GRASP: ADAPTIVE CONTEXT WITH REGULATORY PARTNER SUPERVISION FOR GRN INFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Supervised gene regulatory network inference from single cell transcriptomic data seeks to rank unobserved transcription factor-target links by combining expression profiles with partial interaction references. Graph-based approaches exploit such references to provide relational context for prediction. However, local message passing remains bounded by incomplete reference topology. Moreover, same gene contrastive alignment does not explicitly exploit training positive TF-target partners when defining positive comparisons. To address these issues, we propose a framework named GRASP, which couples adaptive context construction with partner-based contrastive supervision for GRN inference. Technically, GRASP uses expression guided edge retention to adapt local propagation to the measured cellular context, and further constructs a teacher induced higher-order view to provide context beyond recorded graph neighborhoods. To exploit relational supervision already available in training interactions, neighborhood contrastive learning augments same gene correspondence with fixed training partners as positives within and across the two views, while a role-specific decoder produces ordered TF-target scores. We additionally employ pair-disjoint evaluation to prevent held out gene-pair connectivity from entering the training graph. Across 42 evaluation settings, GRASP achieves the highest average AUPRC among eight evaluated baselines. Code is available at https://anonymous.4open.science/r/GRASP-302C/.

1 INTRODUCTION

Gene regulatory network (GRN) describes how transcription factors (TFs) regulate gene expression and shape cellular identity and state. The increasing availability of single cell RNA sequencing has provided high resolution measurements for studying these regulatory programs, but inferring a reliable GRN from expression measurements remains challenging because regulatory interactions are only indirectly reflected in observational transcriptomes and available reference networks provide incomplete and context dependent interaction knowledge (Pratapa et al., 2020; Kamimoto et al., 2023). A practically important setting is therefore to combine cell-specific expression profiles with available interaction references to infer previously unobserved regulatory relationships. In this work, we rank ordered TF-target pairs for supervised GRN inference under partial reference supervision.

Early computational approaches infer regulatory interactions primarily from statistical dependencies in gene expression. GENIE3 (Huynh-Thu et al., 2010), for example, decomposes GRN inference into target-specific regression problems and ranks candidate regulators by their predictive importance, while GRNBoost2 (Moerman et al., 2019) improves the scalability of tree based network inference for single cell data. Although such methods extract regulatory signals directly from transcriptomic measurements, they do not explicitly exploit previously observed gene relationships as relational context for representation learning. This has motivated supervised graph-based GRN inference, which combines expression profiles with partial interaction references to propagate relational information and predict unobserved TF-target links. For example, GENELink (Chen & Liu, 2022) uses graph attention to aggregate information along connections in the reference graph. GRANet (Zhou et al., 2025) incorporates richer expression information into graph aggregation, combining standardized, smoothed, and discretized features through residual graph attention. GCLink (Yu et al., 2025) introduces graph perturbation and cross-view contrastive learning to improve representation consistency across alternative graph views. Together, these methods combine expression features with relational context from reference graph connectivity to learn gene representations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Despite this progress, two limitations remain in how graph-based GRN models construct gene context and define contrastive supervision. **First, local message passing and edge deletion views remain constrained by the supplied interaction topology.** Genes acquire relational context by aggregating information along available propagation edges (Gilmer et al., 2017), so missing reference interactions also remove potential communication routes. For example, edge deletion augmentation creates alternative local views by removing observed connections, but cannot introduce additional routes beyond the supplied graph (Yu et al., 2025). Moreover, uniform removal assigns the same deletion probability to observed interactions regardless of their expression association in the cellular context. Such views therefore create alternative local propagation patterns within an incomplete reference, but cannot provide context beyond its recorded connectivity or use expression association to determine which routes to retain. **Second, same gene contrastive alignment overlooks training supported TF-target partnerships when defining positive comparisons.** GRN inference ranks relationships between different genes, and training positive TF-target pairs provide partner information beyond the identity of individual genes. This information can complement same gene correspondence by specifying which cross gene relationships receive positive supervision across views. However, instance level contrastive alignment rewards only the same gene's cross-view match. For a TF anchor, its recorded training targets remain non-positive competitors alongside other genes, despite their observed partner relationship with that TF (Yu et al., 2025). The auxiliary objective therefore promotes consistency of individual gene identities without explicitly reinforcing training supported partnerships through its positive comparisons. Perturbing local routes or introducing higher-order views does not by itself supply this relational supervision, because it changes the representations being compared rather than the positive pair definition. Together, these limitations motivate adaptive and complementary gene context coupled with contrastive supervision that explicitly incorporates training positive partners.

To address these challenges, we propose Graph-based Regulatory inference with Adaptive context and Stable Partner supervision (GRASP), a framework that couples adaptive context construction with partner-based contrastive supervision for GRN inference. Specifically, GRASP uses expression association to guide student side edge retention, allowing local propagation to adapt to the measured cellular context. Because this local view remains bounded by recorded connectivity, a momentum teacher further induces an auxiliary higher-order representation that provides context beyond explicit graph neighborhoods. To exploit the cross gene partner information available in training interactions within the contrastive objective, Neighborhood Contrastive Learning augments same gene correspondence with fixed training partners as positives within and across the local and higher-order views. A role-specific decoder then transforms the learned representations into ordered TF-target scores. We further adopt a pair-disjoint evaluation protocol that prevents held out gene-pair connectivity from entering the training graph under symmetrized message passing. Across 42 settings spanning seven cell types, three reference networks, and two gene set sizes, GRASP achieves the highest average AUPRC among eight evaluated baselines.

2 METHOD

Notations. Let V contain N genes, $T \subseteq V$ the TFs, and $X \in \mathbb{R}^{N\times C}$ their expression across C cells, standardized per gene. An ordered candidate $(u, v) \in T \times V$, $u \neq v$, has label $y_{uv} \in \{0, 1\}$. Training positives $\mathcal{E}_{\text{tr}}^{+}$ are reference links; negatives are sampled unobserved candidates. Let $A \in \mathbb{R}^{N\times N}$ denote the symmetric nonnegative training adjacency matrix.

Design principle. GRASP (Figure 1) combines local and higher-order context. Expression guided local propagation adapts to cellular context, while a teacher induced higher-order view provides context beyond recorded neighborhoods. Neighborhood Contrastive Learning augments same gene correspondence with training positive partners as symmetric positives across changing views. A role-specific decoder produces ordered TF-target scores.

2.1 ADAPTIVE LOCAL CONTEXT FROM EXPRESSION EVIDENCE

The observed training graph provides useful local relational context, but it treats all stored interactions as equally available communication routes. Expression measurements offer an additional signal for adapting this local context to the cellular condition. We therefore construct the student view through expression guided edge retention. For each training positive, let $r_{uv} \in (0, 1]$ denote the normalized

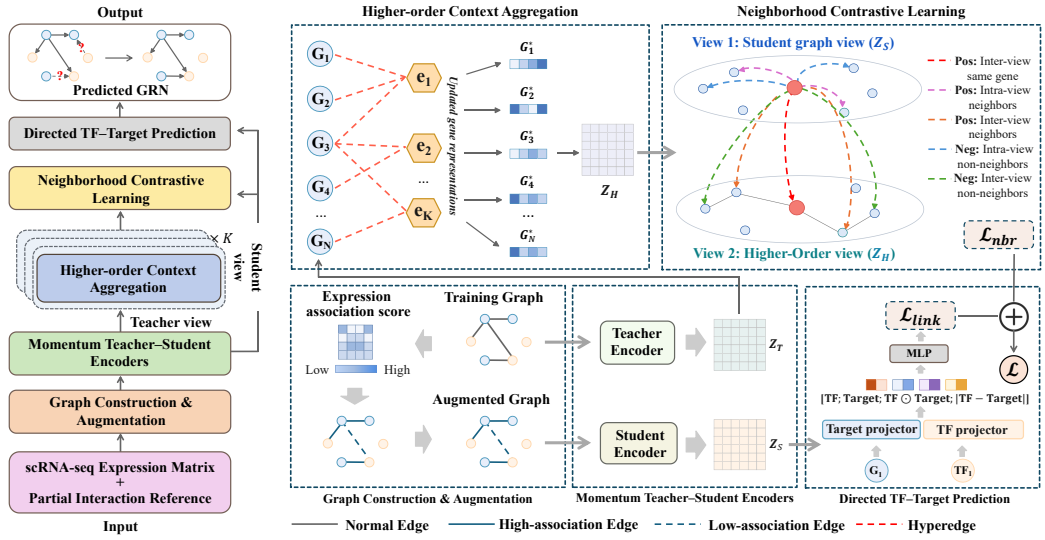


Figure 1: Illustration of the proposed framework GRASP.

percentile rank of its absolute Pearson correlation. With removal scale $p_e \in [0, 1]$ and exponent $\gamma > 0$, we sample an edge retention mask M and form the student adjacency:

$$r_{uv} = \mathrm{RankPct}_{\mathcal{E}_{\mathrm{tr}}^+}(|\,\mathrm{corr}(X_u, X_v)|),$$
$$M_{uv} \sim \mathrm{Bernoulli}(1 - p_e(1 - r_{uv})^\gamma)\,, \qquad \widetilde{A} = A \odot M. \tag{1}$$

Stronger expression associations yield higher retention probabilities, and setting $\gamma = 0$ recovers uniform edge removal with retention probability $1 - p_e$. Resampling changes student communication routes while leaving the training positive relation set unchanged.

The student produces $Z_s = f_\theta(X, \widetilde{A}) \in \mathbb{R}^{N\times d}$ using a two layer multihead GAT (Veličković et al., 2018). Appendix C details sampling and attention over the retained edges.

2.2 Higher-order context beyond recorded neighborhoods

Expression guided sampling adapts the local context, but it still operates over relationships already present in the training graph. To complement this topology bounded view, we construct an auxiliary higher-order context from stable teacher representations.

Stable teacher representations. The momentum teacher shares the student's architecture but encodes the unperturbed training topology, $Z_t = \mathrm{stopgrad}(f_{\bar{\theta}}(X, A)) \in \mathbb{R}^{N\times d}$, with exponential-moving-average (EMA) weights and dropout disabled (Thakoor et al., 2022).

Context pooling. Projection $W \in \mathbb{R}^{d\times K}$ learns gene associations with K latent hyperedges:

$$S = Z_t W \in \mathbb{R}^{N\times K}, \qquad U^{(0)} = S^\top Z_t \in \mathbb{R}^{K\times d}, \tag{2}$$

where S_{ik} weights gene i in hyperedge k during pooling and redistribution.

Higher-order interaction. We refine the pooled representations through L_h residual mappings:

$$U^{(\ell+1)} = U^{(\ell)} + \rho(R_\ell U^{(\ell)} + b_\ell \mathbf{1}_d^\top), \qquad \ell = 0, \ldots, L_h - 1, \tag{3}$$

where $R_\ell \in \mathbb{R}^{K\times K}$ mixes information across hyperedges, $b_\ell \in \mathbb{R}^K$ is a bias vector, $\mathbf{1}_d \in \mathbb{R}^d$ is the all ones vector, and ρ applies LeakyReLU elementwise in each residual update.

Context redistribution. The associations S redistribute the refined context to genes:

$$Z_h = \rho(S U^{(L_h)}) \in \mathbb{R}^{N\times d}. \tag{4}$$

Genes associated with shared factors exchange information through Z_h even without a direct training edge. Contrastive gradients update W, R_ℓ, and b_ℓ through Z_h, while the teacher representations Z_t are held fixed during each optimization step of the student.

Unlike local message passing, which is confined to the edges of the propagation graph, the higher-order mapping is not bounded by its connectivity. We formalize this distinction in Theorem 1, where G denotes the composition of the residual maps in Equation (3).

Theorem 1 (Equivariance and nonlocal dependence)**.** *Let*

$$F(Z) = \rho\big(ZW\, G(W^\top Z^\top Z)\big)\,, \qquad Z \in \mathbb{R}^{N\times d},$$

denote the higher-order mapping in Equations (2)–(4).
(i) *For any fixed parameters and any gene permutation matrix P,*

$$F(PZ) = PF(Z).$$

(ii) *For $N, d, K \geq 2$, there exist parameters and representations $Z, Z' \in \mathbb{R}^{N\times d}$ differing only in the row of gene v such that, for another gene $u \neq v$,*

$$\frac{F_u(Z)}{\|F_u(Z)\|_2} \neq \frac{F_u(Z')}{\|F_u(Z')\|_2},$$

with nonzero outputs, even for genes disconnected in the graph.

The proof is in Appendix A.1. Theorem 1 shows that the higher-order mapping remains equivariant to gene ordering while admitting dependencies beyond training graph connectivity. This view complements local context without adding observed edges.

2.3 Neighborhood Contrastive Learning across context views

Expression guided sampling modifies local communication, while higher-order aggregation introduces context beyond recorded neighborhoods. Same gene alignment defines positives through gene identity alone, leaving training positive partners as non-positive comparisons in the contrastive objective. Neighborhood Contrastive Learning (NCL) instead promotes these partners to symmetric positives, explicitly incorporating training positive partner relations into the contrastive positive set. Training interactions fix partner membership across changing views.

Training neighbor positives. For gene i, the neighbor set is

$$\mathcal{N}_i = \{j \in V \setminus \{i\} : (i,j) \in \mathcal{E}_{\text{tr}}^+ \text{ or } (j,i) \in \mathcal{E}_{\text{tr}}^+\}, \tag{5}$$

where $\mathcal{N}_i$ contains recorded targets and regulators of i, giving symmetric partner membership from $\mathcal{E}_{\text{tr}}^+$ independently of $\widetilde{A}$. Prediction labels retain TF-to-target order.

Contrastive objective. Let (a, b) be either (Z_s, Z_h) or (Z_h, Z_s), where a_i is the anchor, and define $s(a_i, b_j) = \exp(\cos(a_i, b_j)/\tau)$ with temperature $\tau > 0$. The positive and total sums are

$$\begin{aligned} P_i(a,b) &= s(a_i,b_i) + \sum_{j\in\mathcal{N}_i} [s(a_i,a_j) + s(a_i,b_j)], \\ D_i(a,b) &= \sum_{j\in V} s(a_i,b_j) + \sum_{j\in V\setminus\{i\}} s(a_i,a_j), \end{aligned} \tag{6}$$

where the numerator includes the same gene cross-view match and within- and cross-view neighborhood positive matches, while the denominator excludes only the anchor's within-view self-match. With $q_i = 1 + 2|\mathcal{N}_i|$ positives, the per anchor loss and its symmetric average are

$$\begin{aligned} \ell_i(a,b) &= -\log \frac{P_i(a,b)/q_i}{D_i(a,b)}, \\ \mathcal{L}_{\text{nbr}} &= \frac{1}{2N}\sum_{i\in V} [\ell_i(Z_s,Z_h) + \ell_i(Z_h,Z_s)]. \end{aligned} \tag{7}$$

Both views serve as anchors. With same gene alignment alone, a training partner appears only as a non-positive comparison; NCL instead promotes it to a positive within and across views. Proposition 1 formalizes the invariance of this membership to the student graph.

Proposition 1 (Invariance of neighborhood contrastive positives). *For every sampled student edge mask M, the NCL positive set is unchanged. A neighborhood positive $j \in \mathcal{N}_i$ remains positive within and across views even when its propagation edge is removed. Let k index comparisons in D_i, $\mathcal{I}_i^+$ denote positive comparison indices, and t_{ik} be cosine similarity divided by τ. For finite logits and at least one non-positive comparison,*

$$\frac{\partial \ell_i}{\partial t_{ik}} = \begin{cases} e^{t_{ik}} \left(\frac{1}{D_i} - \frac{1}{P_i} \right) < 0, & k \in \mathcal{I}_i^+, \\ \frac{e^{t_{ik}}}{D_i} > 0, & k \notin \mathcal{I}_i^+. \end{cases}$$

The proof is in Appendix A.2. Proposition 1 formalizes the persistence of this partner-based contrastive supervision: training partners are promoted to positives independently of the sampled communication graph, and remain positives as student routes change.

2.4 Directed TF-target prediction and optimization

Representation alignment uses symmetric neighborhood relations, while GRN inference requires ordered TF-target scores. Two independent two layer projections encode regulator and target roles, $a_u = g_{\text{TF}}(z_{s,u})$ and $b_v = g_{\text{target}}(z_{s,v})$. A two layer MLP h_ψ scores their ordered features:

$$\widehat{y}_{uv} = \sigma(h_\psi([a_u \| b_v \| a_u \odot b_v \| |a_u - b_v|])) \,, \tag{8}$$

where σ is sigmoid, and ψ collects the MLP and role projection parameters. The projections use LeakyReLU and dropout. The independent role projections and ordered concatenation allow different prediction scores for the two orientations of a candidate TF-target pair.

Let $\mathcal{L}_{\text{link}}$ be the mean binary cross entropy over a labeled candidate minibatch. The joint objective is

$$\mathcal{L} = \mathcal{L}_{\text{link}} + \beta \mathcal{L}_{\text{nbr}}, \qquad \beta \geq 0. \tag{9}$$

After supervised warm-up, joint training updates the student, decoder, and hypergraph parameters with Adam. The teacher follows $\bar{\theta} \leftarrow m\bar{\theta} + (1-m)\theta$ after each optimizer step, where $0 \leq m < 1$. Inference uses $f_\theta(X, A)$ and the directed decoder. Appendix C provides implementation details, and Algorithm C1 summarizes the complete training procedure.

Computational complexity. Latent hyperedges provide higher-order context without explicitly forming an $N \times N$ gene relation matrix. Adding partner positives preserves the asymptotic cost of all gene contrast. The auxiliary modules operate only during training and add no inference computation. Appendix D provides the detailed time and space complexity analysis.

3 Experiments

3.1 Experimental setup and evaluation protocol

We evaluate GRASP on 42 settings spanning seven cell types, TF+500/TF+1000 gene scopes, and three interaction references: cell-specific ChIP-seq (Specific), non-specific ChIP-seq (Non-Specific), and STRING functional interactions (Pratapa et al., 2020; Chen & Liu, 2022; Yu et al., 2025).

Eight baselines share expression inputs, candidates, and splits. Six are supervised: GENELink (Chen & Liu, 2022) uses graph attention; GCLink (Yu et al., 2025) adds graph contrastive learning; CNNC (Yuan & Bar-Joseph, 2019) convolves pairwise expression histograms; GNE (KC et al., 2019) jointly embeds expression and interactions; GRANet (Zhou et al., 2025) uses residual graph attention; and RegGAIN (Guan et al., 2026) uses dual role graph contrastive learning. GENIE3 (Huynh-Thu et al., 2010) and GRNBoost2 (Moerman et al., 2019) are two expression based baselines that rank regulators by tree based regression without reference link labels.

The task is transductive: genes and expression profiles are shared across partitions, but evaluated TF-target pairs are unseen during training. Our pair-disjoint split assigns both orientations of each gene pair to one partition while retaining directed labels. Pseudo-negatives exclude known positive pairs, and propagation uses only training positives. Section 3.4 examines this protocol under symmetrized message passing using matched split comparisons.

Table 1: Test AUPRC across 42 benchmark settings (seven cell types × three references × two gene scopes). Entries show TF+500 / TF+1000 means, omitting leading zeros; bold and underline mark the highest and second highest displayed AUPRC scores.

Reference	Cell	GENE Link	GCLink	CNNC	GNE	GRANet	GENIE3	GRN Boost2	Reg GAIN	**GRASP**
Specific	hESC	.51/.51	.54/.57	.48/.52	.33/.31	.59/.60	.16/.14	.15/.15	.21/.18	**.60/.62**
	hHEP	.70/.70	.78/.77	.67/.68	.53/.56	**.83**/.83	.38/.37	.37/.35	.40/.39	**.83/.84**
	mDC	.11/.12	.13/.13	.10/.11	.08/.08	.14/.16	.05/.05	.05/.05	.05/.06	**.16/.18**
	mESC	.77/.77	.78/.80	.78/.79	.59/.61	**.89/.89**	.32/.31	.30/.30	.37/.39	.88/.88
	mHSC-E	.89/.89	.91/.92	.89/.90	.78/.77	**.94/.95**	.56/.54	.58/.56	.63/.61	**.94/.95**
	mHSC-GM	.90/.90	.91/.92	.87/.88	.74/.76	.93/.94	.54/.54	.56/.56	.52/.53	**.94/.95**
	mHSC-L	.81/.81	.84/.83	.76/.75	.67/.65	.87/.85	.51/.49	.51/.50	.54/.51	**.88/.87**
Non-Specific	hESC	.03/.03	.03/.03	.02/.02	.02/.03	.03/.03	.02/.01	.02/.01	.02/.02	**.04/.04**
	hHEP	.03/.02	.04/**.04**	.02/.01	.03/.02	.03/.03	.02/.01	.01/.01	.01/.01	**.05**/.03
	mDC	.03/.04	.04/**.05**	.02/.02	.03/.02	.04/.04	.03/.02	.02/.02	.02/.02	**.05/.05**
	mESC	.02/.02	.02/.02	.02/.02	.02/.02	.02/**.03**	.02/.02	.02/.02	.02/.01	**.03/.03**
	mHSC-E	.05/.09	.09/.09	.04/.04	.04/.04	.07/.06	.06/.04	.04/.03	.02/.02	**.18/.15**
	mHSC-GM	.12/**.30**	.17/.20	.06/.08	.05/.08	.13/.22	.15/.11	.08/.06	.05/.03	**.23**/.25
	mHSC-L	**.15**/.10	.13/.10	.07/.06	.05/.05	.11/.09	.13/**.13**	.10/.10	.06/.05	.12/.09
STRING	hESC	.07/.01	.12/.02	.04/.02	.05/.04	**.14**/.02	.05/**.05**	.04/.04	.04/.04	.13/.02
	hHEP	.04/.01	.06/.02	.03/.02	.05/**.04**	.07/.02	.05/**.04**	.04/.03	.04/.03	**.08**/.02
	mDC	.06/.02	.13/.02	.04/.02	.07/.04	.14/.02	.05/.04	.05/.04	.06/**.05**	**.15**/.03
	mESC	.08/.05	.17/.09	.05/.04	.07/.05	**.26**/.17	.05/.05	.04/.03	.05/.03	.25/**.20**
	mHSC-E	.07/.19	.07/.19	.05/.08	.06/.08	.09/.16	.15/.12	.08/.07	.09/.06	**.19/.31**
	mHSC-GM	.27/.35	.28/.23	.15/.24	.09/.13	.20/.38	.38/.25	.20/.12	.05/.04	**.42/.44**
	mHSC-L	.07/.11	.05/.06	.09/.08	.07/.05	.06/.08	**.55/.45**	.35/.33	.14/.12	.13/.15
Average AUPRC		.27/.29	.30/.29	.25/.26	.21/.21	.31/.31	.20/.18	.17/.16	.16/.15	**.35/.34**

AUPRC (average precision) is the primary metric; AUROC is secondary. Validation AUPRC selects checkpoints. In the main benchmark, all methods use three seeds per split. We average seeds within each of five fixed splits, then splits within settings, weighting all 42 settings equally. Appendix E.1 details datasets, candidates, class balance, and evaluation.

3.2 Overall GRN inference performance

In Table 1, we compare GRASP with eight baselines across 42 evaluation settings. GRASP achieves the highest average AUPRC at both TF+500 and TF+1000. Compared with GRANet, the strongest baseline overall, it delivers a 9.1% relative improvement in overall AUPRC and a higher mean AUROC of 0.75 versus 0.69. The absolute AUPRC gains over GRANet are particularly pronounced on Non-Specific and STRING, reaching 0.037/0.019 and 0.057/0.046 at TF+500/TF+1000, respectively, while remaining positive on Specific at both gene scopes. These positive margins across all six reference scope groups show that the aggregate advantage extends across interaction sources rather than depending on a single reference network or gene set size. The results establish the effectiveness of the complete framework under different forms of partial reference supervision. We next examine how adaptive context construction and partner-based contrastive supervision contribute individually and jointly to these gains through matched factorial ablations in Section 3.3. Appendix F.1 provides complete AUPRC/AUROC results, reference level margins, and split level variation.

3.3 Context and partner-based contrastive supervision

We conduct a complete 2^3 factorial study on STRING/TF+1000 to disentangle three factors in GRASP: expression guided edge removal augmentation (Augment), teacher induced higher-order context (Hypergraph), and Neighborhood Contrastive Learning (NCL). Disabling Augment sets $\gamma = 0$ while keeping $p_e = 0.1$, yielding uniform random removal; disabling Hypergraph replaces Z_h with the clean teacher representation Z_t as the auxiliary view; and disabling NCL removes training partners from the positive set while retaining same gene cross-view alignment. All eight variants use identical data partitions, decoder, optimization protocol, and remaining hyperparameters. Table 2 reports mean test AUPRC across seeds, fixed splits, and seven cell types.

Context construction. The base variant without neighborhood positives reaches an AUPRC of 0.1296. Expression guided augmentation increases this to 0.1347, higher-order context to 0.1394, and combining both to 0.1424. Adapting observed propagation routes and introducing context beyond recorded neighborhoods provide complementary context gains. The benefit of higher-order context

is larger when training partners are included in the contrastive positive set: with expression guided augmentation, adding the higher-order view improves AUPRC by 0.0077 without neighborhood positives, compared with 0.0235 when they are included. This larger gain supports including training partners as positives when aligning higher-order representations.

Contrastive supervision. With context construction fixed, neighborhood positives improve all four matched configurations, showing that training positive TF-target partners provide useful contrastive supervision beyond same gene alignment. The gain is largest when both expression guided augmentation and higher-order context are active, raising AUPRC from 0.1424 to 0.1668. The complete configuration performs best, supporting the combination of adaptive context construction and partner-based contrastive supervision for inference.

Table 2: Factorial ablation on STRING/TF+1000. ✓: enabled; –: replacement.

Augment	Hypergraph	NCL	Mean test AUPRC	Drop (%)
–	–	–	0.129614	22.30
✓	–	–	0.134679	19.26
–	✓	–	0.139353	16.46
–	–	✓	0.138825	16.78
✓	✓	–	0.142410	14.63
✓	–	✓	0.143307	14.09
–	✓	✓	0.149430	10.42
✓	✓	✓	**0.166814**	**0.00**

3.4 BIDIRECTIONAL-AWARE PAIR-DISJOINT EVALUATION

Here, we examine how edge partitioning controls the connectivity available during training. BEELINE (Pratapa et al., 2020) provides reference networks for GRN benchmarking but does not prescribe a unified supervised train/validation/test edge split; supervised GRN link prediction methods therefore define their own edge partitions (Chen & Liu, 2022; Mao et al., 2023; Yu et al., 2025). Under symmetrized message passing, independently splitting the two directed edges introduces bidirectional *label leakage*. For a bidirectional pair, $u \to v$ may be assigned to test while $v \to u$ remains in training; symmetrization then leaves u and v directly connected in the training graph. The encoder thus sees the test pair $\{u, v\}$ despite its held out label, turning missing link prediction into a trivial direction completion task for that pair. OpenBioLink (Breit et al., 2020) likewise excludes test relations that can be inferred from reverse or inverse links observed during training.

To prevent this label leakage, we partition unordered gene pairs before graph construction. All observed orientations of a pair are assigned to the same partition, pseudo-negatives exclude known positive pairs, and only training positives are used to construct the propagation graph. Hence, neither orientation of a validation or test pair can enter training time connectivity.

Protocol comparison. We compare the legacy directed edge split with the pair-disjoint protocol under the same model and training configuration. Table 3 reports seven-cell means at each scope and the equally weighted overall mean; Δ is legacy minus pair-disjoint, computed before rounding. Specific remains close to its legacy performance at both gene scopes, whereas the performance gap is substantially larger on Non-Specific and especially STRING (Table 3). Together with the connectivity audit, these higher legacy scores are consistent with overly optimistic evaluation inflated by bidirectional label leakage. Eliminating this leakage requires GRASP to rank a held out TF-target interaction without either orientation of the corresponding gene pair entering the training graph, enabling genuine missing link prediction on unseen gene pairs (Appendix G).

Table 3: Effect of the evaluation protocol.

Reference	Scope	AUPRC			AUROC		
		Legacy	Pair	Δ	Legacy	Pair	Δ
Specific	TF+500	0.74	0.75	-0.008	0.88	0.88	-0.006
	TF+1000	0.75	0.75	-0.003	0.89	0.90	-0.002
Non-Specific	TF+500	0.17	0.10	+0.075	0.77	0.68	+0.087
	TF+1000	0.14	0.09	+0.052	0.76	0.71	+0.057
STRING	TF+500	0.42	0.19	+0.230	0.92	0.68	+0.235
	TF+1000	0.44	0.17	+0.271	0.92	0.64	+0.279
Overall		0.44	0.34	+0.103	0.86	0.75	+0.108

3.5 DIRECTIONALITY OF TF-TARGET SCORING

Pair-disjoint evaluation holds out test pair connectivity, but does not by itself establish whether GRASP distinguishes regulator and target roles. We therefore compare the score of each annotated TF-target link with its eligible reverse orientation on hESC Specific-500/1000. Both genes must be eligible as TFs and targets, so either ordering is a valid candidate. This tests whether scoring preserves the annotated orientation despite symmetrized message passing.

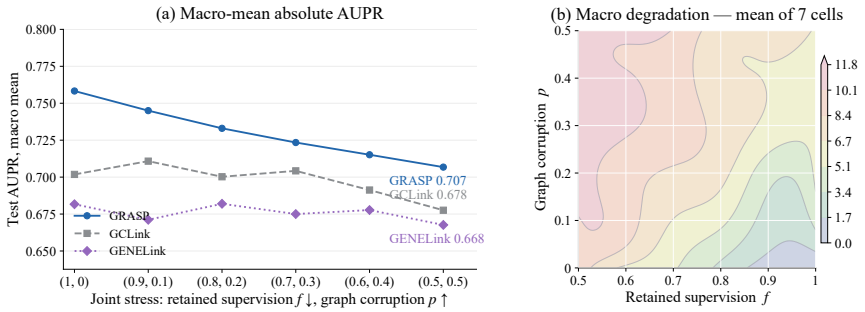


Figure 2: Dual axis robustness on Specific/TF+1000. (a) Macro mean test AUPRC along the diagonal stress. (b) Macro mean relative degradation (%) over grid (GRASP).

Figure 3a pools forward-minus-reverse scores across both scopes and training repeats. Dots show individual comparisons and the lower curve their density; values right of the dashed zero line favor the annotated direction. The pooled mean is $+0.18$ (SD 0.39) and median $+0.15$. Positive margins occur in 62% and 72% of TF+500 and TF+1000 comparisons, respectively (Appendix H). The positive median indicates a forward preference for the typical comparison. These results support endpoint role distinction across both evaluated gene scopes in hESC.

3.6 Robustness to imperfect reference knowledge

In practical scenarios, biological reference networks are inherently noisy and incomplete. We test resilience to incomplete and unreliable references along two axes: reducing retained training supervision and replacing graph edges with spurious interactions. At each supervision fraction, corruption changes the propagation and partner structure while supervised candidate labels remain fixed. Corruption thus affects both propagation routes and partner evidence. The Specific/TF+1000 evaluation covers seven cell types and a 6×6 grid with retained fractions $f = 0.5$–1.0 and replacement rates $p = 0$–0.5. Validation and test candidates remain fixed across conditions.

Dual axis stress. Along the diagonal stress path, GRASP remains ahead of GCLink at every evaluated level (Figure 2a). Even at the most stressed corner, its macro average degradation is approximately 12% (Figure 2b). Across the 252 paired per cell conditions in the full grid, GRASP outperforms GCLink in 242 cases, confirming this. The non-contrastive GENELink control has lower macro mean AUPRC than GCLink at all six diagonal stress levels. GRASP outperforms GENELink in 241 of 252 paired per cell conditions (mean $+0.051$). Along this diagonal, macro mean AUPRC decreases from 0.7583 to 0.7068, while the margin over GCLink remains 0.0292 at the joint corner. This retained advantage indicates that GRASP can still rank missing links when fewer training partners and less reliable propagation routes are available. Absolute AUPRC measures ranking quality on matched candidates; normalized retention measures the fraction of each cell's clean reference performance retained. Together, they characterize performance under reduced reference coverage and reliability across the tested stress conditions.

GRASP's advantage across the full grid supports link recovery as reference coverage and reliability decline (Appendix I.1). Sweeps of eight hyperparameters change validation AUPRC by at most 0.47% in aggregate over the displayed ranges (Appendix I.2). These small changes indicate stable aggregate ranking performance. They suggest that GRASP tolerates variation in augmentation, contrastive learning, and higher-order context within the tested ranges.

3.7 Cross-cell generalization

We test whether mESC pretrained representations can adapt to six target cell types with limited local supervision. Figure 3b compares GRASP and GCLink across target label budgets from 0% to 30%. At each budget, macro-AUPRC averages training runs within each setting, then weights the six target cells and six reference scope groups equally. The curves are nearly coincident at zero and 1% supervision, while GRASP has higher macro-AUPRC at 5%, 10%, and 30%. At the largest budget,

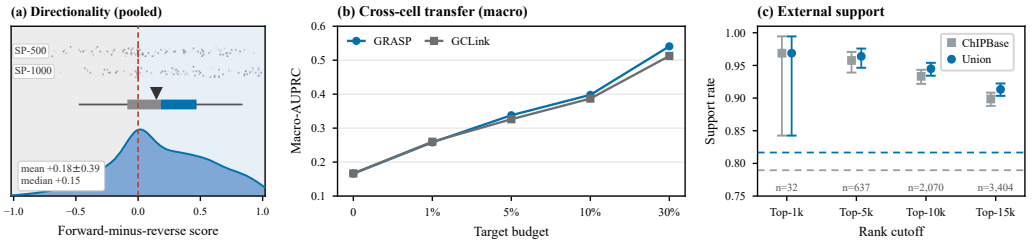


Figure 3: Directionality, cross-cell transfer, and external support. (a) Forward-minus-reverse scores. (b) Test AUPRC by target label budget. (c) External annotation support.

GRASP reaches 0.541 versus 0.513 for GCLink, a gap of 0.028. Both methods improve with larger budgets, while GRASP gains more over the evaluated supervision range.

Appendix J.1 details reference- and target-specific results; Appendix J.2 separately evaluates genes withheld during training. GRASP leads in 11 of 12 Specific cell scope settings at 10% supervision and all 12 STRING settings at 30%, spanning human and mouse targets. Together with the similar zero budget scores, the growing separation indicates that GRASP effectively adapts transferred representations to target cell regulatory structure as local labels become available. These results support GRASP's advantage in prioritizing missing regulatory links across cellular contexts, using partial local reference knowledge to guide target-specific GRN completion.

3.8 External annotation support

To assess external annotation support for GRASP's ranked candidates, we compare top ranked hESC Specific-1000 predictions against TF-target annotations from ChIPBase and Harmonizome. We exclude all training and validation candidate pairs, including positives and pseudo-negatives, leaving a matched background of 9,959 eligible pairs. Figure 3c compares ChIPBase and union support at four nominal Top-K cutoffs. Excluded pairs are not replenished: the cutoff specifies selection depth, while n is the retained set size used as the support denominator. The background uses the same exclusions. Dashed lines mark background support rates, and error bars show 95% Wilson confidence intervals for each retained set of eligible predictions.

Both support curves remain above their matched backgrounds and rise toward stricter rank cutoffs. At Top-5k, 96.39% of 637 eligible predictions have union support; at Top-15k, this is 91.33% of 3,404, compared with 81.66% in the matched background. The close ChIPBase and union curves show that ChIPBase accounts for most of the concordance; Appendix K provides counts and top ranked pairs. The enrichment at stricter cutoffs shows that GRASP concentrates documented TF-target relationships near the top of its predictions. This supports its practical value for prioritizing regulatory candidates from expression profiles and partial interaction references.

4 Conclusion

In this paper, we propose GRASP for supervised GRN inference under partial interaction references. GRASP adapts local graph propagation to the measured expression context through expression guided edge retention. It then constructs a teacher induced higher-order view to complement the context available from recorded graph neighborhoods. Based on these complementary views, Neighborhood Contrastive Learning incorporates training positive TF-target partners into the contrastive positive set, while a role-specific decoder produces ordered regulatory scores. Experiments across 42 settings spanning diverse cellular contexts show that GRASP achieves the highest average AUPRC among eight evaluated baselines, with ablation studies supporting the contributions of adaptive context construction and partner-based contrastive supervision. We further use pair-disjoint evaluation to isolate held out gene-pair connectivity under symmetrized message passing. Additional analyses examine robustness to degraded references, cross-cell transfer, and external annotation support. Future work can extend GRASP with perturbational and chromatin accessibility measurements and investigate broader inductive GRN inference settings.

AI USE STATEMENT

We used AI tools for manuscript preparation and revision, including English editing and Chinese-to-English translation. The authors take responsibility for the final content.

REFERENCES

Anna Breit, Simon Ott, Asan Agibetov, and Matthias Samwald. OpenBioLink: a benchmarking framework for large-scale biomedical link prediction. *Bioinformatics*, 36(13):4097–4098, 2020. doi: 10.1093/bioinformatics/btaa274.

Guangyi Chen and Zhi-Ping Liu. Graph attention network for link prediction of gene regulations from single-cell RNA-sequencing data. *Bioinformatics*, 38(19):4522–4529, 2022. doi: 10.1093/bioinformatics/btac559.

Yu Chen, Lingfei Wu, and Mohammed J. Zaki. Iterative Deep Graph Learning for Graph Neural Networks: Better and Robust Node Embeddings. In *Advances in Neural Information Processing Systems*, 2020.

Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are AllSet: A Multiset Function Framework for Hypergraph Neural Networks. In *International Conference on Learning Representations*, 2022.

Ido Diamant, Daniel J. B. Clarke, John Erol Evangelista, Nathania Lingam, and Avi Ma'ayan. Harmonizome 3.0: integrated knowledge about genes and proteins from diverse multi-omics resources. *Nucleic Acids Research*, 53(D1):D1016–D1028, 2025. doi: 10.1093/nar/gkae1080.

Johannes Gasteiger, Stefan Weißenberger, and Stephan Günnemann. Diffusion Improves Graph Learning. In *Advances in Neural Information Processing Systems*, 2019.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 2017.

Qiyuan Guan, Jiating Yu, Jieyi Pan, Fan Yuan, Jiadong Ji, Rusong Zhao, Zhi-Ping Liu, Bingqiang Liu, Ling-Yun Wu, and Duanchen Sun. Inferring Gene Regulatory Networks From Single-Cell RNA Sequencing Data by Dual-Role Graph Contrastive Learning. *Advanced Science*, 13(9):e18277, 2026. doi: 10.1002/advs.202518277.

Junhong Huang, Wujian Zheng, Ping Zhang, Qiao Lin, Zhirong Chen, Jiajia Xuan, Chang Liu, Di Wu, Qiaojuan Huang, Lingling Zheng, Shurong Liu, Keren Zhou, Lianghu Qu, Bin Li, and Jianhua Yang. ChIPBase v3.0: the encyclopedia of transcriptional regulations of non-coding RNAs and protein-coding genes. *Nucleic Acids Research*, 51(D1):D46–D56, 2023. doi: 10.1093/nar/gkac1067.

Siqi Huang, Yanchen Xu, Hongyuan Zhang, and Xuelong Li. Learn Beneficial Noise as Graph Augmentation. In *International Conference on Machine Learning*, 2025.

Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE*, 5(9):e12776, 2010. doi: 10.1371/journal.pone.0012776.

Kenji Kamimoto, Blerta Stringa, Christy M Hoffmann, Kunal Jindal, Lilianna Solnica-Krezel, and Samantha A Morris. Dissecting cell identity via network inference and in silico gene perturbation. *Nature*, 614(7949):742–751, 2023. doi: 10.1038/s41586-022-05688-9.

Kishan KC, Rui Li, Feng Cui, Qi Yu, and Anne R. Haake. GNE: a deep learning framework for gene network inference by aggregating biological information. *BMC Systems Biology*, 13(Suppl 2):38, 2019. doi: 10.1186/s12918-019-0694-y.

Emre Kulkul, Tolga Çukur, and Aykut Koç. scHyperLink: Revealing Cell-Type-Specific Gene Regulation With Hypergraph Neural Networks. *IEEE Journal of Biomedical and Health Informatics*, 30(8):6474–6487, 2026. doi: 10.1109/JBHI.2025.3650661.

Guo Mao, Zhengbin Pang, Ke Zuo, Qinglin Wang, Xiangdong Pei, Xinhai Chen, and Jie Liu. Predicting gene regulatory links from single-cell RNA-seq data using graph neural networks. *Briefings in Bioinformatics*, 24(6):bbad414, 2023. doi: 10.1093/bib/bbad414.

Thomas Moerman, Sara Aibar Santos, Carmen Bravo González-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35(12):2159–2161, 2019. doi: 10.1093/bioinformatics/bty916.

Aditya Pratapa, Amogh P Jalihal, Jeffrey N Law, Aditya Bharadwaj, and T M Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17(2):147–154, 2020. doi: 10.1038/s41592-019-0690-6.

Xiao Shen, Dewang Sun, Shirui Pan, Xi Zhou, and Laurence T. Yang. Neighbor Contrastive Learning on Learnable Graph Augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023. doi: 10.1609/aaai.v37i8.26168.

Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Mehdi Azabou, Eva L. Dyer, Rémi Munos, Petar Veličković, and Michal Valko. Large-Scale Representation Learning on Graphs via Bootstrapping. In *International Conference on Learning Representations*, 2022.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. In *International Conference on Learning Representations*, 2018.

Pi-Jing Wei, Huai-Wan Jin, Zhen Gao, Yansen Su, and Chun-Hou Zheng. GAEDGRN: reconstruction of gene regulatory networks based on gravity-inspired graph autoencoders. *Briefings in Bioinformatics*, 26(3):bbaf232, 2025. doi: 10.1093/bib/bbaf232.

Songyang Wu, Kui Jin, Mingjing Tang, Yuelong Xia, and Wei Gao. Inference of Gene Regulatory Networks Based on Multi-view Hierarchical Hypergraphs. *Interdisciplinary Sciences: Computational Life Sciences*, 16(2):318–332, 2024. doi: 10.1007/s12539-024-00604-3.

Songyang Wu, Mingjing Tang, Tong Zi, Lin Liu, Shaojie Qiao, Guiguang Ding, and Wei Gao. Motif-Based Hypergraph Representation Learning: Transductive and Inductive Inference for Gene Regulatory Networks. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2026. doi: 10.1109/TNNLS.2026.3685617.

Lianghao Xia, Chao Huang, Yong Xu, Jiashu Zhao, Dawei Yin, and Jimmy Xiangji Huang. Hypergraph Contrastive Collaborative Filtering. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022. doi: 10.1145/3477495.3532058.

Teng Xiao, Huaisheng Zhu, Zhengyu Chen, and Suhang Wang. Simple and Asymmetric Graph Contrastive Learning without Augmentations. In *Advances in Neural Information Processing Systems*, 2023.

Teng Xiao, Huaisheng Zhu, Zhiwei Zhang, Zhimeng Guo, Charu C. Aggarwal, Suhang Wang, and Vasant G Honavar. Efficient Contrastive Learning for Fast and Accurate Inference on Graphs. In *International Conference on Machine Learning*, 2024.

Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. HyperGCN: A New Method For Training Graph Convolutional Networks on Hypergraphs. In *Advances in Neural Information Processing Systems*, 2019.

Dengju Yao, BinBin Zhang, Xiaojuan Zhan, Wentao Wang, and Ning Liang. CaHoT-GRN: context-aware high-order topology learning for robust single-cell gene regulatory network inference. *Briefings in Bioinformatics*, 27(2):bbag202, 2026. doi: 10.1093/bib/bbag202.

Siyu Yi, Zhengyang Mao, Yifan Wang, Yiyang Gu, Zhiping Xiao, Chong Chen, Xian-Sheng Hua, Ming Zhang, and Wei Ju. Hypergraph Consistency Learning with Relational Distillation. *IEEE Transactions on Multimedia*, 27:7028–7039, 2025. doi: 10.1109/TMM.2025.3543068.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph Contrastive Learning with Augmentations. In *Advances in Neural Information Processing Systems*, 2020.

Weiming Yu, Zerun Lin, Miaofang Lan, and Le Ou-Yang. GCLink: a graph contrastive link prediction framework for gene regulatory network inference. *Bioinformatics*, 41(3):btaf074, 2025. doi: 10.1093/bioinformatics/btaf074.

Ye Yuan and Ziv Bar-Joseph. Deep learning for inferring gene relationships from single-cell expression data. *Proceedings of the National Academy of Sciences*, 116(52):27151–27158, 2019. doi: 10.1073/pnas.1911536116.

Junliang Zhou, Ningji Gong, Yanjun Hu, Hong Yu, Guoyin Wang, and Hao Wu. GRANet: a graph residual attention network for gene regulatory network inference. *Briefings in Bioinformatics*, 26 (4):bbaf349, 2025. doi: 10.1093/bib/bbaf349.

Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep Graph Contrastive Representation Learning. *arXiv preprint arXiv:2006.04131*, 2020.

Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph Contrastive Learning with Adaptive Augmentation. In *Proceedings of the Web Conference 2021*, 2021. doi: 10.1145/3442381.3449802.

A Shared Context and Neighborhood Contrastive Learning

A.1 Proof of Theorem 1: equivariance and nonlocal dependence

We analyze the hypergraph view as a function of teacher representations $Z = Z_t \in \mathbb{R}^{N\times d}$, with row z_i for gene i. Let $G : \mathbb{R}^{K\times d} \to \mathbb{R}^{K\times d}$ be the composition of the fixed residual maps in Equation (3), and let ρ act elementwise. Pooling and redistribution give

$$F(Z) = \rho\big(ZW\, G(W^\top Q(Z))\big), \qquad Q(Z) = Z^\top Z = \sum_{i\in V} z_i^\top z_i. \tag{10}$$

Thus, each gene's association row $z_u W$ accesses the shared second moment matrix $Q(Z)$ assembled from all teacher representations across the gene set.

Proof. (i) Since $(PZ)^\top(PZ) = Z^\top Z$ and $(PZ)W = P(ZW)$, the pooled input to G is unchanged by any permutation. Elementwise activation commutes with row permutation, so applying P to the input permutes the output rows identically: $F(PZ) = PF(Z)$, proving (i).

(ii) If only row v changes from z_v to $z_v + \delta$, expanding its outer product gives

$$\begin{aligned} Q(Z') &= Q(Z) + z_v^\top\delta + \delta^\top z_v + \delta^\top\delta, \\ F_u(Z') &= \rho\big(z_u W\, G(W^\top Q(Z'))\big), \end{aligned} \tag{11}$$

where z_u is fixed ($u \neq v$), so changes reach u only through shared context.

For an explicit nonlocal map, let $e_1, e_2 \in \mathbb{R}^d$ and $h_1, h_2 \in \mathbb{R}^K$ be coordinate vectors. Choose $W = e_1h_1^\top + e_2h_2^\top$, $z_u = e_1^\top + e_2^\top$, $z_v = ce_1^\top$ with $c > 0$, and set the remaining rows to zero. Set $R_\ell = 0$, $b_\ell = 0$ in every residual layer, so G is the identity for any number of layers. LeakyReLU preserves the nonnegative output, yielding

$$F_u(Z) = (c^2+2)e_1^\top + 2e_2^\top, \qquad \frac{F_{u,1}(Z)}{F_{u,2}(Z)} = \frac{c^2+2}{2}. \tag{12}$$

Write this construction as $Z(c)$. Taking $Z = Z(c)$ and $Z' = Z(c')$ for distinct positive c, c' changes only row v. The displayed ratio differs, so the two nonzero outputs differ after normalization. The construction applies even when the two genes lie in disconnected components of the observed interaction graph that provides propagation edges. □

Shared aggregation can therefore change the direction of a gene's hypergraph target using information beyond its recorded neighborhood. This provides a nonlocal context channel for cosine based alignment when the reference GRN is incomplete.

A.2 Proof of Proposition 1

Fix an anchor a_i in either view order (a, b) from Equation (7). Index comparisons by $k = (r, j)$, where $v_{(0,j)} = a_j$ and $v_{(1,j)} = b_j$. The comparison and positive sets are

$$\begin{aligned} \mathcal{C}_i &= (\{0\}\times(V\setminus\{i\}))\cup(\{1\}\times V), \\ \mathcal{I}_i^+ &= \{(1,i)\}\cup(\{0,1\}\times\mathcal{N}_i). \end{aligned} \tag{13}$$

For $k \in \mathcal{C}_i$, write $t_{ik} = \langle \widehat{a}_i, \widehat{v}_k\rangle/\tau$, where $\widehat{x} = x/\max(\|x\|_2, \epsilon)$ with $\epsilon > 0$ is the stabilized normalization used to compute cosine scores. The sums in Equation (6) are $P_i = \sum_{k\in\mathcal{I}_i^+} e^{t_{ik}}$ and $D_i = \sum_{k\in\mathcal{C}_i} e^{t_{ik}}$, with $q_i = |\mathcal{I}_i^+| = 1 + 2|\mathcal{N}_i|$ positive comparisons.

Proof. Expanding Equation (7) and differentiating its log sums gives

$$\ell_i = \log D_i - \log P_i + \log q_i, \qquad \frac{\partial\ell_i}{\partial t_{ik}} = \frac{e^{t_{ik}}}{D_i} - \mathbf{1}_{\{k\in\mathcal{I}_i^+\}}\frac{e^{t_{ik}}}{P_i}, \tag{14}$$

where q_i depends only on the fixed training neighbor set. If a non-positive comparison exists, $D_i > P_i > 0$. For $k \in \mathcal{I}_i^+$, the derivative is $e^{t_{ik}}(P_i - D_i)/(D_iP_i) < 0$; otherwise it is $e^{t_{ik}}/D_i > 0$. If $\mathcal{I}_i^+ = \mathcal{C}_i$, every comparison is positive and all derivatives are zero.

With same gene positives alone, $\mathcal{I}_i^+ = \{(1, i)\}$, so both $(0, j)$ and $(1, j)$ for a neighborhood positive $j \neq i$ appear only in the denominator. Finally, $\mathcal{N}_i$ is defined by $\mathcal{E}_{\text{tr}}^+$ in Equation (5), independently of M. Augmentation changes the representations and their logits while preserving positive membership. The argument applies to each anchor in both view orders. □

B Related Work

GRN inference. GRN inference from single cell transcriptomics has been studied through both expression driven and graph-based formulations. Expression based methods such as GENIE3 (Huynh-Thu et al., 2010) and GRNBoost2 (Moerman et al., 2019) rank candidate regulators from predictive dependencies in gene expression without explicitly propagating information over a supplied interaction graph. Supervised graph methods instead incorporate known interactions into representation learning. GENELink (Chen & Liu, 2022) applies graph attention to expression profiles for regulatory link prediction, while GNNLink (Mao et al., 2023) reconstructs regulatory relationships with graph neural networks. More recent methods enrich graph-based inference through graph perturbation, contrastive objectives, residual attention, role aware representations, and directed decoders (Yu et al., 2025; Zhou et al., 2025; Wei et al., 2025; Guan et al., 2026). These approaches use partial interaction references for supervision and structural context in GRN inference.

Context learning beyond fixed graph topology. Graph context can be adapted through augmentation, learned connectivity, and diffusion. Adaptive augmentation conditions local perturbations on node features or topology (Zhu et al., 2021; Huang et al., 2025). Structure learning jointly refines node embeddings and similarity based connectivity (Chen et al., 2020), allowing node attributes and learned representations to propose connections absent from the supplied adjacency. Graph diffusion uses personalized PageRank or heat kernels to aggregate information along multiple graph steps beyond immediate neighbors (Gasteiger et al., 2019); its expanded neighborhoods remain derived from paths in the input graph. Hypergraph models encode group relations or learn latent associations (Yadati et al., 2019; Chien et al., 2022; Xia et al., 2022; Yi et al., 2025). Higher-order GRN context has also been explored through hypergraphs and auxiliary biological information (Wu et al., 2024; Kulkul et al., 2026; Wu et al., 2026; Yao et al., 2026). Beyond structural context, a partial reference also provides observed relational supervision. GRASP explicitly decouples these roles: incomplete context can be adapted and enriched, while observed training interactions define fixed contrastive partner positives within and across local and higher-order views.

Graph contrastive learning with neighborhood positives. Graph contrastive learning varies in view construction and positive selection. GraphCL (You et al., 2020) aligns graph representations across augmentations, whereas GRACE (Zhu et al., 2020) matches nodes across structural and feature perturbations. BGRL (Thakoor et al., 2022) predicts momentum encoder targets without negatives. Relational objectives extend beyond node identity by including graph neighbors as positives within and across views learned jointly with the encoder, predicting neighborhood representations asymmetrically without augmentation, or aligning MLP representations of central nodes with GNN representations of their neighbors (Shen et al., 2023; Xiao et al., 2023; 2024). These choices distinguish invariance to view perturbations from the relational information used to define positive comparisons. In GRN inference, GCLink (Yu et al., 2025) combines contrastive learning with link prediction, while RegGAIN (Guan et al., 2026) uses role-specific representations. GRASP uses observed training interactions to define fixed partner positives within and across views, while adapting and enriching the context used to encode genes for regulatory link prediction.

C Implementation Details

Algorithm C1 summarizes training; ϕ collects W, R_ℓ, and b_ℓ.

Augmentation. Expression associations are ranked over directed training positives; bidirectional records retain separate rank entries. Both stored directions inherit the same rank and are sampled independently once per epoch; enabled self-loops receive rank one. For stored support $\Omega = \{(i, j) : A_{ij} > 0\}$, the expected removal fraction is $\bar{p}_e = p_e|\Omega|^{-1} \sum_{(i,j)\in\Omega}(1 - r_{ij})^\gamma$. At $\gamma = 0$, $\bar{p}_e = p_e$; expression weighting governs routes and mean perturbation.

Algorithm C1 GRASP training and prediction with a momentum teacher

Require: X, training candidates, A, $\{\mathcal{N}_i\}$; warm-up and joint epochs E_w, E_j
Initialize θ, ψ, ϕ and compute ranks using Eq. (1)
Warm up θ, ψ with minibatch $\mathcal{L}_{\text{link}}$ for E_w epochs, resampling $\widetilde{A}$ each epoch
Copy $\bar{\theta} \leftarrow \theta$ and set the teacher to evaluation mode
for $e = 1, \ldots, E_j$ **do**
    Sample $\widetilde{A}$ using Eq. (1)
    for each shuffled candidate minibatch **do**
        Encode all genes to obtain Z_s and detached Z_t
        Construct Z_h using Eqs. (2)–(4)
        Update θ, ψ, ϕ using Eq. (9); update the teacher by EMA
    end for
end for
Prediction: encode with $f_\theta(X, A)$ and score ordered pairs using Eq. (8)

Training and inference. After BCE warm-up, we copy the student into the EMA teacher, which encodes the unaugmented graph with dropout disabled. Each candidate minibatch recomputes full graph embeddings and all gene contrast; gradients update the hypergraph through detached teacher inputs. Inference uses the student on the clean training graph and the decoder.

Attention heads. For input adjacency A_{in}, each attention head transforms features by $h = XW_{\text{att}}$ and scores supported pairs by

$$s_{ij} = \text{LeakyReLU}(a_L^\top h_i + a_R^\top h_j; \alpha), \tag{15}$$

where a_L, a_R are attention vectors and α is the LeakyReLU slope. The condition $A_{\text{in},ij} > 0$ determines the attention mask; expression ranks control edge sampling through Equation (1). Masked row-wise softmax and attention dropout give weights α_{ij}. Each head outputs $o_i = \text{Normalize}_{L2}(\text{LeakyReLU}(\sum_j \alpha_{ij} h_j; \alpha)) + b$. The first layer concatenates heads followed by ELU; the second layer averages the attention head outputs.

D Computational Complexity

Scope and notation. We analyze the dense implementation used in our experiments. Let N be the number of genes, C the expression dimension, E the number of training positives, K the number of latent hyperedges, and L_h the number of hyperedge mappings. A minibatch contains B labeled candidate pairs, with Q pairs per training epoch. We treat the two GAT layers and their head counts as fixed, and take all hidden and decoder widths to be $O(d)$. Costs below count arithmetic operations rather than elapsed time on a particular hardware platform.

Expression guided retention. Computing absolute Pearson correlations on training positives costs $O(EC)$, and percentile ranking costs $O(E \log E)$. These ranks are cached before optimization. Each epoch samples retention decisions over $O(E + N)$ stored entries, including optional self-loops; sparse tensor assembly may incur additional indexing costs. Expression correlations are therefore not recomputed for successive candidate minibatches.

Local and higher-order views. The encoder forms dense attention scores, masks them by the graph support, and multiplies the resulting attention matrix by node features. One encoder pass costs $O(NCd + Nd^2 + N^2 d)$. The student and teacher each perform a pass during joint training; the teacher has no backward pass. For the higher-order view, forming $S = Z_t W$, pooling $S^\top Z_t$, and redistributing $SU^{(L_h)}$ each cost $O(NKd)$. The residual maps add $O(L_h K^2 d)$, giving $O(NKd + L_h K^2 d)$ in total. For fixed K, d, and L_h, this module scales linearly in N and never materializes an $N \times N$ gene relation matrix. Dense attention still incurs quadratic cost in N.

Contrastive and link objectives. NCL computes within- and cross-view similarity matrices over all genes, costing $O(N^2 d)$ time and $O(N^2)$ matrix storage. Symmetric evaluation changes only constant factors. Relative to same gene alignment with the same all gene denominator, partner positives add masking and reductions over existing scores, preserving the asymptotic order. The decoder costs $O(Nd^2)$ for role projection and $O(Bd^2)$ for scoring B candidate pairs.

Training time. The time complexity of a joint optimization step is

$$T_{\text{step}} = O\big(NCd + Nd^2 + N^2d + NKd + L_hK^2d + Bd^2\big)\,, \tag{16}$$

where ordinary backpropagation, Adam updates, and teacher EMA updates do not change the asymptotic order. Each candidate minibatch recomputes full graph embeddings and all gene contrast, so an epoch costs $O(\lceil Q/B \rceil T_{\text{step}})$, in addition to the once-per-epoch graph resampling and validation. Warm-up uses only the student, decoder, and link loss.

Training space. Excluding stored expression features, candidate lists, and parameter states, the working graph masks and intermediate activations require

$$\mathcal{M}_{\text{work}} = O\big(N^2 + NK + L_hKd + Nd + Bd\big)\,, \tag{17}$$

where dense attention and contrast account for the N^2 term. Expression inputs and training candidates add $O(NC + Q)$ storage. Model parameters occupy $O(Cd + d^2 + dK + L_hK^2)$ space; gradients and Adam states add constant multiples of that storage. Thus, the implementation retains quadratic memory dependence on gene count even with a fixed hyperedge count.

Inference. For Q_{test} candidate pairs, one student encoding and two role projections are followed by directed scoring, giving the total inference time complexity

$$T_{\text{infer}} = O\big(NCd + Nd^2 + N^2d + Q_{\text{test}}d^2\big)\,, \tag{18}$$

where teacher encoding, higher-order aggregation, and NCL are omitted. These auxiliary modules add no inference operations; the student incurs $O(N^2d)$ for dense attention.

Scaling implications. With fixed d and L_h, increasing K raises gene hyperedge pooling and redistribution linearly, but raises hyperedge mixing quadratically. Keeping K fixed as N grows therefore limits the extra context cost, although dense attention and NCL retain their N^2 terms. For fixed Q, increasing B reduces repeated full graph passes while total candidate scoring remains $O(Qd^2)$. This saves repeated computation at the cost of larger decoder activations; it does not reduce the all gene similarity matrices retained for contrastive learning.

E Experimental Settings

E.1 Data, candidates, and evaluation

The seven cell types are hESC, hHEP, mDC, mESC, mHSC-E, mHSC-GM, and mHSC-L. Each setting includes the selected TFs and either 500 or 1000 selected genes, following the benchmark (Pratapa et al., 2020; Chen & Liu, 2022; Yu et al., 2025). Table E1 summarizes the expression datasets and original reference candidate pools. Values report TF+500 (TF+1000); TFs and targets count distinct regulators and targets in the reference edges; Pos/Neg gives positive and unrecorded pairs in the original directed candidate pool, with P/N their ratio. The candidate, split, and aggregation rules below apply to all eight baselines introduced in Section 3.1.

Candidates and splits. Candidates are directed links from an included TF to another gene in the selected gene set. Reference interactions supply positive labels. Candidate pairs are grouped by their two gene identities and assigned together to training, validation, or test, following Section 3.1. Five fixed splits are used per setting. The propagation graph is built from training positives after partitioning and then symmetrized, keeping both orientations of held out pairs absent. Candidate labels and decoder scores retain TF-to-target order.

Allocation sizes follow the benchmark's per TF proportions: approximately one third of Non-Specific and STRING positives are reserved for test and one fifth of the remainder for validation; Specific uses $67\%/10\%/23\%$ for training/validation/test. Pairs remain intact.

Pseudo-negatives and class balance. Pseudo-negatives are drawn from unrecorded TF-target candidates after excluding self-links, duplicates, and pairs with a known interaction in either direction. Non-Specific and STRING use equal positive and negative counts for training and validation. Their test negative counts are chosen to match the benchmark positive fractions in Table E2 (Yu et al., 2025). Specific allocates unrecorded candidates by the same per TF rule.

Table E1: Benchmark data and reference candidate pools; paired entries: TF+500 (TF+1000).

Dataset	Cell type	Cells	TFs	Targets	Pos/Neg	P/N
Specific	hESC	758	34(34)	815(1260)	4545/26361 (7084/40822)	0.172(0.174)
	hHEP	425	30(31)	874(1331)	9939/18471 (15558/29299)	0.538(0.531)
	mDC	383	20(21)	443(684)	756/15644 (1193/26527)	0.048(0.045)
	mESC	421	88(89)	977(1385)	29613/68859 (42795/101296)	0.430(0.422)
	mHSC-E	1071	29(33)	691(1177)	11557/8830 (21975/17724)	1.309(1.240)
	mHSC-GM	889	22(23)	618(1089)	7364/6518 (14135/11878)	1.130(1.190)
	mHSC-L	847	16(16)	525(640)	4398/4546 (5180/5876)	0.967(0.882)
Non-Specific	hESC	758	283(292)	753(1138)	3441/253806 (4617/406811)	0.014(0.011)
	hHEP	425	322(332)	825(1217)	4129/300805 (5351/475053)	0.014(0.011)
	mDC	383	250(254)	634(969)	3067/201933 (3918/331362)	0.015(0.012)
	mESC	421	516(522)	890(1214)	6893/570511 (8030/837088)	0.012(0.010)
	mHSC-E	1071	144(147)	442(674)	1425/99807 (1960/174881)	0.014(0.011)
	mHSC-GM	889	82(88)	297(526)	743/50999 (1358/98170)	0.015(0.014)
	mHSC-L	847	35(37)	164(192)	279/19286 (317/25250)	0.014(0.013)
STRING	hESC	758	343(351)	511(695)	4257/307530 (5149/489410)	0.014(0.011)
	hHEP	425	409(414)	646(874)	7523/379800 (9003/590055)	0.020(0.015)
	mDC	383	264(273)	479(664)	4815/211665 (5898/354462)	0.023(0.017)
	mESC	421	495(499)	638(785)	7762/546143 (8479/799402)	0.014(0.011)
	mHSC-E	1071	156(161)	291(413)	1371/108297 (1826/191857)	0.013(0.010)
	mHSC-GM	889	92(100)	201(344)	748/57304 (1311/111789)	0.013(0.012)
	mHSC-L	847	39(40)	70(81)	137/21664 (154/27486)	0.006(0.006)

Table E2: Test set positive fractions for Non-Specific and STRING.

Cell	Non-Specific 500	Non-Specific 1000	STRING 500	STRING 1000
hESC	0.016	0.014	0.024	0.021
hHEP	0.015	0.013	0.028	0.024
mDC	0.019	0.016	0.038	0.032
mESC	0.015	0.013	0.024	0.021
mHSC-E	0.022	0.020	0.029	0.027
mHSC-GM	0.030	0.029	0.040	0.037
mHSC-L	0.048	0.043	0.048	0.045

Metrics and aggregation. AUPRC is average precision over the directed candidate ranking, and AUROC measures discrimination between positive and negative score assignments. Following Section 3.1, all methods use three seeds for each of five data splits in the main benchmark. We average seeds within each split, compute means and SDs across the five split averages, and weight all 42 settings equally using unrounded scores to compute the overall benchmark mean.

F Additional Benchmark Results

F.1 Benchmark performance and variation

Across 42 settings, mean AUROC is 0.75 for GRASP and 0.69 for GRANet (Table F1), consistent with the AUPRC ordering. GRASP also leads GRANet in mean AUPRC in all six reference scope groups: +0.005/+0.007 on Specific, +0.037/+0.019 on Non-Specific, and +0.057/+0.046 on STRING at TF+500/TF+1000, respectively (Table 1). The positive margins at both gene scopes show that the average advantage extends across all three reference sources. The larger gains on Non-Specific and STRING highlight the value of GRASP for recovering regulatory links when the supplied reference is not tailored to the target cellular context.

We report AUROC and AUPRC as mean ± SD of five split level scores, each obtained by averaging three seeds for every method (Tables F1 and F2). These SDs quantify variation across data splits. Paired gene scope rows support within cell comparisons of ranking performance and discrimination between positives and negatives. Bold and underline mark each row's highest and second highest two decimal means; ties share marks. Unrounded results accompany the paper.

Table F1: AUROC mean ± SD across all 42 benchmark settings. Each cell type has paired TF+500 and TF+1000 rows; tinted rows denote TF+1000. Bold and underline mark the highest and second highest displayed means within each row; ties share marks.

Cell	Scope (TF+)	GENE Link	GCLink	CNNC	GNE	GRANet	GENIE3	GRN Boost2	Reg GAIN	GRASP
Specific										
hESC	500	.84 ±.00	.85 ±.01	.81 ±.00	.72 ±.02	**.87** ±.01	.50 ±.01	.49 ±.01	.65 ±.04	**.87** ±.01
	1000	.84 ±.01	.86 ±.00	.84 ±.01	.70 ±.02	**.88** ±.00	.49 ±.00	.49 ±.01	.57 ±.05	**.88** ±.00
hHEP	500	.84 ±.03	.87 ±.01	.81 ±.01	.68 ±.01	.89 ±.00	.54 ±.01	.52 ±.00	.62 ±.02	**.90** ±.00
	1000	.84 ±.02	.86 ±.01	.82 ±.01	.71 ±.01	.89 ±.00	.53 ±.01	.50 ±.00	.60 ±.02	**.90** ±.00
mDC	500	.71 ±.02	**.77** ±.01	.71 ±.02	.64 ±.03	.74 ±.02	.51 ±.01	.52 ±.01	.55 ±.02	.76 ±.01
	1000	.72 ±.02	.77 ±.02	.75 ±.02	.65 ±.01	**.78** ±.01	.52 ±.01	.51 ±.02	.61 ±.05	**.78** ±.02
mESC	500	.89 ±.01	.91 ±.00	.89 ±.00	.77 ±.01	**.94** ±.00	.52 ±.00	.51 ±.00	.61 ±.04	**.94** ±.00
	1000	.89 ±.01	.91 ±.00	.90 ±.00	.79 ±.00	**.95** ±.00	.51 ±.00	.51 ±.00	.65 ±.02	.94 ±.00
mHSC-E	500	.88 ±.02	.90 ±.01	.87 ±.00	.74 ±.01	.92 ±.00	.51 ±.01	.52 ±.01	.62 ±.03	**.93** ±.00
	1000	.89 ±.01	.91 ±.01	.90 ±.00	.75 ±.01	**.94** ±.00	.50 ±.00	.51 ±.00	.63 ±.02	**.94** ±.00
mHSC-GM	500	.89 ±.01	.91 ±.01	.88 ±.00	.73 ±.01	.92 ±.00	.53 ±.01	.53 ±.01	.54 ±.06	**.93** ±.00
	1000	.89 ±.02	.91 ±.01	.89 ±.01	.76 ±.01	.93 ±.00	.51 ±.01	.53 ±.01	.54 ±.03	**.94** ±.00
mHSC-L	500	.83 ±.03	.85 ±.01	.77 ±.02	.69 ±.03	.87 ±.01	.53 ±.01	.53 ±.01	.60 ±.03	**.88** ±.00
	1000	.83 ±.01	.85 ±.01	.78 ±.01	.69 ±.01	.87 ±.00	.54 ±.01	.54 ±.01	.60 ±.02	**.88** ±.00
Non-Specific										
hESC	500	.64 ±.01	.66 ±.01	.55 ±.01	.54 ±.02	.63 ±.02	.47 ±.01	.47 ±.00	.62 ±.05	**.69** ±.01
	1000	.66 ±.02	.65 ±.03	.54 ±.01	.64 ±.03	.64 ±.01	.47 ±.00	.48 ±.01	.61 ±.06	**.70** ±.00
hHEP	500	.65 ±.02	.65 ±.01	.55 ±.01	.63 ±.06	.62 ±.01	.45 ±.01	.46 ±.00	.48 ±.02	**.69** ±.01
	1000	.65 ±.01	.65 ±.01	.54 ±.01	.62 ±.04	.61 ±.02	.45 ±.01	.48 ±.01	.47 ±.02	**.68** ±.01
mDC	500	.63 ±.02	.66 ±.02	.51 ±.01	.58 ±.03	.62 ±.02	.52 ±.01	.51 ±.00	.60 ±.01	**.69** ±.01
	1000	.70 ±.01	.70 ±.01	.51 ±.01	.55 ±.02	.67 ±.01	.51 ±.01	.51 ±.01	.61 ±.02	**.72** ±.01
mESC	500	.58 ±.02	.59 ±.01	.56 ±.01	.58 ±.03	.60 ±.02	.52 ±.00	.51 ±.00	.55 ±.01	**.64** ±.01
	1000	.64 ±.01	.68 ±.01	.58 ±.01	.62 ±.03	.66 ±.02	.53 ±.01	.51 ±.00	.53 ±.04	**.72** ±.01
mHSC-E	500	.66 ±.01	.66 ±.02	.61 ±.02	.62 ±.02	.64 ±.02	.59 ±.01	.57 ±.00	.54 ±.03	**.71** ±.02
	1000	.70 ±.03	.71 ±.02	.63 ±.02	.63 ±.03	.68 ±.01	.58 ±.01	.56 ±.01	.58 ±.03	**.74** ±.01
mHSC-GM	500	.65 ±.02	.65 ±.02	.60 ±.04	.61 ±.05	.67 ±.05	.69 ±.02	.60 ±.01	.63 ±.05	**.70** ±.04
	1000	.71 ±.05	.66 ±.01	.65 ±.03	.68 ±.02	.71 ±.02	.65 ±.01	.59 ±.01	.53 ±.03	**.76** ±.02
mHSC-L	500	**.66** ±.09	.55 ±.03	.57 ±.03	.52 ±.03	.58 ±.04	.61 ±.01	.56 ±.03	.57 ±.05	.63 ±.03
	1000	.61 ±.08	.52 ±.03	.61 ±.04	.54 ±.02	.57 ±.05	.62 ±.02	.59 ±.02	.59 ±.04	**.63** ±.02
STRING										
hESC	500	.54 ±.02	.59 ±.01	.57 ±.01	**.66** ±.01	.58 ±.02	.63 ±.00	.58 ±.00	.65 ±.02	**.66** ±.01
	1000	.32 ±.02	.35 ±.03	.47 ±.02	.66 ±.01	.31 ±.03	.64 ±.00	.59 ±.00	**.68** ±.01	.42 ±.02
hHEP	500	.39 ±.02	.41 ±.02	.50 ±.01	**.65** ±.00	.40 ±.02	.60 ±.00	.55 ±.00	.54 ±.01	.48 ±.02
	1000	.25 ±.05	.29 ±.03	.48 ±.01	**.64** ±.00	.37 ±.01	.58 ±.00	.55 ±.00	.54 ±.00	.44 ±.02
mDC	500	.46 ±.00	.50 ±.02	.45 ±.01	**.62** ±.02	.50 ±.02	.52 ±.00	.54 ±.00	.60 ±.02	.54 ±.01
	1000	.34 ±.01	.30 ±.02	.41 ±.01	.54 ±.02	.31 ±.03	.54 ±.00	.54 ±.00	**.61** ±.00	.46 ±.03
mESC	500	.69 ±.01	.72 ±.01	.65 ±.00	.68 ±.01	.74 ±.01	.60 ±.00	.54 ±.00	.70 ±.01	**.78** ±.01
	1000	.58 ±.01	.60 ±.01	.64 ±.00	.65 ±.00	.61 ±.01	.59 ±.00	.54 ±.00	.64 ±.03	**.69** ±.00
mHSC-E	500	.64 ±.01	.57 ±.03	.64 ±.02	.65 ±.03	.60 ±.03	.66 ±.01	.63 ±.01	**.79** ±.02	.77 ±.02
	1000	.77 ±.04	.76 ±.03	.70 ±.02	.72 ±.01	.78 ±.03	.65 ±.01	.61 ±.01	.72 ±.07	**.84** ±.01
mHSC-GM	500	.74 ±.05	.71 ±.05	.71 ±.12	.68 ±.06	.73 ±.04	.77 ±.01	.69 ±.01	.63 ±.04	**.81** ±.02
	1000	.82 ±.03	.79 ±.02	.79 ±.01	.74 ±.02	.83 ±.02	.70 ±.01	.64 ±.01	.60 ±.06	**.87** ±.01
mHSC-L	500	.54 ±.08	.37 ±.04	.67 ±.04	.54 ±.03	.45 ±.03	.76 ±.01	.73 ±.01	**.80** ±.01	.73 ±.04
	1000	.56 ±.17	.44 ±.07	.63 ±.08	.52 ±.05	.51 ±.08	.72 ±.02	.68 ±.02	**.76** ±.05	.74 ±.02
Average AUROC		.68	.68	.67	.65	.69	.57	.55	.61	**.75**

Table F2: AUPRC mean ± SD across all 42 benchmark settings. Each cell type has paired TF+500 and TF+1000 rows; tinted rows denote TF+1000. Bold and underline mark the highest and second highest displayed means within each row; ties share marks.

Cell	Scope (TF+)	GENE Link	GCLink	CNNC	GNE	GRANet	GENIE3	GRN Boost2	Reg GAIN	GRASP
Specific										
hESC	500	.51 ±.02	.54 ±.02	.48 ±.01	.33 ±.01	.59 ±.02	.16 ±.01	.15 ±.00	.21 ±.04	**.60** ±.01
	1000	.51 ±.04	.57 ±.03	.52 ±.01	.31 ±.02	.60 ±.01	.14 ±.00	.15 ±.00	.18 ±.05	**.62** ±.01
hHEP	500	.70 ±.04	.78 ±.02	.67 ±.01	.53 ±.01	**.83** ±.01	.38 ±.01	.37 ±.00	.40 ±.02	**.83** ±.01
	1000	.70 ±.04	.77 ±.00	.68 ±.01	.56 ±.01	.83 ±.01	.37 ±.00	.35 ±.00	.39 ±.02	**.84** ±.00
mDC	500	.11 ±.00	.13 ±.01	.10 ±.01	.08 ±.01	.14 ±.01	.05 ±.00	.05 ±.00	.05 ±.01	**.16** ±.01
	1000	.12 ±.02	.13 ±.01	.11 ±.01	.08 ±.00	.16 ±.02	.05 ±.00	.05 ±.00	.06 ±.01	**.18** ±.02
mESC	500	.77 ±.02	.78 ±.01	.78 ±.00	.59 ±.01	**.89** ±.00	.32 ±.00	.30 ±.00	.37 ±.04	.88 ±.00
	1000	.77 ±.02	.80 ±.01	.79 ±.00	.61 ±.01	**.89** ±.00	.31 ±.00	.30 ±.00	.39 ±.03	.88 ±.01
mHSC-E	500	.89 ±.02	.91 ±.00	.89 ±.00	.78 ±.01	**.94** ±.00	.56 ±.01	.58 ±.01	.63 ±.03	**.94** ±.00
	1000	.89 ±.02	.92 ±.00	.90 ±.00	.77 ±.01	**.95** ±.00	.54 ±.01	.56 ±.00	.61 ±.02	**.95** ±.00
mHSC-GM	500	.90 ±.01	.91 ±.01	.87 ±.01	.74 ±.01	.93 ±.00	.54 ±.01	.56 ±.01	.52 ±.04	**.94** ±.00
	1000	.90 ±.02	.92 ±.01	.88 ±.01	.76 ±.01	.94 ±.00	.54 ±.00	.56 ±.01	.53 ±.02	**.95** ±.00
mHSC-L	500	.81 ±.04	.84 ±.02	.76 ±.02	.67 ±.03	.87 ±.01	.51 ±.01	.51 ±.01	.54 ±.02	**.88** ±.01
	1000	.81 ±.01	.83 ±.01	.75 ±.01	.65 ±.02	.85 ±.00	.49 ±.01	.50 ±.01	.51 ±.03	**.87** ±.00
Non-Specific										
hESC	500	.03 ±.00	.03 ±.00	.02 ±.00	.02 ±.00	.03 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	**.04** ±.00
	1000	.03 ±.01	.03 ±.00	.02 ±.00	.03 ±.00	.03 ±.00	.01 ±.00	.01 ±.00	.02 ±.00	**.04** ±.00
hHEP	500	.03 ±.00	.04 ±.01	.02 ±.00	.03 ±.01	.03 ±.00	.02 ±.00	.01 ±.00	.01 ±.00	**.05** ±.00
	1000	.02 ±.00	**.04** ±.00	.01 ±.00	.02 ±.01	.03 ±.00	.01 ±.00	.01 ±.00	.01 ±.00	.03 ±.00
mDC	500	.03 ±.01	.04 ±.00	.02 ±.00	.03 ±.00	.04 ±.00	.03 ±.00	.02 ±.00	.02 ±.00	**.05** ±.00
	1000	.04 ±.00	**.05** ±.00	.02 ±.00	.02 ±.00	.04 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	**.05** ±.00
mESC	500	.02 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	**.03** ±.00
	1000	.02 ±.00	.02 ±.00	.02 ±.00	.02 ±.00	**.03** ±.00	.02 ±.00	.02 ±.00	.01 ±.00	**.03** ±.00
mHSC-E	500	.05 ±.01	.09 ±.02	.04 ±.00	.04 ±.00	.07 ±.02	.06 ±.01	.04 ±.01	.02 ±.00	**.18** ±.02
	1000	.09 ±.04	.09 ±.01	.04 ±.01	.04 ±.01	.06 ±.04	.04 ±.00	.03 ±.00	.02 ±.00	**.15** ±.03
mHSC-GM	500	.12 ±.07	.17 ±.06	.06 ±.01	.05 ±.02	.13 ±.06	.15 ±.02	.08 ±.01	.05 ±.01	**.23** ±.07
	1000	**.30** ±.03	.20 ±.03	.08 ±.02	.08 ±.01	.22 ±.03	.11 ±.01	.06 ±.01	.03 ±.00	.25 ±.04
mHSC-L	500	**.15** ±.05	.13 ±.03	.07 ±.01	.05 ±.00	.11 ±.01	.13 ±.01	.10 ±.01	.06 ±.01	.12 ±.03
	1000	.10 ±.02	.10 ±.01	.06 ±.02	.05 ±.00	.09 ±.03	**.13** ±.01	.10 ±.02	.05 ±.01	.09 ±.01
STRING										
hESC	500	.07 ±.02	.12 ±.03	.04 ±.00	.05 ±.00	**.14** ±.03	.05 ±.00	.04 ±.00	.04 ±.01	.13 ±.01
	1000	.01 ±.00	.02 ±.00	.02 ±.00	.04 ±.00	.02 ±.00	**.05** ±.00	.04 ±.00	.04 ±.00	.02 ±.00
hHEP	500	.04 ±.00	.06 ±.01	.03 ±.00	.05 ±.00	.07 ±.01	.05 ±.00	.04 ±.00	.04 ±.00	**.08** ±.00
	1000	.01 ±.00	.02 ±.00	.02 ±.00	**.04** ±.00	.02 ±.00	**.04** ±.00	.03 ±.00	.03 ±.00	.02 ±.00
mDC	500	.06 ±.01	.13 ±.02	.04 ±.00	.07 ±.01	.14 ±.02	.05 ±.00	.05 ±.00	.06 ±.00	**.15** ±.01
	1000	.02 ±.00	.02 ±.00	.02 ±.00	.04 ±.00	.02 ±.00	.04 ±.00	.04 ±.00	**.05** ±.00	.03 ±.00
mESC	500	.08 ±.01	.17 ±.02	.05 ±.00	.07 ±.00	**.26** ±.01	.05 ±.00	.04 ±.00	.05 ±.00	.25 ±.02
	1000	.05 ±.01	.09 ±.01	.04 ±.00	.05 ±.00	.17 ±.02	.05 ±.00	.03 ±.00	.03 ±.00	**.20** ±.01
mHSC-E	500	.07 ±.03	.07 ±.03	.05 ±.01	.06 ±.01	.09 ±.03	.15 ±.01	.08 ±.00	.09 ±.02	**.19** ±.05
	1000	.19 ±.08	.19 ±.06	.08 ±.01	.08 ±.01	.16 ±.04	.12 ±.00	.07 ±.00	.06 ±.02	**.31** ±.03
mHSC-GM	500	.27 ±.12	.28 ±.12	.15 ±.07	.09 ±.02	.20 ±.08	.38 ±.01	.20 ±.01	.05 ±.01	**.42** ±.08
	1000	.35 ±.14	.23 ±.04	.24 ±.03	.13 ±.01	.38 ±.01	.25 ±.02	.12 ±.01	.04 ±.01	**.44** ±.06
mHSC-L	500	.07 ±.01	.05 ±.01	.09 ±.02	.07 ±.02	.06 ±.01	**.55** ±.02	.35 ±.03	.14 ±.01	.13 ±.02
	1000	.11 ±.07	.06 ±.02	.08 ±.02	.05 ±.01	.08 ±.04	**.45** ±.03	.33 ±.03	.12 ±.03	.15 ±.02
Average AUPRC		.28	.30	.25	.21	.31	.19	.17	.16	**.34**

G Evaluation Protocol Comparison

We compare five splits per protocol across 42 settings with model and training configuration fixed (Section 3.4). Test pair label leakage in the training graph averages 0.786% under legacy splitting and reaches 2.021%; pair-disjoint splitting yields zero leakage and zero cross partition pair overlap. Legacy-to-pair candidate count ratios are 0.975–1.015, positive prevalence gaps are at most 0.0301 percentage points, and test set Jaccard similarity is 0.127–0.913. Thus, candidate membership changes at similar counts and class balance: the protocols test different gene pairs drawn from similarly sized candidate pools. Under legacy splitting, a held out link from TF u to target v remains visible if $v \to u$ is retained in the symmetrized training graph; pair isolation removes that direct route. Prediction must then use expression and the remaining network to reconstruct the missing connection. Table G1 gives the complete 42-setting results summarized in Table 3. Δ is legacy minus pair-disjoint before rounding; the table reports four decimal places.

Specific mean AUPRC changes by less than 0.01 at either scope. The larger changes occur on Non-Specific and STRING, with the STRING mean AUROC also decreasing from 0.92 at both scopes to 0.68 and 0.64, respectively. The agreement between AUPRC and AUROC shows that the STRING difference extends to discrimination across the candidate set. In contrast, Specific preserves its ranking performance when each evaluated pair is fully withheld, supporting recovery from the remaining cell matched regulatory context. Overall, higher legacy AUPRC and AUROC are consistent with optimistic evaluation of missing link recovery inflated by bidirectional label leakage.

Table G1: Protocol comparison across all 42 settings. Paired rows show TF+500 and TF+1000 (tinted); Δ is Legacy minus Pair before rounding.

Cell	Scope (TF+)	AUPRC Legacy	AUPRC Pair	AUPRC Δ	AUROC Legacy	AUROC Pair	AUROC Δ
Specific							
hESC	500	0.5910	0.6038	−0.0127	0.8708	0.8720	−0.0012
	1000	0.6069	0.6166	−0.0097	0.8769	0.8838	−0.0069
hHEP	500	0.8161	0.8301	−0.0139	0.8872	0.8953	−0.0081
	1000	0.8364	0.8385	−0.0021	0.9018	0.9028	−0.0011
mDC	500	0.1531	0.1587	−0.0056	0.7488	0.7557	−0.0070
	1000	0.1751	0.1775	−0.0025	0.7833	0.7845	−0.0012
mESC	500	0.8759	0.8834	−0.0075	0.9365	0.9408	−0.0043
	1000	0.8775	0.8826	−0.0051	0.9399	0.9432	−0.0034
mHSC-E	500	0.9414	0.9440	−0.0026	0.9213	0.9253	−0.0040
	1000	0.9503	0.9500	+0.0003	0.9361	0.9368	−0.0007
mHSC-GM	500	0.9299	0.9359	−0.0059	0.9201	0.9269	−0.0068
	1000	0.9469	0.9488	−0.0019	0.9351	0.9372	−0.0021
mHSC-L	500	0.8663	0.8762	−0.0099	0.8681	0.8785	−0.0104
	1000	0.8650	0.8674	−0.0024	0.8775	0.8781	−0.0006
Non-Specific							
hESC	500	0.0700	0.0413	+0.0286	0.7345	0.6945	+0.0401
	1000	0.0492	0.0390	+0.0102	0.7194	0.6953	+0.0241
hHEP	500	0.0814	0.0469	+0.0345	0.7523	0.6943	+0.0580
	1000	0.0516	0.0331	+0.0185	0.7285	0.6763	+0.0522
mDC	500	0.2242	0.0504	+0.1738	0.8305	0.6873	+0.1432
	1000	0.1662	0.0483	+0.1180	0.8038	0.7151	+0.0887
mESC	500	0.1945	0.0296	+0.1650	0.8537	0.6411	+0.2126
	1000	0.1392	0.0330	+0.1061	0.8448	0.7194	+0.1254
mHSC-E	500	0.2268	0.1763	+0.0505	0.7775	0.7137	+0.0638
	1000	0.2129	0.1472	+0.0658	0.7759	0.7415	+0.0345
mHSC-GM	500	0.3013	0.2279	+0.0734	0.7509	0.7036	+0.0473
	1000	0.2596	0.2468	+0.0129	0.7923	0.7587	+0.0336
mHSC-L	500	0.1259	0.1247	+0.0012	0.6723	0.6286	+0.0437
	1000	0.1180	0.0875	+0.0305	0.6715	0.6323	+0.0392
STRING							
hESC	500	0.3844	0.1287	+0.2557	0.9339	0.6584	+0.2755
	1000	0.4147	0.0210	+0.3936	0.9339	0.4151	+0.5188
hHEP	500	0.4610	0.0754	+0.3856	0.9415	0.4810	+0.4604
	1000	0.4788	0.0194	+0.4595	0.9432	0.4364	+0.5068
mDC	500	0.5569	0.1508	+0.4060	0.9476	0.5413	+0.4064
	1000	0.5840	0.0277	+0.5562	0.9461	0.4602	+0.4859
mESC	500	0.4532	0.2539	+0.1993	0.9487	0.7800	+0.1687
	1000	0.4997	0.2030	+0.2967	0.9494	0.6901	+0.2593
mHSC-E	500	0.3354	0.1919	+0.1435	0.9107	0.7676	+0.1431
	1000	0.3611	0.3115	+0.0496	0.9085	0.8442	+0.0643
mHSC-GM	500	0.4813	0.4194	+0.0619	0.8930	0.8149	+0.0781
	1000	0.4275	0.4366	−0.0091	0.8966	0.8733	+0.0233
mHSC-L	500	0.2880	0.1286	+0.1594	0.8432	0.7336	+0.1097
	1000	0.3014	0.1485	+0.1528	0.8331	0.7367	+0.0965

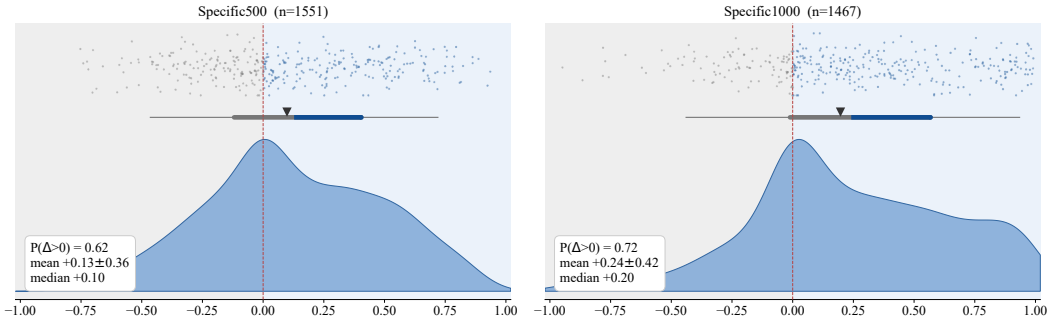


Figure H1: Distribution of forward-minus-reverse scores pooled across training repeats on eligible hESC links: Specific-500 (left panel) and Specific-1000 (right panel).

H Endpoint order sensitivity

Table H1: Endpoint order sensitivity on hESC Specific.

Metric	Specific-500	Specific-1000
Orientation AUROC	0.629 ± 0.048	0.725 ± 0.050
Forward win rate	0.617 ± 0.041	0.721 ± 0.047
Mean margin	0.128 ± 0.053	0.241 ± 0.064

For each eligible held out link $u \to v$, we compare its score with that of $v \to u$; a positive margin favors the annotated direction. Table H1 reports the forward win rate, mean margin, and orientation AUROC for distinguishing annotated from reversed TF-target links.

We evaluate hESC Specific-500 and Specific-1000 on the first fixed split, reporting means $\pm$ standard deviations across training runs. Eligible links are recorded in one direction only, with both endpoints qualifying as TFs and targets. The settings contain 517 and 489 eligible links per repeat, respectively. Figure H1 separates the two scope distributions pooled in Figure 3a.

Figure H1 pools edge repeat observations (n). Dots show margins, the thick bar the interquartile range, the grey/blue junction the mean, and the triangle the median. The lower curve shows density, and the dashed zero line marks equal forward and reverse scores. Insets summarize the pooled observations. The annotated direction scores higher in 62%/72% of observations for Specific-500/Specific-1000, with median margins of 0.10/0.20 and mean margins of 0.13/0.24. Positive means and medians indicate a preference for annotated directions in both gene scopes.

I Robustness and Hyperparameter Sensitivity

I.1 Robustness to reference perturbations

Perturbation design. On Specific/TF+1000 across seven cell types, we retain a fraction f of training positive gene pairs, stratified by source and target degree, and sample negatives at the original class ratio. We replace a fraction p of retained edges with equally many spurious edges between training visible genes, excluding self-pairs and both orientations of known positive, existing graph, and validation/test candidate pairs. The perturbed reference defines propagation and partner membership in Equation (5), while link labels retain the original supervision at each f. Partners vary between conditions but remain fixed across student edge resampling within each condition. Validation/test candidates are fixed over the 6×6 grid, $f \in \{0.5, 0.6, \ldots, 1.0\}$ and $p \in \{0, 0.1, \ldots, 0.5\}$. Results average training runs on one fixed split, with checkpoints selected by validation AUPRC. The non-contrastive GENELink control shares GCLink's encoder, features, partitions, and evaluation protocol. Figure I1 plots AUPRC against f and p with per cell color scales.

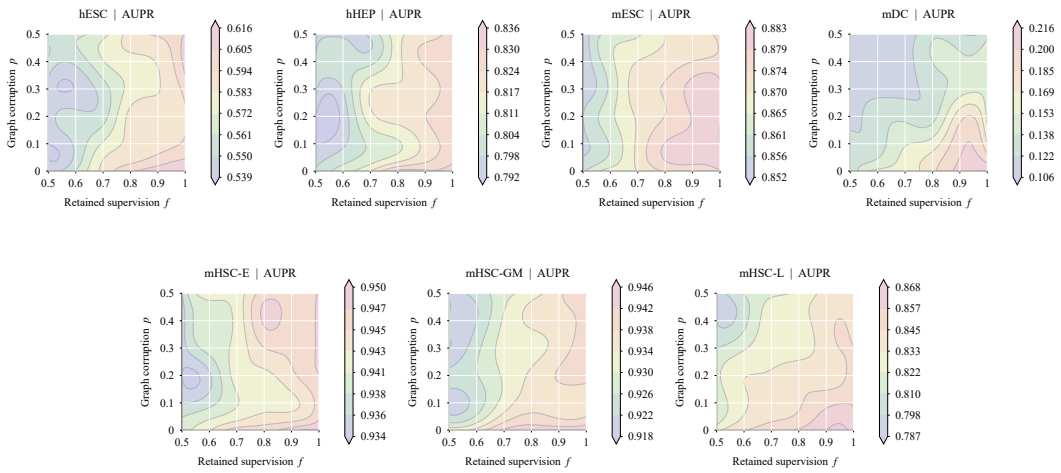


Figure I1: Robustness on Specific/TF+1000 across seven cell types.

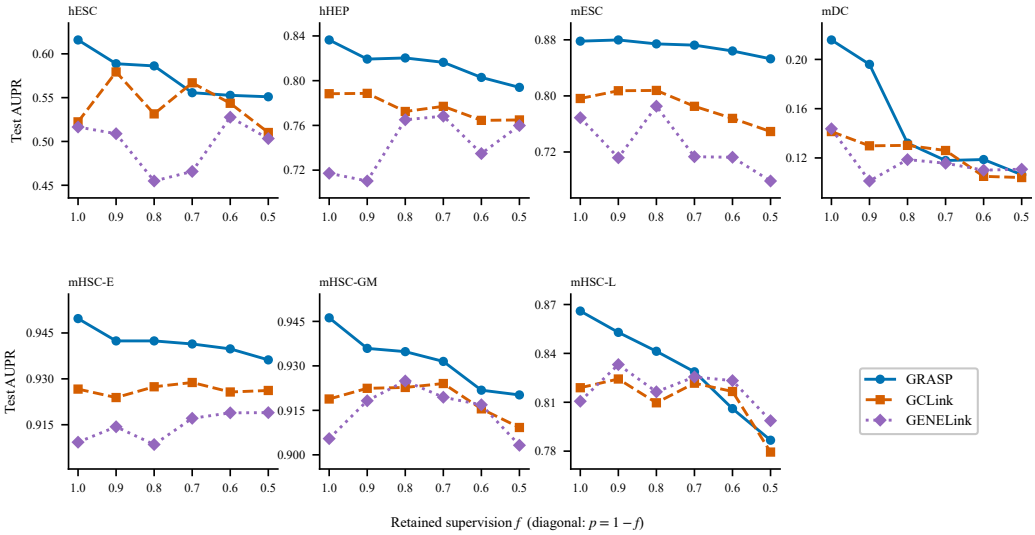


Figure I2: Per cell absolute test AUPRC along the full diagonal stress (f from 1.0 to 0.5 with $p = 1 - f$), GRASP, GCLink, and the non-contrastive GENELink control.

Performance retention. For cell c with mean test AUPRC $S_c(f, p)$, retention is $R_c(f, p) = 100 S_c(f, p)/S_c(1, 0)$. Figure 2(a) equally averages absolute AUPRC across seven cells along the diagonal; panel (b) shows GRASP's mean relative decrease, $100 - \frac{1}{7}\sum_c R_c(f, p)$, across the full stress grid, using each cell's clean performance as its reference.

At $(f, p) = (0.5, 0.5)$, six cells retain 89.5–98.6% of their clean AUPRC. The exception is mDC, whose retention falls to 61.2% at $(0.8, 0.2)$ and 49.3% at $(0.5, 0.5)$ (Figure I1). Its reference contains only 21 TFs and 1,193 positive pairs (Table E1), suggesting that reduced supervision and corrupted propagation and partner evidence are particularly detrimental for mDC.

Per cell comparisons. Figure I2 uses separate vertical ranges to complement the aggregate results in Section 3.6. GRASP leads each baseline at 39 of 42 diagonal points. At the joint corner, it exceeds GCLink in all seven cells by $+0.002$ (mDC) to $+0.104$ (mESC), and GENELink in five cells, including a $+0.174$ gap in mESC. The latter comparison reverses in mDC (-0.004) and mHSC-L (-0.012). Thus, the aggregate advantage covers most cell stress conditions while allowing cell-specific exceptions under the strongest perturbation.

I.2 HYPERPARAMETER SENSITIVITY

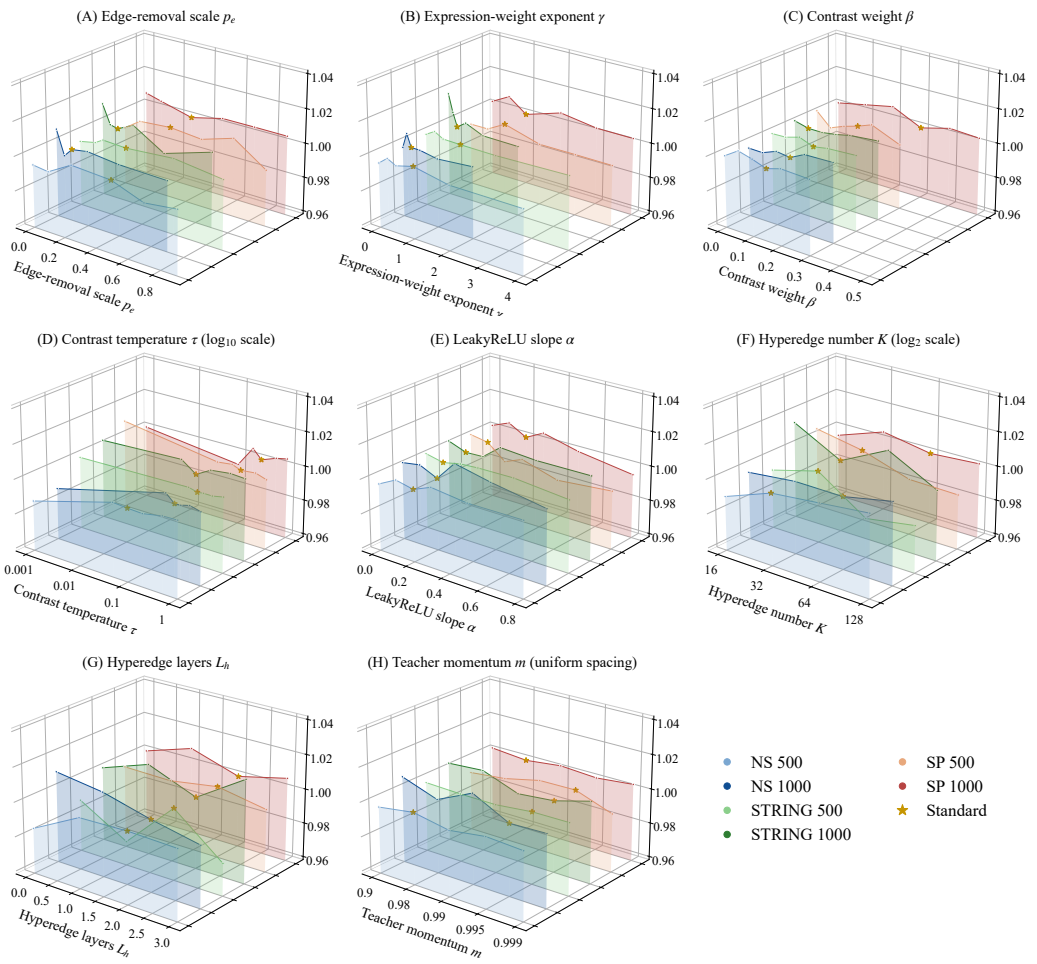


Figure I3: Hyperparameter and hyperedge sensitivity across six reference scope groups. Seven-cell mean validation AUPRC is normalized by each group's reference configuration, in gold. Panels (A–E) sweep five training parameters; panels (F–H) sweep hyperedge number K ($\log_2$ positions), hyperedge layers L_h, and teacher momentum m (uniform positions).

Parameter sweep. We vary one parameter at a time, holding the others at each group's reference values. Six values are tested for each training parameter: edge removal scale p_e, expression weight exponent γ, contrastive weight β, temperature τ, and LeakyReLU slope α. Additional sweeps use $K \in \{16, 32, 64, 128\}$ latent hyperedges, $L_h \in \{0, 1, 2, 3\}$ hyperedge layers, and teacher momentum $m \in \{0.9, 0.98, 0.99, 0.995, 0.999\}$. Each configuration is trained for 240 epochs on the first fixed split of seven cells in all six reference scope groups. Figure I3 normalizes seven-cell mean validation AUPRC at validation best checkpoints by the group's reference value. Heights relative to 1 show improvement or decline; gold stars mark the reference. Axes use $\log_{10}$ positions for τ, $\log_2$ for K, and uniform spacing for labeled momentum values.

Average response across references. For the five training parameters, we linearly interpolate normalized responses within each group's tested range and average values covered by at least three groups. The mean departs from 1 by at most 0.0047 (0.47%), with maximum between group SD 0.0091. The remaining three sweeps share grids across all groups, allowing direct averaging at their 13 tested values in total. Their maximum mean departure is 0.0031 (0.31%), with maximum between group SD 0.0132, indicating small aggregate changes.

Stability across parameter ranges. Contrastive weight and teacher momentum curves are the flattest, with per group departures around 0.01 or less (Figure I3). STRING-500 changes little across the five training parameters but shows the largest overall departure at $L_h = 1$ (about 0.972); $L_h = 0$ does not collapse. Varying K yields a 1.9% decline for STRING-500 at $K = 64$ and a 1.5% gain for STRING-1000 at $K = 16$. Across the eight sweeps, normalized validation AUPRC remains close to the reference. GRASP thus maintains stable ranking as graph perturbation, contrastive supervision, and higher-order context configuration vary within the tested parameter ranges.

J Cross-Cell and Unseen-Gene Generalization

J.1 Cross-cell transfer with limited supervision

Setup. We pretrain on all five fixed mESC data splits and adapt to six target cell types. Each target uses one fixed data split with nested finetuning budgets of 0%, 1%, 5%, 10%, and 30%. At each budget, a subset of gene pairs designated for finetuning is used for training, while the remaining pairs are reserved for evaluation. Both methods share candidate partitions, SVD-300 expression features, as well as training and model selection conditions. At zero budget, we use the source checkpoint. Across all six reference scope groups, boxes show medians and interquartile ranges over training runs; whiskers reach the extreme values within 1.5×IQR.

Cross-cell transfer performance. GRASP uses partial target cell supervision to adapt GRN ranking across cellular contexts. Figure 3b summarizes macro-AUPRC across 36 target, reference, and scope settings; Figure J1 shows variation across targets and runs. On Specific, GRASP exceeds GCLink at a 10% budget in all six TF+500 targets and five of six TF+1000 targets. At 30%, it leads in all six Non-Specific/TF+500 targets and all twelve STRING cell scope settings. STRING mean AUPRC is 0.61/0.62 at TF+500/TF+1000, versus 0.54/0.54 for GCLink, supporting transfer across the diverse human and mouse cellular contexts evaluated here.

Implications for target GRN completion. The STRING advantage at 30% suggests that local interaction labels help specialize transferred representations to the target context. Separation across hESC, hHEP, and mDC repeats supports consistent ranking across runs. Zero budget results use only the source checkpoint, while the higher budget curves show adaptation as target-specific evidence becomes available for predicting missing links in the target cell.

J.2 Inductive prediction for unseen genes

Prediction tasks. On STRING/TF+500/hESC, we test whether expression features enable GRASP to extend a GRN to genes absent during training. Target-cold predicts links from known TFs to new targets; source-cold predicts links from new TFs to known targets; node-disjoint predicts links between a new TF and a new target. At inference, new genes are represented by expression features, with their reference connections held out of the propagation graph.

Evaluation. Held out genes are excluded from training in both roles, and validation and test use separate unseen genes. Each task averages five node based splits and their training runs, with checkpoints selected by validation AUPRC. Candidates follow the gene-pair isolation and pseudo-negative rules in Appendix E.1. Target-cold and source-cold use approximately balanced validation classes; test candidate counts and positive fractions are reported as ranges across splits in Table J1, with positive prevalences as approximate reference values.

Table J1: Inductive prediction of regulatory links involving unseen genes.

Metric	Target-cold	Source-cold	Node-disjoint
Mean test AUPRC	0.0779	0.1329	0.4241
Mean test AUROC	0.7541	0.8157	0.6175
Test candidates	3,417–3,500	2,375–2,625	81–100
Test positives	82–84	57–63	18–32
Positive prevalence	∼2.4%	∼2.4%	∼24.8%

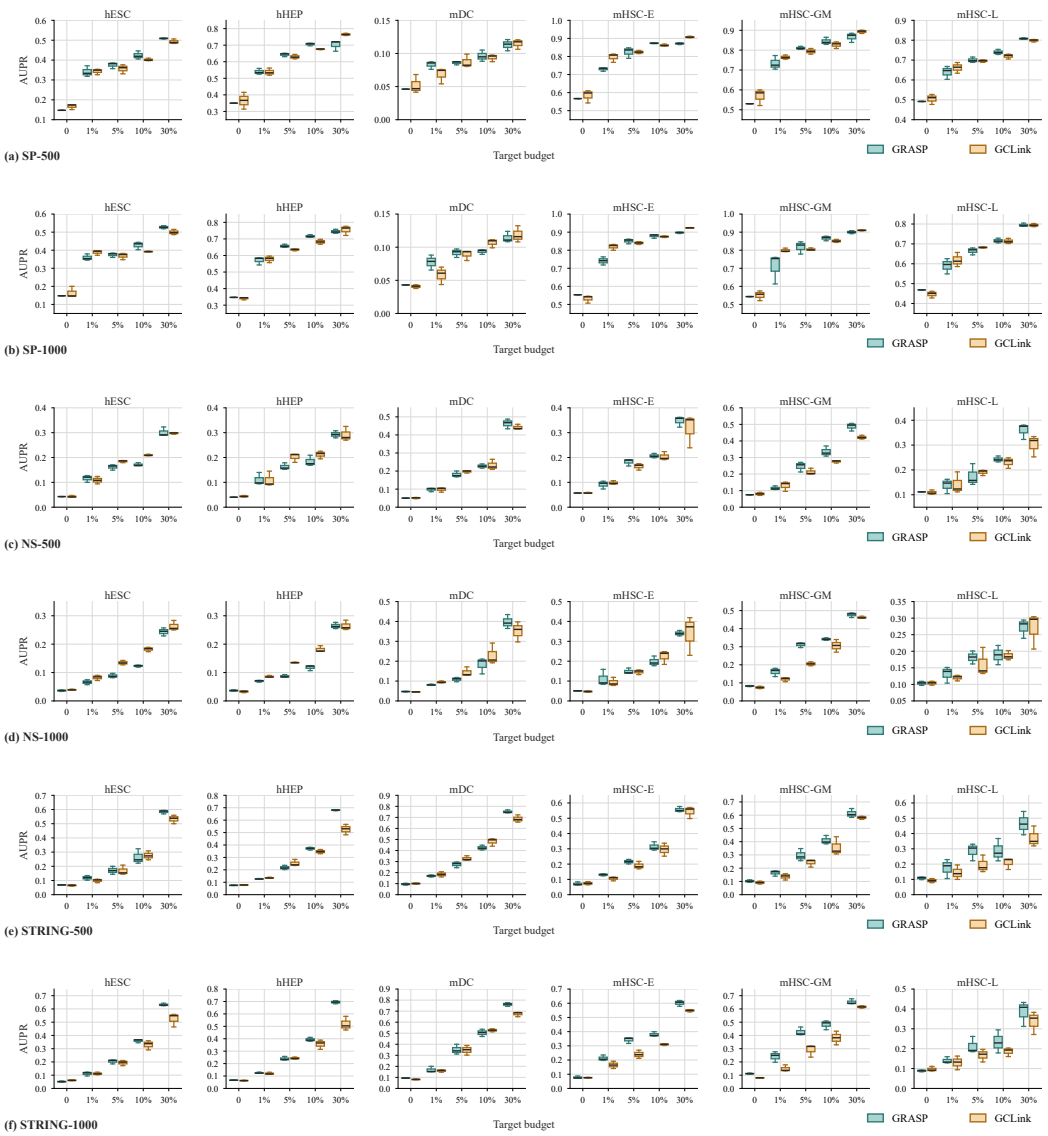


Figure J1: Cross-cell transfer of GRASP and GCLink from mESC to six target cell types across interaction references, gene scopes, and target label budgets.

Expanding the GRN through expression. GRASP prioritizes interactions involving unseen genes in all three tasks, with mean AUPRC above the corresponding positive fraction. Prediction is strongest for new regulators: source-cold reaches 0.82 AUROC and target-cold 0.75 at similar positive fractions. When both endpoints are unseen, mean AUROC falls to 0.62, with wider run-to-run variation (0.40–0.80) on the smaller 81–100-candidate test sets. These sets contain only 18–32 positive pairs, versus 57–84 in the one-unseen-endpoint tasks (Table J1). Each positive therefore contributes more to the AUROC estimate, making it more sensitive to individual link rankings. The larger node-disjoint AUPRC should also be read alongside its higher positive fraction (about 24.8%, versus 2.4% in the other tasks), so raw AUPRC alone does not establish that predicting two unseen endpoints is easier. Across these tasks, expression profiles support GRN expansion beyond training visible genes. With one unseen endpoint, candidate partners are observed genes; with both endpoints unseen, neither has training time graph context. The latter setting therefore warrants greater caution when interpreting individual predicted regulatory connections involving unseen endpoints.

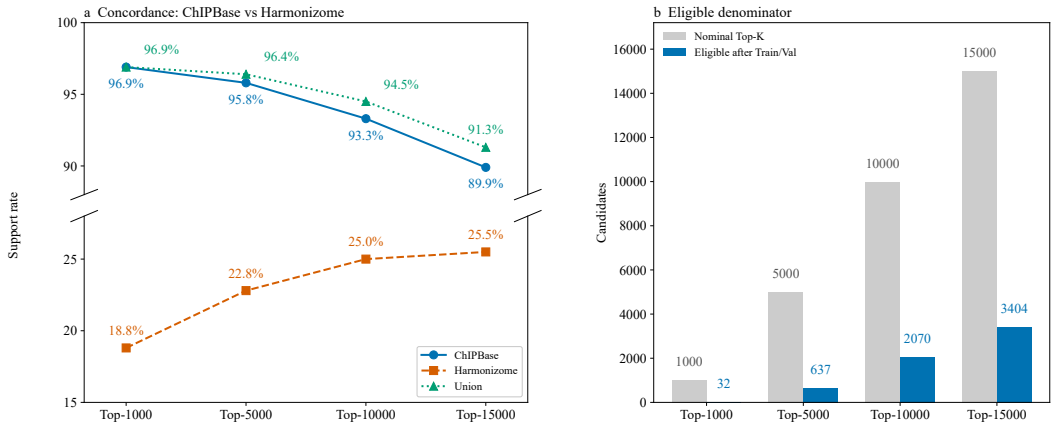


Figure K1: External database concordance across four Top-K cutoffs. (a) ChIPBase, Harmonizome, and union support among eligible pairs. (b) Nominal versus eligible counts.

K External Database Concordance

Table K1: External database support.

Measure	Top-1,000	Top-5,000	Top-10,000	Top-15,000
Eligible	32	637	2,070	3,404
ChIPBase	31 (96.88%)	610 (95.76%)	1,932 (93.33%)	3,059 (89.86%)
Harmonizome	6 (18.75%)	145 (22.76%)	518 (25.02%)	869 (25.53%)
Union	31 (96.88%)	614 (96.39%)	1,956 (94.49%)	3,109 (91.33%)

Table K2: Top ten eligible predictions.

TF	Target gene	External support
TFAP2A	ACSS3	ChIPBase
TFAP2A	AGL	ChIPBase
TFAP2A	ATP2B4	ChIPBase
TFAP2A	ZNF219	ChIPBase
TFAP2A	PPM1H	ChIPBase
TFAP2A	NECAP2	ChIPBase
TFAP2A	TSPAN12	ChIPBase
TFAP2A	TFB2M	ChIPBase
TFAP2A	ANKRD50	ChIPBase
TFAP2A	MRPS35	ChIPBase

Annotation matching. For the hESC Specific-1000 external analysis, we restrict the original 47,906-pair candidate pool (Table E1) to the 1,260 genes recorded as targets in the reference positives. The resulting 42,806 directed candidates across 34 TFs are scored by frozen validation best models. At each nominal Top-K cutoff, we select candidates by model score, remove all training and validation candidate pairs, including positives and pseudo-negatives, and retain the remaining pairs without replenishment. Support is the fraction of eligible pairs recorded in ChIPBase v3 (GRCh38) (Huang et al., 2023), the Harmonizome ChEA Transcription Factor Targets 2022 collection (Diamant et al., 2025), or their union. Applying the same exclusion to the full candidate universe leaves 9,959 eligible pairs, the matched background for ranked predictions. Within it, ChIPBase supports 78.95% (7,863 pairs), Harmonizome 25.28% (2,518), and their union 81.66% (8,133).

Concordance across cutoffs. Figure 3c summarizes ChIPBase and union support; Figure K1 additionally separates Harmonizome support and shows the reduction from nominal to eligible counts. Score based selection enriches for externally annotated regulatory links: union support reaches 96.39% for 637 eligible pairs and 91.33% for 3,404 (Table K1), the latter 9.67 percentage points above the matched background. ChIPBase supplies most matches, whereas Harmonizome adds 50 pairs at the largest cutoff. Its individual support rate stays near 25% at the two largest cutoffs, close to its matched background coverage (Figure K1). Furthermore, Harmonizome covers 27 of 34 candidate TFs, presenting a narrower coverage relative to ChIPBase.

Highest ranked TF-target connections. Table K2 lists the ten highest scoring pairs after excluding all training and validation candidate pairs. These held out connections all involve TFAP2A and have mean prediction probabilities between 0.996 and 0.999; each is recorded in ChIPBase. These pairs recover documented targets using expression and partial network context.