论文

CRISPR/Cas9-sgRNA 全基因组文库筛选人单核细胞白血病功能性促癌/抑癌基因体系的建立与优化

陈晨^①, 郝莎^{①*}, 白杨^①, 张健萍^①, 张孝兵^{①②③⑤}, 程涛^{①②③④*}^① 中国医学科学院北京协和医学院血液病医院血液学研究所, 实验血液学国家重点实验室, 天津 300020;^② 中国医学科学院干细胞医学中心, 天津 300020;^③ 北京协和医学院干细胞与再生医学系, 天津 300020;^④ 肿瘤医学协同创新中心, 天津 300020;^⑤ Department of Medicine, Loma Linda University, Loma Linda 92350, USA

* 联系人, E-mail: haosha@ihcams.ac.cn; chengtao@ihcams.ac.cn

收稿日期: 2016-03-23; 接受日期: 2016-05-19

国家自然科学基金(批准号: 81421002, 81300374, 31301206)和北京协和医学院“协和青年科研基金”(批准号: 3332015183)资助

摘要 CRISPR/Cas9-sgRNA(single guide RNA)全基因组文库筛选技术是在哺乳动物中进行系统基因组分析的有力工具. 本文通过在人急性单核白血病细胞系 THP1 上优化 CRISPR/Cas9-sgRNA 人类全基因组慢病毒文库感染体系, 构建裸鼠皮下移植肿瘤模型, 并通过优化测序建库方法, 对肿瘤组织进行二代测序以及利用生物信息学分析获得筛选结果. 裸鼠肿瘤模型表明, 感染 sgRNA 文库后的 THP1 细胞成瘤能力下降, 推测可能与某些促癌基因的敲除有关; 通过对二代测序数据进行生物信息学分析得到了一些候选基因. 实验结果表明, 在人单核细胞白血病细胞系上进行 CRISPR/Cas9-sgRNA 全基因组文库筛选是可行的, 这项工作在其它白血病细胞系或原代细胞上的筛选应用奠定了一定基础.

关键词 CRISPR/Cas9, sgRNA, THP1, 基因筛选, 二代测序

肿瘤基因组存在复杂的突变和各种遗传畸变, 破解肿瘤基因组的关键是从众多“伴随突变”(passenger mutation)中区分出驱使肿瘤进化的“驱动突变”(driver mutation)^[1]. 白血病是一类发生在造血干细胞(hematopoietic stem cell, HSC)或造血祖细胞(hematopoietic progenitor cell, HPC)水平的恶性转化

的克隆性疾病, 其中急性髓细胞白血病(acute myeloid leukemia, AML)在急性白血病中超过 50%, 是成年人最常见的急性白血病. 已知 AML 主要的“驱动突变”包括 Flt3, c-kit, MLL, TET2, RUNX1 等, 其中 MLL 突变中 t(9; 11)(p22; q23)遗传学改变, 即 MLL-AF9 融合基因是 AML 一类独特的染色体易位^[2],

引用格式: 陈晨, 郝莎, 白杨, 等. CRISPR/Cas9-sgRNA 全基因组文库筛选人单核细胞白血病功能性促癌/抑癌基因体系的建立与优化. 中国科学: 生命科学, 2016, 46: 839–850
Chen C, Hao S, Bai Y, et al. CRISPR/Cas9-sgRNA screening system in THP1 cell line for functional oncogenes and tumor suppressor genes. Sci Sin Vitae, 2016, 46: 839–850, doi: 10.1360/N052016-00148

常表现为 FAB 分型中的 M4 或 M5. 伴有 *MLL* 基因重排的血液系统恶性疾病表现出对常规治疗疗效不佳的生物学特点.

THP1, 即人急性单核白血病细胞系(human acute monocytic leukemia cell line), 最初来源于急性单核细胞性白血病患者的外周血. THP1 含有 *MLL-AF9* 融合基因, 是较为常用的一类白血病细胞系, 因此可用于进行 AML 中功能性促癌/抑癌基因的筛选.

大规模基因筛选为剖析生物学过程和人类疾病中潜在的基因功能和通路提供了重要的工具, 然而全基因组基因筛选的低通量特点使在二倍体多细胞生物上得到全基因组纯合子突变文库耗时耗力^[3]. RNA 干扰(RNA interfere, RNAi)技术在过去数年是功能基因组学采用的主要方法^[4,5], 然而 RNAi 技术也具有其内在局限性, 一方面这种转录后水平对靶基因表达的抑制通常不能充分解释敲除表型, 另外也存在对其他 mRNA 的脱靶效应. 近几年, CRISPR/Cas9 系统已逐渐成为在多种模式生物的靶基因位点引入双链断裂(double strands break, DSB)的强大工具^[6], 该系统包括 sgRNA 和 Cas9 核酸内切酶两种组分. 17~20 bp 的 sgRNA 特异地与靶向序列通过 Watson-Crick 碱基配对结合, 继而引导 Cas9 到达基因组特定区域进行 DNA 切割, 断裂的 DNA 双链可通过非同源末端连接(non-homologous end joining, NHEJ)进行有效修复^[7]. NHEJ 修复导致基因编码区发生插入或缺失(insertion-deletion, Indel)突变, 进而引起移码或终止密码子的提早出现, 导致基因敲除. 因此, 这项技术使得在全基因组规模上构建敲除载体十分简便, CRISPR/Cas9 文库为在人类等哺乳动物细胞上进行大规模敲除筛选开创了一个功能基因组学研究的新时代^[8-11].

目前, 利用 CRISPR/Cas9-sgRNA 全基因组文库筛选已在多种细胞系及原代细胞上被用于毒性基因筛选^[8]、耐药基因筛选^[10]等, 但是在白血病中的研究还不多^[12,13]. 因此选择 THP1 细胞, 利用本研究组构建的包含 54489 个 sgRNA、靶向 18163 个人类基因的 CRISPR/Cas9-sgRNA 全基因组文库, 筛选该类 AML 中功能性促癌或抑癌基因. 本研究主要通过优化条件以提高筛选效率, 并且摸索适用于此种文库筛选所需的 PCR 产物文库构建方法以获得高质量的二代测序数据, 从而建立起可行的 CRISPR/Cas9-

sgRNA 全基因组文库筛选体系, 以用于白血病功能性促癌/抑癌基因的筛选.

1 材料与方法

1.1 实验小鼠及白血病细胞

实验小鼠(*Mus musculus*): 5 周龄雌性 BALB/c-Nu 小鼠, 饲养于中国医学科学院北京协和医学院血液病医院(血液学研究所)无特定病原(specific pathogen free, SPF)级动物房, 置于正压空气清洁房间无菌鼠笼内, 垫料、饲料及饮水均经高压灭菌处理, 垫料每周更换一次, 饲料与饮水不予限制. 所有涉及小鼠的实验均得到中国医学科学院北京协和医学院血液病医院(血液学研究所)动物伦理委员会的管理与批准.

细胞: THP1 细胞购自北京协和医学院细胞中心, RPMI Medium 1640(A10491-01, Gibco, 美国), 10%胎牛血清(fetal bovine serum, FBS; Hyclone, 美国)培养于 37℃, 5% CO₂ 培养箱中, 2 天左右换液传代.

全基因组慢病毒文库: CRISPR/Cas9-sgRNA 全基因组慢病毒文库(699 lenti-U6-sghu-SFFV-cas9-2a-puro-wpre)由张孝兵教授建立. 该文库包含 54489 个 sgRNA, 其中每 3 条不同序列的 sgRNA 靶向一个基因, 可共靶向人类 18163 个基因.

1.2 文库感染体系的建立与优化

取 24 孔板, 以 $1 \times 10^6/\text{mL}$, 1 mL/孔的密度接种 THP1 细胞, 分别以感染复数(multiplicity of infection, MOI)为 0.5 和 1.0 的 CRISPR/Cas9-sgRNA 全基因组慢病毒文库感染 THP1 细胞 24 和 48 h, 再分别用 1 和 2 $\mu\text{g}/\text{mL}$ 的嘌呤霉素(puromycin, puro)筛选感染后细胞, puro 筛选前需将细胞密度调整至 $1 \times 10^6/\text{mL}$. 在感染及筛选的不同时间点对细胞进行计数及拍照, 统计比较进而获得最佳病毒感染及筛选体系. MOI 计算公式如下:

$$\text{MOI} = \frac{\text{病毒滴度} \times \text{病毒体积}(\mu\text{L})}{\text{细胞数量}}$$

定量 PCR 法测定有活性病毒的滴度为 3.5×10^7 , 包装病毒和测定病毒生物学滴度的方法详见以前发表的文章^[14].

1.3 病毒感染后 sgRNA 的 PCR 扩增及纯化

DNA 提取: 收集 CRISPR/Cas9-sgRNA 全基因组慢病毒文库感染并经 puromycin 筛选后的 THP1, 使用 QIAampDNA Mini Kit(51304, QIAGEN, 德国)提取细胞基因组 DNA, 通过细胞裂解、DNA 吸附、洗涤及洗脱等步骤, 最后将 DNA 溶于适量无酶水中. 用 NanoDrop 1000 微量紫外可见分光光度计(E112352, Thermo, 美国)检测 DNA 浓度与纯度, 用于后续 PCR 验证.

PCR 扩增: 根据慢病毒载体的序列, 在 sgRNA 插入位点两端分别设计 3 对 PCR 引物, 具体序列见表 1. PCR 扩增按照 Phusion 超保真 PCR Master Mix(M0531L, Biolabs, 英国)操作手册进行操作. 取 DNA 200 ng, sgRNA 上下游引物各 1 μ L, 2 $\times$ phusion Mix 10 μ L, 用无酶水补至 20 μ L, 进行如下 PCR 程序: 预变性 98 $^{\circ}$ C, 30 s; 然后进入 35 个扩增循环, 98 $^{\circ}$ C, 10 s 变性, 62 $^{\circ}$ C, 20 s 复性, 72 $^{\circ}$ C, 20 s 延伸; 最后于 72 $^{\circ}$ C 保持 5 min 使产物延伸完整. 然后取 PCR 产物 5 μ L 在 1.5% 琼脂糖凝胶上行电泳, 观察比对条带位置进行产物验证.

sgRNA 纯化: 为满足测序文库覆盖要求, 需将 PCR 体系调整为 100 μ L, 3 μ g DNA/体系. 分别对 PCR 扩增的反应条件, 包括循环数(28, 30, 35)和退火温度(58 $^{\circ}$ C, 60 $^{\circ}$ C, 62 $^{\circ}$ C, 63 $^{\circ}$ C, 64 $^{\circ}$ C, 65 $^{\circ}$ C)进行比较和优化, 以减少 PCR 偏倚性和非特异性对测序结果的影响. 同时利用胶纯化及样本纯化磁珠(sample purification beads, SPB, 15041032, Illumina, 美国)纯化去除目的片段外的干扰 DNA.

1.4 裸鼠皮下成瘤实验

为确定 THP1 皮下移植裸鼠的成瘤性及合适细胞数量等, 在裸鼠前肢腋下及后肢背部分别皮下接种 5×10^6 及 1×10^7 THP1 细胞, 观察成瘤时间及肿瘤生长速度, 以确定最佳细胞接种量及接种部位.

表 1 PCR 扩增引物序列

引物序号	引物类型	引物序列
sg1	sg1-F	5'-TTTCTTGCGTAGTTTGACGTTTT-3'
	sg1-R	5'-CGGACTAGCCTTATTTAACTTGC-3'
sg2	sg2-F	5'-TTTCTTGCGTAGTTTGACGTTTT-3'
	sg2-R	5'-CGGTGCCACTTTTTCAAGTT-3'
sg3	sg3-F	5'-GAATTAATTTGACTGTAAACACAAAGA-3'
	sg3-R	5'-CGGACTAGCCTTATTTAACTTGC-3'

条件确定后, 皮下移植 5×10^6 的对照组 THP1 细胞及 CRISPR/Cas9-sgRNA 全基因组慢病毒文库感染的 THP1 细胞于裸鼠前肢腋下, 每 3~4 天观察一次成瘤情况并测量记录肿瘤大小, 30 天左右处死小鼠, 进行血相分析及脏器(脾脏、肝脏、肾脏、股骨)病理分析, 检测是否发生肿瘤转移.

剥取原位肿瘤, 取少许肿瘤组织, 使用 RNeasy Mini Kit(74104, QIAGEN)提取 RNA 实验的步骤参照使用说明书, 主要分为细胞裂解、RNA 吸附、洗涤及洗脱几个步骤, 最终将 RNA 溶于适量无酶水中. 使用 NanoDrop 1000 微量紫外可见分光光度计检测 RNA 浓度及纯度, 按照 TransScript First-Strand cDNA Synthesis SuperMix(AT301, 北京全式金生物技术有限公司)说明书将提取的 RNA 逆转录成 cDNA, -20 $^{\circ}$ C 保存用于功能基因验证, 剩余肿瘤提取基因组 DNA 用来进行二代测序.

1.5 建库及测序

建库前 PCR 扩增: 按照 300 $\times$ 覆盖度进行计算, 测序的肿瘤组织需要约 120 μ g 基因组 DNA, 因此分别取肿瘤 E1(DNA 约 130 μ g)及 E2(DNA 约 150 μ g)进行 PCR 扩增. E1 和 E2 依次为实验组中成瘤体积最小和次小的两个样本. 取样本基因组 DNA 3 μ g/管, 分别进行 PCR 反应, 将各管 PCR 产物混匀.

PCR 产物的磁珠纯化及建库: 取 50 μ L PCR 混合产物用 0.6 及 1.5 比例 SPB 对 PCR 产物进行片段筛选及纯化, 纯化产物取 60 ng 进行建库. 建库采用 Truseq Nano DNA LT sample Prep 试剂盒(FC-121-4001, Illumina), 主要分为末端修复, 即加磷酸、末端加 A 尾、连接接头及 PCR 扩增富集文库等几个步骤. 文库制备完成后用 Qubit3.0 荧光定量仪(Q33216, Invitrogen, 美国), Agilent 2100 生物分析仪(G2939A, Agilent, 美国)及 KAPA 文库定量试剂盒(KK4824, KapaBiosystems, 美国)来对文库进行质控分析.

MiSeq 二代测序: 取建好的文库产物稀释至 15 pmol/L, 为了提高 PCR 产物文库的多样性, 加入 40% Phix control(FC-110-3001, Illumina). 用 MiSeq v2-300cycles 试剂盒(MS-102-2002, Illumina)在 MiSeq 测序仪(Illumina)上进行 PE150 测序分析. 取移植前感染文库细胞 T(THP1)和肿瘤细胞 E1, E2 构建的测序文库进行上机测序.

1.6 统计学处理

采用 GraphPad Prism 5 统计软件对实验数据进行统计和作图, 组间比较采用两独立样本均数 Student *t* test, 以 $P < 0.05$ 表示有统计学显著差异. 对差异基因进行 GO 信号通路分析(<https://david.ncifcrf.gov/home.jsp>).

2 结果

2.1 慢病毒文库感染体系的建立与优化

病毒文库感染主要需要考虑病毒的 MOI 及感染时间, 而针对本课题组载体上的抗 puromycin 筛选标记, 筛选过程中又需要考虑所用 puromycin 的浓度及筛选时间. 因此, 本实验分别对几个关键参数进行组合筛选, 其中包括 2 个 MOI(0.5, 1.0), 2 个感染时间(24, 48 h), 2 个 puromycin 筛选浓度(1, 2 $\mu\text{g/mL}$)及 2 个 puromycin 筛选时间(24, 48 h), 组合筛选条件如下: MOI 0.5 感染 24 h, 1 $\mu\text{g/mL}$ puromycin 筛选 24 h, MOI 1.0 感染 24 h, 1 $\mu\text{g/mL}$ puromycin 筛选 24 h 等, 共计 12 个实验组, 如统计图 1B 和 C. 在不同感染及筛选时间点, 进行细胞计数并于显微镜下观察细胞凋亡及存活状态, 最后根据计数结果绘制统计图, 确定最优体系.

未感染对照细胞随着 puromycin 加入浓度及时间的延长, 死亡细胞数量呈递增趋势, 而当 puromycin 浓度较低时, 细胞增殖速度若超过凋亡速度, 则会出现存活数目反而增加的现象. 感染细胞 MOI 0.5 组加入 puromycin 后细胞死亡情况较严重, 说明 MOI 0.5 的感染复数较低, 不能达到较高的感染效率. 观察发现, MOI 1.0 感染 48 h 组在用 puromycin 处理后, 根据与对照组未感染细胞杀伤情况比较, 当对照组细胞凋亡 80%时, 实验组以病毒 MOI 1.0 感染 THP1 细胞 48 h 后, 用 2 $\mu\text{g/mL}$ 筛选 24 h 剩余细胞最多, 且细胞存活状态较好, 形态规则, 大小均匀(图 1A), 因此选此条件为最佳条件, 即 MOI=1.0, 感染 48 h; puromycin 浓度=2 $\mu\text{g/mL}$, 筛选 24 h.

2.2 sgRNA PCR 产物扩增及纯化

本课题组设计了 3 对 PCR 引物, 首先需为后续实验选取一对合适的引物. 取感染细胞提取 DNA 并进行 PCR 反应, 通过琼脂糖凝胶电泳检测引物特异性. 由电泳图可以看出 3 对引物 sg1, sg2, sg3 均可

PCR 扩增出对应大小(180, 205 和 242 bp)的条带(图 2A), 考虑匹配测序用双端 2 $\times$ 150 测序检测试剂盒, 故选用 sg2 作为后续实验的引物.

为了满足测序样本覆盖 sgRNA 文库的完整性, 需将 PCR 体系调整为 100 μL , 3 μg DNA/体系, PCR 程序不变. 用 sg2 引物行 PCR 扩增后进行琼脂糖凝胶电泳, 发现加样孔上方有弥散条带(2D), 进行 2100 生物分析后, 发现 PCR 产物中还存在 24 bp 左右的引物杂峰, 同时在目标 sgRNA 产物附近还有一条 260 bp 左右的非特异杂峰(图 2E). 由于 2100 芯片(DNA 1000 kit, Agilent)大小的限制, 看不到超过 1500 bp 的片段, 但结合凝胶图分析可知, PCR 产物中除含有 sgRNA 产物外, 还有大片段的基因组 DNA, 引物 DNA 及非特异 260 bp 干扰片段. 这些非目的序列 DNA 的存在不仅会占用测序数据量, 还会对测序的准确度造成干扰, 因此, 测序前必须去除这些干扰 DNA 片段.

首先, 为了降低 PCR 产物非特异性, 减少 PCR 偏倚性以获得高质量的测序数据, 可通过降低循环数及升高退火温度来降低非特异产物的扩增. 比较不同的循环数(28 和 30)及不同的退火温度(58 $^{\circ}\text{C}$, 60 $^{\circ}\text{C}$, 62 $^{\circ}\text{C}$, 63 $^{\circ}\text{C}$, 64 $^{\circ}\text{C}$, 65 $^{\circ}\text{C}$), 本课题组发现, 与 30 个循环数相比, 循环数减至 28 时, 得不到较好的扩增产物(图 2B); 而退火温度的梯度筛选结果显示, 60 $^{\circ}\text{C}$ 时加样孔上方基因组 DNA 弥散最少, 高于 62 $^{\circ}\text{C}$ 时加样孔上方弥散条带明显(图 2C). 各个 PCR 条件 2100 生物分析均含 260 bp 非特异条带, 通过 PCR 方法无法去除此带. 因此, 最终选择 60 $^{\circ}\text{C}$ 退火温度, 30 个循环数作为测序用 PCR 条件.

接下来利用切胶回收纯化法以期去除出目标片段以外的杂带, 结果表明, 胶回收可去除基因组及引物片段, 但 260 bp 的杂带仍然去不掉. 此外, 胶回收容易引入外源 DNA 的污染, 因此考虑采用其他纯化方法.

SPB 可通过调整为不同的比例去除各种大小的 DNA 片段, 因此根据目标产物大小与干扰片段的大小差异, 利用 0.6 及 1.0 比例 SPB 和 0.6 及 1.5 比例 SPB 去除干扰片段. 结果发现, 两种纯化条件在电泳图上加样孔附近的基因组弥散大片段均被去除, 且 0.6 及 1.0 纯化产物较 0.6 及 1.5 纯化产物少, 而 2100 生物分析结果显示引物小片段也被去除, 然而 260 bp

图1 慢病毒感染及筛选体系的建立与优化

A: 不同筛选条件细胞状态及密度. 对照组为未感染病毒细胞, 从左至右依次为未加 puro 细胞, 加 puro 24 h 细胞, 加 puro 48 h 细胞. MOI 1.0 感染组依次为感染 48 h 未加 puro 细胞, 感染 48 h 加 puro 筛选 24 h 细胞, 感染 24 h 加 puro 筛选 48 h 细胞. 显微镜放大倍数为 10 $\times$; B: 不同浓度 puro 对不同 MOI 及感染时间细胞的杀伤情况柱状图. 从左至右柱图依次表示未感染细胞(control, CON), MOI 0.5 感染细胞, MOI 1.0 感染细胞. CON 表示未感染细胞; 0.5, 1.0 表示病毒 MOI; MOI 后 24, 48 表示感染时间(h); 1, 2 表示 puro 浓度(μ g/mL); puro 浓度后 24, 48 表示 puro 筛选时间; C: 不同浓度 puro 对不同 MOI 及感染时间细胞的杀伤情况. Y 轴起始点代表各条件均从 1×10^6 细胞数开始, 各条折线第 2 个及第 3 个点分别代表 puro 处理后 24 及 48 h 细胞数量

杂峰仍然存在, 但含量有所减少(图 2E). 因此, 最终选择 0.6 及 1.5 比例磁珠纯化法对后续样本进行纯化以进行测序. 同时, 利用纯化好的样本进行建库测试, 根据 Truseq Nano DNA Library Prep 操作手册进行, 2100 结果显示, PCR 产物片段加上接头序列后大小约 336 bp, 主峰清晰, 虽伴随少量 432 bp 片段, 但对后续测序影响不大(图 2E).

2.3 sgRNA 全基因组慢病毒文库感染后 THP1 成瘤性检测

为获得最佳接种部位, 分别于裸鼠前肢腋下及后肢背部皮下接种 5×10^6 及 1×10^7 THP1 细胞, 观察发现均在 7 天左右成瘤, 接种细胞数量大肿瘤生长也相应更快, 为了便于后续观察, 选择 5×10^6 细胞量作为皮下移植细胞数.

图2 sgRNA PCR产物扩及纯化

A: 感染细胞用3对引物PCR后凝胶电泳图. M代表Marker, sg1, sg2, sg3分别代表设计的3对引物. 180, 205和242 bp分别代表对应PCR产物片段大小; B: PCR循环数优化. 28和30代表两个循环数条件; C: PCR退火温度优化. 58, 60, 62, 63, 64, 65代表退火温度(°C); D: SPB纯化条件测试. PCR代表PCR产物, 1代表0.6及1.0纯化组合, 1.5代表0.6及1.5纯化组合; E: PCR产物、0.6及1.5比例SPB纯化产物及建库产物Agilent 2100生物分析仪分析结果. M1和M2分别代表15和1500 bp的Marker片段, sgRNA代表扩增目的产物

正式成瘤实验中, 只在小鼠前肢腋下一个接种点接种 5×10^6 的细胞, 以避免同一个体不同肿瘤之间的干扰. 取10只裸鼠, 5只接种未感染对照细胞, 5只接种感染后细胞, 观察及记录肿瘤的发生发展情况. 肿瘤生长结果显示, 文库感染后细胞的成瘤速度相比未感染的细胞有所减慢(图3B), 33天后处死小鼠, 感染组小鼠的原位肿瘤平均体积小于未感染组(图3C), 但血相分析未见异常(图3D). 对肿瘤、脾脏、肾脏、肝脏、股骨、胫骨观察拍照(图3A), 并做病理

检测, 病理结果未见肿瘤发生转移(图3E).

提取肿瘤组织基因组DNA, 取少量进行PCR, 结果显示肿瘤细胞均含有目的sgRNA, 剩余DNA待测序使用.

2.4 肿瘤组织DNA建库及测序

取裸鼠皮下肿瘤E1及E2, 提取DNA后作为测序样本, 建库流程以E1为例说明. 图4为E1的PCR产物、纯化产物及建库产物的条带分布及含量, 从电

图 3 裸鼠皮下移植实验结果分析

A: 对照组和实验组肿瘤及相关脏器照片。对照组为接种未感染细胞组, 实验组为接种病毒文库感染细胞组; B: 裸鼠皮下移植后成瘤体积随时间变化($n=10$)。X轴代表移植后天数, Y轴为肿瘤各时间点测量的肿瘤体积($ab^2/2$, a 代表肿瘤长轴长度, b 代表短轴长度); C: 裸鼠处死后剥取肿瘤重量比较($n=10$); D: 外周血象分析统计图。X轴依次表示红细胞、血小板、白细胞及淋巴细胞, Y轴表示 1 μ L 外周血中对应细胞的数量; E: 肿瘤、脏器及骨病理组织学检查石蜡切片及 H.E.染色

泳图可见如感染细胞 PCR 产物一样, 肿瘤细胞 PCR 产物除 205 bp 特异条带外, 在加样孔上方及条带下方有弥散性条带(图 4A), Agilent 2100 生物分析仪分析结果发现有引物小片段的杂峰(图 4D), 通过 0.6 及 1.5 比例的磁珠纯化法, 成功去除了这些大片段和小片段, 纯化后在电泳图及 Agilent 2100 生物分析仪分析中均看不到大片段与小片段(图 4B 和 D)。

纯化后, 样本 T 及 E 的 Qubit 定量结果如表 2, 用 T 代表感染后移植前细胞, E 代表肿瘤细胞。根据纯化后浓度计算出建库所用体积(60 ng/纯化产物浓度), 选择合适的测序接头, 依照 Truseq Nano DNA

表 2 Qubit 定量样本浓度^{a)}

测序批次	浓度(ng/ μ L)					
	Tpcr	Epcr	T 纯	E 纯	T	E
1	17.10	39.80	5.14	11.00	80.20	70.40
2	18.20	30.40	6.06	5.04	85.70	80.60

a) Tpcr, Epcr 代表 PCR 产物; T 纯, E 纯代表纯化产物; T, E 代表建库产物

Library Prep 操作手册完成建库。电泳图显示, 测序文库条带与之前测试的文库条带大小一致(图 4C), 与 PCR 产物片段大小加接头序列的长度吻合, Agilent 2100 生物分析仪分析结果也验证了此结果

图4 样本E1 PCR产物、纯化产物和建库产物分析

A: 肿瘤组织E1 PCR产物凝胶电泳图. 图中为部分PCR产物图; B: 0.6及1.5比例SPB纯化产物凝胶电泳图. Tpcr及Epcr代表PCR产物, T纯及E纯代表纯化产物. 205 bp为PCR产物条带对应大小; C: 建库产物凝胶电泳图. C1和C2为感染细胞建库对照, T和E1代表建库产物. 360 bp为PCR产物加接头序列所得建库产物的片段大小; D: 样本T及E1 PCR产物、纯化产物、建库产物Agilent 2100生物分析仪分析结果

(图4D). 进行KAPA实时定量计算出文库中加上接头的产物浓度. 文库质量及含量满足测序条件, 在Miseq二代测序仪上机测序. 第一次上机样本为T和E1, 第二次为T和E2.

2.5 测序结果分析

第一次测序(T和E1)共得到2.6 G的数据量, 对照组与实验组各占1/2, 与预估数据量基本吻合. 第二次测序(T和E2)共得到2.6G数据量, 对照组与实验组数据量比例约为1:2. 第一次数据分布较为均衡, 第二次分布较不均匀. 对两次测序结果进行对比分析, 发现对照组的数据较为稳定, 说明测序系统得

到的结果可信度也较高.

将二代测序结果与人基因组CRISPR/Cas9 sgRNA文库中的序列进行比对分析, 进而获得T和E中sgRNA的绝对reads读数. 通过与对照组(T)比较, 结果发现E1肿瘤中相对移植前细胞富集2倍以上的sgRNA对应的基因共有6382个, E2中共有4295个, 二者共同富集的sgRNA对应的有1219个基因. 这些富集的sgRNA对应的基因中有1086个只有对应的1条sgRNA被富集, 122个有对应的2条sgRNA, 而有11个基因其对应的3条sgRNA都富集了2倍以上(图5A). 3条sgRNA都富集的基因更有可能在肿瘤发生中起重要的调控作用.

图5 测序富集 sgRNA 对应基因分析

A: 富集 2 倍以上 sgRNA 对应基因数量及同一基因 sgRNA 富集条数统计. 左边的韦恩图代表 E1 和 E2 中分别富集的 sgRNA 对应的基因个数, 其交集代表 2 个样本中均富集的 sgRNA 对应的基因个数. 右边图表示共同富集 sgRNA 对应基因中 sgRNA 富集的情况, 每个基因有 3 条 sgRNA; B: 富集 2 倍以上 sgRNA 对应基因聚类分析. 左边代表基因调控的生命过程, 右边的数字代表与该过程有关的基因数量

进一步, 对这些富集的基因进行 GO 聚类分析 (图 5B), 发现与细胞增殖、生长及调控细胞周期等过程有关的一些基因, 这些基因可能是影响肿瘤形成的关键点. 分析肿瘤中高度富集的 sgRNA 以期发现重要的基因, E1 中富集 100 倍以上的 sgRNA 对应的基因及 E2 中富集 40 倍以上 sgRNA 对应的基因各有 10 个, 其中, 二者共同富集 10 倍以上的 sgRNA 共有 30 个.

这 30 个候选基因中, 据报道有一些是与肿瘤形成发展及细胞凋亡等相关的. 例如, *TUBG2* 过表达可能促进肿瘤形成与发展^[15], 而 *BAG4* 可能发挥抗细胞凋亡的作用^[16], *RHOB* 在肺腺癌、胶质瘤等肿瘤中可促进肿瘤的形成与转移^[17], 这些与促肿瘤生成有关的基因在实验组中表达受抑, 因此肿瘤生长受抑. 可选择这几个基因进行后续功能验证, 观察其单独是

否可在 THP1 细胞系中引起抗癌作用, 其他未有相关报道的基因也可一一进行验证, 以期发现功能性促癌基因.

3 讨论

急性白血病是血液系统常见的恶性肿瘤之一, 病程短, 死亡率高. 目前化疗是其主要治疗手段之一^[18]. 虽然大部分急性白血病患者化疗后能够获得较高的缓解率, 但多数患者在缓解后一年内复发, 导致其 5 年生存率仍较低. 究其原因可能是存在“驱动突变”的肿瘤干细胞未被完全杀死, 一旦有机会便会卷土重来, 造成白血病的复发难治^[19]. 骨髓移植虽然能拯救部分化疗效果差或高危病人, 但仍存在移植

后白血病复发和移植相关并发症等问题。另外,白血病的种类繁多,且存在高度异质性。一般认为白血病的发生至少有两个阶段:(i)各种原因所致的单个细胞原癌基因决定性的突变;(ii)进一步启动一个或多个癌基因的激活或抑癌基因的失活,导致细胞凋亡或衰老受阻、恶性增殖或分化阻滞。因此,发现白血病发生发展中功能性促癌/抑癌基因,探索白血病的发病机制,以寻找特异、高效和低毒的靶向治疗方法,是临床针对病人异质性采用个体化治疗的基础,也是白血病研究中的一大热点。

CRISPR-Cas9 文库筛选技术有许多优势。Chen 等人^[20]利用“小鼠全基因组 CRISPR 敲除文库 A”(mGeCKOa),以及 Cas9 DNA 切割酶对来自非小细胞肺癌(non-small cell lung cancer, NSCLC)小鼠模型的细胞进行处理,随后将这些细胞移植到小鼠体内,发现用这一基因敲除文库处理的一些细胞形成了高转移性肿瘤。利用新一代测序鉴别出了在原发肿瘤及转移灶中敲除的基因,其中包含一些在人类肿瘤中众所周知的肿瘤抑制基因,包括 *PTEN*, *CDKN2A* 和 *NF2*,也涵盖了一些从前未与癌症关联的基因。在耐药性筛选应用中,该研究团队构建了一个全基因组范围内的 CRISPR 敲除文库(GeCKO),并在黑色素瘤模型中鉴定了对维罗非尼产生抵抗的突变,包含之前曾发现的 *NF1* 和 *MED12*,以及未曾发现过的 *NF2*, *CUL3*, *TADA2B* 和 *TADA1*^[11]。但与此同时,体内筛选受到许多因素的影响,如文库的复杂性、病毒感染或细胞移植的限制性、细胞在动物体内与其他细胞相互作用的复杂性等,因此,在不同的细胞系或原代细胞上进行筛选体系优化是很有必要的,这使得在 THP1 细胞系上的筛选体系优化对其他类似白血病细胞系的基因组研究具有一定指导意义。

本研究对慢病毒文库感染体系进行优化并建立起相应的小鼠肿瘤模型,对二代测序中建库方法进行优化,建立起 CRISPR/Cas9-sgRNA 全基因组文库筛选人单核细胞白血病功能性促癌/抑癌基因体系,并发现了一些可能的促癌形成基因,如 *TUBG2*, *BAG4* 及 *RHOB* 等。

需要提出的是,本研究对照组采用的是未经处理的 THP1 细胞,观察到的表型为感染病毒后的细胞成瘤能力减弱,降低的成瘤性可能是高度富集的 sgRNA 造成的某些基因敲除所导致的,但也不排除病毒感染本身对细胞造成的损伤,因为慢病毒转导后,通常有一定毒性,会让细胞生长速度减慢。另外,细胞表达 Cas9 等后,会消耗细胞的资源,增加额外的负担,使细胞生长速度受到影响。在接下来的实验中,本课题组会考虑构建带有不靶向基因的随机序列的病毒感染细胞作为对照,以消除病毒感染的影响。在本实验体系中有两套对照,一套为空载对照,另一套为感染后移植前的 THP1 细胞,而 THP1 细胞群这个对照是测序所采用的对照,是获得数据分析结果的主要来源。另外,裸鼠皮下移植成瘤平均值在对照组与感染组有差异,但各组内肿瘤大小差异很大,说明在每个肿瘤中起主要作用的 sgRNA 也许并不相同。因此,分析多个样本中的 sgRNA 富集情况,合理分析,即不忽略个别潜在基因的重要性,也要积极排除假阳性,才能获得更加有意义的结果。测序获得的目标基因需要后续功能验证,即发现的重要基因需要克隆单独的 sgRNA 和过表达慢病毒载体进行下一步研究,这样可以从正反两方面来证实某基因对特定功能的促进或抑制作用。除此以外,若发现的新基因有小分子抑制剂,可用其替代 sgRNA,观察是否有类似的表型。

致谢 北京协和医学院血液学研究所杨铭老师给予了皮下注射支持,北京协和医学院血液学研究所董芳博士对本课题提供了宝贵意见,北京协和医学院血液学研究所刘翠翠博士对本文的编辑、讨论及建议提供了帮助,在此一并致谢。

参考文献

- 1 Anosha P, Huang L T, Sakthivel R, et al. Discrimination of driver and passenger mutations in epidermal growth factor receptor in cancer. *Mutat Res*, 2015, 780: 24-34
- 2 Marschalek R. MLL leukemia and future treatment strategies. *Arch Pharm (Weinheim)*, 2015, 348: 221-228

- 3 Shen Z, Ou G. CRISPR-Cas9 knockout screening for functional genomics. *Sci China Life Sci*, 2014, 57: 733–734
- 4 Schramek D, Sendoe A, Segal J P, et al. Direct *in vivo* RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. *Science*, 2014, 343: 309–313
- 5 Shao D D, Xue W, Krall E B, et al. KRAS and YAP1 converge to regulate EMT and tumor survival. *Cell*, 2014, 158: 171–184
- 6 Charpentier E, Doudna J A. Biotechnology: rewriting a genome. *Nature*, 2013, 495: 50–51
- 7 Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, 1994, 14: 8096–8106
- 8 Zhou Y, Zhu S, Cai C, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*, 2014, 509: 487–491
- 9 Koike-Yusa H, Li Y, Tan E P, et al. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol*, 2014, 32: 267–273
- 10 Wang T, Wei J J, Sabatini D M, et al. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, 2014, 343: 80–84
- 11 Shalem O, Sanjana N E, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 2014, 343: 84–87
- 12 Wallace J, Hu R, Mosbrugger T L, et al. Genome-wide CRISPR-Cas9 screen identifies microRNAs that regulate myeloid leukemia cell growth. *PLoS One*, 2016, 11: e0153689
- 13 Shi J, Wang E, Milazzo J P, et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*, 2015, 33: 661–667
- 14 Meng X, Baylink D J, Sheng M, et al. Erythroid promoter confines FGF2 expression to the marrow after hematopoietic stem cell gene therapy and leads to enhanced endosteal bone formation. *PLoS One*, 2012, 7: e37569
- 15 Cho E H, Whipple R A, Matrone M A, et al. Delocalization of gamma-tubulin due to increased solubility in human breast cancer cell lines. *Cancer Biol Ther*, 2010, 9: 66–76
- 16 Frisch S M. Evidence for a function of death-receptor-related, death-domain-containing proteins in anoikis. *Curr Biol*, 1999, 9: 1047–1049
- 17 Ma Y, Gong Y, Cheng Z, et al. Critical functions of RhoB in support of glioblastoma tumorigenesis. *Neuro Oncol*, 2015, 17: 516–525
- 18 Stein E M, Tallman M S. Emerging therapeutic drugs for AML. *Blood*, 2016, 127: 71–78
- 19 Cheng T. Toward “SMART” stem cells. *Gene Ther*, 2008, 15: 67–73
- 20 Chen S, Sanjana N E, Zheng K, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*, 2015, 160: 1246–1260

Establishment and optimization of genome-wide CRISPR/Cas9-sgRNA screening system in THP1 cell line for functional oncogenes and tumor suppressor genes

CHEN Chen¹, HAO Sha¹, BAI Yang¹, ZHANG JianPing¹, ZHANG XiaoBing^{1,2,3,5}
& CHENG Tao^{1,2,3,4}

*1 State Key Laboratory of Experimental Hematology, Blood Disease Hospital & Institute of Hematology,
Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China;*

2 Stem Cell Medical Center, Chinese Academy of Medical Sciences, Tianjin 300020, China;

3 Department of Stem Cell and Regenerative Medicine, Peking Union Medical College, Tianjin 300020, China;

4 Collaborative Innovation Center for Cancer Medicine, Tianjin 300020, China;

5 Department of Medicine, Loma Linda University, Loma Linda 92350, USA

Genome-wide CRISPR/Cas9-sgRNA screen is a powerful tool for systematic genetic analysis in mammalian cells. In this study, we optimized the lentivirus-based CRISPR/Cas9-sgRNA system in the THP1 cell line and established the nude mouse tumor model for functional gene screen. Through the optimization of the sequencing library construction, we performed the next generation sequencing on the tumor cells and obtained the screening results by bioinformatics analyses. The mouse tumor model showed decreased tumorigenic ability of sgRNA transduced cells, which may be due to knock out of some oncogenes. We confirmed some candidate genes through the sequencing analysis. Therefore, the genome-wide CRISPR/Cas9-sgRNA screening system in THP1 cell line is feasible. This work sets the stage for our extended screen for functional genes in other leukemia cell lines or primary cells.

CRISPR/Cas9, sgRNA, THP1, gene screen, next generation sequencing

doi: 10.1360/N052016-00148