DOI:10.16016/j.2097-0927.202112095

正常肺上皮细胞致瘤性转化中抑癌基因的全基因组筛选及数据库验证

赵斌瑜[1],张俊磊[2],王加琪[2],阮 艳[2],张 月[4],刘恋恋[2],寨 锐[2],熊加祥[3],姚忠祥[1] 400038 重庆,陆军军医大学(第三军医大学)基础医学院:生理学教研室[1],组织学与胚胎学教研室[2],教学实验中心[3];400038 重庆,陆军军医大学(第三军医大学)第一附属医院眼科[4]

[摘要] **目的** 利用 *CRISPR/Cas*9 敲除文库在人正常肺上皮细胞致瘤性转化过程中进行全基因组筛选,以发现新的人肺癌抑癌基因,并利用临床非小细胞肺癌(non-small cell lung cancer, NSCLC)样本数据库进行初步验证。**方法** 构建稳定表达 Cas9 蛋白的人正常肺上皮 BEAS-2B 细胞系,在该细胞中感染含有 77 441 个单链向导 RNA(sgRNA)、靶向 19 114 个人类基因的 Brunello 慢病毒文库,于裸鼠皮下移植。提取全瘤体细胞(Tumor)及移植前细胞(Input)的全基因组 DNA,扩增 sgRNA 序列,建库进行深度测序。通过分析 Tumor 中 sgRNA 读数占比及 Tumor 与 Input 中 sgRNA 读数差异倍数的改变,筛选功能缺失后诱导 BEAS-2B 细胞致瘤性转化的抑癌基因,并利用公共数据平台进行初步验证。**结果** 成功在稳定表达 Cas9 蛋白的 BEAS-2B 细胞中进行了文库感染及裸鼠皮下成瘤实验。只有感染文库的细胞在部分接种位点成瘤。测序结果显示 Tumor 中只检测到少数 sgRNA,选取 sgRNA 的读数占总读数的比例>1%或有 2 个及以上 sgRNA 被显著富集的 38 个基因为有效的人肺癌候选抑癌基因,包括 *NF*2、*PTEN* 等已知抑癌基因及未在肺癌中报道过的候选抑癌基因,如 *AP*2*M*1、*PSENEN* 等。富集分析发现候选抑癌基因在 Notch 和 Hippo 信号等参与癌症发生的关键生物学通路中富集。应用临床 NSCLC 样本数据库证实 38 个候选抑癌基因在 NSCLC 样本中均发生了突变,其中突变率>2%的 17 个基因中有 14 个与 *KRAS* 或 *TP*53 发生共突变,3 个突变与预后相关;30 个基因在 NSCLC 样本中的表达水平与正常癌旁组织有明显差异,17 个基因的表达水平与患者的预后密切相关。**结论** 利用体内全基因组 CRISPR/Cas9 技术筛选出 38 个人肺癌候选抑癌基因,并在临床 NSCLC 样本数据库中得到了初步验证。

[关键词] 致瘤性转化;CRISPR/Cas9;抑癌基因;全基因组筛选;肺癌
[中图法分类号] R394-33;R394.3;R734.2 [文献标志码] A

Genome-wide screening of tumor suppressor genes in oncogenic transformation of normal lung epithelial cells and preliminary validation in sample databases

ZHAO Binyu[1], ZHANG Junlei[2], WANG Jiaqi[2], RUAN Yan[2], ZHANG Yue[4], LIU Lianlian[2], JIAN Rui[2], XIONG Jiaxiang[3], YAO Zhongxiang[1] [1]Department of Physiology, [2]Department of Histology and Embryology, [3]Teaching Experiment Center, College of Basic Medical Sciences, Army Medical University (Third Military Medical University), Chongqing, 400038; [4]Department of Ophthalmology, First Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing, 400038, China

[**Abstract**] **Objective** To identify new human lung cancer suppressors in a genome-wide screening library during neoplastic transformation of normal human lung epithelial cells using CRISPR/Cas9-mediated genome editing, and preliminarily validate the results in non-small cell lung cancer (NSCLC) sample databases. **Methods** Human normal lung epithelial BEAS-2B cell line with stable expression of Cas9 protein was constructed, and then infected with the Brunello lentiviral library encompassing 77 441 single-guide RNAs (sgRNAs) targeting 19 114 genes. Then the obtained cells were transplanted subcutaneously into nude mice. Genomic DNA of tumor cells (Tumor) and pre-transplantation cells (Input) were extracted. The extracted

[基金项目] 国家自然科学基金面上项目(32071014)
[通信作者] 姚忠祥,E-mail: yaozhx@yahoo.com

2097-0927/Copyright © 2022, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循 CC BY 许可协议(https://creativecommons.org/licenses/by/4.0/)

DNA and plasmid library were amplified for deep sequencing. By analyzing the proportion of sgRNA reads in Tumor cells compared to the Input cells and the Log Fold Change (LFC) of sgRNA reads, tumor suppressor genes that can induce oncogenic transformation of BEAS-2B cells after knockout were screened, and preliminary validation were performed using public data platforms. **Results** Library infections in BEAS-2B cells stably expressing Cas9 protein and subcutaneous tumorigenesis in nude mice were successfully established. Only cells infected with the library formed tumors at the inoculation sites. Sequencing results showed that only a small part of sgRNAs were detected in tumor cells. Here, 38 genes with sgRNA reads accounting for more than 1% of the total reads or with 2 or more significantly enriched sgRNA were selected as effective candidate human lung tumor suppressor genes, including known tumor suppressor genes such as NF2 and PTEN *etc*, and some candidate tumor suppressor genes (AP2M1 and PSENEN) had not been previously reported in lung cancer. Enrichment analysis revealed that these candidate tumor suppressor genes were enriched in key biological pathways involved in cancer, such as Notch and Hippo signaling pathways. It is confirmed in the clinical NSCLC sample databases that 38 candidate tumor suppressor genes were mutated in NSCLC samples. Among the 17 genes with mutation rates greater than 2%, 14 genes were co-mutated with KRAS or TP53, and 3 ones were associated with the prognosis of NSCLC patients. There were 30 candidate genes with their expression in the samples of NSCLC different from those in para-tumor normal tissues, and the expression of 17 candidate genes were correlated with the prognosis of NSCLC patients. **Conclusion** By using an *in vivo* genome-wide CRISPR/Cas9 screening library, 38 human lung cancer candidate suppressor genes are identified and preliminarily verified in the clinical NSCLC sample databases.

[**Key words**] neoplastic transformation; CRISPR/Cas9; tumor suppressor genes; genome-wide screening; lung cancer

Supported by the General Program of National Natural Science Foundation of China (32071014). Corresponding author: YAO Zhongxiang, E-mail: yaozhx@yahoo.com

根据2020年全球癌症统计数据,肺癌在恶性肿瘤中发病率居第二,病死率居第一[1]。肺癌也是我国最常见且病死率最高的恶性肿瘤[2],通常只有16%的患者能于早期诊断,大多数患者诊断即为晚期。因此,肺癌患者的5年生存率很低[3],亟需提高其早期诊断及治疗效果。细胞转化是癌症发生的第一步,癌症的发生是癌基因和抑癌基因异常表达导致正常细胞致瘤性转化的结果。因此,正常肺上皮细胞致瘤性转化机制的研究对肺癌早期诊断和干预治疗至关重要。

已知癌基因和抑癌基因遗传改变的积累促使正常细胞向恶性肿瘤细胞转化[4]。基因组测序发现了许多潜在的癌基因和抑癌基因[5-6],但癌细胞中大部分突变为"乘客"突变(passenger mutations),对癌细胞的生物学机制并没有明显影响。只有极少部分基因的突变可直接刺激癌细胞的生长和其他生物学行为,此即"司机"突变(driver mutations)[7]。目前,功能基因组筛选技术可有效地从众多突变中识别出驱动肿瘤形成及发展的"司机"突变,常用的主要方法为RNA干扰技术,但近年来逐步被特异性更高、脱靶效率更低的CRISPR/Cas9技术取代。已有研究利用体内全基因组CRISPR/Cas9技术,筛选出功能缺失促进肿瘤生长及转移的抑癌基因[8-9]。例如,文献报道利用CRISPR全基因组敲除文库,对具有致癌作用的*Kras*基因伴纯合子*p*53和杂合子*Dicer*1功能丧失的鼠非小细胞肺癌(non-small cell lung cancer, NSCLC)细胞系KPD($Kras^{G12D/+}$;$p53^{-/-}$;$Dicer1^{+/-}$)在裸鼠皮下肿瘤形成及转移过程中的抑癌基因进行了筛选[8];还有研究在纯合子*p*53功能丧失及具有致癌作用的*Myc*基因的小鼠胚胎肝祖细胞中,对肿瘤生长过程中的肝癌抑癌基因进行了全基因组筛选[9]。以上两种细胞系均可在裸鼠皮下成瘤,具有致瘤性,但均缺乏对正常细胞致瘤性转化过程的研究。已知癌症发生和发展是两个不同的过程,其机制有巨大的差异,而正常细胞的致瘤性转化是癌症发生的起始,因此,有必要对正常细胞致瘤性转化的过程与机制进行深入研究。

为了在人正常肺上皮细胞致瘤性转化过程中筛选功能缺失的抑癌基因,本文利用CRISPR/Cas9全基因组敲除文库技术,在人正常肺上皮BEAS-2B细胞[10]的致瘤性转化过程中对功能缺失的抑癌基因进行筛选,筛选出38个有效的人肺癌候选抑癌基因,其中包含一些之前未见报道的人肺癌抑癌基因。在此基础上,进一步应用临床NSCLC样本数据库对候选抑癌基因的突变和表达水平进行初步验证,并分析其与患者预后的关系,为探讨人正常肺上皮细胞致瘤性转化机

制及肺癌早期诊治提供新的线索。

1 材料与方法

1.1 材料

1.1.1 实验动物 6~8周龄雌性BALB/c裸鼠16只购自江苏集萃药康生物科技有限公司。动物实验获得陆军军医大学实验动物福利伦理审查委员会批准(AMUWEC2020549)。

1.1.2 实验材料 细胞:BEAS-2B细胞系[美国模式菌种收集中心(ATCC)];质粒:lenti-Cas9-Blast(Addgene:#52962),lenti-Guide(Addgene:#52963),Brunello文库(Addgene:#73179)[11];DMEM高糖培养基(Hyclone);胎牛血清、胰蛋白酶、杀稻瘟菌素、嘌呤霉素(Gibco);聚凝胺(Sigma);Blood & Cell Culture Mini Kits试剂盒(QIAGEN)。

1.2 方法

1.2.1 体外BEAS-2B细胞培养 含有10%胎牛血清的DMEM培养基于37 ℃的恒温细胞培养箱中培养。当细胞密度达到80%时使用胰酶消化并传代,每2~3天换液。

1.2.2 建立表达Cas9酶的细胞 利用慢病毒包装,磷酸钙沉淀法转染质粒lenti-Cas9-Blast,得到浓缩表达Cas9酶的病毒液(含杀稻瘟菌素抗性),将病毒液与感染组BEAS-2B细胞混匀感染,未感染组BEAS-2B细胞不加病毒液作为对照。24 h后,两组均加入含有5 μg/mL杀稻瘟菌素的10%胎牛血清的DMEM培养基,挑选单克隆,WB验证单克隆细胞中高表达Cas9蛋白的细胞,将该细胞命名为BEAS-2B-CAS9细胞用于后续实验。

1.2.3 文库感染及筛选 用含有Brunello文库的慢病毒载体转染BEAS-2B-CAS9细胞,MOI=0.5。用嘌呤霉素(2 μg/mL)筛选细胞2 d。每2天传代1次,每次取4×10^7在筛选后2 d消化收集细胞,移植于裸鼠;另取4×10^7个移植前细胞提取基因组DNA作为测序对照(Input)。

1.2.4 小鼠皮下成瘤实验 分别收集感染与未感染文库的BEAS-2B-Cas9细胞,用PBS+5%胎牛血清重悬,使浓度为1×10^8个细胞/mL。取400 μL细胞悬液与400 μL matrigel混匀,以100 μL(5×10^6个细胞)/点的量移植于裸鼠两侧髂窝皮下,感染与未感染文库的BEAS-2B-Cas9细胞各自分别移植4只裸鼠共8个点。每3~4天观察1次肿瘤生长情况,25 d左右处死小鼠,解剖肿瘤,测量瘤体体积。肿瘤体积计算公式:体积=0.5×长径×短径^2。

1.2.5 全基因组DNA提取、扩增、测序及统计学分析 采用Blood & Cell Culture Mini Kits试剂盒提取移植前细胞(Input)和瘤体细胞(Tumor)的基因组DNA(genetic DNA,gDNA),提取的DNA和质粒文库(Plasmid)按参考文献方法扩增[12],简述如下:对每个DNA样本,使用KOD FX DNA聚合酶(TOYOBO)对7 μg gDNA进行30次100 μL的分离反应,然后将扩增得到的扩增产物进行合并。扩增sgRNA的引物序列为CRISPR seqF:5′-TTGTGGAAAGGACGAAACACCG-3′和CRISPR seqR:5′-TCTACTATTCTTTCCCCTGCACTGT-3′。PCR产物使用凝胶提取试剂盒(Omega)进行纯化。样本由上海凌恩生物科技有限公司提供的NextSeq机器(Illumina)进行测序。采用MAGeCK软件[58]得到LFC(Log Fold Change)值,LFC指的是Tumor相对Input在正筛选检测中sgRNA改变倍数的log2为底的值。以FDR<0.05为差异具有统计学意义。

1.2.6 富集分析 Metascape[13](http://metascape.org)是一个基因注释和分析工具。本研究利用Metascape在线分析网站,分析了有效的人肺癌候选抑癌基因在基因本体(Gene Ontology,GO):生物过程(biological process,BP)与京都基因和基因组百科全书(Kyoto Encyclopedia of Genes and Genomes,KEGG)中的富集情况。以$P<0.01$,最小计数为3,富集因子>1.5为临界值。

1.2.7 候选抑癌基因在临床NSCLC样本数据库中的突变分析及临床相关性分析 cBioPortal[14](http://cbioportal.org)是基于TCGA数据库开发的一款集数据挖掘、数据整合及可视化等多功能于一体的综合性开放网络平台,涵盖了28 000例临床样本的数据。本研究利用cBioPortal在线分析网站,分析了人肺癌候选抑癌基因在7 338例临床NSCLC样本中的突变率,以及突变率与患者预后的相关性(采用Logrank检验,以$P<0.05$为差异具有统计学意义)。已知基因的共突变诱导肿瘤的发生[15-17],所以本研究进一步检测了这些候选抑癌基因与临床NSCLC样本中高突变率基因如*EGFR*、*KRAS*、*TP*53及*STK*11等基因[18-19]的突变相关性(采用Fisher精确概率检验,以$q<0.05$为差异具有统计学意义;log2比值比>0为共突变)。

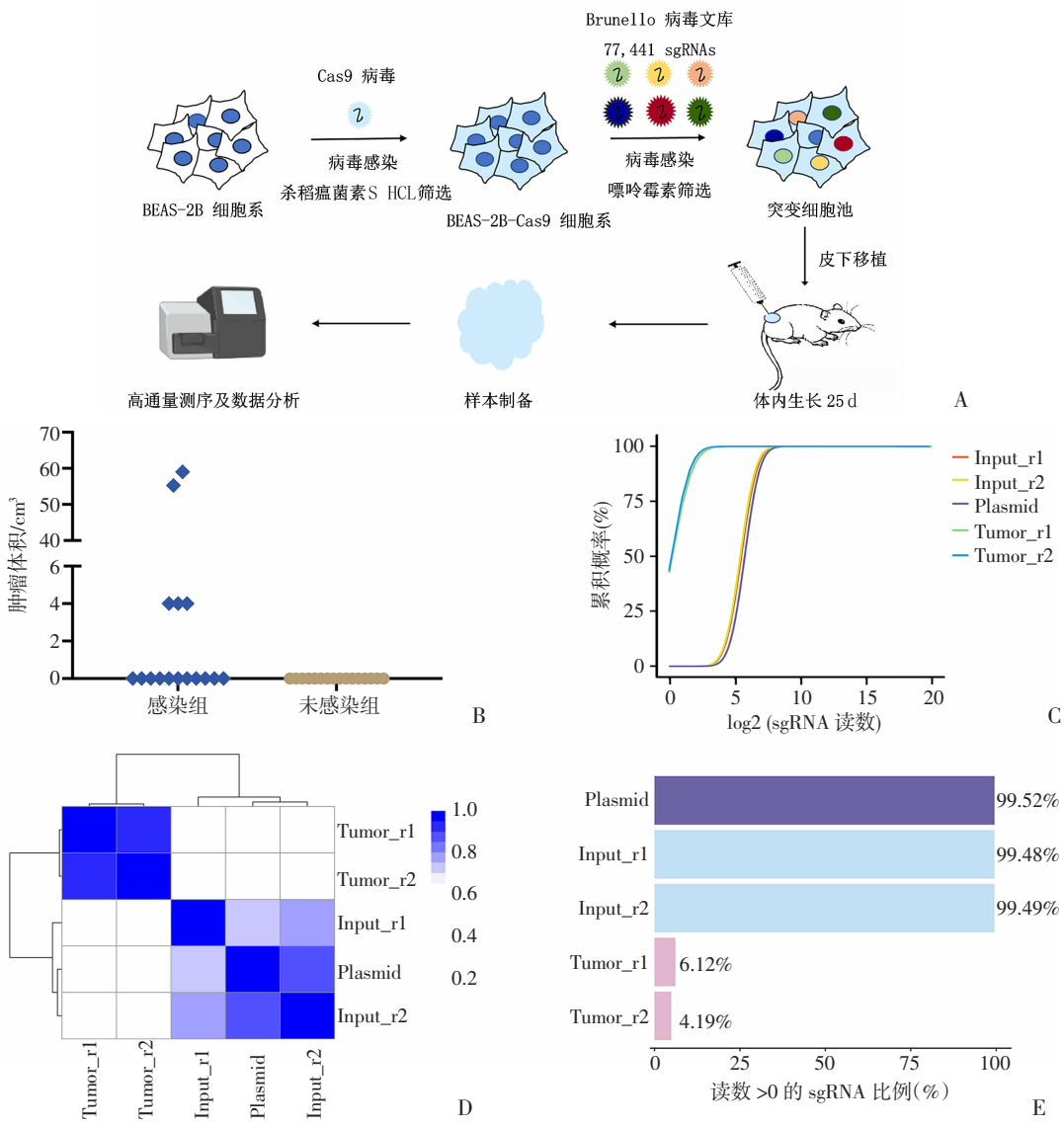
1.2.8 候选抑癌基因在临床NSCLC样本数据库中的mRNA表达水平及临床相关性分析 肺癌探索者(lung Cancer Explorer,LCE)[20]数据库(http://lce.biohpc.swmed.edu/lungcancer/)包括56项研究、超过6 700名患者的表达数据和临床资料数据。利用LCE对人肺癌候选抑癌基因在临床NSCLC样本和癌旁正常肺组织中的mRNA表达水平及与其与患者预后的相关性进行Meta分析。

2 结果

2.1 抑癌基因的全基因组筛选结果及质量控制

为了在人正常肺上皮细胞致瘤性转化过程中筛选功能缺失的肺癌抑癌基因,对永生化的人正常肺上皮BEAS-2B细胞进行了全基因组CRISPR/Cas9基因敲除筛选。第一步是构建稳定表达Cas9的BEAS-2B细胞系(BEAS-2B-Cas9),然后将含有77 441个sgRNA的基因敲除文库导入BEAS-2B-CAS9细胞系。体外培养4 d后,将未感染文库的细胞和感染文库的细胞均以5×10^6/点移植于裸鼠两侧髂窝皮下,25 d后收集瘤体细胞用于后续测序等(图1A)。

分别取质粒(Plasmid)、Input和Tumor的基因组DNA样本扩增并进行高通量测序,检测其sgRNA的分布情况。上述实验步骤从文库感染开始各重复1次。两次重复实验结果显示未感染文库的细胞不能成瘤,而感染文库的细胞有5个移植点成瘤并检测了瘤体大小(图1B);质粒文库和Input样本中包含99.5%的sgRNA类型,读数小于100次的sgRNA比例小于10%,表明文库覆盖度和DNA测序深度均符合标准(图1C)。两次Input样本与质粒文库高度相关($r=0.81$,$n=2$,图1D);Tumor样本中检测到的sgRNA类型仅占总sgRNA的4.91%~6.12%(图1E),但两次sgRNA的分布总体相关性非常强($r=0.93$),具有较好重复性(图1D)。综上所述,文库质检结果符合标准,实验可重复性较高,测序数据有效,可用于后续分析。



A:筛选流程图;B:感染文库及未感染文库的BEAS-2B-CAS9细胞($n=8$)裸鼠皮下成瘤情况;C:质粒(Plasmid)、移植前细胞(Input)和瘤体细胞(Tumor)中的sgRNA的累积概率分布;D:Plasmid、Input和Tumor中sgRNA分布情况的相关性分析;E:Plasmid、Input和Tumor中sgRNA读数>0的sgRNA占比

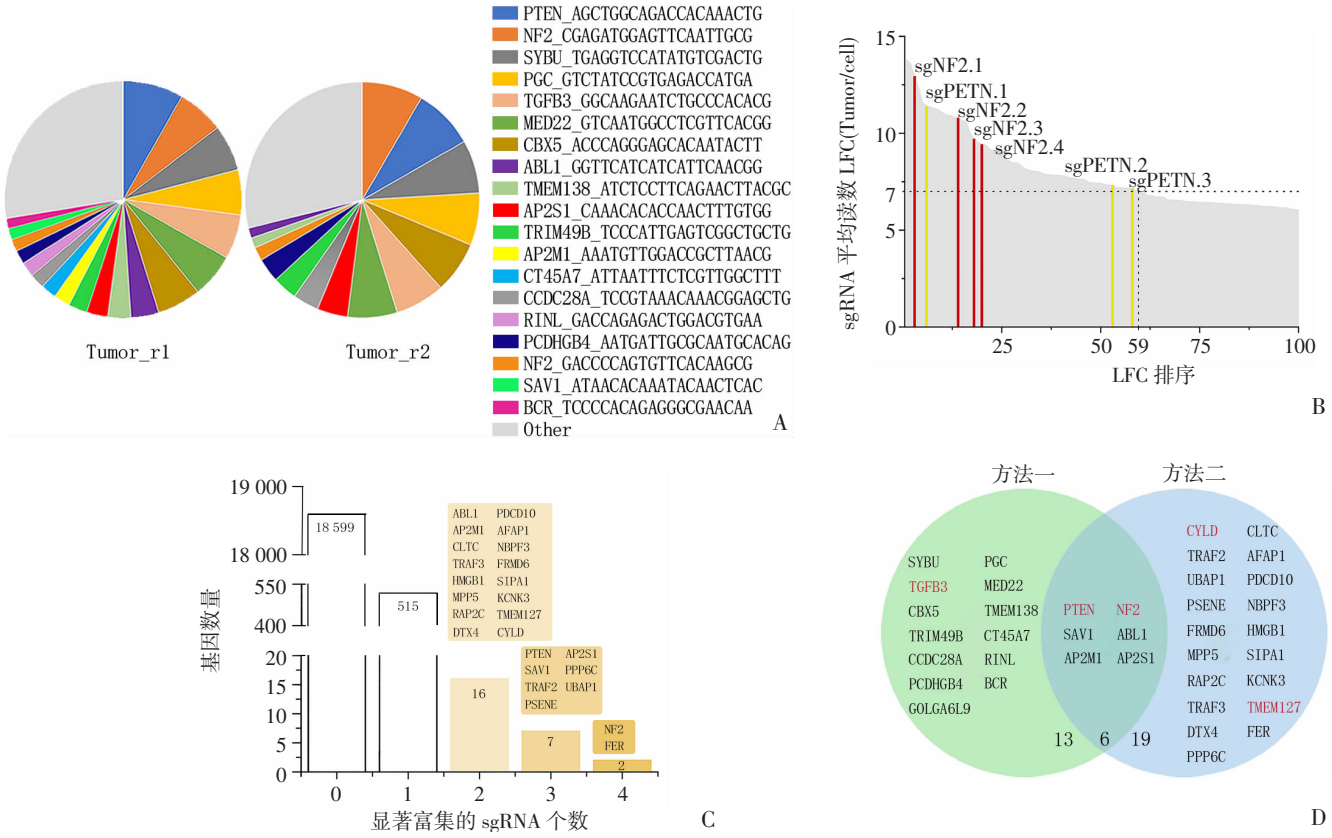
图1 抑癌基因的全基因组筛选结果及质量控制

2.2　人肺癌抑癌基因的筛选

采用 MAGeCK 软件[58] 对测序数据进行分析,结果显示,Tumor 样本中仅检测到少量的 sgRNA 类型(Tumor_r1 中 4 741 个,占 6.12%;Tumor_r2 中 3 799 个,占 4.90%)。其中,极少量 sgRNA 的读数占总读数的比例大于 1%(在 Tumor_r1 中共 19 个,Tumor_r2 中共 14 个,图 2A),而 93.88%~95.09% 的 sgRNA 并没有被检测到。因此,接下来同时使用两种方法筛选 Tumor 样本中显著富集的基因。方法一:sgRNA 靶向的读数比例超过 1% 的基因;方法二:单个基因有 2 个及以上显著富集的 sgRNA(sgRNA LFC>1 且 FDR<0.05,图 2B、C),分别得到 19 个和 25 个候选抑癌基因。对两组基因的比较发现,两组的共有基因是 6 个(图 2D),包括已知的抑癌基因 *NF2*、*PTEN*(靶向 *NF2* 的 4 条,sgRNA 及靶向 *PTEN* 的 3 条,LFC>7)(图 2B);特有的基因集分别包含 13 个和 19 个,其中也含有已知抑癌基因 *TGFB3*、*CYLD* 和 *TMEM127*,提示两种方法都有可能筛选到潜在的抑癌基因(图 2D)。因此,本研究将二者合并,共获得 38 个有效的人肺癌候选抑癌基因(图 2D)。

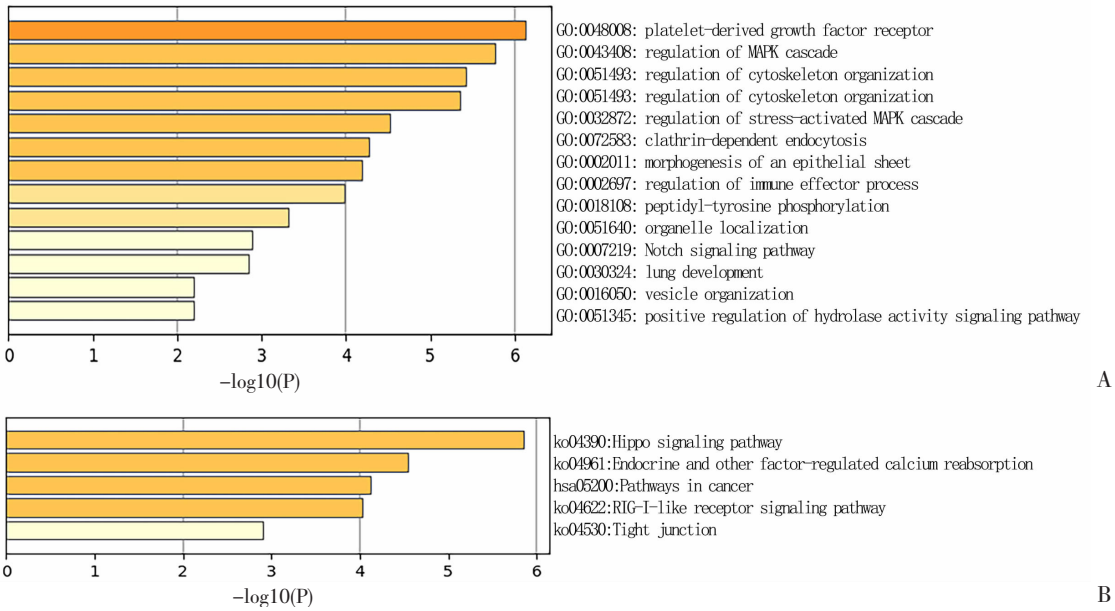
2.3　人肺癌候选抑癌基因的功能学特性分析

利用 METASCAPE 对 38 个人肺癌候选抑癌基因进行 GO-BP 和 KEGG 功能富集分析(图 3),结果提示这些基因参与了多种细胞转化过程中的关键生物学过程及分子信号通路,包括与细胞增殖、分化和凋亡相关的 Notch、MAPK、p38 MAPK 和 Hippo 信号通路;参与了肿瘤侵袭和转移的网格蛋白依赖的内吞(clathrin-dependent endocytosis)、囊泡组织(vesicle organization)和细胞骨架结构调控(regulation of cytoskeleton organization)等生物学过程;参与了免疫逃逸密切相关的免疫效应过程调节(regulation of immune effector process)和 RIG-I 样受体信号通路(RIG-I-like receptor signaling pathway);参与了血管生成诱导的 PDGFR 信号通路(platelet-derived growth factor receptor signaling pathway)等。此外,还与钙吸收调节相关的生物学过程密切相关,提示其也可能参与肿瘤的发生过程。



A:瘤体细胞(Tumor)中 sgRNA 读数占比　每个 sgRNA 的面积对应于 sgRNA 的读数占总读数的比例,比例>1% 的被标出;B:sgRNA LFC 值排序(前 100);C:显著富集的 sgRNA(sgRNA LFC>1,FDR<0.05)的个数对应的基因数量　显著富集的 sgRNA 数量≥2 的基因在方框中列出;D:方法一和方法二的筛选结果

图 2　人肺癌候选抑癌基因的筛选结果



A:GO-BP 功能富集分析; B:KEGG 功能富集分析

图 3 人肺癌候选抑癌基因的功能富集分析

表 1 与临床 NSCLC 样本的其他高突变基因存在共突变的人肺癌候选抑癌基因

A	B	A 与 B 均未突变	A 突变但 B 未突变	B 突变但 A 未突变	A 与 B 均突变	log2 比值比[a]	*q* 值
CLTC	*KRAS*	3 062	89	822	40	0.743	0.041
FRMD6	*KRAS*	3 057	94	816	46	0.874	0.009
SAV1	*KRAS*	3 101	50	835	27	1.004	0.027
AP2M1	*TP53*	1 444	87	1 946	536	2.193	<0.001
BCR	*TP53*	1 506	25	2 377	105	1.412	<0.001
CYLD	*TP53*	2 093	36	2 800	130	1.433	<0.001
NF2	*TP53*	2 832	32	3 589	93	1.197	<0.001
PDCD10	*TP53*	1 454	77	2 014	468	2.134	<0.001
PGC	*TP53*	1 507	24	2 393	89	1.224	0.001
PSENEN	*TP53*	1 479	52	2 352	130	0.653	0.025
PTEN	*TP53*	2 788	76	3 367	315	1.779	<0.001
RINL	*TP53*	1 486	45	2 362	120	0.746	0.015
TRAF3	*TP53*	1 507	24	2 395	87	1.19	0.002
UBAP1	*TP53*	1 500	31	2 389	93	0.914	0.01

a:比值比=(A 与 B 均未突变 * A 与 B 均突变)/(A 突变但 B 未突变 * B 突变但 A 未突变);共突变:log2 比值比>0,q<0.05

2.4 人肺癌候选抑癌基因在临床 NSCLC 样本中的突变情况及其与常见 NSCLC 高突变基因的相关性分析

抑癌基因突变是促进正常细胞致瘤性转化的重要因素,为进一步了解人肺癌候选抑癌基因在临床 NSCLC 样本中的突变情况,接下来利用 Cibioportal 数据库分析了人肺癌候选抑癌基因在临床 NSCLC 样本(7 338 例)中的突变情况。结果显示,38 个候选抑癌基因均发生了不同程度的突变,*AP2M1*、*PDCD10*、*PTEN*、*SYBU*、*PSENEN*、*RINL*、*FRMD6*、*FER*、*CYLD*、*PGC*、*PCDHGB4*、*CLTC*、*TRAF3*、*BCR*、*UBAP1*、*SAV1* 及 *NF2* 共 17 个基因突变率大于 2%,其中 *PDCD10*、*AP2M1* 的突变率大于 15%,提示这些基因可能通过突变失活诱导了正常细胞的致瘤性转化。

在癌症样本中发生显著共突变的基因配对一般倾向于分布在不同生物学分子信号通路,它们的共突变导致了这些通路的共同扰动,从而诱导肿瘤发生。因此,进一步分析突变率>2%的 17 个人肺癌候选抑癌基因与临床 NSCLC 样本中发生高突变的其他基因如 *EGFR*、*KRAS*、*TP53*、*STK11* 之间的相关性。结果显示,有 14 个候选抑癌基因与这些高突变基因存在共突变(log2 比值比>0,q<0.05),其中 *AP2M1*、*BCR*、*CYLD*、*NF2*、*PDCD10*、*PGC*、*PSENEN*、*PTEN*、*RINL*、*TRAF3* 及 *UBAP1* 等基因与 *TP53* 存在共突变;*CLTC*、*FRMD6* 及 *SAV1* 等基因与 *KRAS* 存在共突变(表 1)。

2.5　人肺癌候选抑癌基因在临床 NSCLC 和癌旁正常肺组织样本中表达水平的 Meta 分析

利用 LCE 在线分析网站中 Meta 分析模块,比较了 38 个人肺癌候选抑癌基因在临床 NSCLC 和癌旁正常肺组织样本中的 mRNA 表达水平,具体分析肺腺癌(lung adenocarcinoma,LUAD)与肺鳞癌(lung squamous cell carcinoma,LUSC)两种 NSCLC 类型。结果显示,在 LUAD 中,*FER*、*PTEN*、*SAV*1、*HMGB*1、*MPP*5、*SNX*5、*KCNK*3、*TMEM*127、*DTX*4、*CYLD*、*PGC* 及 *PPP*6*C* 等基因的表达水平明显低于癌旁正常肺组织;在 LUSC 中,*FER*、*PTEN*、*SAV*1、*HMGB*1、*MPP*5、*SNX*5、*KCNK*3、*TMEM*127、*DTX*4、*CYLD*、*PGC*、*PCDHGB*4、*ABL*1 及 *CCDC*28*A* 等基因的表达水平明显低于癌旁正常肺组织。它们均与癌症发生呈负相关(图 4)。

2.6　人肺癌候选抑癌基因在临床 NSCLC 样本中的突变及表达水平与患者预后的关系

为了探究人肺癌候选抑癌基因是否与参与了肺癌发展的调控,进一步分析 38 个人肺癌候选抑癌基因的突变(cibioportal 数据库)及表达水平(LCE 数据库)与患者预后的相关性。结果显示,11 个肺癌候选抑癌基因与患者预后密切相关,其中 *AP2M*1 及 *PSENE* 等基因的突变预示着不良预后,且突变率大于 2%(图 5A、B)。17 个肺癌候选抑癌基因与患者预后密切相关,其中 *KCNK*3、*CYLD*、*PGC*、*CCDC*28*A*、*NBPF*3、*SYBU*、*GOLGA*6*L*9 及 *DTX*4 等基因的高表达与 LUAD 患者预后呈正相关,*AP2M*1 的高表达与 LUSC 患者预后呈正相关,提示这些基因的高表达可能抑制癌症的发展。*PCDHGB*4、*RAP*2*C*、*AP*2*S*1、*FRMD*6、*AFAP*1 及 *AP2M*1 等基因的高表达与 LUAD 患者预后呈负相关,*FER*、*DTX*4 及 *PCDHGB*4 等基因的高表达与 LUSC 患者预后呈负相关,提示这些基因的高表达可能促进癌症的发展(图 5C)。

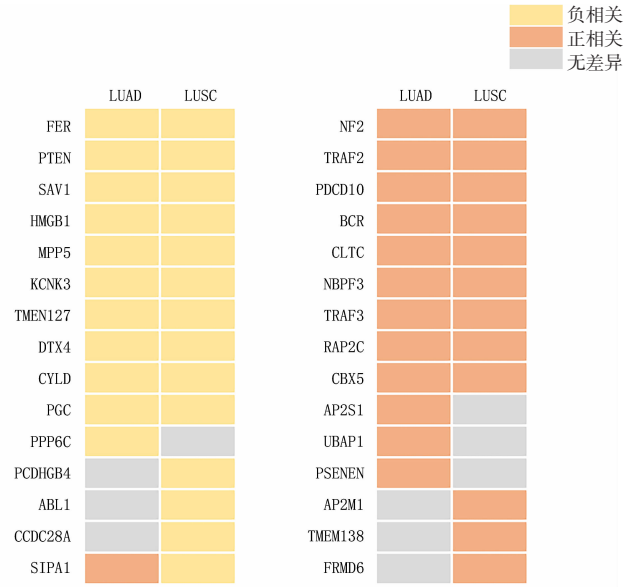
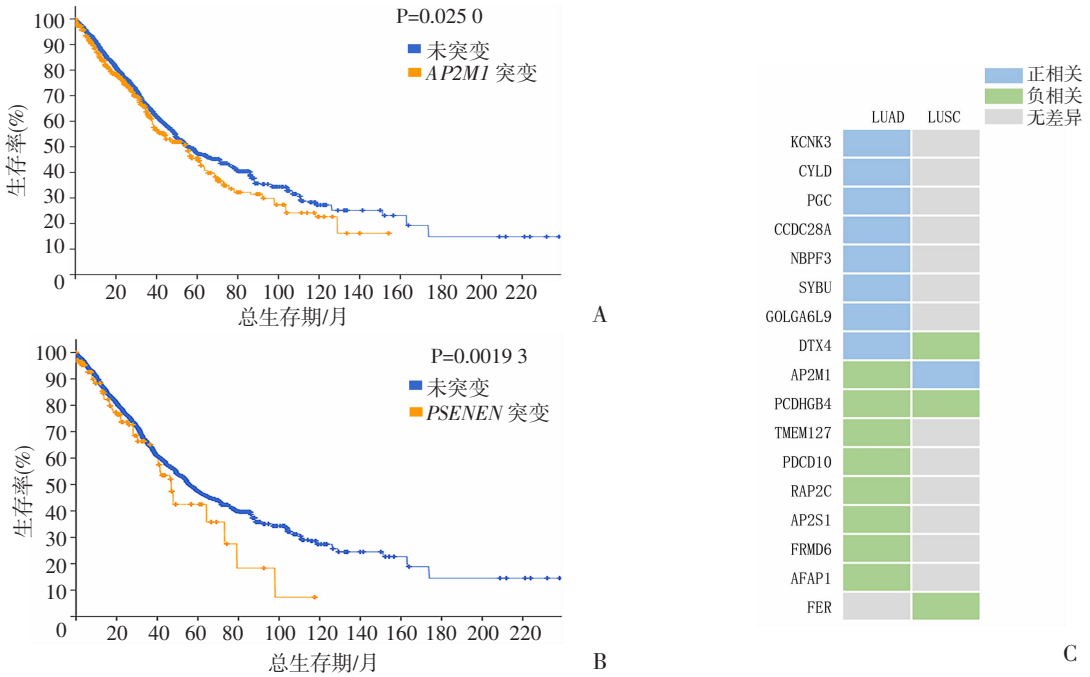


图 4　人肺癌候选抑癌基因在临床 NSCLC 和癌旁正常肺组织样本中 mRNA 表达水平的 Meta 分析



A:*AP2M*1 基因突变与患者预后的关系;B:*PSENEN* 基因突变与患者预后的关系;C:人肺癌候选抑癌基因对两种 NSCLC(LUAD、LUSC)患者预后的判断价值

图 5　人肺癌候选抑癌基因突变或表达水平与 NSCLC 患者预后的关系

3 讨论

正常细胞的致瘤性转化是癌症发生的起始,了解细胞致瘤性转化过程对研究癌症发病机制至关重要,而寻找诱导细胞致瘤性转化的新驱动基因可以为临床癌症早期筛查及诊断提供新思路。在肺癌的组织发生学上,大多数起源于上皮细胞,尤其是支气管上皮细胞。BSAE-2B 细胞是 SV40 病毒转染的人支气管上皮细胞,体外容易培养和传代,具有无限生长能力,保留了正常人支气管上皮细胞的特性和生理功能;体外连续传代 100 次以上不会发生衰老和死亡,而且不具有锚定独立生长能力和裸鼠成瘤性等恶性转化特征,表明它们是恶性前而非恶性细胞,但具有在外来致癌因素作用时获得完全恶性表型的能力[21],是研究正常细胞致瘤性转化的理想细胞模型[22]。

识别抑制肿瘤发生的基因(抑癌基因)已被证明是特别具有挑战性的,因为它们具有隐性的本质。简而言之,失活的抑癌基因的等位基因在细胞中起隐性作用,即一个拷贝失活,另一个拷贝仍以野生型存在,使细胞呈正常表型。例如,*Rb* 基因只有当野生型拷贝失活后才导致肿瘤发生[23],抑癌基因的失活机制(包括基因拷贝数和结构改变、点突变和表观遗传学改变等[24])使抑癌基因更难发现且更加复杂化。因此,本文的目的聚焦于发现在肺正常上皮细胞致瘤性转化过程中的关键抑癌基因。

本研究通过 CRISPR 全基因组敲除文库感染 BEAS-2B-Cas9 细胞,将其移植于裸鼠皮下,以获得致瘤性转化的细胞。与传统的 RNAi 技术相比,本文所使用的 CRISPR/Cas9 技术主要具有以下优势:①RNAi 技术是在 RNA 水平进行抑制,而 CRISPR/Cas9 技术是在 DNA 水平抑制基因表达,因此,CRISPR/Cas9 技术可以研究需要基因敲除才能显现的表型,也可以用于研究非转录区和非编码 RNA,这是 RNAi 技术难以实现的;②通过设计 sgRNA 混合文库,CRISPR 切割的靶点可以覆盖全基因组,而且其特异性更高,比 RNAi 技术具有更低的脱靶率与更高的性价比[11, 25]。由此可见,使用 CRISPR/Cas9 技术极大地提高了在正常人肺上皮细胞的致瘤性转化过程中识别肺癌抑癌基因的可能性。

虽然使用全基因组测序及功能基因组筛选已发现了大量的、与肿瘤发生相关的基因变异,并对其功能进行了注释[5],但由于肿瘤基因组突变的复杂性、不同类型肿瘤间的异质性及肿瘤本身的异质性,例如,每一个肿瘤细胞群体内还有许多亚群,导致许多具有促进正常细胞致瘤性转化的关键调控基因并未被发现。本研究筛选出的人肺癌候选抑癌基因,既包括了 *NF*2、*PTEN* 等已知的肺癌抑癌基因,也发现了一些未见报道的人肺癌候选抑癌基因,包括 *PSENEN*、*AP*2*M*1 等。NSCLC 发病率占肺癌总发病率的约 85%[26],因此,本研究进一步分析了这些肺癌候选抑癌基因在临床 NSCLC 样本的变化情况及其与患者预后的相关性。结果发现,*AP*2*M*1 在 NSCLC 中的突变率高达 18%,且其突变与患者的不良预后密切相关;*KCNK*3、*DTX*4、*CYLD* 及 *PGC* 在 NSCLC 中的 mRNA 表达水平明显低于正常组织,且其高表达与患者的良好预后密切相关。以上结果表明了本研究进行的功能基因筛选的有效性,为正常肺上皮细胞的致瘤性转化机制研究提供新的线索及依据。

基因的表达水平涉及转录、翻译及翻译后修饰,它们参与了许多癌基因或抑癌基因的表达调控,从而影响癌症的发生发展。本研究在 mRNA 表达水平的分析中,发现包括 *NF*2[27] 和 *TRAF*3 等在内的部分候选抑癌基因在 NSCLC 中的表达水平高于癌旁正常肺组织或无明显差异,提示这些候选基因可能是通过蛋白水平的变化导致基因功能缺失,进而促使正常细胞发生致瘤性转化。研究表明,*AP*2*M*1 通过与 *EGFR* 胞内段的 YXXΦ 共有序列结合诱导 *EGFR* 内化,从而使信号猝灭,已知 *AP*2*M*1 蛋白的功能与翻译后修饰有关,只有在 Thr156 位磷酸化的 *AP*2*M*1 才可发挥诱导 *EGFR* 内化的生物学功能,*BMP*2*K* 通过磷酸化 AP2M1-Thr156 从而在体内和体外抑制胆囊癌细胞的增殖和迁移[28]。因此,后续研究将进一步分析在蛋白水平的改变并进行功能实验验证。

在筛选出的候选抑癌基因与患者临床预后相关性分析结果显示,部分候选抑癌基因如 *PDCD*10、*AP*2*M*1 等基因的高表达与 LUSC 患者的预后呈负相关,这似乎与抑癌基因的功能矛盾。结合文献分析其可能原因如下:一方面,本研究分析的是全基因组转录本,因此,这里面可能包含了同一条基因的功能是完全相反的两条不同转录本(即正义链和反义链),如基因 *ASCC*3 的两条 RNA 功能完全相反[29];另外,在基因突变和非突变的情况下,同一基因的转录本功能也可能是完全相反的,如野生型 *P*53 和突变型 *P*53[30],但是在转录组测序时,检测的是 *P*53 的总和,而在大部分肿瘤中高表达的是突变型 *P*53,利用 KM-plotter(http://kmplot.com/analysis/)在线生存分析的网站对 *P*53 进行分析发现[31],在肺癌、卵巢癌及胃癌中,高表达 *P*53 的患者预后更差(总生存期更短)。所以,下一步要检测候选

抑癌基因是否发生了突变及是否存在功能不同的转录本,以排除前述因素的影响。另一方面,癌症的发生及发展是两个不同的过程,而癌症患者的临床预后主要与癌症的深入发展相关,同样的基因在癌症的不同发展阶段的作用完全可能是相反的。例如,*TGFβ3* 在健康组织和早期肿瘤中限制细胞增殖并诱导细胞凋亡,从而起到肿瘤抑制作用;但是,已经恶变的肿瘤细胞却丧失对 *TGFβ3* 的反应,反而通过促进 *TGFβ3* 的分泌而促进肿瘤细胞的增殖和侵袭,加速肿瘤的进展[32]。本研究是在正常细胞致瘤性转化即癌症发生起始阶段中进行筛选,因此,筛选到的候选抑癌基因完全可能在癌症的深入发展阶段发挥其他作用。

筛选功能性基因的常用策略主要分为功能获得型(gain-of-function)和功能缺失型(loss-of-function)筛选。功能获得型筛选策略通常利用 cDNA 文库使基因过表达来筛选功能性基因;功能缺失型(loss-of-function)筛选策略是通过抑制基因表达或通过基因敲除来筛选功能性基因[33]。CRISPR-Cas9 文库高通量筛选常用模式可以分为阳性筛选(positive selection screen)和阴性筛选(negative selection screen)[34]。阳性筛选是对已成功整合 sgRNA 的细胞文库施加一定的筛选压力,仅使少数目的表型的细胞能够存活,达到富集关键基因的目的;阴性筛选与之相反,存活的细胞并不是目的表型细胞,因此需要比较不同时间点 sgRNA 的丰度,找出差异 sgRNA 来确定关键基因[34]。已有报道利用 Cas9 通过功能获得型筛选来发现调节癌症转移的癌基因[35]。本研究对已整合 Brunello 敲除文库的 BEAS-2B-CAS9 细胞施加一定的筛选压力,即将其移植到裸鼠皮下,使少数获得致瘤性的细胞能够存活,从而富集关键的抑癌基因。

综上,本研究成功地利用 CRISPR/Cas9 全基因组敲除文库技术,筛选出了促使人正常肺上皮细胞致瘤性转化的 38 个人肺癌候选抑癌基因,这些抑癌基因主要富集在与细胞增殖、分化和凋亡相关的 Notch、MAPK、p38 MAPK 和 Hippo 等信号通路;进一步在临床 NSCLC 样本数据中初步验证了它们均发生了不同程度的突变,并与其他高突变基因存在共突变;还证明了其中部分基因的突变及表达水平变化与患者预后密切相关。这为进一步研究人肺癌的发生机制及临床肺癌早期筛查诊断与治疗提供新的依据和靶标。

参考文献:

[1] SUNG H, FERLAY J, SIEGEL R L, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249. DOI: 10.3322/caac.21660.

[2] CHEN W Q, ZHENG R S, BAADE P D, *et al*. Cancer statistics in China, 2015[J]. CA Cancer J Clin, 2016, 66(2): 115-132. DOI: 10.3322/caac.21338.

[3] SMITH R A, ANDREWS K S, BROOKS D, *et al*. Cancer screening in the United States, 2019: a review of current American Cancer Society guidelines and current issues in cancer screening[J]. CA Cancer J Clin, 2019, 69(3): 184-210. DOI: 10.3322/caac.21557.

[4] CICEK Y, KOSAR P, ÖZTÜRK Ö. Molecular genetics of lungcancer[J]. Eurasian J Pulmonol, 2018, 20(3): 111-117. DOI: 10.4103/ejop.ejop_33_18.

[5] CHIN L, HAHN W C, GETZ G, *et al*. Making sense of cancer genomic data[J]. Genes Dev, 2011, 25(6): 534-555. DOI: 10.1101/gad.2017311.

[6] LI Y Y, KANG K, KRAHN J M, *et al*. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data[J]. BMC Genomics, 2017, 18(1): 508. DOI: 10.1186/s12864-017-3906-0.

[7] PON J R, MARRA M A. Driver and passenger mutations in cancer[J]. Annu Rev Pathol, 2015, 10: 25-50. DOI: 10.1146/annurev-pathol-012414-040312.

[8] CHEN S D, SANJANA N E, ZHENG K J, *et al*. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis[J]. Cell, 2015, 160(6): 1246-1260. DOI: 10.1016/j.cell.2015.02.038.

[9] SONG C Q, LI Y X, MOU H W, *et al*. Genome-wide CRISPR screen identifies regulators of mitogen-activated protein kinase as suppressors of liver tumors in mice[J]. Gastroenterology, 2017, 152(5): 1161-1173.e1. DOI: 10.1053/j.gastro.2016.12.002.

[10] REDDEL R R, KE Y, GERWIN B I, *et al*. Transformation of human bronchial epithelial cells by infection with SV40 or adenovirus-12 SV40 hybrid virus, or transfection via strontium phosphate coprecipitation with a plasmid containing SV40 early region genes[J]. Cancer Res, 1988, 48(7): 1904-1909.

[11] SANSON K R, HANNA R E, HEGDE M, *et al*. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities[J]. Nat Commun, 2018, 9(1): 5416. DOI: 10.1038/s41467-018-07901-8.

[12] DOUDNA J A, CHARPENTIER E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9[J]. Science, 2014, 346(6213): 1258096. DOI: 10.1126/science.1258096.

[13] ZHOU Y Y, ZHOU B, PACHE L, *et al*. Metascape pro-

vides a biologist-oriented resource for the analysis of systems-level datasets[J]. Nat Commun, 2019, 10(1): 1523. DOI: 10.1038/s41467-019-09234-6.

[14] GAO J J, AKSOY B A, DOGRUSOZ U, *et al*. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal[J]. Sci Signal, 2013, 6(269): pl1. DOI: 10.1126/scisignal.2004088.

[15] GREENMAN C, STEPHENS P, SMITH R, *et al*. Patterns of somatic mutation in human cancer genomes[J]. Nature, 2007, 446(7132): 153-158. DOI: 10.1038/nature05610.

[16] CUI Q H, MA Y, JARAMILLO M, *et al*. A map of human cancer signaling[J]. Mol Syst Biol, 2007, 3: 152. DOI: 10.1038/msb4100200.

[17] DING L, GETZ G, WHEELER D A, *et al*. Somatic mutations affect key pathways in lung adenocarcinoma[J]. Nature, 2008, 455(7216): 1069-1075. DOI: 10.1038/nature07423.

[18] NETWORK C G A R. Comprehensive genomic characterization of squamous cell lung cancers[J]. Nature, 2012, 489(7417): 519-525. DOI: 10.1038/nature11404.

[19] NETWORK C G A R. Comprehensive molecular profiling of lung adenocarcinoma[J]. Nature, 2014, 511(7511): 543-550. DOI: 10.1038/nature13385.

[20] Correction: LCE: an open web portal to explore gene expression and clinical associations in lung cancer[J]. Oncogene, 2020, 39(3): 718-719. DOI: 10.1038/s41388-019-1000-6.

[21] AMSTAD P, REDDEL R R, PFEIFER A, *et al*. Neoplastic transformation of a human bronchial epithelial cell line by a recombinant retrovirus encoding viral Harvey ras[J]. Mol Carcinog, 1988, 1(3): 151-160. DOI: 10.1002/mc.2940010303.

[22] 秦爱平. 人支气管上皮细胞在致癌机制研究中的应用[J]. 卫生毒理学杂志, 2003, 17(3): 188-191. DOI: 10.16421/j.cnki.1002-3127.2003.03.020.
QIN A P. Application of human bronchial epithelial cells in the study of carcinogenesis mechanism[J]. J Heal Toxicol, 2003, 17(3): 188-191. DOI: 10.16421/j.cnki.1002-3127.2003.03.020.

[23] KNUDSON A G JR. Mutation and cancer: statistical study of retinoblastoma[J]. Proc Natl Acad Sci USA, 1971, 68(4): 820-823. DOI: 10.1073/pnas.68.4.820.

[24] BALMAIN A, GRAY J, PONDER B. The genetics and genomics ofcancer[J]. Nat Genet, 2003, 33(S3): 238-244. DOI: 10.1038/ng1107.

[25] EVERS B, JASTRZEBSKI K, HEIJMANS J P M, *et al*. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes[J]. Nat Biotechnol, 2016, 34(6): 631-633. DOI: 10.1038/nbt.3536.

[26] CHEN Z, FILLMORE C M, HAMMERMAN P S, *et al*. Non-small-cell lung cancers: a heterogeneous set of diseases[J]. Nat Rev Cancer, 2014, 14(8): 535-546. DOI: 10.1038/nrc3775.

[27] PETRILLI A M, FERNÁNDEZ-VALLE C. Role of merlin/NF2 inactivation in tumor biology[J]. Oncogene, 2016, 35(5): 537-548. DOI: 10.1038/onc.2015.125.

[28] SONG X L, LI M L, WU W G, *et al*. Regulation of BMP2K in AP2M1-mediated EGFR internalization during the development of gallbladder cancer[J]. Signal Transduct Target Ther, 2020, 5(1): 154. DOI: 10.1038/s41392-020-00250-3.

[29] WILLIAMSON L, SAPONARO M, BOEING S, *et al*. UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene[J]. Cell, 2017, 168(5): 843-855.e13. DOI: 10.1016/j.cell.2017.01.019.

[30] DUFFY M J, SYNNOTT N C, CROWN J. Mutant p53 as a target for cancer treatment[J]. Eur J Cancer, 2017, 83: 258-265. DOI: 10.1016/j.ejca.2017.06.023.

[31] NAGY Á, LÁNCZKY A, MENYHÁRT O, *et al*. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets[J]. Sci Rep, 2018, 8(1): 9227. DOI: 10.1038/s41598-018-27521-y.

[32] GÓMEZ-GIL V. Therapeutic implications of TGFβ in cancer treatment: a systematic review[J]. Cancers, 2021, 13(3): 379. DOI: 10.3390/cancers13030379.

[33] 李欢欢, 黄承浩. 基于CRISPR-Cas9的功能基因筛选研究进展[J]. 生物工程学报, 2018, 34(4): 461-472. DOI: 10.13345/j.cjb.170348.
LI H H, HUANG C H. Functional genetic screening using CRISPR-Cas9 system[J]. Chin J Biotechnol, 2018, 34(4): 461-472. DOI: 10.13345/j.cjb.170348.

[34] XU H, XIAO T F, CHEN C H, *et al*. Sequence determinants of improved CRISPR sgRNA design[J]. Genome Res, 2015, 25(8): 1147-1157. DOI: 10.1101/gr.191452.115.

[35] GILBERT L A, HORLBECK M A, ADAMSON B, *et al*. Genome-scale CRISPR-mediated control of gene repression and activation[J]. Cell, 2014, 159(3): 647-661. DOI: 10.1016/j.cell.2014.09.029.

(收稿:2021-12-13;修回:2022-01-04)

(编辑 张艺霖)