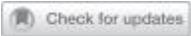


文章



分子诊断学

基于机器学习的病理组学签名可以作为一个透明细胞肾细胞癌患者新的新型预后标志物

陈思腾^{1, 4}, 江立人^{2, 4}, 冯高^{2, 4}张恩成¹陶王¹张宁³王翔¹郑俊华¹

©作者(s), 在施普林格自然有限公司2021年的独家许可下

背景：病理学家通过肉眼进行的传统组织病理学检查无法准确预测透明细胞肾细胞癌（ccRCC）的生存。
方法：回顾性分析来自3个患者队列的483张全载玻片图像（WSIs）数据。我们采用机器学习算法来识别最优的数字病理特征，并为ccRCC患者构建了基于机器学习的病理组学签名（MLPS）。预后模型的预后性能也在两个独立的验证队列中得到了验证。
结果：MLPS能显著区分高生存风险的ccRCC患者，在三个独立队列中，危险比分别为15.05、4.49和1.65。Cox回归分析显示，MLPS可作为ccRCC患者的独立预后因素。基于MLPS、肿瘤分期系统和肿瘤分级系统的整合列线图提高了ccRCC患者当前的生存预测精度，1、3、5、10年无病生存预测的曲线下面积分别为89.5%、90.0%、88.5%和85.9%。
讨论：基于机器学习的病理组学特征可以作为ccRCC患者的一种新的预后标志物。然而，对多中心患者队列的前瞻性研究仍需要进一步验证。
英国癌症杂志（2022）126：771-777；https://doi.org/10.1038/s41416-021-01640-2

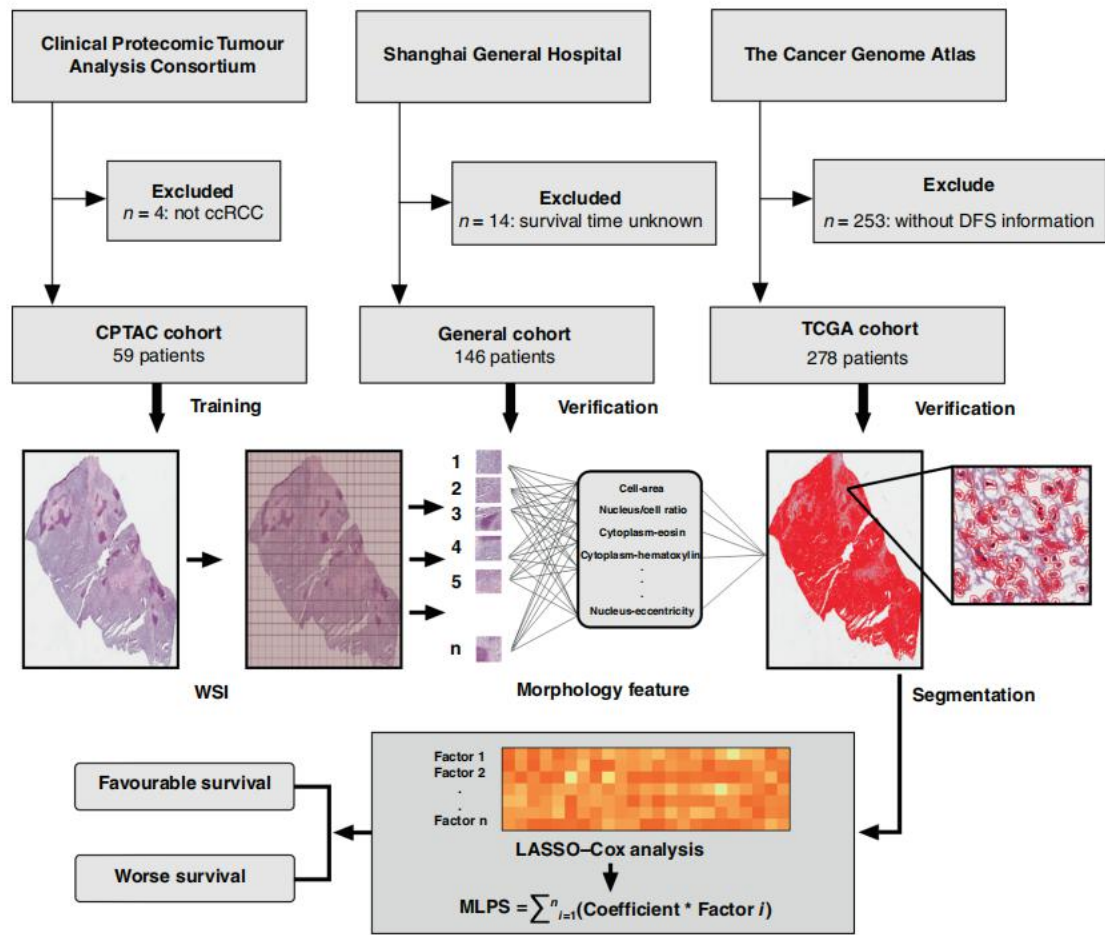
背景
肾癌是2021年美国男性第六大常见恶性肿瘤。在多种肾癌中，透明细胞肾细胞癌（ccRCC）占[2]的主要病例。不同分期的ccRCC患者的预后不同，部分低期肿瘤患者即使在[3, 4]根治性肾切除术后治愈率也较低。TNM分类是最经典的预后参数，但仍有局限性，因为相同TNM分期的患者可能表现出不同的临床结果[5, 6]。此外，国际泌尿系统病理学学会（ISUP）分级系统主要针对组织学核形态学[4, 7]。因此，一个更全面的预后预测可能是一个有趣的话题。
由于新的高通量技术，组织学特征识别的人工智能在临床结果预测[8, 9]方面显示出巨大的潜力。通过对组织学形态学的综合分析，计算病理学被广泛应用于各种恶性肿瘤的诊断和生存预测[10-14]。然而，这些研究大多是针对传统的光学显微镜，这可能会导致

与全幻灯片成像（WSI）相比，视图问题较窄，在不同的放大[15]下可以显示肿瘤组织的组织和细胞结构。
在本研究中，我们从WSIs中构建了一个基于机器学习的病理组学签名（MLPS）来分析和预测ccRCC患者的临床结果，这可能为临床医生提供促进个性化治疗过程的个体化治疗的智能建议。

方法
患者队列和数据资源
本研究中招募的患者队列来自三个独立的数据来源——临床蛋白质组学肿瘤分析联盟（CPTAC队列）；上海总医院（普通队列）和癌症基因组图谱（TCGA队列）。我们从上海总医院招募了146名ccRCC患者，他们在2012年1月至2018年12月期间接受了根治性或部分肾切除术。所有纳入的患者均应符合以下纳入标准：（i）病理诊断为ccRCC，无其他类型的恶性肿瘤；（ii）有未经摘要的临床和病理信息和无病生存期（DFS）随访数据；（iii）可获得相应的肿瘤样本。DFS被定义为时间从

¹上海交通大学医学院上海总医院泌尿科，中国上海。²上海交通大学医学院上海总医院病理科，中国上海。³上海交通大学医学院瑞金医院泌尿科，中国上海。⁴ 翔教育中心
这些作者贡献相同：陈贤、李仁、丰高。电子邮件：zn12235@rjh...wangl@shgh.cn；郑jh0471@sjtu..

收到：2021年5月5日 修订：2021年10月26日 接受：2021年11月10日
在线发布时间：2021年11月25日



基于机器学习的分析管道

图1. 本研究中组织病理学图像处理和机器学习分析的工作效率。WSI整个幻灯片图像。

初次手术，首次肿瘤复发，进展，死亡或最后一次随访审查患者。对肿瘤标本进行副纤维包埋，切片制作福尔马林混合和副纤维包埋（FFPE）切片，进一步进行苏木精-伊红（H&E）染色。我们的研究已获得上海总医院研究伦理委员会的伦理批准。

从CPTAC数据库中共获得59例ccRCC患者，可获得组织病理学H&E图像和DFS随访信息（<https://proteomics.cancer.gov/programs/cptac>）数据库和癌症成像档案[16]，它遵守了知识共享署名3.0未移植许可。另外278例符合上述前两个标准的ccRCC患者，并开放获取来自TCGA的组织病理学H&E图像，也被纳入本研究。从TCGA数据库中下载了原始的H&E图像和详细的临床资料，包括DFS随访数据（<https://portal.gdc.cancer.gov/>）。

苏木精伊红染色

FFPE切片首先用苏木精（西格玛-奥德里奇）和1%酸乙醇（西格玛-奥德里奇）染色，然后在蒸馏水中冲洗。其次，切片用伊红（Sigma-Aldrich）染色3 min，并进一步用分级酒精脱水。所有h&e染色的切片均通过徕卡AperioAT2（徕卡生物系统公司，法兰克福，德国）进行WSI扫描。

整个幻灯片图像处理和管道

将未经任何预处理的H&E染色切片的WSIs快速分割成瓷砖，然后导入QuPath数字病理软件[17]中进行检测，进行细胞和核分割。通过应用分析模块中的细胞检测功能，使细胞核

基于分割参数[18]，通过流域细胞检测进行对象，包括细胞、核和细胞质的形态特征。设置参数设置为苏木精OD，像素大小为0.5 μm 。对于核参数的背景半径和sigma分别设置为8 μm 和1.5 μm 。强度参数的阈值为0.1，最大背景强度设为2。其他参数也被设置为它们各自的默认值。在wsi的过程中也应用了平滑边界和参数测量。随后，对每个分段瓷砖中的每个被识别的物体，共检测并测量了43个不同的病理特征（表S1）。

我们构建了一个分析管道来分析在训练队列（CPTAC队列）中不同临床应用中检测到的H&E图像特征。我们通过glmnet软件包[19]进行最小绝对收缩和选择算子，以识别最优的数字病理特征，并计算MLPS模型中每个特征的系数。对于MLPS，十倍交叉验证的误差和不同程度的惩罚进展如图所示。S1和S2。组织病理学图像处理和管道的工作效率如图所示。1。

统计分析

在这项研究中，r3.6.2用于进行数据分析和可视化。采用Kaplan-Meier（KM）曲线分析，采用比（lag-rank），确定不同的临床结果。采用单因素和多因素Cox回归分析来确定ccRCC患者的预后标志物。通过受试者工作特征（ROC）曲线和曲线下面积（AUC）值来评价预测模型的特殊性和敏感性。

表1. 来自三个独立患者队列的患者的基本临床特征。

| | CPTAC队列 (59) | 将军 队列 (146) | TCGA队列 (2 78) |
|---|-----------------|----------------|------------------|
| 年龄 (年) | | | |
| ≥65 | 28 (47.5%) | 45 (30.8%) | 93 (33.5%) |
| <65 | 31 (52.5%) | 101 (69.2%) | 185 (66.5%) |
| 性别 | | | |
| 男 | 37 (62.7%) | 115 (78.8%) | 182 (65.5%) |
| 女性的 | 22 (37.3%) | 31 (21.2%) | 96 (34.5%) |
| 等级 | | | |
| G1 | 3 (5.1%) | 14 (9.6%) | 7 (2.5%) |
| G2 | 37 (62.7%) | 79 (54.1%) | 119 (42.8%) |
| G3 | 15 (25.4%) | 48 (32.9%) | 116 (41.7%) |
| G4 | 4 (6.8%) | 5 (3.4%) | 36 (13.0%) |
| 阶段 | | | |
| i | 30 (50.8%) | 128 (87.6%) | 134 (48.2%) |
| ii | 6 (10.2%) | 9 (6.2%) | 29 (10.4%) |
| iii | 18 (30.5%) | 9 (6.2%) | 73 (26.3%) |
| iv | 5 (8.5%) | 0 | 42 (15.1%) |
| T阶段 | | | |
| T1 | 32 (54.2%) | 128 (87.6%) | 137 (49.3%) |
| T2 | 7 (11.9%) | 9 (6.2%) | 36 (12.9%) |
| T3 | 19 (32.2%) | 9 (6.2%) | 103 (37.1%) |
| T4 | 1 (1.7%) | 0 | 2 (0.7%) |
| N级 | | | |
| N1 | / | 2 (1.4%) | 8 (2.9%) |
| N0 | / | 144 (98.6%) | 270 (97.1%) |
| M期 | | | |
| M1 | / | 0 | 41 (14.7%) |
| M0 | / | 146 (100%) | 237 (85.3%) |
| DFS状态 | | | |
| 是的, 45 (76.3%) 131 (89.7%) 185 (66个。5%) | | | |
| 不 | 14 | 15 | 93 (33.5%) |

CPTAC临床蛋白质组学肿瘤分析联盟, TCGA癌症基因组图谱, DFS无病生存期。

结果

临床和病理特征

CPTAC队列 (n = 59)、一般队列 (n = 146) 和TCGA (n = 278) 队列的临床和病理特征见表1, 平均随访时间为26.4 ± 16.8、54.9 ± 分别为27.8和43.2 ± 30.6个月。在所有三个队列中, 男性患者所占比例均高于女性 (62.7%、78.8%和65.5%)。在三个队列中也观察到更多的G2/G3肿瘤患者 (88.1%、87.0%和84.5%)。而有淋巴转移或远处转移的患者在这些队列中很少见。

在套索分析中的预后图像因素和

MLPS分类系统的开发

如图所示。S1, 左竖线等于最小十倍交叉验证误差达到5, 表明ive图像因子被筛选为ccRCC患者最多的预后因素。所选图像因子包括核圆度、核最小卡尺、核苏木素OD平均值、核苏木素OD最小和细胞伊红OD标准差。

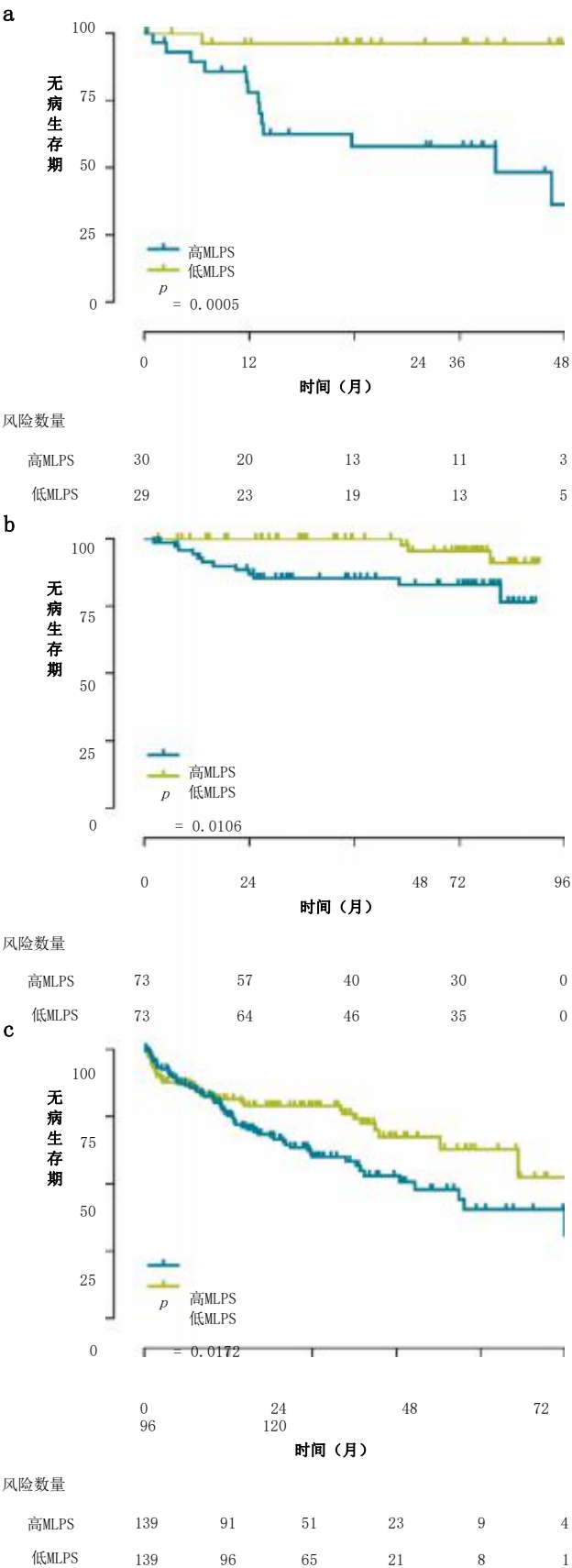


图2. MLPS预测ccRCC患者无病生存期的Kaplan-Meier生存期分析。a CPTAC队列; b一般队列; c TCGA队列。ccRCC透明细胞肾细胞癌, 基于MLPS机器学习的病理组学签名, CPTAC临床蛋白质组学肿瘤分析联盟, TCGA癌症基因组图谱。

表2. 单变量 和 多变量cox回归 三个独立的队列的预后标志物分析。

| | | 单变量、多变量 | | | |
|----------|----------------|-----------------------|---------|--------------------|---------|
| | | HR (95% CI) p | | p | |
| CPTAC队列 | 等级 | | | | |
| | G3/4 vs G1/2 | .681 (0.58 - 4.85) | 0.337 | | / |
| | iii/iv vs i/ii | 9.94 (2.22 - 44.52) | | | 0.013 |
| | 男性与女性 | 0.81 (0.27 - 2.43) | 0.711 | | / |
| | ≥65vs<65 | 0.64 (0.21 - 1.90) | 0.419 | | / |
| | 高与低 | 15.28 (2.00 - 117.05) | 0.009 | | 0.026 |
| 一般队列 | 等级 | | | | |
| | G3/4 vs G1/2 | 11.15 (3.04 - 40.92) | | | 0.014 |
| | ii/iii vs i | 7.90 (2.85 - 21.87) | | | 0.003 |
| | 男性与女性 | 29.28 (0.15 - >100) | | | / |
| | ≥65vs<65 | 2.79 (1.01 - 7.74) | | | 0.026 |
| | 高与低 | 4.5 (1.27 - 15.97) | 0.02 | | 0.003 |
| TCGA队列等级 | | | | | |
| | G3/4 vs G1/2 | 3.11 (1.92 - 5.02) | <0.0001 | 1.99 (1.21 - 3.27) | 0.007 |
| | 阶段 | | | | |
| | iii/iv vs i/ii | 5.82 (3.63 - 9.33) | <0.0001 | 4.92 (3.02 - 8.04) | <0.0001 |
| | 性别 | | | | |
| | 男性与女性 | .291 (0.83 - 2.00) | 0.261 | / | / |
| | 年龄 | | | | |
| | ≥65vs<65 | 1.10 (0.72 - 1.69) | 0.656 | / | / |
| | 百万分之一 | | | | |
| | 高与低 | 1.65 (1.09 - 2.51) | 0.018 | 1.78 (1.17 - 2.71) | 0.007 |

HR风险比, CI证据间隔, CPTAC临床蛋白质组学肿瘤分析联盟, 基于MLPS机器学习的病理组学签名, TCGA癌症基因组图谱, DFS无病生存期。

粗体值表示有统计学意义, $p < 0.05$ 。

从图中的套索分析中提取每个选定图像因子的回归系数(β)。S2(β 核周长= -23.8919605, β 核最小卡尺= -0.6106737, β 苏木精OD平均核= -1.7673163, β 苏木精OD最小核= -2.1720564, β 细胞伊红OD std dev= -7.2948452)。然后建立公式MLPS: MLPS=核循环* (-23.8919605) +核Min caliper* (-0.6106737) +核苏木素OD mean* (-1.7673163) +核苏木素OD min* (-2.1720564) +细胞伊红OD std dev* (-7.2948452)。

生存分析和Cox回归分析基于

在独立的队列中的MLPS分类

以MLPS的临界值为每个患者队列的中位数, 将每个队列的患者分为高MLPS组和低MLPS组。分为高或低MLPS的CPTAC队列的KM生存曲线如图所示。2a. 与低MLPS患者相比, 高MLPS亚组患者的临床预后明显更差(HR = 15.05, 95% CI: 5.27 - 43.04, $p = 0.0005$)。在一般队列和TCGA队列中进一步验证KM生存曲线分析。MLPS分类系统能够

在两个独立队列中区分与较差DFS相关的患者, 在一般队列中HR为4.49 (95%CI: 1.63-12.38, $p=0.0106$)。2b)和1.65(95%CI: 1.10 - 2.48, $p=0$)。在TCGA队列中(图. 2c)。

Cox回归分析结果见表2。在CPTAC队列中, 肿瘤分期系统和MLPS分类系统是ccRCC患者DFS的独立预后因素。在一般队列中, 肿瘤分期系统、肿瘤分级系统、患者年龄和MLPS分类系统被证明与ccRCC患者的DFS独立相关。而在TCGA队列中, 肿瘤分期系统、肿瘤分级系统和MLPS分类系统被发现是独立的预后因素。在CPTAC和TCGA队列中, MLPS分类系统的HR比其他队列都要大。

积分列线图提高了当前的生存率

对ccRCC患者的预测准确性

为了提高目前ccRCC患者的生存预测精度, 我们建立了基于MLPS分类系统和可预测的临床和临床和可预测的整合列线图

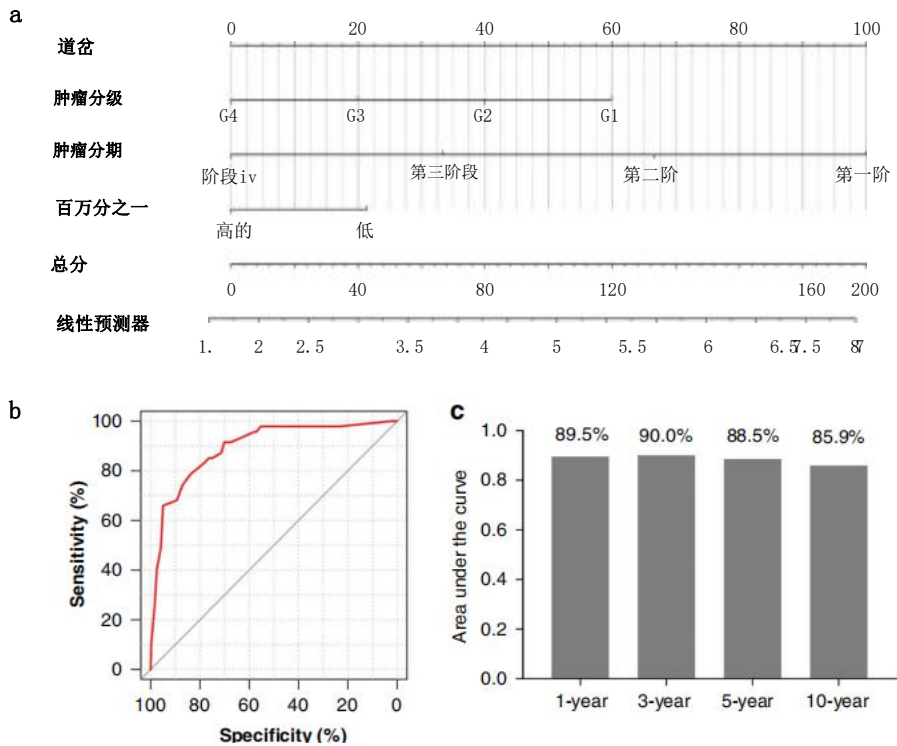


图3制定并评估了整个患者队列的预后列线图。a基于MLPS和临床病理因素的ccRCC患者无病生存预测的列线图。b ROC分析，用于评估整个患者队列中列线图的1年无病生存预测。cROC分析的曲线下面积，用于评价1、3、5、10年无病生存预测。ccRCC透明细胞肾细胞癌，ROC受试者工作特征，基于MLPS机器学习的病理组学特征。

表3. 受试者工作特征曲线分析，用于评估三个独立患者队列中预后模型的准确性。

| | | AUC 95% CI p | | |
|----------|-----|--------------|---------------|---------|
| CPTAC队列 | 1年 | 0.952 | 0.862 – 0.991 | <0.0001 |
| | 3年 | 0.941 | 0.848 – 0.986 | <0.0001 |
| | 5年 | 0.864 | 0.750 – 0.940 | <0.0001 |
| | 十年 | 0.864 | 0.750 – 0.940 | <0.0001 |
| 一般队列 | 1年 | 0.896 | 0.835 – 0.940 | <0.0001 |
| | 3年 | 0.923 | 0.867 – 0.960 | <0.0001 |
| | 5年 | 0.898 | 0.837 – 0.942 | <0.0001 |
| | 十年 | 0.814 | 0.741 – 0.874 | <0.0001 |
| TCGA队列1年 | | 0.871 | 0.826 – 0.908 | <0.0001 |
| | 3年 | 0.878 | 0.833 – 0.914 | <0.0001 |
| | 5年 | 0.877 | 0.833 – 0.913 | <0.0001 |
| | 10年 | 0.853 | 0.806 – 0.892 | <0.0001 |

AUC曲线下面积，CI证据区间，CPTAC临床蛋白质组学肿瘤分析联盟，TCGA癌症基因组图谱。

病理因素(图。3a)在整个患者队列中。ROC曲线分析表明，1年、3年、5年和10年DFS预测的AUC值分别为89.5%、90.0%、88.5%和85.9%，显示了积分列线图的良好性能。3b, c)。进一步的亚组分析也显示，在这三个独立的cccCC患者的预后列线图预测具有很高的准确性(表3)。

讨论
传统的病理图像的显微视图可以通过细胞结构，特别是重要的核结构来区分。然而，组织结构，如在周围的间充质细胞中的肿瘤内滤液，是缺失的。WSIs包含多种组织学特征，不仅包括细胞信息，还包括组织排列[15]。此外，WSIs可以根据肿瘤内滤在深度和范围上预测ccRCC的pT分期，从而获得更好的预后价值。因此，WSI上的整合病理特征表明其与传统图像显微视角不可比拟的优势。
人工智能技术的出现严重影响了传统的病理诊断和预后预测[20, 21]。随着WSIs的不断应用，多种癌症的自动组织学分析显示了其在诊断和临床结果预测[22, 23]方面有前景的蓝图。此外，通过整合WSIs的信息，病理组学极大地提高了计算病理学[15]的性能。然而，关于ccRCC中WSIs计算病理的研究目前鲜有报道。
在本研究中，我们通过应用WSIs，构建了一个基于机器学习的cc-cc-RCC患者生存预测预后模型。MLPS在三个独立的患者队列中显示出较高的预测准确性。在不同的时间尺度中，1年和3年无病生存预测的准确性最高。Cox回归分析结果表明，MLPS可作为ccRCC患者DFS的独立预后因素。整合列线图结合MLPS和临床病理因素，提高了目前的生存预测的准确性，1年、3年、5年和10年DFS预测的AUC值分别为89.5%、90.0%、88.5%和85.9%，显示了积分列线图的良好性能。

所选的图像因子包括核圆度、核最小卡尺、核苏木素OD均值、核苏木素OD最小和细胞伊红OD std dev, 主要与细胞核的形状和染色深度有关, 反映了癌细胞核异形性与临床预后的密切联系。因此, 我们基于机器学习的预测模型可能会研究细胞核的组织病理学表现, 并区分与ccRCC患者不良临床预后显著相关的特征。

TNM分类作为最经典的预后评估系统, 已被用于预测所有癌症患者[5, 24]的生存和辅助治疗设计。TNM分期的标准主要是关于宏观特征的。然而, 同一器官内不同的病理诊断也可能表现出不同的预后, 如高级别ccRCC和高级别集合管癌均属于肾细胞肿瘤, 但[2]预后不同。即使是相同级别的ccRCC患者, 由于细胞上的差异, [2, 5, 24]也可能表现出不同的预后。

此外, 基于ISUP分级系统的ccRCC核异常, 证明了其对ccRCC患者[4]的生存预测的价值。较高的分级与较差的临床结果[4]相关。然而, ISUP分级系统并不评价宏观特征。在临床实践中, 临床医生通常结合TNM和ISUP分类, 以获得更好的治疗决策。在本研究中, 应用WSIs和临床病理因素的结合可以帮助临床医生更好地进行可视化的生存预测。

本研究存在一定的局限性。首先, 将MLPS的临界值设置为三个队列之间的中位数, 而不是混合值。此外, 尽管我们的预测模型在区分高生存风险患者方面表现良好, 但该研究是回顾性的, 该系统的临床应用需要进一步的前瞻性验证。此外, 考虑到不同患者队列中不同的H&E染色方案, WSIs可能表现出一定的颜色异质性, 这可能会影响分析。

结论

总之, 在本研究中, 我们开发并验证了一种基于来自WSIs的机器学习方法的病理组学签名。MLPS可作为ccRCC患者良好的生存预测因子, 有助于临床决策。然而, 对多中心患者队列的前瞻性研究仍需要进一步验证。

数据可用性

支持本研究内容的数据可在补充信息中获得, 也可在合理的要求下从作者处获得。

代码可用性

用于幻灯片图像处理和特征提取的代码已经在<https://qupath.github.io/> [17]上公开提供。

参考文献

1. 西格尔, 米勒, 福克斯, 杰玛尔A. 癌症统计, 2021年。CA癌症J Clin. 2021;71:7-33.
2. 汉弗莱, 莫奇, 库比拉, 乌尔布赖特, 路透社。2016年世界卫生组织泌尿系统和男性生殖器官肿瘤分类-B部分: 前列腺和膀胱肿瘤。Eur Urol. 2016;70:106-19.
3. 刘磊, 齐L, 李Y, 陈M, 李C, 程X, 等。中度复杂腹膜肿瘤的肾部分切除术: 中国单一机构的手术技术和三组结果。J腹腔镜检查技术A部分。2017;27:812-7.

4. 易世、肖问、曾福、尹赫、李子、钱C等。计算机断层扫描放射组学预测肾细胞癌的病理分级。前面的协议。2020;10:570396.
5. 内村K, 藤原T, 和田Y, 异香川T, 琼斯MH, 维尔塔宁C, 等。通过分层聚类和非负矩阵分解, 确定了具有不同基因表达谱和预后的肺鳞状细胞癌的两个亚类。致癌基因。2005;24:7105-13.
6. 陈思, 张N, 江伦, 高F、邵杰、王T等。机器学习组织病理学图像特征在透明细胞肾细胞癌诊断和生存预测中的临床应用。Int J癌症。2021;148:780-90.
7. 德拉亨特B, 切维尔JC, 马蒂诺尼, 汉弗莱, 马加卢齐, 麦肯尼, 等。国际泌尿系统病理学会 (ISUP) 肾细胞癌分级系统和其他预后参数。我是Surg Pathol. 2013;37:1490-504.
8. 克里莫夫, 米美美, 格蒂奇, 江Y, 托斯女士, Rida P, 等。一种基于全幻灯片图像的机器学习方法来预测导管原位癌 (DCIS) 的复发风险。乳腺癌Res. 2019;21:83.
9. 路易斯DN, 戈伯GK, 男爵JM, Bry L, Dighe AS, Getz G, 等。计算病理学: 一种新兴的缺陷。拱形Pathol实验室医学。2014;138:1133-8.
10. Gertych A, 内格N, Ma Z, Fuchs TJ, 萨尔曼S, 莫汉蒂S, 等。用机器学习的方法来分析根治性前列腺切除术组织中组织的组织学图像。Comput医学成像图。2015;46:197-208.
11. B, B, B, B, B, 等, 等。深度学习算法检测乳腺癌女性淋巴结转移的诊断评估。JAMA. 2017;318:2199-210.
12. Yu KH, 张C, 贝瑞GJ, Altman RB, Re C, Rubin DL, 等。利用全自动显微镜病理图像特征预测非小细胞肺癌的预后。纳特Commun. 2016;7:12474.
13. 踢球者P, 伊森西F, 图尔苏诺娃I, 彼得森J, 纽伯格U, 博内坎普D, 等。利用人工智能神经网络评估神经肿瘤学中MRI的自动定量肿瘤反应: 一项多中心、回顾性研究。《柳叶刀》。2019;20:728-40.
14. 布林克TJ, 赫克勒A, Enk AH, 伯金C, 哈弗坎普S, 豪斯查尔德A, 等。深度神经网络在黑色素瘤图像分类方面优于皮肤科医生。Eur J癌症。2019;119:11-7.
15. 曹瑞、杨福、马思、刘乐、赵洋、李等。一种基于病理组学的结肠直肠癌微卫星不稳定性预测模型的开发和解释。治疗学。2020;10:11080-91.
16. 克拉克K, Vendt B, 史密斯K, 弗雷曼J, Kirby J, Koppel P, 等。癌症成像档案 (TCIA): 维护和操作一个公共信息存储库。J数字成像。2013;26:1045-57.
17. 班克海德P, 劳夫雷MB, 费尔南德斯JA, 东布罗夫斯基, M卡特DG, 邓恩PD等。QuPath: 用于数字病理图像分析的开源软件。科学代表。2017;7:16878.
18. 库尔卡尼总理, 罗宾逊EJ, 沙林普拉丹J, 加特雷尔-科拉多路, Rohr BR, Trager MI等。基于原发性黑色素瘤肿瘤的标准H&E图像的深度学习可以识别有内脏复发和死亡风险的患者。克林癌症Res. 2020;26:1126-34.
19. 弗里德曼, 哈斯蒂, 蒂希希拉尼。广义线性模型的正则化路径。J Stat软w. 2010;33:1-22.
20. 舒特福特虚拟机, 广场B, 溜冰场M, 竞争E, 伊斯兰教旧金山。泌尿外科的病理学。库尔Opin Urol. 2020;30:823-31.
21. 姜勇、金C、于H、吴J、陈C、袁Q等。开发和验证一种深度学习CT信号来预测胃癌的生存和化疗: 一项多中心、回顾性研究。安Surg. 2021;274: e1153-61.
22. Esteva A, Kuprel B, Novoa RA, Ko J, 湿特SM, Blau HM, 等。深度神经网络对皮肤癌的皮肤病水平分类。自然2017;542:115-8.
23. 赵、李、姚、王Y、吴X、徐Z等。人工智力定量肿瘤间质比率是可切除结肠直肠癌总生存率的独立预测因子。生物医学。2020;61:103054.
24. 啤酒DG, Kardias SL, 黄CC, 佐丹奴TJ, Levin AM, Misek DE, 等。基因表达原体能预测肺癌患者的生存期。Nat地中海。2002;8:816-24.

致谢

我们感谢QuPath数字病理学软件的贡献。本出版物中使用的开放数据由美国国家癌症研究所临床蛋白质组学肿瘤分析联盟 (CPTAC) 和TCGA数据库生成。

作者贡献

NZ、XW、JZ和SC参与了研究的构思和设计、数据收集、数据分析，审阅了论文并批准了inal草案提交。LJ、FG、EZ和TW参与了数据收集，审阅了论文，并批准了inal草案提交。

资金信息

国家自然科学基金资助项目（81972393和82002665）。资金来源在研究的设计中没有任何作用；收集、分析或解释数据；撰写报告；提交发表的决定。

伦理批准和同意参与

本研究的伦理批准已获得上海总医院研究伦理委员会的伦理批准（批准号：2021SQ121）。不需要进一步的伦理批准，因为所有的切片图像来自CPTAC队列和TCGA队列，都是公开的，可用于研究目的。

同意出版

不适用。

竞争利益

作者声明没有任何相互竞争的利益。

附加信息

补充信息在线版本包含补充材料，可在[https://doi获得。org/10.1038/s41416-021-01640-2](https://doi.org/10.1038/s41416-021-01640-2)。

信件和材料要求应寄给张宁、王翔或郑俊华。

重印和许可信息可在<http://www>上获得。[自然com/重印](#)

出版商的说明《施普林格自然》对已出版的地图和机构机构中的管辖权主张保持中立。