Computational Self-Organisation of Hematopoietic Latent Architectures via Causal Directionality in D-SPIN Manifolds

A Proposed Framework for Resolving Frustrated Regulatory States via Interventional VAE Logic

Oluwatobi Baruwa

1. Manifold Parameterisation and the Developmental Prior The proposed framework addresses the transcriptomic "measurement revolution" by utilising a Variational Autoencoder (VAE) to project high-dimensional single-cell RNA sequencing (scRNA-seq) counts (X with ~20,000 dimensions) onto a parsimonious latent manifold (z with roughly 20 dimensions).

- **Front-End Integration:** The model utilises PopAlign to build Gaussian Mixture Models (GMMs) of healthy hematopoietic populations sourced from Villani et al. (2017). These compressed representations serve as a Developmental Prior, anchoring the latent space to biophysical constraints and mitigating the "catastrophic forgetting" inherent in static neural architectures.
- **Symmetry-Breaking Logic:** Unlike standard autoencoders, the Mesengenic AI architecture is dynamic. By monitoring reconstruction error across the manifold, the network initiates "Symmetry-Breaking" events—branching new latent layers (increasing the number of dimensions) to mirror the differentiation of Hematopoietic Stem Cells (HSCs) into specialised lineages.

2. D-SPIN Integration: Orienting Frustrated Spins D-SPIN recovers undirected gene couplings Jij within an Ising-model Hamiltonian. The Energy (E) of a cell state is defined by the sum of interactions between genes (s) and external perturbations (h):

$$E(s) = -\sum_{i<j} J_{ij} S_i S_j - \sum_i h_i s_i$$

A primary computational challenge in this framework is **"Frustration"**—states where mutual inhibition ($J_{ij} < 0$) creates a rugged energy landscape. My framework (VAE v3) introduces **Causal Directionality** to resolve these tensions:

- **The do-operator Logic:** Using the Pearl framework for causal inference, we distinguish between observation P(B|A) and intervention P(B | do(A)). By applying Sparse Mechanism Shift (SMS) to the Norman et al. (2019) CRISPRa dataset, we measure the directional asymmetry in gene to gene interactions.
- **The Operation:** If the impact of forcing Gene A on Gene B - P(B | do(A)) - is significantly greater than the inverse, the model orients the undirected coupling as a directed edge (A -> B).This transforms the undirected spin-glass into a Directed Acyclic Graph (DAG), identifying the "Causal Pivot Genes" that act as the master repressors for lineage commitment.

3. Inverse Manifold Simulation and Precision Reprogramming The model acts as a scalable tool for Algorithmic Tissue Engineering, specifically targeting the causal rewiring found in Myelodysplastic Syndromes (MDS) and Clonal Hematopoiesis, utilising the Van Galen et al. (2019) bone marrow hierarchies.

- **Localisation:** By performing a manifold comparison between healthy priors and diseased states, the VAE localises the coordinate of divergence (z-fail) where the HSC trajectory is biased toward pro-inflammatory branching.
- **The Causal Nudge:** The generative engine performs Inverse In-Silico Simulation, predicting the minimal perturbation do(X) required to flip the identified Causal Pivot Genes. This effectively "nudges" the system back across the Waddington ridge toward a healthy regulatory attractor.

4. Implementation Stack

- **Frameworks:** PyTorch, PyTorch Geometric (for MPNN-based graph modeling).
- **Kernels:** GPflow for Gaussian Process-informed latent transitions.
- **Inference:** PHATE for high-fidelity visualisation of branching differentiation trajectories.

Citations

1. Jiang J, Thomson M. A guide to D-SPIN: constructing regulatory network models from single-cell RNA-seq perturbation data. bioRxiv. 2025:2025.07.28.667326.

2. Chen S, Rivaud P, Park JH, Tsou T, Charles E, Haliburton JR, et al. Dissecting heterogeneous cell populations across drug and disease conditions with PopAlign. Proceedings of the National Academy of Sciences. 2020;117(46):28784-94.

3. Jiang J, Chen S, Tsou T, McGinnis CS, Khazaei T, Zhu Q, et al. D-SPIN constructs gene regulatory network models from multiplexed scRNA-seq data revealing organizing principles of cellular perturbation response. bioRxiv. 2025:2023.04.19.537364.

4. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, et al. Visualizing structure and transitions in high-dimensional biological data. Nature Biotechnology. 2019;37(12):1482-92.

5. Neuberg L. Causality: models, reasoning, and inference, by judea pearl, cambridge university press, 2000. Econometric Theory. 2003;19:675-85.

6. Norman TM, Horlbeck MA, Replogle JM, Ge AY, Xu A, Jost M, et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science. 2019;365(6455):786-93.

7. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science. 2017;356(6335).

8. van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell. 2019;176(6):1265-81.e24.

9. Hay SB, Ferchen K, Chetal K, Grimes HL, Salomonis N. The Human Cell Atlas bone marrow single-cell interactive web portal. Exp Hematol. 2018;68:51-61.