

人类肺腺癌中特异性 可变剪接事件的分析与识别*

高洁¹, 高源源², 张利绒¹

(1. 内蒙古大学物理科学与技术学院, 呼和浩特 010021; 2. 内蒙古自治区中蒙医药研究院, 呼和浩特 010020)

摘要:前体mRNA(pre-mRNA)的异常剪接会导致癌症, 研究证实可变剪接(Alternative splicing, AS)在人类肺腺癌(Lung adenocarcinoma, LUAD)的增殖和迁移中起着至关重要的作用。基于LUAD组织和癌旁正常组织的RNA-Seq数据, 构建了3种常见模式的AS事件集, 识别并分析了LUAD和正常组织中的3类高频AS事件, 即筛选得到了一致性好($H=0$)、高重复率($m>18$)的51个事件作为LUAD的特异性AS事件。进一步, 对这些AS事件及其相应的特征基因进行功能分析和通路分析, 发现11个基因与癌症的发生发展密切相关, 14个基因在其分子功能上可能参与癌症的迁移和增殖, 其中NUMB和POSTN两个基因已被证实与肺癌有关。这些AS事件和基因可能适合作为LUAD的潜在诊断生物标志物。

关键词:可变剪接; 肺腺癌; 信息熵

中图分类号:Q61 **文献标志码:**A

在RNA剪接过程中, 同一个pre-mRNA会因为选择了不同的剪接位点, 从而产生不同的剪接异构体, 这个过程被称为可变剪接(Alternative splicing, AS), 也常被称为选择性剪接^[1]。有研究指出, 人类基因组中约95%的多外显子基因存在可变剪接现象^[2-3]。在可变剪接中, 断裂基因的GT-AG法则是识别剪接信号的一个重要依据。断裂基因的一个重要特点是每个外显子和内含子的连接处有一段高度保守的序列, 特别是几乎所有的内含子5'端都以GT二联体为起始, 3'端以AG二联体为结束, 这种连接方式在真核生物中普遍存在, 因此被人们称为GT-AG法则^[4]。可变剪接与多种疾病的发生发展密切相关, 它可以通过作用于癌细胞的不同生物学过程来影响肿瘤的进展, 既往研究显示可变剪接事件发生在癌组织中的概率高于正常组织, 可变剪接在外显子和内含子上均可发生^[5-10]。因此, 疾病相关可变剪接事件的鉴定有助于癌症的早期发现和治疗。

肺癌是一种全球范围内发病率和死亡率都极高的肿瘤, 2022年发表于《中华医学杂志》的一项研究指出肺癌是我国癌症死亡的主要原因^[11]。肺癌主要分为小细胞肺癌和非小细胞肺癌, 其中, 肺腺癌(Lung adenocarcinoma, LUAD)是非小细胞肺癌的一种常见亚型, 它的发病率和致死率位居前列。在非小细胞肺癌中, 基因BCL2L1的不同剪接异构体会影响Bcl-XL蛋白中的BH1、BH2和BH3结构域, 进而调控癌细胞的凋亡过程^[12-14]。Zhang等^[15]发现位于基因MYH14和ESYT2中的两个显著差异外显子跳跃事件与LUAD密切相关。Gorgoulis等^[16]发现基因MDM2促进肺癌的发生发展, 该基

* 收稿日期:2025-08-29; 修回日期:2025-11-09

基金项目:内蒙古自然科学基金项目(2024MS03023);2024年内蒙古大学研究生科研创新项目(11200-5223737)

作者简介:高洁(1998—),女,内蒙古乌兰察布人,2023级博士研究生。主要从事理论生物物理与生物信息学研究。E-mail:15734743671@163.com

通信作者:张利绒(1972—),女,内蒙古乌兰察布人,教授,博士。主要从事理论生物物理研究。E-mail:py-zlr@imu.edu.cn

因的可变剪接对癌细胞中的p53蛋白有重要的调节作用。Moll等^[17]指出基因*MDM2*本身由p53诱导。Lin等^[18]发现*PRSS3*转录本及其亚型在肺癌中表达不同,通过靶向不同的下游基因表现出相反的功能和临床结果,通过位点特异性iCpGI去甲基化可使*MZF1*上调*PRSS3-V3*的表达,在肺腺癌细胞中发挥抗肿瘤作用。*RBM10*已被证明通过跳过外显子9来影响Notch调节基因*NUMB*的选择性剪接,*NUMB*中外显子9的保留(促增殖)是肺癌中最常见的选择性剪接之一^[19-21]。Yan等^[22]发现*RAC1B*能够促进LUAD细胞增殖和肿瘤生长,而在LUAD中经常突变的剪接因子*RBM10*被进一步鉴定为*RAC1B*的负调节因子。因此,基因的可变剪接在癌症的发生发展过程中起着重要的调控作用。

基于GEO数据库(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40419>)^[23]的77个LUAD样本和癌旁正常样本对的RNA-Seq数据,识别了其中发生的差异AS事件,通过分析这些差异可变剪接事件在LUAD样本和癌旁正常样本中的分布,筛选得到在LUAD中特异性的AS事件,并对这些AS事件进行了功能分析。

1 数据与方法

1.1 基于RNA-Seq数据构建AS事件集

本文所用的RNA-Seq数据源自GEO数据库的GSE40419数据集,该数据集提供了87个LUAD样本和77个癌旁正常样本的表达谱数据,其中一一匹配成对的有77对样本。基于这77对RNA-seq数据构建了可变剪接事件集。

如图1所示,针对3种常见的AS模式:可变3'剪接位点(A3SS)、可变5'剪接位点(A5SS)和外显子跳跃(SE),使用rMATS软件对样本的AS事件进行定量分析,得到样本中AS事件中的剪接率(IncLevel),即A3SS和A5SS事件选择近端剪接位点(Proximal)和SE事件选择包含中间外显子的概率,差异剪接率(IncLevelDifference)即LUAD样本和癌旁正常样本之间的剪接率差值,以及Benjamini-Hochberg校正后的P值和错误发现率(False discovery rate, FDR)。在识别AS事件过程中,设定发生可变剪接的阈值为0.1, FDR值为0.5,得到LUAD样本和癌旁正常样本对的差异剪接事件。

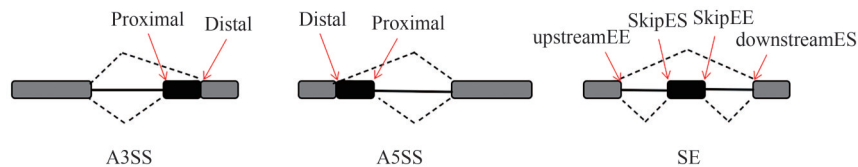


图1 A3SS、A5SS和SE事件中AS位点的命名

Fig. 1 Names of AS sites in A3SS, A5SS and SE events

通过整合77对样本的差异剪接事件,构建了AS事件集1(Set1)。在Set1中,剔除X/Y染色体上发生的AS事件,建立AS事件集2(Set2)。进一步,对AS事件去重,最终得到LUAD样本与癌旁正常样本差异的AS事件集合3(Set3),结果如表1所示。

表1 77个LUAD-癌旁正常样本对中3种AS事件数

Table 1 Number of three AS events in 77 LUAD-adjacent normal pairs

事件类型	Set1	X/Y	Set2	重复事件数	Set3
A3SS	10 633	374	10 259	6 998	3 261
A5SS	8 490	267	8 223	5 837	2 386
SE	64 534	2 140	62 394	40 828	21 566

注:同一AS事件可在多个样本对中发生。Set1为77对样本中全部的AS事件集;Set2为剔除X/Y染色体上发生的AS事件集,包括重复的AS事件;Set3为去重后的AS事件集。

1.2 特异性AS事件的识别方法

为了观察AS事件在不同样本对中的发生情况,统计了3类差异AS事件中任意两组LUAD-癌旁正常样本对中AS事件的交集,称为AS事件重叠率;以及3类差异AS事件中每一个事件在不同样本中出现的频次,称为AS事件的样本重复率。统计方法如下。

(1) AS事件重叠率:对任意一类AS事件,假设有两组LUAD-癌旁正常样本对,分别为 S_1 和 S_2 ,样本对 S_1 有 N_1 个事件,样本对 S_2 有 N_2 个事件,两组样本对事件集的交集有 n 个事件,则在样本对 S_1 中重叠率为 n/N_1 ,在样本对 S_2 中重叠率为 n/N_2 。

(2) AS事件的样本重复率:对任意一类AS事件,假设77个LUAD-癌旁正常样本对中共发生了 M 个事件,其中,第 i 个事件在 m 个样本对中出现,则记该事件的样本重复率为 m 。

本研究使用信息熵来判断AS事件在LUAD样本和癌旁正常样本中剪接率的一致性。定义AS事件的信息熵(H)为 $H_i = -p_1 \log_2 p_1 - p_2 \log_2 p_2$,其中, p_2 为第 i 个AS事件满足差异剪接率小于0的样本占全部样本的比值, $p_1 = 1 - p_2$ 。 H 的取值范围为 $[0, 1]$,当 $H=0$ 时,认为该AS事件满足一致性;当 $H=1$ 时,认为该AS事件完全不一致;当 $0 < H < 1$ 时,认为该AS事件不完全满足一致性。

2 结果分析与讨论

2.1 可变剪接事件剪接率和差异剪接率分布

基于Set2数据集,针对3类AS模式,统计了LUAD样本和癌旁正常样本中剪接率和差异剪接率的分布,如图2所示。结果显示,LUAD样本和癌旁正常样本中绝大多数AS事件的剪接率具有显著的偏向性,事件集中分布在 $[0, 0.1]$ 和 $(0.9, 1.0]$ 区间内,如图2(a)–(c)所示,占总事件的比值约为25%~45%。在3类模式的分析中,剪接率均集中分布于两端,表明差异可变剪接事件对剪接位点的选择偏好明显。通过比较癌症和癌旁正常样本的差异剪接率,发现3类模式AS事件的差异剪接率在 $[-1.0, -0.9)$ 和 $(0.9, 1.0]$ 范围内分布最多,如图2(d)–(f)所示,占比均大于40%,表明癌症和癌旁正常样本具有明确的剪接位点选择性。

此外,分析了Set3集合中的剪接率和差异剪接率分布情况。通过对比发现,AS事件在不同样本对之间存在一定差异,而这种差异性对AS事件剪接率整体分布的影响不大,因此可得出如下推断:第一,AS事件在不同个体之间存在一定的差异,但也有部分AS事件是在多个个体中普遍存在的;第二,剪接位点的选择与是否患癌关系密切,即同一个AS事件在多个样本中剪接位点的选择方式与是否患癌较为一致。

根据Set3集合中的AS事件,对可变剪接位点是否满足GT-AG法则进行了分析,计算了满足规则的AS位点比例,结果如图3所示。在A3SS剪接模式中,99.9%的AS位点为AG;在A5SS剪接模式中,99.4%的AS位点为GT;在SE剪接模式中,100%的3'端AS位点为AG,99.9%的5'端AS位点为GT。3类模式的AS事件中,与AS位点配对供体和受体端的剪接位点为GT/AG的比例基本一致。

2.2 高频AS事件的识别

在Set3集合中任意取77个样本对中的两组样本对进行统计,并分析AS事件重叠率分布情况,结果如图4所示。第一,3类AS事件中的样本重叠率呈右偏分布,且均低于0.25;第二,3类AS事件中,重叠率分布最多的区间均为 $(0.06, 0.07]$;第三,SE事件的重叠率分布最为集中,主要分布在 $(0.05, 0.09]$,A3SS事件的重叠率分布较为集中,A5SS事件的重叠率分布较为分散。根据样本对之间的重叠率可以发现,AS事件的发生在不同的个体中存在一定的差异,但也有部分AS事件是在多个个体中普遍存在的。

综上所述,在77个LUAD癌旁正常样本对中,AS事件发生的差异性与样本的来源不同有一定的关系,因此,想要在理论上判断出可作为LUAD检测的生物标志物,需要将差异AS事件集按其不同样本对中出现频次进行分类讨论。

通过统计AS事件重复率,发现在LUAD中3类差异AS事件大多只在一个样本中出现,只有少

量差异 AS 事件在多个样本中重复出现,3类 AS 事件在 77 对样本中的实际最大重复样本率(m)分别为 A3SS($m_{\max}=31$), A5SS($m_{\max}=35$)和 SE($m_{\max}=55$),结果如图 5 所示。按差异 AS 事件在 77 个 LUAD-癌旁正常样本对中出现的频次将 Set3 集合分为两类:高频 AS 事件和低频 AS 事件,并分析其剪接率分布。在图 5 中可以看出, $m=18$ 是一个在 3 类 AS 事件类型中均较为明显的高重复率的分界点,因此, $m>18$ 的部分认为是高频 AS 事件, $m\leq 18$ 的部分认为是低频 AS 事件,我们认为高频 AS 事件是对 LUAD 有普遍影响的事件。最终,得到的高频 AS 事件包括 6 个 A3SS 事件、11 个 A5SS 事件、64 个 SE 事件。

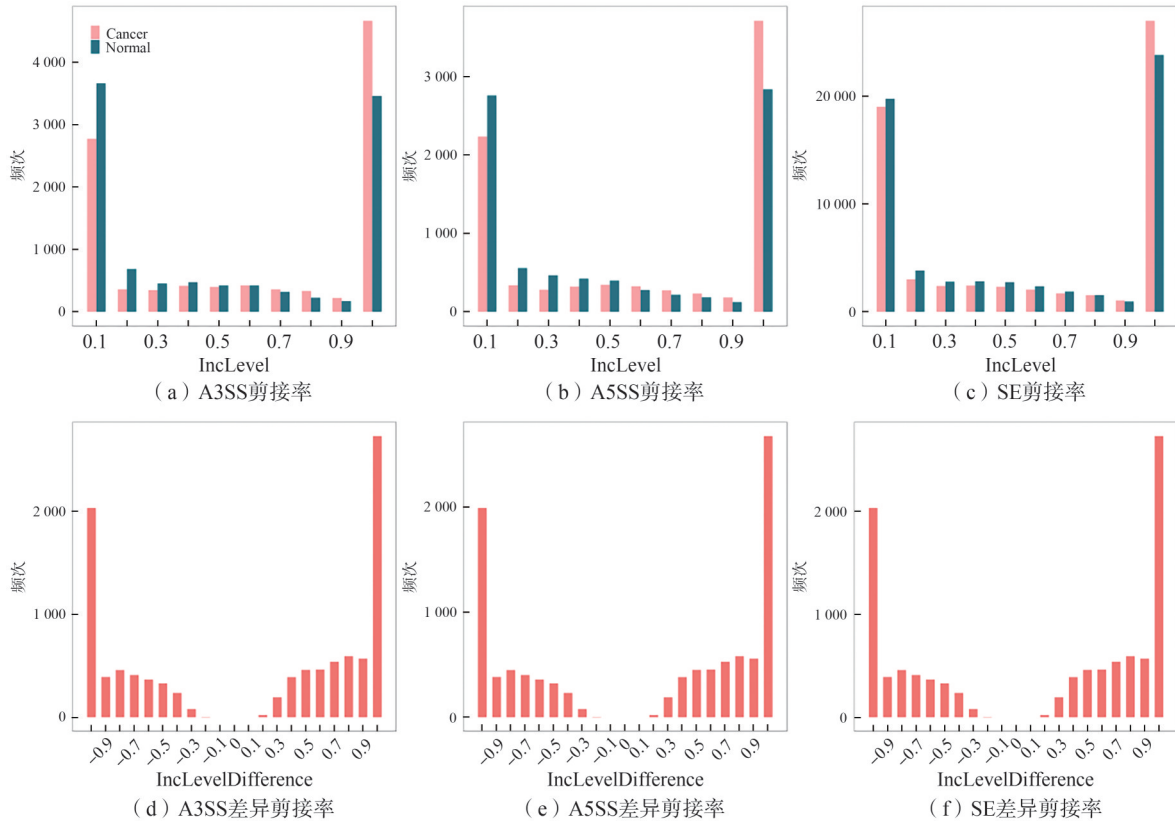


图2 Set2集合中3类AS事件的剪接率和差异剪接率分布

Fig. 2 Distributions of IncLevel and IncLevelDifference values for three kinds of AS events in Set2

对于3类 AS 事件的高频 AS 事件集,分析讨论了 LUAD 中每一个 AS 事件在样本对中剪接率的分布规律,图 6 展示了部分 AS 事件的剪接率分布,共发现有 76 个 AS 事件在 LUAD 与癌旁正常样本中的剪接率有显著差异($P<0.01$),即剪接位点的选择与是否患癌有较为明显的偏好性。IncLevelDifference 小于 0 表明该事件在正常样本中倾向于选择近端剪接模式、包含中间 exon 的剪接模式,IncLevelDifference 大于 0 表明该事件在正常样本中倾向于选择远端剪接模式、不包含中间 exon 的剪接模式;共有 5 个 AS 事件在 LUAD 与癌旁正常样本中剪接率无显著差异,结果如表 2 所示。

2.3 特异性 AS 事件的识别与分析

针对 77 对 LUAD 样本中在多个样本对发生的 AS 事件,进一步分析了 AS 事件剪接率一致性(H)的分布规律,结果如图 7 所示。

选取 3 类 AS 事件中完全满足一致性的高频 AS 事件作为 LUAD 的特异性事件,即认为同时满足 $H=0$ 和 $m>18$ 的 AS 事件为 LUAD 特异的可变剪接事件,最终筛选得到 LUAD 特异性 AS 事件集,结果如表 3 所示。

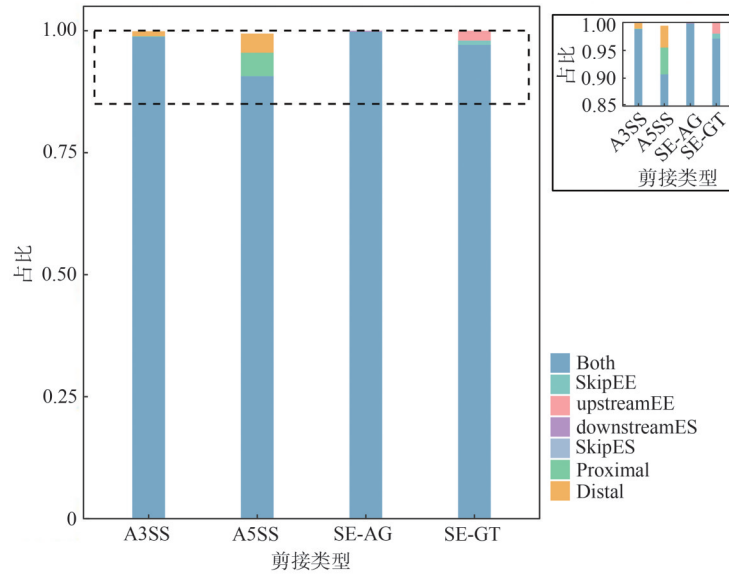


图 3 3类模式 AS 事件中剪接位点满足 GT/AG 的占比

Fig. 3 Percentage of splicing sites satisfied GT/AG rules in three AS patterns

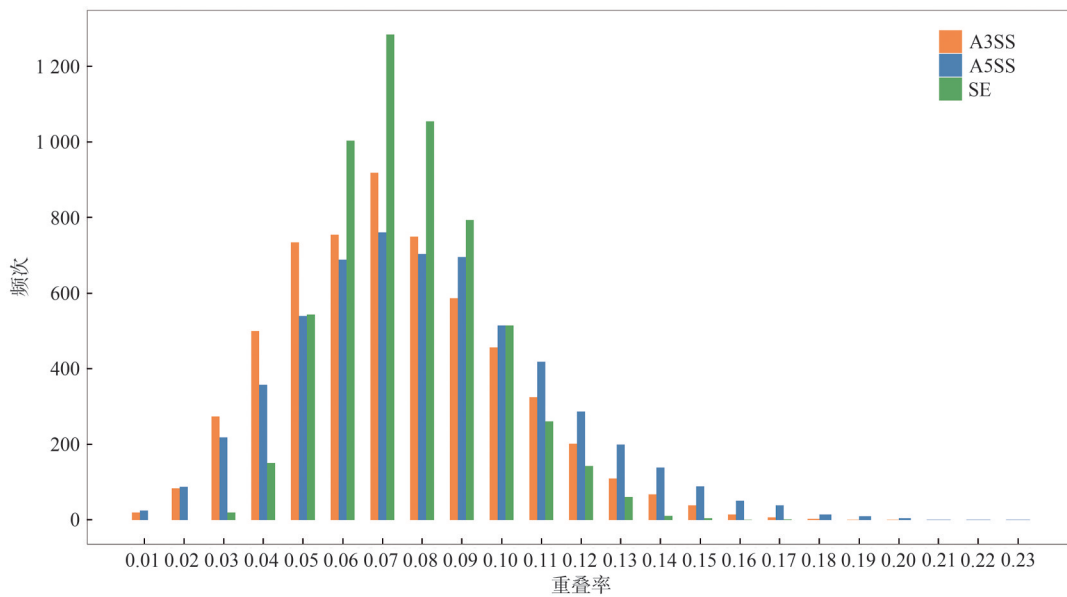


图 4 3类 AS 事件重叠率分布

Fig. 4 Distributions of overlapping ratio in three AS events

2.4 特征基因功能分析

通过比对 AS 事件与基因组注释文件,获取发生了特异性 AS 事件的基因,将这些与 LUAD 特异性 AS 事件对应的基因称为特征基因。由于同一个基因可能发生不同的 AS 事件,因此,51 个特异性 AS 事件共发生在 42 个特征基因。

为了探究这些特征基因的生物学功能,首先利用 GeneCards 数据库(<https://www.genecards.org/>)进行功能注释,发现数据集中有 11 个特征基因与癌症的发生发展密切相关。这 11 个基因分别是 *LRRFIP2*(SE)、*DVL1*(A5SS)、*POSTN*(SE)、*NUMB*(SE)、*CLK1*(SE)、*MUC1*(A3SS)、*ITGB4*(SE)、*CTNND1*(SE)、*VEGFA*(SE)、*MAP3K7*(SE)和 *MYO18A*(SE),其中有 9 个基因涉及 SE 事件,这暗示着它们的可变剪接异构体可能在蛋白质结构和功能层面对基因表达调控产生较大影响。

为进一步系统地揭示这些基因的分子机制,进行了通路富集(GO,KEGG)和蛋白质互作网络(PPI)分析,结果如图8所示,发现其中有14个基因,即NCOR2(A5SS)、DVL1(A5SS)、NUMB(SE)、CTNND1(SE)、LMO7(SE)、MAP3K7(SE)、TPM1(A5SS,SE)、SYNE2(SE)、MYH14(SE)、LIMCH1(SE)、MYO6(SE)、FN1(A3SS,SE)、VEGFA(SE)和SPAG9(SE),显著富集于多个与癌症进展(特别是细胞迁移、侵袭和增殖)相关的生物学通路中,且这些基因在同一个PPI网络中。

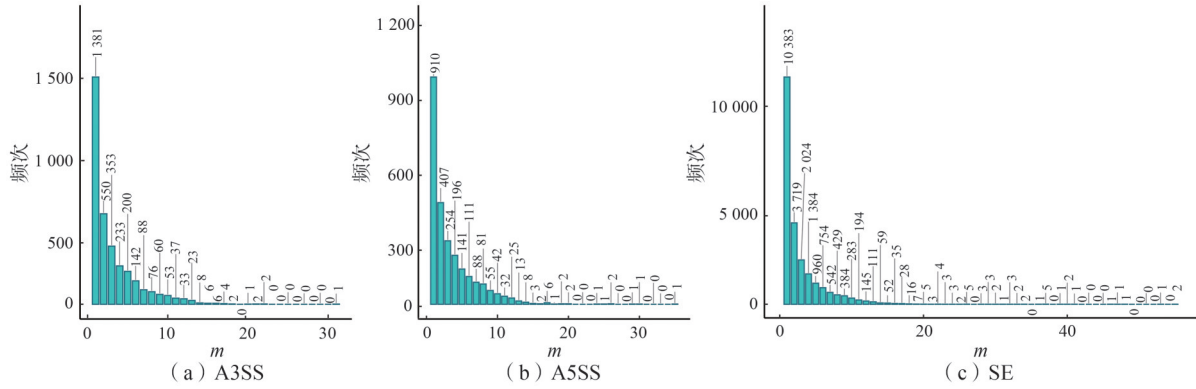


图5 3类AS事件在样本对中的重复频次分布

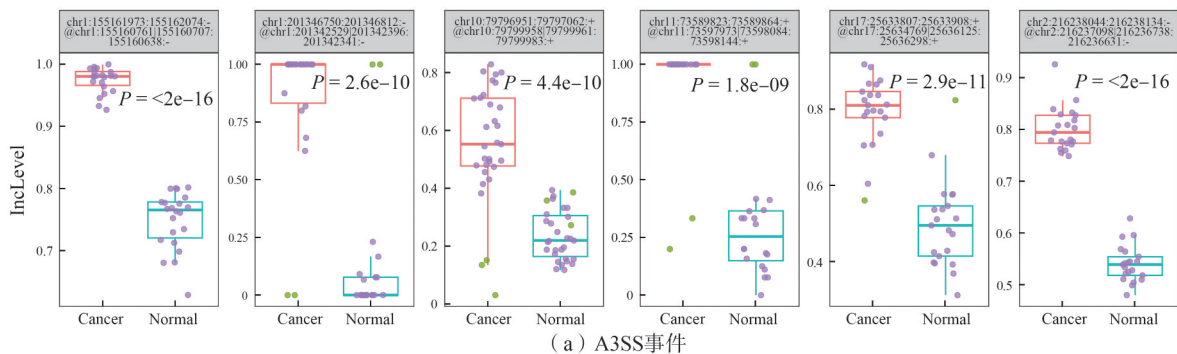
Fig. 5 Number of repetition frequency of three kinds of AS events in sample pairs

表2 高频AS事件剪接率的差异显著性

Table 2 Significance of the differences in the splicing rates of high-frequency AS events

AS类型	AS事件数	$P < 0.01$	InclLevelDifference 大于0	InclLevelDifference 小于0
A3SS	6	6	6	0
A5SS	11	9	7	2
SE	64	61	30	31

首先,在信号通路方面,基因NCOR2、DVL1和NUMB与Notch信号通路相关,该通路在细胞正常形态发生的多个过程中起作用,包括多能祖细胞的分化、细胞的增殖和凋亡以及细胞边界的形成;同时,基因LRRFIP2和DVL1可以激活Wnt信号通路,进而调控癌症的进展。其次,在细胞黏附与细胞骨架方面,基因CTNND1、LMO7和MAP3K7参与上皮细胞间或上皮细胞和细胞外基质的黏着连接;基因TPM1、SYNE2、MYH14、LIMCH1和MYO6与肌动蛋白结合,和肌动蛋白丝结合活性等生物学过程相关,且显著性较高,在分子功能和细胞组分两个方面均排前五,而肌动蛋白可能对癌症的迁移和增殖有一定影响。2018年,李炯等^[24]发现肌动蛋白通过调节分子CAP1对肺癌细胞的迁移和增殖起促进作用,并推测可能与CAP1参与细胞肌动蛋白骨架和细胞周期调控有关。此外,基因集还显著参与调控细胞形态的发生(FN1、DVL1、TPM1、MYH14、VEGFA和SPAG9等, $P = 1.43 \times 10^{-5}$)。



(a) A3SS事件

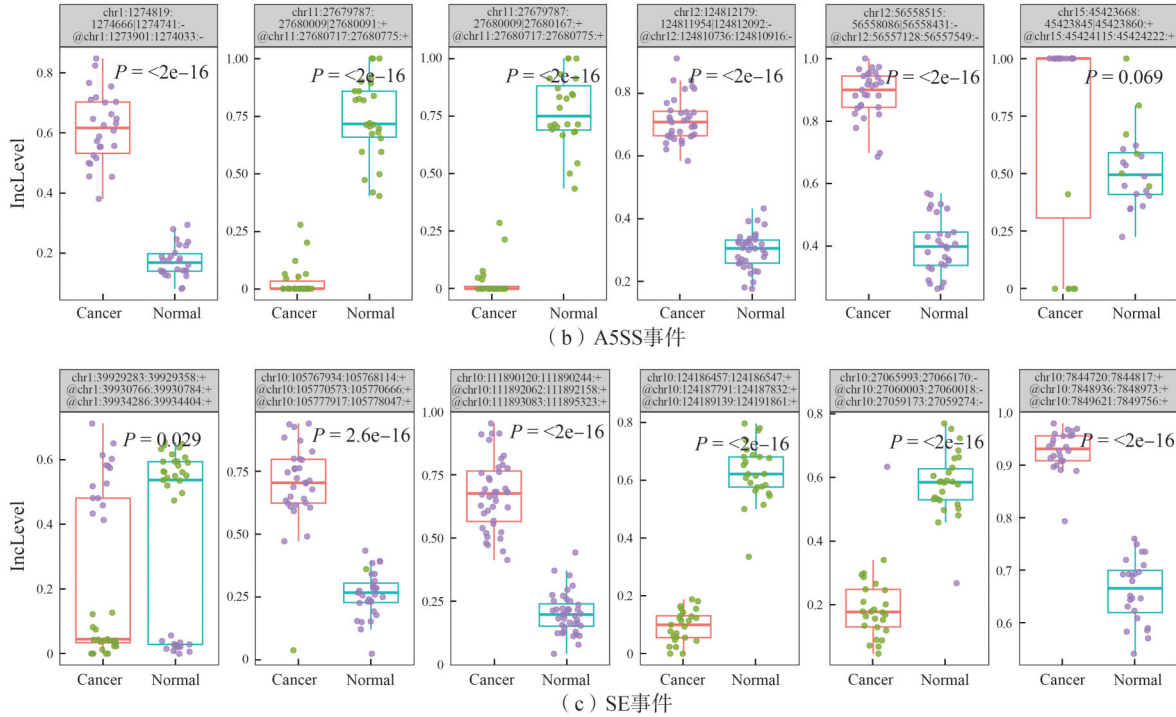


图 6 高频 AS 事件剪接率分布

Fig. 6 Distributions of IncLevel values in high-frequency AS events

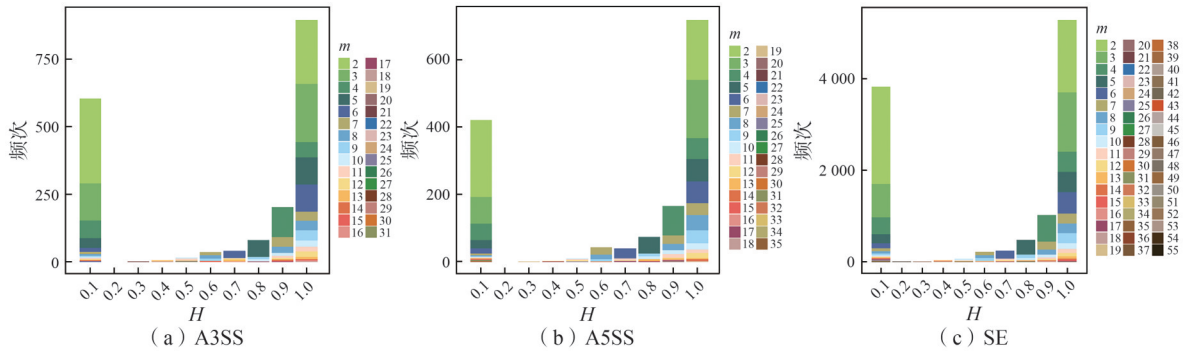


图 7 AS 事件的 H 值分布

Fig. 7 Distributions of H for AS events

表 3 LUAD 特异性 AS 事件集

Table 3 LUAD specific AS event sets

事件类型	Set3	$m \geq 2$	$H=0$	$m > 18$	$H=0 \& m > 18$	IncLevelDifference	
						大于 0	小于 0
A3SS	3 261	1 880	604 (32.1%)	6	2	2	0
A5SS	2 386	1 476	422 (28.6%)	11	6	4	2
SE	21 566	11 183	3 832 (34.3%)	64	43	20	23

注:括号中的数值表示满足 $H=0$ 的 AS 事件数占比。

以上通路富集结果与已报道的单个基因功能高度一致。例如,在 Notch 通路中富集的 *NUMB* 作为抑癌基因对肺癌等多种癌症均具有调控作用,*NUMB* 的可变剪接可以产生 6 种不同的功能异构体,在癌症进程中起重要作用,其外显子 9 的保留可以促进癌细胞增殖,且是肺癌中最常见的可变剪接之一^[20,25]。*POSTN* 被报道参与肺癌的迁移与侵袭^[26]。同时,有研究发现基因 *CLK1* 的编码蛋白可

能在控制剪接位点的选择中起间接作用,数据表明其可变剪接与胃癌密切相关^[27]。基因 *MUC1*、*ITGB4*、*CTNND1*、*MAP3K7* 和 *VEGFA* 等也分别与 TP53 介导的转录、癌症侵袭、细胞凋亡和癌症分期与进展等过程有关,基因 *MYO18A* 可能参与调控巨噬细胞的表达和激活。

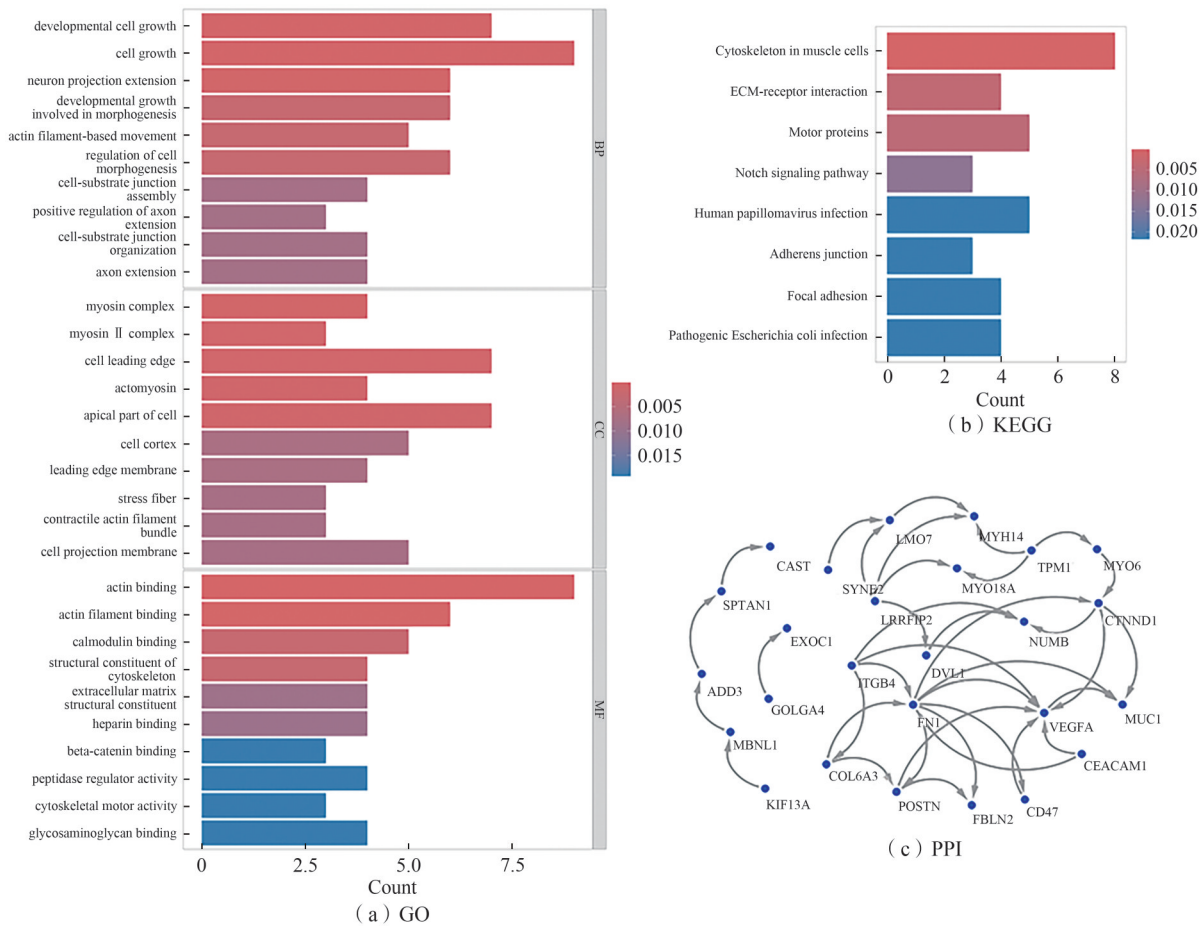


图8 特征基因的GO分析、KEGG分析和PPI

Fig. 8 GO analysis, KEGG analysis and PPI of target genes

3 结论

本研究旨在通过分析LUAD中AS事件剪接率的分布规律,找到与LUAD发生发展密切相关的AS事件与基因,即特异性AS事件与特征基因。识别了77对样本的3类AS事件,首先按照同一AS事件在不同样本中出现的频次(m)筛选出高频AS事件($m > 18$),然后根据剪接率的信息熵(H)筛选出满足剪接率一致性的高频AS事件,将其定义为LUAD的特异性AS事件集,并对LUAD的特异性AS事件和特征基因进行功能分析。最终,通过GeneCards数据库发现了11个基因与癌症的发生发展密切相关,其中 *NUMB* 和 *POSTN* 已有研究证明与肺癌有关,且 *NUMB* 的可变剪接(外显子9的保留)可以促进肺癌细胞增殖;通过通路分析发现了14个基因显著富集于多个与癌症进展(特别是细胞迁移、侵袭和增殖)相关的生物学通路中,包括Notch信号通路、Wnt信号通路和细胞黏附与细胞骨架等。

本研究也存在一定的局限性。第一,本研究所用样本量仅77对,未来需要更大规模的数据来验证这些特异性AS事件和特征基因的普遍性。第二,本研究结论主要基于生物信息学的理论分析,对于筛选出来的特异性AS事件和特征基因作为LUAD潜在生物标志物的可能性,仍迫切需要后续的生物实验进行深入的功能验证和机制探讨。

参考文献:

- [1] TAYLOR M R. Student study guide for biology[M]. 6th ed. San Francisco: Benjamin Cummings, 2002.
- [2] SINGH R K, COOPER T A. Pre-mRNA splicing in disease and therapeutics[J]. Trends in Molecular Medicine, 2012, 18(8): 472-482.
- [3] JIA Z C, YANG X, HOU X X, et al. The importance of a genome-wide association analysis in the study of alternative splicing mutations in plants with a special focus on maize[J]. International Journal of Molecular Sciences, 2022, 23(8): 4201.
- [4] 关华, 高洁, 石璐璐, 等. 人类基因组跳跃外显子组蛋白修饰特征分析[J]. 内蒙古大学学报(自然科学版), 2022, 53(3): 275-282.
- [5] RODRIGUEZ-AGUAYO C, MONROIG P D C, REDIS R S, et al. Regulation of hnRNPA1 by microRNAs controls the miR-18a-K-RAS axis in chemotherapy-resistant ovarian cancer[J]. Cell Discovery, 2017, 3(1): 17029.
- [6] YIN H J, WANG J S, LI H, et al. Extracellular matrix protein-1 secretory isoform promotes ovarian cancer through increasing alternative mRNA splicing and stemness[J]. Nature Communications, 2021, 12(1): 4230.
- [7] PELLARIN I, DALL'ACQUA A, GAMBELLI A, et al. Splicing factor proline- and glutamine-rich (SFPQ) protein regulates platinum response in ovarian cancer-modulating SRSF2 activity[J]. Oncogene, 2020, 39(22): 4390-4403.
- [8] HAN C H, ZHAO R, KROGER J, et al. Caspase-2 short isoform interacts with membrane-associated cytoskeleton proteins to inhibit apoptosis[J]. PLoS One, 2013, 8(7): e67033.
- [9] ANUFRIEVA K S, SHENDER V O, ARAPIDI G P, et al. Therapy-induced stress response is associated with downregulation of pre-mRNA splicing in cancer cells[J]. Genome Medicine, 2018, 10(1): 49.
- [10] GAMBINO G, TANCREDI M, FALASCHI E, et al. Characterization of three alternative transcripts of the *BRCA1* gene in patients with breast cancer and a family history of breast and/or ovarian cancer who tested negative for pathogenic mutations[J]. International Journal of Molecular Medicine, 2015, 35(4): 950-956.
- [11] XIA C F, DONG X S, LI H, et al. Cancer statistics in China and United States, 2022: Profiles, trends, and determinants[J]. Chinese Medical Journal, 2022, 135(5): 584-590.
- [12] REEVE J G, XIONG J, MORGAN J, et al. Expression of apoptosis-regulatory genes in lung tumour cell lines: Relationship to *p53* expression and relevance to acquired drug resistance[J]. British Journal of Cancer, 1996, 73(10): 1193-1200.
- [13] YIN X M, OLTVAI Z N, KORSMEYER S J. BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax[J]. Nature, 1994, 369(6478): 321-323.
- [14] LINDENBOIM L, BORNER C, STEIN R. Bcl-x_s can form homodimers and heterodimers and its apoptotic activity requires localization of Bcl-x_s to the mitochondria and its BH3 and loop domains[J]. Cell Death & Differentiation, 2001, 8(9): 933-942.
- [15] ZHANG L, LI S Y, CHOI Y L, et al. Systematic identification of cancer-related long noncoding RNAs and aberrant alternative splicing of quintuple-negative lung adenocarcinoma through RNA-seq[J]. Lung Cancer, 2017, 109: 21-27.
- [16] GORGOULIS V G, ZOUMPOURLIS V, RASSIDAKIS G Z, et al. A molecular and immunohistochemical study of the MDM2 protein isoforms and *p53* gene product in bronchogenic carcinoma[J]. The Journal of Pathology, 1996, 180(2): 129-137.
- [17] MOLL U M, PETRENKO O. The MDM2-p53 interaction[J]. Molecular Cancer Research, 2003, 1(14): 1001-1008.
- [18] LIN S Y, XU H L, QIN L, et al. UHRF1/DNMT1-MZF1 axis-modulated intragenic site-specific CpGI methylation confers divergent expression and opposing functions of PRSS3 isoforms in lung cancer[J]. Acta Pharmaceutica Sinica B, 2023, 13(5): 2086-2106.
- [19] REYES A, AFKHAMI M, MASSARELLI E, et al. RBM10 mutation as a potential negative prognostic/predictive biomarker to therapy in non-small-cell lung cancer[J]. Clinical Lung Cancer, 2024, 25(8): e411-e419.

- [20] BECHARA E G, SEBESTYÉN E, BERNARDIS I, et al. RBM5, 6, and 10 differentially regulate *NUMB* alternative splicing to control cancer cell proliferation[J]. *Molecular Cell*, 2013, 52(5):720-733.
- [21] MISQUITTA-ALI C M, CHENG E, O'HANLON D, et al. Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer[J]. *Molecular and Cellular Biology*, 2011, 31(1):138-150.
- [22] YAN Y R, WANG N, XING B W, et al. Splicing shift of *RAC1* accelerates tumorigenesis and defines a potent therapeutic target in lung cancer[J]. *Advanced Science*, 2025, 12(33):e03322.
- [23] SEO J S, JU Y S, LEE W C, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma[J]. *Genome Research*, 2012, 22(11):2109-2119.
- [24] 李炯, 周颢, 白平, 等. 肌动蛋白调节分子 CAP1 对肺癌细胞迁移和增殖的影响[J]. *第三军医大学学报*, 2018, 40(16):1500-1506.
- [25] CHOI H Y, SEOK J, KANG G H, et al. The role of NUMB/NUMB isoforms in cancer stem cells[J]. *BMB Reports*, 2021, 54(7):335-343.
- [26] RATAJCZAK-WIELGOMAS K, KMIECIK A, DZIEGIEL P. Role of periostin expression in non-small cell lung cancer; Periostin silencing inhibits the migration and invasion of lung cancer cells via regulation of *MMP-2* expression[J]. *International Journal of Molecular Sciences*, 2022, 23(3):1240.
- [27] BABU N, PINTO S M, BISWAS M, et al. Phosphoproteomic analysis identifies CLK1 as a novel therapeutic target in gastric cancer[J]. *Gastric Cancer*, 2020, 23(5):796-810.

(责任编辑 刘俊杰)

Analysis and Identification of Specific Alternative Splicing Events in Human Lung Adenocarcinoma

GAO Jie¹, GAO Yuanyuan², ZHANG Lirong¹

(1. *School of Physical Science and Technology, Inner Mongolia University, Hohhot*
010021, China;

2. *Inner Mongolia Traditional Chinese & Mongolian Medical Research Institute, Hohhot*
010020, China)

Abstract: Abnormal splicing of pre-mRNA can lead to cancer, and previous studies have confirmed that alternative splicing (AS) plays a crucial role in the proliferation and migration of human lung adenocarcinoma (LUAD). Based on LUAD-adjacent normal sample RNA-Seq data, three common patterns AS event sets were constructed, and three types of high-frequency AS events in LUAD and adjacent normal samples were analyzed and identified. 51 events with good consistency ($H=0$) and high repetition rate ($m>18$) were selected as LUAD-specific AS events, and the functional analysis of these AS events and their related characteristic genes was further conducted. Finally, through functional inquiries and pathway analyses respectively, it was found that 11 characteristic genes were closely related to the occurrence and development of cancer, and 14 characteristic genes may participate in the migration and proliferation of cancer in their molecular functions. It has been proved that *NUMB* and *POSTN* genes are related to lung cancer. These AS events and genes may be suitable as potential diagnostic biomarkers for LUAD.

Key words: alternative splicing; lung adenocarcinoma; information entropy