**Model description**

We developed SPICESS (Spatially-informed Protein Imputation using Constrained Embeddings from Shared latent Spaces), a joint graph variational autoencoder based contrastive learning algorithm that leverages spatial CITE-seq data to predict protein expression levels from gene expression data

We use two over-parameterized graph convolutional neural networks to generate separate gene and protein feature spaces. The input data consists of a normalized gene count matrix (Y), a normalized protein expression count matrix (P) and spatial coordinates of each spot (S). We construct a graph representation of each modality based on nearest neighbors where each spot is a node and gene or protein expression are node features. The generated graph embeddings capture both modalities as well as the spatial context. We employ a graph contrastive representation learning approach based on maximizing the agreement of different graph views at the node level [1]. Specifically, our model learns by generating graph views using two schemes - removing edges randomly and masking corresponding mRNA and protein features - and then applying a contrastive loss to maximize the agreement of two perturbed graphs. As the contrastive learning pulls together of the representation from two corrupted views, it enforces the model to learn essential biological information that is insensitive to perturbation in the embedding space. The graph-encoders of gene and protein are jointly trained using this self-supervised approach to generate spatially-consistent embeddings which are then aggregated into a shared latent space, Z. To ensure that this learned shared representation maximally captures relevant signals from each input modality, we use three separate decoders that reconstruct spot adjacency, gene and protein expression matrices.

To maximize generalizability, SPICESS employs 2 levels of built-in batch correction mechanisms. Firstly, by expanding the hidden layer sizes (and therefore the shared embedding space) to be larger than the input feature dimensions both graph encoders learn to correct for sample-sample variations by always mapping new samples to the training reference anchor points in the latent space [6-8]. Secondly, when multiple training samples are available, SPICESS uses an additional adversarial discriminator network that incentivizes the graph encoders to generate embeddings that carry no information about which batch a randomly sampled spot originated from [8], therefore removing any batch effect signal.

After training the model, the gene encoder and protein decoder can be coupled to impute spot-specific protein expression from gene expression for new spatial transcriptomics datasets. Additionally, protein expression can also be imputed directly from new H&E images using a query-reference technique similar to SeuratV3 [9]. Input image patches (queries) are projected into the shared latent space to index the 10 nearest embedding vectors that are linearly combined and used as input for the protein decoder to generate the imputed protein profiles for each image patch.

1. Zhu, Yanqiao, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. 2020. Deep Graph Contrastive Representation Learning. arXiv. <http://arxiv.org/abs/2006.04131>.
2. Liu, Zhuang, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. 2022. A ConvNet for the 2020s In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. <https://doi.org/10.1109/cvpr52688.2022.01167>.
3. Xie, Ronald, Kuan Pang, Sai W. Chung, Catia T. Perciani, Sonya A. MacParland, Bo Wang, and Gary D. Bader. 2023. Spatially Resolved Gene Expression Prediction from H&E Histology Images via Bi-Modal Contrastive Learning. arXiv. <http://arxiv.org/abs/2306.01859>.
4. Anand, Deepak, Goutham Ramakrishnan, and Amit Sethi. 2019. Fast GPU-Enabled Color Normalization for Digital Pathology. In 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE. <https://doi.org/10.1109/iwssip.2019.8787328>.
5. Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. 2021. Learning Transferable Visual Models from Natural Language Supervision. arXiv. <http://arxiv.org/abs/2103.00020>.
6. Radhakrishnan, Adityanarayanan, Mikhail Belkin, and Caroline Uhler. 2020. Overparameterized Neural Networks Implement Associative Memory. Proceedings of the National Academy of Sciences of the United States of America 117 (44).
7. Jain, Saachi, Adityanarayanan Radhakrishnan, and Caroline Uhler. 2021. A Mechanism for Producing Aligned Latent Spaces with Autoencoders. arXiv. <http://arxiv.org/abs/2106.15456>.
8. Zhang, Xinyi, Xiao Wang, G. V. Shivashankar, and Caroline Uhler. 2022. Graph-Based Autoencoder Integrates Spatial Transcriptomics with Chromatin Images and Identifies Joint Biomarkers for Alzheimer's Disease. Nature Communications 13 (1): 7480.
9. Satija, Rahul, Jeffrey A. Farrell, David Gennert, Alexander F. Schier, and Aviv Regev. 2015. Spatial Reconstruction of Single-Cell Gene Expression Data. Nature Biotechnology 33 (5).