Johnson&Johnson Innovative Medicine

Gene regulatory network (GRN) inference from scRNA-seq data:

Supervised learning (with known interactions):



$$p_{ik} = f(x_i, x_k), y_{ik} \sim \text{Bernoulli}(p_{ik}).$$

Unsupervised learning:



DeepSEM for GRN inference:

Generative Model:

 $Z \sim N(0, \sigma^2)$; $A \sim Laplace(0, \sigma_a^2)$;

$$\hat{Z} = (I - A)^{-1} Z;$$

 $X \sim N\left(f_{\theta}(\hat{Z}), g_{\theta}(\hat{Z})\right).$



InfoSEM - DeepSEM with informative priors

InfoSEM-B:

Informative prior of *A* with gene embeddings *H*:

 A_{ik} ~Laplace($[h_i, h_k]w, \sigma_a^2$).

Gist:

1. Gene embedding represent gene functions;

2. Genes with similar function have similar interactions.

InfoSEM-BC:

 $A_{ik} = A^e_{ik} \times A^p_{ik}.$

Informative prior of *A* with1. gene embeddings *H*;2. known interaction *Y*.





$$\operatorname{Logit}(A_{ik}^{p}) \sim \begin{cases} N(\operatorname{Logit}(0.95), \sigma_{p}^{2}), & \text{if } Y_{ik} = 1; \\ N(\operatorname{Logit}(0.05), \sigma_{p}^{2}), & \text{if } Y_{ik} = 0; \\ N(\operatorname{Logit}(0.50), \sigma_{p}^{2}), & \text{if } Y_{ik} \text{ is unknown}; \end{cases}$$



InfoSEM: A Deep Generative Model with Informative Priors for Gene Regulatory Network Inference Tianyu Cui, Song-Jun Xu, Artem Moskalev, Shuwei Li, Tommaso Mansi, Mangal Prakash*, Rui Liao*

oskalev, Shuwei Li, Tommaso Mansi, Mangal Prakash*, Rui Liao* J&J Innovative Medicine





Current DL-Based GRN Benchmarks Focus on Seen Genes, Neglecting Unseen Genes Critical for Biomarker Discovery



(a) train set (known GRN for training)

DL-Based supervised methods rely on gene-specific biases, mimicking 1-hot LR for Seen Genes setup without generalizing to unseen genes





(0) 2000 8000 000 000 (0) 1

sed nes	
not LR	
omp	
EAT	
ELink	
vised	
ies	
Boost	
SEM	
5	
EM-B	
EM-BC	

Key contributions:

New benchmark for unseen gene interactions aligning with real-world applications such as biomarker discovery

InfoSEM, a SOTA unsupervised model for GRN inference using gene embeddings and GT labels, improving performance by 38.5% over existing models





SL methods achieve high accuracy on seen gene test set by learning gene-specific class imbalance levels, which do not generalize to unseen gene test set.

Should evaluate on unseen gene test sets!

Our generative framework achieved state-of-the-art accuracy on unseen gene test sets by incorporating various priors!

InfoSEM is flexible to support gene embeddings from any resources:

