

· 应用研究 ·

基于增强子调控化疗相关基因的结直肠癌分型研究*

哈尔滨医科大学卫生统计学教研室(150081) 田伟 严光灿 张秋菊 刘美娜[△]

【摘要】目的 利用增强子 RNA 调控的化疗相关基因对结直肠癌患者进行分型,为精准医疗提供依据。**方法** 通过 TCGA、GEO 数据库获取结直肠癌患者基因表达数据,筛选增强子调控的化疗相关基因,利用非负矩阵分解方法进行直肠癌患者分型,并在独立数据集中验证分型结果的稳定性。通过生存分析、GSVA、SubMap 方法,比较各亚型患者在生存结局、代谢通路活性、免疫浸润细胞浸润等方面的差异。收集结直肠癌新辅助化疗患者血浆,检测血浆蛋白表达量,分析增强子 RNA(eRNA)调控基因与结果化疗敏感性的相关性。**结果** 本研究收集了 581 人的 TCGA 基因表达数据,筛选 50 个 eRNA 调控的化疗相关基因,通过非负矩阵分解获得三个结直肠癌亚型:eRNA1、eRNA2、eRNA3,GEO 验证集的分型结果与 TCGA 基因表达数据一致,均分为 3 个亚型。生存分析、代谢通路分析、免疫浸润分析提示:eRNA2,生存结局最差,32 个代谢相关通路异常激活、免疫浸润程度明显偏低。在蛋白水平上,eRNA 调控基因与化疗敏感性存在关联性。**结论** 在缺乏有效结直肠癌临床分型的前提下,本研究通过 eRNA 调控的化疗相关靶基因可以进行结直肠癌稳定有效的分型;各亚型在生存结局、免疫细胞浸润、代谢相关通路研究中的差异,可为临床个体化治疗提供指导,并为相关机制研究提供线索。

【关键词】 结直肠癌 增强子 RNA 化疗 代谢通路**【中图分类号】** R195.1**【文献标识码】** A**DOI** 10.11783/j.issn.1002-3674.2024.01.009

结直肠癌发生率高、异质性强。研究结直肠癌的肿瘤异质性,进行病人合理的亚型划分,提供针对性治疗方案成为当下实现精准医疗、改善治疗结局的重要手段。已有研究通过全基因组测序、表观遗传学分析、基因表达谱分析、肿瘤免疫微环境分析等对结直肠癌异质性进行多方面探索,如基因突变、染色体突变、癌症相关基因的启动子甲基化、微卫星不稳定性,肿瘤微环境构成等^[1-2]。增强子作为重要的基因调控元件,不仅直接参与基因的表达调控,也可直接转录出相应的 RNA 序列,即增强子 RNA(enhancer RNA, eRNA)。eRNA 在各类组织、器官中广泛表达,且表达具有组织、器官特异性^[3-4],eRNA 应成为研究结直肠癌亚型的重要工具。化疗是影响结直肠癌生存结局的重要因素,其效果受多方因素影响,如化疗相关基因活性、肿瘤免疫微环境等。eRNA 作为调节靶基因活性的重要转录调控因子,亦可通过影响化疗相关基因的活性,影响化疗效果及治疗结局。因此,本研究基于 eRNA 调控的化疗相关靶基因,探索结直肠癌新亚分型及亚型生物差异特征,进行合理准确的结直肠癌分型及化疗敏感性相关机制研究,为促进个体化治疗、精准医疗及相关机制研究提供依据。

原理和方法

1. 数据收集及样本筛选

下载癌症基因组图谱数据库(the cancer genome atlas, TCGA; <https://portal.gdc.cancer.gov/>) 结直

肠癌的基因表达、生存及临床表型数据,其中基因表达数据作为训练集,用于癌症分型。下载基因表达数据库(gene expression omnibus, GEO; <https://www.ncbi.nlm.nih.gov/geo/>) 多个结直肠癌数据集,包括:GSE62080, GSE39582, GSE14333, GSE17536, GSE17537, GSE33113 和 GSE37892。GSE62080 用于寻找化疗相关基因,GSE39582 作为第一个验证集,后 5 个数据集经批次校正后合并,作为第二个验证集(GEO-5 数据集)。eRNA 调控基因则由 eRic(<https://hanlab.uth.edu/eRic/>) 数据库提供。病人纳入标准:①原发性肿瘤;②总生存期大于 30 天。排除标准:①生存数据不完整;②eRNA 表达数据缺失。

2. NMF 法进行结直肠癌分型

通过 limma 包计算化疗相关差异基因,并根据 eRic 数据库提供的 eRNA-靶基因调控关系,获取 eRNA 调控的化疗相关基因。基于非负矩阵分解(nonnegative matrix factorization, NMF)方法寻找结直肠癌的潜在亚型,并根据 Cophenetic 参数确定最优亚组数,具体标准为:Cophenetic 参数下降最快的亚组数。

3. 免疫微环境及代谢通路活性分析

获得分型后,通过 CIBERSROT 算法评估各亚型患者肿瘤微环境中免疫细胞的浸润程度,该算法以支持向量回归为核心,并引入去卷积化过程,实现基因表达数据对免疫细胞成分的估算^[5]。基因集变异分析算法(gene set variation analysis, GSVA)实现,其为非参数、无监督的分析方法,根据基因表达数据估计通路活性的变化,如代谢相关通路^[6]。

4. 免疫治疗敏感性分析

* 基金项目:国家自然科学基金面上项目(82073666)

[△]通信作者:刘美娜, E-mail: liumeina369@163.com

各亚组的免疫治疗效果评估主要从两个方面进行：第一，肿瘤免疫功能障碍和排斥 (tumor immune dysfunction and exclusion, TIDE) 评分；第二，基于真实免疫治疗队列的子类映射算法 (subclass mapping, submap) 的关联性分析, 包括 GSE78220, GSE100797 和 Roh 队列^[7]。

5. 结直肠癌患者血浆蛋白比较

收集 17 名新辅助结直肠癌患者的血浆样本, 根据其化疗敏感性将其分为敏感组、耐药组。通过非靶向 TMT (tandem mass tag) 标记定量蛋白质组技术, 检测患者血浆蛋白表达量丰度。使用非参数的 Wilcoxon 秩和检验比较新辅助敏感与耐药病人间的 eRNA 调控的化疗相关基因在血浆蛋白水平上的差异。本文中, 所有统计分析均以 $\alpha = 0.05$ 为检验水准。

结 果

1. 结直肠癌 eRNA 调控基因筛选

按纳入排除标准, 选取 581 名结直肠癌患者为训练集; 验证集两个, 包括: 573 名 GSE39582 患者、663 名 GEO-5 患者。结直肠癌共表达 433 个 eRNA, 包括: ENSR00000000137、ENSR00000000138、ENSR00000018454、ENSR00000018455、ENSR00000018456 等; 调控 1081 个基因, 包括: TAS1R3、CDK11A、DFFB、DISP2、UBE2I 等。

2. 结直肠癌分型

GSE62080 数据集中共有 21 例病人, 其中 9 人对化疗敏感, 12 人对化疗不敏感; 化疗方案为: 亚叶酸 + 5-氟尿嘧啶 + 伊立替康 (folinic acid, fluorouracil

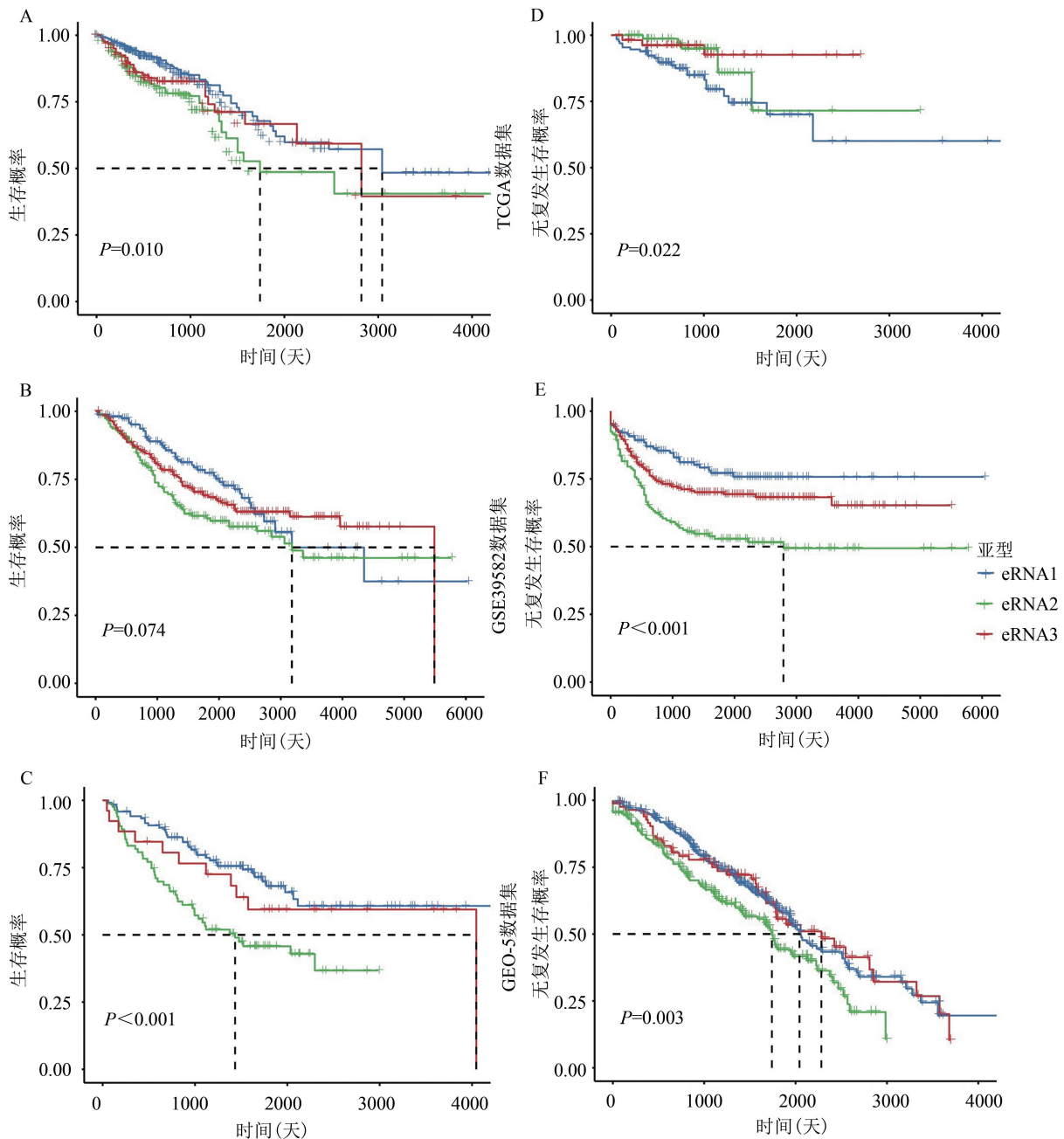


图 1 各亚组的总生存曲线与无复发生存曲线

and irinotecan, FOLFIRI), 为结直肠癌一线化疗方案。在检测的 23521 个基因中, 以 limma 包分析结果中的 $P < 0.05$ 为标准筛选出 1073 个差异基因, 其中 50 个基因由 eRNA 调控, 包括: PEX10、ZSCAN20、PIAS3、TAS1R3、CDK11A 等。以上述 50 个靶基因为基准, 分别对 TCGA、GSE39582 和 GEO-5 三个数据集的患者进行 NMF 亚型聚类, 结果均显示 3 个亚型为最优, 分别为 eRNA1, eRNA2 和 eRNA3。经 NMF 分析后, 在三个数据集中分别绘制 3 个亚型的生存曲线、无复发生存曲线, eRNA2 组的总生存率较低、eRNA1 组的总生存率较高, 见图 1A~C; 患者的无复发生存也有类似的情况, 见图 1E~F; TCGA 数据中 eRNA2 亚型出现无复发生存较高, 见图 1D, 这可能与 TCGA 复发数据缺失过多有关。

3. 免疫微环境分析

eRNA2 与 eRNA1、eRNA3 相比, 在 eRNA2 的肿瘤免疫微环境中, 大多数免疫细胞的浸润程度都较低, 如: 活化 CD4 T 细胞, 活化 CD8 T 细胞, 记忆 B 细胞, 调节性 T 细胞, 自然杀伤(nature killer, NK) 细胞等; 仅部分细胞的浸润程度较高, 如: 单核细胞, CD56dim NK 细胞等, 见图 2A。但在与单核细胞, CD56dim NK 细胞相关的免疫调节因子 CD16、NKG2D 的基因比较发现: 这两个辅助因子在 eRNA2 亚型中低表达, 见图 2B~C。

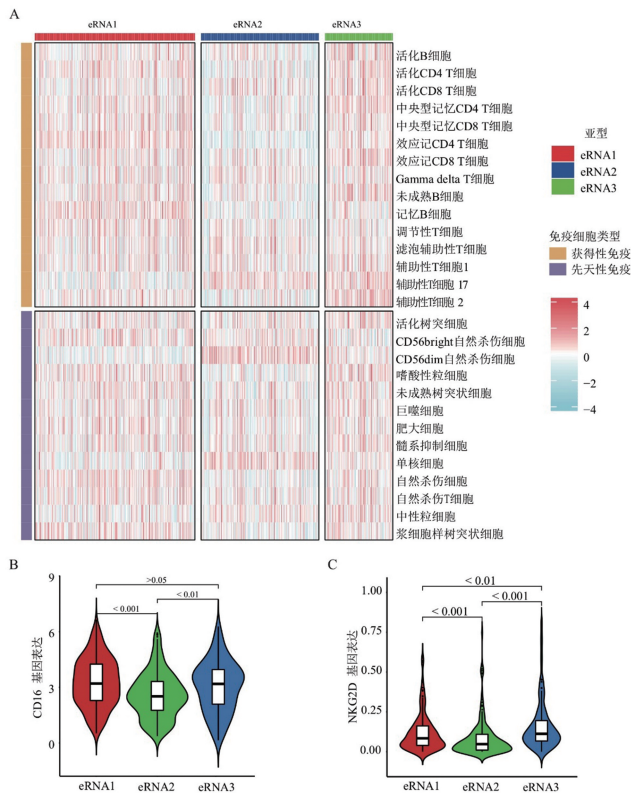


图 2 免疫细胞浸润分析

4. 通路活性分析

在代谢相关通路中, eRNA-2 亚型过度激活的通路最多, 为 32 个, 包括心磷脂代谢, 嘧啶合成等, 见图 3; eRNA-3 亚型过度激活的通路次之, 为 12 个, 包括: 环氧合酶花生四烯酸代谢, 视黄醇代谢等; eRNA1 组过度激活的通路最少, 仅有嘌呤合成、糖原合成两个代谢通路。

5. 免疫治疗效果评估

在 TIDE 方法评估的亚型潜在免疫治疗效果中, eRNA2 亚型可能有效的患者比例低于另两个亚型, 见图 4A; SubMap 的分析结果同样不支持 eRNA2 会从中受益的结论。GSE78220 的 SubMap 分析结果提示 eRNA1 亚型组可能从 PD1 治疗中受益, 见图 4B; Roh 队列的分析结果可以得出类似结论, 见图 4D。

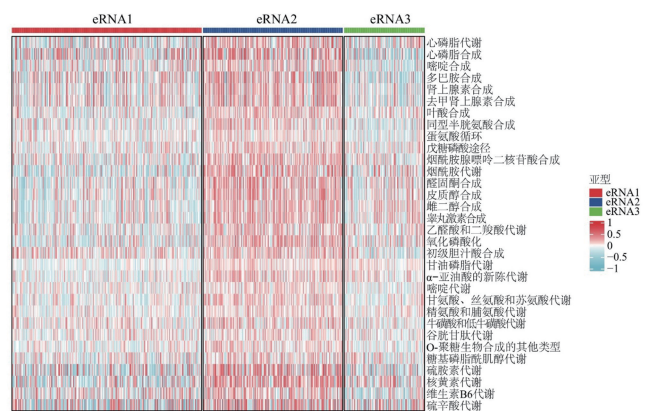


图 3 eRNA2 亚型异常激活代谢通路

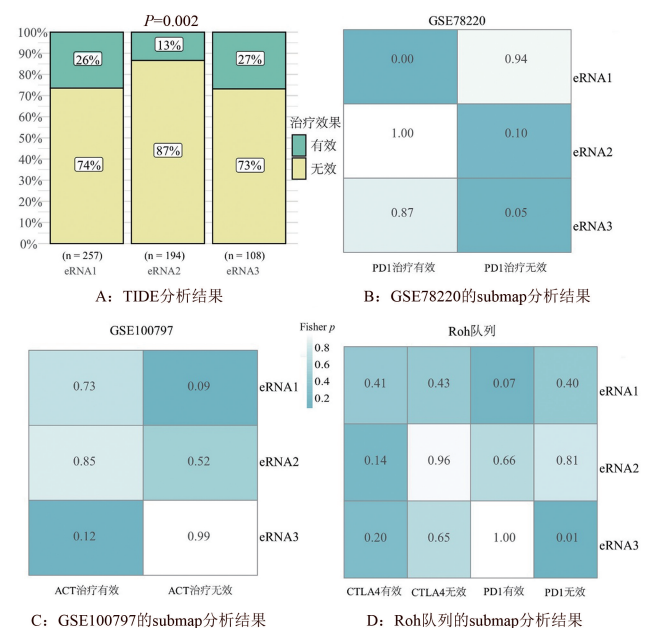


图 4 免疫治疗效果评价

6. eRNA 调控化疗相关基因血浆蛋白比较

17 位接受新辅助化疗的结直肠癌患者中, 新辅助化疗耐药组 6 人、敏感组 11 人。常见的肿瘤标志物 CEA 和 CA199, 在敏感组与耐药组间差别无统计学意义; 仅有与评价化疗敏感性直接相关的肿瘤直径、肿瘤厚度等指标, 在两组间差异具有统计学意义 (Wilcoxon 秩和检验), 见表 1。TMT 共检测出 934

个蛋白,其中两个为 eRNA 调控化疗相关基因所表达的蛋白,即 NAP1L4 与 PROZ;两蛋白在化疗敏感组与耐药组间存在一定差异,见图 5。

表 1 新辅助化疗患者基本信息

变量	水平	耐药组	敏感组	P
人数		6	11	
性别	男	3	5	
	女	3	6	
年龄		52.67(7.66)	52.64(7.26)	0.88
术前 CEA		13.02(17.82)	8.57(7.47)	0.96
术后 CEA		6.86(9.48)	3.21(2.78)	0.66
术前 CA199		29.49(28.56)	22.94(17.78)	0.88
术后 CA199		13.59(10.20)	21.83(25.75)	0.73
术前肿瘤直径		55.17(7.73)	39.55(13.49)	0.02
术后肿瘤直径		46.50(17.44)	26.25(13.31)	0.02
术前肿瘤厚度		18.17(3.19)	13.18(3.60)	0.02
术后肿瘤厚度		16.33(4.93)	8.62(2.72)	0.01

注:表中连续型变量表示为:均值(方差);离散型变量表示为:样本量。

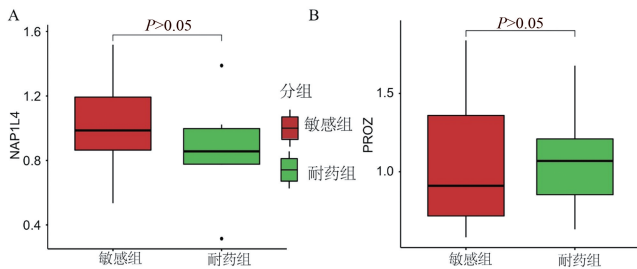


图 5 新辅助化疗敏感组与耐药组血浆蛋白比较

讨论

结直肠癌是具有高度异质性的消化系统常见肿瘤,导致其高异质性的原因很多,包括基因突变、染色体变异、肿瘤微环境改变,以及肠道微生物群落改变等^[2]。eRNA 作为高度特异性的 DNA 序列表达产物,其在调节基因表达、影响肿瘤免疫微环境、影响生存方面的作用已经得到多方面证实^[8]。有研究表明,eRNA 调控基因在胶质瘤亚型研究中表现突出^[9]。FOLFIRI 作为结直肠癌的一线化疗方案已在临床实践中得到广泛应用,患者生存结局得到显著改善,患者中位生存期提高至 15 个月以上^[10];但是,仍有 50% 左右的患者对化疗药物不敏感,不仅不能从中受益,反而承受过重的经济负担与健康损害。因此,对结直肠癌患者的准确分型、化疗相关机制的深入研究显得极为重要。已有研究者关注化疗相关基因在癌症分型中的重要作用,通过化疗相关基因对癌症患者实现了更为精细化的分型,提供了探讨可能影响生存、复发、化疗敏感性的潜在机制^[2, 11]。但是,以靶基因为研究起点,而没有分析上游调控机制,会使得相关的机制研究不完整,本研究是对该缺失的重要补充。

通过 50 个 eRNA 调控的化疗相关基因,实现稳

健的结直肠癌患者分型,三个数据集的分析结果均提示结直肠癌中存在三个亚型,且分型结果具有明确的临床意义。在 TCGA 数据集中,eRNA2 组复发风险较低,与其余数据集中的表现不同,这可能是由于 TCGA 数据集中复发数据缺失过多造成的。在 eRNA 调控的基因中,NAP1L4、PROZ 蛋白在新辅助化疗敏感人群与耐药人群中得到验证。虽然由于样本量较少,差异检验结果在统计学上没有意义,但仍可直观看出 NAP1L4 在敏感组高表达、PROZ 在敏感组低表达的直观现象,这对于未来研究结直肠癌化疗敏感性,提供了一定参考。

肿瘤免疫微环境是影响结直肠癌治疗效果和生存预后的重要因素,探索不同亚型的肿瘤免疫微环境差异,利于理解免疫微环境中各类因子在癌症发生、发展过程中的作用。对于生存预后最差、复发风险最高的 eRNA2 亚型组,其在重要的机体免疫防御细胞中存在着免疫细胞浸润不足的问题。NK 细胞、活化 CD8 T 细胞作为重要的杀癌细胞成分,在 eRNA2 亚型中的缺失可能是导致该亚型生存期短、复发风险高的原因之一。虽然有部分原本可以发挥重要功能的免疫细胞在该亚型组织中富集,但是由于这部分免疫细胞的调节因子在癌组织中低表达,而使得癌细胞免疫逃逸增强,异常细胞被无法识别。如,CD56dim NK 细胞,虽是最为重要的功能性 NK 细胞,其免疫应答过程无需经过抗原提呈过程,可直接裂解靶细胞,但是其功能的发挥依赖 CD16 基因的高表达,而 eRNA2 亚型的 CD16 基因却低表达,这可能是 eRNA2 亚型患者 CD56dim NK 细胞免疫细胞无法发挥其功能^[12-13],生存结局差、复发风险高的原因之一。NKG2D 基因的低表达同样会增强肿瘤细胞的免疫逃逸功能^[14],在 eRNA2 亚型患者存在同样的问题。

综上,本研究基于 eRNA 调节的化疗相关基因,对结直肠癌患者进行癌症分型研究。经两个独立大型外部数据验证,结直肠癌患者可分为三个分子亚型,分型结果稳定可靠。不同亚型的免疫相关分析、代谢通路活性分析,发现导致结直肠癌患者生存预后不佳的潜在相关机制信息,可为后续相关机制研究提供参考。

参考文献

[1] Fridman WH, Pagès F, Sautès-Fridman C, et al. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer*, 2012, 12(4): 298-306.

[2] Zhu X, Tian X, Ji L, et al. A tumor microenvironment-specific gene expression signature predicts chemotherapy resistance in colorectal cancer patients. *NPJ Precis Oncol*, 2021, 5(1): 7.

[3] Lee JH, Xiong F, Li W. Enhancer RNAs in cancer: regulation, mechanisms and therapeutic potential. *RNA Biol*, 2020, 17(11): 1550-1559.

(下转第 52 页)