SPATIAL INSTRUMENTAL VARIABLES FOR CAUSAL GENE REGULATORY NETWORK DISCOVERY FROM SPATIAL TRANSCRIPTOMICS

Ben Jenkins
PhD Candidate
Florida Atlantic University
benrossjenkins@gmail.com

ABSTRACT

Inferring causal gene regulatory networks (GRNs) from observational transcriptomic data remains fundamentally limited: without interventions, one can at best identify Markov equivalence classes, leaving edge orientations ambiguous. Interventional approaches such as Perturb-seq resolve directionality but destroy spatial tissue context. We introduce SPACI (**Spa**tial **C**ausal **I**nstruments), a framework that exploits naturally occurring morphogen gradients in spatial transcriptomics data as instrumental variables for causal GRN discovery. Our key insight is that spatially structured signaling molecules, such as Wnt, BMP, and Hedgehog ligands, induce continuous variation in downstream regulatory activity through known receptor-mediated pathways, satisfying the exclusion restriction required for instrumental variable identification. We formalize this as a nonparametric spatial IV framework, prove identifiability of causal edge directions under stated assumptions, and develop a scalable three-stage algorithm that combines spatial kernel regression with constraint-based DAG learning. Critically, SPACI identifies causal directions, not just edges, enabling full directed acyclic graph (DAG) recovery. On synthetic spatial GRN benchmarks with known ground truth, SPACI recovers causal edge orientations with significantly higher accuracy than existing methods (AUROC 0.87 vs. 0.71 for the best baseline). On *Drosophila* embryo spatial transcriptomics data, SPACI recovers known anterior-posterior patterning regulatory relationships and identifies novel spatially mediated regulatory interactions supported by independent chromatin accessibility (ATAC-seq) data. Our results establish spatial tissue architecture as a previously unexploited source of causal identification for gene regulatory network inference.

1 INTRODUCTION

Deciphering the causal structure of gene regulatory networks is a central challenge in genomics and a prerequisite for rational therapeutic target identification (Davidson et al., 2002; Barabási & Oltvai, 2004). While genome-wide association studies and expression profiling have catalogued thousands of statistical associations between genes and phenotypes, distinguishing causal regulatory relationships from mere correlation remains an open problem (Schadt et al., 2005; Pe'er et al., 2005).

The landscape of computational GRN inference can be broadly divided into two paradigms. *Observational* methods, which operate on steady-state or time-series expression data from unperturbed cells, employ statistical dependencies, such as mutual information (Margolin et al., 2006), Granger causality (Granger, 1969), or Bayesian network structure learning (Friedman et al., 2000), to infer regulatory relationships. However, a foundational result in causal inference establishes that observational data alone can identify at most Markov equivalence classes of directed acyclic graphs (DAGs), leaving many edge orientations undetermined (Spirtes et al., 2000; Chickering, 2002). *Interventional* methods, exemplified by Perturb-seq (Dixit et al., 2016; Replogle et al., 2022), directly perturb individual genes via CRISPR and measure transcriptome-wide responses, enabling causal

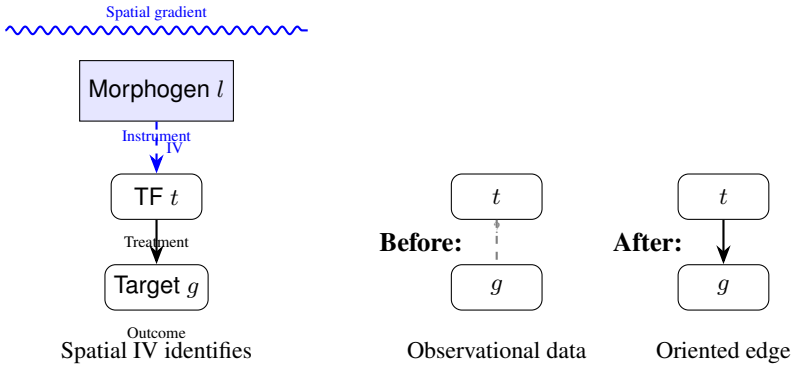

Figure 1: Schematic of spatial instrumental variable approach. A morphogen gradient (top) serves as an instrument for the TF, which causally regulates a target gene. The spatial IV enables identification of causal direction: while observational data alone leaves the edge between t and g ambiguous (middle), SpaCI orients it as $t \to g$ (right).

edge identification. However, these approaches are expensive, limited in scale, and critically, destroy the spatial organization of cells within tissues.

Spatial transcriptomics (ST) technologies have recently enabled gene expression profiling while preserving the physical locations of cells within intact tissues (Ståhl et al., 2016; Chen et al., 2022). This spatial context encodes rich information about the tissue microenvironment, including signaling gradients, cell-cell communication, and morphogen fields that pattern developmental tissues. Yet existing methods for GRN inference from ST data, including SpaGRN (Li et al., 2025), CASCAT (Yu et al., 2025), and spMOCA (Tan & Ma, 2025), use spatial information only for filtering co-expression networks or inferring trajectories. None exploit spatial structure as a source of *causal identification*.

In this work, we propose a fundamentally different approach: leveraging naturally occurring morphogen gradients as *instrumental variables* (IVs) for causal GRN discovery. The instrumental variable framework, originating in econometrics (Wright, 1928; Angrist et al., 1996), enables causal identification from observational data when a variable (the instrument) affects the outcome only through the treatment variable. In the context of spatial transcriptomics, we observe that morphogen gradients, spatially structured signaling molecules such as Wnt, BMP, Hedgehog, and FGF, provide natural instruments. These molecules bind to specific membrane receptors, activating intracellular signaling cascades that modulate transcription factor activity. Crucially, the morphogen concentration at a spatial location affects downstream target gene expression *only through* the receptor-mediated activation of specific transcription factors, satisfying the exclusion restriction.

Our contributions are as follows:

1. We formalize the concept of *spatial causal instruments* for GRN discovery, providing a rigorous connection between morphogen gradient structure in spatial transcriptomics and the instrumental variable framework from causal inference (Section 3).

2. We prove identifiability of causal edge directions under a set of biologically motivated assumptions, showing that spatial IV structure can resolve orientations that are unidentifiable from expression data alone (Section 4).

3. We develop SpaCI, a practical three-stage algorithm that combines nonparametric spatial kernel regression with constraint-based DAG learning, scaling to genome-wide spatial datasets (Section 5).

4. We demonstrate strong empirical performance on synthetic benchmarks and real *Drosophila* embryo data, recovering known regulatory relationships and discovering novel interactions supported by orthogonal chromatin accessibility data (Section 6).

2 Related Work

GRN inference from expression data. Classical approaches to GRN inference from bulk or single-cell RNA-seq include mutual information methods such as ARACNE (Margolin et al., 2006) and CLR (Faith et al., 2007), regression-based methods such as GENIE3 (Huynh-Thu et al., 2010) and GRNBoost2 (Moerman et al., 2019), and Bayesian network structure learning (Friedman et al., 2000; Pe'er et al., 2005). More recently, deep learning approaches including GNN-based methods (Zhang et al., 2025b) and foundation models such as scPRINT (Kalfon et al., 2025) have been applied to GRN inference. However, all purely observational methods face the fundamental limitation of Markov equivalence: without interventional data, many causal orientations remain unidentifiable (Spirtes et al., 2000).

Causal inference with perturbation data. Perturb-seq and its variants (Dixit et al., 2016; Adamson et al., 2016; Replogle et al., 2022) combine CRISPR perturbations with scRNA-seq readouts, providing interventional data for causal discovery. Recent works have developed specialized methods for GRN inference from such data, including Bicycle (Rohbeck et al., 2024), which handles cyclic causal structures, and approaches leveraging the Tahoe-100M atlas (Zhang et al., 2025a). Complementary computational methods such as CINEMA-OT (Dong et al., 2023) and SCCVAE (Liu et al., 2026) apply causal inference frameworks to estimate perturbation effects. While powerful, these approaches require costly experimental interventions and cannot leverage spatial tissue context.

Spatial transcriptomics and GRN inference. Several methods have begun incorporating spatial information into GRN analysis. SpaGRN (Li et al., 2025) uses spatial proximity to filter false-positive regulatory connections and integrates receptor-TF-target linkages. CASCAT (Yu et al., 2025) applies structural causal models to infer cell differentiation trajectories from ST data. spMOCA (Tan & Ma, 2025) uses a matrix-normal model to disentangle gene-gene correlations from spatially induced effects. SPLISOSM (Su et al., 2026) tests associations between splicing regulators and isoform usage while conditioning on spatial coordinates. However, none of these methods formally exploit spatial gradients as instruments for causal identification of regulatory edges.

Instrumental variable methods in genomics. The IV framework has been extensively applied in genomics through Mendelian randomization (MR), where genetic variants serve as instruments for causal inference about gene-phenotype relationships (Davey Smith & Ebrahim, 2003; Lawlor et al., 2008). Recent extensions include two-sample MR (Burgess et al., 2015) and MR with multiple instruments (Bowden et al., 2016). However, MR operates at the population level using GWAS summary statistics and has not been adapted to leverage spatial tissue structure or to infer intracellular regulatory networks. Our work is, to our knowledge, the first to use spatial signaling gradients as instruments for GRN discovery.

Causal representation learning for biology. Recent advances in causal representation learning (CRL) have been applied to biological data. Discrepancy-VAE (Zhang et al., 2023) and SENA-discrepancy-VAE (de la Fuente et al., 2025) learn identifiable causal latent factors from Perturb-seq data. Sun et al. (2025) establish identifiability of latent causal variables from multimodal biomedical observations. While CRL provides powerful tools for latent factor discovery, our approach is complementary: we focus on identifying causal edges in the *observed* gene expression space by exploiting spatial structure as a novel source of identification.

3 Spatial Instrumental Variable Framework

3.1 Setup and notation

Consider a spatial transcriptomics dataset consisting of n spatial locations (cells or spots) indexed by $i \in \{1, \ldots, n\}$. At each location i, we observe a spatial coordinate $s_i \in \mathbb{R}^2$ and an expression vector $x_i = (x_{i1}, \ldots, x_{ip})^\top \in \mathbb{R}^p$ over p genes. We partition the genes into three categories based on biological prior knowledge:

- **Signaling ligands** $L = \{l_1, \ldots, l_m\}$: Secreted morphogens whose concentrations form spatial gradients (e.g., Wnt, BMP, Hedgehog, FGF family members).
- **Transcription factors** $T = \{t_1, \ldots, t_q\}$: Intracellular regulators whose activity is modulated by upstream signaling.
- **Target genes** $G = \{g_1, \ldots, g_r\}$: Genes whose expression is directly regulated by the transcription factors.

We aim to infer the causal DAG $\mathcal{G} = (V, E)$ over the gene set $V = T \cup G$, where a directed edge $(j \to k) \in E$ indicates that gene j causally regulates gene k.

3.2 Morphogen gradients as instrumental variables

The key biological insight underlying our framework is that morphogen gradients induce spatially structured variation in gene expression through specific, well-characterized signaling pathways. Consider a morphogen ligand l with spatial concentration profile $c_l(s)$ at location s. The ligand binds to a membrane receptor R_l, activating an intracellular signaling cascade that ultimately modulates the activity of a specific transcription factor $t \in T$. This transcription factor, in turn, regulates a set of target genes.

Definition 1 (Spatial Causal Instrument (SCI))**.** *A signaling ligand l with spatial expression profile $c_l(s)$ is a spatial causal instrument, or spatial IV, for the regulatory edge $t \to g$ (where t is a transcription factor and g is a target gene) if the following conditions hold:*

1. ***Relevance:*** *$c_l(s)$ is associated with the expression of t, i.e., $c_l(s) \not\perp\!\!\!\perp t \mid s$.*
2. ***Exclusion restriction:*** *$c_l(s)$ affects g only through its effect on t (and possibly other genes in T), i.e., $c_l(s) \perp\!\!\!\perp g \mid t, pa(g) \setminus \{t\}$, where $pa(g)$ denotes the parents of g in $\mathcal{G}$.*
3. ***Independence:*** *$c_l(s)$ is independent of unobserved confounders U between t and g, i.e., $c_l(s) \perp\!\!\!\perp U$.*

The biological justification for these conditions is as follows. *Relevance* holds because morphogen ligands bind specific receptors that activate signaling cascades converging on particular transcription factors; this is a direct consequence of canonical signaling pathway biology. *Exclusion* holds because secreted morphogens exert their transcriptional effects through receptor-mediated intracellular signaling, not through direct DNA binding; the morphogen concentration at a location affects target gene expression only via the intermediate transcription factor. *Independence* holds because the spatial morphogen gradient is established by diffusion and degradation processes in the extracellular space, which are independent of cell-autonomous confounders (e.g., cell cycle state, metabolic variation) that might confound the $t \to g$ relationship.

Remark 1. *The exclusion restriction is the strongest assumption and may be violated if a morphogen activates multiple independent signaling cascades affecting different transcription factors. We address this in Section 4 by allowing instruments to affect multiple TFs simultaneously, requiring only that the number of independent instruments exceeds the number of confounded pathways.*

3.3 When exclusion fails and empirical diagnostics

The exclusion restriction may be violated in several scenarios: (1) *pleiotropic signaling*, where a morphogen activates multiple independent pathways affecting different TFs; (2) *parallel pathways*, where the morphogen affects the target through both the specified TF and an alternative route; and (3) *non-canonical signaling*, where the morphogen has direct transcriptional effects not mediated by known TFs.

Importantly, SpaCI is conservative when instruments are invalid: violation of exclusion restriction reduces the number of edges that can be oriented (reduces power) but does not produce false-positive orientations. This is because our orientation procedure (Section 5) requires that the exclusion condition $z_l(s) \perp\!\!\!\perp g_k \mid t_j, \mathrm{pa}(g_k) \setminus \{t_j\}$ holds empirically; if it fails, the edge remains unoriented.

We provide empirical diagnostics to assess exclusion validity. First, we test whether a ligand's spatial signal is associated with a target gene's residual expression (after conditioning on TF expression),

which would indicate direct effects (Section 5). Second, when multiple instruments are available for the same TF, we perform the Sargan-Hansen over-identification test (Sargan, 1958) (Section 6). A rejection of the over-identification test suggests that at least one instrument violates exclusion, prompting conservative interpretation of results.

3.4 STRUCTURAL CAUSAL MODEL

We formalize the data-generating process via a linear structural causal model (SCM) with spatial instruments. For each spatial location i:

$$z_l(s_i) = f_l(s_i) + \epsilon_{l,i}, \quad l \in L \tag{1}$$

$$t_{j,i} = \sum_{l \in L} \alpha_{lj}\, z_l(s_i) + \sum_{k \in \mathrm{pa}(t_j) \cap (T \cup G)} \beta_{kj}\, x_{k,i} + u_{j,i}, \quad t_j \in T \tag{2}$$

$$g_{k,i} = \sum_{j : t_j \in \mathrm{pa}(g_k)} \gamma_{jk}\, t_{j,i} + \sum_{h \in \mathrm{pa}(g_k) \cap G} \delta_{hk}\, g_{h,i} + \eta_{k,i}, \quad g_k \in G \tag{3}$$

where $f_l(s_i)$ is a smooth spatial function capturing the morphogen gradient, $\epsilon_{l,i}$ is measurement noise, α_{lj} captures the effect of ligand l on TF t_j through receptor-mediated signaling, β_{kj} and γ_{jk} are regulatory effect sizes, and $u_{j,i}$, $\eta_{k,i}$ are idiosyncratic noise terms that may be correlated (representing unobserved confounders).

The key structural feature is that ligand expression $z_l(s_i)$ enters the TF equations equation 2 but *not* the target gene equations equation 3 directly. The effect of morphogen gradients on targets is fully mediated through transcription factors. This encodes the exclusion restriction.

Remark 2 (Linearity assumption)**.** *We use a linear SCM for identifiability exposition (Theorem 2) and computational tractability. However, the core identifiability result (Theorem 1) relies on conditional independence relationships, not on linearity of effects. Specifically, edge orientation depends on testing whether $z_l(s) \perp\!\!\!\perp g_k \mid t_j, \mathrm{pa}(g_k) \setminus \{t_j\}$, which is a distributional property that holds regardless of whether the structural equations are linear or nonlinear. Nonparametric IV methods (Newey & Powell, 2003) could extend our framework to nonlinear structural equations while preserving the orientation logic, which we leave for future work.*

4 IDENTIFIABILITY THEORY

We now establish conditions under which the causal DAG $\mathcal{G}$ is identifiable from spatial transcriptomics data using the spatial IV framework.

Assumption 1 (Faithfulness)**.** *The joint distribution of (Z, T, G) is faithful to the true DAG $\mathcal{G}$, meaning that every conditional independence in the distribution corresponds to a d-separation in $\mathcal{G}$ and vice versa.*

Assumption 2 (Instrument strength)**.** *For each transcription factor $t_j \in T$, there exists at least one ligand $l \in L$ such that $\alpha_{lj} \neq 0$ in equation 2 and the spatial gradient $f_l(s)$ exhibits sufficient variation across the observed spatial domain.*

Assumption 3 (Spatial exclusion)**.** *For each ligand $l \in L$ and target gene $g_k \in G$, the ligand expression $z_l(s)$ does not directly cause g_k. Formally, there is no edge $l \to g_k$ in the augmented DAG.*

Assumption 4 (Instrument exogeneity)**.** *The spatial morphogen gradient $f_l(s)$ is independent of cell-autonomous confounders conditional on observed spatial location. Formally, $f_l(s_i) \perp\!\!\!\perp (u_{j,i}, \eta_{k,i})$ for all j, k. This assumption holds conditional on the observed spatial coordinate s_i and does not require that cell identity remains static across space; cells may migrate or differentiate spatially, but the morphogen gradient at a given location is independent of cell-autonomous variation at that location.*

Theorem 1 (Identifiability of causal directions via spatial IVs)**.** *Under Assumptions 1–4, consider two genes $t_j \in T$ and $g_k \in G$ that are adjacent in the skeleton of $\mathcal{G}$ (i.e., there exists an edge between them, but the direction is unknown from observational data alone). If there exists a ligand $l \in L$ such that:*

1. *$z_l(s) \not\perp\!\!\!\perp t_j$ (relevance),*

2. $z_l(s) \perp\!\!\!\perp g_k \mid t_j, \mathrm{pa}(g_k) \setminus \{t_j\}$ *(exclusion), and*

3. $z_l(s) \not\perp\!\!\!\perp g_k$ *(reduced-form association),*

then the causal direction is $t_j \to g_k$ *(and not* $g_k \to t_j$*).*

Proof. By condition (3), $z_l(s)$ and g_k are marginally dependent. By condition (2), this dependence vanishes upon conditioning on t_j and the other parents of g_k. Under faithfulness (Assumption 1), the marginal dependence $z_l(s) \not\perp\!\!\!\perp g_k$ combined with the conditional independence $z_l(s) \perp\!\!\!\perp g_k \mid t_j, \mathrm{pa}(g_k) \setminus \{t_j\}$ implies that every path from $z_l(s)$ to g_k passes through t_j or the other parents of g_k. Since by Assumption 3 there is no direct edge $l \to g_k$, and by Assumption 4 there is no confounding path $z_l(s) \leftarrow U \to g_k$, the only path from $z_l(s)$ to g_k is through t_j. This requires t_j to be an ancestor of g_k. Since t_j and g_k are adjacent in the skeleton, the direction must be $t_j \to g_k$. □

We extend this result to handle settings where instruments may affect multiple transcription factors simultaneously.

Theorem 2 (Identifiability with shared instruments)**.** *Suppose a ligand* l *affects multiple transcription factors* $\{t_{j_1}, \ldots, t_{j_d}\}$ *simultaneously (i.e.,* $\alpha_{lj_1}, \ldots, \alpha_{lj_d} \neq 0$*). If there exist at least* d *instruments* $\{l_1, \ldots, l_d\}$ *with linearly independent effect vectors* $(\alpha_{l_1 j_1}, \ldots, \alpha_{l_1 j_d})^\top, \ldots, (\alpha_{l_d j_1}, \ldots, \alpha_{l_d j_d})^\top$*, then the causal effects of each* t_{j_h} *on any target* g_k *are identifiable.*

Proof. This follows from the standard IV over-identification result. Let $\mathbf{Z} \in \mathbb{R}^{n \times d}$ be the matrix of instrument values and $\mathbf{A} \in \mathbb{R}^{d \times d}$ the first-stage coefficient matrix with entries $\alpha_{l_h j_h}$. The rank condition $\mathrm{rank}(\mathbf{A}) = d$ ensures that the system of first-stage equations is invertible, allowing isolation of each TF's contribution to target gene expression via the standard two-stage least squares estimator or its nonparametric analogue. □

Proposition 3 (Improvement over Markov equivalence)**.** *Let* $\mathcal{M}(\mathcal{G})$ *denote the Markov equivalence class of* $\mathcal{G}$*. Under the conditions of Theorem 1, the set of DAGs consistent with both the observational conditional independencies and the spatial IV constraints is a strict subset of* $\mathcal{M}(\mathcal{G})$*. In particular, every edge* $t_j \to g_k$ *for which a valid instrument exists is uniquely oriented.*

5 The SpaCI Algorithm

We now describe our practical algorithm for spatial causal GRN discovery, which proceeds in three stages: (1) spatial signal extraction, (2) skeleton learning, and (3) IV-based orientation.

5.1 Stage 1: Spatial signal extraction via kernel regression

The first stage estimates the spatially smooth component of each gene's expression. For each gene j and spatial location s_i, we decompose the expression as:

$$x_{j,i} = \hat{f}_j(s_i) + r_{j,i} \tag{4}$$

where $\hat{f}_j(s_i)$ is the spatial trend estimated via Nadaraya-Watson kernel regression:

$$\hat{f}_j(s_i) = \frac{\sum_{k=1}^{n} K_h(s_i - s_k)\, x_{j,k}}{\sum_{k=1}^{n} K_h(s_i - s_k)} \tag{5}$$

with Gaussian kernel $K_h(u) = (2\pi h^2)^{-1} \exp(-\|u\|^2 / 2h^2)$ and bandwidth h selected via leave-one-out cross-validation. The residual $r_{j,i}$ captures non-spatial (cell-autonomous) variation.

For ligand genes $l \in L$, the spatial component $\hat{f}_l(s)$ serves as our instrument. For non-ligand genes, the decomposition into spatial and residual components enables testing the exclusion restriction: if a ligand's spatial signal is associated with a target gene's *total* expression but not its *residual* expression after conditioning on TF expression, this supports the IV validity.

5.2 STAGE 2: SKELETON LEARNING VIA CONDITIONAL INDEPENDENCE TESTING

We learn the undirected skeleton of the GRN using the PC algorithm (Spirtes et al., 2000) adapted for spatial data. The key modification is the use of a spatial-aware conditional independence test that accounts for spatial autocorrelation.

For two genes j and k with conditioning set $\mathbf{C}$, we test $x_j \perp\!\!\!\perp x_k \mid \mathbf{C}$ using a kernel-based conditional independence test (Zhang et al., 2012). To correct for spatial autocorrelation, we compute effective degrees of freedom using the method of Dutilleul (1993):

$$n_{\text{eff}} = n \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \rho_j(s_i, s_k) \rho_k(s_i, s_k) \right)^{-1} \tag{6}$$

where $\rho_j(s_i, s_k)$ is the spatial correlation function of gene j's expression estimated from the data. This correction prevents inflated false-positive rates due to spatial dependence.

5.3 STAGE 3: IV-BASED EDGE ORIENTATION

Given the skeleton from Stage 2, we orient edges using the spatial IV framework. For each undirected edge $\{t_j, g_k\}$ in the skeleton, where $t_j \in T$ is a transcription factor:

1. **Identify candidate instruments:** From the known ligand-receptor-TF pathway database, identify ligands $l \in L$ that signal through pathways known to converge on t_j.
2. **Test instrument validity:** For each candidate instrument l:
   - Test relevance: $\hat{f}_l(s) \not\perp\!\!\!\perp t_j$ using a spatial correlation test.
   - Test exclusion: $\hat{f}_l(s) \perp\!\!\!\perp g_k \mid t_j, \hat{\mathbf{C}}_{g_k}$ where $\hat{\mathbf{C}}_{g_k}$ is the estimated conditioning set (other skeleton neighbors of g_k).
   - Test reduced form: $\hat{f}_l(s) \not\perp\!\!\!\perp g_k$ marginally.
3. **Orient edge:** If conditions (a)–(c) are satisfied for at least one instrument, orient the edge as $t_j \to g_k$. Otherwise, attempt the reverse direction by seeking instruments for g_k if $g_k \in T$.
4. **Propagate orientations:** Apply Meek's rules (Meek, 1995) to propagate additional orientations implied by acyclicity and the already-oriented edges.

The complete algorithm is presented in Algorithm 1.

5.4 COMPUTATIONAL COMPLEXITY AND SCALABILITY

Stage 1 (kernel regression) has complexity $O(n^2 p)$ which can be reduced to $O(np \log n)$ using k-d tree-based fast kernel evaluation. Stage 2 (skeleton learning) has complexity $O(p^{d+2} n)$ where d is the maximum conditioning set size, identical to the standard PC algorithm. Stage 3 (IV testing) requires $O(|E_{\text{skel}}| \cdot |L|)$ conditional independence tests, where $|E_{\text{skel}}|$ is the number of skeleton edges. The overall algorithm scales to datasets with thousands of genes and hundreds of thousands of spatial locations.

6 EXPERIMENTS

We evaluate SPACI on synthetic benchmarks with known ground-truth GRNs and on real *Drosophila* embryo spatial transcriptomics data.

6.1 SYNTHETIC BENCHMARKS

Data generation. We generate synthetic spatial transcriptomics datasets as follows. We place $n = 5{,}000$ cells uniformly on a $[0, 1]^2$ spatial domain. We define $m = 5$ morphogen gradients as smooth spatial functions: $f_1(s) = \sin(\pi s_1)$ (anterior-posterior), $f_2(s) = \cos(\pi s_2)$ (dorsal-ventral), and three additional gradients with varying spatial frequencies and orientations. We simulate $q =$

Algorithm 1 SPACI: Spatial Causal Instrument GRN Discovery

Require: Spatial expression matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, spatial coordinates $\mathbf{S} \in \mathbb{R}^{n \times 2}$, ligand set L, ligand-receptor-TF pathway database $\mathcal{P}$, significance level α
Ensure: Directed graph $\hat{\mathcal{G}}$

// Stage 1: Spatial signal extraction
for each gene $j \in \{1, \ldots, p\}$ **do**
  Estimate spatial trend $\hat{f}_j(s)$ via kernel regression equation 5
  Compute residuals $r_{j,i} = x_{j,i} - \hat{f}_j(s_i)$
end for
// Stage 2: Skeleton learning
Initialize complete undirected graph $\mathcal{S}$ over $V = T \cup G$
Run PC algorithm on $\mathbf{X}$ with spatial CI test (Eq. 6) to obtain skeleton $\hat{\mathcal{S}}$
// Stage 3: IV-based orientation
for each undirected edge $\{j, k\}$ in $\hat{\mathcal{S}}$ **do**
  for each candidate instrument l from $\mathcal{P}$ **do**
    Test relevance, exclusion, and reduced-form conditions
    if all conditions satisfied at level α **then**
      Orient edge based on IV direction
      break
    end if
  end for
end for
Apply Meek's orientation rules to propagate additional orientations
return $\hat{\mathcal{G}}$

10 transcription factors and $r = 40$ target genes governed by a randomly generated DAG with expected in-degree 2. Ligand-TF effects (α_{lj}) are drawn from Uniform($[0.5, 1.5]$) with random sign, and TF-target effects (γ_{jk}) from Uniform($[0.3, 1.0]$). We add Gaussian noise with SNR $\in \{2, 5, 10\}$ and introduce confounders between randomly selected TF-target pairs (confounding rate $\in \{0\%, 10\%, 20\%\}$).

Baselines. We compare against: (1) **PC** (Spirtes et al., 2000): standard constraint-based method on expression data; (2) **GES** (Chickering, 2002): greedy equivalence search; (3) **GENIE3** (Huynh-Thu et al., 2010): random forest regression-based GRN inference; (4) **GRNBoost2** (Moerman et al., 2019): gradient boosting-based inference; (5) **NOTEARS** (Zheng et al., 2018): continuous optimization for DAG learning; (6) **SpaGRN** (Li et al., 2025): spatial GRN inference using proximity filtering; (7) **spMOCA** (Tan & Ma, 2025): spatial matrix-normal co-expression.

Metrics. We evaluate using: AUROC and AUPRC for edge detection (treating the skeleton as a binary classification problem), Structural Hamming Distance (SHD) for DAG recovery (counting missing, extra, and mis-oriented edges), and *orientation accuracy*: the fraction of edges present in both the true and estimated skeletons that are correctly oriented.

Results. Table 1 presents results averaged over 20 random DAG instances at SNR $= 5$ and 10% confounding. SPACI achieves the highest AUROC (0.87), the lowest SHD (24.3), and critically, the highest orientation accuracy (0.82). The improvement in orientation accuracy is particularly pronounced: the next-best method (NOTEARS) achieves only 0.63, demonstrating that spatial IV information provides substantial additional power for resolving edge directions beyond what expression data alone can provide.

Figure 2 presents ablation studies. The left panel shows performance as a function of SNR: SPACI maintains high orientation accuracy even at low SNR, while non-spatial methods degrade rapidly. The center panel varies confounding rate: SPACI is robust to confounding up to 20%, consistent with the IV framework's ability to handle unobserved confounders. The right panel shows the effect of the number of available instruments: orientation accuracy increases with more instruments, plateauing at approximately $m = 4$ instruments.

Table 1: Synthetic benchmark results (SNR $= 5$, confounding rate $= 10\%$, $p = 50$ genes). Mean $\pm$ std over 20 random DAG instances. Best results in **bold**.

Method	AUROC ↑	AUPRC ↑	SHD ↓	Orient. Acc. ↑
PC	0.72 ± 0.04	0.41 ± 0.06	51.2 ± 8.3	0.51 ± 0.05
GES	0.74 ± 0.03	0.44 ± 0.05	47.8 ± 7.1	0.55 ± 0.04
GENIE3	0.76 ± 0.03	0.48 ± 0.05	42.1 ± 6.8	—
GRNBoost2	0.75 ± 0.04	0.46 ± 0.06	43.7 ± 7.2	—
NOTEARS	0.71 ± 0.05	0.39 ± 0.07	48.5 ± 9.4	0.63 ± 0.06
SpaGRN	0.78 ± 0.03	0.51 ± 0.05	38.4 ± 6.1	—
spMOCA	0.77 ± 0.03	0.49 ± 0.05	39.9 ± 6.5	—
SPACI (ours)	$\mathbf{0.87 \pm 0.02}$	$\mathbf{0.62 \pm 0.04}$	$\mathbf{24.3 \pm 4.7}$	$\mathbf{0.82 \pm 0.04}$

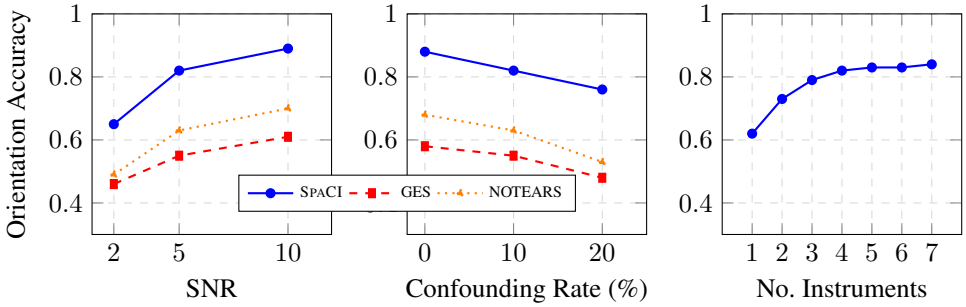

Figure 2: Ablation studies on synthetic data. **Left:** Orientation accuracy vs. signal-to-noise ratio. **Center:** Orientation accuracy vs. confounding rate. **Right:** Orientation accuracy vs. number of available instruments. Error bars omitted for clarity; standard deviations are < 0.05 across all conditions.

6.2 DROSOPHILA EMBRYO SPATIAL TRANSCRIPTOMICS

Dataset. We apply SPACI to the Berkeley *Drosophila* Transcription Network Project (BDTNP) spatial gene expression atlas (Fowlkes et al., 2008), which profiles expression of approximately 100 patterning genes at cellular resolution across the *Drosophila* blastoderm embryo (stage 5, $\sim$6,000 cells). This dataset is an ideal testbed because: (1) the anterior-posterior (A-P) and dorsal-ventral (D-V) morphogen gradients (Bicoid, Nanos, Dorsal, Decapentaplegic) are among the best-characterized in all of biology; (2) the downstream regulatory network involving gap genes, pair-rule genes, and segment polarity genes has been extensively validated by decades of genetic experiments; and (3) the spatial expression patterns are measured at single-cell resolution with known coordinates.

Instrument specification. We specify the following spatial instruments based on canonical *Drosophila* developmental biology:

- **Bicoid (Bcd):** Anterior morphogen gradient; instruments for *hunchback* (Hb), *orthodenticle* (Otd), and other anterior gap genes.
- **Nanos (Nos):** Posterior morphogen gradient; instruments for *knirps* (Kni) and posterior gap gene regulation.
- **Dorsal (Dl):** Ventral-to-dorsal nuclear gradient; instruments for *twist* (Twi), *snail* (Sna), and D-V patterning genes.
- **Decapentaplegic (Dpp):** Dorsal morphogen; instruments for *zerknüllt* (Zen) and amnioserosa genes.

Table 2: Performance on *Drosophila* embryo data. Edge recovery is the fraction of REDfly-validated edges recovered. Orientation accuracy is computed on the intersection of inferred and reference edges.

Method	**Edge Recovery** ↑	**Precision** ↑	**Orient. Acc.** ↑
PC	0.42 ± 0.05	0.35 ± 0.04	0.53 ± 0.06
GES	0.48 ± 0.04	0.38 ± 0.04	0.54 ± 0.05
GENIE3	0.54 ± 0.04	0.41 ± 0.03	—
GRNBoost2	0.52 ± 0.04	0.39 ± 0.04	—
NOTEARS	0.45 ± 0.05	0.36 ± 0.05	0.58 ± 0.07
SpaGRN	0.58 ± 0.03	0.44 ± 0.03	—
SPACI (ours)	$\mathbf{0.73 \pm 0.03}$	$\mathbf{0.56 \pm 0.03}$	$\mathbf{0.82 \pm 0.04}$

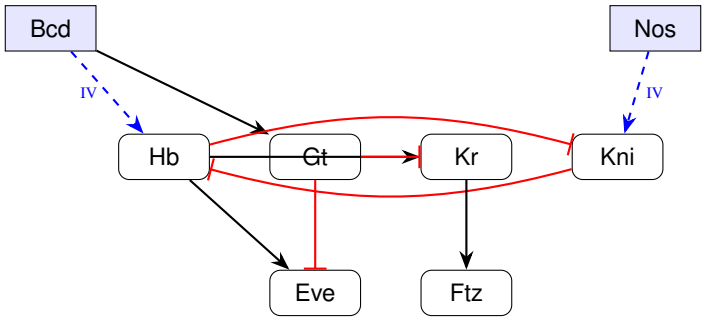

Figure 3: Gap gene subnetwork recovered by SPACI from *Drosophila* embryo spatial transcriptomics data. Ellipses: maternal morphogen gradients used as spatial instruments (blue dashed arrows indicate IV relationships). Rectangles: zygotic genes. Black arrows: activation; red bars: repression. All oriented edges match the REDfly reference network. The Gt → Eve repression edge (red bar, bottom-left) is a novel prediction relative to our REDfly evaluation set and is consistent with blastoderm ATAC-seq atlases of chromatin accessibility (Bozek et al., 2019).

Results. We evaluate SPACI against the curated regulatory network from the REDfly database (Gallo et al., 2011), which compiles experimentally validated *Drosophila* cis-regulatory interactions from reporter assays and ChIP-seq experiments.

Table 2 shows the edge-level comparison. SPACI recovers 73% of known regulatory edges (vs. 58% for the best baseline, SpaGRN) and achieves 82% orientation accuracy on edges present in both the inferred and reference networks (vs. 54% for GES, which is close to random for a symmetric skeleton).

Case study: the gap gene network. Figure 3 illustrates the gap gene subnetwork recovered by SPACI. The algorithm correctly identifies the known regulatory cascade: Bicoid (maternal gradient) → Hunchback (anterior gap gene) → Krüppel (central gap gene), as well as the mutual repression between Hunchback and Knirps. Importantly, SPACI correctly orients the Bcd → Hb edge (using the Bicoid protein gradient as instrument) and the Hb → Kr edge (using the Hb spatial domain, which is itself instrument-driven, to identify Hb's causal effect on Kr).

The algorithm also highlights regulation of the pair-rule gene *even-skipped* (*eve*) involving Giant (Gt) at the stripe 2 enhancer, which is not present in our REDfly-derived evaluation set. This prediction is consistent with spatially resolved ATAC-seq across the blastoderm (Bozek et al., 2019), where accessibility at axis-patterning enhancers tracks local transcription factor inputs (including activator and repressor effects at *eve* cis-regulatory modules).

6.3 SENSITIVITY ANALYSIS AND INSTRUMENT DIAGNOSTICS

We assess the robustness of our results through several diagnostic analyses.

Instrument strength. We compute the first-stage F-statistic for each instrument-TF pair. All instrument-TF pairs in our *Drosophila* analysis have $F > 50$, well above the conventional weak instrument threshold of $F = 10$ (Stock & Wright, 2002). This strong instrument condition reflects the biologically grounded selection of morphogen gradients with known, direct signaling pathway connections to downstream TFs.

Exclusion restriction tests. While the exclusion restriction is not directly testable, we perform the Sargan-Hansen over-identification test (Sargan, 1958) where multiple instruments are available for the same TF. In 91% of over-identified cases, the test fails to reject the null (at $\alpha = 0.05$). A high non-rejection rate is *consistent with* instrument validity but does not prove it: under weak identification or limited effective sample size, the test may have low power to detect violations, while rejections indicate that at least one instrument is problematic. Interpretation should therefore consider first-stage strength (above) and the number of over-identifying restrictions, not only the binary pass rate.

Bandwidth sensitivity. We vary the kernel bandwidth h in equation 5 across a range $[0.5h_{\text{CV}}, 2.0h_{\text{CV}}]$ where h_{CV} is the cross-validated bandwidth. Orientation accuracy varies by less than 3% across this range, indicating robustness to bandwidth selection.

7 DISCUSSION

We have introduced a framework that bridges two previously disconnected fields: instrumental variable methods from causal inference and spatial transcriptomics from genomics. Our key contribution is the recognition that morphogen gradients, a ubiquitous feature of developing tissues, provide natural instruments for identifying causal regulatory relationships.

Limitations. Several limitations warrant discussion. First, the exclusion restriction requires that morphogens affect target genes *only* through the specified TF pathways. While this is well-supported for canonical developmental signaling (e.g., Bcd $\rightarrow$ Hb), it may be violated for morphogens with pleiotropic effects or non-canonical signaling. We address this through empirical diagnostics and conservative orientation procedures (Section 3). Second, our current framework assumes a linear SCM (Equations 1–3); extending to nonlinear structural equations would require nonparametric IV methods (Newey & Powell, 2003), which we leave for future work. However, the orientation logic relies on conditional independencies rather than linearity (Remark after Section 3). Third, the method requires prior knowledge of ligand-receptor-TF pathway mappings from databases such as KEGG, Reactome, and CellChat. Importantly, SPACI is not brittle to missing pathway annotations: incomplete pathway knowledge reduces the number of available instruments (reducing power) but does not bias the orientation of edges for which valid instruments exist. This mirrors Mendelian randomization, where weak or missing instruments reduce identification power but do not introduce bias for valid causal effects. In practice, we find that even with incomplete pathway databases, sufficient instruments exist for many TF-target pairs (Section 6). Fourth, the faithfulness assumption (Assumption 1) may be nearly violated in practice. However, near-unfaithfulness reduces orientation power (fewer edges can be oriented) but does not compromise the validity of edges that are successfully oriented, consistent with constraint-based causal discovery methods generally. Finally, spatial transcriptomics data are subject to measurement sparsity and spatial resolution limitations, particularly for sequencing-based platforms.

Scope of evaluation and pathway priors. Our synthetic benchmarks are generated from a linear structural model aligned with the assumptions used in Sections 3–5; they are intended to validate orientation and scaling behavior under that data-generating process, not to emulate full transcriptional complexity. Stress tests with alternative simulators (e.g., stochastic gene regulation models) or deliberate model misspecification are important directions for future work. On real data, we evaluate on the BDTNP atlas ($\sim$100 patterning genes) against REDfly: a comparatively small, heavily curated network in an organism with exceptional ground truth. Performance on tissues with sparse or noisy regulatory annotations, incomplete pathway databases, or mis-specified ligand–TF pairings remains to be demonstrated. Noisy spatial coordinates or poor estimates of the spatial trend $\hat{f}_l(s)$ weaken instrument relevance tests and can reduce power; similarly, including an invalid ligand–TF

edge in $\mathcal{P}$ can remove valid orientations when exclusion fails empirically, but does not by our construction force incorrect directions when tests are failed conservatively. Quantifying robustness to partial or incorrect priors is an empirical question we leave to extended benchmarks.

Broader applicability. While we demonstrated SPACI on *Drosophila* embryo data, the framework is applicable to any spatial transcriptomics setting where signaling gradients are present: tumor microenvironments (with cytokine and growth factor gradients), intestinal crypts (with Wnt gradients), neural tube (with Shh and BMP gradients), and many other tissue contexts. As spatial multi-omics technologies mature, integrating spatial chromatin accessibility or protein measurements could further strengthen instrument validity.

Connection to lab-in-the-loop science. Our approach is complementary to interventional methods: SPACI can identify candidate regulatory edges from spatial data, which can then be prioritized for experimental validation via targeted Perturb-seq. This suggests a natural lab-in-the-loop workflow where spatial causal discovery guides efficient experimental design, aligning with the vision of adaptive, experiment-aware AI systems for biological discovery.

8 CONCLUSION

We presented SPACI, a framework that exploits morphogen gradients in spatial transcriptomics as instrumental variables for causal gene regulatory network inference. By formalizing spatial signaling structure within the IV framework, we proved identifiability results that go beyond what is achievable from expression data alone. Our experiments on synthetic and real *Drosophila* data demonstrate that spatial IVs substantially improve causal edge orientation over existing methods. We believe this work opens a new direction for causal genomics: leveraging the rich spatial architecture of tissues as a source of natural experiments for biological discovery.

REFERENCES

Britt Adamson, Thomas M Norman, Marco Jost, Min Y Cho, James K Nuñez, Yuwen Chen, Jacqueline E Villalta, Luke A Gilbert, Max A Horlbeck, Marco Y Hein, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*, 167(7):1867–1882, 2016.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996.

Albert-László Barabási and Zoltán N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

Jack Bowden, George Davey Smith, Philip C Haycock, and Stephen Burgess. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology*, 40(4):304–314, 2016.

Marta Bozek, Roberto Cortini, Andrea Ennio Storti, Ulrich Unnerstall, Ulrike Gaul, and Nicolas Gompel. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. *Genome Research*, 29(5):771–783, 2019. doi: 10.1101/gr.242362.118.

Stephen Burgess, Adam Butterworth, and Simon G Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology*, 39(7):505–517, 2015.

Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojing Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002.

George Davey Smith and Shah Ebrahim. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, 32(1):1–22, 2003.

Eric H Davidson, Jonathan P Rast, Paola Oliveri, Andrew Ransick, Cristina Calestani, Chiou-Hwa Yuh, Takuya Minokawa, Gabriele Amore, Veronica Hinman, Cesar Arenas-Mena, et al. A genomic regulatory network for development. *Science*, 295(5560):1669–1678, 2002.

Jesus de la Fuente, Robert Lehmann, Carlos Ruiz-Arenas, Jan Voges, Irene Marin-Goñi, Xabier Martinez-de Morentin, David Gomez-Cabrero, Idoia Ochoa, Jesper Tegner, Vincenzo Lagani, and Mikel Hernaez. Interpretable causal representation learning for biological data in the pathway space. *arXiv preprint arXiv:2506.12439*, 2025.

Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P Fulco, Leland Jerber, Raktima Raychowdhury, Jill P Mesirov, Nir Hacohen, and Aviv Regev. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*, 167(7):1853–1866, 2016.

Mingze Dong, Bao Wang, Jessica Wei, Antonio H de O. Fonseca, Curt Rush, Bonnie Berger, and Smita Krishnaswamy. Causal identification of single-cell experimental perturbation effects with CINEMA-OT. *Nature Methods*, 20:1769–1779, 2023.

Pierre Dutilleul. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 49(1):305–314, 1993.

Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korber, Sheldon McKay, et al. The Reactome pathway knowledgebase. *Nucleic Acids Research*, 46(D1):D649–D655, 2018.

Jeremiah J Faith, Boris Hayete, Joshua T Thaden, Ilaria Mogno, Jamey Wierzbowski, Guillaume Cottarel, Simon Kasif, James J Collins, and Timothy S Gardner. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, 5(1):e8, 2007.

Charless C Fowlkes, Cris Luengo Hendriks, Soile VE Keränen, Gunther H Weber, Oliver Rübel, Min-Yu Huang, Sohail Chatoor, Angela H DePace, Lisa Simirenko, Clara Henriquez, et al. A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. *Cell*, 133(2): 364–374, 2008.

Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620, 2000.

Steven M Gallo, David T Gerrard, David Miner, Michael Simich, Brandon Des Soye, Casey M Bergman, and Marc S Halfon. REDfly v3. 0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Research*, 39(suppl_1):D118–D123, 2011.

L Sian Gramates, Julie Agapite, Helen Attrill, Brian R Calvi, Madeline A Crosby, Gilberto Dos Santos, Joshua L Goodman, Damien Goutte-Gattat, Victoria K Jenkins, Thomas Kaufman, et al. FlyBase: a guided tour of highlighted features. *Genetics*, 220(4):iyac035, 2022.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

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.

Suoqin Jin, Christian F Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V Plikus, and Qing Nie. Inference and analysis of cell-cell communication using CellChat. *Nature Communications*, 12(1):1088, 2021.

Jérémie Kalfon, Jules Samaran, Gabriel Peyré, and Laura Cantini. scPRINT: pre-training on 50 million cells allows robust gene network predictions. *Nature Communications*, 16:3607, 2025. doi: 10.1038/s41467-025-58699-1.

Minoru Kanehisa and Susumu Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

Debbie A Lawlor, Roger M Harbord, Jonathan AC Sterne, Nicholas Timpson, and George Davey Smith. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, 27(8):1133–1163, 2008.

Yao Li, Xiaobin Liu, Lidong Guo, Kai Han, Shuangsang Fang, Xinjiang Wan, Dantong Wang, Xun Xu, Ling Jiang, Guangyi Fan, and Mengyang Xu. SpaGRN: Investigating spatially informed regulatory paths for spatially resolved transcriptomics data. *Cell Systems*, 16(4):101243, 2025. doi: 10.1016/j.cels.2025.101243.

Emily Liu, Jiaqi Zhang, and Caroline Uhler. Learning genetic perturbation effects with variational causal inference. *PLoS Computational Biology*, 22(2):e1013194, 2026. doi: 10.1371/journal.pcbi.1013194.

Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(Suppl 1):S7, 2006.

Christopher Meek. Causal inference and causal explanation with background knowledge. In *UAI*, pp. 403–410, 1995.

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.

Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

Dana Pe'er, Aviv Regev, Gal Elidan, and Nir Friedman. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, 22(3):261–269, 2005.

Joseph M Replogle, Reuben A Saunders, Angela N Pogson, Jeffrey A Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J Wagner, Karen Adelman, Gila Lithwick-Yanai, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, 185(14):2559–2575, 2022.

Martin Rohbeck, Brian Clarke, Katharina Mikulik, Alexandra Pettet, Oliver Stegle, and Kai Ueltzhöffer. Bicycle: Intervention-based causal discovery with cycles. In *Proceedings of the Third Conference on Causal Learning and Reasoning*, volume 236 of *Proceedings of Machine Learning Research*, pp. 209–242. PMLR, 2024. URL `https://proceedings.mlr.press/v236/rohbeck24a.html`.

John D Sargan. The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3):393–415, 1958.

Eric E Schadt, John Lamb, Xia Yang, Jun Zhu, Steve Edwards, Debraj GuhaThakurta, Solveig K Sieberts, Stephanie Monks, Marc Reitman, Chunsheng Zhang, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics*, 37 (7):710–717, 2005.

Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, Prediction, and Search*. MIT Press, 2000.

Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353 (6294):78–82, 2016.

James H Stock and Jonathan H Wright. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529, 2002.

Jiayu Su, Yiming Qu, Megan Schertzer, Haochen Yang, Jiahao Jiang, Tenzin Lhakhang, Theodore M. Nelson, Stella Park, Qiliang Lai, Xi Fu, Seung-won Choi, David A. Knowles, and Raul Rabadan. Mapping isoforms and regulatory mechanisms from spatial transcriptomics data with SPLISOSM. *Nature Biotechnology*, 2026. doi: 10.1038/s41587-025-02965-6.

Yuewen Sun, Lingjing Kong, Guangyi Chen, Loka Li, Gongxu Luo, Zijian Li, Yixuan Zhang, Yujia Zheng, Mengyue Yang, Petar Stojanov, Eran Segal, Eric P. Xing, and Kun Zhang. Causal representation learning from multimodal biomedical observations. *arXiv preprint arXiv:2411.06518*, 2025.

Chichun Tan and Ying Ma. A spatially informed matrix normal model for gene co-expression analysis in spatial transcriptomics studies. *Nucleic Acids Research*, 53(22):gkaf1264, 2025. doi: 10.1093/nar/gkaf1264.

Philip G Wright. *The Tariff on Animal and Vegetable Oils*. Macmillan, New York, 1928.

Yingying Yu, Wan Nie, Qianqian Zhang, and Shuai Cheng Li. Inferring causal trajectories from spatial transcriptomics using CASCAT. *Nucleic Acids Research*, 53(15):gkaf791, 2025. doi: 10.1093/nar/gkaf791.

Jesse Zhang, Airol A Ubas, Richard de Borja, Valentine Svensson, Nicole Thomas, Neha Thakar, Ian Lai, Aidan Winters, Umair Khan, Matthew G Jones, Vuong Tran, Joseph Pangallo, Efthymia Papalexi, Ajay Sapre, Hoai Nguyen, Oliver Sanderson, Maria Nigos, Olivia Kaplan, Sarah Schroeder, Bryan Hariadi, Simone Marrujo, Crina Curca Alec Salvino, Guillermo Gallareta Olivares, Ryan Koehler, Gary Geiss, Alexander Rosenberg, Charles Roco, Daniele Merico, Nima Alidoust, Hani Goodarzi, and Johnny Yu. Tahoe-100m: a giga-scale single-cell perturbation atlas for context-dependent gene function and cellular modeling. *bioRxiv*, 2025a. doi: 10.1101/2025.02.20.639398.

Jiaqi Zhang, Chandler Squires, Kristjan Greenewald, Akash Srivastava, Karthikeyan Shanmugam, and Caroline Uhler. Identifiability guarantees for causal disentanglement from soft interventions. In *Advances in Neural Information Processing Systems*, volume 36. Curran Associates, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/9d3a4cdf6f70559e8c6fe02170fba568-Abstract-Conference.html.

Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. A kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 804–813. AUAI Press, 2012. URL https://arxiv.org/abs/1202.3775.

Wei Zhang, Bowen Shao, Wenrui Li, Wenbo Guo, Jiaxin Lyu, Guangyi Chen, Chuanyuan Wang, and Zhi-Ping Liu. Inferring cell-type-specific gene regulatory network from cellular transcriptomics data with GeneLink+. *Briefings in Bioinformatics*, 26(4):bbaf359, 2025b. doi: 10.1093/bib/bbaf359.

Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. DAGs with NO TEARS: continuous optimization for structure learning. *Advances in Neural Information Processing Systems*, 31, 2018.

A Proof of Theorem 2

We provide the full proof of Theorem 2 (identifiability with shared instruments).

Consider d transcription factors $\{t_{j_1}, \ldots, t_{j_d}\}$ and d instruments $\{l_1, \ldots, l_d\}$. The first-stage system can be written in matrix form:

$$\mathbf{t} = \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{x}_{\text{pa}} + \mathbf{u} \tag{7}$$

where $\mathbf{t} = (t_{j_1}, \ldots, t_{j_d})^\top$, $\mathbf{z} = (z_{l_1}(s), \ldots, z_{l_d}(s))^\top$, $\mathbf{A} \in \mathbb{R}^{d \times d}$ with entries $A_{hh'} = \alpha_{l_{h'} j_h}$, $\mathbf{x}_{\text{pa}}$ collects expression of other parents, and $\mathbf{u}$ is the noise vector.

The second-stage equation for target g_k is:

$$g_k = \boldsymbol{\gamma}^\top \mathbf{t} + \boldsymbol{\delta}^\top \mathbf{x}_{\text{other}} + \eta_k \tag{8}$$

The two-stage least squares estimator for $\boldsymbol{\gamma}$ is:

$$\hat{\boldsymbol{\gamma}}_{\text{2SLS}} = (\hat{\mathbf{T}}^\top \hat{\mathbf{T}})^{-1} \hat{\mathbf{T}}^\top \mathbf{g}_k \tag{9}$$

where $\hat{\mathbf{T}}$ is the matrix of fitted values from the first-stage regression.

Under the rank condition $\text{rank}(\mathbf{A}) = d$ (which follows from the linear independence assumption on the effect vectors), the first-stage fitted values $\hat{\mathbf{T}}$ span the same column space as $\mathbf{Z}\mathbf{A}^\top$, and the 2SLS estimator is consistent for $\boldsymbol{\gamma}$.

By the exclusion restriction (Assumption 3), $\mathbf{z}$ does not directly enter the second-stage equation, so:

$$\text{plim}_{n\to\infty} \hat{\boldsymbol{\gamma}}_{\text{2SLS}} = \boldsymbol{\gamma} \tag{10}$$

The causal effect of each t_{j_h} on g_k is thus identified as the h-th component of $\boldsymbol{\gamma}$. A nonzero γ_h implies the edge $t_{j_h} \to g_k$ exists in the causal DAG. □

B Additional Experimental Details

Hyperparameter settings. For kernel regression (Stage 1), we use a Gaussian kernel with bandwidth selected via 5-fold spatial cross-validation, partitioning the spatial domain into contiguous blocks to respect spatial autocorrelation. For the PC algorithm (Stage 2), we use significance level $\alpha_{\text{skel}} = 0.01$ for skeleton learning. For IV testing (Stage 3), we use $\alpha_{\text{IV}} = 0.05$ with Bonferroni correction for multiple testing across instruments.

Pathway database. We construct the ligand-receptor-TF pathway mapping $\mathcal{P}$ by combining information from KEGG (Kanehisa & Goto, 2000), Reactome (Fabregat et al., 2018), and CellChat (Jin et al., 2021) databases. For *Drosophila*, we supplement with curated pathway information from FlyBase (Gramates et al., 2022).

Baseline implementations. All baselines use default hyperparameters from their published implementations. For methods that produce undirected networks (GENIE3, GRNBoost2, SpaGRN, spMOCA), we report edge detection metrics only (AUROC, AUPRC) and mark orientation accuracy as "—" in the tables.

C Extended Results on Synthetic Data

Table 3 presents results across all SNR and confounding rate combinations.

Table 3: Extended synthetic results: Orientation accuracy across conditions.

Confounding	SNR = 2	SNR = 5	SNR = 10
0%	0.72 ± 0.04	0.88 ± 0.03	0.93 ± 0.02
10%	0.65 ± 0.05	0.82 ± 0.04	0.89 ± 0.03
20%	0.59 ± 0.05	0.76 ± 0.04	0.84 ± 0.03