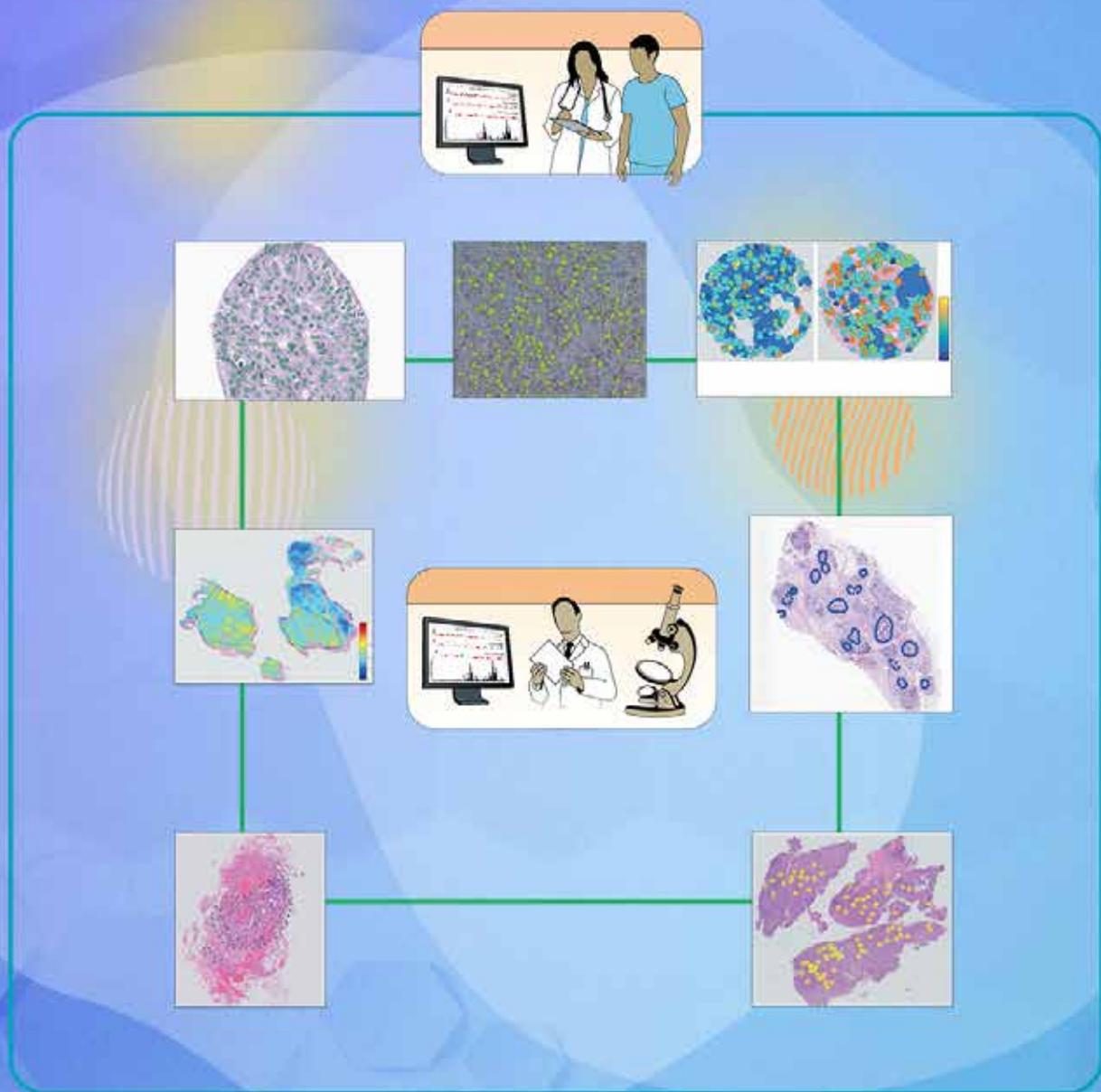


生命奥秘

总 128 期 / 2020/9

LIFEOMICS

人工智能应用于肿瘤
诊疗的最新进展



无奇不有

生命世界

解读生命

走进科学

目录 CONTENTS

专题: 人工智能应用于肿瘤诊疗的最新进展

前言	01
一、人工智能应用于肿瘤诊疗领域的概况	02
1. 人工智能应用于肿瘤诊疗的主要领域	02
2. 相关领域的重大进展	02
二、用于RNA测序试验的新型人工拼接基因内部参照物	05
1. 基于人工特征的方法	05
2. 基于神经网络的方法	07
3. 人工智能方法中病理学和肿瘤学的整合	10
三、人工智能在肿瘤筛查中的应用——以一项跨国乳腺癌筛查研究为例	13
1. 乳腺癌筛查评估人工智能系统的构建	14
2. 乳腺癌预测中人工智能系统和临床工作者的比较	16
四、讨论	18
1. 人工智能在肿瘤诊疗领域临床应用的前景	18
2. 人工智能在肿瘤诊疗领域临床应用的挑战	19

专题

人工智能应用于 肿瘤诊疗的最新进展

特约编辑：叶景佳，女，博士，研究方向：肿瘤生物治疗及发病机制

前言

近十年，精准肿瘤学的进展大大增加了预测性分析的需求，这种分析使对患者进行筛选和分层治疗成为可能。虽然介导细胞间相互作用的信号传导和转录网络的繁复差异，使基于单个基因或蛋白质功能进行肿瘤标记物开发的研究更加复杂，但组织切片图像的数字化使人工智能(Artificial Intelligence, AI)和机器学习(Machine Learning, ML)工具可应用于数字病理学，这使亚视觉形态学表型的数据挖掘成为可能，并有望最终改善患者管理。本文对各种基于人工智能的数字病理计算方法进行了综述，将人工智能和机器学习工具纳入临床肿瘤学，并讨论了与使用人工智能相关的一些挑战，以及精准肿瘤学未来发展的机遇。

一、人工智能应用于 肿瘤诊疗领域的概况

1. 人工智能应用于 肿瘤诊疗的主要领域

数字病理学包括将组织病理切片扫描成数字化的图像，并用计算处理对这些数字化图像进行分析的过程。这个概念最初是指使用先进的全玻片数字扫描技术将病理切片图像数字化的过程，现在还包括使用基于人工智能(AI)的方法，对这些数字化的全玻片图像(Whole slide images, WSI)进行检测、分割、诊断和分析的过程。

在20世纪50年代，人工智能(AI)指的是计算机科学的一个分支，在这个分支中，尝试用机器的方法预测——模拟一个智能人在同样的情况下可能会做什么。基于机器学习(ML)的方

法，包括从输入数据中进行机器“学习”以做出预测，这属于人工智能的大范畴。深度学习(Deep Learning, DL)则是基于人工神经网络的进步而发展起来的一种特殊的ML方法。DL网络通常包括多层人工神经网络，并倾向于包括输入层、输出层和多个隐藏层。有趣的是，隐藏层还用于生成新的图像表征，并且通过足够的训练实例，可以用于识别最能区分所关注类别的表征。在过去的15年里，随着多个研究团队的研究成果和计算处理能力的提高，人们越来越关注数字语言在包括数字病理在内的不同领域中的应用。

2. 相关领域的重大进展

在过去数十年中，技术的进步使全玻片图像的高效数字化成为可能，从而简化了世界各地病理实验室的病理工作流程。而玻片数字化

又帮助人们创建了大规模的数字玻片库(其中最受欢迎的可能是癌症基因组Atlas)，使世界各地的研究人员都能够自由地访问一个经过充

分整理和注释的，与临床、疗效和基因组信息相关的病理学图像数据集，这反过来也促进了大量数字病理学和肿瘤学研究进入人工智能领

域。
相关领域的重大进展可参见下图：

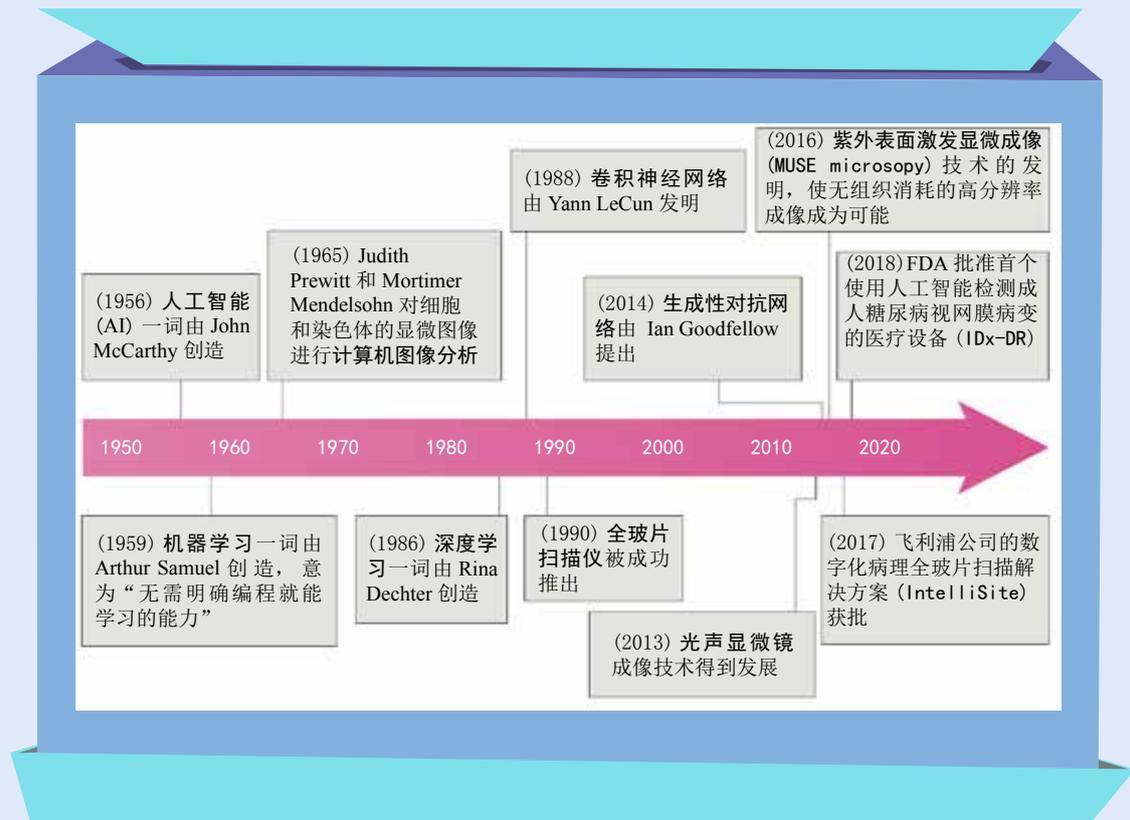


图1. 计算病理学的里程碑。

图片来源: Kaustav Bera, *et al.* (2019) Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, 16: 703–715.

在病理学和肿瘤学中，开发基于人工智能的图像分析新方法的工作由计算机工程师和数据科学家主导，他们开发和应用人工智能工具来完成各种任务，包括帮助提高诊断准确性，鉴别可应用在精准肿瘤学中的新的生物标志方

法等。病理医生和肿瘤医生则是这些图像分析方法的主要终端用户。在常规的临床实践中，病理医生(特别是解剖病理医生)的组织学诊断需要对所分析的样本进行视觉识别、半定量和多种形态学特征的整合，并结合基础疾病的进

程背景进行综合考虑。

虽然系统化培训和标准化指南的使用可加强分析过程的规范化和诊断结果的准确性，但组织病理学分析仍受到其固有的主观性局限，而且独立观察者之间也存在着视觉感知、数据整合和判断方面的天然差异。即使是接受相同培训的病理医生之间也会出现意见分歧，而这会导致诊断不一致和非最优患者护理的可能性。

此外，目前广泛使用的无创或微创操作，大大降低了所获取的诊断样本的规格和质量，这使病理医生的工作更具挑战性。

再加上，在癌症中，影响细胞信号传导和

细胞间相互作用的基因变化非常错综复杂，这些都会影响疾病的生物学进程以及治疗措施的作用。

基于人工智能的方法可以帮助肿瘤医生和病理医生应对这些挑战，因为其可靠并可重复。这一点已得到了各种研究结果的支持，这些研究表明，基于人工智能的方法与病理学专家的判断有着相似的准确度，更重要的是，当该方法与标准流程同时应用于检测和诊断场景时，可以提高人类读片者的表现。接下来将介绍如何将基于人工智能的方法集成到病理医生和肿瘤医生的工作流程中，并讨论了在癌症患者的日常管理中此类工具所面临的挑战。

资讯 · 频道

www.LifeOmics.com



二、人工智能在肿瘤病理学中的应用

在数字病理学中，人工智能方法已经在图像处理和分类任务中广泛应用，包括聚焦目标识别的低级任务，如目标图像的检测和分割，以及高级任务，如根据一定的模式在图像中预测疾病的诊断和治疗反应的预后。

另外，有不少人工智能应用程序致力于满足病理医生的需求——将一些耗时的任务自动化，从而使他们能够在高级决策任务上投入更多的时间，特别是那些疾病表型特征更加复杂的任务。此外，人工智能方法也已越来越多地应用于帮助解决肿瘤医生面临的问题，例如，通过发展预后分析来评估疾病的严重程度和后果，并预测治疗反应。

1. 基于手工特征的方法

特征工程是描述和开发ML算法的构建模块的过程，既可以利用病理医生和肿瘤医生的专业知识(基于特定领域的特征)，也可以不使用专业知识(不基于特定领域的特征)，构建基于手工特征的ML途径。在手工ML途径中，人们尝试构建可以锚定到问题领域的新特性——也就是说，算法通常针对特定的癌症或组织类型，所聚焦的特征可能不适用于广泛用途。

同时，手工特征也可以包括不基于特定领域的特征，例如组织的亚视觉结构异质性测量或细胞核形状和大小的测量，这些可以应用于不同的疾病和组织类型。其他广泛使用的不基

于特定领域的特征包括基于图形的方法，这些方法能够量化单个类型的离散组织元素或基元(例如，细胞核、淋巴细胞或腺体结构)或不同组织特异基元之间的空间分布、排列和结构。在此，重点介绍一些正在开发的基于手工特征的方法。这些基于手工特征(包括不基于特定领域的特征和基于特定领域特征)的方法已被用于不同类型的肿瘤的诊断、分级、预后以及预测不同肿瘤亚型对治疗的反应，涉及的肿瘤包括前列腺癌、乳腺癌、口咽癌和脑瘤。

在诊断应用方面，早在2010年，Osareh等人就提出了一个监督ML模型，该模型专注

于由乳腺病理专家提出的10个细胞特征，用于在穿刺活检样本图像中区分乳腺肿瘤的良恶性。

在Lee等人于2017年发表的一项有关前列腺癌的研究中，介绍了在肿瘤和肿瘤相关良性区域中，一些关于细胞核形态、方向和结构紊乱的新手工特征。这些特征随后与利用H&E染色的手术标本开发的ML分类器(classifier)结合使用，可用于预测手术后5年内生化复发的可能性。

在2018年发表的两项研究中，Lu等人提出，结合细胞核的形状和取向特征，可在雌激素受体(Oestrogen Receptor, ER)阳性乳腺癌患者中，将短期生存(<10年)与长期生存(>10年)的患者区别开来，Whitney等人还指出，细胞核的形状、组织和结构的定量特征可独立预测ER阳性乳腺肿瘤患者的复发风险(基于21个基因表达的伴随诊断试验Oncotype DX)。

在关于预后应用方面，人们正在应用图形方法来评估不同类型组织成份的空间排列和结构，从而预测临床结果。在2018年发表的一篇文章中，Saltz等人描述了利用卷积神经网络(Convolutional Neural Network, CNN)结合病理医生的反馈，自动检测来自“癌症和肿瘤基因图谱”(The Cancer Genome Atlas, TCGA)的组织切片图像中肿瘤浸润淋巴细胞(Tumour-Infiltrating Lymphocytes, TIL)的空间组织，继而发现这一特征是13种不同癌症亚型的预后指标。在一项相关研究中，Corredor等人发现在早期非小细胞肺癌(Non-Small-Cell Lung Cancer, NSCLC)中，TIL簇的空间排列比单纯的TIL密度更能预测复发风险，这是通过计算相邻的TIL之间以及TIL和癌细胞核之间的图形导出关系而得出的。

在Yuan等人对三阴性乳腺癌的全玻片数字图像进行肿瘤细胞间淋巴细胞空间分布的建模和研究分析的基础上，Heindl等人发现免疫

细胞的空间分布也与ER阳性乳腺癌的晚期复发有关。

虽然许多手工ML方法主要集中在分析肿瘤内上皮来源的细胞，但越来越多的研究者试图确定肿瘤基质中的预后模式。在Beck等人的一项开创性研究中，从乳腺癌患者标本的数字化WSI中提取了6642个与形态学、空间关系以及全局图像特征有关的，来自上皮和间质区域的特征，研究表明，从间质区提取的特征($P = 0.004$)比从上皮区提取的特征($P = 0.02$)具有更强的预后价值。

此外，研究者还发现，一个只基于基质特征的预后模型，在经过一组非裔美国前列腺癌患者训练后，与同时接受非裔美国患者和白人患者训练的模型相比，在独立测试中，前者对一组非裔美国患者的术后生化复发的预测精度更高——平均曲线下面积(Average area Under the Curve, AUC)为0.82(前者)对0.68(后者)。

在药物开发应用方面，过去数年间，人们对使用基于ML方法进行药物开发的兴趣日益增加。事实上，有相当比例的患者，对某些疗法——如细胞毒性药物或免疫检查点抑制剂(Immune-Checkpoint Inhibitors, ICIs)——并没有反应，这增加了人们通过结合AI与数字病理学来识别最有可能获得治疗益处的患者的兴趣。Wang等人建立了一种方法——利用细胞核和核周特征(形状、取向和空间排列)将早期非小细胞肺癌患者(仅接受手术治疗)按照疾病复发的风险分为两组；高危组患者被认为是可能受益于辅助化疗的潜在候选者。手工ML方法还致力于探索特定药物，包括靶向药物、化疗药物和ICIs的治疗反应。Wang等人已经开发出一种方法，通过细胞核或TIL的空间排列来预测晚期NSCLC患者对于抗细胞程序性死亡蛋白1(Programmed cell Death 1, PD-1)抗体nivolumab的反应性。

2. 基于神经网络的方法

由于深度学习(DL)方法不依赖于设计特征，可以直接从原始数据中学习表型，已经被越来越多地投入应用，并被改造得适应数字病理学领域。DL方法通常涉及具有一定类别标签的图像的学习集(例如，良性或恶性肿瘤)以及随后的无预设标签的新输入数据的询问。这个过程包括生成并学习最优图像特征，以便最

有效地区分感兴趣的类别。与手工特征工程相比，基于DL的方法更易于应用，并且准确度也更高，这些都是它被广泛接受的原因。由此，也有不少DL模型或算法被开发应用于病理图像分析。手工特征工程和基于DL方法的比较如下表：

表1. 手工特征工程和基于DL的方法的比较

项目	手工特征工程	基于DL的方法
开发	通常与病理学和肿瘤学专家密切合作开发；该方法基于特定专业知识，开发可能复杂而且耗时。	通过无监督的特征学习来开发，依赖于学习集的存在和所感兴趣类别的注释样本；网络设计通常关注微调算法，以最大限度地提高精度，同时最小化处理时间。
普遍适用性	该方法通常为特定的癌症亚型和/或特定起源的组织定制。例如，与腺体形态有关的特征只适用于具有大量小管和腺体的疾病。	使用转移学习等方法，一个针对特定疾病亚型训练的网络也可以应用于其它疾病亚型。
训练数据集	采用已知相关特征或感兴趣标记的数据集，规模较小即可满足需要。	通常需要大量注释良好的数据，其中包含来自不同目标类别或感兴趣种类的多个范例。
可解释性	其特征通常与疾病领域的属性相关，与基于DL的方法相比，往往具有更高的可解释性和更强的形态学和生物学基础。	其表征通常难以解释，尽管出现了一些旨在提供清晰度的新方法(如视觉注意力地图)，但在很大程度上仍被认为是“黑箱”方法。
临床应用	由于手工特征比基于DL的方法更具有可解释性，因此更可能用于高层决策，例如与疾病预后相关的决策或从治疗中获益的预测。	可能更适用于不太需要“解释为何作此决定”的情况，例如包括目标图像的检测或分割在内的一些低级任务。

表格来源：Kaustav Bera, *et al.* (2019) Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, 16: 703–715.

2.1 卷积神经网络(CNN)

卷积神经网络是迄今为止应用最广泛的DL算法，已经被应用于各种病理图像分析软件。CNN是一种深度前馈网络，它包括多个层以便从输入(例如图像)推断输出(通常是特定类别)，并且由于多个卷积片(sheet)的存在而得名。就是在如此构造的CNN中，网络使用输入层和输出层之间的过滤器从图像中学习和提取特征图谱。

基于CNN的DL方法已被用于基于图像的检测和分割任务，以识别和量化细胞(如中性粒细胞、淋巴细胞和母细胞)、组织学特征(如细胞核、有丝分裂图像、基质和腺结构)或感兴趣的区域(如肿瘤或瘤周区域)。此外，Senaras等人开发了基于CNN的DeepFocus系统，用于自动检测和分割数字化WSI中的失焦和模糊区域，平均准确率达到93.2%(±9.6%)。

在基于人工智能的诊断方法研究中，CNN取得了与人类专家评估相当的结果。例如在2019年发表的一项研究中，Tschandl等人比较了各种基于CNN的DL模型相对于人类读片者的诊断准确性。在用皮肤镜数字图像将色素

性皮损正确分类到7个预先确定的疾病类别中时，研究人员发现，DL模型始终优于人类医生——在511个人类读片者与139个DL算法的比较中，算法平均多出2.01个正确诊断(17.91个对19.92个； $P < 0.0001$)。

在预后软件研究方面，Nagpal等人利用一个DL系统，在检测根治性前列腺切除术标本WSI中的癌变区域后，自动给出了Gleason评分(Gleason分级是目前广泛采用的前列腺癌组织学分级方法，也是制定前列腺癌治疗方案的重要参考指标)。在含331张玻片的验证集中，参考由一位泌尿生殖系统病理专家制定的诊断标准(该专家还得到了至少三位普通病理医生的初步诊断意见)，DL方法在预测Gleason评分方面的准确度为0.70，而29位病理医生的平均准确度为0.61。

Kather等人使用CNN生成“深间质评分(deep stroma score)”，以预测结直肠癌患者的生存期，并在含409名患者标本的独立验证集中发现该指标还可独立判断无复发生存率(HR 1.92, CI 1.34-2.76)和总体生存率(HR 1.63, CI 1.14-2.33)。

2.2 完全卷积网络(Fully convolutional networks, FCN)

另一种流行的DL模型是完全卷积网络(Fully convolutional network, FCN)，它缺乏完全连接的层，只包含卷积层的层次。与CNN

用于在局部聚集信息以进行全局预测不同，FCN可用于从每个像素学习表型，因此，有可能用于对整个病理图像中仅稀疏出现的元素或

特征进行检测。此属性使FCN能够进行像素水平的预测，这点可能比CNN更具优势，后者要从整个图像中重复出现的特征中学习。在数字病理学研究方面，2019年Rodner等人利用FCN在头颈部肿瘤标本的组织病理学图像中区分癌变区域和非恶性上皮细胞——将头颈癌切片用于标准组织病理学，并与来自相同切片并经多种显微镜技术分析获得的多模态图像联合使用。利用FCN对图像进行分析，可将WSI

分为四类：癌组织、非癌上皮组织、背景和其它组织。Cruz-Roa等人在2018年发表的一项研究中，使用来自349名患者的500张图像训练FCN，然后在癌症基因组图谱的195项研究中进行了独立测试，以检测WSI上的浸润性乳腺癌区域。与乳腺病理学专家的评估相比，结果显示FCN的预测和病理医生对侵袭性癌症的注释(基本事实)之间具有良好的一致性，检测准确率为71%(Sørensen-Dice系数)。

2.3 递归神经网络(Recurrent neural networks, RNN)

与CNN和FCN局限于分析单个时间点的数据不同，RNN可以在不同的时间间隔或时间点上存储输入，以便按顺序处理它们，并从(通常数以百万计)早前的不连续步骤中学习。由此RNN可以处理在不同时间点输入的状态，从而表现动态行为。

长短时记忆(Long short-term memory, LSTM)网络是RNN的一种，它通过引入循环门——也称为遗忘门(forget gates)——来增强学习和保存长期信息的能力。循环门使CNN能够通过回顾传播误差来进行任务学习。

2018年，Bychkov等人建立了一个结合

LSTM和CNN的网络，利用H&E染色的结直肠癌组织芯片(Tissue Microarray, TMA)标本的图像预测疾病复发的风险(基于回顾性DFS数据的高低)。该LSTM网络的预测性能(平均AUC为0.69)高于仅基于组织学分级(平均AUC为0.57)和经三位病理学专家达成一致的基于视觉的风险评估(平均AUC为0.58)。对在不同时间点获得的组织图像进行分析，例如对处于积极监测下的前列腺癌患者的连续随访活检样本进行分析，是基于RNN的AI方法的另一个潜在用途。

2.4 生成性对抗网络(Generative adversarial networks, GAN)

除了CNN之外，基于生成性对抗网络(GAN)的方法在基于DL的数字病理学方法中表现出越来越大的潜力，包括特征分割和染色转移(一个校正颜色变化的术语)。GAN的工作原理是同时实现两个相互竞争的神经网络。一个网络是生成器，根据提供给网络的训练样本生成合成数据，而第二个网络则评估生成的数据与原始数据之间的一致性。GAN的目标是降低第二网络的分类误差，使生成的图像更接近原始图像。Gadermayr等人利用GAN从切除的小鼠肾脏(显示出与人类肾脏高度相似)获得的病理标本图像中切割肾小球图像。基于GAN的方法也被用于训练DL方法，以自动评分非小细胞肺癌(NSCLC)活检样本图像中肿瘤细胞程序性死亡配体1(Programmed cell death

1 ligand 1, PD-L1)的表达。该方法有助于最小化必要的病理医生注释数量，从而弥补活检标本中可用组织的不足。类似地，Xu等人提出了一种新的基于GAN的方法，将WSI的H&E染色转化为基于细胞角蛋白18和19的虚拟免疫组织化学染色，这种方法有可能避免破坏性的免疫组化检测。今年5月Liu等人发表在自然出版集团旗下刊物《科学报告》(Scientific Reports)上的一项研究中，GAN方法还被用于提高利用胶质瘤组织病理切片图像预测异柠檬酸脱氢酶(Isocitrate dehydrogenase, IDH)突变的准确性，由于IDH是胶质瘤重要的预后、诊断和治疗的生物标志物，该方法作为对免疫组化和基因测序鉴定法的补充，可帮助医生更方便地进行胶质瘤IDH突变的预测。

3. 人工智能方法中病理学和肿瘤学的整合

成功的基于人工智能的数字病理学方法需要将病理医生的工作与肿瘤医生的工作紧密结合起来(图2)。例如，大多数有监督的人工智能算法都非常依赖于对仔细标注的数字化病理切片图像进行分类器训练。病理医生在为参与

设计和开发人工智能算法的计算机科学家和数据科学家提供特定领域的知识方面发挥着重要作用。他们还经常提供诊断参考或金标准，与ML算法进行比较。

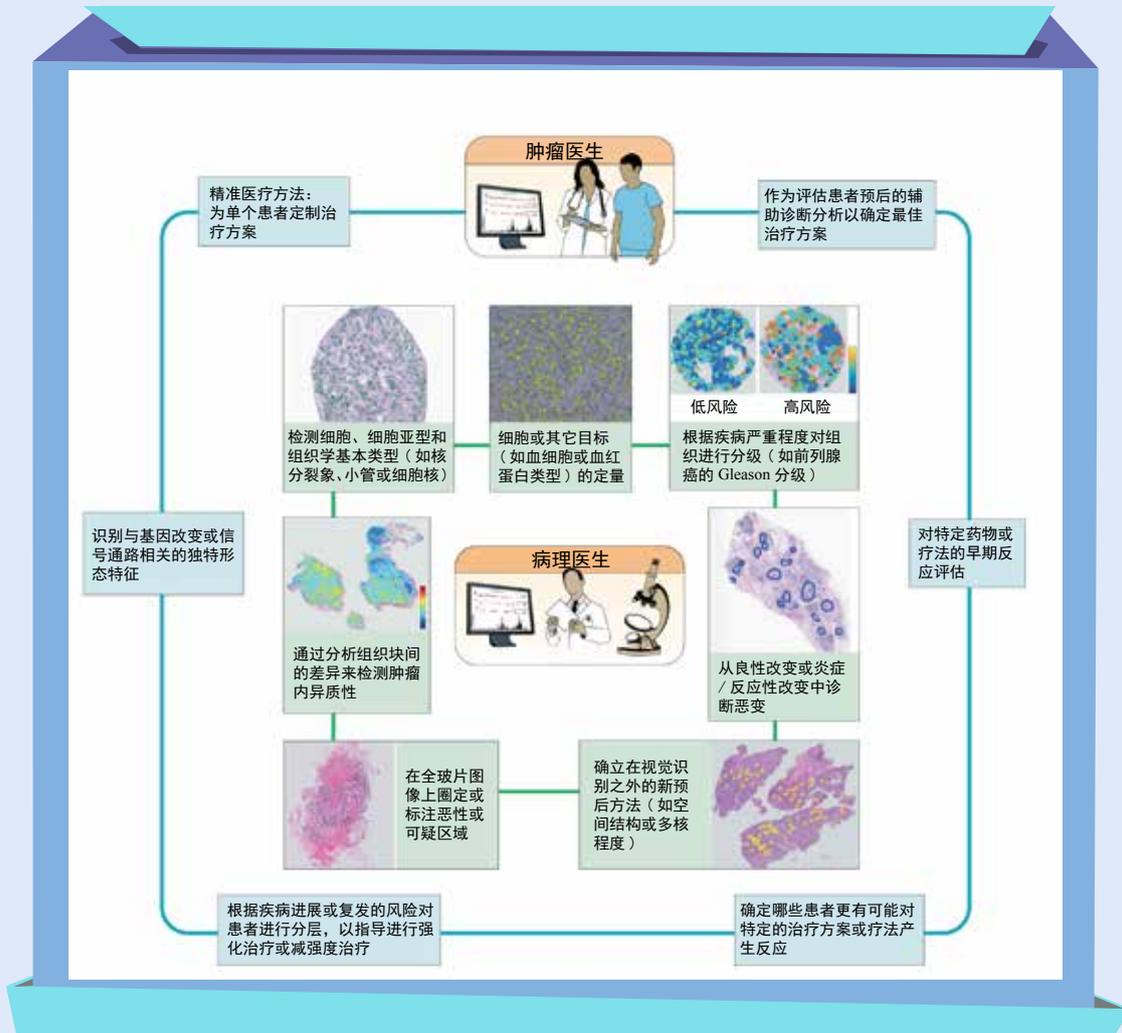


图2. 人工智能(AI)和机器学习方法可补充专业知识, 并为病理医生和肿瘤医生提供支持。本文介绍了一些现有的正被病理医生用来分析肿瘤图像的人工智能方法。对于执业肿瘤医生来说, 人工智能方法可提供癌症患者各方面管理的辅助决策。

图片来源: Kaustav Bera, *et al.* (2019) Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, 16: 703–715.

肿瘤医生面临的一个主要难题是，当为个体患者定制治疗方案时，在是否提供一种特定的疗法方面缺乏共识。人工智能方法除了作为决策的支持系统之外，还可以作为以患者为中心的精准医疗的辅助工具。

例如，当基于人工智能的预测方法根据手术标本的组织形态学分析，提示患者具有高复发风险时，临床医生将更有可能建议化疗。除此之外，人工智能还可以帮助肿瘤医生在处理目前生物标志物(基于基因或组织)辅助诊断分析时的一些局限。

与基因组检测相关的挑战之一是不同类型的癌症都存在肿瘤内和肿瘤间的异质性，即对不同位置样本的检测可以产生不同的预后信息

(例如，与组织采样坐标相对应的肿瘤遗传特征的差异)。基于人工智能的问诊工具可以帮助克服这一局限性，其可以通过分析个体从肿瘤到正常组织的所有病理切片，生成涵盖整个病变的完整统一的特征。

一些研究小组正在探索将肿瘤的分子特征和形态学特征结合起来，以提高ML方法的预后和预测性能。

优化的基于人工智能的生物标记物方法可以成为肿瘤医生的一个物美价廉的工具，使他们有可能预选出真正能从新药治疗中受益的患者，这样不仅可为患者节省不必要的费用，还能避免患者在没有任何实质性生理改善的情况下承受不必要的药物毒性。

三、人工智能在肿瘤筛查中的应用——以一项跨国乳腺癌筛查研究为例

2018年报道的一项涉及1992例不同肿瘤标本以及16名外科病理医生的研究表明，在初步诊断表现上，数字化WSI(不一致率分别4.9%)并不亚于传统的镜检方法(不一致率为4.6%)。

作为利用人工智能诊断癌症的早期例子，纽约大学朗格尼分校(NYU Langone)的珀尔马特癌症中心(Perlmutter Cancer Center)于2019年10月获准使用其人工智能分类器进行诊断试验，通过全基因组甲基化分析对肿瘤进行分类，可以提高诊断的准确率。世界各地的研究人员正在开发类似的系统，以帮助病理学家更准确地诊断癌症。今年3月发表在《自然·医学》(*Nature Medicine*)杂志上的一项研究发现：在造血疾病的诊断方面，基于对71例临床样本的研究，结合了机器学习和形态标记的髓细胞计数法，要优于流式细胞术，并可与专家显微镜检查相媲美，这使复杂的造血疾病的诊断更加灵敏和自动化。

在利用人工智能诊断乳腺癌方面，Google团队的研究者进行了大量的工作。乳房X线筛查的目的是实现乳腺癌早诊早治，以提高治疗的成功率。尽管世界各地都有筛查项目，但乳房X光片的解读仍受到较高的假阳性和假阴性率的影响。Google团队的研究者在今年1月的《自然》(*Nature*)杂志上发表了他们的最新研究成果。在这项研究中，他们提出的人工智能(AI)系统能够在乳腺癌预测方面超越人类专家。为了评估其在临床环境中的表现，研究者分别收集了来自英美两国的大数据集，结果显示假阳性和假阴性率在两国的数据集中都有所下降。而且在一项对6名放射科医生的独立研究中，人工智能系统的表现优于所有人类读片者。当使用人工智能系统来模拟英国乳腺癌诊断采用的双人读片过程(double-reading process)时，发现人工智能系统保持了良好的性能，并将第二位读片者的工作量减少了88%。这些对人工智能系统切实的评估为旨在提高乳腺癌筛查的准确性和效率的临床试验铺平了道路。

1. 乳腺癌筛查评估人工智能系统的构建

针对以往研究中存在的种种局限：1. 多数研究基于小而丰富的数据集和有限的随访，很少有研究将AI的表现与真正从事临床实践的读片者(病理医生)进行比较——取代在实验室里模拟的读片环境。2. 到目前为止，几乎没有证据表明人工智能系统能够在没有额外训练数据的情况下，在不同的筛查人群和设置之间进行转换。3. 至关重要的是，普遍采用的随访时间间隔不超过12个月，这意味着更多症状轻微的癌症可能被忽略了——直到下一次筛查（才会被发现）。

在本研究中，研究者使用了来自英美两国的两个具有代表性的大型临床数据集来评估一个新的预测乳腺癌的人工智能系统的性能，并将该系统的预测结果与在常规临床实践中读片

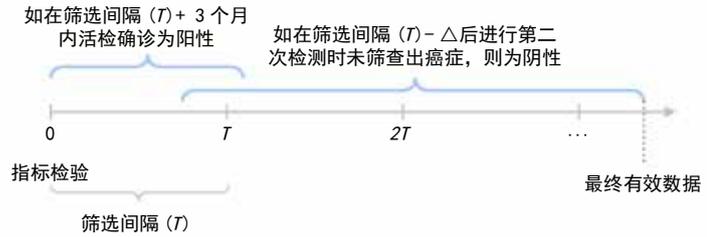
者所做的预测进行了比较，结果显示其性能超过了放射科医生个人。这些结果被一项独立进行评估的读片者研究证实。此外，该研究还展示了如何将该系统整合到筛选工作流程中，并提供了证据，证明该系统可以跨洲推广。图3展示了项目的概况。

乳腺癌的预后是在多年随访的基础上确定的。研究者根据每个数据集所在国家的筛选间隔来选择随访时间，考虑到安排的变化和随访的延迟，为每个间隔增加了三个月的缓冲期。被认定为癌症的病例都有随访期间的活检确诊。标记为癌症阴性的病例在至少一个随访期间未筛查出癌症；没有随访的病例被移出测试数据集。

检测数据

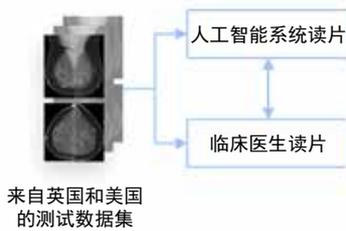
		
妇女人数	25,856	3,097
筛查流程	二次读片	一次读片
筛查间隔	3年	1或2年
癌症随访	39个月	27个月
癌症数量	414 (1.6%)	686 (22.2%)

正确数据定义

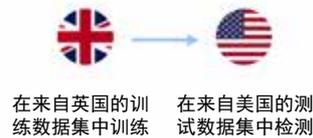


评估

与回顾性临床表现相比较



跨数据集泛化能力



独立进行的读片者研究

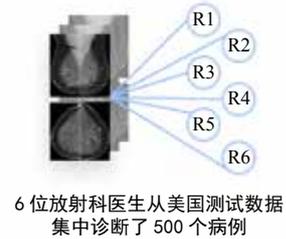


图3. 开发一个通过筛查乳房X线照片检测乳腺癌的人工智能系统。代表英国和美国乳腺癌筛查人群的数据集来自英国的三个筛查中心和美国的一个中心, 结果来自活检记录和纵向随访。以一组筛查乳房的X线照片对人工智能系统进行了培训, 使之识别乳腺癌, 并通过三种方式进行评估: 首先, 将人工智能预测与临床上已经作出的诊断进行比较; 其次, 为了评估其不同人群间的普遍性, 在其中的一个版本上只使用英国的数据训练, 并在美国的数据上重新检测; 最后, 使用美国测试集的一个子集, 将人工智能系统的表现与6名独立工作的放射科医生的表现进行比较。

图片来源: Scott Mayer McKinney, *et al.* (2020) International evaluation of an AI system for breast cancer screening *Nature*, 577:89–94.

2. 乳腺癌预测中人工智能系统和临床工作者的比较

2.1 与回顾性临床结果比较

该研究使用经活检确诊的乳腺癌结果来评估人工智能系统的预测和人类读片者在临床实践中做出的原始判断。根据临床医生召回患者进行进一步诊断调查的决定来衡量他们的表现。

在英国，每一张乳房X光片都由两位医生解读，如果有分歧，仲裁程序将引入第三种意见。每个读片者都能参考前人的意见。这些决策的记录为癌症预测提供了人类表现的三个基准。

与第一位读片者相比，人工智能系统的绝对特异性和绝对敏感性分别提高了1.2%(95%置信区间(CI) 0.29%, 2.1%; $P = 0.0096$ 优越性)和2.7%(95% CI 3%, 8.5%; $P = 0.004$ 非

劣效应预设界值5%)。与第二位读片者相比，人工智能系统在特异性、敏感性，以及在特异性和敏感性的一致判断上均表现出非劣效能。

在美国的标准筛查流程中，每张乳房X光片由一名放射科医生解读。该研究使用在最初的筛查流程中每个病例的乳腺影像报告和数据库(Breast Imaging Reporting and Data System, BI-RADS)评分作为人类癌症预测的指标。与常规读片者相比，人工智能系统的表现有统计学意义上的改善，绝对特异性提高5.7% (95% CI 2.6%, 8.6%; $P < 0.001$)，绝对敏感性提升9.4% (95% CI 4.5%, 13.9%; $P < 0.001$)。

2.2 不同人群之间的泛化

为了评估人工智能系统在不同人群和筛选设置之间的泛化能力，该研究仅使用英国数据集训练系统，再将其应用于美国测试数据集。即使没有接触到美国的训练数据，人工智能系

统的ROC曲线也包含了美国放射科医生的平均表现。同样，与放射科医生相比，人工智能系统显示出更好的特异性(+3.5%, $P=0.0212$)和敏感性(+8.1%, $P=0.0006$)。

2.3与读片者研究进行比较

在一项由外部临床研究机构主持的研究中，6名经认证的放射科医生从美国测试数据集中随机抽取500份乳房X线照片进行诊断，在参考包括患者年龄、乳腺癌病史和以往的乳腺检查记录在内的临床信息的基础上，使用美国放射学会的强化版BI-RADS量表对每个病例进行评分，并将BI-RADS评分与已知的真实结果进行比较，据此绘制每个读片者的受试者操作特征(receiver operating characteristic, ROC)曲线。人工智能系统给出的评分也以同样的方式处理。

与这6名放射科医生的表现相比，人工智能系统的表现显著优于医生们的平均水准(曲线下面积变化(Δ AUC)=+0.115, 95%CI 0.055,

0.175; P=0.0002)。

此外，人工智能系统还被设计用来标志疑似恶变的特定区域。同样，放射科医生也提供了相关区域标注。经分析，在这方面人工智能系统的表现也大大超过了人类(Δ pAUC=+0.0192, 95%CI 0.0086, 0.0298; P=0.0004)。

值得一提的是，在这项研究中，既有被6名放射科医生遗漏却被AI正确识别的病例，也有被AI遗漏却被6位医生正确识别的病例。虽然研究者尚无法确定其中的机制或模式，但这样的结果提示在癌症的识别诊断方面，AI系统和人类读片者可通过互补以提高准确率。

四、讨论

1. 人工智能在肿瘤诊疗领域临床应用的前景

人工智能系统可以在保持诊断水准的前提下，减少病理诊断过程中的工作量。以AI在双人读片过程中的应用为例，在Scott Mayer McKinney等人的研究中，当人工智能系统的判断与第一位读片者一致时，就省略后续步骤，以第一位读片者的意见为准；当判断不一致时，再依照惯例引入第二位读片者和仲裁程序，以获得最终诊断意见。这种人机结合的方法，在性能上与传统的双人读片过程相当，但是节省了第二位读片者88%的工作量。

人工智能系统也可以用于在病例筛选中提供自动的、即时的反馈。

研究表明，通过调低AI系统的判断阈值，可以排除掉约三四成的正常病例。在使用双人读片法时，每182人中检出1例癌症属于正常范围。而使用人工智能系统，这个阈值可以下调至每千人至每万人中检出1例，以保障正常病例的高可信度。同理，通过调高判断阈值，还可以筛选出三四成需要优先确认的癌症病例，

这些筛选出的病例中，大约6次随访中有5次最终被诊断为癌症。相比之下，Scott Mayer McKinney等人的研究发现，只有22.8%通过双人读片过程召回的英国病例和4.9%通过单次读片召回的美国病例，最终被诊断为癌症。

由于人工智能具有泛化能力，可以实现跨人群筛选和跨诊断方案筛选，在未来的临床应用中，它既能提供强大的基准性能，又能通过当地数据进行调整和提升。

人工智能系统的特异性优势表明，它可以帮助降低病人的召回率和不必要的活检，也能够比医疗标准更早地发现癌症，还能尽早标记可疑区域，供专家审查。值得注意的是，人工智能系统能额外识别的这些癌症往往是侵袭性的，而非原位疾病。

上述研究表明这项技术有巨大的应用前景，可以在医疗资源不足或劳动力短缺的情况下，以可持续的方式迅速提供筛查结果。既能提供跨人群跨诊疗方案的基础筛查，又能针对

特定人群或其它参数进行优化以提高一定范围内的筛查精度。在癌症的识别诊断方面，AI系统和临床医生的判断可互补以提高诊断的准确

率。还可以通过评估患者的预后，帮助制定个性化治疗方案。有必要开展更多临床研究，以明确这项技术能在多大程度上使患者受益。

2.人工智能在肿瘤诊疗领域临床应用的挑战

人工智能虽然在临床应用上拥有广阔的前景，但面临的挑战也为数不少，其中之一就是监管障碍。因为采用人工智能方法进行数字化病理的临床应用首先要获得监管机构的批准。在大多数国家，指导批准过程的关键原则是要求“解释软件如何工作”——这对于基于DL的人工智能方法至关重要，因为这种方法常被认为是一个“黑匣子”，缺乏可解释性。这增加了AI产品审批和监管的难度。

同时，数据质量也是一个问题，基于人工智能方法的性能主要取决于输入数据的数量和质量。用于训练人工智能算法的数据需要干净、准确、有最大信噪比，并且尽可能精确和全面，以获得最强的预测性能。如何推进医疗数据电子化、规范化，在加强医疗数据隐私保护的同时促进数据交流和共享，都是当前面临的挑战。

尽管人工神经网络具有很高的准确性和易用性，但也存在针对人工神经网络缺乏可解释性和不同领域的直观性的批评，这些都可能构成其临床应用的障碍。目前已经有一些研究试图为DL工具提供生物学解释，包括利用事后分析方法或有监督的ML模型，来解释DL模型做出的预测。

在临床应用之前，还需要进行算法验

证——采用多机构数据充分验证基于人工智能和ML的工具，以确保方法的通用性。在过去的几年里，研究者也在一直致力于建立全面的质量控制和标准化工具。

要在临床上使用人工智能辅助分析或决策支持系统，成本补偿也是需要解决的主要问题之一。探索可行的商业模式，尝试将AI产品和服务纳入现有的保险和收费体系，还有很长的路要走。

对于病理医生来说，人工智能工具将主要用于检测数字化WSI中他们感兴趣的结构或特定区域。因此，开发此类应用程序的关键要素是数字幻灯片扫描仪、相关应用程序、方便共享图像和预测结果的巨大带宽和通信通道。此外，为了让病理医生能够放心地使用人工智能算法，它们必须达到的性能阈值也是一个亟需明确的问题。

对于肿瘤医生来说，基于数字病理学的辅助诊断测试可以为疾病风险分层和患者选择靶向治疗提供额外的有价值的信息，其中一个重要的问题是确定常规临床应用所需的辅助诊断分析的最低性能和准确性。同样，基于人工智能的预后和预测分析可能需要一定级别的证据来支持，并能够在临床相关应用中得到验证。鼓励临床医生参与相关产品的设计、研发以及

交流使用经验，这有助于相关AI产品更好地服务于临床。

从人工智能的发展历史来看，我国人工智能领域的开发研究稍晚于发达国家，但从人工智能战略布局来看，各国的差距并不明显。自2017年7月我国发布《新一代人工智能发展规划》(比《美国国家人工智能研究和发展战略计划》(The National Artificial Intelligence Research and Development Strategic Plan)的发布只晚了9个月)以来，人工智能医疗器械领域的研发工作和市场需求都发展迅速。截止2019年1月，我国已有89家人工智能医疗企业，国家药监局网站上共能查询到两项已注册的国产人工智能医疗器械产品：皮肤病影像人工智能辅助识别系统(湘械注准20202210769)和诊断图像处理软件(赣械注准20182210153)。

2019年7月17日，人工智能医疗器械创新推进会在京召开(图4)，这次会议宣布了人工智能医疗器械创新合作平台成立，旨在积极应对人工智能快速发展带来的各方面的挑战，加快人工智能科技成果在我国医疗器械领域的转化应用。2019年10月12日，国家药品监督管理局发布了《国家药监局关于成立人工智能等3个医疗器械标准化技术归口单位的公告》(2019年第82号)，提出了人工智能医疗器械标准化技术归口单位组成方案，包含人工智能医疗器械所涉及的术语和分类、数据集质量管理、基础共性技术、质量管理体系、产品评价流程、专用方法等行业医疗器械标准制修订工作归口单位和专家组名单，这将有力促进产学研协同创新和人工智能相关产品落地，并应对在此过程中的各项挑战。



图4. 2019年7月17日，人工智能医疗器械创新推进会在京召开。

图片来源：<http://www.nmpa.gov.cn/WS04/CL2056/339061.html>

主要参考文献

1. Kaustav Bera, *et al.* (2019) Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, 16: 703–715.
2. Scott Mayer McKinney, *et al.* (2020) International evaluation of an AI system for breast cancer screening. *Nature*, 577: 89–94.
3. Sidong Liu, *et al.* (2020) Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. *Sci Rep*, 10: 7733.

特约编辑招聘启事

为了及时收集生命科学最新资讯、提高《生命奥秘》办刊质量，现面向从事生命科学或对这学科有浓厚兴趣的科研人员、学生诚聘特约编辑（兼职）。

职位职责：

独立完成《生命奥秘》专题的策划：对基因组学、蛋白组学、生物信息学和细胞生物学等学科的发展以及生物医学领域相关技术（例如基因诊断技术、干细胞和克隆技术、生物芯片技术等）的应用进行翻译及深入评述。

选题要求内容新颖、评述精辟、注重时效和深入浅出。尤其欢迎以自身系统研究为基础的高水平译述与评论，结合所从事的科研工作提出自己的见解、今后设想或前瞻性展望。

要求：

- 1.具备基因组学、蛋白组学、生物信息学、细胞生物学等生命科学学科背景；
- 2.具备良好的生命科学前沿触觉；
- 3.具备较高的外文文献翻译、编译水平；
- 4.具备较强的选题策划、资料搜集、组织能力，以及专业稿件撰写能力；
- 5.具有高级职称；或者拥有（正在攻读）该领域的最高学位。

有意者请将个人简历发送至 editor@lifeomics.com



BlazeTaq™ SYBR® Green mRNA qPCR 系列

• 高灵敏度，高特异性



扫码查看产品详情

• 抗体修饰法 mRNA qPCR 试剂

买一赠一

活动时间: 2020年9月1日-11月30日

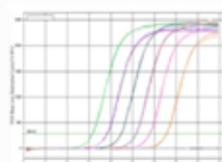
RT-qPCR 实验简图

两步法 RT-qPCR 反应

RNA

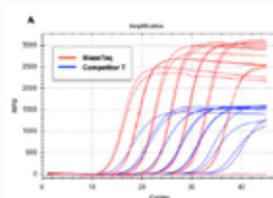
SureScript™ First-Strand cDNA Synthesis Kit

cDNA

BlazeTaq™ Probe qPCR Mix &
BlazeTaq™ SYBR Green qPCR Mix2.0 &
All-in-One™ qPCR Primers

一步法 RT-qPCR 反应

RNA

BlazeTaq™ Probe One-Step RT-qPCR Kit &
BlazeTaq™ One-Step SYBR Green RT-qPCR Kit &
All-in-One™ qPCR Primers

BlazeTaq™ 系列产品列表 **买一赠一**

产品名称	货号	规格	目录价
BlazeTaq™ SYBR® Green qPCR mix 2.0 (with ROX)	QP031 QP032 QP033	200/600/1200 rxns	¥555 起
BlazeTaq™ SYBR® Green qPCR mix 2.0 (without ROX)	QP041 QP042 QP043	200/600/1200 rxns	¥555 起
BlazeTaq™ One-Step SYBR® Green RT-qPCR Kit (with ROX)	QP071 QP072 QP073	200/600/1200 rxns	¥1498 起
BlazeTaq™ One-Step SYBR® Green RT-qPCR Kit (without ROX)	QP081 QP082 QP083	200/600/1200 rxns	¥1498 起

*另外, 我们还提供探针法 RT-qPCR 试剂 BlazeTaq™ Probe One-Step RT-qPCR Kit (QP076 和 QP086) 和 BlazeTaq™ Probe qPCR Mix (QP036 和 QP046)

配套产品列表

产品名称	产品描述	货号	规格	目录价	优惠
SureScript™ First-Strand cDNA Synthesis Kit	第一链合成试剂盒	QP056 QP057	20/60 rxns	¥360 起	买一赠一
All-in-One™ mRNA qPCR 内参引物	mRNA内参引物	各异, 官网可查	20 μL×100 rxns	¥100	¥60
All-in-One™ mRNA qPCR Primer	mRNA验证引物		20 μL×100 rxns	¥200	¥120

All-in-One™ miRNA qPCR系列

灵敏度高、特异性强、性价比高

优惠价：**6**折起



活动时间：2020年9月1日 - 11月30日 扫码查看产品详情

miRNA 定量检测

All-in-One™ miRNA 定量检测试剂盒基于 SYBR® Green 的 qPCR 技术，可对低至 20 pg 的 Total RNA（或者低至 10 pg 的 small RNA）进行快速、准确定量。

产品名称	货号	规格	目录价	优惠价 (6折起)
All-in-One™ miRNA First-Strand cDNA Synthesis Kit 2.0	QP113 QP114	20/60rxns	¥2180 / ¥4980	¥1308 / ¥2988
All-in-One™ miRNA qRT-PCR Detection Kit 2.0	QP115 QP116	200/600rxns	¥2980 / ¥7280	¥1788 / ¥4368
All-in-One™ miRNA qPCR Kit	QP010 QP011 QP012	200/600/1200rxns	¥1100 起	¥880 起
All-in-One™ miRNA qPCR 内参引物	各异, 官网可查	20μL×100rxns	¥100	¥80
All-in-One™ miRNA qPCR 验证引物	各异, 官网可查	20μL×100rxns	¥200	¥160

qPCR 检测阵列的服务对象

人鼠全基因组 miRNA 研究

疾病（癌症）研究

信号通路研究

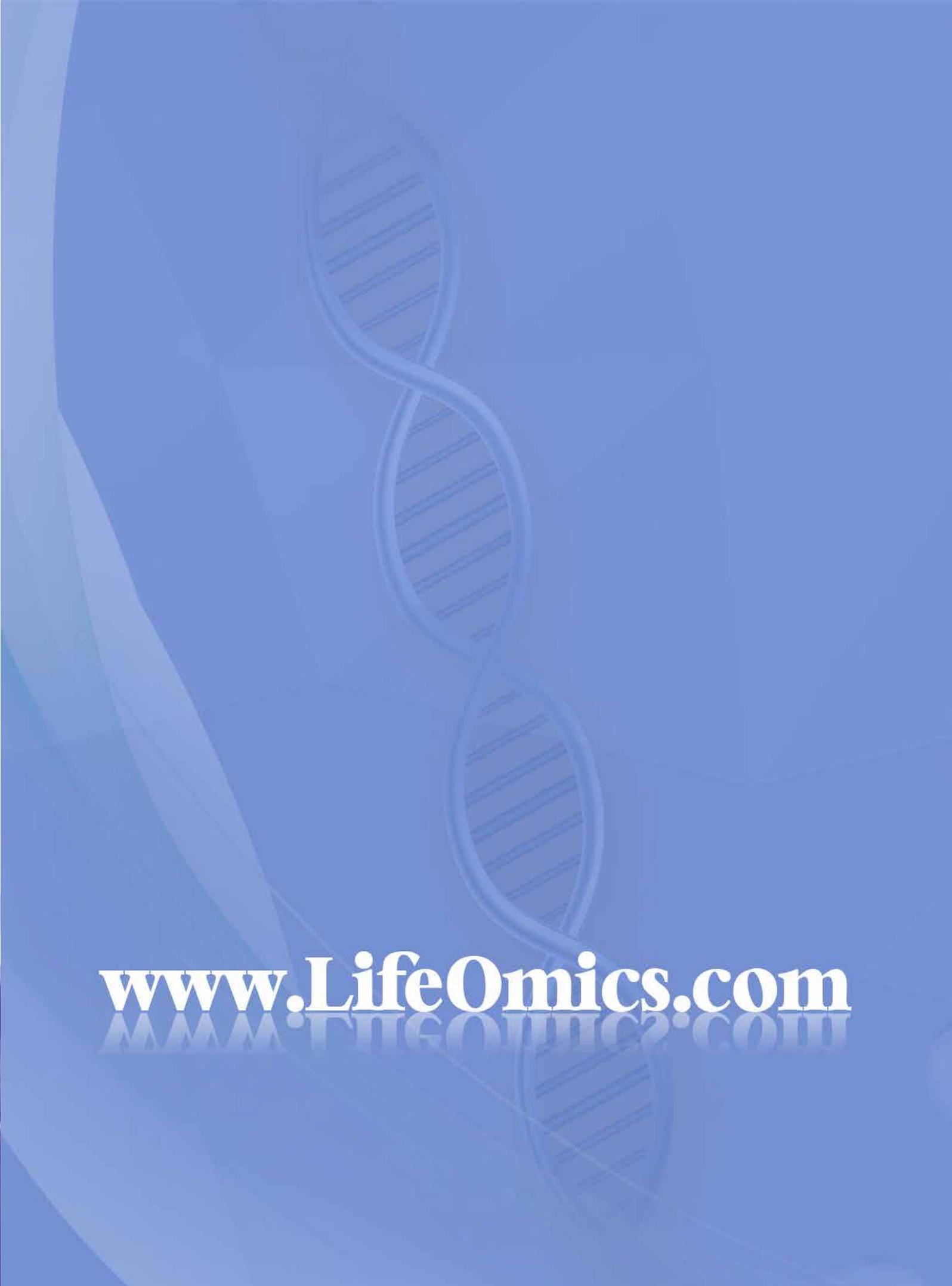
芯片和测序结果、
基因功能确证

类别	产品名称	描述
基因表达量检测阵列	ExProfile™ Pathway qPCR arrays	61 种信号通路相关基因表达差异定量检测阵列
	ExProfile™ Cancer Gene qPCR arrays	21 种不同癌症（肿瘤）相关基因表达差异定量检测阵列
	ExProfile™ Disease and gene group qPCR arrays	17 种疾病（或特殊的基因功能群体）相关基因表达差异定量检测阵列
miRNA 表达量检测阵列	miProfile™ miRNAome miRNA qPCR arrays	高通量人类、小鼠 miRNA 全基因组 miRNA 表达差异定量检测阵列
	miProfile™ Cancer qPCR arrays	27 种不同癌症（肿瘤）相关 miRNA 表达差异定量检测阵列
	miProfile™ Disease and Focus-Group miRNA qPCR arrays	11 种疾病（或特殊的基因功能群体）相关 miRNA 表达差异定量检测阵列
外泌体 miRNA 表达量检测阵列	miProfile™ Exosome miRNA qPCR arrays	来源于 MSC、血液、尿液、乳汁、癌症、肺癌、白血病、肝癌、结肠癌、鳞状细胞癌、卵巢癌、膀胱癌、前列腺癌、乳腺癌、肾癌、胰腺癌等 16 种组织的外泌体
表达量检测阵列定制服务	Custom-made Gene or miRNA qPCR arrays	96 和 384 孔板供选择，客户定制 array 阵列和 qPCR array 检测服务

A group of people are performing a human pyramid against a cloudy sky. The pyramid consists of four people standing on the ground, two people standing on their shoulders, and one person standing on the shoulders of the two people in the middle. The people are wearing dark jackets and light-colored pants. The sky is filled with soft, white clouds, and a bright sun is visible in the upper left corner, creating a lens flare effect. The overall scene conveys a sense of teamwork and achievement.

合办专题专刊
网站广告合作
邮件群发推广

请致电 (020) 32051255



www.LifeOmics.com