Comparative Analysis of Systemic lupus Disease Using VEGA and scRNA-Seq Data

Zuzanna Smektała

June 2025

Abstract

Systemic lupus erythematosus (SLE) is a complex autoimmune disease characterized by widespread immune dysregulation, often driven by type I interferon (IFN) signaling. This study utilizes single-cell RNA sequencing (scRNA-Seq) data from peripheral blood mononuclear cells (PBMCs) of SLE patients, both untreated and treated with interferon-beta (IFN-β), to investigate cell-type-specific pathway activity. The VEGA model (Variational autoEncoder enhanced by Gene Annotations), a deep learning approach incorporating biological pathway information into a sparse and interpretable latent space, is applied to the dataset. Compared to a standard variational autoencoder (VAE), VEGA shows improved separation of cell types and enables the identification of condition-specific transcriptional programs, particularly in response to IFN-β treatment. Pathway-level analysis indicates that IFN-α/β signaling and downstream antiviral gene programs are selectively activated in myeloid cells—especially monocytes and dendritic cells—following IFN-β stimulation. These results demonstrate the potential of pathway-informed generative models to elucidate immune dysregulation in SLE and to identify treatment-responsive cellular subpopulations and candidate biomarkers.

1. Introduction

Systemic lupus erythematosus (SLE) is a chronic, autoimmune inflammatory disease characterized by the immune system attacking the body's own tissues. It can affect various organs, including the skin, joints, kidneys, heart, lungs, and nervous system. The etiology of SLE involves a complex interplay of genetic, hormonal, and environmental factors. A hallmark of SLE is the overactivation of the type I interferon (IFN) pathway, which contributes to widespread immune dysregulation. Symptoms range from fatigue, joint pain, and rashes to severe organ damage. The disease course is typically relapsing-remitting, and early diagnosis is key to improving outcomes [2].

Recent advances in single-cell RNA sequencing (scRNA-Seq) allow for highly detailed cellular state characterization. Artificial Neural Networks (ANNs), particularly autoencoders (AEs), are powerful tools for scRNA-Seq analysis, handling tasks like dimensionality reduction, clustering, and data denoising. More recently, deep generative models like variational autoencoders (VAEs) (scVI, scGen) excel at probabilistic modeling of single-cell transcriptomes by learning data distributions. However, a major limitation is their lack of interpretability, making it difficult to relate latent representations to biological meaning (scGen's latent perturbation vectors aren't linked to gene module variations).

To address this, integrating prior biological knowledge has been explored. DCell incorporates hierarchical molecular subsystem information for supervised tasks but is limited to predicting phenotypic outcomes. f-scLVM, a Bayesian hierarchical model, uses prior biological knowledge to infer gene module activity, offering interpretability, but it's computationally expensive and can't infer for out-

of-sample data. This leads to the proposal of VEGA (VAE enhanced by gene annotations), a Variational Autoencoder (VAE) featuring a sparse linear decoder informed by biological networks. Its primary objectives are twofold: first, to encode data into an interpretable latent space that reflects pathway status or transcriptional regulator activity, and second, to infer gene module activities for out-of-sample data.

VEGA is a novel VAE architecture designed for interpretability. Unlike standard VAEs where latent variables are hard to interpret, VEGA guides its decoder connections using gene module membership information from databases like Gene Ontology or Reactome. While a linear decoder offers some interpretability by directly connecting latent variables to genes, it still requires further statistical tests. VEGA, in contrast, uses a sparse architecture that explicitly reflects biological knowledge about gene regulation. Genes within gene modules work together and are often correlated. It incorporates these modules as latent variables (Z). The key is its gene membership mask (M) applied to the decoder layer's weights. This mask ensures a link from a gene module variable (GMV) to an output gene only if that gene is annotated as a member of that specific module.

The design offers two main advantages: direct interpretability of latent variables as biological module activity, and flexibility in gene module specification, which allows for generalization to various biological abstractions such as pathways, gene regulatory networks (GRNs), and cell types derived from curated databases like MSigDB and Reactome [1].

2. Results

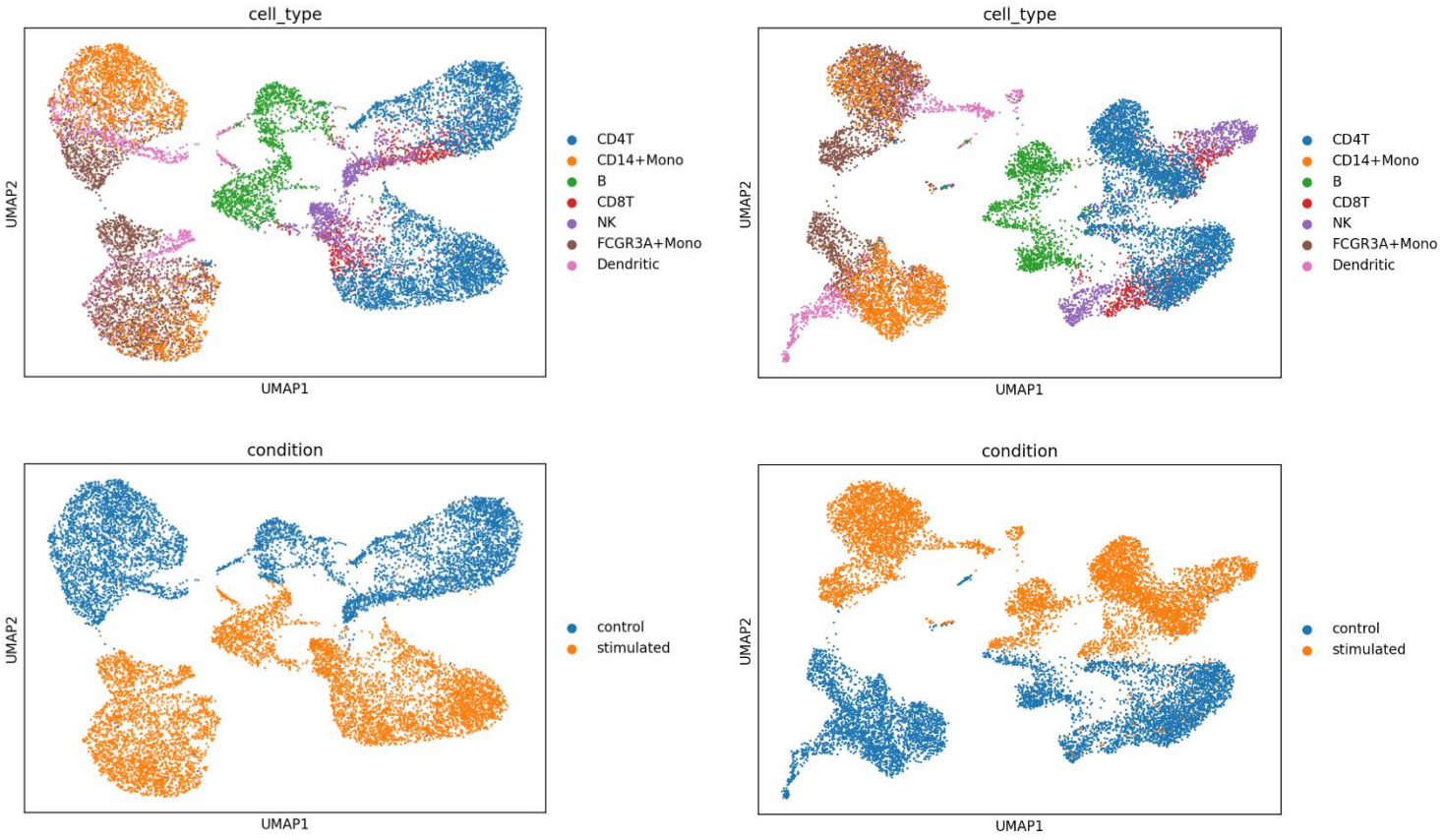

Fig. 1 Comparative analysis evaluates UMAP visualizations of single-cell transcriptomes from lupus patients, contrasting the performance of a standard VAE model against the pathway-aware VEGA model. Both visualizations display 7 immune cell populations (CD4T, CD8T, B cells, monocyte subsets, NK, and dendritic cells) across control (untreated) and stimulated (interferon-treated) conditions.

In terms of cell type separation, the VEGA visualization consistently outperforms the VAE. While the VAE shows partial overlap between CD4T/CD8T cells and intermixed CD14+ and FCGR3A+ monocyte subsets, VEGA achieves sharper separation for these populations. Dendritic cells, diffuse in the VAE plot, form more compact groupings with VEGA. Overall, the VAE produces broader, more overlapping clusters, whereas VEGA yields tighter, more distinct populations [Fig.1].

Regarding condition-specific effects, VEGA demonstrates a stronger ability to reveal the impact of interferon stimulation. The VAE visualization shows only mild condition-based dispersion and subtle shifts in treatment effects, with mostly overlapping control and stimulated samples. In contrast, VEGA exhibits stronger condition-specific clustering, clear subgrouping within cell types due to treatment, and more distinct spatial separation between control and stimulated samples.

2.1 Myeloid-restricted interferon pathway activation in IFN-β-treated lupus patients

To investigate the transcriptional effects of interferon-beta (IFN-β) treatment in systemic lupus erythematosus (SLE), pathway-level analysis were performed using single-cell RNA sequencing data from peripheral blood mononuclear cells (PBMCs) of IFN-β–treated and untreated patients. UMAP embeddings were generated to visualize global transcriptomic variation across cells. The enrichment scores were overlaid for pathways:

- → nNef mediated downregulation of mhc class i complex cell surface expression,
- → Antigen processing cross presentation,
- → Cross presentation of soluble exogenous antigens endosomes,
- → Il 7 signaling,
- → Growth hormone receptor signaling,
- → Antiviral mechanism by ifn stimulated genes,
- → Prolactin receptor signaling,
- → Signaling by scf kit,
- → Signaling by erbb4,
- → Signalling by ngf,
- → Interferon alpha beta signaling

onto the UMAP projection to assess pathway activation across cell types and treatment conditions.

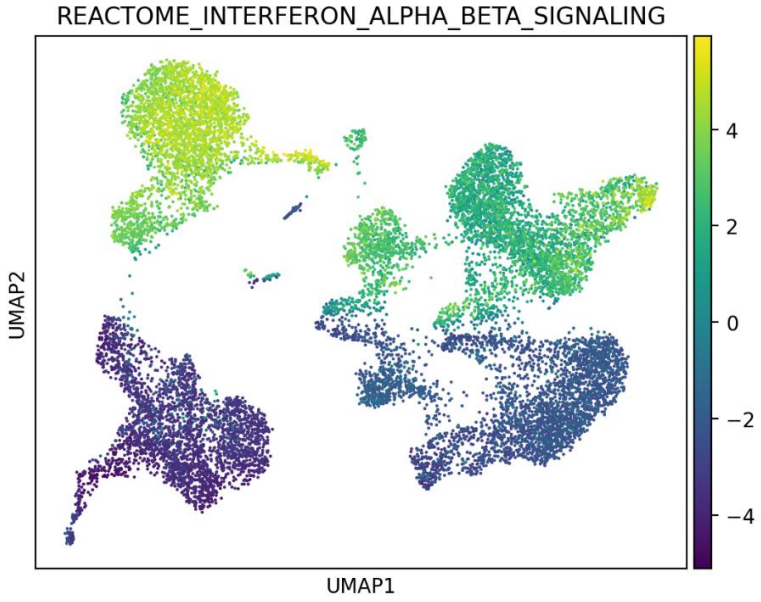

Fig.2 Single-cell mapping of interferon-α/β signaling activity in lupus PBMCs

UMAP visualization of interferon alpha beta signaling pathway activity across immune cell populations

As shown in Figure 2, pathway activity was heterogeneously distributed across the UMAP space, with several distinct clusters exhibiting markedly elevated IFN-α/β signaling scores (yellow regions). Integration with cell type annotations (Figure 2, top) revealed that these clusters predominantly consisted of CD14+ monocytes, FCGR3A+ monocytes, and dendritic cells, indicating that the IFN response is largely confined to the myeloid compartment. In contrast, T cells (CD4+ and CD8+), B cells, and NK cells showed relatively lower pathway activity.

When stratified by treatment condition (Figure 2, bottom), it is observed that the regions of highest IFN-α/β signaling activity coincided with cells derived from IFN-β–treated individuals (orange dots, Fig.1), while control (untreated) samples displayed consistently lower pathway scores (blue dots, Fig.1). These results demonstrate that exogenous IFN-β administration selectively enhances the expression of type I interferon-responsive genes, particularly within innate immune subsets.

This finding is consistent with the known role of monocytes and dendritic cells as primary responders to type I interferons, given their high expression of interferon receptors and downstream signaling components . It also aligns with prior reports of elevated interferon signatures in SLE patients and underscores the importance of the type I interferon axis in disease pathogenesis [3,4].

2.2 Antiviral ISG program shows myeloid-specific activation in IFN-β-treated SLE

To further delineate the downstream effects of type I interferon signaling in systemic lupus erythematosus (SLE), the expression of genes involved in the *REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIMULATED_GENES* pathway were examined. This gene set encompasses a broad range of interferon-stimulated genes (ISGs) that mediate antiviral immunity via mechanisms such as RNA degradation, inhibition of viral replication, and immune effector activation.

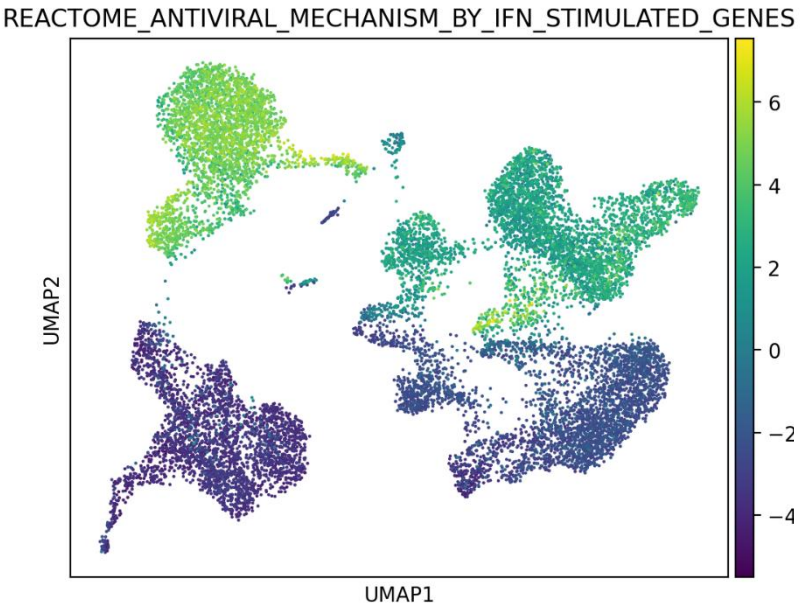

Fig.3 Single-cell mapping of antiviral mechanism by IFN stimulated genes activity in lupus PBMCs

UMAP visualization of antiviral mechanism by IFN stimulated genes pathway activity across immune cell populations

As shown in Figure 3, pathway activity across the UMAP embedding reveals discrete cellular subpopulations with elevated ISG signatures (yellow to green regions), particularly in the upper and upper-right clusters. These areas overlap with the same regions previously shown to express high levels of interferon alpha/beta signaling, suggesting a coordinated induction of the antiviral state in response to IFN-β treatment.

When mapped onto cell type annotations, the strongest antiviral responses are localized to CD14+ monocytes, FCGR3A+ monocytes, and dendritic cells, consistent with their role as first-line immune sentinels and potent producers and responders to type I interferons [5,6]. In contrast, lymphoid populations such as T cells and B cells exhibit markedly lower ISG pathway activity, highlighting a selective engagement of the innate immune arm in this antiviral program.

This pattern was again found to correlate with treatment condition: IFN-β–treated cells displayed significantly higher activation of this antiviral signature compared to untreated controls, consistent with pharmacologic stimulation of ISG expression through the type I IFN pathway.

2.3 Interferon pathway gene weight analysis reveals dominant ISG contributors and proteasomal signatures

To identify the most influential genes within the *INTERFERON_ALPHA_BETA_SIGNALING* pathway, gene weight magnitudes derived from the VEGA latent space were ranked. As shown in Figure 4a, ISG20, ISG15, and IFIT3 exhibited the highest absolute weights, indicating their dominant contribution to the learned latent representation of interferon signaling. These genes are well-characterized interferon-stimulated genes (ISGs) involved in antiviral defense, RNA degradation, and immune regulation [8,9]. The top 10 contributors were largely composed of canonical ISGs, including IFI6, IFIT1/2, and IFITM3, which are known to inhibit viral entry, replication, and promote apoptosis in infected cells [10]. In contrast, genes such as *IFNB1*, *C21orf67*, and *TGB2* showed the lowest ranked weights, suggesting a more limited role in driving the latent interferon response signal captured by the model.

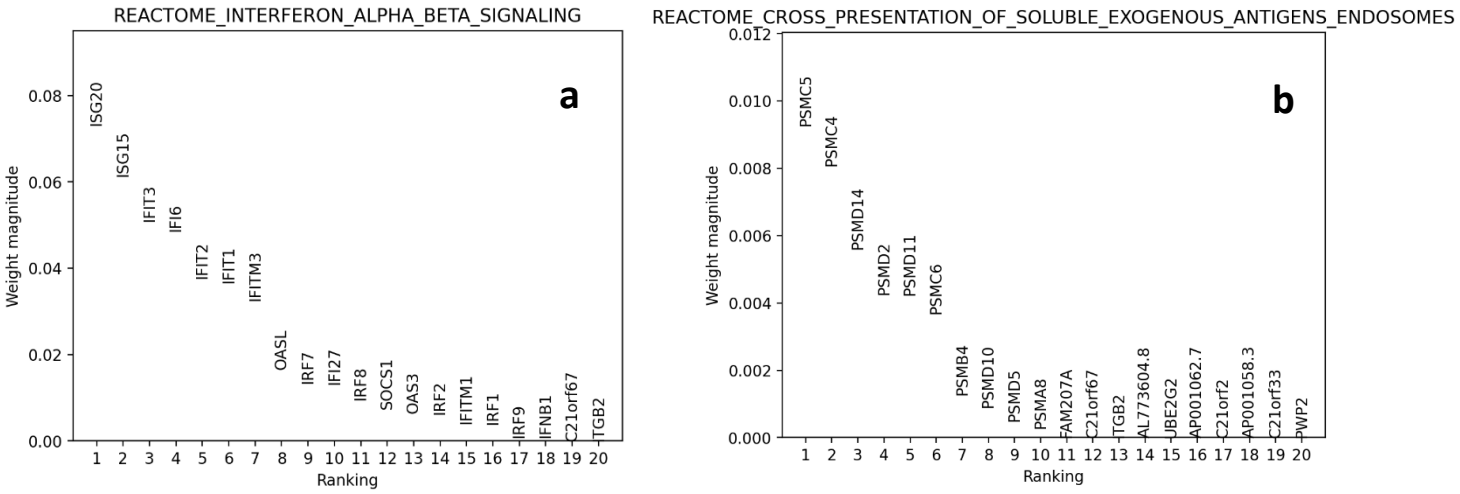

Fig.4 Ranked gene weights from the VEGA model decoder layer showing the top 20 contributors to **a)** Interferon alpha beta signaling pathway activity and **b)** Cross presentation of soluble exogenous antigens endosomes

These results demonstrate that VEGA assigns higher weights to genes with established roles in type I IFN responses, effectively prioritizing biologically meaningful features without manual curation. This supports the utility of the model in dissecting complex immune signaling programs at single-cell resolution.

As shown in Figure 4b, genes encoding proteasomal subunits emerged as the most influential contributors, with PSMC5, PSMC4, and PSMD14 ranking highest based on absolute weight magnitude. These genes are essential components of the 19S regulatory particle of the 26S proteasome, which is

responsible for the ATP-dependent unfolding and translocation of ubiquitinated proteins into the catalytic core for degradation [11]. Their prominence in SLE samples suggests heightened proteasomal engagement, potentially reflecting increased antigen processing and altered immunoproteasome activity, a known feature of SLE [12]. Further high-ranking genes included PSMD2, PSMD11, and PSMC6, reinforcing the role of the regulatory subunit in antigen processing. These findings align with prior studies showing dysregulated proteasome function in autoimmune settings and its impact on MHC class I antigen presentation [13]. Lower-weighted genes in the ranked list included PSMA8, FAM207A, and multiple non-coding or poorly characterized transcripts (e.g., AL773604.8, AP001062.7, C21orf2). Their low impact in the VEGA model suggests limited involvement in pathway modulation or possible non-specific transcriptional noise in the disease context.

Notably, the stratification of these weights was consistent across treated and untreated SLE subgroups, implying that antigen cross-presentation machinery remains transcriptionally active regardless of immunosuppressive therapy, possibly reflecting a persistent autoimmune drive at the level of the proteasome and endosomal compartments.

3. Discussion

This project demonstrates the utility of biologically informed deep generative models, specifically VEGA, in elucidating immune dysregulation in systemic lupus erythematosus (SLE). By incorporating pathway-level annotations into the variational autoencoder framework, VEGA provides interpretable latent spaces that not only improve cell type separation but also enhance the detection of treatment-induced transcriptional changes—particularly those induced by interferon-beta (IFN-β).

A key finding of this analysis is the myeloid-specific activation of type I interferon pathways in response to IFN-β stimulation. Both interferon-α/β signaling and downstream antiviral ISG programs were enriched predominantly in CD14+ and FCGR3A+ monocytes, as well as dendritic cells. This selective response aligns with known biology, where myeloid cells act as primary sensors and effectors in the type I IFN response due to their high receptor expression and regulatory gene networks [3,4,5]. The low activation observed in lymphoid populations, such as T and B cells, suggests a compartmentalized immune response and reinforces the notion that innate immune dysregulation plays a pivotal role in SLE pathogenesis [2].

The observed treatment-induced transcriptional shift was clearly detectable in the VEGA latent space but was poorly resolved using a standard VAE. This highlights the advantage of incorporating prior biological knowledge into deep learning frameworks, enabling condition-specific stratification and improving biological interpretability. Such capability is critical in autoimmune diseases like SLE, where cellular heterogeneity and subtle transcriptional alterations complicate biomarker discovery and therapeutic targeting [1].

Further, the VEGA model effectively identified dominant gene contributors to the interferon and antigen presentation pathways. High-ranking interferon-stimulated genes (ISGs) such as ISG20, IFIT3, and ISG15 are well-known antiviral effectors involved in RNA degradation, viral restriction, and immune modulation [6,8,9]. Their prominence in the latent representation underscores the model's capacity to prioritize biologically relevant features without manual curation, a task often constrained in traditional bioinformatics approaches.

The pathway "cross-presentation of soluble exogenous antigens" was associated with proteasomal subunits (e.g., PSMC5, PSMD14), suggesting ongoing or enhanced antigen processing activity in SLE PBMCs regardless of IFN-β treatment. This observation aligns with evidence linking altered proteasomal function to autoimmune processes, potentially contributing to aberrant MHC class I antigen

presentation and autoantigen exposure [11,12,13]. Notably, the persistent expression of these components in both treated and untreated samples hints at a chronic immunological state that may underlie refractory disease mechanisms.

Overall, the results support the growing consensus that type I interferon signaling is central to SLE pathogenesis and may serve as a biomarker axis for both disease activity and therapeutic response [3,4]. The VEGA model not only reinforces this paradigm but also offers a scalable framework for dissecting heterogeneous immune states in other complex inflammatory conditions. Future work may extend this approach to longitudinal datasets and incorporate additional omics layers for more comprehensive mechanistic insight.

4. Methods and Materials

4.1 Data Preparation

A preprocessed single-cell RNA-seq dataset is loaded using scanpy.read_h5ad. This dataset contains peripheral blood mononuclear cells (PBMCs) from lupus patients and controls. The input data is converted to a dense array format and loaded into a PyTorch DataLoader with a batch size of 128 for efficient mini-batch training.

4.2 VAE Model Architecture

A Variational Autoencoder (VAE) is composed of two main parts: an encoder and a decoder, both utilizing fully connected layers. The encoder features two hidden layers, each with 800 units, employing ReLU activation and a dropout rate of 0.3. This part of the VAE is responsible for outputting the parameters (μ and σ) that define a Gaussian distribution in the latent space, which in this case has 50 dimensions. To sample latent vectors (z), the encoder applies the reparameterization trick, where $z=\mu+\sigma\cdot\epsilon$, with ϵ being sampled from a standard normal distribution, i.e., $\epsilon\sim N(0,1)$. To ensure that the latent space adheres to a Gaussian distribution, the encoder also computes the Kullback-Leibler (KL) divergence loss. The decoder mirrors the encoder's structure with two hidden layers, each containing 800 units, also utilizing ReLU activation and a 0.3 dropout rate. Its primary function is to reconstruct the original input data from the sampled latent vectors.

4.3 VAE Training

The Variational Autoencoder (VAE) is trained using the Adam optimizer with a learning rate of 0.0001. The loss function is a weighted sum of two components: the reconstruction loss, measured by the Mean Squared Error (MSE) between the original input and its reconstruction , and the Kullback-Leibler (KL) divergence. During each training epoch, the following steps occur: gradients are first reset using opt.zero_grad(). Then, a forward pass is executed to compute both the reconstructed output (x^) and the KL divergence loss. Finally, backpropagation is performed via loss.backward() to update the model's weights. Throughout the training process, the total loss, MSE loss, and KL loss are meticulously logged for each epoch, providing insights into the model's performance and convergence.

4.4 Data Projection and Visualization

First, the to_latent() function plays a crucial role in encoding the input single-cell RNA-seq data, typically found in adata.X, into a compact, low-dimensional latent space using a pre-trained Variational Autoencoder (VAE). This function takes the trained VAE model and an AnnData object as input. The process involves converting the sparse count matrix into a dense NumPy array and then into a PyTorch tensor. Importantly, the VAE is set to evaluation mode during this step, which disables features like dropout and batch normalization to ensure consistent inference. The VAE's encoder network then

computes the latent representation, denoted as z, for each individual cell. Finally, these latent embeddings are moved back to the CPU, concatenated, and returned as a NumPy array. These valuable latent embeddings are then stored within adata.obsm for subsequent analysis.

Following the projection into latent space, a k-nearest-neighbor (kNN) graph is constructed to capture similarities between cells. This is achieved using sc.pp.neighbors() from the Scanpy library. It specifically utilizes the 50-dimensional VAE latent space and considers n_neighbors=15 to strike a balance between local and global structural relationships.

Next, Uniform Manifold Approximation and Projection (UMAP) is applied for nonlinear dimensionality reduction, allowing for a 2D visualization of the data. The sc.tl.umap() function is used for this purpose, with random_state=13 ensuring the reproducibility of the visualization.

Finally, the cell clusters are visualized on the UMAP plot, colored according to metadata annotations. The sc.pl.umap() function is employed to display the data, with cells stratified by biological labels such as cell types and disease states using color=["cell_type", "condition"]. The size=10 parameter adjusts the point size for better clarity, and the resulting figure is exported as _celltype_condition.png.

4.5 Mask Creation

The create_mask function is designed to generate a binary matrix that establishes a link between genes and their associated biological pathways. This function is then data, specifically utilizing Reactome pathway definitions. In this application, one additional node is added to accommodate unannotated genes. Consequently, the resulting mask matrix will have dimensions of (number of genes × number of pathways + 1).

4.6 VEGA Model Architecture

The Encoder Module within VEGA is built upon a base Encoder class. Its primary function is to map high-dimensional gene expression data into a lower-dimensional latent space. The architecture of this encoder consists of two fully-connected layers, each containing 800 neurons and utilizing ReLU activation. To promote regularization and prevent overfitting, a dropout rate of p=0.3 is applied. The encoder outputs the mean and log-variance parameters that characterize the latent distribution. The Kullback-Leibler (KL) divergence for this module is also calculated. The Pathway-Aware Decoder (DecoderVEGA) is a distinctive feature of the VEGA model, implemented through a SparseLayer. This SparseLayer is initialized with a binary pathway mask, which has dimensions of (genes × pathways). This mask plays a crucial role in enforcing sparsity by performing element-wise multiplication with the decoder's weights. The decoder's weight matrix itself has dimensions of (pathways × genes) and includes a trainable bias term. A custom Autograd Function, SparseLayerFunction, is employed to manually implement the forward and backward passes, ensuring that the mask constraints are preserved during gradient updates. Furthermore, non-negativity enforcement is applied to the decoder's weights; these weights are clamped to be greater than or equal to zero via a positive_weights() method, which is executed after each optimization step.

4.7 VEGA Training

The VEGA model is trained using the Adam optimizer . Its loss function, combines Mean Squared Error for reconstruction with a KL divergence regularization term. During training, data is processed in 128-sample batches on the GPU. For each batch, gradients are zeroed, the encoder maps input x to latent z, and the decoder reconstructs $\hat{x}$. Loss is then computed. Backpropagation updates parameters, followed by enforcing non-negative decoder weights. Training progress is monitored by logging epoch-averaged total, MSE, and KL losses.

4.8 Latent Space Projection

Cell embeddings in a pathway-informed latent space using a pre-trained VEGA model were generated. The to_latent() function converts the sparse gene expression data into a dense tensor, sets the VEGA model to evaluation mode, and then computes these latent embeddings. The output is a NumPy array representing (number of cells × number of pathways), where each dimension corresponds to a biological pathway or an unannotated node.

Next, dimensionality was reduced to enhance visualization and preserve data structure. A k-nearest-neighbor (kNN) graph is built using sc.pp.neighbors(), setting k=15 and using Euclidean distances within the latent space to preserve local manifold structure. Finally, UMAP projects the cells into 2D, with random_state=13 ensuring reproducible visualization that maintains global relationships between cell clusters.

4.9 Visualization & Interpretation

UMAP plots are generated using sc.pl.umap(), colored by cell_type to identify biological clusters and condition to reveal disease associations. To analyze specific pathway activity, pathway names are retrieved from VEGA metadata, and latent dimensions, representing pathway activity scores, are extracted. Scores for the pathway of interest are isolated and stored. Finally, a UMAP plot colored by this pathway's activity visualizes cell subsets with varying levels of signaling, revealing condition-specific patterns.

To identify key genes influencing a biological pathway (e.g., interferon signaling) using a trained VEGA model, the decoder's gene-pathway weight matrix are extracted. Then biological annotations to identify the target pathway's column index are mapped. Genes are ranked by their descending weight magnitudes for that pathway.

References

1.Seninge, L., Anastopoulos, I., Ding, H. *et al.* VEGA is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. *Nat Commun* **12**, 5684 (2021)

2. Tsokos GC. Systemic lupus erythematosus. N Engl J Med. 2011 Dec 1;

3. Crow MK. Type I interferon in organ-targeted autoimmune and inflammatory diseases. Arthritis Res Ther. 2010;

4. Rönnblom L, Leonard D. Interferon pathway in SLE: one key to unlocking the mystery of the disease. Lupus Sci Med. 2019 Aug 13;

5. Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. Curr Opin Virol. 2011 Dec;

6. Schneider WM, Chevillotte MD, Rice CM. Interferon-stimulated genes: a complex web of host defenses. Annu Rev Immunol. 2014;

7. Goubau D, Deddouche S, Reis e Sousa C. Cytosolic sensing of viruses. Immunity. 2013 May 23;

8. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, Rice CM. A diverse range of gene products are effectors of the type I interferon antiviral response. Nature. 2011 Apr 28;

9. Fensterl V, Sen GC. Interferon-induced Ifit proteins: their role in viral pathogenesis. J Virol. 2015 Mar;

10. Sadler AJ, Williams BR. Interferon-inducible antiviral effectors. Nat Rev Immunol. 2008 Jul;

11. Tanaka K. The proteasome: overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci. 2009;

12. Mishto M, Santoro A, Bellavista E, Bonafé M, Monti D, Franceschi C. Immunoproteasomes and immunosenescence. Ageing Res Rev. 2003 Oct;

13. Basler M, Kirk CJ, Groettrup M. The immunoproteasome in antigen processing and other immunological functions. Curr Opin Immunol. 2013 Feb;