Extending the scVAE Framework: A Hierarchical Mixture-of-Mixtures Approach for Single-Cell Gene Expression Data

Raphaël RUBRICE, Adam KEDDIS, Tiffney AINA

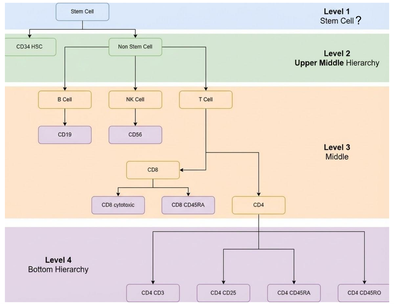
école normale supérieure paris-saclay — MATHÉMATIQUES VISION APPRENTISSAGE

université PARIS-SACLAY — Université Paris Cité

Introduction

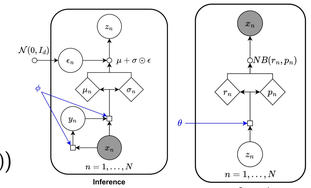
- **scVAE** (Grønbech et al., 2020) allows clustering and modeling of single cell count data.
- Cell-types are hierarchically organized, but scVAE's flat mixture prior cannot explicitly enforce nested structure: We **extend the scVAE** framework to **model hierarchies**.
- Using **PBMC** (Peripheral Blood Mononuclear Cell) datasets, we built a reference four-level cell hierarchy (from stem to fine subtypes) and introduced two new models:
  - **IndMoMVAE** (independent mixture branches)
  - **MoMixVAE** (hierarchical mixture-of-mixtures)

Cell-type hierarchy used in our experiments:



Single Cell Variational Auto-Encoder (scVAE)

- Models each cell's gene-expression **vector (x)** through a **latent representation (z)** and discrete **cluster label (y)**



- **Generative** process (common to all models):

$$y \sim \mathrm{Cat}(\pi), \qquad z \sim \mathcal{N}(\mu_y, \sigma_y^2 I), \qquad x \sim \mathrm{NB}\big(r_\theta(z), p_\theta(z)\big)$$

- Introduces a **Gaussian mixture prior** over (z) to enable unsupervised discovery of cell-type clusters
- **Inference** network $q_\phi(y, z|x)$ approximates the intractable posterior (modified in our extended models)
- Training maximizes the **Evidence Lower Bound (ELBO)** balancing reconstruction and **KL regularization**:

$$\mathcal{L}(x) = \mathbb{E}_{q_\phi(z,y|x)}[\log p_\theta(x|z)] - \mathrm{KL}\big(q_\phi(z, y|x) \,\|\, p_\theta(z, y)\big)$$

- Provides a probabilistic alternative to heuristic clustering and serves as the foundation for our hierarchical extensions **IndMoMVAE** and **MoMixVAE**

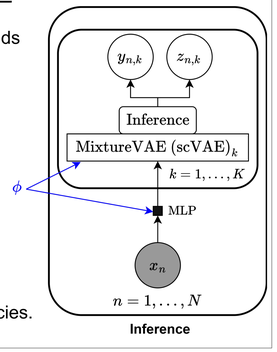
Our Methods

Independent Mixture VAE

- Multiple **independent mixture views**: each branch yields a separate clustering of observations. The generative probability of **x** is formulated as:

$$p(x) = \sum_k \left[\int_z \left(\sum_y p(x|z)\, p(z|y)\, p(y|k)\, p(k) \right) dz \right]$$

- Variational approximation ***independently*** infers latent variables and cluster labels per branch, serving as a baseline to evaluate the impact of hierarchical dependencies.



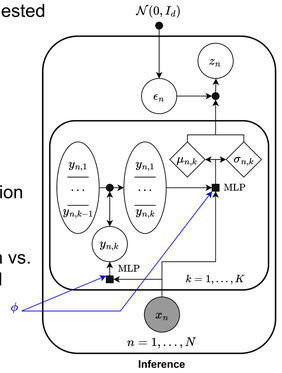
Mixture of Mixture (MoMix)

- Introduces **mixture-of-mixtures** via **dependent mixture** layers: each hierarchical level is conditioned on previous levels
- Captures **soft hierarchical** relationships through the joint factorization:

$$z \sim \sum_{k_1=1}^{K_1} \pi_{k_1} \left(\cdots \left(\sum_{k_{L-1}=1}^{K_{L-1}} \pi_{k_{L-1}} \left(\sum_{k_L=1}^{K_L} \pi_{k_L} Law(\theta) \right) \right) \cdots \right)$$

$$\begin{aligned} p(x, z, y_1, \ldots, y_L) &= p(x \mid z)\, p(z \mid y_L, \ldots, y_1) \\ &\times p(y_L \mid y_{L-1}, \ldots, y_1) \\ &\cdots \\ &\times p(y_2 \mid y_1)\, p(y_1) \end{aligned}$$

- Variational posterior mirrors this structure enabling nested clustering rather than rigid tree branches.
- Optimizes a hierarchical ELBO that averages KL divergences across levels, with beta scaling and **marginal regularization** to prevent component collapse.
- To fight component ***marginal collapse:*** Regularization schedule for a marginal-usage KL term
- Learns smooth, probabilistic hierarchies (e.g., Stem vs. Non-stem) subdivide into finer immune lineages and cell-state subtypes



Results

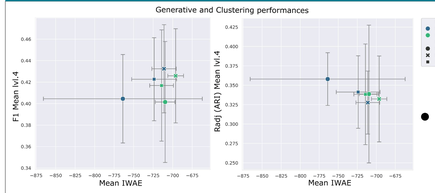


Figure 1 - Performance Comparison. (F1, ARI) across models and priors at level 4.

- Shows slight trade-offs between generative performance (IWAE) and clustering accuracy.

- On average:
  - Student's t prior yields higher IWAE: better reconstruction quality. Normal prior tends to yield better clustering quality
  - MoMixVAE: Best balance of stability and accuracy.

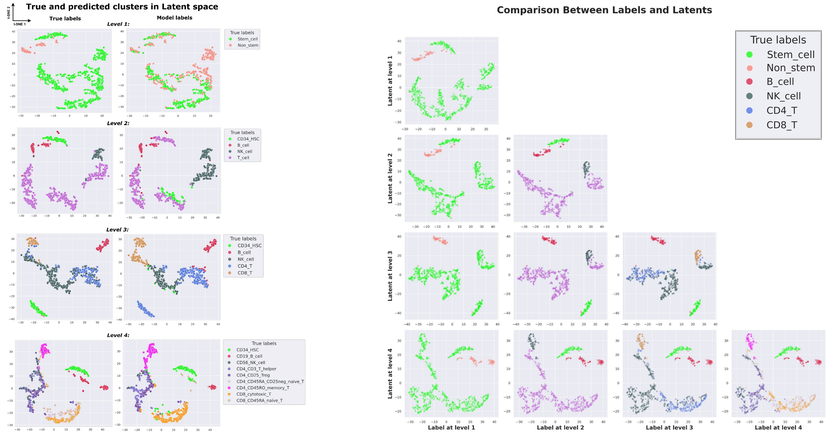


Figure 2 - True vs Predicted Hierarchical Labels. Compares true biological labels and MoMixVAE-predicted clusters across all four hierarchy levels.

- Demonstrates strong alignment from coarse Stem/Non-stem splits to fine immune subtypes, confirming hierarchical consistency.

Figure 3 - Hierarchical Latent Space Matrix. Visualizes the **coarse-to-fine structure** of the learned latent hierarchy.

- Latent spaces progressively subdivide into nested, biologically meaningful clusters, revealing consistent lineage organization.

Conclusion

Hierarchical Mixture-of-Mixtures VAEs show improvements in clustering accuracy and biological interpretability by capturing multi-level cell-type structures better than the MixtureVAE and IndMoMVAE models. This approach provides a powerful framework for **modeling complex cellular hierarchies**.

However, training remains computationally demanding and occasionally prone to marginal component collapse at deeper levels. **Future work** could focus on scaling hierarchical inference and integrating conditional generation to improve efficiency and biological controllability.