空间转录组学中空间异质因果结构的贝叶斯学习

Kyungwook Back，高丽大学统计学系；Junsouk Choi†高丽大学统计学系

arXiv:2611.33094v1 [stat.ME] 24 Nov 2026

摘要

空间转录组学保留了每个表达谱的位置信息，使研究组织内基因调控网络的差异成为可能。在组织的不同区域，调控关系可能发生变化，甚至反转方向，因此，识别特定病理区域中的因果关系有助于发现治疗靶点。我们提出 SpaDAG，这是一种用于学习空间变化因果有向无环图的贝叶斯方法。通过在组织的三角剖分上定义图，SpaDAG 允许每条边具有自身的空间模式，甚至能在不同区域反转方向。我们证明，这些图仍然完全可识别，并开发了一种马尔可夫链蒙特卡罗采样器，通过一次移动在整个连通区域内反转一条边，从而高效地探索这些图。在有方向反转和无方向反转的模拟中，SpaDAG 的表现均优于其他方法。在一份涵盖肾细胞癌及其邻近肾脏的切片中，SpaDAG 发现两种组织中的关系存在显著差异，且肿瘤中的关系集中于与治疗靶点相关的基因。在其中两个基因之间，它识别出一种在肿瘤边界两侧方向反转的关系，据此提出了关于它们区域特异性耦合的假设。本文的补充材料和代码可在线获取。

关键词：因果发现；有向无环图；高斯过程；空间尖峰与板块先验；空间分子谱分析；可识别性

†通讯作者：junsoukchoi@korea.ac.kr

1 引言

科学背景。从表达数据中恢复基因调控网络是基因组学领域由来已久的难题。了解哪些基因调控哪些其他基因，有助于解释细胞状态如何得以维持或发生改变，以及这种调控在疾病中何处出现异常。批量 RNA 测序测量的是组织样本中多个细胞的汇总表达，因此难以区分不同细胞类型特有的调控关系。单细胞 RNA 测序可提供单个细胞的表达谱，从而能够研究不同细胞类型和状态下的调控网络（Aibar 等人，2017）。然而，传统的单细胞 RNA 测序需要对组织进行解离，因此无法保留细胞原本的位置。这限制了人们研究调控关系如何随局部组织背景而变化，例如细胞与肿瘤边界的距离。空间转录组学是一项近期进展，它通过测量完整组织切片中的表达谱来保留每个表达谱的位置；根据所用平台的不同，该表达谱来自单个细胞，或来自覆盖切片的网格上的一个点（Rao 等人，2021年）。因此，它揭示了基因表达的空间异质性，使人们不仅能够研究哪些基因之间存在关联，还能探究这种关联在组织中的具体位置。

组织切片包含细胞组成和表达程序各不相同的区域，例如肿瘤边界的两侧、不同的皮层层次，或发生免疫浸润的部位。这些区域中的基因间调控关系也可能有所不同。某个区域中存在连接的基因对，在另一区域中可能并无连接；对于某些基因对，哪个基因位于上游也取决于所在区域。例如，在斑马鱼胚胎中，血细胞调控因子 gata1 和 pu.1 就是这样一对基因：在红细胞最初形成的区域，gata1 抑制 pu.1；而在白细胞最初形成的区域，pu.1 抑制 gata1（Monteiro 等人，2011年）。我们旨在确定组织中哪些位置存在调控关系，以及这些关系的方向是否因位置而异

存在于组织中，以及它们的方向是否因位置而异 s.

启发性案例研究：透明细胞肾细胞癌肿瘤-肾脏界面。 肿瘤边界是研究这类差异的自然场景，因为肿瘤细胞与同一器官中相邻的正常细胞位于同一切片内。我们研究透明细胞肾细胞癌（ccRCC）中的这一场景。透明细胞肾细胞癌是肾细胞癌（RCC）最常见的亚型，起源于肾小管上皮细胞（Hsieh 等人，2017）。

大多数透明细胞肾细胞癌肿瘤都存在肿瘤抑制因子 VHL 的遗传改变或表观遗传改变（Nickerson 等人，2008）。在氧气充足的细胞中，VHL 蛋白会标记缺氧诱导因子（HIFs）的α 亚基，使其降解，从而阻止这些转录因子激活低氧反应相关基因（ Kaelin，2008）。当 VHL 失活时，由 EPAS1 基因编码的缺氧诱导因子-2α α 亚基会积累，促进肿瘤生长，并提高 VEGFA 等缺氧诱导因子靶基因的表达（Raval 等，2005）。京都基因与基因组百科全书中的肾细胞癌通路（hsa05211，Kanehisa 和 Goto，2000）包含参与这一过程及与之相关的生长因子信号传导的基因。针对晚期透明细胞肾细胞癌的疗法作用于该通路的多个环节，包括 VEGF信号通路和缺氧诱导因子-2α （Choueiri 和 Kaelin，2020；Choueiri 等人，2024）。然而，该通路描述的是疾病整体层面各基因之间的关系，并未考虑其在组织中的位置。我们探究肿瘤与相邻肾组织之间的这些调控关系有何不同，包括关系是否存在，以及哪个基因处于上游。

我们分析了一张横跨透明细胞肾细胞癌（ccRCC）及其邻近肾组织的组织切片，该切片由 Li 等人（2022）生成，使用的是随 CELINA 分析（Shang 等人，2025）一同发布的计数数据。该切片采用 10x Visium 平台进行分析，该平台可测量覆盖在组织上的空间点中的基因表达，每个空间点通常覆盖多个细胞。我们将每个空间点标记为

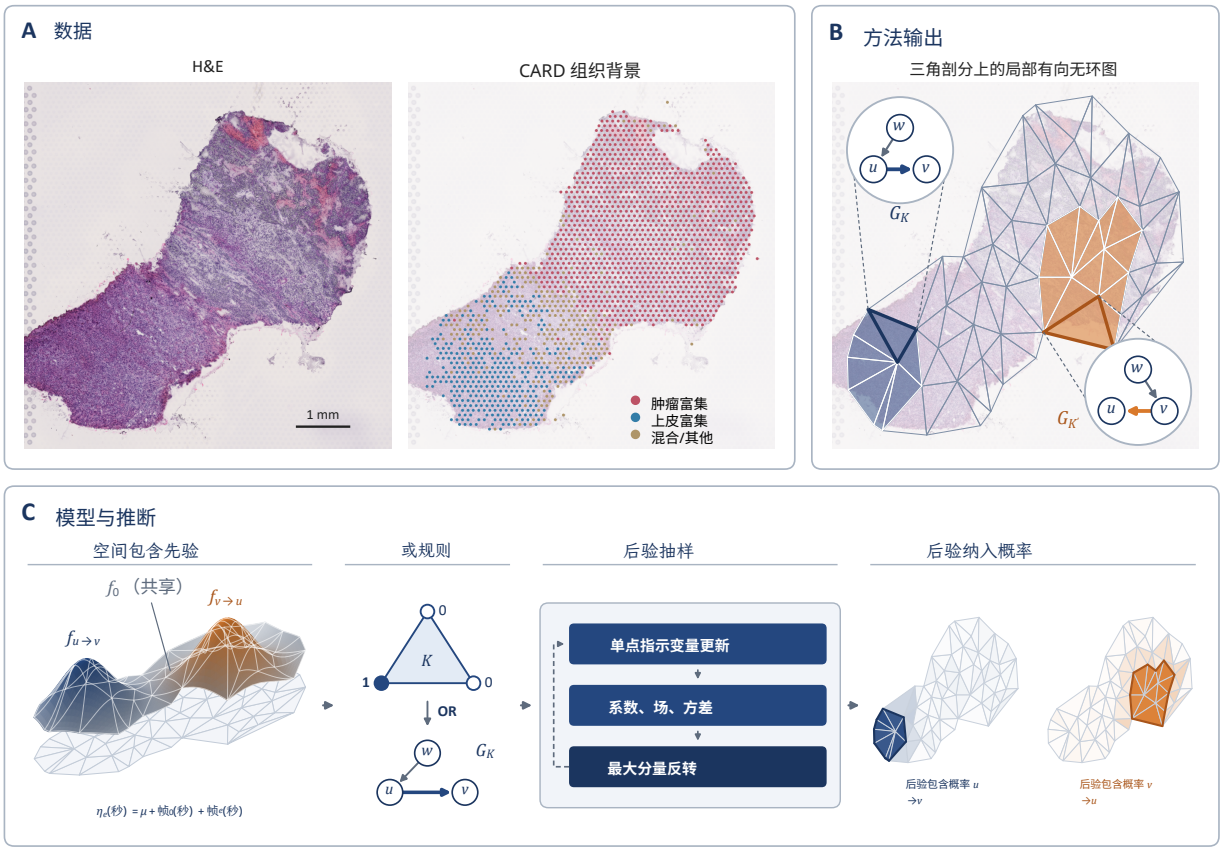

图 1：SpaDAG 概览。(A) 透明细胞肾细胞癌肿瘤–肾脏界面切片：苏木精-伊红图像，以及 1,885 个保留空间点的 CARD 组织背景。比例尺：1 mm。(B) 切片三角剖分上的局部有向无环图；着色三角形分别携带 $u \to v$ （蓝色）或 $v \to u$（橙色）。(C) 每个方向的纳入倾向由共享场 f_0 和方向特异性场共同决定；顶点指示变量通过或规则给出局部有向无环图，而每个方向的后验纳入概率（PIP）则对采样得到的局部有向无环图取平均。图 B 和图 C 为示意图。

根据细胞类型比例，将其标记为肿瘤富集、上皮富集或混合/其他；细胞类型比例由去卷积方法 CARD 估计（Ma 和 Zhou，2022）；图 1A 在苏木精和伊红染色（H&E）图像旁显示了这些标签。这些标签并非我们方法的输入；我们的方法仅使用空间点的表达谱和坐标，它们的作用是帮助解释该方法估计的关系。

空间异质因果结构与现有方法。哪个基因位于另一个基因的上游，以及扰动某个基因会带来哪些变化，这些都是因果问题。我们在因果图模型框架下处理这些问题，其中每个位置的调控结构都表示为具有因果解释的有向无环图（DAG）。这种表示建立在因果马尔可夫假设之上（Spirtes 等人，2000）：给定某个基因的直接原因（父节点），该基因与其非后代条件独立。因果解释则源于这样一种设定：将给定父节点时每个基因的条件分布视为一种机制，即使对其他基因进行干预，该机制仍保持不变（Pearl，2009）。从观察性数据中估计这类图，称为因果结构学习；它一直是基因组学中生成因果假设的一种工具（Maathuis 等人，2010）。

在每个位置都有一个有向无环图时，调控结构可能以两种方式在组织不同区域之间发生变化：某条边可能在一个区域存在而在另一个区域不存在，或者在两个区域中指向相反方向。然而，一些联合估计多个有向无环图的方法只允许第一种变化，因为它们施加了共享的拓扑排序，即一种变量排序，使每个父节点都排在其子节点之前（Ni 等，2019；Wang 等人，2020）。在这种排序下，只要某条边存在，它在任何位置的指向都相同。因此，我们旨在研究的区域性方向变化，只有在不对整个组织施加单一排序的情况下才能表示。

此外，组织中的邻近位置往往具有相似的细胞组成和表达程序，因此它们的调控关系也可能相似。在每个位置估计一个有向无环图时应当利用这种相似性，因为相对于候选边的数量而言，一个小邻域包含的观测数量很少。然而，不同的调控关系不一定以相同方式在组织中变化。一种关系可能贯穿整个切片，而另一种关系可能仅限于肿瘤区域，或在其他位置改变方向。因此，我们希望在邻近位置之间共享信息，同时允许每条边具有其自身的空间模式。

现有方法满足了其中一些要求。空间转录组学的网络方法利用空间信息来估计在组织中变化的图，但输出的是无向图（Acharyya 等，2025）或不具因果性的调控网络（Li 等，2026）。其他方法允许有向无环图随单元或上下文变化，且不施加共同的排序。有些方法会为每个单元（如受试者）估计一个有向无环图（Oates 等，2016）；而 ContextualDAG（Thompson 等，2024）则通过共享神经网络将上下文变量映射到有向无环图。应用于空间数据时，ContextualDAG 将坐标作为上下文变量，仅通过神经网络隐式地关联邻近位置的图，且无法为因果解释提供可识别性保证。这些方法都未能将每个位置上的因果有向无环图与空间依赖的显式模型结合起来。

即使不同位置之间共享信息，数据仍然有限，且许多图对数据的解释几乎同样好，因此关键在于数据对每条边及其方向的支持力度。贝叶斯结构学习通过对图的后验分布进行平均，而不是只选取一张图，来解决这一问题（Madigan 和 York，1995）。针对若干组图的贝叶斯方法已被提出，这些图估计到马尔可夫等价类为止（Castelletti 等，2020）；此外，还有适用于随协变量变化的有向无环图的方法（Ni 等，

2019），以及适用于由潜在伪时间索引的有向无环图的方法，沿着潜在伪时间，边可能出现、消失或改变方向（Zhou 等人，2023）。在空间转录组学中，BNP-DCGx（Dawn 和 Ni，2025）为每个空间点簇估计一张图，其中可能包含有向环。它依据取决于空间点坐标的先验对空间点进行划分，并让所有边共享这一划分，因此，给定该划分后，每种关系只能在相同的簇之间变化。这些方法能够量化图的不确定性，但没有一种方法能针对每个位置的因果有向无环图做到这一点，同时还为每条边显式建模空间依赖。

我们的贡献。 我们提出 SpaDAG（空间变化有向无环图），这是一种贝叶斯方法，可根据空间转录组学数据推断组织切片中每个位置的因果有向无环图，且无需强加共享的拓扑排序。SpaDAG 在组织三角剖分的每个三角形上放置一个局部有向无环图，并且只要求每个三角形内部无环，因此同一基因对的方向可以在不同三角形之间有所不同，如图1B所示。如图1C所示，针对每条边设置的空间先验会在邻近位置之间共享信息，同时允许每条边具有各自的空间模式；此外，采样器可以通过一次移动，在一个连通区域内反转一条边，从而得出每条局部边存在与方向的后验概率。当每个基因的误差方差在空间上保持恒定时，局部有向无环图以先验概率1可识别。在模拟中，SpaDAG 对局部有向图的恢复准确度高于其他竞争方法。在透明细胞肾细胞癌切片中，肿瘤富集组织与上皮富集组织所支持的有向关系几乎完全不同。EPAS1 与生长因子基因 HGF 之间的估计方向，在肿瘤富集区域内外有所不同，而采用共享排序的模型无法表示这种差异。

第2节介绍模型和可识别性结果，第3节介绍后验推断，

第4节介绍模拟，第5节介绍透明细胞肾细胞癌分析，第6节介绍局限性和扩展。证明、实现细节及其他结果详见补充材料，其中所有章节、公式、图和表的编号均带有前缀 S。实现 SpaDAG 的代码可从 `https://github.com/KyungWookBack/SpaDAG` 获取。

2 空间变化因果有向无环图的贝叶斯学习

SpaDAG 是一种贝叶斯结构方程模型，其因果有向无环图随空间变化。第2.1节在组织域的三角剖分上定义该模型，并在每个三角形上设置一个局部有向无环图。第2.2节表明，当误差方差在空间上保持恒定、结构系数随空间变化时，这些有向无环图可通过观察性数据完全识别。第2.3节提出了一种先验，明确对局部有向无环图之间的空间依赖进行建模。

2.1 SpaDAG模型

我们观测了有界二维定义域 $\Omega \subset \mathbb{R}^2$ 中 n 个空间位置 $s_1, \ldots\ldots, s_n$ 处 p 个基因的表达。令 $X_i = (X_{i1}, \ldots\ldots, X_{ip})^\top$ 表示位置 s_i 处的表达向量，并将 $\mathcal{D} = \{(s_i, X_i): i = 1, \ldots\ldots, n\}$ 记为数据。为允许调控结构随空间变化，我们引入空间变化有向无环图模型：在每个位置 $s \in \Omega$，基因之间的因果调控关系由有向无环图（DAG）$G(s) = (\mathcal{V}, E(s))$ 表示。节点集 $\mathcal{V} = \{1, \ldots, p\}$ 对基因进行索引，边集 $E(s) \subseteq \mathcal{V} \times \mathcal{V}$ 由构成有向环的有向边组成，并且可能随位置 s 变化。我们用 $\mathrm{pa}_{G(s)}(j)$ 表示 $G(s)$ 中节点 j 的父节点集，用 $\mathrm{ch}_{G(s)}(k)$ 表示节点 k 的子节点集。该模型假设，在给定

位置 s_i时， X_i 的联合分布根据因果有向无环图 $G(s_i)$进行因子分解（Lauritzen, 1996；Pearl，2009）：

$$p(X_i \mid s_i) = \prod_{j=1}^{p} p\big(X_{ij} \mid X_{i,\mathrm{pa}_{G(s_i)}(j)},\, s_i\big), \tag{1}$$

其中 $X_{i,\mathrm{pa}_{G(s_i)}(j)} = \{X_{ik} : k \in \mathrm{pa}_{G(s_i)}(j)\}$ 收集了基因 j 的父节点的表达。

由于 $G(s)$ 表示位置 s 处的因果结构，因此分解式（1）等价于位置层面的因果马尔可夫假设：在给定 s_i 并给定 $G(s_i)$ 中的父节点（即其直接调控因子）的条件下，每个基因都独立于其非后代，即独立于其既不直接调控也不间接调控的其他基因。这是关于给定位置时分布的陈述。在 Ω 上汇集的表达分布混合了各个位置的信息，因此不必相对于任何图 $G(s)$ 满足马尔可夫性质。按照惯例，因果马尔可夫假设以因果充分性为基础（Spirtes 等人，2000）；我们也要求逐个位置满足这一条件：在给定 s_i 的情况下， p 基因不存在未测量的共同原因。这并不排除由位置决定的未测量因素，因为它们的影响被吸收到条件分布（1）对 s_i 的依赖中。全文中，我们始终假设位置层面的因果充分性。

(1) 中每个 X_{ij} 的条件分布由空间结构方程模型（SEM）指定，该模型通过随空间变化的结构系数和空间基线，将每个基因与其余基因联系起来：

$$X_{ij} = \alpha_j(s_i) + \sum_{k \neq j} \beta_{jk}(s_i) X_{ik} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{\text{ind}}{\sim} N(0, \sigma_j^2). \tag{2}$$

这里，结构系数 $\beta_{jk}(s)$ 对应方向 $k \to j$，用于量化位置 s处基因 k 对基因 j 的直接因果效应。基线 $\alpha_j(s)$ 用于描述平均表达的空间变化，包括与组织组成和表达程序相关的变化。误差方差 σ_j^2 具有基因特异性，且在定义域 Ω 上保持不变。

在空间结构方程模型下，空间变化有向无环图 $G(s)$ 由位置 s处哪些结构系数非零决定：我们假设，当且仅当 $k \to j \in E(s)$ 时， $\beta_{jk}(s) \neq 0$ 。

我们通过三角剖分上的空间回归样条表示系数曲面和基线（Sangalli等人，2013）。设 $\mathcal{T}$为 Ω 的一个三角剖分，其节点为 ξ_1、……、ξ_H；设 $\{\psi_h\}_{h=1}^{H}$ 为 $\mathcal{T}$上由连续分片多项式函数构成的节点基，且$\psi_h(\xi_g) = \mathbf{1}\{h = g\}$。对于分片线性基，节点是 $\mathcal{T}$的顶点，而 ψ_h是帽函数：它在 ξ_h 处取值为 1，在其他所有顶点处取值为 0；对于分片二次基，节点还包括三角形边的中点，因此每个三角形包含六个节点。接着，我们令

$$\beta_{jk}(s) = \sum_{h=1}^{H} b_{jkh}\psi_h(s), \qquad \alpha_j(s) = a_{j0} + \sum_{h=1}^{H} a_{jh}\psi_h(s). \tag{3}$$

综合式 (1)–(3) 即定义了 SpaDAG模型，其中，学习空间异质因果图 $G(s)$ 可归结为对有限组基系数 b_{jkh}进行推断。

节点性质使得 b_{jkh} 就是 β_{jk} 在节点 ξ_h处的取值。在每个三角形上，曲面由该三角形各节点处的取值确定；相邻三角形共享其公共边上的节点，因此曲面连续。因此，当且仅当某个三角形各节点处的取值全为零时，曲面在该三角形上恒为零。这一性质构成了第 2.3.1 节局部图表示的基础，该表示将有向无环图学习转化为系数选择问题。为使行文简明，本文始终采用分片线性基函数。此时，节点即为 $\mathcal{T}$的顶点，且每个曲面在每个三角形上都是仿射的。也可以采用更高次的节点基。

三角剖分仅根据空间信息构建，并在后验抽样前固定；本文采用的构建方法见 S1.2节。三角剖分还决定了局部图的空间分辨率，其敏感性

网格构建将在第4节中进行评估。

2.2 可识别性

如果观测变量的分布能够唯一确定其图，那么因果DAG模型就是可识别的。一般来说，情况并非如此：蕴含相同条件独立关系的有向无环图构成一个马尔可夫等价类；如果不对结构方程或误差分布加以限制，它们会生成相同的分布集合（Verma和Pearl，1990）。同一等价类中的成员连接相同的变量对，但部分边的方向可能不同，因此观察性数据无法确定这些方向。要识别边的方向，模型就需要具备某种不对称性。在非线性加性噪声模型中，正向拟合和反向拟合通常可以区分（Hoyer等，2009）；在线性模型中，非高斯误差则提供了这种不对称性（Shimizu等，2006）。

为说明我们所采用的不对称性，我们首先考虑定义在单个固定DAG上的结构方程模型（SEM），其中系数允许变化。设 G 是定义在 $\mathcal{V}$ 上的DAG， $\mathcal{U}$ 是任意集合。在每个 $u \in \mathcal{U}$处，

$$X_j = \alpha_j(u) + \sum_{k \in \mathrm{pa}_G(j)} \beta_{jk}(u) X_k + \epsilon_j, \qquad \epsilon_j \overset{\text{ind}}{\sim} N(0, \sigma_j^2), \qquad j = 1, \ldots, p, \tag{4}$$

其中，截距 α_j 和系数 β_{jk} 是定义在 $\mathcal{U}$上的实函数，并且对于 G的任意边 $k \to j$， β_{jk} 在 $\mathcal{U}$ 上不恒等于零。因此，该图在每个 u处都相同，而其边上的系数可能随 u 变化，误差方差则不会变化。

条件1。对于每个 $\ell \in \mathcal{V}$ 以及每个非空子集 $S \subseteq \mathrm{ch}_G(\ell)$，和$\sum_{j \in S} \beta_{j\ell}(u)^2/\sigma_j^2$关于 $u \in \mathcal{U}$都不是常数。

命题1。在条件*1*下，有向无环图 G 以及截距、系数和误差

X 在 $u \in \mathcal{U}$处的分布可以识别(4) 的方差。

在 BN-LTE（Zhou 等人，2023）中，u 是一个潜在伪时间，DAG 会沿着它发生变化；在每组共享同一个 DAG 的伪时间上，模型的形式为（4）。此外，假设忠实性成立，BN-LTE 通过分别识别每组伪时间上的 DAG 来识别其各个有向无环图，这要求每个有向无环图都满足条件1。

在 SpaDAG 中，图 $G(s)$ 在三角形 $K \in \mathcal{T}$ 的每个位置都相同；我们将此图记为 G_K 。因此，在每个三角形上，模型（2）的形式为（4），其中 $\mathcal{U} = K$ 和$G = G_K$，且所有三角形共享误差方差。

定理1。 如果条件*1*在某个三角形 K_0 上成立，那么对于所有 $K \in \mathcal{T}$，有向无环图 G_K 以及模型 （2） 的截距、系数和误差方差，都可以根据给定 s、 $s \in \Omega$ 时 X 的条件分布识别。

关于 K_0 的条件1确定了每个误差方差。由于这些方差是共享的，因此在其他每个三角形中都可将其视为已知方差，从而使图和系数具有可识别性。因此，定理1只要求在单个三角形上满足条件1，而非对每个图都满足。命题1和定理1的证明见S2节。

该结果基于这样一个假设：每个误差方差在空间上均为常数。该结果可推广至按所有基因共有的因子变化的方差，$\sigma_j^2(s) = c(s)\sigma_j^2$（S2.4节），但不能推广至基因间噪声水平变化各不相同的情形。

2.3 先验设定

该先验表达了我们的假设：局部图是稀疏的，并且在邻近位置往往相似；但单条边仍可能从一个区域变为另一个区域，或反转方向。

2.3.1 通过尖峰与厚尾先验构建局部有向无环图

SpaDAG 根据三角形上相应系数是否恒等于零来定义局部边的存在性。我们使用二元指示变量对每个节点系数的结构活跃性进行编码。

令 $\mathcal{E}_c = \{(j,\ k)$ 表示候选方向，其中 $e = (j,\ k)$ 表示$k \to j$，而 $\bar{e} = (k,\ j)$ 表示其反向。在顶点 ξ_h 处，指示变量 $\gamma_{eh} \in \{0,1\}$ 记录相应系数 $b_{eh} = b_{jkh}$ 是否在结构上处于激活状态。点质量尖峰与厚尾先验将未激活的系数设为零，并为激活的系数指定高斯 slab：

$$b_{eh} \mid \gamma_{eh}, \nu_1 \sim (1-\gamma_{eh})\delta_0 + \gamma_{eh} N(0, \nu_1). \tag{5}$$

这里，δ_0 是集中于零点的点质量，ν_1 是所有方向和顶点共用的厚尾方差。

由于从厚尾分布中抽取的节点系数以概率 1 不等于零，因此，当且仅当某个系数曲面在一个三角形上的三个节点指示变量全为零时，该曲面在该三角形上几乎必然恒为零。因此，SpaDAG 通过对这些指示变量应用或规则来定义三角形层级的边存在性。令 $V(K)$ 表示三角形 K 的顶点索引，

$$\begin{aligned} z_{eK}(\gamma) &= \max_{h \in V(K)} \gamma_{eh}, \\ E_K(\gamma) &= \{k \to j : z_{(j,k)K}(\gamma) = 1\}, \qquad G_K(\gamma) = (\mathcal{V}, E_K(\gamma)). \end{aligned} \tag{6}$$

因此，图在每个三角形内保持不变，并在整个定义域上分段常数，而系数曲面则连续且分段线性。

为确保每个位置上的图都是有向无环图，我们将指示变量限制为满足每个诱导三角形图均为无环的配置：

$$\mathcal{A} = \left\{\gamma \in \{0,1\}^{|\mathcal{E}_c| \times H} : G_K(\gamma) \text{ is a DAG for every } K \in \mathcal{T}\right\}. \tag{7}$$

我们令 $G(s) = G_K(\gamma)$ 对于 $s \in K$ 三角形边界之外的部分，该集合的测度为零。该约束分别适用于每个三角形，无需在整个定义域内采用共同的拓扑排序

。随后，结构方程模型在每个位置都定义一个联合高斯分布，其均值和协方差随空间连续变化（第 S3.1 节）。

局部无环性和系数连续性也会约束方向如何随空间变化。对于一个连续系数场，若每个位置的图都是无环图，那么连接一对方向相反的边（u，v）的任何路径，都会经过一个两种方向系数均为零的点。在 SpaDAG 中，这种转变可以发生在两个相邻三角形的共享边界上。如果 K 和 K' 分别携带 $u \to v$ 和 $v \to u$，那么在它们的共享顶点处，两个方向都不能激活，因为在该处激活的方向会同时属于两个三角形图；在点质量尖峰先验下，两种系数随后都会沿共享边界归零。每个方向仍可在其所属三角形中与该边界相对的顶点处激活，因此两个方向可以分别位于相邻三角形中，而它们之间没有任何一个三角形会使这对边不相连。图2展示了两个相邻三角形上的这一约束：面板A中两个三角形图均为无环图，而在面板B中，共享顶点处激活 $u \to v$ 会在 $G_{K'}$中形成一个二环。

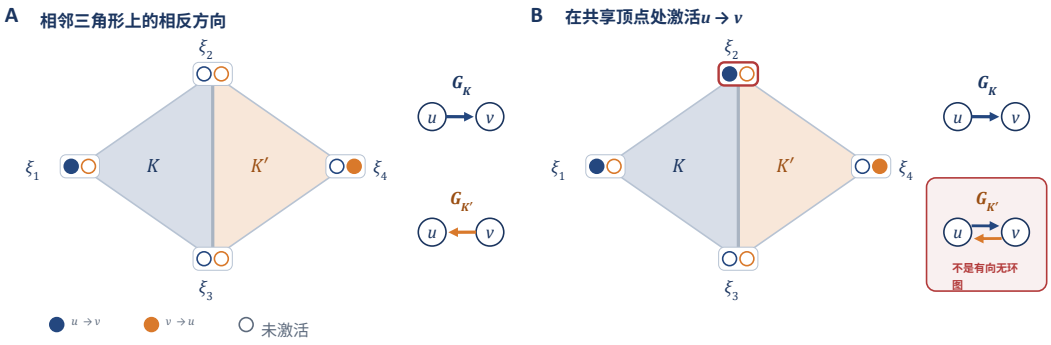

图2：相邻三角形 K 和 K' 上（u，v）的相反方向。（A）共享顶点处的两个方向均未激活时，G_K 包含 $u \to v$，且 $G_{K'}$ 包含 $v \to u$。（B）在 ξ_2 处激活 $u \to v$，通过或规则将其添加到 $G_{K'}$ 中，从而形成二环。

这种图表示通过共享顶点，将相邻三角形上的系数值与边支撑联系起来。接下来，我们将介绍一种空间包含先验，以便在更广泛的空间范围内共享有关边存在性的信息。

2.3.2 空间纳入先验

尖峰与厚尾先验 (5) 以节点指示变量为条件，下面我们将指定这些指示变量本身的先验。首先介绍不施加无环约束 (7) 的情形，该约束将在本小节末尾施加。SpaDAG 为每个候选方向$e \in \mathcal{E}_c$ 分配一个随空间变化的纳入倾向 $q_e(s) \in (0, 1)$, $s \in \Omega$；每个节点指示变量 γ_{eh} 服从以其顶点处的纳入倾向为参数的伯努利先验 ξ_h,

$$\gamma_{eh} \mid q_e \overset{\text{ind}}{\sim} \text{Bernoulli}\{q_e(\xi_h)\}. \tag{8}$$

由于纳入倾向因方向而异，因此每条边都可以具有各自的空间模式。

每个纳入倾向的逻辑斯谛变换 $\eta_e(s) = \text{logit}\{q_e(s)\}$，由一个总体纳入水平和两个空间分量构成：

$$\eta_e(s) = \mu + f_0(s) + f_e(s). \tag{9}$$

参数 μ 控制总体纳入水平。场 f_0 由所有候选方向共享；由于它会同步改变这些方向的纳入倾向，因此可标示图相对稠密或稀疏的区域。场 f_e 则针对方向 e，用于刻画该方向偏离这一共同趋势的程度。

我们为这些场指定零均值高斯过程先验，即 $f_0 \sim \text{GP}(0, \sigma_0^2\kappa)$ 和 $f_e \sim \text{GP}(0, \tau^2\kappa)$, 并施加中心化约束，以区分 μ、f_0 和 f_e（S1.1节）。公共核 $\kappa(s, s')$ 是 Ω 上的空间协方差核，在该核下，场在两个位置的取值之间的相关性会随着位置距离的减小而增强。因此，给定方向的倾向在邻近顶点处相似，并且该先验倾向于邻近位置共享边的配置。这些倾向的

场方差 σ_0^2 和 τ^2 从数据中学习得到，它们决定倾向在空间中的变化程度。变化的空间尺度由核函数确定（S1.1节）。本文使用的核函数见第3节。

(8)中的独立伯努利分布并不强制无环性：对于三角形图包含有向环的配置，它们赋予了正概率。令 ϑ 汇集总体纳入水平、场在网格顶点处的取值以及场方差。我们通过将上述分布限制在 $\mathcal{A}$中的指示变量配置上，定义$(\gamma,\ \vartheta)$的联合先验，其中 $\mathcal{A}$是(7)中定义的集合：

$$\pi(\gamma,\vartheta)\propto\pi_{\text{base}}(\vartheta)\,\mathbf{1}\{\gamma\in\mathcal{A}\}\prod_{e\in\mathcal{E}_c}\prod_{h=1}^{H}q_e(\xi_h)^{\gamma_{eh}}\{1-q_e(\xi_h)\}^{1-\gamma_{eh}},\tag{10}$$

其中，双重乘积是在(8)下 γ 的概率，而 $\pi_{\text{base}}(\vartheta)$是限制之前 ϑ的先验，由上述高斯过程先验以及 μ 和场方差的先验共同给出（S1.1节）。限制$(\gamma,\ \vartheta)$的联合分布，而非给定 ϑ 时 γ 的条件分布，可以避免关于 $\mathcal{A}$的归一化常数依赖于这些场。

2.3.3 SpaDAG先验下的可识别性

在第2.3.1–2.3.2节的SpaDAG先验下，(6)中的三角形图 $G_K(\gamma)$几乎必然是第2.2节中的有向无环图 G_K 。

推论1。在 *SpaDAG*先验下，每个三角形上的条件*1*几乎必然成立，且局部有向无环图、系数和误差方差均可识别。

在我们的空间回归样条表示（3）下，系数在每个三角形上都是多项式；当且仅当不存在某条边的系数

在该三角形上为常数时，条件1成立（S2.4节）。只有当某条边在该三角形的每个节点处均处于激活状态且所有节点值都相等时，该系数才会在三角形上为常数；在高斯 slab 下，这一事件的概率为零。该论证对指示变量的任何先验分布都成立；证明见 S2.3节。

3 后验推断

本节介绍所提出的 SpaDAG 模型的后验推断。由于后验分布难以解析求解，我们使用马尔可夫链蒙特卡罗（MCMC）从中采样。主要难点在于，采样器必须探索在每个三角形上均无环、且在相邻顶点之间相互耦合的指示变量配置。为高效探索这一受约束空间，我们开发了一种 Metropolis-within-Gibbs采样器，通过两种互补的转移更新指示变量：单点Gibbs更新，以及在连通区域内反转一条边的分块 Metropolis–Hastings移动。其余参数则从各自的全条件分布中抽取（第 S3 节）。

最大分量反转。 单点更新只能缓慢地改变连通区域内一条边的方向。当某个顶点的关联三角形中的任一顶点上当前方向处于激活状态时，反向就无法在该顶点处激活（图2B）。因此，逐个顶点地反转某个区域内的一条边，会经过一些配置：该边在区域的部分范围内缺失。难点在于，当数据支持该区域内始终存在某种有向关系时，这些中间配置的后验概率可能远低于支持任一方向的配置。

下面的阻塞式移动绕过了这一困难：它作用于节点指示变量，在一次提议中反转整个连通分量上的方向。两个网格

顶点若共享一个三角形，则彼此相邻。从一个方向 e 处于激活状态的顶点开始，令 C 为包含该顶点的 e激活顶点的极大连通分量。由于$\gamma \in \mathcal{A}$， $\bar{e}$ 在 C的每个顶点处都处于非激活状态，因此我们提议

$$(\gamma_{eh}, \gamma_{\bar{e}h}) = (1, 0) \longrightarrow (\gamma'_{eh}, \gamma'_{\bar{e}h}) = (0, 1), \qquad h \in C, \tag{11}$$

保持所有其他指示变量不变。根据 C 的极大性，每个与 C 中某个顶点共享三角形的 e-活跃顶点都属于 C，因此在第2.3.1节的或规则下，该反转不会在 e 和 $\bar{e}$ 之间产生二环。任何在三角形中产生更长有向环的提议都会被拒绝。接受概率对 e 和 $\bar{e}$ 的系数向量进行积分消去；无论结果如何，这些系数向量都会重新抽样，并且该转移相对于我们的后验分布满足细致平衡（命题 S1）。

高斯过程近似。 为了在高斯过程先验下实现可扩展的后验计算，我们通过对其协方差核进行特征分解所得到的低秩基展开来表示每个高斯过程。Mercer 定理给出 $\kappa(s, s') = \sum_{\ell \geq 1} \lambda_\ell \phi_\ell(s) \phi_\ell(s')$，其特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots > 0$，相应的正交归一特征函数为 $\{\phi_\ell\}_{\ell \geq 1}$。 因此，具有核 κ 的零均值高斯过程可表示为 Karhunen–Lo ` eve 展开式 $\sum_{\ell \geq 1} \theta_\ell \sqrt{\lambda_\ell}\, \phi_\ell(s)$，其中系数 θ_ℓ 相互独立且服从高斯分布。仅保留前 L 项，我们将每个场近似为

$$f_0(s) = \sum_{\ell=1}^{L} \theta_{0\ell} \sqrt{\lambda_\ell}\, \phi_\ell(s), \qquad f_e(s) = \sum_{\ell=1}^{L} \theta_{e\ell} \sqrt{\lambda_\ell}\, \phi_\ell(s), \tag{12}$$

且 $\theta_{0\ell}$ 和 $\theta_{e\ell}$ 先验独立且服从高斯分布。因此，对每个场的推断均可简化为对有限维高斯系数向量的推断。在模拟和真实数据分析中，我们均使用修正平方指数核函数

$$\kappa(s, s') = \exp\{-a(\|s\|^2 + \|s'\|^2) - b\|s - s'\|^2\}, \qquad a, b > 0, \tag{13}$$

其 Mercer 展开 $\kappa(s,\ s')\ =\sum_{\ell\geq 1}\lambda_\ell\phi_\ell(s)\phi_\ell(s')$ 具有闭式特征值和特征函数（Lin 等人，2024）。核参数和截断级别 L 均预先固定（S1.1节）。

后验汇总。 由于局部图定义在三角形上，我们在三角形层面汇总后验。对于每个方向 e 和三角形 K，令 $z_{eK}(\gamma^{(m)})\ =\max_{h\in V(K)}\gamma_{eh}^{(m)}$ 表示第 m 个保留后验抽样中的或规则指示变量，$m\ =1$，……，N_{ret}。后验纳入概率（PIP）通过对这些指示变量取平均来估计，

$$\widehat{\rho}_{eK}=\frac{1}{N_{\text{ret}}}\sum_{m=1}^{N_{\text{ret}}}\max_{h\in V(K)}\gamma_{eh}^{(m)}. \tag{14}$$

这些概率构成空间图，展示各个方向的后验支持度。由于没有抽样得到的三角形图同时包含一对方向，$\widehat{\rho}_{eK}+\widehat{\rho}_{\bar{e}K}\leq 1$；它们的和是该方向对在 K 上相连的后验概率，而比较这两个概率则能显示该概率如何分配到两个方向。

4 模拟研究

我们首先考察 SpaDAG 能否恢复方向在定义域内发生反转的局部图。我们在包含 50 个三角形的 $[0,1]^2$ 规则三角剖分上指定真实节点指示变量，并在从定义域中均匀抽取的位置，根据模型 (1)–(3) 生成数据。我们考虑这样的图场：定义域的两个区域中，部分边的方向相反，但每个局部图仍然无环，因此定义域内不存在共同的拓扑排序。我们对系数进行缩放，使总体信噪比的中位数（公式 (S17)）等于 1。以 $n\ =\ 500$ 和 $p\ =\ 50$ 为基准，我们每次改变一个因素，即 $p\in\{25,75\}$ 和 $n\in\{250,1000\}$，每种配置生成 30 个数据集

。每种配置生成 30 个数据集。我们在生成三角剖分上拟合 SpaDAG，使用一条从空图初始化、迭代 50,000 次的 MCMC 链，并根据中位概率模型选择边（即 PIP ≥ 0.5）。我们通过各三角形上所选边的真正率 (TPR)、错误发现率 (FDR) 和 F_1 分数评估恢复效果；只有方向也正确时，边才计为正确（公式 (S13)）。设计与拟合方案见 S4 节。

我们将 SpaDAG 与以下方法进行比较：分别对每个三角形拟合的 Independent NOTEARS（Zheng 等，2018）；通过组惩罚在各三角形间汇集边选择结果的 Joint NOTEARS；在共同拓扑排序下估计三角形图的 JointGES（Wang 等，2020）；以及将空间坐标作为共享神经网络上下文的 ContextualDAG（Thompson 等，2024）。与其作者的模拟设置相同，ContextualDAG 使用真实边数进行调参；我们还报告了一个使用真实位置特定排序的变体。

图3展示了五种配置下各方法的真正例率、假发现率和 F_1 分数。在每种配置中，SpaDAG 的平均 F_1 分数均为最高，同时真正例率最高、假发现率最低。尽管使用了真实边数，表现最强的对比方法 ContextualDAG 在全部 150 个数据集上的 F_1 分数都低于 SpaDAG；使用真实排序的变体在每种配置下的平均 F_1 分数也同样较低。差异主要体现在发现连接方面。SpaDAG 找回了 62–96% 的真实连接（无论方向），而 ContextualDAG 为 32–58%；对于找回的连接，SpaDAG 有 97–98% 判断出了正确方向，而 ContextualDAG 为 80–83%，JointGES 为 53–55%（表 S4）。分量反转移动有助于找回连接：在五个基准数据集上，一个执行相同计算但始终拒绝反转提议的采样器，其平均 F_1 分数低了 0.16（第 S4.3 节）。

另外两项研究考察了这一优势是否取决于方向反转

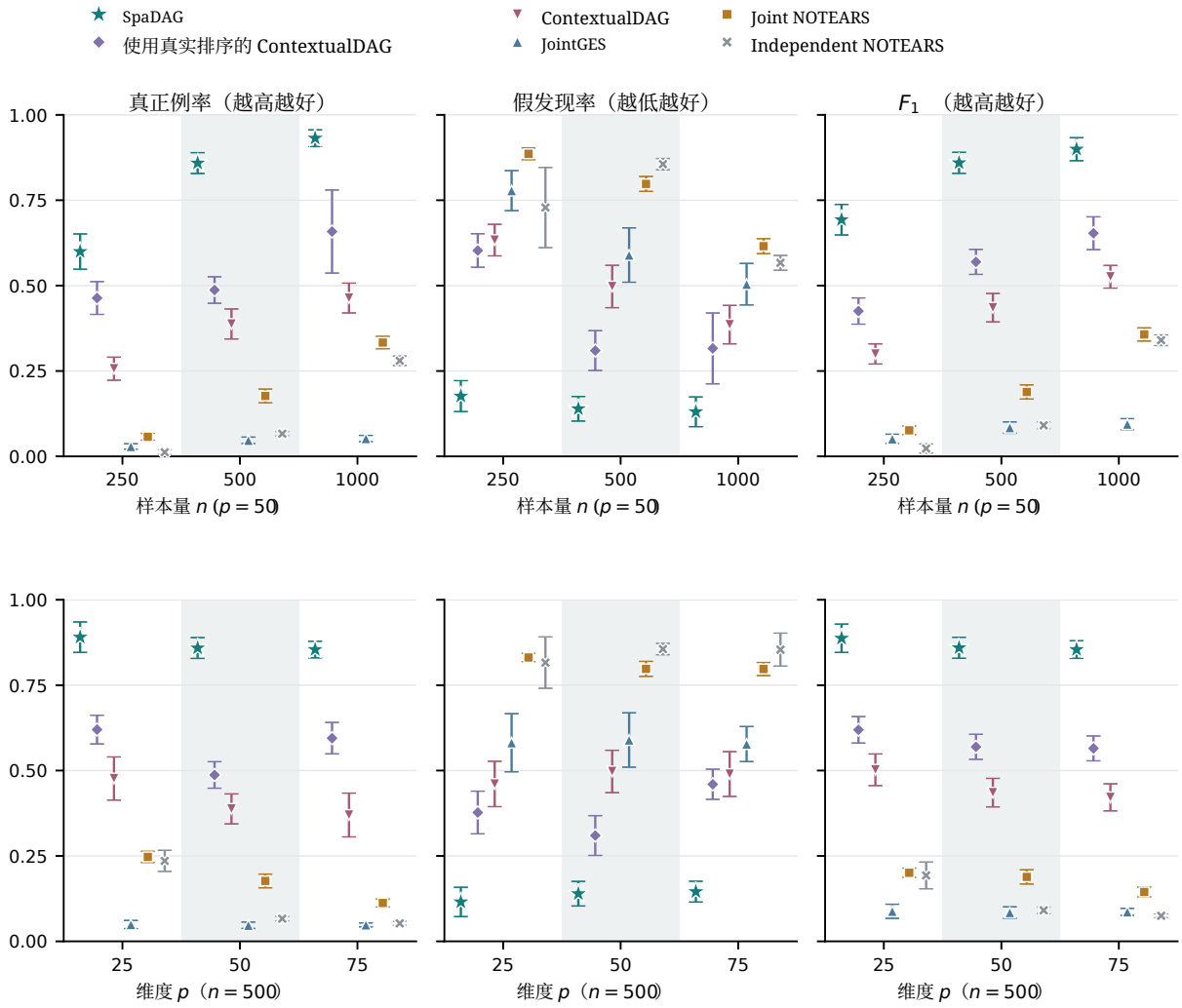

图3：空间方向反转条件下局部有向图的恢复：每种配置下基于30个数据集计算的平均真正例率、假发现率和 F_1 得分，误差条表示数据集间标准差的1倍。上排，n 在 $p = 50$处变化；下排，p 在 $n = 500$处变化。阴影配置 $n = 500$ 和 $p = 50$为两排所共有。

以及当三角剖分依据数据构建时，恢复效果会如何变化。在第一项研究中，我们将每个局部图生成为单个有向无环图的子图，因此所有位置共享相同的拓扑排序，正如 JointGES 所假设的那样。此时，各位置的边仅在是否存在方面有所不同；我们考察了三个层次的空间平滑性，从边存在于连续区域到边存在于分散斑块（S4.4节）。SpaDAG 再次比所有对比方法（包括 JointGES）更准确地恢复了局部图。

第二项研究根据观测到的位置构建三角剖分，就像在空间转录组学数据的实际应用中通常会做的那样，采用与生成三角剖分分辨率相近的受约束德劳内三角剖分（S1.2节）。由此得到的三角形会跨越真实图场的边界，因此，即使是已知图场的预言方法，在这些三角形上的平均 F_1 score 也无法超过 0.894。SpaDAG 的 F_1 score 从 0.86 降至 0.72，但仍显著高于所有对比方法；尽管如此，对每个三角形估计一个图的对比方法获得了真实三角剖分。更粗和更细的三角剖分比较见 S4.5节。

5 透明细胞肾细胞癌空间转录组学分析

现在，我们转而分析第 1 节介绍的透明细胞肾细胞癌数据，考察该通路的调控关系在肿瘤与邻近肾组织之间有何差异。针对晚期透明细胞肾细胞癌的几种疗法，作用于京都基因与基因组百科全书肾细胞癌通路的组成部分：VEGF信号通路，包括 VEGFA；缺氧诱导因子-2α，由 EPAS1 编码；以及 MET，即 HGF 的受体（Choueiri 和 Kaelin，2020；Choueiri 等人，2015，2024）。由于这些疗法针对肿瘤起效，我们主要关注肿瘤组织中存在的关系，而邻近肾组织则显示其中哪些关系是肿瘤特有的。肿瘤组织中存在的关系表明，如果某个

靶基因的表达发生变化，哪些基因会在该组织中作出响应，以及哪些基因会作用于该靶基因。因此，我们要探究哪些基因在肿瘤富集组织的局部图中连接度较高，以及它们在该组织之外是否也有较高的连接度（第 5.2 节）；还要探究哪些关系仅存在于肿瘤富集组织中（第 5.3 节）。鉴于靶基因发生变化时，只有其下游基因会受到影响，我们还要考察基因对的方向是否会随组织环境而变化（第 5.4 节）。

5.1 预处理、组织背景与推断

我们分析了京都基因与基因组百科全书肾细胞癌通路中的49个基因，这些基因在质量控制前的2,048个空间点中至少有20%被检测到；分析对象为质量控制后保留的1,885个空间点。计数被转换为解析皮尔逊残差（Lause等人，2021年），以校正文库大小差异（即一个空间点中唯一分子标识符的总计数），并对每个基因的残差进行标准化。因此，下文的因果陈述涉及经文库大小校正的转录本丰度，系数的单位为标准差。我们将分析网格构建为受约束德劳内三角剖分，其中包含85个三角形，每个三角形的空间点数中位数为21（S5.1节和S1.2节）。

根据 CARD 比例（见图 1A），我们将每个空间点标记为肿瘤富集、上皮富集或混合/其他；上皮富集点位主要由邻近肾脏的肾小管上皮构成，而非肿瘤细胞。每个三角形采用其所含空间点的多数标签。图4A展示了苏木精-伊红图像上的三角剖分，图4B展示了空间点的标签，图4C展示了三角形的标签：肿瘤富集三角形形成一个连续区域，而上皮富集三角形和混合三角形则位于界面的另一侧。下文所述肿瘤富集组织以外的组织，包括所有非肿瘤富集三角形，其中大多数为上皮富集或混合/其他。标签的定义及其对阈值的敏感性见第 S5.1 节

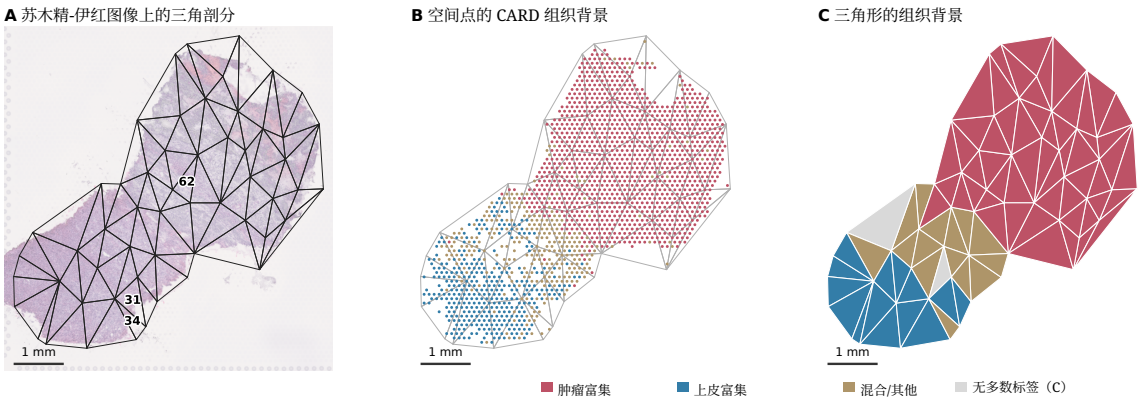

图4：透明细胞肾细胞癌切片的组织背景。（A）H&E图像上的三角剖分，三角形编号用于补充材料。（B）1,885个保留空间点的CARD组织背景。（C）每个三角形的组织背景，即其空间点的多数标签。

及S5.5。

我们运行了第3节提出的马尔可夫链蒙特卡罗采样器，共迭代 50,000 次，丢弃前 25,000 次作为预热期，此后每隔五次保留一次抽样（第 S5.2 节）。我们通过运行四条独立链来评估收敛性，其中两条以空图初始化，另外两条以随机图初始化；结果见第 S5.3 节。根据第3节中的纳入概率（PIPs），我们在贝叶斯错误发现率（BFDR；Ha 等人，2021年）控制在 0.1 的条件下筛选局部边（第 S3.5 节），最终得到 78 个有向关系。

图5A和B按组织背景汇总了选出的关系。面板A展示了至少在一个肿瘤富集三角形中选出的45个关系，面板B则展示了至少在一个上皮富集三角形中选出的30个关系。这两个图仅共享两个关系：HIF1A→PDGFB 和 MAP2K1→SOS1，因此，这两种组织环境支持的有向关系几乎完全不同。就每个三角形而言，肿瘤富集三角形和上皮富集三角形中选出的关系数量都约为5个。

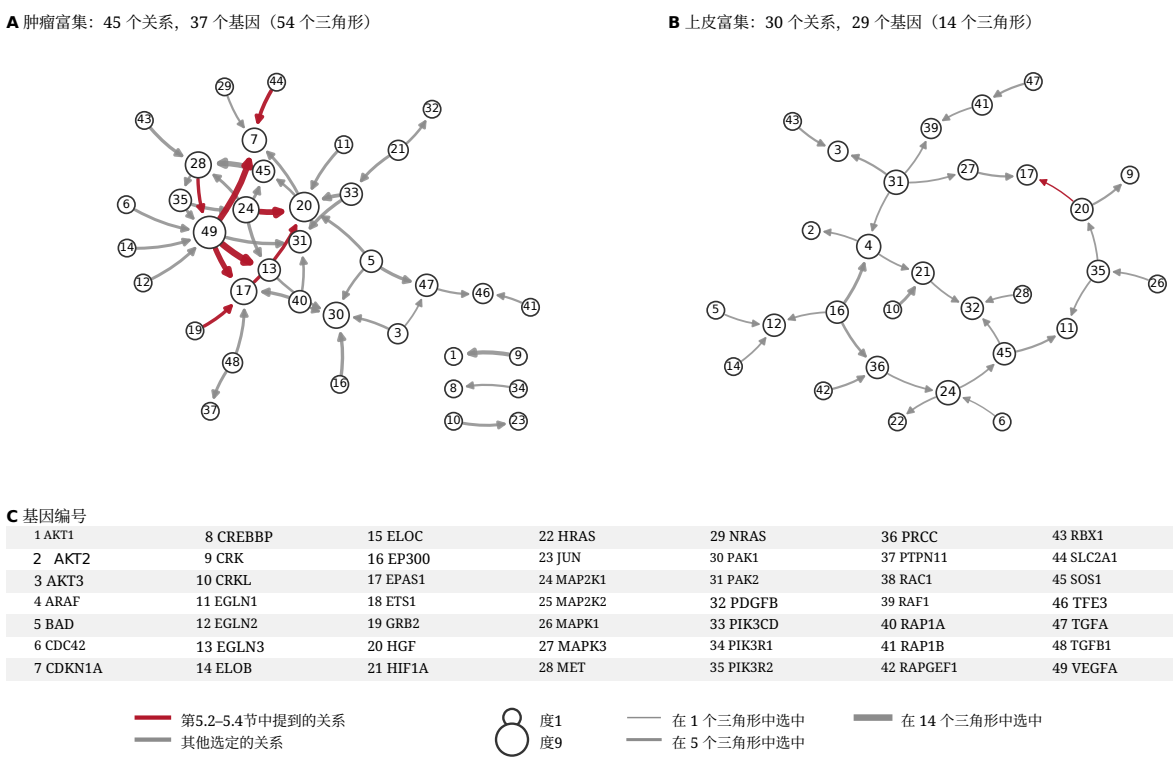

图5：按组织背景展示在 BFDR-0.1 截止值下选出的关系。（A、B）分别展示至少在一个肿瘤富集三角形和至少在一个上皮富集三角形中选出的关系；节点大小随基因的度增大，边宽则随选出该关系的三角形数量增加。红色边表示第5.2–5.4节中提及的关系。每个图汇集了在多个三角形中选出的关系，因此未必是有向无环图，因为无环性仅在每个三角形内施加。混合/其他三角形中选出的关系见图S3。（C）基因编号。

5.2 肿瘤富集组织中的高度连接的基因

图5A显示，在肿瘤富集组织中，VEGFA、HGF、EPAS1和MET属于具有最多入选关系的基因。这些基因与三个治疗靶点相关。在肿瘤富集组织中选定的关系约有一半涉及这四个基因中的一个，而在该组织之外约为六分之一，这与它们在所有候选关系中的占比相同（S5.4节）。在肿瘤富集组织之外，没有哪个基因以这种方式脱颖而出；选定的关系分散在许多基因之间。

图5中的图结合了某一组织背景下所有三角形中选定的关系。在这样的图中具有许多关系的基因，不一定在任何一个三角形中都有许多关系：它可能在一些三角形中有一个子节点，而在另一些三角形中有不同的子节点。为了考察一个基因在单个局部有向无环图中的连接程度，我们统计每个三角形中每个基因的选定子节点数，并将其在拥有多个子节点的三角形中称为枢纽（S5.4节）；枢纽的表达发生变化会同时影响多个基因。图6A和B显示了整个切片中VEGFA和EP300的这一计数结果，后者编码转录共激活因子p300。这些基因作为枢纽覆盖的区域最广：VEGFA仅在肿瘤富集组织中，而EP300仅在该组织之外。在肿瘤富集组织中，VEGFA作用于EPAS1、EGLN3和CDKN1A。VEGFA及其两个子节点是HIF应答基因：HIF诱导VEGFA、EGLN3，并且在透明细胞肾细胞癌中诱导EPAS1本身（del Peso等，2003年；Carroll和Ashcroft，2006；Naas等，2026年）。因此，肿瘤富集组织的枢纽连接了HIF应答基因，其中包括与两个治疗靶点相关的基因，即HIF-2α 和VEGF信号通路。

一个基因也可能作为子节点高度连接，成为调控输入汇聚的点。对于一个靶基因而言，多个父节点意味着不止一个上游基因能够改变其表达。因此，我们还统计了每个三角形中每个基因所选定的父节点数量。在肿瘤富集组织中，在最广泛区域内具有两个或更多父节点的基因

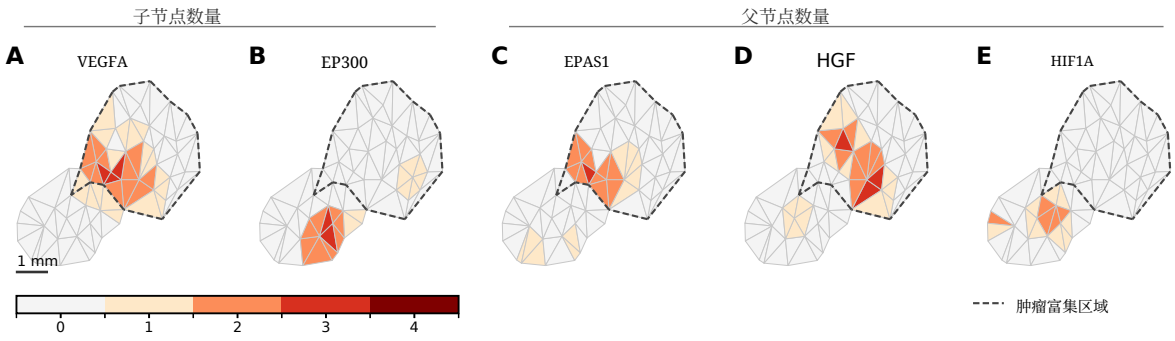

图6：在 BFDR-0.1 截止值下，每个三角形中筛选出的子节点和父节点数量。（A、B）每个三角形中 VEGFA 和 EP300 的筛选子节点数量。（C–E）EPAS1、HGF 和 HIF1A 的筛选父节点数量。虚线圈出了肿瘤富集三角形；同一三角剖分中的所有组织环境见图 4C。

是 HGF 和 EPAS1。图6C和D展示了每个三角形中 EPAS1 和 HGF 所选定的父节点数量。HGF 的父节点包括 MAP2K1（MAPK 级联中的一个基因），而 EPAS1 的父节点包括衔接蛋白基因 GRB2；这两个基因都参与生长因子信号传导，并具有正向估计效应。EPAS1 本身是 HGF 的一个父节点，这种关系将在第 5.4 节中进行考察。因此，在肿瘤富集组织中，通路中的多个基因在同一位置作用于 HGF 和 EPAS1。

在肿瘤富集组织之外的任何三角形中，HGF 和 EPAS1 的父节点数量都不超过一个。如图6E所示，在肿瘤富集组织中，在最广区域拥有多个父节点的基因之一是 HIF1A，它编码缺氧诱导因子-1α （S5.4节）。在肾小管上皮细胞中，缺氧（低氧）会诱导 HIF-1α 积累，而 VHL 失活则会促使 HIF-2α 表达（Schietke 等，2012）。HIF-2α 支持 VHL 缺陷型透明细胞肾细胞癌模型的生长（Kondo 等，2003）。在我们的分析中，编码这两个亚基的基因在界面的两侧各自有多个被选中的父节点：EPAS1 位于

肿瘤富集组织中，HIF1A 位于其外。

5.3 仅存在于肿瘤富集组织中的关系

在仅于肿瘤富集组织中选出的关系中，我们介绍其中两种：一种连接两个治疗靶点，另一种位于缺氧诱导因子-2α的下游。第一种是 MET→VEGFA。在肿瘤富集区域中，MET 是 VEGFA 的父节点，而在肿瘤富集组织之外则未选出该关系。该关系在六个三角形中被选出（图 S4A）；在这六个三角形中，MET 每增加一个标准差，VEGFA 均会增加 0.33 至 0.41 个标准差（对每个三角形的后验均值系数取平均）。这一方向与 ACHN 肾癌细胞中过表达野生型 MET 后 VEGF-A 分泌增加的现象一致（Miyata 等，2003）。

第二组是 SLC2A1→CDKN1A。SLC2A1 编码葡萄糖转运蛋白 GLUT1，而缺氧诱导因子-2α 会在 VHL 缺陷型肾癌细胞中诱导其表达（Raval 等，2005）；CDKN1A 则编码细胞周期抑制剂 p21。在肿瘤富集区域内，SpaDAG 以接近 1 的后验概率支持 SLC2A1→CDKN1A；而在肿瘤富集组织之外，它不支持这两个基因之间的任一方向关系。该关系在 7 个三角区中被选中，且全部为肿瘤富集型（图 S4B）。在这 7 个三角区中，估计的直接效应均为正：SLC2A1 每增加 1 个标准差，CDKN1A 就增加约 0.2 个标准差（按各三角区平均的后验均值系数为 0.19 至 0.21）。两个基因都依赖缺氧诱导因子-2α 会产生相反的符号，因为缺氧诱导因子-2α 会诱导 SLC2A1，却会降低肾癌细胞和透明细胞肾细胞癌肿瘤中的 p21 水平（Gordan 等，2007、2008）；因此，仅凭这种共同依赖关系无法解释正系数。SLC2A1 指向 CDKN1A 的方向，与肾小管细胞中葡萄糖摄取后 p21 水平升高（Kitada 等，2014），以及大鼠肾系膜细胞中 GLUT1 过表达的结果（Zheng 等，2008）相符。

5.4 EPAS1–HGF 的空间变化方向

最明显的空间方向变化涉及 EPAS1 和 HGF；这是唯一一个在两个方向上都被选中的基因对。SpaDAG 在肿瘤富集区域内估计出 EPAS1→HGF，而在一个上皮富集三角区和一个混合三角区中估计出反向关系 HGF→EPAS1。图7A和B展示了全部 85个三角区中 EPAS1→HGF 和 HGF→EPAS1 的局部 PIP，并用轮廓标出了各方向关系被选中的三角区。图7C和D显示这两项估计的符号也不同：在 6 个肿瘤富集三角区中，EPAS1 每增加 1 个标准差，HGF 就降低 0.32 至 0.36 个标准差；而在肿瘤富集组织之外的两个三角区中，HGF 每增加 1 个标准差，EPAS1 就增加 0.36 和 0.38 个标准差（按各三角区平均的后验均值系数）。

后验概率还显示了每个方向获得支持的确定程度。在六个肿瘤富集三角形中，EPAS1→HGF 的后验概率接近 1。在肿瘤富集组织之外的两个三角形中，HGF→EPAS1 的后验概率约为 0.76。对整个组织施加单一排序的模型只能表示这两个方向中的一个。

EPAS1和HGF连接了两个治疗靶点：HIF-2α 和MET信号通路。研究在人内皮细胞系中观察到HGF与HIF-2α 之间存在功能耦联：敲低HIF-2α 会减少HGF诱导的IL-8分泌（Makarevich等人，2018）。估计得到的关系涉及它们的转录本，并提出了新的假设，探讨HIF-2α 与MET信号通路在各组织背景中如何耦联。

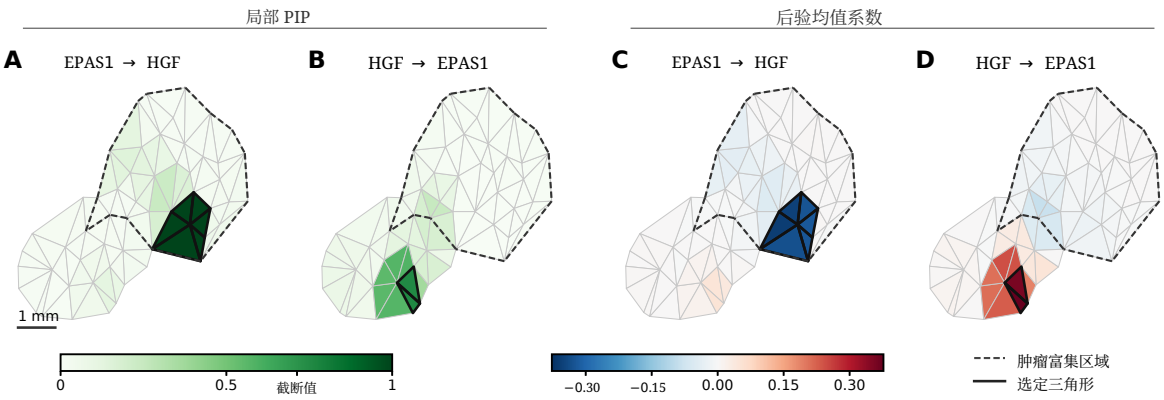

图7：EPAS1–HGF 的局部方向相反。（A、B）所有 85个三角区中各方向的局部 PIP。（C、D）各方向的后验均值系数，按每个三角形取平均。实线轮廓标示在 BFDR-0.1 截止值下选定的三角形，虚线圈出肿瘤富集三角形。

5.5 坐标置换对照

上述局部模式可能反映了表达在组织中的分布位置，也可能仅由空间先验和选择规则产生。为了区分这两种情况，我们随机置换斑点坐标一次后重新拟合SpaDAG。此操作保留了每个表达谱，但将这些表达谱随机分配到不同位置，因此每个区域的表达谱都来自同一组汇总的空间点（第S5.5节）。

去掉空间排列后，拟合图失去了局部结构。在使用观测坐标选出的 78 个关系中，有 68 个关系（包括 EPAS1–HGF 的两个方向）在置换后没有在任何三角形中被选中（见第 S5.5 节和表 S12）。三角形之间后验概率的变化也基本消失，而两个基因在整个切片中共同变化的关系（如 VEGFA→EGLN3）几乎处处都获得了支持（图 S5）。第 5.2–

5.4 节的局部模式因此取决于表达的空间排列。

6 讨论

本文提出了 SpaDAG，该方法能够推断组织切片每个位置的因果有向无环图，并给出每条边存在及其方向的后验概率。将 SpaDAG 应用于一张透明细胞肾细胞癌切片后，我们发现，肿瘤富集组织与上皮富集组织几乎完全支持不同的有向关系；此外，在一个肿瘤富集区域与其外部的两个三角形区域之间，EPAS1 与 HGF 的估计方向也存在差异。这些关系是转录本水平上的假设，可在各自的组织背景中通过蛋白质水平实验进行检验。

我们的解释基于位置层面的因果充分性（第2节 2.1）。未测量的调节因子（例如在蛋白质水平上发挥作用的调节因子）可能是建模基因的共同原因；例如，VEGFA→EGLN3 可能反映了 HIF 对这两个基因的共同调控。在本文所分析的这类基于点位的数据中，一个空间点内的细胞类型混合也可能发挥相同作用。我们的可识别性结果假设误差方差在空间上保持不变（第2节 2.2）。在这类数据中，不同空间点之间的测序深度和细胞组成存在差异，因此噪声水平可能在组织内部发生变化。

SpaDAG 在采样前固定三角剖分；若在采样器中更新三角剖分，三角形的边界便能随局部图的变化而调整。还可以扩展采样器，加入可在连通区域内添加或删除边的操作。尽管 SpaDAG 的设计初衷是用于空间转录组学，但在空间结构方程模型的假设适用时，它也可用于其他空间分辨多变量测量，例如空间蛋白质组学。

参考文献

Acharyya, S., Kang, J. 和 Baladandayuthapani, V.（2025），《空间转录组癌症数据的空间变化图估计》，bioRxiv 预印本。Aibar, S., Gonz´alez-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G. 等（2017），《SCENIC：单细胞调控网络推断与聚类》，*Nature Methods***14(11)**, 1083–1086。Carroll, V. A. 和 Ashcroft, M.（2006），《缺氧诱导因子（HIF）-1α 与 HIF-2α 在缺氧、胰岛素样生长因子-I 或冯·希佩尔-林道功能缺失时调节 HIF 靶基因中的作用：对靶向 HIF 通路的启示》，*Cancer Research* **66(12)**, 6264–6270。Castelletti, F., La Rocca, L., Peluso, S., Stingo, F. C. 和 Consonni, G.（2020），《基于观察性数据的多个有向网络的贝叶斯学习》，*Statistics in Medicine***39(30)**, 4745–4766。Choueiri, T. K., Escudier, B., Powles, T., Mainwaring, P. N., Rini, B. I., Donskov, F. 等（2015），《卡博替尼对比依维莫司治疗晚期肾细胞癌》，*New EnglandJournal of Medicine***373(19)**, 1814–1823。Choueiri, T. K. 和 Kaelin, Jr., W. G.（2020），《靶向肾细胞癌中的 HIF2–VEGF轴》，*Nature Medicine* **26(10)**, 1519–1530。Choueiri, T. K., Powles, T., Peltola, K., de Velasco, G., Burotto, M., Suarez, C. 等（2024），《贝组替凡对比依维莫司治疗晚期肾细胞癌》，*New EnglandJournal of Medicine***391(8)**, 710–721。Dawn, T. 和 Ni, Y.（2025），《基于贝叶斯非参数协变量依赖有向循环图模型的空间变化基因调控网络》，arXiv 预印本 arXiv:2512.11732。

del Peso, L., Castellanos, M. C., Temes, E., Mart´ın-Puig, S., Cuevas, Y., Olmos, G. 等（2003），《希佩尔-林道/缺氧诱导因子（HIF）通路调控低氧条件下HIF-脯氨酰羟化酶基因的转录》，《生物化学杂志》**278**(49), 48690–48695。Gordan, J. D., Bertout, J. A., Hu, C.-J., Diehl, J. A. 和 Simon, M. C.（2007），《HIF-2α通过增强c-Myc转录活性促进缺氧细胞增殖》，*CancerCell***11**(4), 335–347。Gordan, J. D., Lal, P., Dondeti, V. R., Letrero, R., Parekh, K. N., Oquendo, C. E. 等（2008），《HIF-α 对c-Myc的影响区分了两种散发性VHL缺陷型透明细胞肾癌亚型》，*Cancer Cell* **14**(6), 435–446。Ha, M. J., Stingo, F. C. 和 Baladandayuthapani, V.（2021），《多层基因组网络中的贝叶斯结构学习》，《美国统计学会杂志》**116**(534), 605–618。Hoyer, P. O., Janzing, D., Mooij, J. M., Peters, J. 和 Sch¨olkopf, B.（2009），《基于加性噪声模型的非线性因果发现》，见《Advances in Neural Information Processing Systems 21》，第689–696页。Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M. 等（2017），《肾细胞癌》，《自然综述：疾病导论》**3**, 17009。Kaelin, Jr., W. G.（2008），《希佩尔-林道肿瘤抑制蛋白：氧$_2$ 感知与癌症》，《自然综述：癌症》**8**(11), 865–873。Kanehisa, M. 和 Goto, S.（2000），《KEGG：京都基因与基因组百科全书》，《核酸研究》**28**(1), 27–30。Kitada, K., Nakano, D., Ohsaki, H., Hitomi, H., Minamino, T., Yatabe, J. 等（2014），《高血糖在糖尿病肾病早期通过依赖SGLT2和p21的通路导致近端小管细胞衰老》，《糖尿病及其

并发症》**28**(5), 604–611。Kondo, K., Kim, W. Y., Lechpammer, M. 和 Kaelin, Jr., W. G.（2003），《抑制HIF2α 足以抑制pVHL缺陷型肿瘤生长》，*PLoS Biology* **1**(3), e83。Lauritzen, S. L.（1996），图模型，第17卷，牛津统计科学丛书，牛津大学出版社，牛津。Lause, J., Berens, P. 和 Kobak, D.（2021），《用于单细胞RNA测序UMI数据归一化的解析皮尔逊残差》，*Genome Biology* **22**, 258。Li, R., Ferdinand, J.R., Loudon, K.W., Bowyer, G.S., Laidlaw, S., Muyas, F.等（2022），《绘制肾癌肿瘤内及相关区域的单细胞转录组》，*Cancer Cell* **40**(12), 1583–1599.e10。Li, Y., Chen, J., Lu, T., Tsai, N.-P. 和 Wang, H.（2026），《基于空间转录组学的空间变异基因调控网络推断》，*Bioinformatics Advances* **6**(1), vbag230。Lin, Z., Si, Y. 和 Kang, J.（2024），《图像对标量回归中的潜在亚组识别》，*The Annals of Applied Statistics* **18**(1), 468–486。Ma, Y. 和 Zhou, X.（2022），《面向空间转录组学的空间信息指导细胞类型反卷积》，*Nature Biotechnology* **40**(9), 1349–1359。Maathuis, M. H., Colombo, D., Kalisch, M. 和 B¨uhlmann, P.（2010），《根据观察性数据预测大规模系统中的因果效应》，*Nature Methods* **7**(4), 247–248。Madigan, D. 和 York, J.（1995），《离散数据的贝叶斯图模型》，*International Statistical Review***63**(2), 215–232。Makarevich, P. I., Dergilev, K. V., Tsokolaeva, Z. I., Boldyreva, M. A., Shevchenko, E. K., Gluhanyuk, E. V. 等（2018），《VEGF165和HGF联合基因治疗在心肌梗死大鼠模型中的促血管生成及多效性效应》，*PLoS ONE***13**(5), e0197566。Miyata, Y., Ashida, S., Nakamura, T., Mochizuki, Y., Koga, S., Kanetake, H. 等（2003），

《肝细胞生长因子受体在肾癌细胞中的过表达间接刺激体内肿瘤生长》，《生物化学与生物物理研究通讯》**302**(4), 892–897. Monteiro, R., Pouget, C. 和 Patient, R. (2011), ‘$gata1/pu.1$ 谱系命运范式在不同血液细胞群之间存在差异，并受$tif1\gamma$调节’，《欧洲分子生物学组织杂志》**30**(6), 1093–1103. Naas, S., Kr¨uger, R., Grampp, S., Lauer, V., Kraus, A., Naas, J 等 (2026), ‘HIF维持肾透明细胞癌中EPAS1表达的转录调控回路’，《自然-通讯》**17**(1), 1764. Ni, Y., Stingo, F. C. 和 Baladandayuthapani, V. (2019), ‘贝叶斯图回归’，《美国统计学会杂志》**114**(525), 184–197. Nickerson, M. L., Jaeger, E., Shi, Y., Durocher, J. A., Mahurkar, S., Zaridze, D. 等 (2008), ‘改进冯·希佩尔-林道基因在肾透明细胞肿瘤中的变异识别’，《临床癌症研究》**14**(15), 4726–4734. Oates, C. J., Smith, J. Q., Mukherjee, S. 和 Cussens, J. (2016), ‘多个有向无环图的精确估计’，《统计与计算》**26**(4), 797–811. Pearl, J. (2009),《因果关系：模型、推理与推断》，第2版，剑桥大学出版社，剑桥。Rao, A., Barkley, D., Fran¸ca, G. S. 和 Yanai, I. (2021), ‘利用空间转录组学探索组织结构’，《自然》**596**(7871), 211–220. Raval, R. R., Lau, K. W., Tran, M. G. B., Sowter, H. M., Mandriota, S. J., Li, J.-L. 等 (2005), ‘缺氧诱导因子1（HIF-1）与HIF-2在冯·希佩尔-林道相关肾细胞癌中的不同特性’，《分子与细胞生物学》**25**(13), 5675–5686. Sangalli, L. M., Ramsay, J. O. 和 Ramsay, T. O. (2013), ‘空间样条回归模型’，

英国皇家统计学会杂志：*B*辑（统计方法学）**75**(4), 681–703. Schietke, R. E., Hackenbeck, T., Tran, M., G¨unther, R., Klanke, B., Warnecke, C. L. 等 (2012), ‘肾小管HIF-2α 表达需要VHL 失活，并导致纤维化和囊肿’，*PLoS ONE* **7**(1), e31034. Shang, L., Wu, P. 和 Zhou, X. (2025), ‘空间转录组学中细胞类型特异性空间变异基因的统计识别’，《自然-通讯》**16**, 1059. Shimizu, S., Hoyer, P. O., Hyv¨arinen, A. 和 Kerminen, A. (2006), ‘用于因果发现的线性非高斯无环模型’，《机器学习研究杂志》**7**, 2003–2030. Spirtes, P., Glymour, C. 和 Scheines, R. (2000),《因果、预测与搜索》，第2版，麻省理工学院出版社，马萨诸塞州剑桥。Thompson, R., Bonilla, E. V. 和 Kohn, R. (2024), 上下文有向无环图，载于‘第二十七届人工智能与统计国际会议论文集’,《机器学习研究论文集》第238卷，PMLR，第2872–2880页。Verma, T. 和 Pearl, J. (1990), 因果模型的等价性与综合，载于‘第六届人工智能不确定性大会论文集’，第220–227页。Wang, Y., Segarra, S. 和 Uhler, C. (2020), ‘多个高维有向高斯图模型的联合估计’，《电子统计学杂志》14(1), 2439–2483. Zheng, J.-M., Zhu, J.-M., Li, L.-S. 和 Liu, Z.-H. (2008), ‘大黄酸通过抑制己糖胺通路，逆转过表达葡萄糖转运蛋白（GLUT1）的系膜细胞糖尿病表型’，《英国药理学杂志》**153**(7), 1456–1464. Zheng, X., Aragam, B., Ravikumar, P. K. 和 Xing, E. P. (2018), DAGs with NO TEARS：用于结构学习的连续优化，载于‘神经信息处理系统进展’，第31卷。Zhou, F., He, K. 和 Ni, Y. (2023), ‘基于潜在轨迹的个体化因果发现

嵌入式贝叶斯网络’， *Biometrics***79(4)**, 3191–3202。