an added protein likelihood term, trained **end-to-end**.

Generative model. Keep RNA model $p(x_n \mid z_n, l_n, s_n)$ (ZINB) and add an ADT decoder:

$$z_n \sim \mathcal{N}(0, I), \qquad \log l_n \sim \mathcal{N}(l_\mu^{(s_n)}, (l_\sigma^{(s_n)})^2),$$
$$a_{np} \mid z_n, s_n \sim \mathrm{NB}(\mu_{np}^a(z_n, s_n),\ \phi_p), \qquad \mu_n^a = f_a(z_n, s_n).$$

Inference.

$$q(z_n, l_n \mid x_n, a_n, s_n) = q(z_n \mid x_n, a_n, s_n)\, q(l_n \mid x_n, s_n).$$

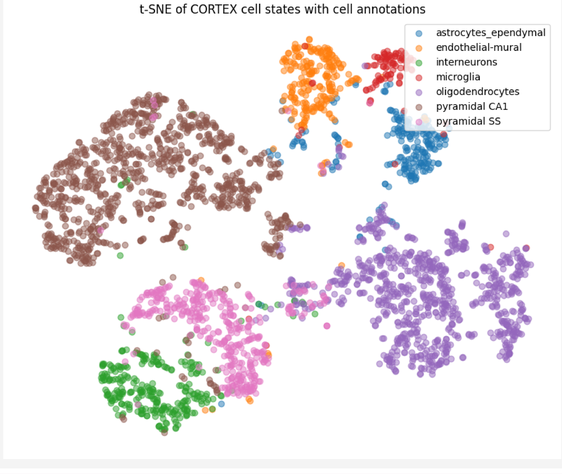
Training objective.

$$\mathcal{L}(x_n, a_n) = \mathbb{E}_{q(z_n, l_n \mid x_n, a_n, s_n)}\Big[\log p(x_n \mid z_n, l_n, s_n) + \log p(a_n \mid z_n, s_n)\Big] - \mathrm{KL}(q(z_n \mid x_n, a_n, s_n) \,\|\, p(z_n)) - \mathrm{KL}(q(l_n \mid x_n, s_n) \,\|\, p(l_n \mid s_n)).$$

Takeaway. A **single latent space** captures biology; proteins sharpen separation of protein-defined subpopulations (CD proteins).

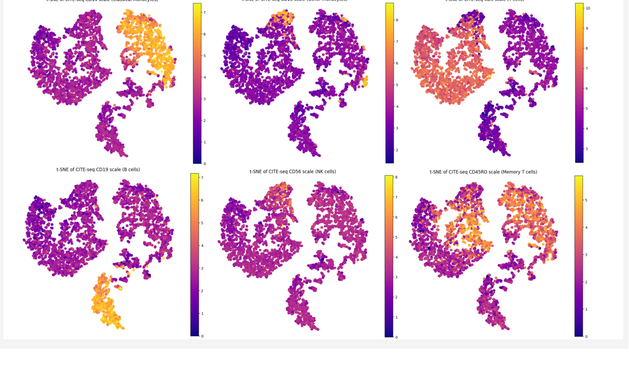
Results 1: CORTEX (RNA-only)

RNA latent space captures annotated cell types. Train scVI on CORTEX and visualize posterior means with t-SNE: separation of neuronal and glial populations.



Results 2: scVI recovers cell types in CITE-seq (RNA-only)

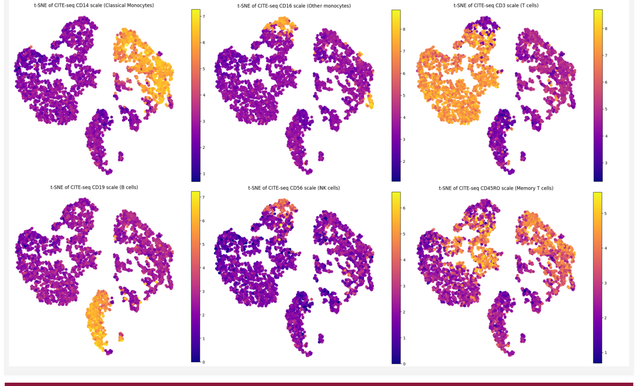
scVI latent space recovers major immune cell types. t-SNE of the RNA latent space learned by scVI on CITE-seq PBMC data, colored by surface protein markers (CD14, CD16, CD3, CD19, CD56, CD45RO). Distinct regions correspond to classical and non-classical monocytes, T cells, B cells, NK cells and memory T cells, even though protein information is **not used during training**.



Results 3: CITE-seq (RNA+ADT)

Protein markers validate and refine the learned embedding. t-SNE of the joint RNA+ADT latent space, colored by surface protein expression (CD14, CD16, CD3, CD19, CD56, CD45RO). Each marker highlights a well-localized region corresponding to known immune cell populations: classical and non-classical monocytes (CD14/CD16), T cells (CD3, CD45RO), B cells (CD19), and NK cells (CD56).

Interpretation. The multimodal model aligns transcriptomic structure with protein-defined cell identities, yielding sharper and more biologically consistent separation than RNA-only embeddings.



Conclusion / Takeaways

- **scVI re-implementation:** count-aware VAE learns robust RNA embeddings (CORTEX) and aligns with protein markers (PBMC).
- **CITE-seq integration:** adding ADT via a jointly learned module improves match with protein-defined subtypes (CD14/CD16).
- **Practical:** probabilistic, batch-aware latent space for visualization and interpretation.

References / Contact

References.

- Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., Yosef, N. (2018). *Deep generative modeling for single-cell transcriptomics.* Nature.
- Stoeckius, M. et al. (2017). *Simultaneous epitope and transcriptome measurement in single cells.* Nature.
- Zeisel, A. et al. (2015). *Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* Science.
- van der Maaten, L., Hinton, G. (2008). *Visualizing High-Dimensional Data Using t-SNE.* JMLR.

Contact.

Etienne Lefranc
Email: etiennelefranc@yahoo.com
Phone: +33 7 82 65 42 06
Valentin Badea
Email: valentin.badea@etu.u-paris.fr
Phone: +33 6 95 16 03 88