

# 解析肿瘤免疫微环境: 宫颈癌成纤维细胞的单细胞和空间转录组学见解

林志恒<sup>1†</sup>, 周有为<sup>2†</sup>, 刘振然<sup>2†</sup>, 聂文阳<sup>3†</sup>, 曹恒杰<sup>1</sup>, 李胜男<sup>1</sup>, 李璇玲<sup>1</sup>, 朱立军<sup>1</sup>, 林光耀<sup>1</sup>, 丁艳宇<sup>4,5</sup>, 姜毅<sup>6\*†</sup>, 顾祖熙<sup>7\*†</sup>, 徐连伟<sup>1\*†</sup>, 赵志杰<sup>8\*†</sup>, 蔡华宝<sup>9,10\*†</sup>

## 摘要

背景尽管在筛查、HPV 疫苗接种和治疗策略方面取得了进展, 但宫颈癌 (CC) 仍然是一项重大的全球健康挑战。由表观遗传修饰驱动的肿瘤异质性影响免疫逃逸、转移和治疗反应。癌症相关成纤维细胞 (CAFs) 在 CC 进展和治疗耐药中起关键作用。单细胞测序提供了新的见解, 但在 CC 研究中仍未得到充分利用。本研究整合单细胞 RNA 测序 (scRNA-seq)、空间转录组学和反卷积分析, 以识别关键基因和免疫治疗靶点。通过构建预后模型并探索免疫微环境, 我们旨在为 CC 发病机制和潜在治疗策略提供新的见解。

方法我们利用单细胞 RNA 测序、空间转录组学、反卷积分析和伪时间轨迹映射来描绘宫颈癌肿瘤免疫微环境 (TIME) 中的成纤维细胞亚型。系统地分析了功能注释、差异基因表达谱、细胞间通信途径和转录因子网络。基于批量 RNA 测序数据构建了一个预后模型, 并通过生存分析进行验证, 同时分析了其与免疫微环境特征的相关性。功能实验研究了 SDC1 (成纤维细胞与肿瘤相互作用的关键介质) 的作用。此外, 还进行了成纤维细胞与肿瘤细胞共培养

---

† 林志恒、周有为、刘振然和聂文阳作为共同第一作者对本工作贡献相同。

† 姜毅、顾祖熙、徐连伟、赵志杰和蔡华宝作为共同通讯作者对本工作贡献相同。

\* 通讯作者:

姜毅

jiangyi930902@163.com

顾祖熙

guzuxi1972@163.com

徐连伟

xu\_lianwei2800@shutcm.edu.cn

赵志杰

zhaozhijie@sjtu.edu.cn

蔡华宝

ahmudrtsai@163.com

作者信息完整列表见文章末尾

---

系统和功能测定用于研究 SDC1 的旁分泌作用。通过荧光激活细胞分选 (FACS) 分离出 CAF MYH11<sup>+</sup> 亚群。对培养细胞和人宫颈癌组织样本进行多重免疫荧光和免疫组织化学分析, 以表征 MYH11 在基质重组过程中的空间分布和动态重塑。

结果鉴定出六种不同的成纤维细胞亚型, 包括 C0 MYH11<sup>+</sup> 成纤维细胞, 其在干性维持、代谢活性和免疫调节中发挥独特作用。空间和功能分析表明, C0 亚型对于肿瘤 - 成纤维细胞相互作用至关重要, 特别是通过 MDK - SDC1 信号轴。包含成纤维细胞特异性标志物的预后模型对患者生存结果具有强大的预测能力。此外, 体外敲低 SDC1 显著抑制 CC 细胞增殖、迁移和侵袭。成纤维细胞表现出空间调节的异质性, 激活标志物在肿瘤区域富集, 而 MYH11 在正常区域最高, 表明基质的动态重塑。C0 MYH11<sup>+</sup> CAF 通过可溶性 SDC1 促进肿瘤细胞增殖、迁移并抑制凋亡。

结论我们的结果在某些方面说明了 CAFs 在 CC TIME 中可能的免疫调节和肿瘤支持作用，并强调了 MDK-SDC1 通路作为一个有前景的治疗靶点的可能性。本研究不仅促进了对 CC 中时间异质性的部分新理解，还为开发新的生物标志物和免疫治疗方法以改善临床结果提供了一个可能的参考依据。

关键词宫颈癌、肿瘤免疫微环境、单细胞 RNA 测序、空间转录组学、癌症相关成纤维细胞、SDC1、免疫调节

## 引言

宫颈癌 (CC) 是全球女性中第四大最常被诊断出的癌症，也是癌症相关死亡的第四大主要原因。在转型国家，其发病率和死亡率仅次于乳腺癌 [1]。CC 的主要组织学亚型包括鳞状细胞癌 (SCC)、腺癌 (AC) 和腺鳞癌 (ASC)。尽管 CC 筛查取得了进展，HPV 疫苗接种得到广泛应用，诊断和治疗策略也有所改进，但 CC 仍然是一个重大的公共卫生挑战 [2]。因此，了解宫颈癌发生的分子机制对于开发新的治疗方法和改善患者生存结果至关重要。

宫颈癌的异质性是有效预防和精准治疗的基础。肿瘤异质性体现在多个维度，包括免疫逃逸、复发、转移和耐药性 [3]。肿瘤发生发展过程中的表观遗传修饰会影响肿瘤细胞的增殖、侵袭及治疗反应，进而塑造肿瘤微环境 (TME) 的异质性 [4,5]。这种异质性直接影响诊断准确性、治疗效果和预后 [6,7]。因此，全面了解 TME 异质性对于改善宫颈癌预后至关重要。目前，国际上关于宫颈癌的研究在发病机制、免疫治疗、新型治疗靶点和多组学整合方面取得了重要突破。研究发现，人乳头瘤病毒 (HPV) DNA 整合到宿主基因组中可导致 DNA 断裂、基因扩增以及致癌基因表达激活，如 MYC 和 TERT [8,9]。它还通过 DNA 甲基化和组蛋白修饰调节宿主基因表达，从而影响致癌过程 [10]。宫颈癌组织中 PD-L1 的高表达和 T 细胞耗竭标志物 (如 TIM-3、LAG-3) 的上调提示联合免疫治疗的可能性 [11]。蛋白质组学和转录组学分析表明，NOTCH1 和 WNT 通路的异常激活可能与宫颈癌细胞侵袭性增加有关 [12]。基因组学和代谢组学分析表明，糖酵解途径在宫颈癌患者中高度激活，提示代谢干预可能是一种新的治疗策略 [13]。

单细胞测序技术的出现为揭示肿瘤微环境 (TME) 的异质性提供了强大工具，为治疗创新开辟了新途径 [14 - 16]。然而，单细胞测序在肿瘤学中的应用仍局限于少数肿瘤类型。使用单细胞测序对 CC 的研究很少，导致对其肿瘤内异质性的了解有限。

癌症相关成纤维细胞 (CAFs) 作为宫颈癌肿瘤微环境的关键组成部分，在肿瘤进展、治疗抵抗和免疫逃逸中发挥着关键作用。例如，Takuya 等人通过将人宫颈癌细胞与 CAFs 共同移植到无胸腺小鼠体内，重建了转移性肿瘤微环境，导致 40% 的病例发生淋巴结转移，而单独的宫颈癌细胞则未能转移 [17]。这些发现突出了 CAFs 在克服抗转移屏障中的作用，并强调了它们作为治疗靶点的潜力。同样，Liu 等人通过加权基因共表达网络分析 (WGCNA) 确定了与 CAF 相关的关键基因——COL4 A1、LAMC1、RAMP3、POSTN 和 SERPINF1，并基于这些标志物开发了一个预测模型 [18]。此类研究强调了 CAFs 在宫颈癌发展中的重要性及其作为治疗靶点的潜力。

在本研究中，我们旨在通过整合单细胞 RNA 测序 (scRNA-seq)、反卷积、空间转录组学和伪时间分析，挖掘可能与 CC 发生和进展相关的关键基因和免疫治疗靶点。此外，我们将基于大量 RNA-seq 数据构建一个预后模型，探索 CC 免疫微环境，并通过一系列实验探索许多先前参考数据的可能参考价值。我们的目标是促进 CC 的免疫治疗策略提供新的可能见解和参考，希望最终能够为改善患者预后做出贡献。

## 方法

### 数据收集与处理

本研究的综合工作流程和方法总结于图 1。单细胞 RNA 测序 (scRNA-seq) 数据从 <https://www.ebi.ac.uk/biostudies/array> 获取，登录号为 E-MTAB-12305，批量数据从癌症基因组图谱 (TCGA) 网站 (<https://portal.gdc.cancer.gov/>)、

加州大学圣克鲁兹分校 Xena(<https://xena.ucsc.edu/>) 和基因表达综合数据库 (GEO)(<https://www.ncbi.nlm.nih.gov/geo/>) 获取。来自 10× 网站 (<https://www.10xgenomics.com/>) 的空间转录组数据。

使用 Seurat 软件包 (v4.3.0) 将每个样本的 10× 基因组数据加载到 R 软件 (v4.3.0) 中。最初, 使用 DoubletFinder 程序 (v2.0.3) 过滤潜在的重叠样本和低质量细胞。如果细胞满足以下标准, 则保留用于进一步分析:  $500 < nCount < 50000$ ,  $300 < nFeature < 7000$  且线粒体基因表达低于总计数的 25%。

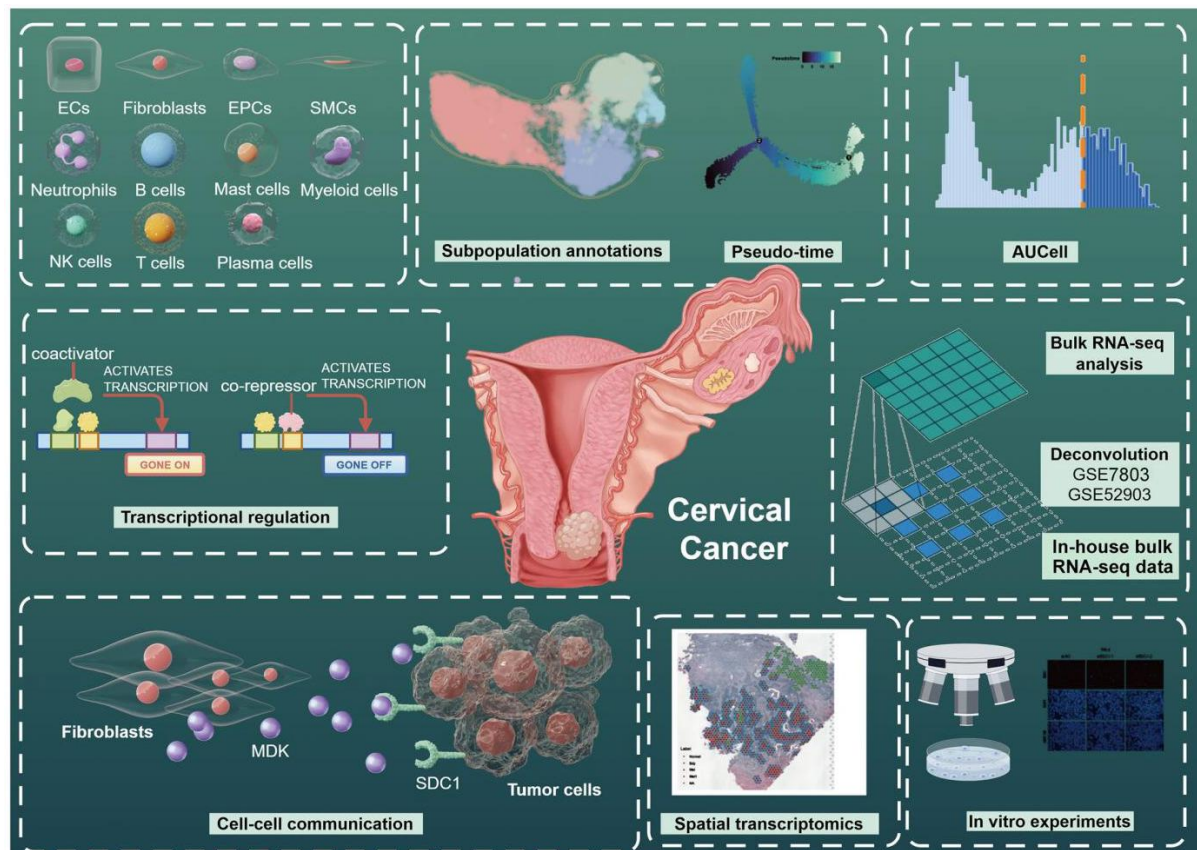


图 1 本研究概述

## 识别细胞类型

对选定样本进行标准化后, 我们从所得表达矩阵中识别出前 2000 个高变异性基因 (HVGs)。ScaleData 函数对数据进行标准化。通过主成分分析 (PCA) 对这些 HVGs 进行分析, 选择前 30 个主成分 (PCs) 进行进一步处理。使用 Harmony 方法 (v1.2.0) 校正样本间的批次效应。使用 FindClusters 和 FindNeighbors 函数识别聚类。然后通过 UMAP 在二维空间中对数据进行可视化, 以促进细胞类型识别。细胞标记物从 CellMarker 数据库 (<http://xteam.xbio.top/CellMarker/>) 获取, 并应用于注释细胞聚类, 使我们能够定义不同的细胞类型并评估其分布和比例。此外, 为了揭示 CC 成纤维细胞之间的异质性, 我们进行了重新聚类, 并根据其特定基因表达模式对每个成纤维细胞亚群进行注释。

## 差异基因表达和 AUCell 的富集分析

为了识别每种细胞类型中的差异表达基因 (DEGs), 我们使用了 “FindAllMarkers” 函数, 采用默认参数的 Wilcoxon 秩和检验 ( $\log_{2} FC > 0.25$ )。通过使用 clusterProfiler(v4.6.2) 和 SCP(v0.4.8) 软件包进行富集

分析, 进一步探索 DEGs 的功能作用。使用 GO-BP 和 GSEA 对每种细胞类型进行通路分析。为了评估单细胞水平的通路活性, 使用 AUCell(v1.22.0) 工具检测单细胞 RNA-seq 数据中的活跃基因。

## 细胞 CNV 水平的测定

我们使用 InferCNV(v1.16.0) 算法计算 CNV 水平。肿瘤发生过程中非整倍体细胞的拷贝数核型分析用于区分非肿瘤细胞和恶性肿瘤细胞。以内皮细胞作为 inferCNV 参考, 以确定其他癌细胞是否表现出显著的染色体拷贝数变异。我们计算了成纤维细胞的 CNV 状态。此外, 我们进一步从 EPCs 中筛选出肿瘤细胞。在从 scRNA-seq 数据衍生的伪时间图谱上, Monocle 有助于识别与 CC 成纤维细胞分化相关的细胞转变。

## Slingshot

为了探索 CC 成纤维细胞不同亚群在发育和分化状态上的差异, 我们首先使用“Slingshot”软件包(v2.8.0) 构建反映每个亚群发育阶段和状态的轨迹。

## 细胞间通讯和 pySCENIC 分析

使用单细胞 RNA 测序数据和 Cell-Chat 软件包(v1.6.1) 预测细胞类型之间的潜在相互作用。CellChatDB.human 是配体-受体相互作用的参考数据库。我们在该数据库中使用 P 值截止为 0.05 的显著性阈值来预测不同细胞类型之间的细胞间相互作用。PySCENIC 是一种基于 scRNA-seq 数据重建基因调控网络 (GRN) 的工具, 特别关注识别稳定的细胞状态。我们在 Python(v3.9.0) 中使用默认参数应用 pySCENIC(v0.1..1) 工具进行调控网络重建并评估转录因子活性。生成 AUCell 矩阵以评估转录因子的富集及其相应调节因子的活性, 从而深入了解控制细胞状态和转变的调控环境。

## 构建 CC 成纤维细胞的预后模型

从 TCGA 数据库 (<https://portal.gdc.cancer.gov/>) 和 UCSC Xena(<https://xena.ucsc.edu/>) 数据库中检索了具有全临床信息的 CC 患者的转录组数据, 以进行进一步分析。为了研究 CC 成纤维细胞作为患者生存预测指标的潜力, 我们采用了一种策略, 即利用成纤维细胞关键亚型的标记基因作为预测基因。使用“survival” R 包进行单变量 Cox 和 Lasso 回归分析, 以识别其他关键的预后基因。随后通过多变量 Cox 回归分析建立预后模型。使用以下公式确定每个样本的风险评分:

## CC 成纤维细胞亚型的伪时间轨迹

使用 Monocle 包 (v2.24.0) 对 CC 细胞进行伪时间轨迹分析。基于

根据 MYH11 + 成纤维细胞风险评分 (MFRS) 的最佳临界值, 将 CC 队列分为两组: 高 MFRS 组和低 MFRS 组。

---

$$\text{风险评分} = (\text{表达基因 } 1 \times \text{系数 } 1) + (\text{表达基因 } 2 \times \text{系数 } 2) + \dots + (\text{表达基因 } N \times \text{系数 } N)。$$

---

## 批量转录组分析

为了评估生存差异, 我们使用生存软件包 (v3.5 - 5) 和 survminer 包 (v0.4.9) 绘制 Kaplan - Meier 曲线进行生存分析。为了评估和校准模型的预测准确性, 使用 “timeROC” 软件包 (v0.4) 生成 1 年、3 年和 5 年间隔的受试者工作特征 (ROC) 曲线。进行主成分分析 (PCA) 以可视化高 MFRS 组和低 MFRS 组之间的转录组差异。为了评估 MFRS 的预后价值, 纳入临床变量进行多变量 Cox 回归分析。使用 rms(v6.7 - 1) 包构建列线图以评估风险评分的预测能力。使用 Pearson 相关分析探索关键分子特征之间的关联。此外, 进行免疫相关分析以表征肿瘤免疫微环境。进行体细胞突变和拷贝数变异 (CNV) 分析以比较两组之间的基因组改变。使用 pRRophetic(v0.5) 包进行药物敏感性分析以预测两组之间的潜在治疗反应。使用 DESeq2(v1.42.1) 包对高 MFRS 组和低 MFRS 组进行差异表达分析。使用 cluster - Profiler 对差异表达基因 (DEG) 进行基因本体 (GO) 和京都基因与基因组百科全书 (KEGG) 通路的富集分析。为了进一步探索生物学相关性, 进行基因集富集分析 (GSEA)。此外, 使用 GSVA(v1.50.5) 对关键通路进行评分并评估其在各个样本中的活性。

## 空间转录组测序分析

我们从 10× 基因组学网站 (<https://www.10xgenomics.com/>) 获得了 CC 空间转录组数据集。使用 Seurat 包进行初始数据处理, 并使用 “SCTransform” 函数对数据集进行归一化和对数转换。使用 “RunPCA” 函数进行降维。为了绘制细胞群体的空间分布, 将整合的 scRNA - seq 数据集用作参考。使用 spacexr(v2.2.1) 包和稳健细胞类型分解 (RCTD) 方法进行细胞类型反卷积, 从而能够估计每个空间点处的细胞类型比例。为了进一步表征肿瘤微环境并在空间转录组切片内划定肿瘤边界, 我们使用了 Cottrazm(v0.1.1) 包。这允许基于转录模式和细胞类型组成对肿瘤与非肿瘤区域进行空间注释。此外, 通过可视化关键基因和通路的空间表达模式进一步检查它们, 突出显示整个组织中特定区域的生物学活性。此外, 使用 stlearn(v0.4.12) 工具分析组织区域之间的空间相互作用, 以研究空间通信和微环境组织。

## 反卷积

用于反卷积的批量数据来自 GEO 数据库。GSE 登录号为 GSE7803 和 GSE52903。此外, 我们还纳入了内部批量 RNA-seq 转录组数据 (三个肿瘤组和三个癌旁对照组)。为了解 CC 的病理过程, 我们以 scRNA-seq 数据为参考, 在多种细胞类型水平上对批量 RNA-seq 样本进行反卷积。将一整套共同基因和选定的标记基因输入 BayesPrism 软件包 (v2.2.2) 以对数据进行反卷积。

## 基因开关分析

为阐明 CC 成纤维细胞激活过程中关键基因表达和功能事件的顺序, 我们使用了基因开关 (v0.1.0) 计算工具, 并结合 scRNA-seq 数据的伪时间轨迹分析。基因开关识别 “开关基因”, 这些基因作为细胞状态之间的开/关调节因子, 并确定这些转变的顺序。基因开关使用 KEGG 和 GO 数据集进行通路分析, 这些数据集是分子特征数据库 (MSigDB) 中特征通路和基因组的集合。基因开关还支持通过识别和可视化共同的开关基因对两个相关的伪时间轨迹进行比较分析。此外, 它突出显示每个轨迹特有的开关基因, 从而深入了解通路或基因组特异性。

## CytoTRACE 和 CytoTRACE 2 分析

我们使用 CytoTRACE(v0.3.3) 和 CytoTRACE2(v1.0.0) 分析 scRNA-seq 数据, 以预测 CC 成纤维细胞的潜在类别和绝对发育潜能。CytoTRACE 对所有成纤维细胞亚群的分化状态进行排名。CytoTRACE2 中的

潜在类别能够根据细胞的发育潜能对细胞进行分类，此外，预测的潜能值提供了从 0(分化) 到 1(全能性) 的发育潜能的连续测量。便于在绝对空间中对发育潜能进行直接的跨数据集比较。

## 细胞培养

CC 细胞系 HeLa 和 SiHa 购自 Pro-cell(中国武汉)。细胞在补充有 10% 胎牛血清 (FBS)、100 U/mL 青霉素和 0.1 mg/mL 链霉素的 MEM 培养基中培养。培养物在湿度培养箱中于 37°C 和 5%CO<sub>2</sub> 条件下维持。收获对数生长期的细胞用于后续实验。此外，HT-3 和 ME-180 宫颈癌细胞系购自中国科学院细胞库(中国上海)。HT-3 细胞在补充有 10% FBS、100 U/mL 青霉素和 0.1 mg/mL 链霉素的 McCoy's 5A 培养基(美国 Gibco) 中培养。ME-180 细胞在含有 10% FBS 和抗生素的 Eagle's Minimum Essential Medium(EMEM; 美国 Gibco) 中在相同的培养条件下维持。所有细胞系均定期检测支原体污染。

## 人体组织标本采集

人宫颈癌及相邻正常组织样本取自安徽医科大学第一附属医院妇产科接受妇科手术的患者。所有患者在样本采集前均提供了书面知情同意书。该研究方案经安徽医科大学第一附属医院伦理委员会审查并批准(批准号:PJ2021-15-28)。新鲜手术标本根据实验要求立即进行原代细胞分离、组织学分析或冷冻保存。所有涉及人体组织的程序均按照《赫尔辛基宣言》的原则和机构伦理标准进行。

## SiRNA 转染

对于实验步骤,将细胞以每孔  $2 \times 10^5$  个细胞的密度接种到 6 孔板中。在 24 h 之后,用从 GenePharma(中国上海) 获得的 siRNA 以终浓度 20  $\mu$ M 转染细胞。使用 RNAiMax(赛默飞世尔科技经销商; 澳大利亚昆士兰州布伦代尔) 按照制造商的说明进行转染。转染后 24 h 收集细胞用于进一步分析。实验中使用的 SDC1 siRNA 序列如下: 制造商的方案。接下来,用 PrimeScript RT 试剂试剂盒(日本东京 Takara) 将 500 ng RNA 逆转录为 cDNA。使用 SYBR<sup>®</sup> Premix Ex Taq<sup>™</sup> (Takara) 进行定量实时 PCR。在 ABI V7 仪器(美国印第安纳州印第安纳波利斯 ABI) 上进行 PCR 扩增。用于扩增的特异性引物针对 SDC1 基因设计如下: 正向引物: 5'-CCAAGC TGACCTTCACACTC-3', 反向引物: 5'-GGC CACTACAGCCGTATTCT-3'

## 细胞活力测定

使用来自日本熊本 DOJINDO 的细胞计数试剂盒-8(CCK-8) 评估细胞活力。将细胞以每孔  $1 \times 10^3$  个细胞的密度接种到 96 孔板中并培养过夜。向每个孔中加入 100  $\mu$ L 检测试剂并孵育 1 h。在 4 天内每天测量 450 nm 处的吸光度。通过将 OD450 值与时间相关联绘制生长曲线以评估细胞活力。

## 克隆形成测定

制备对数生长期细胞的悬浮液并稀释, 每孔接种  $1 \times 10^3$  个细胞到 6 孔板中。细胞培养 10 天, 定期观察。一旦可见菌落, 去除培养基, 用冰冷的 PBS 洗涤细胞两次, 然后在 4% 多聚甲醛中固定 20 min。固定后, 用 0.1% 结晶紫染色 10 min, 并使用凝胶成像分析系统(印度班加罗尔 Syngene, GBOX-F3EE) 计数菌落。

## EdU 分析

制备对数生长期细胞的悬浮液，稀释后以每孔  $1 \times 10^3$  个细胞的密度接种到 6 孔板中。按照制造商的说明 (中国 RiboBio) 进行 EdU 测定。在荧光显微镜下检查处理过的细胞，并通过在至少六个随机选择的视野中计数细胞来量化 EdU 阳性细胞的数量。

## 伤口愈合测定

每孔接种  $2 \times 10^5$  个细胞到 6 孔板中并培养过夜。用垂直于培养物的  $10\mu\text{L}$  移液器吸头在细胞中划一道划痕

siRNA1 : GACUGCUUUGGACCUAAAU; siRNA2 : GCAAAUUGUGGCUACUAAU.

## RNA 提取和定量实时 PCR

根据美国马萨诸塞州沃尔瑟姆的赛默飞世尔科技公司的 TRIzol 试剂提取总 RNA。用 PBS 冲洗平板三次以去除脱落的细胞，然后将新鲜的无血清培养基加入平板中。细胞在  $37^\circ\text{C}$  的培养箱中培养 5%  $\text{CO}_2$ ，并在明场显微镜下于 0 h 和 48 h 获取图像。

## Transwell 迁移和侵袭测定

进行 Transwell 实验以评估细胞迁移和侵袭能力。将细胞接种于 24 孔 Transwell 小室 (美国 BD Biosciences 公司) 中，使用无血清培养基。对于侵袭实验，小室预先用 2% 的基质胶包被。在下室中加入 20% 胎牛血清作为趋化剂。在 48 h 后，去除未迁移至下表面的细胞。然后将下表面迁移的细胞固定，用结晶紫染色，并在显微镜下进行成像和分析。

## MYH11 + 成纤维细胞的分离

使用肿瘤解离试剂盒 (德国美天旎生物技术公司, 产品编号 130 - 095 - 929) 从新鲜的人类肿瘤标本中分离原代癌相关成纤维细胞 (CAF)，该试剂盒包含用于高效组织消化的优化酶试剂 (IV 型胶原酶、透明质酸酶和 DNA 酶 I)。将肿瘤组织切成  $\sim 1 - 3 \text{ mm}^3$  碎片，在  $37^\circ\text{C}$  下用酶解离 45 min，并轻轻搅拌，然后通过  $70\mu\text{m}$  细胞筛过滤以获得单细胞悬液。最初通过差异贴壁法 (6 - 12 小时) 富集 CAF，然后使用抗成纤维细胞磁珠 (加拿大 STEM - CELL Technologies 公司, 产品编号 17,859) 进一步纯化。为了去除污染细胞，使用磁珠 (德国美天旎生物技术公司, 产品编号分别为 130 - 105 - 958、130 - 091 - 935 和 130 - 045 - 801) 去除上皮细胞 (EpCAM +)、内皮细胞 (CD31 +) 和造血细胞 (CD45 +)。为了分离 MYH11 + CAF 亚群，使用 BD Cytotfix/Cytoperm<sup>TM</sup> (美国 BD Biosciences 公司, 产品编号 554,714) 使细胞通透，然后用抗 MYH11 - Alexa Fluor 647 (英国 Abcam 公司, 产品编号 ab53219) 进行细胞内染色，并用抗  $\alpha$  - SMA - PE (美国 R&D Systems 公司, 产品编号 IC1420P)、抗 CD31 - APC、抗 CD45 - FITC 和抗 EpCAM - BV421 (美国 BioLegend 公司, 产品编号 303,115、304,005 和 324,215) 进行表面染色。在 BD FACSAria III (美国 BD Biosciences 公司) 上进行流式细胞术分析，并使用顺序门控策略分选 MYH11 + CAF: 活的单细胞  $\rightarrow \alpha$  - SMA +  $\rightarrow$  MYH11 +。通过对分选群体的重新分析确认纯度。



## 多重免疫荧光

进行多色免疫荧光染色以检测  $\alpha$ -平滑肌肌动蛋白、波形蛋白、钙调蛋白、纤连蛋白和肌球蛋白重链 11。将组织切片或培养细胞用 4% 多聚甲醛 (美国西格玛奥德里奇公司) 在室温下固定 15 分钟。必要时, 使用柠檬酸盐缓冲液 (pH 6.0) 在 95°C 下进行 15 min 的抗原修复。固定后, 样品用 0.3% Triton X-100 (美国西格玛奥德里奇公司) 通透处理 10 min, 然后用 5% 牛血清白蛋白 (BSA, 美国西格玛奥德里奇公司) 在室温下封闭 1 h 以减少非特异性结合。然后将样品在 4°C 下与从英国剑桥艾博抗公司购买的一抗孵育过夜:  $\alpha$ -平滑肌肌动蛋白 (ab32575, 1:200)、波形蛋白 (ab92547, 1:200)、钙调蛋白 (ab46794, 1:200)、纤连蛋白 (ab23750, 1:200) 和肌球蛋白重链 11 (ab133567, 1:200)。用 PBS 洗涤三次后, 在室温黑暗条件下应用 Alexa Fluor 偶联的二抗 (Alexa Fluor 488 偶联的山羊抗兔 IgG [ab150077, 1:500] 和 Alexa Fluor 594 偶联的山羊抗小鼠 IgG [ab150116, 1:500], 艾博抗公司) 孵育 1 h。在用 Fluoromount-G (美国南方生物科技公司) 封片前, 用 DAPI (ab228549, 1:1000, 艾博抗公司) 对细胞核进行复染 10 分钟。使用蔡司 LSM 880 共聚焦显微镜 (德国) 进行成像, 并使用美国国立卫生研究院的 ImageJ 软件通过皮尔逊相关系数进行共定位分析。

## 多重免疫组化

将福尔马林固定、石蜡包埋 (FFPE) 的人肿瘤组织、相邻正常组织和肿瘤边界区组织切成 4  $\mu$ m 薄片。玻片用二甲苯脱蜡, 通过梯度乙醇复水, 并在高压锅中使用 10mM 柠檬酸盐缓冲液 (pH 6.0) 进行热诱导抗原修复。对于多重免疫组化 (mIHC), 进行带有酪胺信号放大 (TSA) 的顺序单抗体染色。用 3%  $H_2O_2$  封闭内源性过氧化物酶活性, 然后用 5% 正常血清封闭。依次应用包括抗  $\alpha$ -平滑肌肌动蛋白 (艾博抗公司, 产品编号 ab5694)、抗波形蛋白 (细胞信号技术公司, 产品编号 5741)、抗钙调蛋白 (艾博抗公司, 产品编号 ab467994)、抗纤连蛋白 (艾博抗公司, 产品编号 ab2413) 和抗肌球蛋白重链 11 (艾博抗公司, 产品编号 ab53219) 在内的一抗。孵育后, 玻片用 HRP 偶联的二抗 (杰克逊免疫研究公司, 种属特异性) 处理, 然后使用单独的荧光团 (珀金埃尔默公司, 产品编号 FP1487001 KT, FP1497001 KT) 进行 TSA 荧光团标记。在每一轮之间使用 0.1 M 甘氨酸缓冲液 (pH 2.0) 在 50°C 下进行抗体剥离。在最后一轮染色后, 用 DAPI (赛默飞世尔科技公司, 产品编号 D1306) 对细胞核进行复染, 玻片用 ProLong<sup>TM</sup> Gold 抗褪色封片剂 (赛默飞世尔科技公司, 产品编号 P36930) 封片。在 Vectra<sup>®</sup> Polaris<sup>TM</sup> 自动定量病理成像系统上进行成像, 并使用 inForm<sup>®</sup> 软件 (Akoya Biosciences 公司) 进行光谱解混和定量分析。

## 细胞转染

按照制造商的说明, 使用 Lipofectamine 3000 试剂盒 (美国加利福尼亚州卡尔斯巴德市英杰公司) 进行转染。将细胞在 6 孔板中培养至 50% 至 60% 汇合度, 并用含有 shRNA 的质粒进行转染。感染 72 小时后, 在 2  $\mu$ g/mL 嘌呤霉素存在下选择细胞, 收集对嘌呤霉素有抗性的细胞并进行培养。稳定细胞系构建完成。转染后 72 h 收获细胞, 随后通过定量实时聚合酶链反应 (qPCR) 或蛋白质免疫印迹分析来确认转染效率。

## RNA 提取与 qPCR

根据制造商的说明, 使用 TRIzol 试剂 (美国马萨诸塞州沃尔瑟姆市英杰公司) 从组织或细胞中获得总 RNA。使用 NanoDrop 2000 仪器 (美国马萨诸塞州沃尔瑟姆市赛默飞世尔科技公司) 评估 RNA 的浓度和纯度。使用 GoScript 逆转录系统 (美国威斯康星州麦迪逊市普洛麦格公司) 从总 RNA 合成第一链 cDNA。使用 GoTaq qPCR Master Mix 试剂盒 (美国威斯康星州麦迪逊市普洛麦格公司) 进行 qPCR。使用 GAPDH 作为内参进行标准化。本研究中使用的所有引物均由金斯瑞定制。引物序列见补充表 1。



## 成球实验

将肿瘤细胞用对照培养基处理，与 CAFMYH11+ 共培养或用指示条件处理 CAFMYH11+ 24 小时后收集的上清液共培养。孵育 72 小时后，将每孔总共 1000 个单细胞接种到含有 B27 补充剂(美国加利福尼亚州卡尔斯巴德市英杰公司旗下 Gibco 产品)、EGF(20 ng/mL; 美国新泽西州皮斯卡塔韦市 Peprotech 公司)和 bFGF(20 ng/mL; 美国新泽西州皮斯卡塔韦市 Peprotech 公司)的 DMEM/F12 培养基(美国密苏里州圣路易斯市西格玛奥德里奇公司)中的超低附着 6 孔板中。7 天后，计数直径为  $> 75\mu\text{m}$  的球体。

## 蛋白质免疫印迹

使用放射免疫沉淀分析 (RIPA) 裂解缓冲液提取总细胞蛋白。使用二辛可宁酸 (BCA) 测定法(中国江苏碧云天)测量蛋白质浓度。等量的蛋白质在 10% 十二烷基硫酸钠 - 聚丙烯酰胺凝胶电泳 (SDS - PAGE) 凝胶上分离。电泳后，将蛋白质转移到聚偏二氟乙烯 (PVDF) 膜上。将膜在室温下用 5% 脱脂牛奶封闭 1 h，然后在  $4^{\circ}\text{C}$  与一抗孵育过夜。之后，将膜在  $37^{\circ}\text{C}$  与适当的二抗孵育 2 h。使用来自英国剑桥市艾博抗公司的以下一抗:SDC1(ab142138, 1:1000),  $\beta$ -肌动蛋白(ab8227, 1:5000)。使用的二抗是辣根过氧化物酶 (HRP) 偶联的抗兔 IgG(ab6721, 1:5000, 艾博抗)。使用增强化学发光 (ECL) 检测系统(美国赛默飞世尔科技公司)检测蛋白质条带，并使用 ImageJ 软件(美国国立卫生研究院)进行图像分析。

## 统计分析

使用 R 软件 (v4.3.0) 和 Python 软件 (v3.9.0) 进行统计分析。威尔科克森检验和 (见下页图。)

图 2 宫颈癌中成纤维细胞亚型的特征。A 所有高质量细胞进行降维处理并聚类到 UMAP 图中，根据差异表达基因对细胞类型进行注释。还展示了各细胞间 G2/M 评分、S 评分、nCount RNA 和 nFeature RNA 的差异。B UMAP 图展示了肿瘤组和肿瘤旁正常组织中细胞的分布。还呈现了所有注释细胞类型内的细胞周期阶段分布，并用柱状图说明其各自的百分比。C 使用气泡图可视化所有细胞的前 5 个标记基因以及两组之间的差异表达基因。D 对所有注释细胞进行 GO - BP 分析。E 对宫颈癌成纤维细胞中上调和下调基因进行 GO - BP 分析。F GSEA 分析确定了成纤维细胞中富集的通路。G 富集网络图基于成纤维细胞基因集说明了通路分析结果。H 根据其标记基因，成纤维细胞被分为六个具有不同特征的亚型。I 使用 UMAP 图展示肿瘤组和肿瘤旁正常组织中成纤维细胞亚型的分布，以及每个亚型内细胞周期阶段的分布。J 堆叠柱状图描绘了不同样本来源中成纤维细胞亚型的百分比组成。K ST 图推断了宫颈癌组织切片内选定位置最可能的细胞类型。L 展示了各成纤维细胞亚型间 nFeature RNA、CNV 评分和 nCount RNA 的差异。M ST 图展示了成纤维细胞亚型的 nFeature RNA、CNV 标签和 nCount RNA。N 热图展示了六个成纤维细胞亚型间差异表达的标记基因。O 对每个成纤维细胞亚型的差异表达基因进行 GO - BP 分析。P 词云突出显示了与六个成纤维细胞亚型中每一个密切相关的术语。Q ST 图可视化了 CO MYH11 + 成纤维细胞。R GO - BP 分析确定了 C0 亚型中肿瘤组和肿瘤旁正常组织中上调的基因。S GSEA 分析揭示了 C0 MYH11 + 成纤维细胞富集的通路

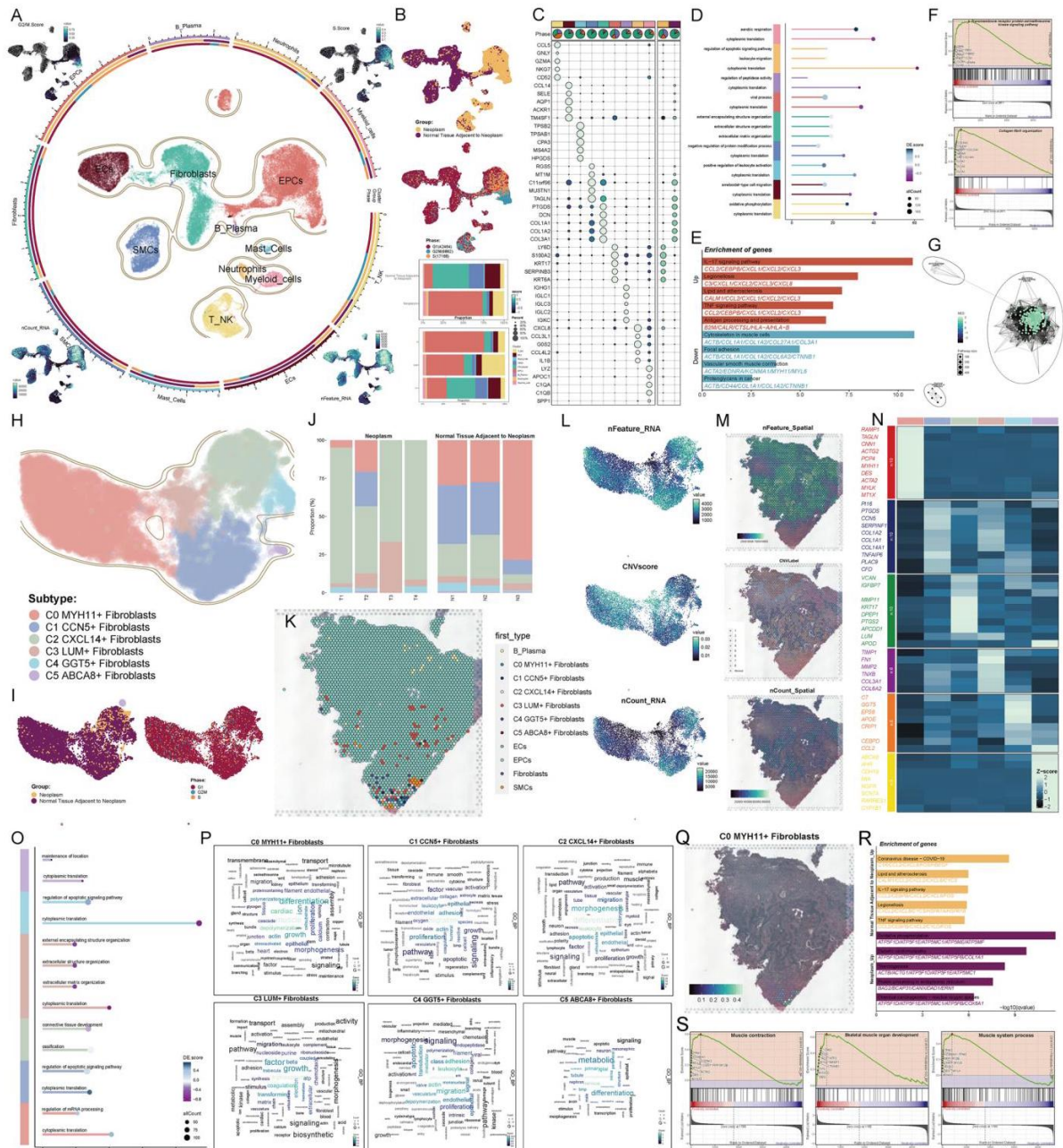


图 2(见上一页面图例)

使用 Pearson 相关系数评估组间差异的显著性。显著性水平定义如下: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , \*\*\*\*  $P < 0.0001$ , 其中 'ns' 表示无显著差异。这些统计检验和显著性阈值用于评估研究结果的可靠性并确保结果的稳健性。

## 结果

### 宫颈癌成纤维细胞亚型图谱

对所选数据集中的所有细胞进行批次校正后, 共获得 69494 个高质量细胞。根据已知的标记基因, 这些细胞被分为 9 种细胞类型, 包括内皮细胞 (ECs)、成纤维细胞、上皮细胞 (EPCs)、平滑肌细胞 (SMCs)、

B 浆细胞、肥大细胞、中性粒细胞、髓样细胞和 T-NK 细胞。需要注意的是,在进行细胞类型注释之前,通过基于 inferCNV 结果选择上皮细胞来识别肿瘤细胞,突出显示那些具有高 CNV 分数的细胞(补充图 1A)。与其他细胞类型相比,EPCs 的 G2/M 和 S 分数显著更高。EPCs、成纤维细胞、ECs、髓样细胞和 SMCs 的 nCount RNA 和 nFeature RNA 值更高(图 2A)。

根据细胞来源将其分为两组:肿瘤组和肿瘤旁正常组织组。UMAP 图显示,EPCs、中性粒细胞、髓样细胞和 T-NK 细胞主要集中在肿瘤组,而 ECs、成纤维细胞和 SMCs 主要位于肿瘤旁正常组织组。柱状图表明,EPCs 在肿瘤组中占比最大,而成纤维细胞是肿瘤旁正常组织组中的主要细胞类型。此外,成纤维细胞主要处于 G1 期,而 EPCs 主要处于 G2/M 和 S 期(图 2B)。关于差异基因表达,成纤维细胞的前 5 个标记基因是 PTGDS、DCN、COL1A1、COL1A2 和 COL3A1。在肿瘤旁正常组织组中表达较高的基因包括这些标记基因以及 TAGLN 和 C11orf96(图 2C)。对每种细胞类型的差异表达基因进行 GO-BP 分析表明,成纤维细胞主要富集在外部包被结构组织、细胞外结构组织和细胞外基质组织等过程中(图 2D)。进一步分析表明,成纤维细胞中上调的基因主要与 IL-17 信号传导、军团菌病、脂质代谢、动脉粥样硬化、TNF 信号传导和抗原加工及呈递等途径相关。相反,下调的基因与肌肉细胞中的细胞骨架、粘着斑、血管平滑肌收缩和癌症中的蛋白聚糖等途径相关(图 2E)。GSEA 分析突出了成纤维细胞的两个重要途径:跨膜受体蛋白丝氨酸/苏氨酸激酶信号通路和胶原纤维组织(图 2F)。富集网络图进一步显示,所有成纤维细胞亚型的基因集在细胞表面受体信号通路和丝氨酸水解酶活性中富集程度最高(图 2G)。

在重新进行批次校正后,我们将来自肿瘤组和肿瘤旁正常组织组的 15843 个成纤维细胞分为六个不同的细胞簇。在注释成纤维细胞亚型之前,我们观察了每个亚型的 CNV(拷贝数变异)模式(补充图 1B)。然后我们根据其特定的标记基因表达对亚型进行命名(补充图 2A),并使用空间转录组学确认了这六个标记基因的表达(补充图 2B)。这六个亚型如下:C0 MYH11 + 成纤维细胞、C1 CCN5 + 成纤维细胞、C2 CXCL14 + 成纤维细胞、C3 LUM + 成纤维细胞、C4 GGT5 + 成纤维细胞和 C5 ABCA8 + 成纤维细胞(图 2H)。除了 C2 亚型有一些来自肿瘤组的细胞外,其他亚型的大多数细胞起源于肿瘤旁正常组织组。所有成纤维细胞亚型中的大多数细胞处于 G1 期(图 2I)。在这些亚型中,C0 MYH11 + 成纤维细胞在肿瘤旁正常组织组的三个样本中占比最大(图 2J)。我们还使用空间转录组学在 CC 组织切片的选定点验证了所有成纤维细胞亚型的存在,确认了所有成纤维细胞亚型的存在(图 2K, 补充图 7A)。

接下来,我们使用 UMAP 图直观展示了不同亚型之间 nFeature RNA、CNV 评分和 nCount RNA 的差异。值得注意的是,C0 亚型表现出较高的 CNV 评分(图 2L)。对 C0 亚型的反卷积分析显示,与肿瘤组相比,来自肿瘤旁正常组织组的细胞丰度更高。除了 EPCs,C0 亚型还包含第二高比例的细胞,这与我们的观察结果一致。在 GSE7803、GSE52903 和内部批量 RNA 测序数据中,MYH11 在相邻正常组织中的表达更高

(见下一页的图。)

图 3 成纤维细胞亚型中干性、代谢和 apCAF 特征的评分。A 使用 UMAP 图和 ST 特征图直观展示成纤维细胞亚型的干性。箱线图说明了各亚型之间干性评分的差异。B 气泡图展示了成纤维细胞亚型中差异表达的干性相关基因的表达水平。C UMAP 图突出显示了成纤维细胞亚型中干性相关基因(包括 KDM5B、CTNBN1、CD44 和 CD34)的表达差异。D 显示了肿瘤组和肿瘤旁正常组织组之间干性相关基因表达的差异。E 确定并展示了每个成纤维细胞亚型和两个样本组的显著代谢途径。F UMAP 图展示了各亚型之间氧化磷酸化和谷胱甘肽代谢的差异。箱线图比较了这两种代谢途径在各亚型细胞周期阶段和样本组之间的差异。(G) 使用 UMAP 图和箱线图直观展示了成纤维细胞亚型之间 pHB、pMT 和 pRP 评分的差异。H UMAP 图显示了各亚型之间 apCAF 评分在表达和密度方面的差异。ST 特征图在空间上展示了 apCAF 评分,箱线图定量比较了成纤维细胞亚型之间的 apCAF 评分比肿瘤组织中的高,这与单细胞分析一致(补充图 2C-E, 补充图 7B)。空间转录组学进一步表明可能存在各种 CNV 标签,CNV 标签 ST 特征图显示了标签 1-8 以及正常标签(图 2M)。这些 CNV 标签与不同的 CNV 评分相关(补充图 2F)。然后,我们检查了所有成纤维细胞亚型之间的差异基因表达(图 2N),并对每个亚型的差异表达基因进行了富集分析。GO-BP 分析表明,C0 亚型主要富集在与 mRNA 加工调控和细胞质翻译相关的过程中(图 2O)。C0 亚型的词云突出显示了诸如肌肉、心脏、解聚、分化和生长等关键术语(图 2P)。基于空间转录组学,我们推断了 C0 亚



型在 CC 组织切片中的可能位置, 如图 2Q 所示。此外, 对肿瘤组和肿瘤旁正常组织组中上调基因的 GO - BP 分析表明, 肿瘤旁正常组织组中上调的基因主要富集在与冠状病毒病 (COVID - 19)、脂质代谢和动脉粥样硬化、IL - 17 信号传导、军团病和 TNF 信号传导相关的途径中 (图 2R)。最后, 对 C0 亚型的 GSEA 分析显示在肌肉收缩、骨骼肌器官发育和肌肉系统过程中显著富集 (图 2S)。

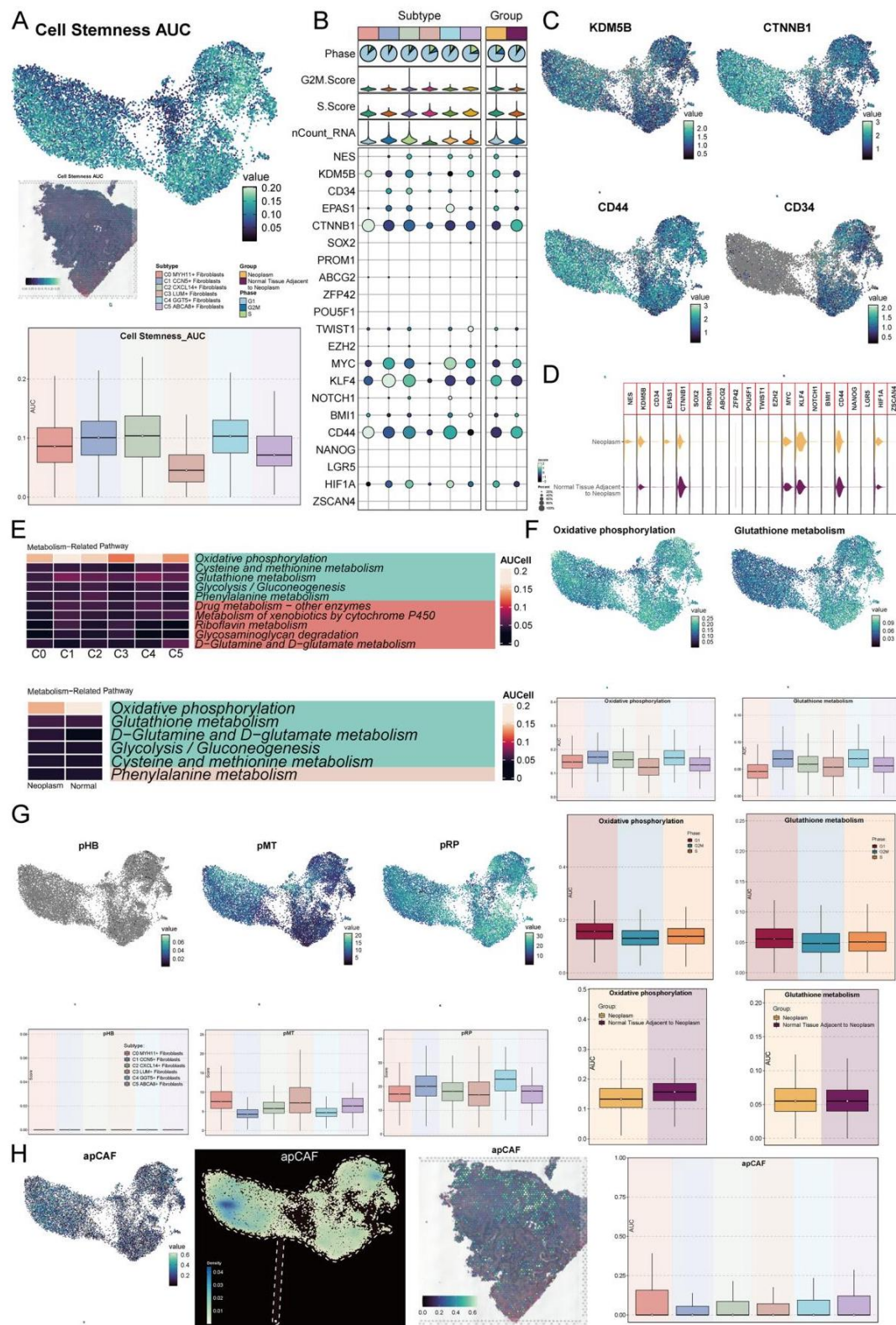


图 3(见上一頁的图例。)

## 成纤维细胞亚型的干性、代谢和 CAF 特征异质性

为了更好地理解成纤维细胞亚型之间的各种差异以及关键亚型 C0 的具体特征，我们进行了更全面的评分和可视化分析。首先，我们使用 UMAP 来展示各亚型间细胞干性的差异，空间转录组学分析证实了细胞干性 AUC 值的最大潜力。箱线图比较了各亚型间的干性水平，结果显示 C2 亚型表现出最高的干性，其次是 C1、C4 和 C0 亚型，而 C3 的干性最低(图 3A)。接下来，气泡图展示了各亚型间以及两个样本组间干性基因的表达差异。在 C0 亚型中高表达的干性基因包括 CTNNB1、CD44 和 KDM5B，而在肿瘤旁正常组织组中，CTNNB1、CD44 和 MYC 高表达(图 3B)。然后，我们使用 UMAP 可视化了在 C0 亚型中高表达的干性基因(KDM5B、CTNNB1 和 CD44)，并与在 C0 亚型中几乎不存在的干性基因 CD34 进行对比(图 3C)。此外，小提琴图用于直观比较两个样本组间干性基因的表达水平(图 3D)。

然后，我们使用热图展示了所有成纤维细胞亚型和两个样本组的显著代谢途径(图 3E)。C0 亚型和肿瘤旁正常组织组均显示氧化磷酸化是最突出的代谢途径，其次是谷胱甘肽代谢。这两条途径被专门可视化。两种代谢途径中的细胞大多处于 G1 期，但氧化磷酸化在肿瘤旁正常组织组中更突出，而谷胱甘肽代谢则不太明显(图 3F)。我们还对成纤维细胞亚型的 pHB、pMT 和 pRP 进行了评分，发现与其他亚型相比，C0 的 pMT 得分最高，而其 pRP 得分最低。所有亚型中的 pHB 几乎为零(图 3G)。最后，为了评估该数据集成纤维细胞亚型的 CAF(癌症相关成纤维细胞)特征，我们对所有亚型的六个相关 CAF 特征进行了评分。气泡图展示了各亚型间 CAF 相关评分的差异(补充图 3A)。我们发现与其他亚型相比，C0 亚型的 apCAF 评分差异最为显著。然后，我们可视化了 apCAF 评分，其在 C0 亚型中确实更高，密度更强。小提琴图也证实了这一观察结果。此外，空间转录组学验证了 apCAF 在 CC 组织切片中的最大潜力(图 3H)。各亚型间其他五个 CAF 相关特征的评分和密度差异也通过 UMAP 图展示(补充图 3B、C)，并通过空间转录组学进一步验证(补充图 3D)。箱线图直观比较了各亚型间 CAF 相关评分的水平(补充图 3E)。

---

(见下页图。)

图 4 成纤维细胞亚型中的伪时间异质性。A - B CytoTRACE 分析揭示了六种成纤维细胞亚型的分化水平。C 鉴定了与 CytoTRACE 呈正相关和负相关的基因。D UMAP 图显示了每个亚型占据的主要伪时间区域。E 构建了从左到右的伪时间轨迹，显示了六种成纤维细胞亚型沿轨迹的分布。F 热图展示了沿伪时间轨迹各亚型间差异表达基因的时间表达模式。G 基于分支点将伪时间轨迹分为五个状态(状态 1 - 5)，饼图说明了每个状态的百分比组成。H 堆叠柱状图描绘了状态 1 - 5 中各亚型的比例代表情况。I GeneSwitches 分析展示了表面蛋白、转录因子(TFs)和所有基因沿伪时间轨迹的分布。J 对伪时间轨迹状态 1 - 4 中不同时间点上调和下调的基因进行了 GO - BP 分析。K 鉴定了在高 C0 和低 C0 伪时间轨迹上共同表达的基因。L 突出显示了沿高 C0 轨迹特异性表达的基因，包括早期的 CST3 和晚期的 COL14A1、VIM 和 S100A4。M 可视化了 COL14A1、VIM 和 S100A4 沿伪时间轨迹的表达模式。N ST 特征图描绘了 COL14A1、VIM 和 S100A4 的空间表达。O 热图说明了干性基因沿伪时间轨迹的差异表达。P CytoTRACE2 分析揭示了具有不同效力的干性基因的表达分布。Q UMAP 图显示了 CytoTRACE2 在不同亚型中的相对密度和评分密度。R Slingshot 分析在亚型间构建了两条轨迹，谱系 1 和谱系 2。沿两条轨迹展示了 C0 MYH11 + 成纤维细胞命名基因 MYH11 的表达。S UMAP 图说明了谱系 1 和谱系 2 在各亚型中的时间进展。(T) 基于不同状态，将 Slingshot 轨迹(谱系 1 和谱系 2)映射到各亚型的分布上。U 可视化了谱系 1 和谱系 2 上干性相关基因(KDM5B、CTNNB1 和 CD44)表达的时间变化

---

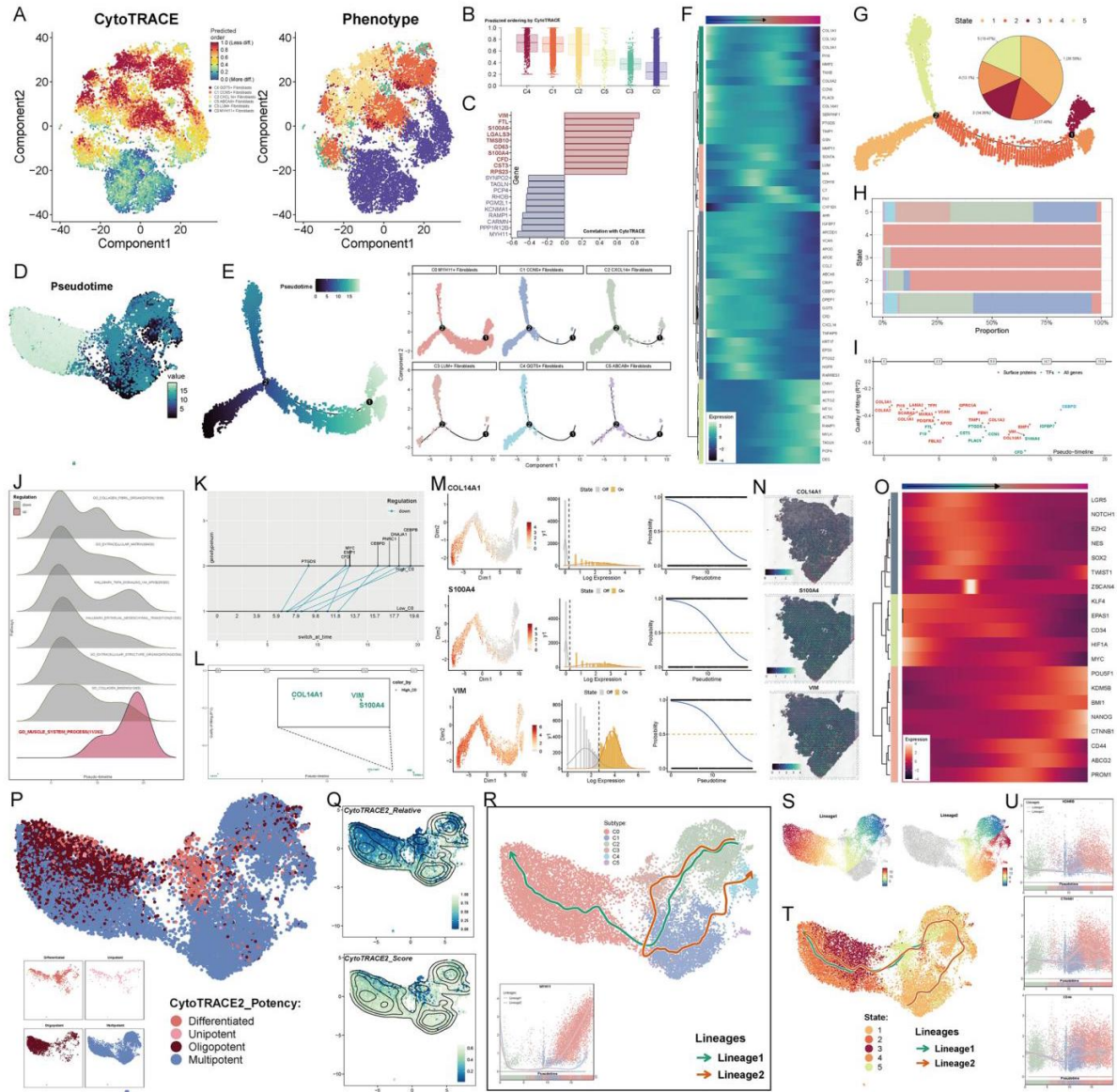


图 4(见上一页图例)

## 伪时间轨迹中 C0 亚型的显著特征

首先，细胞追踪分析揭示了不同成纤维细胞亚型的分化水平，其中 C4 亚型的分化程度最低，C0 亚型的分化程度最高(图 4A、B)。与细胞追踪分析呈正相关的基因包括 VIM、FTL、S100A6、LGALS3、TMSB10、CD63、S100A4、CFD、CST3 和 RPS23(图 4C)。接下来，我们在 UMAP 图上展示了各亚型之间的伪时间顺序分布(图 4D)，并构建了一条从左到右具有时间进程的伪时间轨迹，该轨迹有两个分支点。然后，我们将每个成纤维细胞亚型映射到这条轨迹上，发现大多数 C0 亚型位于右分支(图 4E)。热图显示了随着伪时间推进差异基因的表达差异(图 4F)。为了更好地定义伪时间轨迹上的位置，我们根据两个分支点将整个轨迹分为五个部分，并将它们命名为五个状态。饼图展示了每个状态在整个伪时间轨迹中所占的比例，其中状态 1 占比最大，为 36.58%(图 4G)。堆叠柱状图显示了每个成纤维细胞亚型在每个状态中所占的比例，C0 亚型在状态 2、3 和 4 中占据主导地位(图 4H)。基于 Monocle，我们纳入了基因开关，首先展示了轨迹中的所有表面蛋白、转录因子(TFs)和基因(图 4I)。由于我们将轨迹分为两条主要路径，一条由状态



1 和 5 组成, 另一条由状态 1 - 4 组成, 我们将含有更多 C0 亚型细胞的分支命名为高 C0, 而另一个分支命名为低 C0。然后, 我们对沿高 C0 轨迹上调和下调的基因进行了 GO - BP 分析, 结果如图 4J 所示。上调的基因主要富集在 GO\_MUSCLE\_SYSTEM\_PROCESS 通路中。接下来, 我们研究了高 C0 和低 C0 轨迹之间的异质性。两条轨迹之间的共同基因包括 PTGDS、CFD、EMP1、MYC、CEBPD、PNRC1、DNAJA1 和 CEBPB(图 4K)。在高 C0 轨迹中特异性表达的基因包括 CST3、COL14A1、VIM 和 S100A4(图 4L)。其中, CST3 表达较早, 而其他三个基因表达较晚。因此, 我们选择 COL14A1、S100A4 和 VIM 来可视化它们沿轨迹的表达水平(图 4M)。空间转录组学证实了 COL14A1、S100A4 和 VIM 的最佳存在情况(图 4N)。

此外, 我们特别关注了细胞干性景观以及沿拟时间的干性基因表达。首先, 一张热图展示了随着拟时间推进干性基因的时间表达差异(图 4O)。我们还使用了 CytoTRACE2 来揭示不同成纤维细胞亚型间干细胞的功能差异, UMAP 图显示 C0 亚型主要由多能和寡能干细胞组成(图 4P)。为了验证 CytoTRACE2 的准确性, 我们在 UMAP 图上展示了不同亚型间 CytoTRACE2 的相对密度和评分密度(图 4Q)。

最后, 我们使用 Slingshot 构建了两条轨迹, 即谱系 1 和谱系 2。在谱系 1 中, 发育顺序为 C2 - C3 - C1 - C0, 并且 C0 亚型中命名的基因 MYH11 在两条轨迹的中后期均高表达(图 4R)。随后的 UMAP 图显示了两条轨迹所经过的成纤维细胞亚型范围(图 4S)。此外, 热图展示了两条轨迹的差异基因表达, 随后进行了富集分析。结果表明, 在谱系 1 后期高表达的基因包括 CHRDL2、TAGLN、TPM2、CNN1、ACTC1 和 TENM2。这些差异基因主要富集在肌肉、肾脏、组装、丝状伪足、肌原纤维和心脏等方面(补充图 4A)。接下来, 基于 UMAP 图上的不同状态构建了两条轨迹, 巧合的是, 这两条轨迹大多重叠, 其终点位于 C0 亚型所在的状态 2 和状态 4(图 4T)。最后, 干性基因 KDM5B 在谱系 1 后期高表达, CTNNB1 在谱系 1 后期也显著上调, 并且 CD44 在谱系 1 和谱系 2 的中后期均高表达(图 4U)。

(见下页图。)

图 5 CO MYH11 + 成纤维细胞通过 MDK - SDC1 串扰与肿瘤细胞相互作用。A 圆形图说明了所有成纤维细胞亚型与其他细胞类型之间相互作用的数量(顶部)和强度(底部)。B 气泡图分别显示了靶细胞的传入通信模式(顶部)和分泌细胞的传出通信模式(底部)。C 观察到肿瘤组与肿瘤相邻正常组织组之间信号通路信号强度的差异, 以及两组中细胞相互作用的数量和强度的比较。D 圆形图描绘了 CO MYH11 + 成纤维细胞与其他细胞类型的相互作用数量和强度(左)以及其他细胞与肿瘤细胞的相互作用(右)。E 展示了 MDK 信号通路中所有细胞类型之间的相互作用网络。F MDK 信号通路中所有细胞的旁分泌相互作用以及 MDK 信号通路网络的中心性得分。G 显示了沿 MDK 信号通路不同细胞类型中各种受体蛋白的表达。H 弦图展示了 CO MYH11 + 成纤维细胞与所有其他细胞类型的相互作用。I 呈现了 MDK - SDC1 通路中所有细胞之间的相互作用



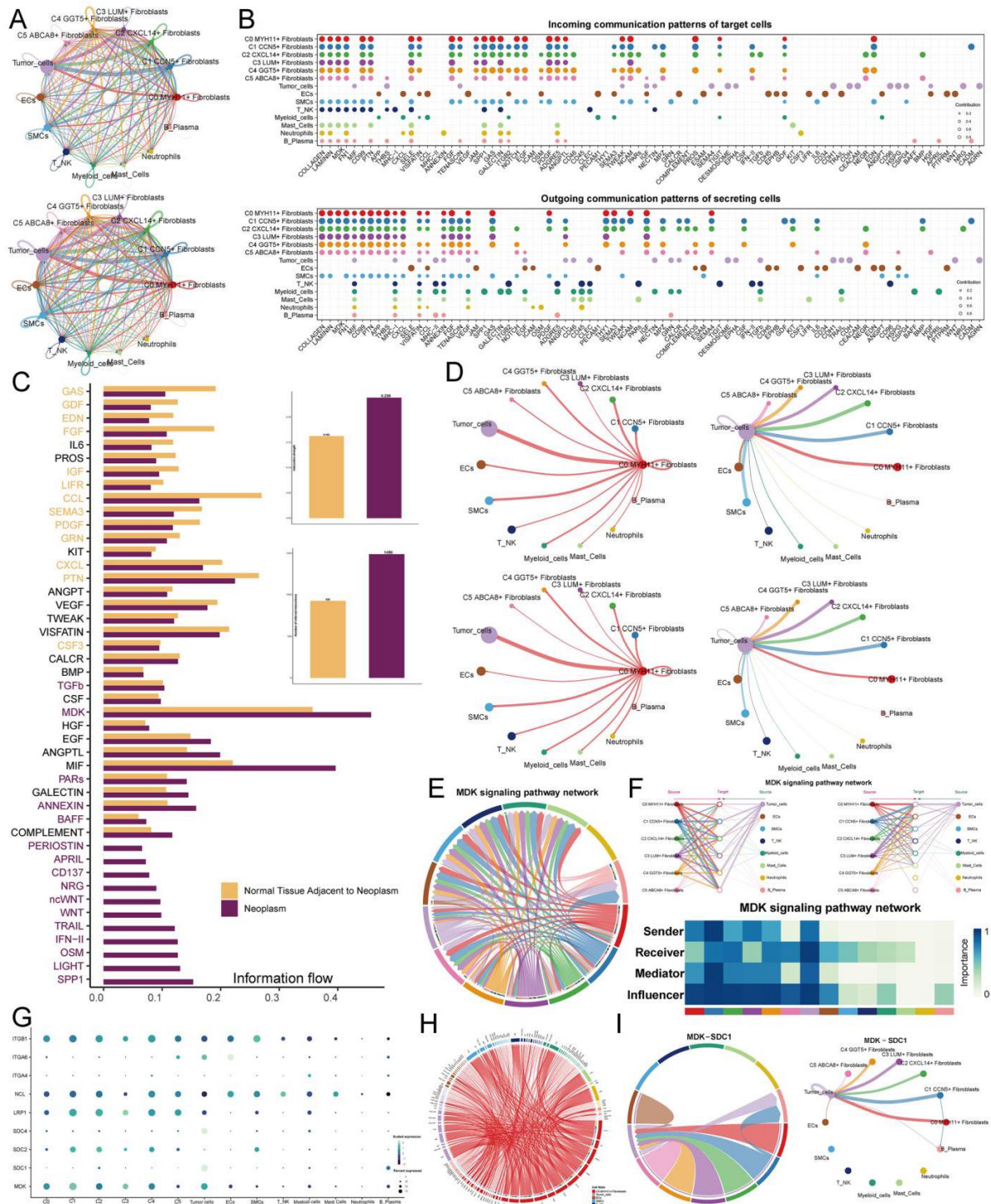


图 5(见上一页图例。)

### C0 亚型与肿瘤细胞相互作用 通过 MDK - SDC1 通路

为了解 C0 亚型与肿瘤细胞之间的相互作用，我们首先使用圆形图展示了所有成纤维细胞亚型与其他细胞之间的相互作用 (图 5A)。无论是强度还是数量，成纤维细胞亚型与各种细胞之间都存在丰富的相互作用。接下来，我们特别展示了靶细胞的传入通信模式和分泌细胞的传出通信模式 (图 5B)。显示了肿瘤组和肿瘤旁正常组织组之间各种信号通路的差异表达强度 (图 5C)。值得注意的是，TGFb、MDK、PARs、膜联蛋白、BAFF、骨膜蛋白、APRIL、CD37、NRG、TRAIL、IFN-II、OSM 等通路在肿瘤组中表达明显更

高(图 5C)。在表达强度和数量上,肿瘤组的值总是高于肿瘤旁正常组织组。CC 中关键的感兴趣亚组与肿瘤细胞之间的相互作用至关重要。为了直观展示这一点,我们使用圆形图展示了 C0 亚型与其他细胞类型之间相互作用的数量和强度(图 5D)。结果清楚地表明,C0 亚型与多种细胞相互作用,与肿瘤细胞的相互作用明显更强且更频繁。此外,另一个圆形图展示了所有细胞类型与肿瘤细胞之间的相互作用,表明成纤维细胞亚型与肿瘤细胞具有显著的相互作用强度和频率。接下来,我们可视化了两组中参与 C0 亚型 - 肿瘤细胞相互作用的受体 - 配体对(补充图 4B)。MDK 相关通路尤为突出,促使我们使用弦图展示该通路内的所有相互作用(图 5E)。我们还绘制了一个层次图,展示了 MDK 信号通路中所有细胞之间的自分泌和旁分泌相互作用,从而了解到 C0 亚型是重要的信号转导传递者,而肿瘤细胞则是其主要受体(图 5F)。MDK 信号通路网络的中心性得分分析表明,C0 亚型主要作为信号的发送者、调节者和影响者发挥作用(图 5F)。为了识别其他关键的受体蛋白,我们分别使用气泡图和弦图可视化了 C0 亚型与其他细胞类型之间的相互作用(图 5G, H)。MDK - SDC1 通路最为显著,因此我们可视化了该通路内的所有细胞相互作用。弦图和圆形图都强调了 MDK - SDC1 通路是 C0 亚型与肿瘤细胞之间相互作用的关键途径(图 5I)。此外,空间转录组学证实了 SDC1 作为受体蛋白的显著存在(补充图 4C)。

## 转录因子调控网络图谱

为了解各亚型之间转录因子(TFs)的差异,我们试图可视化它们的转录调控图谱。首先,我们展示了一个热图,显示了每个亚型的前 5 个 TFs,其中 C0 亚型的前 5 个 TFs 是 CUX1、TEAD1、HOXA10、PBX1 和 FOSL2(图 6A)。接下来,我们根据 RSS 对 C0 亚型的调控因子进行排名,并在 UMAP 图中用深绿色点突出显示。我们还基于 RAS 在 UMAP 上绘制了 C0 亚型调控最高的因子,其分布用深绿色点表示,与热图结果一致(图 6B)。UMAP 图显示 CUX1、TEAD1、HOXA10、PBX1 和 FOSL2 在 C0 亚型中均表达更高(图 6C),箱线图提供了直接比较(图 6D)。空间转录组学验证了这 5 个 TFs 可能的存在(图 6E)。

接下来,我们应用 SCENIC 识别规则和 CSI 矩阵来揭示成纤维细胞亚型的调控模块。基于 AUCell 评分,我们将调控模块分为两个主要组,即 M1 和 M2(图 6F)。当我们在 UMAP 上可视化每个模块的平均活性评分时,我们观察到模块与不同成纤维细胞亚型之间存在明显的关联(图 6G)。C0 亚组的 M1 模块得分最高,而 M2 模块得分最低(图 6H)。随后,我们对 M1 和 M2 中每个亚型的转录因子(TF)活性评分进行了排名。在 M1 中,C0 亚型的 TF 活性明显高于其他亚型,而在 M2 中,C0 的 TF 活性明显低于其他亚型(图 6I)。我们还根据各亚型间的方差解释对 M1 中的前 5 个 TF 进行了排名,发现这些 TF 在排名中处于高位(图 6J)。比较这 5 个 TF 在各亚型中的表达水平,发现它们在 C0 亚型中均有显著表达(图 6K)。最后,我们对 M1 和 M2 中差异表达的基因和 TF 进行了 GO - BP 分析。结果表明,M1 主要富集于多梳抑制复合体途径,而 M2 主要与内质网中的蛋白质加工、癌症中的转录失调、冠状病毒病(COVID - 19)、糖尿病并发症中的 AGE - RAGE 信号通路以及军团菌病等过程相关(图 6L)。

(见下一页的图。)

图 6 成纤维细胞亚型中的 TF 调控网络。A 热图展示了所有成纤维细胞亚型中前 5 个转录因子(TF)的差异表达。B 在 UMAP 图(左)上,CO MYH11 + 成纤维细胞用深绿色点突出显示。基于调控评分(RSS)显示 CO MYH11 + 成纤维细胞中调控因子的排名(中),以及 CO MYH11 + 成纤维细胞中前 5 个 TF 的表达分布。C UMAP 图说明了 CO MYH11 + 成纤维细胞中前 5 个 TF 在不同亚型中的表达强度。D 箱线图直观地比较了不同成纤维细胞亚型中前 5 个 TF 的表达水平。E ST 特征图显示了 CO MYH11 + 成纤维细胞中前 5 个 TF 的表达。F 热图说明了基于 SCENIC 识别模块和 AUCell 评分的调控子模块的相似性。确定了肿瘤组和肿瘤旁正常组织中成纤维细胞亚型的调控子模块,基于规则相似性确定了两个子模块。G UMAP 图显示了 M1 和 M2 模块在成纤维细胞亚型中的分布。H 显示了 M1 和 M2 模块的值。I 显示了 M1 和 M2 模块中各成纤维细胞亚型含量的排名。J 展示了 M1 模块中 TF 表达强度的排名。K 说明了 5 个 TF(CUX1、

TEAD1、HOXA10、PBX1 和 FOSL2) 在不同成纤维细胞亚型中的差异表达水平。L 气泡图展示了基于 M1 和 M2 模块基因集的 GO - BP 分析结果

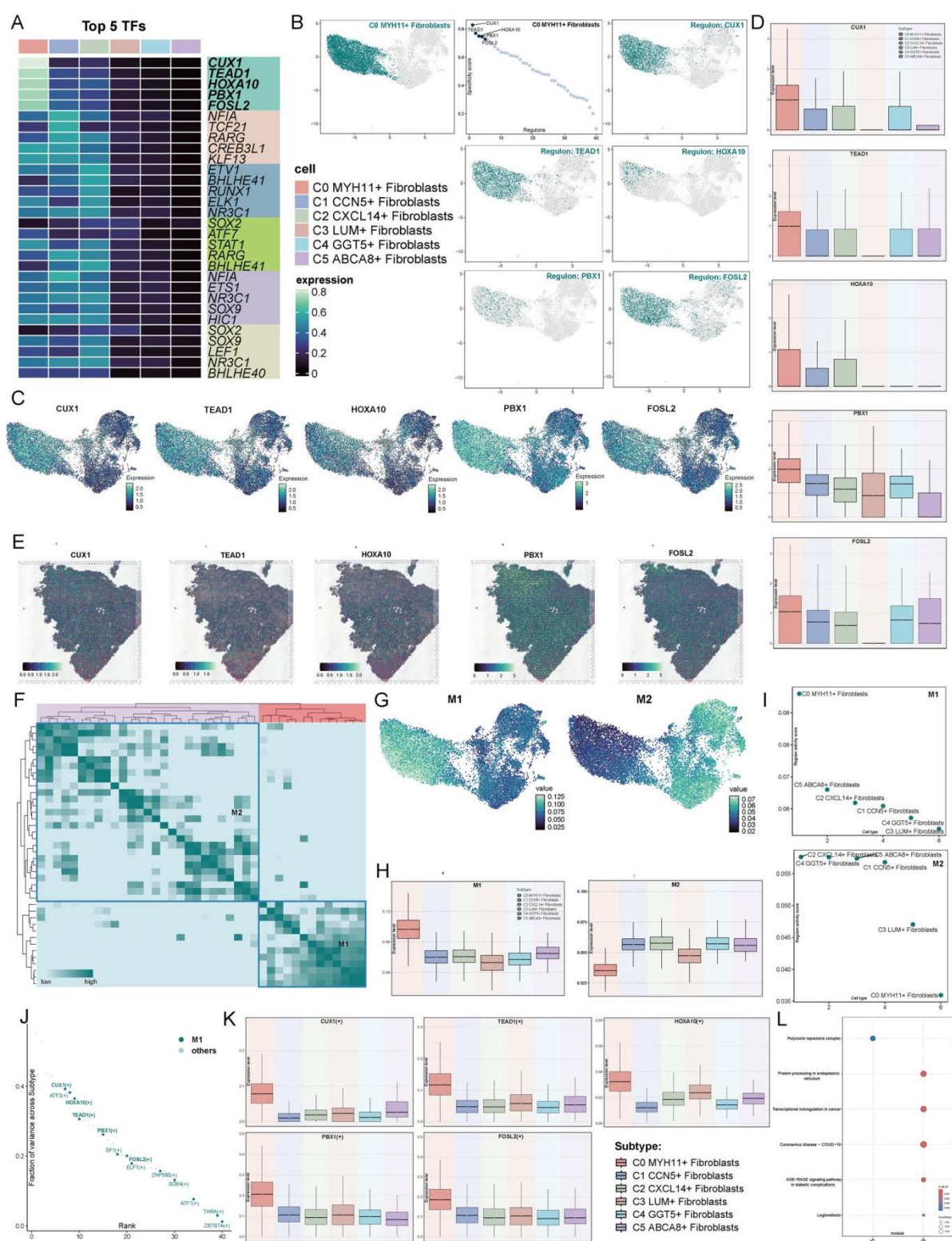


图 6(见上一頁的圖例。)



## 基于 C0 亚型的 CC 预后预测模型

鉴于 SDC1 在宫颈癌中的关键作用，我们基于 C0 亚型开发了一种宫颈癌预后模型。我们的分析确定了 10 个与患者预后显著相关的基因 (图 7A - D)。然后，我们可视化了高 MFRS 组 (MYH11 + 成纤维细胞风险评分) 和低 MFRS 组之间的生存和基因表达差异 (图 7E, F)。结果表明，模型中包含的基因具有很强的预测能力 (图 7G, H)。此外，与低 SDC1 表达组相比，高 SDC1 表达组的总生存率显著更低 ( $p = 0.013$ ) (图 7I)。为了进一步评估 MFRS 的独立预后价值，我们在 TCGA 数据库中对总生存 (OS) 进行了多因素 Cox 分析 (图 7J)。创建了一个列线图，整合了 MFRS 风险评分、年龄类别 (高和低)、种族 (美洲印第安人或阿拉斯加原住民、亚洲人、黑人或非裔美国人、白人) 和肿瘤分级 (G2、G3、GB、GX)，以预测训练队列中的 1 年、2 年和 3 年 OS (图 7K)。此外，我们使用散点图和热图评估并可视化了 10 个预后相关基因、OS 和 MFRS 之间的相关性 (图 7L)。为了进一步验证我们的发现，我们从 10 个风险基因中随机选择了 8 个基因，并比较了它们在高 MFRS 组和低 MFRS 组之间的表达水平。箱线图和散点图显示，与低 MFRS 组相比，所有 8 个基因在高 MFRS 组中的表达水平均显著更高 (图 7M)。

## 免疫浸润分析

为了研究高 MFRS 组和低 MFRS 组之间的异质性，我们对它们的肿瘤免疫微环境进行了广泛分析。最初，生成了一个热图以可视化两组中 22 种不同免疫细胞类型的分布 (补充图 5A)。箱线图显示巨噬细胞 M0、CD4 记忆静止 T 细胞和 CD8 T 细胞的比例更高。此外，与低 MFRS 组相比，高 MFRS 组显示出更多的活肥大细胞、巨噬细胞 M0 和静止 NK 细胞 (补充图 5B)。相关性分析表明，幼稚 B 细胞和浆细胞之间以及 CD8 T 细胞和活化 CD4 记忆 T 细胞之间存在很强的关系 (补充图 5C)。此外，在风险评分基因与 MFRS 组和活肥大细胞、巨噬细胞 M0、静止 NK 细胞、中性粒细胞、静止 CD4 记忆 T 细胞、 $\gamma\delta$  T 细胞和活化 NK 细胞之间发现了正相关 (补充图 5D)。我们进一步利用了

(见下页图。)

图 7 C0 MYH11 + 成纤维细胞在宫颈癌患者中的关键预后意义。A 森林图显示单因素 Cox 分析结果，确定了 13 个与宫颈癌预后相关的 C0 MYH11 + 成纤维细胞相关基因， $P < 0.05$ 。参考线 ( $HR = 1$ ) 区分了保护因素 ( $HR < 1$ ) 和危险因素 ( $HR > 1$ )。B 应用 Lasso 回归选择 10 个对风险评分有贡献的基因，结果显示在 lambda 图中 ( $\lambda_{min} = 0.017$ )。C 森林图描绘了与宫颈癌预后相关的最终 10 个基因。D 柱状图显示了 10 个风险基因的系数值。E 曲线图显示出高 MFRS 组和低 MFRS 组的风险评分，散点图显示高 MFRS 组和低 MFRS 组随时间的生存/死亡事件。F 热图显示高 MFRS 组和低 MFRS 组之间 10 个风险基因的差异表达。G ROC 曲线显示 1 年 ( $AUC = 0.80$ )、3 年 ( $AUC = 0.89$ ) 和 5 年 ( $AUC = 0.85$ ) 生存预测的 AUC 值。H 散点图显示高 MFRS 组和低 MFRS 组基因沿 PC1 和 PC2 的分布。I Kaplan-Meier 生存曲线显示高和低 SDC1 表达组之间的生存差异。J 森林图展示了训练队列中临床因素和风险评分的多因素 Cox 分析结果。K 构建基于 MFRS 的列线图模型，包括种族、年龄和分级。L 散点图和热图显示 10 个预后基因与 OS 和 MFRS 之间的相关性。M 箱线图和散点图说明了高 MFRS 组和低 MFRS 组中 10 个风险基因中的 8 个基因的表达差异，使用 CIBERSORT 和 Xcell 算法估计两个 MFRS 组中免疫浸润的比例 (补充图 5E)。



对于 PPP1R14A、DES 和 CALD1(补充图 5K)。计算了两组的肿瘤突变负荷 (TMB) 值, 并使用小提琴图可视化其分布。尽管两组之间的 TMB 值没有显著差异, 但低 MFRS 组更有可能表现出高 TMB(补充图 5L)。此外, 观察到 TMB 值与风险评分之间存在负相关(补充图 5M), 表明这些因素之间存在潜在关系。基于 TMB 中位数, 将训练队列分为高 TMB 组和低 TMB 组, 并进一步按 MFRS 组分层(补充图 5N、O)。低 TMB 组的生存率较低(补充图 5N), 与其他三组相比, 高风险 - 低 TMB 组的生存率最低(补充图 5O)。

## 关键基因的肿瘤边界特征和 Kaplan-Meier 生存曲线

首先, 我们继续利用空间转录组学生成图 8A 所示的 ST 特征图, 该图更清晰地描绘了所选切片内肿瘤边界、恶性区域和正常组织的分布。此外, 在比较两组时, 高 MFRS 组的总生存率显著低于低 MFRS 组 ( $p < 0.0001$ )(图 8B)。我们对 MYH11、KDM5B、CTNNB1、CD44、CUX1、TEAD1、HOXA10、PBX1 和 FOSL2 等关键基因进行了生存曲线分析。除 MYH11 外, 所有基因在高表达组中的生存率均低于低表达组(补充图 6A)。

我们还可可视化了 10 个预后基因的表达模式, 与之前的观察结果类似, 只有 DES 和 PCP4 在低 MFRS 组中的表达水平相对于高 MFRS 组更高(图 8C)。此外, 我们分析了预后基因在不同状态类别(如年龄、种族、分期和分级)中的分布(图 8D)。最后, 我们检查了高 MFRS 组和低 MFRS 组之间四种免疫相关药物(小白菊内酯、甲磺酸奥巴托卡、FTI.277 和 AZD8055)的 IC50 值(图 8E)。所有四种药物的 IC50 值在高 MFRS 组中较低, 而在低 MFRS 组中较高。

## 高 MFRS 组和低 MFRS 组之间的富集分析

为了进一步探索高 MFRS 组和低 MFRS 组之间的功能差异, 我们分析了两组之间的差异表达基因 (DEGs), 并检查了它们富集的途径。首先, 我们确定了高 MFRS 组和低 MFRS 组之间的 DEGs(图 9A、B)。然后, 我们进行了 GO-BP、GO-CC 和 KEGG 途径分析(图 9C-E)。在 GO-BP 分析中, 途径主要富集在与消化、中间丝组织等相关的过程中。GO-CC 分析显示在质膜的外在成分、刷状缘、基于肌动蛋白的细胞突起簇、连环蛋白复合物、中间丝细胞骨架和膜的外在成分中富集。KEGG 途径分析突出了与青少年成熟型糖尿病、Ras 信号传导、金黄色葡萄球菌感染、碳水化合物消化和吸收以及癌症中的转录失调相关的途径。

此外, 我们进行了 GSEA 分析, 结果表明上调基因主要富集在胶原分解代谢过程、伤口愈合、表皮细胞扩散、

(见下页图。)

图 8 肿瘤边界特征、预后模型差异和免疫药物治疗预测。A ST 特征图显示了所选组织切片中宫颈癌肿瘤边界 (Bdy)、恶性区域 (Mal, Mal1) 和正常组织 (Normal) 的分布。B Kaplan-Meier 生存曲线显示了高 MFRS 组和低 MFRS 组之间的生存差异。C 说明了高 MFRS 组和低 MFRS 组之间 10 个风险基因的表达差异。D 饼图显示了研究中不同状态、年龄、种族、分期和分级的比例。E 箱线图显示了高 MFRS 组和低 MFRS 组中四种免疫药物的 IC50 值内胚层形成和细胞外基质分解。相比之下, 下调基因主要富集在钾离子稳态、杀死其他生物体的细胞、轴丝组装和轴丝组装(图 9F)。

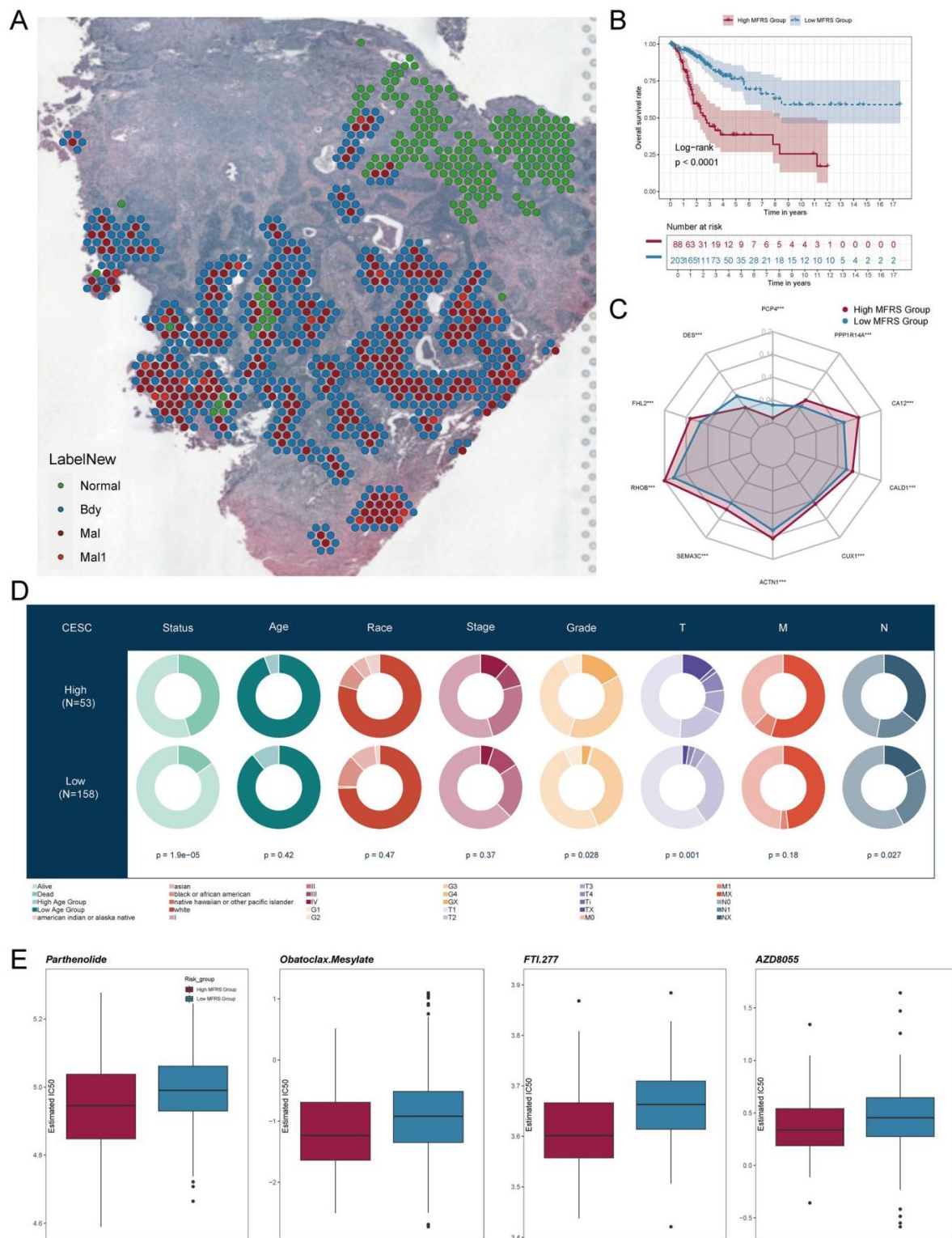


图 8(见上一页图例。)





(图 9H)。

最后，为了探究两组中的免疫细胞格局，我们发现高 MFRS 组中活化的树突状细胞、M0 巨噬细胞、活化的 NK 细胞和静息 NK 细胞更为丰富 (图 9I)。

## SDC1 促进 CC 细胞的增殖、迁移和侵袭

为了研究 SDC1 在 CC 中的具体作用，我们进行了体外实验。我们在两种 CC 细胞系 HeLa 和 SiHa 中敲低 SDC1，并观察到 SDC1 mRNA 的相对表达显著降低 (图 10A)。同样，敲低后 SDC1 蛋白表达也明显下调 (图 10B)。CCK - 8 实验表明，敲低 SDC1 导致 HeLa 和 SiHa 细胞的细胞活力显著下降 (图 10C、D)。EDU 染色实验表明，敲低 SDC1 显著降低了 CC 细胞的增殖能力 (图 10E)。集落形成实验证明，敲低 SDC1 后形成的细胞集落数量显著低于对照组，进一步证实了增殖能力的降低 (图 10F)。最后，Transwell 和伤口愈合实验均表明，SDC1 表达降低显著损害了 CC 细胞的迁移和侵袭能力 (图 10G - I)。

## 成纤维细胞异质性和空间分布的特征

为了研究肿瘤微环境中成纤维细胞的异质性和空间组织，对来自肿瘤区域 (TZ)、交界区域 (JZ) 和正常区域 (NZ) 的福尔马林固定石蜡包埋 (FFPE) 人体组织样本以及分离的正常成纤维细胞 (NFs) 和癌症相关成纤维细胞 (CAFs) 进行了多重免疫组织化学 (mIHC)。评估了一组成纤维细胞相关标志物，包括  $\alpha$  - 平滑肌肌动蛋白 ( $\alpha$  - SMA)、波形蛋白、钙调蛋白、纤连蛋白和肌球蛋白重链 11(MYH11)。

采用流式细胞术定义活的、单个的 CD45-CD31- $\alpha$ -平滑肌肌动蛋白 + 成纤维细胞群体 (图 11A)，用肌球蛋白重链 11(MYH11) 进一步区分亚群。发现与癌相关成纤维细胞 (CAF) 相比，正常成纤维细胞 (NF) 中 MYH11 的表达显著更高。相反，CAF 中钙调蛋白、波形蛋白和纤连蛋白的表达升高，而这些蛋白在 NF 中表达极少或不表达 (图 11B)。

组织切片的空间分析显示，从非肿瘤区 (NZ) 到转化区 (TZ)，成纤维细胞激活标志物的表达逐渐增加 (图 11C)。具体而言， $\alpha$  - 平滑肌肌动蛋白、钙调蛋白、波形蛋白和纤连蛋白在 NZ 中弱表达，在交界区 (JZ) 中中度表达，在 TZ 中强烈共表达。相比之下，MYH11 呈现相反的梯度，在 NZ 中表达强烈，在 JZ 中表达中等，在 TZ 中表达极少。这些发现表明，沿着 NZ-JZ-TZ 轴存在空间调节的成纤维细胞激活轨迹。

高倍和低倍的多重免疫组化 (mIHC) 图像 (图 11D) 证实， $\alpha$  - 平滑肌肌动蛋白、钙调蛋白、波形蛋白和纤连蛋白主要在 TZ 中共定位，在 JZ 中观察到部分重叠，在 NZ 中表达极少。相反，MYH11 主要在 NZ 中共定位，在 JZ 中表达减少，在 TZ 中信号可忽略不计。总体而言，这些数据揭示了一种空间上不同的成纤维细胞激活模式，强调了肿瘤微环境中表型异质性和动态基质重塑。

(见下页图。)

图 10 抑制 SDC1 可抑制宫颈癌细胞的增殖、迁移和侵袭。A 在 HeLa 和 SiHa 细胞中敲低 SDC1 后，检测宫颈癌细胞中 SDC1 mRNA 的相对表达。B 评估 SDC1 敲低后 HeLa 和 SiHa 细胞中 SDC1 蛋白的相对表达。C-D CCK-8 实验显示，敲低 SDC1 后，HeLa 和 SiHa 细胞的细胞活力显著降低。E EDU 染色实验表明，下调 SDC1 会阻碍 HeLa 和 SiHa 细胞的细胞增殖。F 集落形成实验表明，与阴性对照 (NC) 组相比，SDC1 表达降低的细胞形成的集落明显更少。G-H Transwell 实验显示，敲低 SDC1 显著降低了宫颈癌细胞的迁移和侵袭。I 划痕愈合实验表明，SDC1 表达降低显著降低了伤口愈合率。\*P<0.01，\*\*P<0.001，\*\*\*P<0.0001，\*\*\*\*P<0.0001。