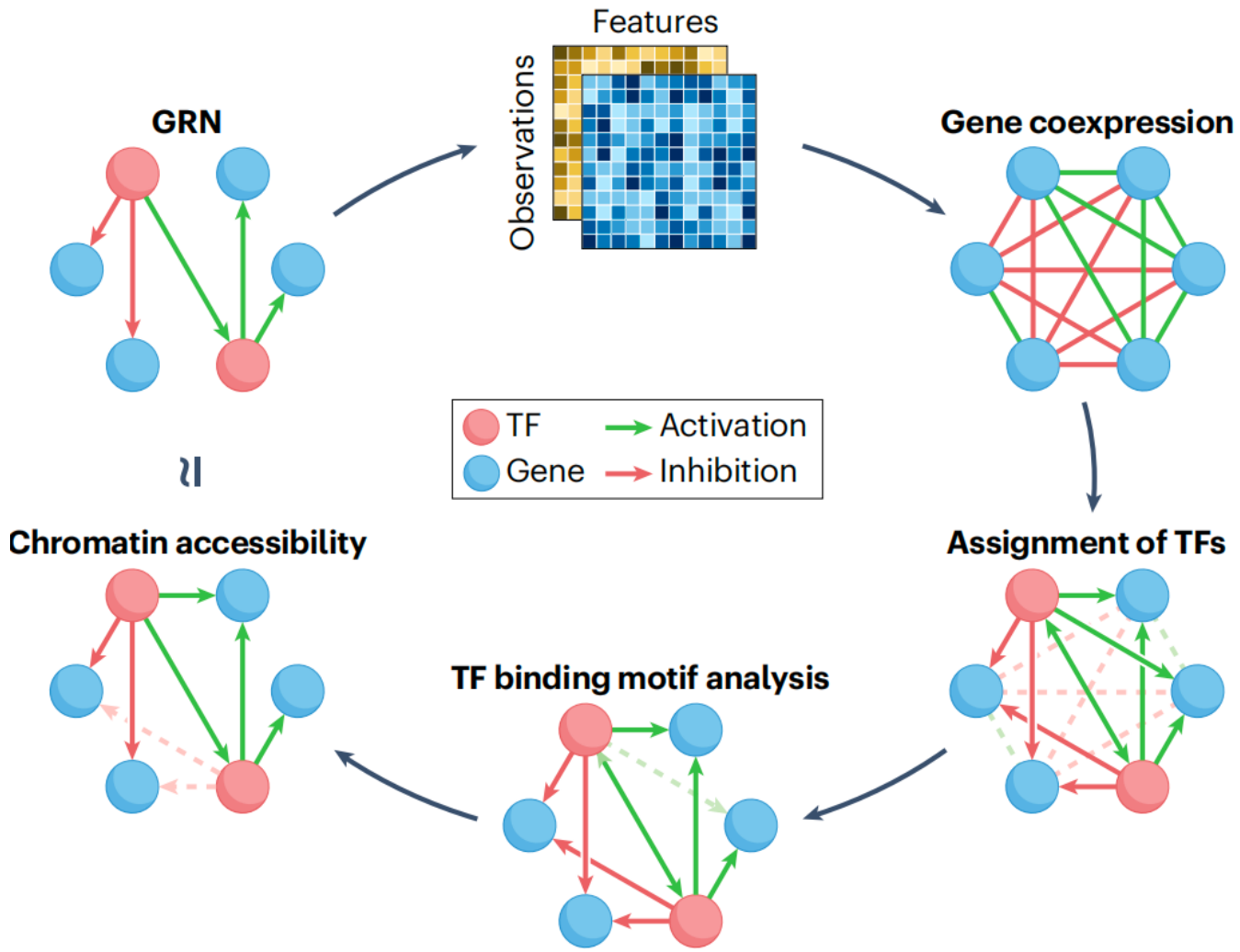
KINDLE: Knowledge-Guided Distillation for Prior-Free Gene Regulatory Network Inference

Rui Peng, Yuchen Lu, Qichen Sun,
Yuxing Lu, Chi Zhang, Ziru Liu, JinzhuoWang

Author contact information:

Gene regulatory network (GRN) inference

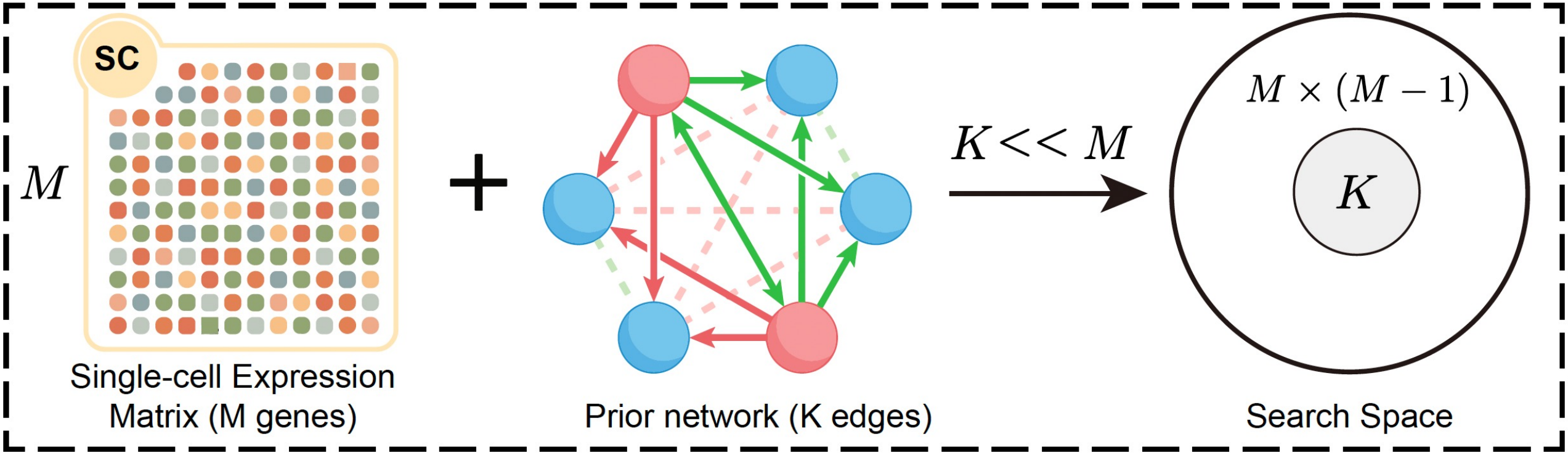


GRN inference refers to the process of summarizing gene regulation — a highly complex and dynamic process — into an interpretable network structure from data using computational methods.

Interactions in GRNs can be directed or undirected , signed (denoting the mode of regulation, positive or negative) or weighted (denoting the strength of the interaction).

Integrating Prior knowledge: chromatin accessibility, TF binding motif analysis

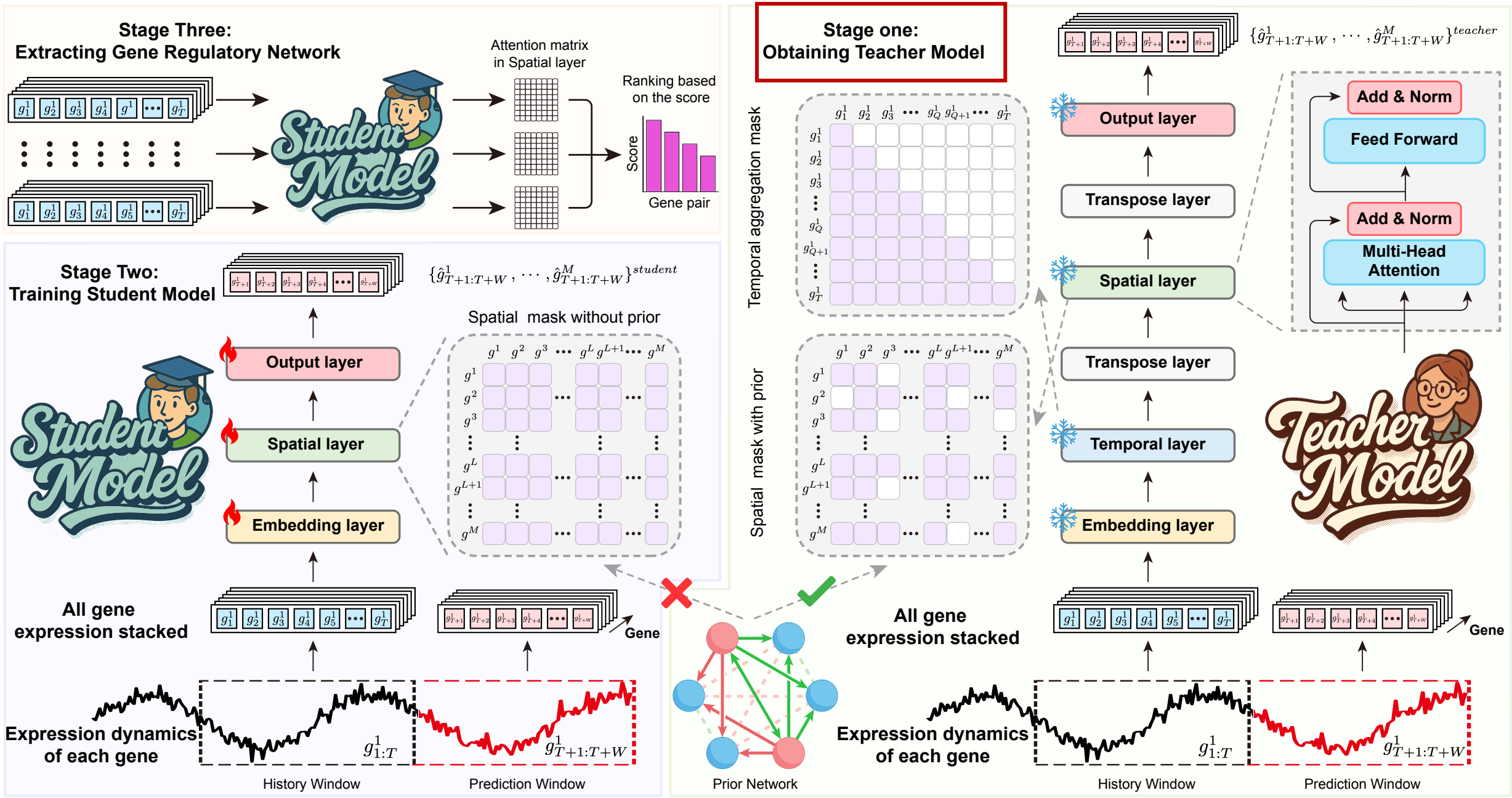
Limitation of prior information



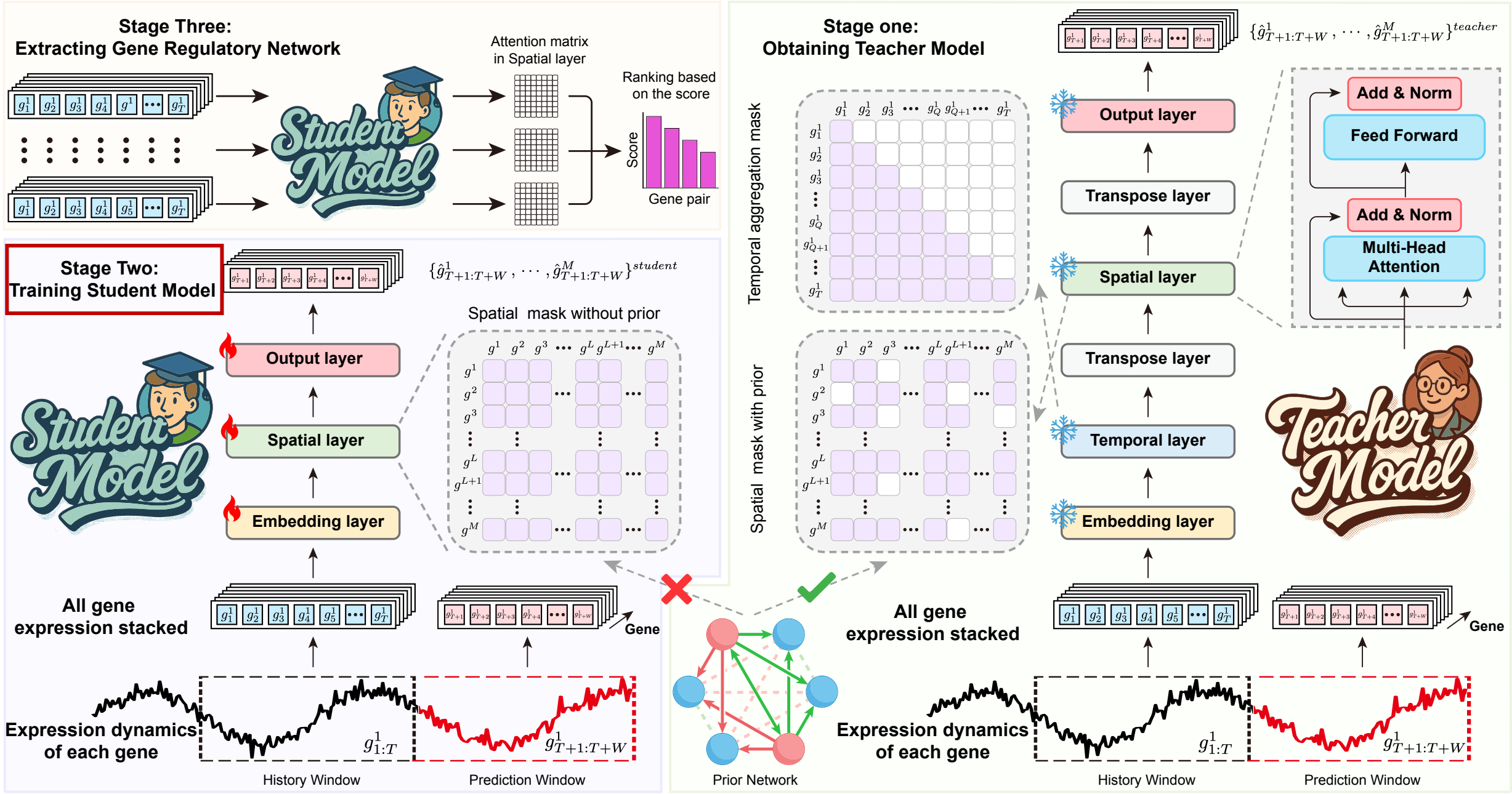
Firstly, its performance depends on the overlap between the prior and the ground truth network.

Secondly, limiting candidate edges to the prior prevents the detection of regulatory interactions absent from it, ruling out the discovery of new regulatory relationships.

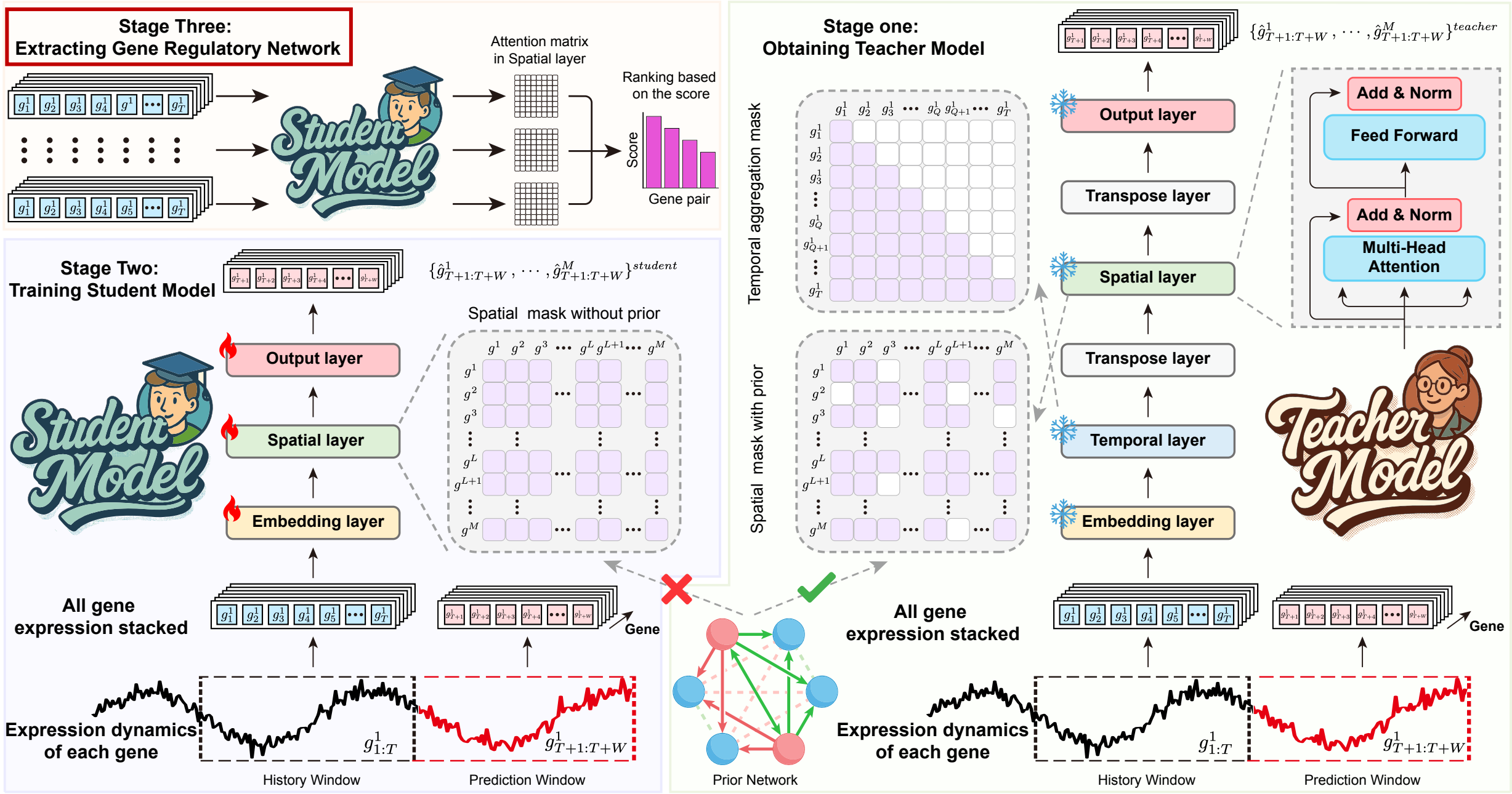
KINDLE: Knowledge-guided network distillation for prior-free GRN inference



KINDLE: Knowledge-guided network distillation for prior-free GRN inference



KINDLE: Knowledge-guided network distillation for prior-free GRN inference



The four variants of KINDLE

Hard Distillation

$$\mathcal{L}_{\text{distill}}\left(f_{\theta_S}(\mathbf{G}_{1:T})_{b,m}, f_{\theta_T}(\mathbf{G}_{1:T})_{b,m}\right) = \left\|f_{\theta_S}(\mathbf{G}_{1:T})_{b,m} - f_{\theta_T}(\mathbf{G}_{1:T})_{b,m}\right\|_2^2$$

Soft Distillation

$$\mathcal{L}_{\text{distill}}\left(f_{\theta_S}(\mathbf{G}_{1:T})_{b,m}, f_{\theta_T}(\mathbf{G}_{1:T})_{b,m}\right) = \mathbf{KL}\left(\sigma\left(\frac{f_{\theta_S}(\mathbf{G}_{1:T})_{b,m}}{\tau}\right)\middle\|\sigma\left(\frac{f_{\theta_T}(\mathbf{G}_{1:T})_{b,m}}{\tau}\right)\right)$$

Bilinear Pool

$$\mathcal{K}_{\text{bilinear}}(f_{\theta_S}(\mathbf{G}_{1:T})_{b,m}, f_{\theta_T}(\mathbf{G}_{1:T})_{b,m}) = (f_{\theta_S}(\mathbf{G}_{1:T})_{b,m})^\top (f_{\theta_T}(\mathbf{G}_{1:T})_{b,m})$$

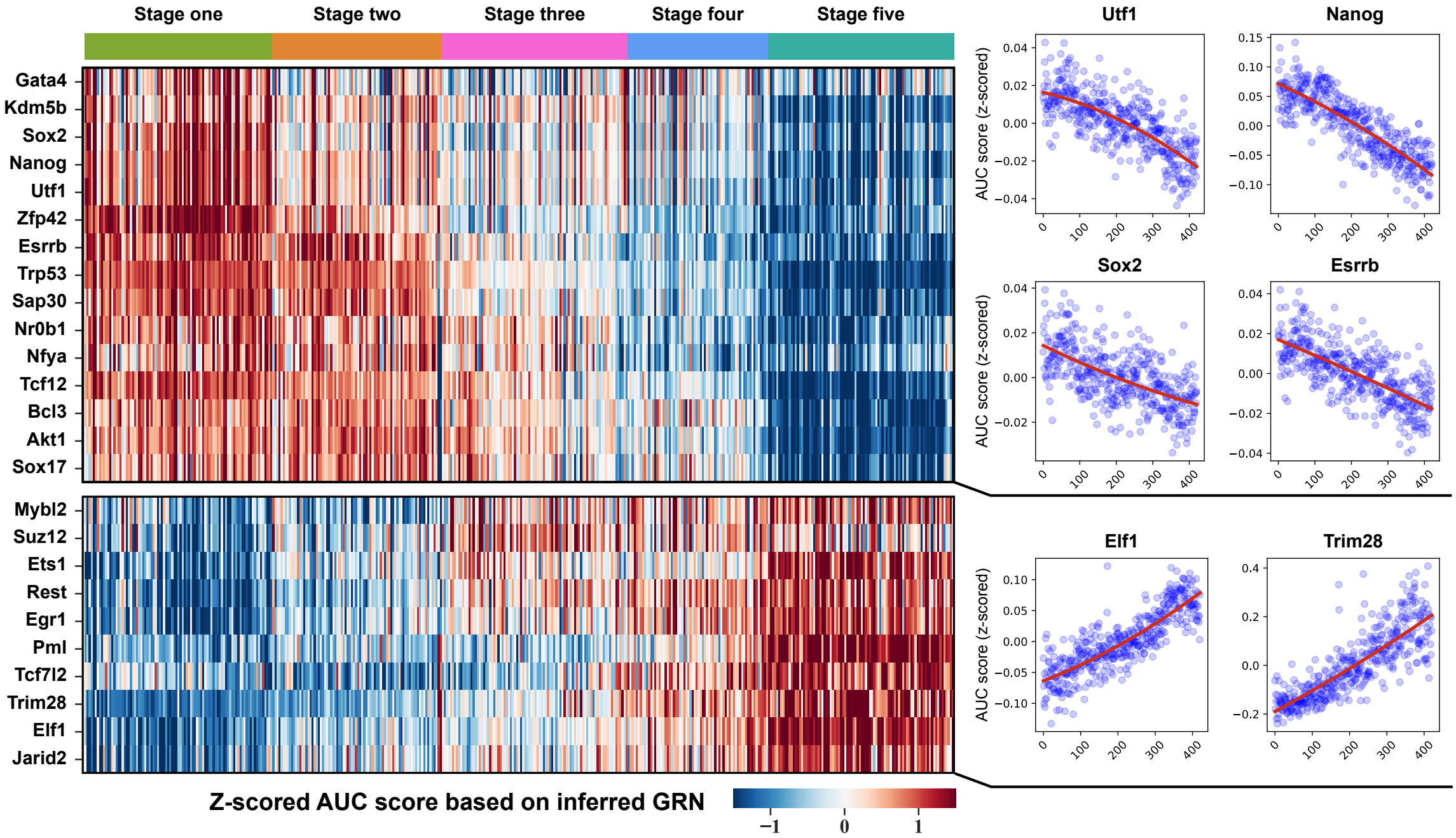
Gaussion RBF

$$\mathcal{K}_{\text{gaussion}}\left(f_{\theta_S}(\mathbf{G}_{1:T})_{b,m}, f_{\theta_T}(\mathbf{G}_{1:T})_{b,m}\right) = exp\left(-\frac{\left\|f_{\theta_S}(\mathbf{G}_{1:T})_{b,m} - f_{\theta_T}(\mathbf{G}_{1:T})_{b,m}\right\|_2^2}{2\lambda^2}\right)$$

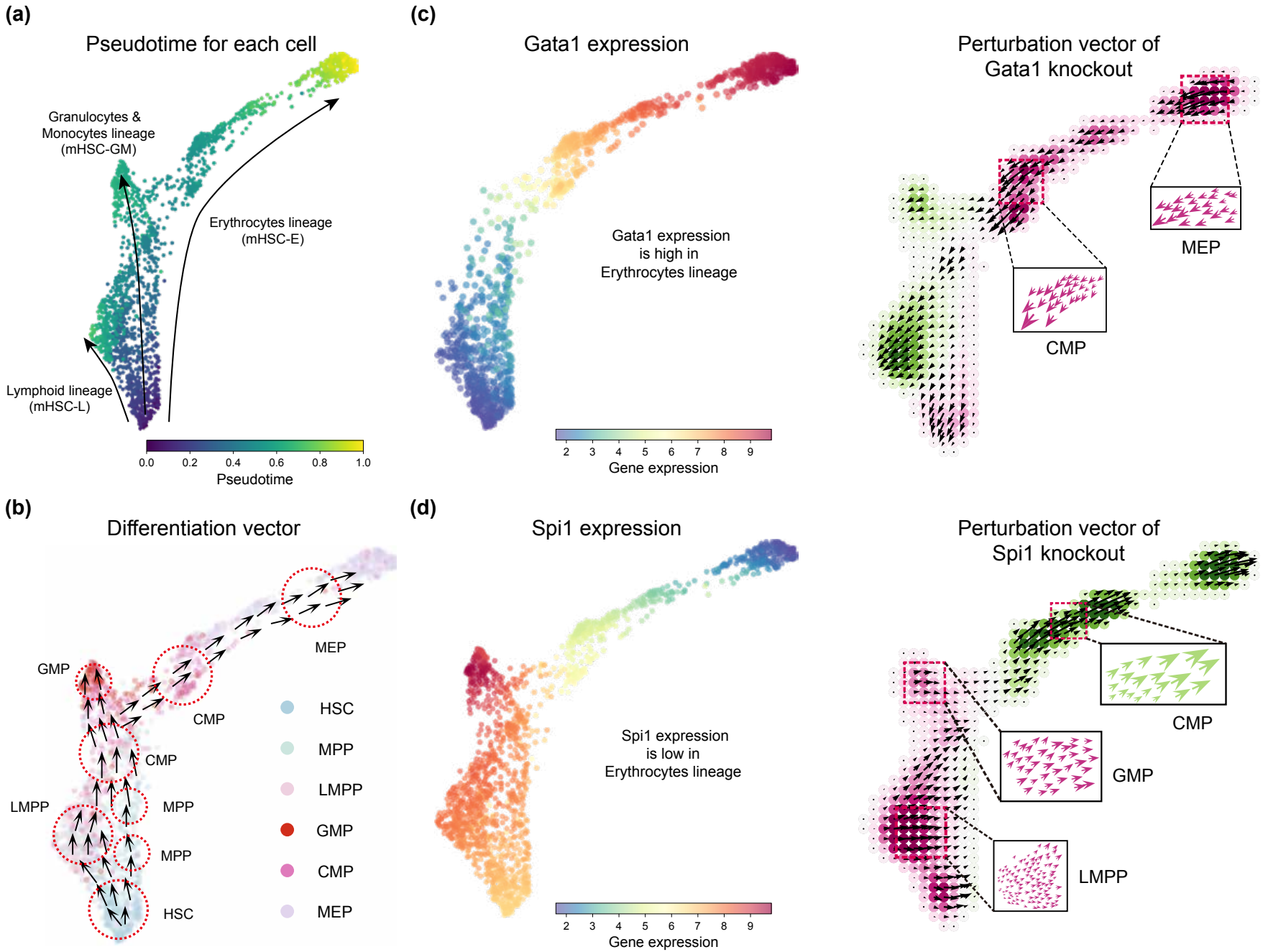
Comparison with other SOTA GRN inference methods

	Methods	mESC			mHSC-E			mHSC-L			mHSC-GM		
		AUROC	AUPRC	F1	AUROC	AUPRC	F1	AUROC	AUPRC	F1	AUROC	AUPRC	F1
Without Prior	**GRNBoost2 [8]**	0.537	0.127	0.203	0.397	0.034	0.087	0.515	0.181	0.297	0.474	0.083	0.146
	GENIE3 [9]	0.545	0.137	0.218	0.381	0.042	0.108	0.486	0.183	0.322	0.437	0.078	0.162
	Random	0.506	0.083	0.152	0.493	0.087	0.161	0.518	0.135	0.227	0.504	0.083	0.154
Prior Guided	**CEFCON [15]**	0.479	0.253	0.429	0.531	0.405	0.551	**0.653**	0.659	0.675	0.457	0.444	0.647
	Celloracle [16]	0.490	0.177	0.305	0.536	0.290	0.420	0.557	0.277	0.368	0.487	0.243	0.401
	NetREX [19]	0.522	0.128	0.217	0.511	0.117	0.211	0.520	0.177	0.282	0.526	0.144	0.219
	Prior_Random	0.498	0.318	0.482	0.492	0.389	0.570	0.522	0.551	0.691	0.509	0.464	0.627
KINDLE	**KINDLE (Soft distillation)**	0.747	0.636	0.519	0.561	0.559	0.691	0.599	0.670	0.752	0.562	0.789	0.864
	KINDLE (Hard distillation)	0.753	0.643	0.526	0.564	0.578	0.711	0.599	0.669	0.757	0.569	0.793	0.871
	KINDLE (Bilinear Pool)	0.751	0.644	0.521	0.551	0.574	0.723	0.567	0.581	0.761	0.561	0.787	0.867
	KINDLE (Gaussian RBF)	**0.757**	**0.646**	**0.529**	**0.594**	**0.601**	**0.731**	0.600	**0.672**	**0.763**	**0.570**	**0.799**	**0.875**

Temporal dynamics of TF regulatory performance



In silico perturbation



Conclusion

We propose KINDLE to eliminate prior dependence in GRN inference by knowledge distillation, which achieves state-of-the-art performance across four benchmark datasets without requiring prior knowledge.

On mouse embryonic stem cell development data, KINDLE successfully identifies key TFs and predicts their functional roles during differentiation processes.

For mouse hematopoietic stem cell development, KINDLE accurately predicts the effects of Gata1 and Spi1 knockouts on cell fate determination, demonstrating its capability to capture critical regulatory mechanisms.