整合组织学图像的乳腺癌空间转录组数据空间域识别*

崔丁童[1] 田雅昕[1,2] 王 彤[1,2] 刘改琴[1] 房瑞玲[1,3] 曾 平[4]
罗艳虹[1,3] 余红梅[1,3] 张岩波[1,3] 曹红艳[1,3Δ]

【提 要】 目的 整合空间转录组数据的组织学图像、基因表达和空间位置信息，识别乳腺癌患者组织切片的不同空间域，筛选空间可变基因（spatially variable genes，SVGs）和空间可变元基因（spatially variable meta genes，meta SVGs）。**方法** 从10× Genomics平台下载有标签的乳腺癌A区第1段数据和无标签的浸润性乳腺导管癌空间转录组学数据，通过设置不同的组织学权重，利用SpaGCN整合组织学图像RGB三通道信息、基因表达数据及二维空间位置坐标，采用Louvain算法进行空间域识别，并与BayesSpace方法进行比较。基于较高组织学权重下识别的空间域结构，筛选每个空间域的SVGs，对于未能识别出SVGs的空间域，进行meta SVGs的检测。进一步利用DAVID数据库对识别出的SVGs分别进行基因本体（gene ontology，GO）功能富集分析及基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）通路富集分析。**结果** SpaGCN识别出有标签的乳腺癌A区第1段数据的20个空间域以及332个SVGs，无标签的浸润性乳腺导管癌数据的9个空间域和439个SVGs。进一步对筛选出的SVGs分别进行GO和KEGG富集分析，分别得到200个GO条目、28个KEGG通路，以及367个GO条目、49个KEGG通路。对乳腺癌A区第1段数据中的10个未识别出SVGs的空间域，检测出15个meta SVGs。**结论** 在乳腺癌空间转录组数据分析中，SpaGCN整合组织学图像、基因表达和空间位置信息，实现了空间域的精准识别。筛选出的SVGs、meta SVGs和富集通路可以为乳腺癌潜在靶点的识别和个性化治疗策略的制定提供重要依据。

【关键词】 空间转录组学 空间域 组织学图像 SpaGCN 乳腺癌
【中图分类号】 Q78；R737.9 **【文献标志码】** A **DOI** 10.11783/j.issn.1002-3674.2026.05.002

Spatial Domain Recognition of Breast Cancer Spatial Transcriptomics Data Integrated with Histological Images

Cui Dingtong, Tian Yaxin, Wang Tong, et al (*Division of Health Statistics, Shanxi Provincial Key Laboratory of Major Diseases Risk Assessment, Shanxi Medical University, Taiyuan* 030001)

【Abstract】 Objective Integrating histological images, gene expression, and spatial location information from spatial transcriptomics data to identify distinct spatial domains in tissue sections of breast cancer patients, and to screen for spatially variable genes (SVGs) and spatially variable meta genes (meta SVGs). **Methods** Data from the 10× Genomics platform, including labeled breast cancer Block A Section 1 data and unlabeled invasive ductal carcinoma spatial transcriptomics data were downloaded. By assigning different histological weights, SpaGCN was used to integrate RGB three-channel histological image data, gene expression data, and 2D spatial coordinates. The Louvain algorithm was applied for spatial domain identification, and the results were compared with the BayesSpace method. Based on the spatial domain structures identified under higher histological weight, SVGs for each spatial domain were selected. For spatial domains where SVGs could not be identified, the expression fold-change threshold between the target domain and adjacent domains was lowered to detect meta-SVGs. Furthermore, the identified SVGs were subjected to gene ontology (GO) functional enrichment analysis and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis using the DAVID database. **Results** SpaGCN identified 20 spatial domains and 332 SVGs in the labeled breast cancer Block A Section 1 data, and 9 spatial domains and 439 SVGs in the unlabeled invasive ductal carcinoma data. Further GO and KEGG enrichment analyses were conducted on the identified SVGs, resulting in 200 GO terms and 28 KEGG pathways for the labeled data, and 367 GO terms and 49 KEGG pathways for the unlabeled data. For the 10 spatial domains in the breast cancer Block A Section 1 data where SVGs were not detected, 15 meta-SVGs were detected. **Conclusion** In the analysis of breast cancer spatial transcriptomics data, SpaGCN integrates histological images, gene expression, and spatial location information to achieve precise identification of spatial domains. The identified SVGs, meta SVGs, and enriched pathways provide valuable insights for the identification of potential targets in breast cancer and the development of personalized treatment strategies.

【Key words】 Spatial transcriptomics; Spatial domain; Histological image; SpaGCN; Breast cancer

* 基金项目：国家自然科学基金（82473739；82273742）；山西省基础研究计划（202303021211130）；山西省回国留学人员科研资助项目（2024-081）；山西省高等教育“百亿工程”科技引导专项；煤炭环境致病与防治教育部重点实验室（山西医科大学）开放课题基金（MEKLCEPP/SXMU-202415）
1. 山西医科大学卫生统计教研室，重大疾病风险评估山西省重点实验室（030001）
2. 山西医科大学医学科学院
3. 煤炭环境致病与防治教育部重点实验室
4. 徐州医科大学公共卫生学院流行病与卫生统计学系
△通信作者：曹红艳，E-mail：caohy@sxmu. edu. cn

乳腺癌是全球女性癌症发病率最高的癌症，据全球癌症统计报告，2022年全球乳腺癌新发病例达到约230万例，其中死亡病例约为70万例[1]。在所有女性恶性肿瘤中，乳腺癌发病例数位居第二[2]，且其所导致的疾病负担呈现逐年上升的趋势[3]。肿瘤异质性（intratumoral heterogeneity，ITH）是致死、治疗失败以及耐药的一个关键因素[4]。ITH包括时间和空间异质性，涵盖基因、细胞和代谢等多个层面的异质性[5]。因此，准确评估肿瘤组织的异质性对于开发有效治疗策略、识别潜在药物靶点及改善预后至关重要。

空间转录组（spatial transcriptomics，ST）技术通过结合传统的转录组分析与组织学切片图像，揭示肿瘤微环境中不同细胞群体的基因表达特征及其空间分布，进而在基因、细胞和代谢等多个层面提供癌症发生机制和病理信息[6]。精准识别乳腺癌空间转录组数据的不同空间域，有效鉴定乳腺癌的早期分子标志物和高危区域，对于提高早期筛查的敏感性和特异性具有重要意义[7]。

现有空间域识别方法如BayesSpace[8]、STAGATE[9]、BASS[10]、DR-SC[11]等仅依赖基因表达和空间位置信息来识别空间域，未能考虑组织学图像信息。组织学图像信息通过提供细胞表型的视觉特征，直观展示了细胞的空间分布，为空间域识别提供了独特的形态学信息[12]。因此，整合图像信息以提高空间域识别准确性尤为重要。SpaGCN[13]通过计算组织学图像的RGB像素值，同时结合不同的组织学权重，充分利用组织学图像、基因表达和空间位置信息，增强空间域识别的精度和连续性，从而更加准确地理解组织结构和功能。

本研究采用SpaGCN方法，识别人类乳腺癌A区第1段数据和无标签的浸润性乳腺导管癌空间转录组学数据的空间域，进一步检测具有显著空间表达模式的空间可变基因（spatially variable genes，SVGs）和空间可变元基因（spatially variable meta genes，meta SVGs）以及相应的GO和KEGG富集通路，为理解乳腺癌组织分区、优化治疗策略和探索新型靶点提供重要理论依据。

数据和方法

1. 数据来源与处理

从10× Genomics平台[14]（https://www. 10xgenomics. com/resources/datasets）下载有标签的人类乳腺癌A区第1段数据，以及无标签的浸润性乳腺导管癌空间转录组学数据，其TNM分期均为ⅡA期。每个数据集均包含组织学图像、基因表达矩阵和空间位置矩阵。其中，基因表达数据是一个包含N个位点和P个基因的$N \times P$矩阵，空间位置数据是一个包含M个位点和二维空间坐标的$M \times 2$矩阵。有标签的人类乳腺癌A区第1段数据包含3798个位点和33601个基因，共有20个空间域，标签信息来源于GitHub上Xu等[15]的手动注释信息（https://github. com/JinmiaoChenLab/SEDR_analyses）。无标签的浸润性乳腺导管癌数据包含4727个位点和33601个基因。

数据预处理步骤：删除空间位置数据中未捕获到基因表达的位点，并与基因表达数据进行整合。筛选出至少在3个位点中表达的基因，并去除外源性控制基因和线粒体基因，采用python中的SCANPY库[16]对每个位点的基因表达量进行归一化和对数转换：

$$X_{log}^{i} = \log\left(\frac{X_{ij}}{\sum X_{ij}} \times 10000 + 1\right) \quad (1)$$

其中，X_{ij}是第i个位点的第j个基因的表达量，$X_{\log}^{i}$是位点i经对数转换后的表达量。

2. SpaGCN方法

SpaGCN通过构建无向加权图来表示所有位点之间的关系，其中每个边权值衡量两个位点之间的关联度，在二维空间位置坐标的基础上，通过计算RGB通道中所有像素的平均颜色值，将其扩展至包含组织学信息的三维坐标，由空间二维坐标(x, y)和从组织学图像中的RGB值获得的z^{*}定义，通过增加组织学图像的权重s，可以更清晰地分离病变与非病变区域。然后利用图卷积网络（graph convolutional networks，GCN）聚合邻近点的基因表达信息[17]，采用Louvain算法[18]进行聚类。

（1）构建无向加权图

①整合组织学图像RGB信息

两个位点之间的空间相似度由两个因素决定：a. 位点之间的物理位置；b. 两个位点对应的组织学图像信息。对于点v，其在组织切片中的物理位置用二维坐标(x_v, y_v)表示，根据点v在组织学图像中对应的像素坐标(x_{pv}, y_{pv})，绘制以(x_{pv}, y_{pv})为中心的正方形，其中包含50×50像素，并计算RGB三通道(r_v, g_v, b_v)中所有像素的平均值，用z_v表示组织学图像特征，定义如下：

$$z_v = \frac{r_v \times V_r + g_v \times V_g + b_v \times V_b}{V_r + V_g + V_b} \quad (2)$$

其中，V_r是r_v的方差，V_g是g_v的方差，V_b是b_v的方差。然后将z_v重新缩放为：

$$z_v^{*} = \frac{z_v - \mu_z}{\sigma_z} \times \max(\sigma_x, \sigma_y) \times s \quad (3)$$

其中，μ_z为z_v的均值，$\sigma_x, \sigma_y, \sigma_z$分别为$x_v, y_v, z_v$的标准差，$s$为比例因子，默认设置为1，通过增大$s$的值，可以提高组织学图像的权重。将点$v$在扩展的三维空间中的坐标设置为$(x_v, y_v, z_v^{*})$，则任意两个位点$u, v$之间的欧氏距离计算为：

$$d(u,v)=\sqrt{(x_u-x_v)^2+(y_u-y_v)^2+(z_u^*-z_v^*)^2} \quad (4)$$

②图结构

每条边(u,v)的权重衡量两位点u,v之间的关联度，图结构G存储在$N\times N$邻接矩阵$A=[w(u,v)]$中，两个位点u,v之间的边权定义为：

$$w(u,v)=\exp\left(-\frac{d(u,v)^2}{2l^2}\right) \quad (5)$$

其中，超参数l称为特征长度尺度，决定了权重随着距离变化的衰减速度。采用二分法迭代地寻找最佳参数l，直到找到使得$A-I$矩阵行和的平均值p最接近0.5的l，其中A指邻接矩阵，I指单位矩阵。对于位点v，以a_v表示$A-I$的行和，表示其他位点对位点v的基因表达的相对贡献。二分查找的起始值设置为0.01。容忍误差设置为0.01，即当两次迭代结果之差小于该值时停止迭代。最大迭代次数设置为100，防止无限循环。

（2）构建图卷积网络

利用主成分分析（principal component analysis PCA）[19]对预处理后的基因表达矩阵进行降维，使用前50个主成分矩阵X作为输入，通过卷积层以整合基因表达信息和G中的边权重（包括空间二维坐标和组织学图像RGB信息）：

$$f(X,A)=\delta(A\times B) \quad (6)$$

其中，X为经PCA降维后的基因表达矩阵，A为图结构G的$N\times N$邻接矩阵，B为卷积层过滤参数的50×50矩阵，$\delta(\cdot)$为非线性激活函数，该层的输出是一个包含基因表达、空间位置和组织学信息的聚合矩阵。在图卷积层输出的基础上，采用Louvain算法得到初始的聚类质心，进一步利用随机梯度下降算法迭代地更新和细化聚类。

3. 空间可变基因检测

首先，寻找每个空间域的相邻空间域：在目标域的每个位点周围画一个预先指定半径的圆，来自其他域但位于圆内的位点为其邻居，设置半径使得目标域中的每个点平均约有10个邻居。目标域中所有位点的邻居构成邻近集，如果非目标域中超过50%的位点在邻近集中，则选择该域作为相邻域。

使用Wilcoxon秩和检验在目标域和相邻域的位点之间进行差异表达（differential expression，DE）分析，检测每个空间域的SVGs。基因筛选标准设置如下：①P_{adj}<0.05的基因；②在目标域内表达该基因的位点所占的百分比（in-fraction）大于80%；③目标域与相邻域表达该基因的位点百分比之比（in/out fraction）大于1；④目标域与相邻域的表达倍数大于1.5。

4. 空间可变基因检测

对于在上述条件下未检测到SVGs的空间域，将目标域与相邻域的表达倍数阈值由1.5降为1.2，以识别出目标域的弱SVGs。若存在多个弱SVGs，随机选择其中一个作为基础基因（即$gene_0$），进一步检测meta SVGs，以增强目标域的空间表达模式。检测步骤如下：①计算$gene_0$在目标域所有位点中的平均表达水平（e_0）；②设置对照组为非目标域中基因表达水平高于e_0的位点；③通过Wilcoxon秩和检验对目标域和对照组的位点进行DE分析，选择P_{adj}最小且在目标域表达量较高的基因为$gene_{0+}$；选择P_{adj}最小且在对照组表达量较高的基因为$gene_{0-}$；meta SVGs的标准化表达量计算为：

$$\log(meta_gene_1)=\log(gene_0)+\log(gene_{0+})-\log(gene_{0-})+C_0 \quad (7)$$

其中，C_0是一个常数，使得$\log(meta_gene_1)$非负，SpaGCN通过迭代可以形成其他更具空间表达模式的meta SVGs，第$(t+1)$次迭代的meta SVGs标准化表达量计算为：

$$\log(meta_{gene_{t+1}})=\log(meta_{gene_t})+\log(gene_{t+})-\log(gene_{t-})+C_t \quad (8)$$

5. GO功能富集分析和KEGG通路分析

通过DAVID数据库[20]，对人类乳腺癌A区第1段数据和无标签的浸润性乳腺导管癌空间转录组学数据筛选出的SVGs分别进行GO功能富集及KEGG通路分析。从以下三个层面：分子功能（molecular function，MF）、细胞成分（cellular component，CC）和生物学过程（biological process，BP）进行GO功能富集分析。使用R语言ggplot2[21]包对DAVID在线分析工具所获得的GO富集分析结果进行可视化。

6. 聚类评价指标（adjusted rand index，ARI）

（1）调整兰德指数

针对有标签的人类乳腺癌A区第1段数据，采用调整兰德指数指标计算预测结果与实际标签之间的一致性[22]，ARI值越大，表明聚类结果的一致性越好。

$$ARI=\frac{RI-E(RI)}{\max(RI)-E(RI)} \quad (9)$$

（2）空间混沌评分

针对无标签的浸润性乳腺导管癌数据，为每个空间域构建一阶最近邻（1NN）图，用w_{kij}表示点i和点j之间的边权重，通过下式的空间混沌评分（spatial chaos score，CHAOS）[23]评价聚类效果，CHAOS分数越小，表明空间域的连续性越好。

$$CHAOS=\frac{\sum_{k=1}^{K}\sum_{i,j}^{n_k}w_{kij}}{N} \quad (10)$$

（3）局部逆辛普森指数（local inverse simpson index，LISI）

LISI得分用于衡量聚类异质性，反映了每个位点

周围的局部邻近区域内类别之间的异质性[24]。得分越小表明局部区域内属于同一类别的位点更加集中，即聚类效果越好，用$p(k)$表示第k个类别在局部邻近区域内出现的频率，LISI计算公式如下：

$$S=\frac{1}{\sum_{k=1}^{K}p(k)} \tag{11}$$

（4）异常点百分比（percentage of abnormal spots，PAS）

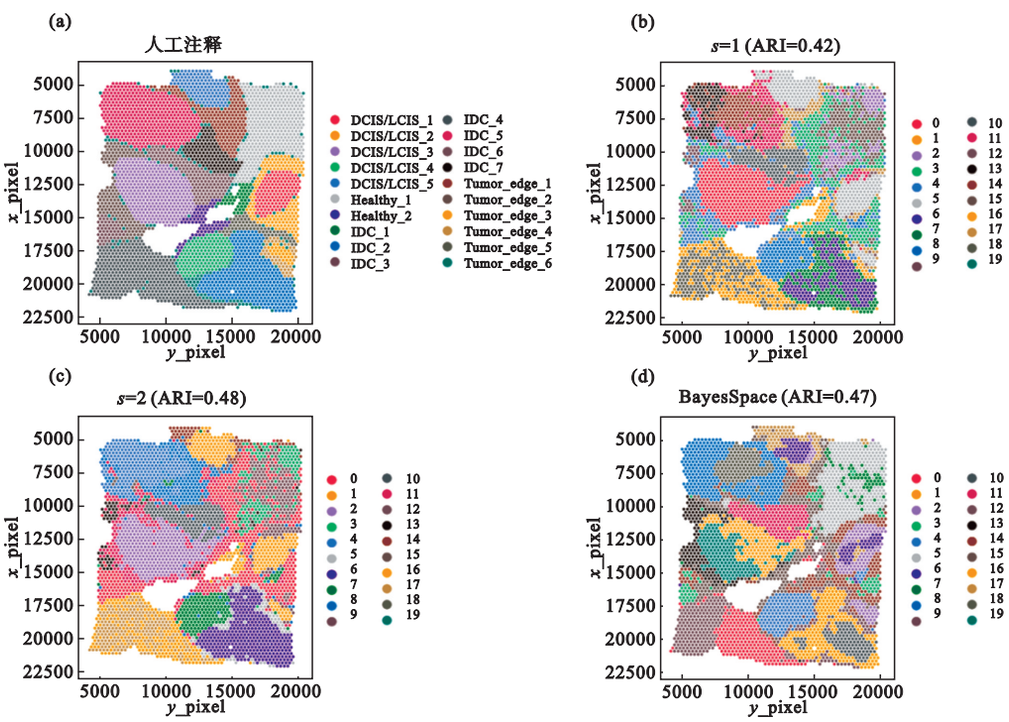
PAS分数衡量聚类中心周围位点的随机性，统计每个位点的10个相邻位点中与该位点属于不同空间域的位点数量，并计算数量大于6的位点所占比例[23]。较小的PAS分数表明空间域内位点的同质性较高。

结　果

1. 空间域聚类结果

（1）乳腺癌A区第1段数据

人工注释提供的20个空间域可视化见图1（a），利用SpaGCN在不同组织学权重下对该数据的3798个位点进行空间域的识别，组织学权重在默认情况下（即s=1），ARI=0.42，见图1（b），当增加组织学权重时（即s=2），聚类效果有所提升（ARI=0.48），见图1（c）。利用BayesSpace方法得到的空间域识别结果（ARI=0.47）见图1（d）。虽然两种方法ARI相差甚微，但相比于BayesSpace，SpaGCN精确识别出肿瘤区域如左下角的IDC_4、右下角的IDC_2以及中间的DCIS/LCIS_3，更清晰地分离病变与非病变区域。



注：（a）乳腺癌（A区第1段）人工注释的空间域可视化；（b）s=1时的空间域可视化结果；（c）s=2时的空间域可视化结果；（d）BayesSpace的空间域可视化结果

图1　乳腺癌A区第1段空间域聚类结果

（2）无标签的浸润性乳腺导管癌

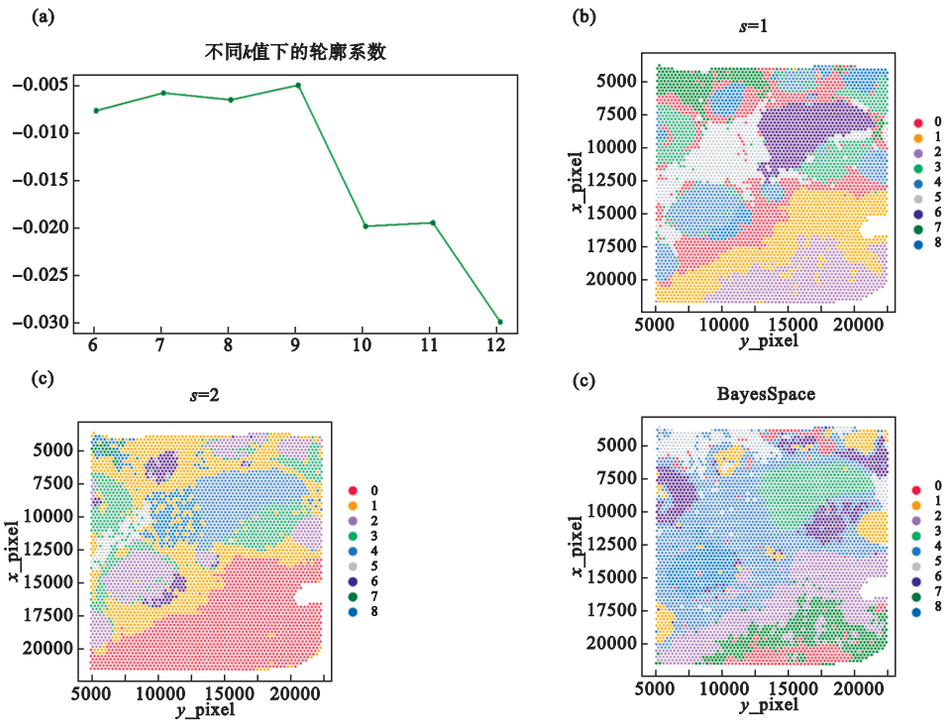
使用SpaGCN对来自10× Genomics平台的浸润性乳腺导管癌数据进行空间域聚类，依据*k-means*算法[25]不同k值下的轮廓系数［见图2（a）］确定聚类数为9，在不同组织学权重下对该数据的4727个位点进行空间域的识别，SpaGCN得到的空间域可视化结果见图2（b）和图2（c），BayesSpace得到的空间域可视化结果见图2（d）。

使用LISI、PAS和CHAOS三个内部指标评价聚类效果（见表1），在较高组织学权重下（即s=2）得到了更好的聚类结果。

2. 空间可变基因

（1）乳腺癌A区第1段数据

选取组织学权重s设置为2所识别的20个空间域，逐一检测SVGs，共检测到10个空间域的398个SVGs，各空间域检测到的SVGs数量及P值排名前三的top基因见表2。其中，*NDUFB9*、*ERLIN2*、*TMEM150C*、*S100G*、*SFRP2*、*ATP6V1G1*、*MLPH*、*ACADSB*共8个基因同时在三个空间域差异表达，*IL6ST*、*WWP1*、*TPD52*、*TBC1D9*、*YWHAE*等50个基因同时在两个空间域差异表达，共识别出332个唯一SVGs。



注：(a)不同k值下的轮廓系数；(b)s=1时的空间域可视化结果；(c)s=2时的空间域可视化结果；(d)BayesSpace的空间域可视化结果

图2 浸润性乳腺导管癌空间域聚类结果

表1 不同组织学权重下的聚类评价指标

	LISI	PAS	CHAOS
s=1	1. 698	0. 147	0. 057
s=2	1. 599	0. 143	0. 055
BayesSpace	1. 686	0. 159	0. 058

表2 乳腺癌A区第1段数据各空间域所检测到的SVGs数量

空间域	SVGs数量	top1基因	top2基因	top3基因
2	9	*SCD*	*NDUFB9*	*IL6ST*
4	75	*TBC1D9*	*YWHAE*	*EEF1A2*
5	12	*IGLC3*	*SFRP4*	*TIMP1*
6	58	*CDH1*	*EFHD1*	*HSBP1*
7	80	*S100G*	*HEBP1*	*CSTA*
9	36	*CYBA*	*AEBP1*	*S100A9*
10	116	*LINC00645*	*MUC5B*	*TMEM150C*
15	9	*SFRP2*	*LUM*	*SFRP4*
16	1	*S100G*	–	–
17	2	*LDLRAD4*	*GFRA1*	–
合计	398	10	9	8

(2)浸润性乳腺导管癌数据

选取组织学权重s设置为2时所识别出的9个空间域，逐一检测SVGs，共检测到9个空间域的637个SVGs，各空间域检测到的SVGs数量及P值排名前三的top基因见表3。其中，*CD24*、*GPRC5A*、*RPS26*、*RPL23A*共4个基因同时在四个空间域差异表达，*MUC1*、*ACTG1*、*KRT8*、*KRT18*、*MIF*等39个基因同时在三个空间域差异表达，*CCND1*、*HSP90AA1*、*CRIP1*、*GALNT6*、*HSPB1*等108个基因同时在两个空间域差异表达，共识别出439个唯一SVGs。

表3 浸润性乳腺导管癌数据各空间域所检测到的SVGs数量

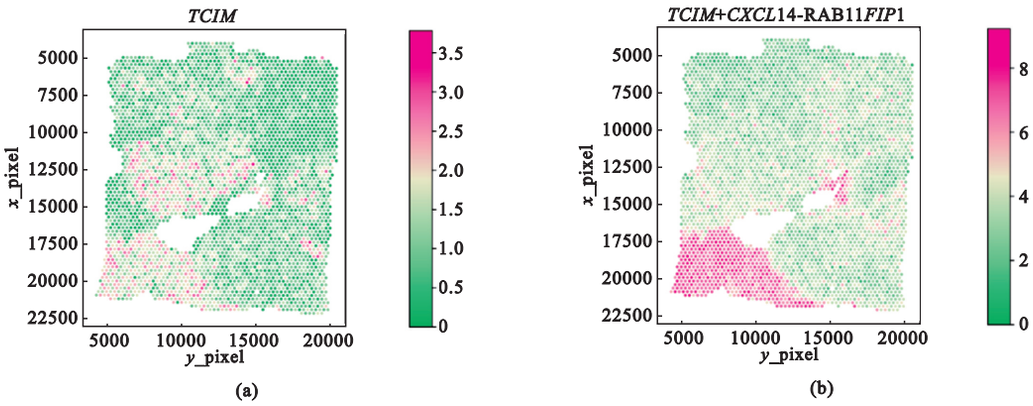
空间域	SVGs数量	top1基因	top2基因	top3基因
0	111	*MUC1*	*GALNT6*	*TCEAL4*
1	13	*IGKC*	*C3*	*IGHG3*
2	23	*ACADSB*	*SNHG25*	*RPL30*
3	42	*SCGB1D2*	*SCGB2A2*	*ADIRF*
4	330	*ZNF703*	*SLC39A6*	*BAMBI*
5	4	*ALB*	*APOD*	*MT-ND5*
6	84	*ACTG1*	*EFNA1*	*CD24*
7	6	*TPR*	*AKAP9*	*BOD1L1*
8	24	*MT-ND2*	*DSP*	*HSP90AB1*
合计	637	9	9	9

3. 空间可变元基因

在人类乳腺癌A区第1段数据中，对于未检测到SVGs的空间域0、1、3、8、11、12、13、14、18、19，将目标域与相邻域的表达倍数阈值由1. 5降为1. 2以识别弱SVGs。其中，空间域1检测到*TGM2*和*TCIM*共2个弱SVGs，空间域8检测到*IGLC1*、*STC2*、*SERPINA3*等7个弱SVGs，空间域11检测到*C1QA*、*APOE*、*SFRP2*等6个SVGs，进一步对弱SVGs进行meta SVGs识别。meta SVGs有更明显的空间表达模式，如图3所示，识别出的meta SVG：*TCIM+CXCL14-RAB11FIP1*比单个基因*TCIM*显示出更清晰的空间表达模式，准确标记了乳腺癌病变区域[即图1(a)的IDC_4区域]。

4. GO和KEGG富集分析

运用DAVID数据库对两个数据筛选出的SVGs进行GO和KEGG富集分析，人类乳腺癌A区第1段数据共得到129条BP条目、79条CC条目、52条MF条目以及28条KEGG通路，分别取P值最小的前5个BP、

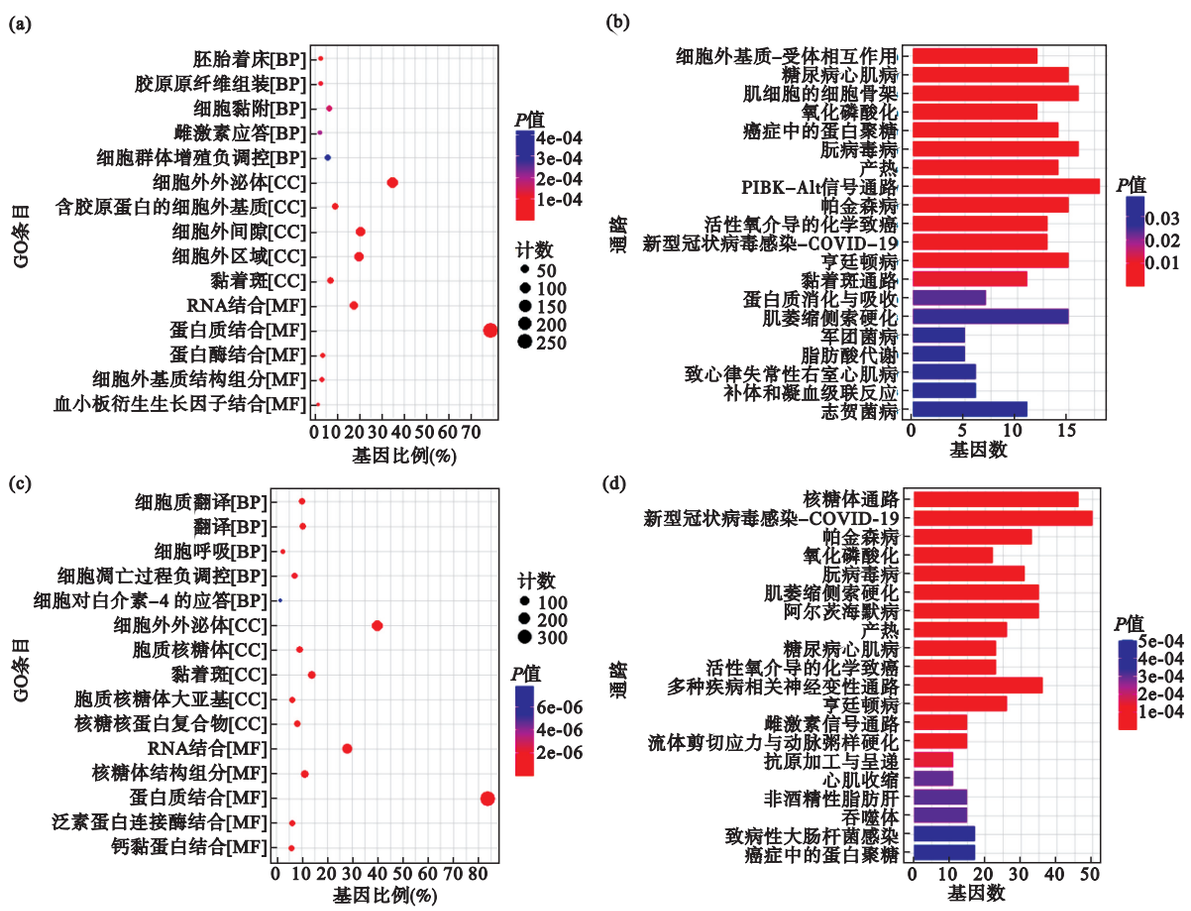


注：(a)SVG：*TCIM*；(b)meta SVG：*TCIM*+*CXCL*14−*RAB*11*FIP*1

图3　乳腺癌A区第1段部分基因空间表达模式

CC和MF共15个条目绘制气泡图，以及*P*值最小的前20条KEGG通路绘制柱状图，见图4(a)和图4(b)。

浸润性乳腺导管癌数据共得到180条BP条目、109条CC条目、78条MF条目以及49条KEGG通路，分别取*P*值最小的前5个BP、CC和MF共15个条目绘制气泡图，以及*P*值最小的前20条KEGG通路绘制柱状图，见图4(c)和图4(d)。



注：(a)乳腺癌A区第1段的GO条目；(b)乳腺癌A区第1段的KEGG通路；(c)浸润性乳腺导管癌的GO条目；(d)浸润性乳腺导管癌的KEGG通路

图4　GO和KEGG富集分析

讨　论

本文采用SpaGCN方法整合空间转录组数据的组织学图像、基因表达和空间位置信息，并在不同组织学权重下对两个乳腺癌数据进行空间域识别，聚类结果表明较高的组织学图像权重能够提升聚类效能，更好地区分病变与非病变区域，揭示了组织学图像在空间域识别中的潜在价值。

本研究对每个空间域进行SVGs检测，对未检测到SVGs的空间域进行弱SVGs及meta SVGs检测。人类乳腺癌A区第1段数据检测出10个空间域的332个SVGs，浸润性乳腺导管癌数据检测出9个空间域的

439 个 SVGs。筛选出的 SVGs 中，如胶原蛋白家族的基因在肿瘤组织边界区域如空间域 9、10、15 差异表达，多项研究表明[26-29]，*COL1A1* 的表达在癌前组织和恶性组织中显著高于正常组织，*COL1A2* 的表达在恶性组织中显著高于癌前组织和正常组织，且与肿瘤大小和浸润深度呈正相关。*CD24* 调节成熟粒细胞和 B 细胞的生长和分化信号，其过表达与乳腺癌患者较差的预后显著相关[30]，是免疫疗法的重要靶点[31]。识别出的 meta SVGs 比单个 SVGs 有更明显的空间表达模式，*CXCL14* 编码参与免疫调节和炎症过程的分泌蛋白，能够促进肿瘤转移[32]和淋巴细胞的浸润[33]。

本研究对筛选出的 SVGs 分别进行了 GO 功能富集分析以及 KEGG 通路富集分析。GO 通路中，两个数据均检测到胞外外泌体（GO:0070062）通路，外泌体 RNA 的定量检测可用于乳腺癌早期诊断[34]。检测出的 KEGG 通路如癌症中的蛋白多糖通路（hsa05205），肿瘤微环境中的许多蛋白多糖已被证明是影响肿瘤进展的各种类型癌症（包括增殖、黏附、血管生成和转移）生物学的关键大分子，其中透明质酸与 CD44 共同作用促进肿瘤细胞生长和迁移[35]。这些结果为乳腺癌的发生机制提供了重要的分子线索，也为乳腺癌的早期诊断、预后评估以及靶向治疗提供了潜在的靶点。

在本研究的基础上，未来可通过对不同亚型乳腺癌（如 Luminal A 型、Luminal B 型、HER2 阳性型和三阴性型）的多组切片进行深入分析，选择适当的方法（如 GraphST[36]、IRIS[37]、SMILE[38]等）以消除批次效应，有效整合多个组织切片，进而开展更为细致的空间转录组学研究，比较各亚型之间空间域特征、SVGs 以及 meta SVGs 的差异。同时，结合更为广泛的临床预后数据（如生存率、复发率等），进一步探讨所识别出的 SVGs 和 meta SVGs 与乳腺癌患者预后的关系，深入探索不同亚型的分子异质性及其临床意义，为乳腺癌的精准分型和个性化治疗提供更加深入的理论支持。

综上所述，SpaGCN 充分利用空间转录组学数据中的组织学图像、基因表达和空间位置信息，通过提高组织学权重，提升空间域识别的精度，从而更加准确地理解组织结构和功能，为肿瘤分型中组织学图像的灵活使用提供了重要理论依据。

参 考 文 献

[1] Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022[J]. Journal of the National Cancer Center, 2024, 4(1): 47-53.

[2] 郑荣寿,陈茹,韩冰峰,等. 2022 年中国恶性肿瘤流行情况分析[J]. 中华肿瘤杂志, 2024, 46(3):221-231.

[3] 崔芳芳,鲍俊哲,王琳琳,等. 1990—2019 年中国女性“两癌”疾病负担变化趋势及预测分析[J]. 中国卫生统计,2022,39(5):647-652.

[4] Jin Y, Zuo Y, Li G, et al. Advances in spatial transcriptomics and its applications in cancer research[J]. Molecular Cancer, 2024, 23 (1): 129.

[5] Almendro V, Marusyk A, Polyak K. Cellular heterogeneity and molecular evolution in cancer[J]. Annual Review of Pathology: Mechanisms of Disease, 2013, 8(1): 277-302.

[6] Ståhl PL, Salmén F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics[J]. Science, 2016, 353(6294): 78-82.

[7] Zhang Y, Gong S, Liu X. Spatial transcriptomics: a new frontier in accurate localization of breast cancer diagnosis and treatment[J]. Frontiers in Immunology, 2024, 15: 1483595.

[8] Zhao E, Stone MR, Ren X, et al. Spatial transcriptomics at subspot resolution with BayesSpace[J]. Nature Biotechnology, 2021, 39 (11): 1375-1384.

[9] Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder[J]. Nature Communications, 2022, 13(1): 1739.

[10] Li Z, Zhou X. BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies[J]. Genome Biology, 2022, 23 (1): 168.

[11] Liu W, Liao X, Yang Y, et al. Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data[J]. Nucleic Acids Research, 2022, 50(12): e72.

[12] Zhou R, Yang G, Zhang Y, et al. Spatial transcriptomics in development and disease[J]. Molecular Biomedicine, 2023, 4(1): 32.

[13] Hu J, Li X, Coleman K, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network[J]. Nature Methods, 2021, 18(11): 1342-1351.

[14] Ji AL, Rubin AJ, Thrane K, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma[J]. Cell, 2020, 182(2): 497-514.

[15] Xu H, Fu H, Long Y, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics[J]. Genome Medicine, 2024, 16(1): 12.

[16] Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis[J]. Genome Biology, 2018, 19: 1-5.

[17] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks[C]. International Conference on Learning Representations, 2017, arXiv:1609. 02907.

[18] Blondel VD, Guillaume JL, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(10): P10008.

[19] Abdi H, Williams LJ. Principal component analysis[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4): 433-459.

[20] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources [J]. Nature Protocols, 2009, 4(1): 44-57.

[21] Wickham H. ggplot2[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2011, 3(2): 180-185.

[22] Kiselev VY, Kirschner K, Schaub MT, et al. SC3: consensus clustering of single-cell RNA-seq data[J]. Nature Methods, 2017, 14(5):

483-486.

[23] Shang L, Zhou X. Spatially aware dimension reduction for spatial transcriptomics[J]. Nature Communications, 2022, 13(1): 7203.

[24] Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony[J]. Nature Methods, 2019, 16(12): 1289-1296.

[25] Kodinariya TM, Makwana PR. Review on determining number of Cluster in K-Means Clustering[J]. International Journal, 2013, 1(6): 90-95.

[26] Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer[J]. World Journal of Surgical Oncology, 2016, 14: 1-5.

[27] Zhang G, Li T, Tan G, et al. Identity of MMP1 and its effects on tumor progression in head and neck squamous cell carcinoma[J]. Cancer Medicine, 2022, 11(12): 2516-2530.

[28] Yu Y, Liu D, Liu Z, et al. The inhibitory effects of COL1A2 on colorectal cancer cell proliferation, migration, and invasion[J]. Journal of Cancer, 2018, 9(16): 2953.

[29] Hayashi M, Nomoto S, Hishida M, et al. Identification of the collagen type 1 alpha 1 gene (COL1A1) as a candidate survival-related factor associated with hepatocellular carcinoma[J]. BMC Cancer, 2014, 14: 1-10.

[30] Wang Z, Wang Q, Wang Q, et al. Prognostic significance of CD24 and CD44 in breast cancer: a meta-analysis[J]. The International Journal of Biological Markers, 2017, 32(1): 75-82.

[31] Barkal AA, Brewer RE, Markovic M, et al. CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy [J]. Nature, 2019, 572(7769): 392-396.

[32] Chang TM, Chiang YC, Lee CW, et al. CXCL14 promotes metastasis of non-small cell lung cancer through ACKR2-depended signaling pathway[J]. International Journal of Biological Sciences, 2023, 19(5): 1455.

[33] Parikh A, Shin JH, Faquin W, et al. Malignant cell-specific CXCL14 promotes tumor lymphocyte infiltration in oral cavity squamous cell carcinoma[J]. Journal for ImmunoTherapy of Cancer, 2020, 8(2):e001048.

[34] Lee JU, Kim WH, Lee HS, et al. Quantitative and specific detection of exosomal miRNAs for accurate diagnosis of breast cancer using a surface-enhanced Raman scattering sensor based on plasmonic head-flocked gold nanopillars[J]. Small, 2019, 15(17): 1804968.

[35] Bourguignon LYW. Hyaluronan-mediated CD44 activation of RhoGTPase signaling and cytoskeleton function promotes tumor progression[C]. Seminars in Cancer Biology. Academic Press, 2008, 18(4): 251-259.

[36] Long Y, Ang KS, Li M, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST[J]. Nature Communications, 2023, 14(1): 1155.

[37] Ma Y, Zhou X. Accurate and efficient integrative reference-informed spatial domain detection for spatial transcriptomics[J]. Nature Methods, 2024, 21(7): 1231-1244.

[38] Chen Y, Zhen C, Mo Y, et al. Multiscale Dissection of Spatial Heterogeneity by Integrating Multi-Slice Spatial and Single-Cell Transcriptomics[J]. Advanced Science, 2025,12(15): e 2413124.

（责任编辑：林美艺）

（上接第646页）

[18] Argüelles MH, Orellana ML, Castello AA, et al. Measles virus-specific antibody levels in individuals in Argentina who received a one-dose vaccine[J]. J Clin Microbiol, 2006, 44(8): 2733-2738.

[19] Gans HA, Arvin AM, Galinus J, et al. Deficiency of the humoral immune response to measles vaccine in infants immunized at age 6 months[J]. JAMA, 1998, 280(6): 527-532.

[20] Brinkman ID, de Wit J, Smits GP, et al. Early Measles Vaccination During an Outbreak in the Netherlands: Short-Term and Long-Term Decreases in Antibody Responses Among Children Vaccinated Before 12 Months of Age[J]. J Infect Dis, 2019, 220(4): 594-602.

[21] Mok DZL, Chan KR. The Effects of Pre−Existing Antibodies on Live−Attenuated Viral Vaccines[J]. Viruses, 2020, 12(5): 12050520.

[22] Amanna IJ, Slifka MK. Mechanisms that determine plasma cell lifespan and the duration of humoral immunity[J]. Immunol Rev, 2010, 236(1): 125-138.

（责任编辑：郭海强）