

# 肿瘤突变特征与病理分型的关联研究

史悦, 许争争, 鲁欢, 慈维敏

中国科学院北京基因组研究所, 中国科学院精准医学重点实验室, 北京 100101

**摘要:** 准确评估肿瘤的病理亚型对诊断、治疗和预后至关重要。以往病理亚型的诊断主要依赖 HE 染色法和免疫组织化学法, 而随着测序技术的不断发展, 对患者进行基因型和表型特点的个体分析成为可能, 将肿瘤病理分型与基因分型结合用于疾病分型、诊治选择和疗效判断的精准医学研究逐渐兴起。不同病理亚型的肿瘤细胞来源、致癌因素和临床表型均不尽相同, 其在基因组上会留下特异“印迹”, 即突变特征。本研究通过整合癌症基因组数据库(The Cancer Genome Atlas, TCGA)中肾癌、肺癌和食管癌的外显子测序数据, 分别对 3 种肿瘤通过肿瘤基因突变特征进行肿瘤病理分型聚类 and 预测。首先通过非监督聚类方法将 3 种肿瘤分别按照 24 种突变特征进行聚类分析, 其次通过随机森林法从 24 种突变特征中进一步选择对于区分不同病理亚型有显著性的突变特征并进行聚类分析, 构建突变特征对 3 种肿瘤病理亚型的分型模型。在肾癌中, 该模型准确率达到 100% (95% confidence interval (CI): 0.93~1.00), 肺癌和食管癌中分别达到了 78% (95% CI: 0.66~0.86) 和 84% (95% CI: 0.60~0.97)。以上研究结果表明, 突变特征作为新型分子标记物, 对肿瘤的病理分型、诊断, 尤其是早诊具有一定的参考意义。

**关键词:** 突变特征; 病理分型; 肾癌; 肺癌; 食管癌

## Correlation studies of distinct mutational signatures with common cancer pathological subtyping

Yue Shi, Zhengzheng Xu, Huan Lu, Weimin Ci

CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

**Abstract:** It holds great promises to precisely stratify cancer subtypes to improve cancer diagnosis, therapy and prognosis. In the past, the diagnosis of pathological subtypes mainly relied on hematoxylin-eosin staining and immunohistochemistry. With the development of sequencing technologies, genotype and phenotype analysis of individuals has become possible and precision medicine is on the rise in healthcare. As different tumor subtypes have different cell-of-origin, risk factors and clinical phenotypes, they generate unique combinations of mutation types, termed “Mutational Signatures”. Herein, using the exome sequencing data from The Cancer Genome Atlas (TCGA), we evaluated the utility of mutational landscape for differentiating cell-of-origin within three common cancers (kidney, lung and esophageal cancers). We found that mutational

收稿日期: 2018-07-19; 修回日期: 2018-10-21

作者简介: 史悦, 博士, 助理研究员, 研究方向: 肿瘤表观基因组学。E-mail: shiyue@big.ac.cn

通讯作者: 慈维敏, 博士, 研究员, 研究方向: 肿瘤表观基因组学。E-mail: ciwm@big.ac.cn

DOI: 10.16288/j.ycz.18-208

网络出版时间: 2018/10/30 18:23:43

URI: <http://kns.cnki.net/kcms/detail/11.1913.R.20181030.1822.006.html>

signatures predicted histological subtypes of kidney cancers, clear cell renal cell carcinoma (KIRC) vs. chromophobe renal cell carcinoma (KICH), which had different cell-of-origin, with 100% accuracy (95% *CI*: 0.93–1.00). The mutational signatures also predicted histological subtypes of lung cancers (lung adenocarcinoma vs. lung squamous cell carcinoma) and esophageal cancers (esophageal adenocarcinoma vs. esophageal squamous cell carcinoma) with 78% (95% *CI*: 0.66–0.86) and 84% accuracy (95% *CI*: 0.60–0.97), respectively. Collectively, mutational signatures-based subtyping is good at pathological classification, personalized diagnosis, especially early detection for common cancers.

**Keywords:** mutational signature; pathological subtypes; kidney cancer; lung cancer; esophageal cancer

癌症被认为是一类源于细胞突变的疾病。细胞分裂次数多的组织，罹患肿瘤的几率相对较高。近期研究发现来源于 DNA 复制错误产生的突变约占人类癌症突变总数的 2/3<sup>[1,2]</sup>。该结论引起科学界的强烈辩论和公众的迷惑。研究发现同种肿瘤内也存在巨大的异质性，不同的病理亚型可能细胞来源不同，其细胞复制的活跃程度不同，肿瘤的病理亚型及临床特征都可能不同<sup>[3]</sup>。而每一种突变过程根据 DNA 损伤、DNA 修复或者复制机制等的不同在癌基因组上都留有一定的“印迹”，产生不同的突变组合，即“突变特征”<sup>[4]</sup>。

传统的肿瘤分型主要依赖于 HE 染色以及免疫组织化学法对癌组织进行病理学分析。这就和“活检”取材的大小、部位，肿瘤的进展程度等有一定关联。临床上获取患者肿瘤组织样本只有手术活检及穿刺活检两种方法。如果是早期肿瘤，可能肿瘤细胞本身量少以及形态学并未发生明显变化而导致传统检测方法无法鉴定出肿瘤。但实际上肿瘤细胞的基因组已积累了大量突变，形成了各自特异的“突变特征”。因为不同肿瘤病理亚型，代表的是不同的肿瘤细胞来源，其肿瘤发生发展的致病因素包括突变特征在内会存在一定差异，因此我们认为可以通过不同突变特征区分不同病理亚型，同时肿瘤的突变图谱还可能作为一类生物标记物为我们对某类癌症的诊断，尤其是早期诊断提供参考及帮助。本文拟探索在不同种肿瘤中突变累积的图谱特征和病理分型的关系。利用 TCGA 肾癌、食管癌和肺癌的测序数据我们发现相同病理亚型的肿瘤趋向聚类到一起。一些突变特征还可进一步将细胞来源不同的癌症亚型区分开，而且其中两种肾癌亚型预测的准确率达到 100%。同样地，突变特征在区分肺腺癌和肺鳞状细胞癌以及食管腺癌和食管鳞状细胞癌的准

确率也分别达到了 78% 和 84%。我们认为突变特征可作为一种新型分子标记物，对肿瘤的病理分型、诊断尤其是早诊具有一定的参考意义。

## 1 材料和方法

### 1.1 肿瘤患者相关数据来源和收集

本研究选择了 3 种常见癌症(肾癌、肺癌和食管癌)，深入探索突变特征对提高癌症病理分型鉴定的作用。已知文献报道肾癌的主要病理类型是起源于近曲小管的透明肾细胞癌(clear cell renal cell carcinoma, KIRC)，而嫌色肾细胞癌(chromophobe renal cell carcinoma, KICH)起源于肾集合小管暗细胞<sup>[5]</sup>。而肺腺癌(lung adenocarcinoma, LUAD)主要起源于支气管粘膜上皮，少数起源于大支气管的粘液腺，肺鳞状细胞癌(lung squamous cell carcinoma, LUSC)主要起源于支气管粘膜上皮，是发展中国家的主要病理类型，居于第二位，约占所有病理类型肺癌的 30%<sup>[6]</sup>。与此类似，食管鳞状细胞癌(esophageal squamous cell carcinoma, ESCC)也是一类非常常见的恶性肿瘤，起源于食管的扁平细胞，最常见于食管的上部和中部。而且在亚洲某些地区食管鳞状细胞癌的发病率要高于起源于食管下 1/3 的 Barrett 粘膜的腺管状的食管腺癌(esophageal adenocarcinoma, EAC)<sup>[7]</sup>。

为了鉴定突变特征是否可预测常见癌症的病理分型，本研究收集 TCGA 数据库中透明肾细胞癌和嫌色肾细胞癌患者的全外显子组单核苷酸突变(single nucleotide variants, SNV)数据，其中透明肾细胞癌 451 例，嫌色肾细胞癌 66 例；肺腺癌 543 例，肺鳞状细胞癌 178 例以及 96 例食管鳞状细胞癌患者和 88 例食管腺癌的全外显子组单核苷酸突变数据。

## 1.2 肿瘤特异突变特征的选择

癌症体细胞突变目录(Catalog of Somatic Mutations in Cancer, COSMIC)数据库目前根据癌症基因组数据库(The Cancer Genome Atlas, TCGA)收录的 40 多种不同肿瘤的外显子组和全基因组测序数据, 共发现 6 种 DNA 碱基的变异(C>A、C>G、C>T、T>A、T>C 和 T>G), 考虑到变异碱基上游和下游各有 4 种碱基类型, 则会出现 96 类三核苷酸的突变类型<sup>[3,8-10]</sup>。针对每一个样本的单核苷酸突变都能计算 96 种突变形式的比例。由于 DNA 复制、DNA 损伤物质暴露、修复缺失等不同原因产生的突变类型具有某种特异性即 96 种突变形式的比例组成不同, 这被称作该种肿瘤的突变特征<sup>[4]</sup>。

## 1.3 突变特征分析方法

突变特征计算: 首先利用 deconstructsig R (<https://cran.r-project.org/web/packages/deconstructSig/index.html>) 对从 TCGA 下载的肾癌、肺癌、食管癌外显子数据进行突变特征分析做突变特征分析<sup>[11]</sup>。考虑到某些肿瘤突变负荷(tumor mutation burden)低, 处于外显子的突变数目可能少于 96 个, 因此本研究进一步将突变类型简化, 将 96 个突变形式中第一个碱基例如 AT>CG, CT>CG, TT>CG, GT>CG 合并为一种 T>CG, 最终成为 24 种突变形式。

突变特征预测模型分析: 对各样本进行随机排序并按照 1:9 模式将数据拆分成测试集合和训练集合。接着利用 randomForest (<https://cran.r-project.org/src/contrib/Archive/randomForest/>) R 包进行建模并用测试集合进行模型检验。最后结合 ROCR 包 (<https://cran.r-project.org/src/contrib/Archive/ROCR/>) 进行 ROC 曲线展示模型预测效果。

# 2 结果与分析

## 2.1 突变特征可以准确、特异地区分两种不同病理亚型的肾癌

本研究下载 TCGA 中肾癌的外显子组数据, 首先关注两种不同病理亚型的肾癌: 来源于近曲小管的透明肾细胞癌和来源于远曲小管的嫌色细胞癌

(图 1A)。每组各选择了 1 个典型病例, 检测对应的 24 个突变形式的分布情况。图 1B 表明, 两个不同病理亚型的肾癌突变特征显著不同, KIRC 以 T>A 变异比例最高, 而 KICH 以 C>T 的变异比例最高。进一步采用非监督聚类的方法, 结果发现虽然同一病理亚型的患者相同突变类型频率存在差异, 但同一病理亚型的患者更趋向聚类到一起(图 1C)。为进一步确定这 24 种突变特征哪些能够将同种癌症不同病理亚型有效分开, 随后用随机森林法进行了特征选择分析。与预期一致, 随机森林法选择出来的突变特征可以很好的将细胞来源不同的两组肾癌亚型 KIRC 和 KICH 分开, 而且准确率达到了 100% (95% confidence interval (CI): 0.93~1.00) (图 1D)。上述结果表明, 突变特征可以准确、特异地区分两种不同细胞来源的肾癌病理亚型。

## 2.2 突变特征可以准确、特异地区分两种不同病理亚型的肺癌和食管癌

鳞癌和腺癌是多种肿瘤中都存在的两个病理亚型, 其治疗方式和对放化疗的敏感性均存在显著差异, 因此病理类型的诊断对于癌症病人来说至关重要。为进一步探究通过突变特征预测肿瘤病理分型的可行性和准确度, 本研究继续选取了两种常见的肿瘤肺癌和食管癌进行突变特征与肿瘤病理亚型(鳞癌和腺癌)的相关性和预测研究。类似地, 在肺癌和食管癌中, 非监督聚类方法显示: 同一病理类型的患者趋向聚类在一起(图 2, A 和 B)。进一步用随机森林法选择可以用来区分两种病理类型的突变特征, 在肺癌中分类准确性达到了 78% (95% CI: 0.66~0.86) (图 2C), 而在食管癌中达到了 84% (95% CI: 0.60~0.97) (图 2D)。并且, 在不同类型肿瘤中相同病理亚型的患者突变特征总体比较相似, 略有不同, 暗示细胞来源以及致癌因素共同刻画了肿瘤中的突变特征(图 2E)。上述结果表明, 突变特征本身在很大程度上就可以代表了常见癌症不同病理亚型的“自然属性”, 从而将其区分开来。

# 3 讨论

病理类型的诊断对于癌症病人来说至关重要,

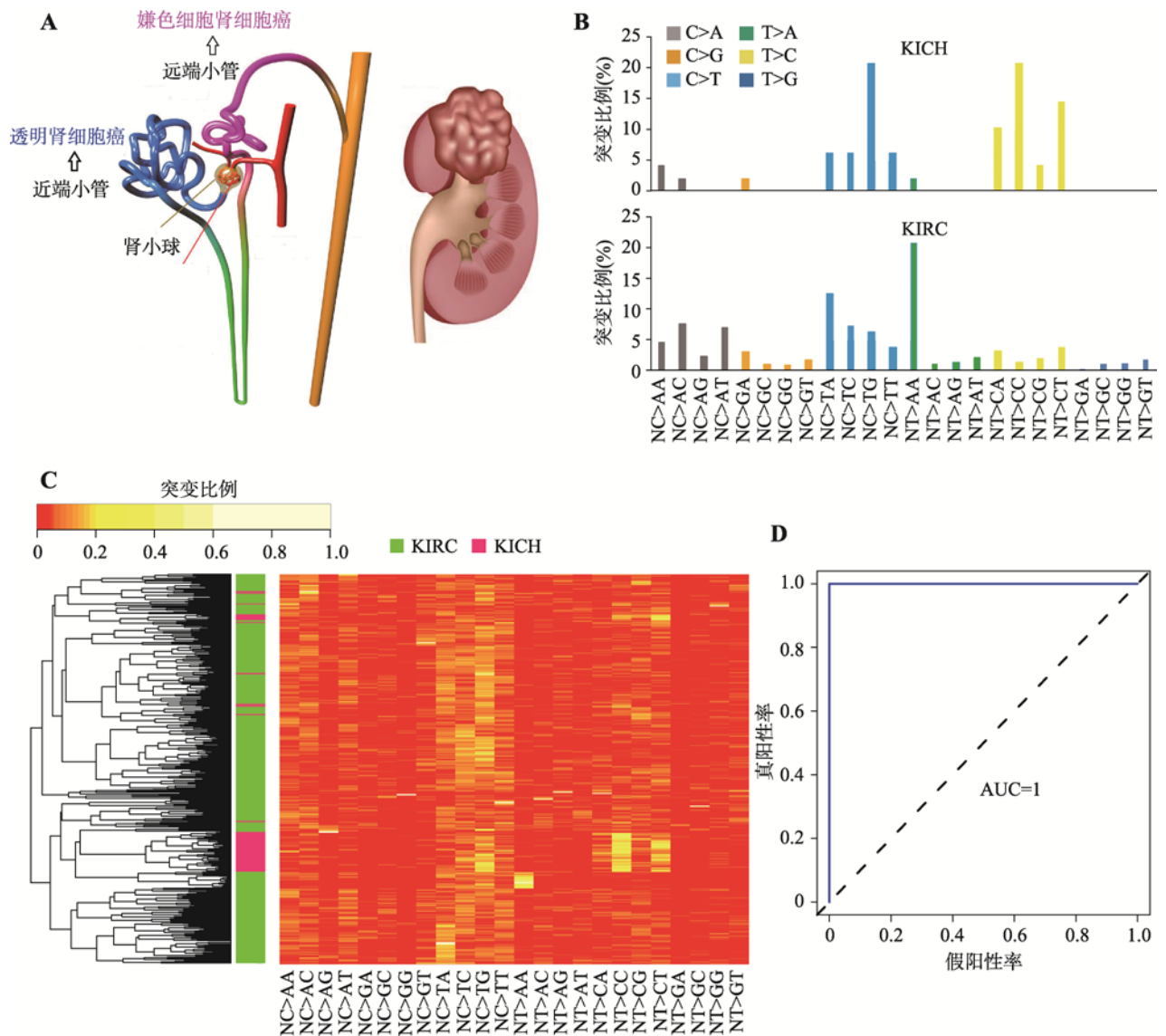


图 1 透明肾细胞癌和嫌色肾细胞癌能够被突变特征准确、特异地区分

Fig. 1 Mutational signatures predict histological subtypes within kidney cancers

A: 肾癌两种不同病理亚型的细胞来源示意图。B: 肿瘤患者的突变特征谱。两个肾癌亚型中分别选取 1 例患者。C: 突变特征谱的非监督聚类。D: 随机森林法选取突变特征对肾癌两种不同亚型分类准确性的 ROC 曲线。KIRC 和 KICH 分别代表透明肾细胞癌和嫌色肾细胞癌。

因为其能预测病人对不同治疗方式的敏感度以及药物的毒性反应。经典的癌症病理类型鉴定是利用 HE 染色以及免疫组织化学法对癌组织进行病理学分析, 而对于穿刺等活检样本(样本的组织空间结构信息不完整)的诊断灵敏性有待提高。如早期肿瘤患者, 有可能由于处于进展早期以及取材的限制, 利用经典的 HE 染色和免疫组化法不一定能观察到肿瘤细胞, 及其组织形态的明显改变。但实际上这些肿瘤细胞已积累了相当多的突变, 形成了各自的“突变

特征”。如果利用基因组测序和分析技术就可以弥补这一不足, 提高病理诊断的灵敏度。本文尝试将病理分型和基因组学相结合, 展示了利用突变特征对 3 种常见肿瘤不同病理亚型进行分类的可行性。研究结果也表明突变特征可以作为辅助手段来帮助更多不同种癌症的鉴定。

美国食品药品监督管理局在 2017 年 5 月 23 号批准同意了首款不依据肿瘤来源, 而是依据生物标志物进行区分的抗肿瘤药物——Keytruda<sup>[12]</sup>, 用于治疗携



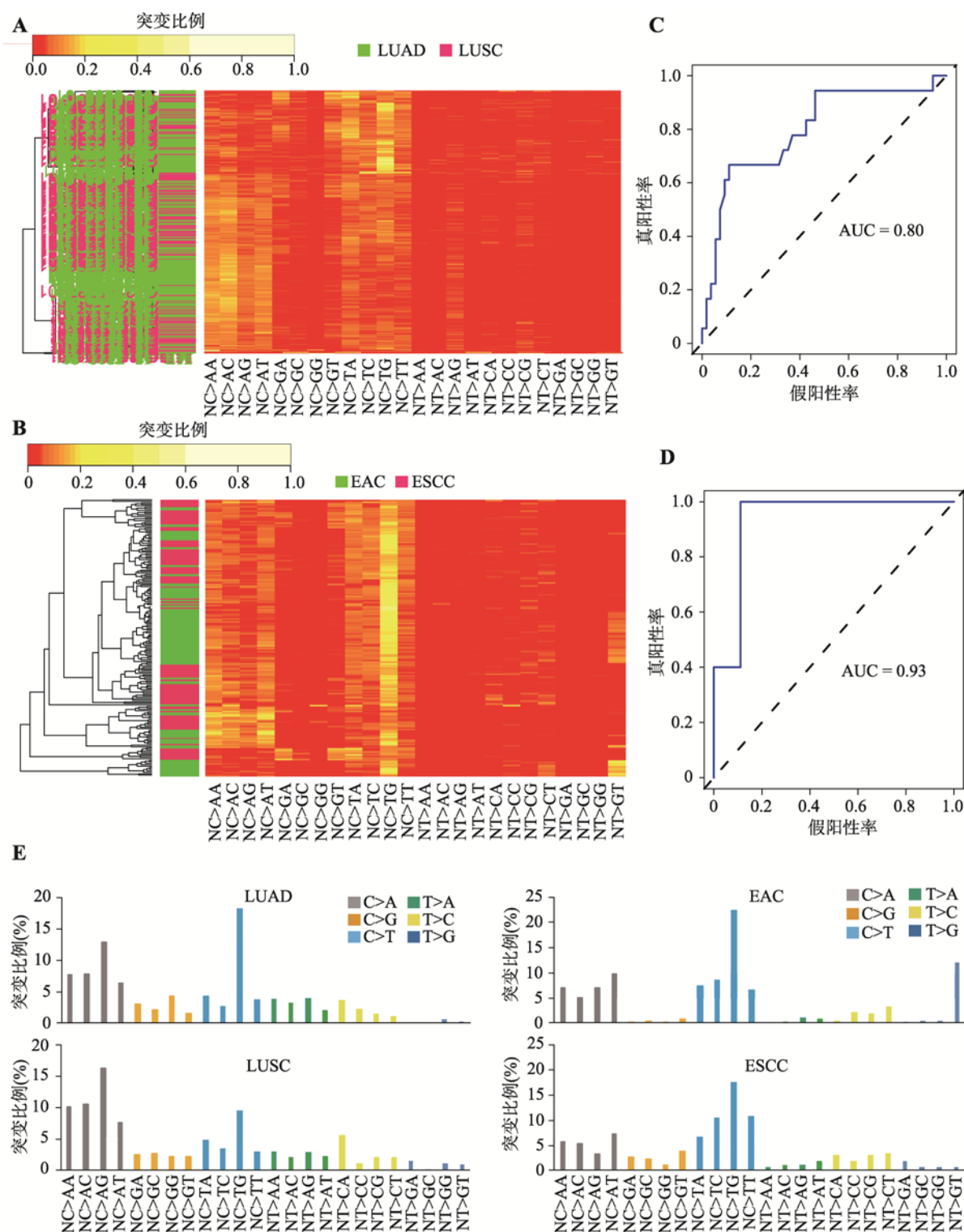


图 2 在肺癌和食管癌中突变特征可以准确、特异地区分各自不同的病理亚型

Fig. 2 Mutational signatures predict histological subtypes within lung cancers and esophageal cancers

A: 肺癌两种不同病理亚型的突变特征谱非监督聚类; B: 食管癌两种不同病理亚型的突变特征谱非监督聚类; C: 随机森林法选取突变特征对肺癌两种不同亚型分类准确性的 ROC 曲线; D: 随机森林法选取突变特征对食管癌两种不同亚型分类准确性的 ROC 曲线; E: 肿瘤患者的突变特征谱。肺癌和食管癌中两种亚型中分别选取 1 例患者。LUAD 和 LUSC 分别代表肺腺癌和肺鳞状细胞癌。EAC 和 ESCC 分别代表食管腺癌和食管鳞状细胞癌。

带某特定基因特征的多种实体瘤。Keytruda 是针对 PD-1/PD-L1 通路发挥的作用。通过阻碍这一通路, Keytruda 可帮助人体免疫系统对抗癌细胞。而以往 FDA 一直依据癌症的发病部位批准治疗方法, 如肺癌, 胃癌等。这是第一次基于肿瘤生物标志物而非肿瘤的发生部位批准的一个新药。这预示着今后深入探索突变特征在癌症病人精准化医疗的临床应用价值将是一个有利方向。例如本文所述的突变特征就可能作为一类早期诊断中液体活检或者临床预后等的生物标记。相反, 如果仅仅只检测体细胞特定基因的突变对鉴定某类癌症的敏感性和特异性均不佳, 不仅突变频率低, 而且重现性突变比例也很低<sup>[13]</sup>。因此, 利用常见的体细胞染色体拷贝数变化或者突变特征作为生物标记分子, 预期可提高早诊的特异性和敏感度。综上所述, 突变特征作为新型分子标记物, 对肿瘤的病理分型、诊断尤其是早诊的预测具有一定的参考意义。

## 参考文献(References):

- [1] Tomasetti C, Vogelstein B. Cancer etiology, Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, 2015, 347(6217): 78–81. [DOI]
- [2] Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*, 2017, 355(6331): 1330–1334. [DOI]
- [3] Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet*, 2014, 15(9): 585–598. [DOI]
- [4] Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilcic T, Imbeaud S, Imielinski M, Jäger N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdés-Mas R, van Buuren MM, van't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MML-Seq Consortium, ICGC PedBrain, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. Signatures of mutational processes in human cancer. *Nature*, 2013, 500(7463): 415–421. [DOI]
- [5] Prasad SR, Narra VR, Shah R, Humphrey PA, Jagirdar J, Catena JR, Dalrymple NC, Siegel CL. Segmental disorders of the nephron: histopathological and imaging perspective. *Br J Radiol*, 2007, 80(956): 593–602. [DOI]
- [6] Ginsberg MS, Grewal RK, Heelan RT. Lung cancer. *Radiol Clin North Am*, 2007, 45(1): 21–43. [DOI]
- [7] Zhao P, Dai M, Chen W, Li N. Cancer trends in China. *Jpn J Clin Oncol*, 2010, 40(4): 281–285. [DOI]
- [8] Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep*, 2013, 3(1): 246–259. [DOI]
- [9] Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev*, 2014, 24: 52–60. [DOI]
- [10] Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, Jia M, Jubb H, Sondka Z, Thompson S, De T, Campbell PJ. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*, 2017, 45(d1): D777–D783. [DOI]
- [11] Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*, 2016, 17: 31. [DOI]
- [12] Garber K. In a major shift, cancer drugs go 'tissue-agnostic'. *Science*, 2017, 356 (6343): 1111–1112. [DOI]
- [13] Kandath C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L. Mutational landscape and significance across 12 major cancer types. *Nature*, 2013, 502(7471): 333–339. [DOI]

(责任编辑: 孙玉洁)