

分泌蛋白特异性基因陷阱的设计与验证*

孙 强 韩 骅**

(第四军医大学医学遗传学与发育生物学教研室, 唐都医院干细胞研究中心, 西安 710032)

摘要 蛋白质分子向细胞外分泌的过程有赖于信号肽的介导. 基因陷阱是功能性基因克隆的一种有效方法, 经典的基因陷阱以 *geo* 作为报告基因, 对所克隆的基因类型没有选择性. 将绿脓杆菌外毒素、2A 序列、IL-2 受体穿膜区以及新霉素磷酸转移酶的基因依次融合构建新型报告基因——*peo*, 该报告基因能够特异性地甄别具有信号肽编码序列的基因. 为验证 *peo* 对信号肽编码序列的特异性甄别作用, 以 *peo* 为报告基因, 构建了 3 种质粒载体, 分别模拟用基因陷阱载体进行筛选时可能出现的 3 种情况, 通过转染 HeLa 细胞、G418 筛选以及碱性磷酸酶检测, 证明 *peo* 能够有效地区分信号肽和非信号肽编码序列, 以 *peo* 为报告基因改造的基因陷阱载体可以用于分泌蛋白基因的筛选.

关键词 基因陷阱, 信号肽, *peo*, 绿脓杆菌外毒素, 新霉素磷酸转移酶

学科分类号 Q785, Q789

分泌蛋白(细胞外蛋白和细胞膜表面受体)在生物体个体发育、生理功能的发挥及各种病理过程的演进中起着重要作用. 克隆分泌蛋白的基因并研究其功能, 有助于我们对生命现象的深入认识和临床新药的开发. 1993 年, Tashiro 等^[1]以 IL-2 受体作为报告基因首先建立了信号肽捕获系统(signal sequence trap, SST), 该系统的最大特点是可以特异性克隆真核细胞 cDNA 文库中编码分泌蛋白的基因序列. 以此为基础, 随后的科学家使用不同报告基因对 SST 进行改进、优化^[2].

基因陷阱(gene trap)方法是一种直接在胚胎干细胞(embryonic stem cell, ES cell)中进行筛选的方法. 该方法的核心是基因陷阱报告基因的设计, 经典的载体将 β -半乳糖苷酶(β -galactosidase, β -gal)与新霉素磷酸转移酶(neomycin phosphotransferase, Neo)融合作为报告基因(*geo*), 然后在 5' 端融合一段拼接受体序列(splicing acceptor, SA), 所构建的载体整合入基因组后, 就可以利用 SA 与上游表达的内源性基因发生 RNA 的选择性剪切拼接, 形成内源性基因与 *geo* 的融合转录本, 由其翻译产生的蛋白质同时具有 β -gal 和 Neo 的活性, 可以用 G418 进行阳性克隆的筛选, 用 LacZ 显色监测其表达情况, 与报告基因 *geo* 融合基因可以通过 cDNA 末端快速扩增法(rapid amplification of cDNA ends, RACE)的办法来克隆^[3]. 然而以 *geo* 作为报告基因构建的载体对所筛选的基因类型没有选择性.

本研究将绿脓杆菌外毒素(PE40)、2A 序列、

IL-2 受体(*hTac*)穿膜区以及 Neo 基因依次融合, 设计了一种新型基因陷阱报告基因——*peo* (PE40-2A-TM-Neo)(图 1), 该报告基因能够特异性甄别具有信号肽编码序列的基因. 为验证 *peo* 对信号肽编码序列的特异性甄别作用, 以 *peo* 为报告基因, 构建了三种质粒载体, 分别模拟用基因陷阱载体进行筛选时可能出现的 3 种情况: a. 没有与基因组上的基因发生 RNA 水平的融合, 或融合不正确——pPeo-IP; b. 发生了正确的融合, 但融合的顺序不能编码信号肽序列——pATG-Peo-IP; c. 发生了正确的融合, 且融合的顺序能够编码信号肽序列——pSS-Peo-IP. 通过转染 HeLa 细胞、G418 筛选以及碱性磷酸酶检测, 证明 *peo* 能够有效区分信号肽和非信号肽编码序列, 以 *peo* 为报告基因改造的基因陷阱载体可以用于分泌蛋白基因的筛选.

1 材料和方法

1.1 细菌、细胞系和质粒

大肠杆菌 DH5 α 、HeLa 细胞以及质粒载体 pCDNA3.0 (含有 Neo 基因)由本室保存; 含有 *hTac* 基因的质粒 pCDL-SR α -*hTac* 由日本东京大学医学部本庶佑教授惠赠; 含有 PE40 基因的载体 pCMV-23escfv-PE40 由本校生物化学教研室杨安钢

* 国家重点基础研究发展规划项目资助 (2001CB509900, 001CB509906).

** 通讯联系人.

Tel: 029-3374490, E-mail: huahan@fmmu.edu.cn

收稿日期: 2003-09-08, 接受日期: 2003-10-31

教授惠赠; 含有人胎盘碱性磷酸酶 (*hPLAP*) 基因的载体 pGT-pfs 由 Skarnes 博士惠赠。

1.2 试剂

寡核苷酸由大连宝生物工程有限公司合成。细胞培养试剂购自 Gibco BRL 公司, 其他分子生物学试剂分别购自上海华舜生物有限公司、大连宝生物工程有限公司、Promega、Sigma 等公司。

1.3 PCR 扩增

使用 Takara 的 LA *Taq* 扩增所有目的片段, *PE40* 的扩增使用 2 × GC 缓冲液。以下列出实验中使用到的引物与 linkers 序列。

PE40: *PE1*, 5' AAGAATTGCGGCCGCGAAA-GTCGACCAGGTGATCCGC 3'; *PE2*, 5' AAGAT-CTCGGCGGTTTGCCGGGCTGGC 3'. *hTac*: *TM1*, 5' ACTAGTGTGGGGCAGATGGTTTATTAT 3'; *TM2*, 5' GATATCTCTACTCTTCCTCTGTCTC 3'. *Neo*: *Neo1*, 5' GATATCATGATTGAACAAGATGG 3'; *Neo2*, 5' CTCGAGCTTGCTCGGTCATTTTCG 3'. *2A linker*: *A1*, 5' GATCT CAACTGCTGAACCTTTGACCTG 3'; *A2*, 5' CTCAAGTTGGCTGGAGATGTGGAGTCCAAC-CCTGGGCCC A 3'; *A3*, 5' CTAGT GGGCCCAGGG-TTGGACTCCAC 3'; *A4*, 5' ATCTCCAGCCAACCTG-AGCAGGTCAAAGTTCAGCAGTTG A 3'.

1.4 细胞培养与基因转染

HeLa 细胞常规培养: DMEM 培养液含 10% FCS, 50 U/ml 青霉素, 50 g/L 链霉素, L-谷氨酰胺。G418 用无血清培养液溶解、过滤后, 直接加入常规培养基中作为筛选培养基。使用 LipofectaMIN™ 进行基因转染, 按照说明书进行操作。

1.5 碱性磷酸酶检测

用武汉博士德公司的 NBT/BCIP 试剂进行碱性磷酸酶活性的检测, 按照说明书进行操作。

1.6 基因组 DNA 提取

目的细胞扩增后, 使用上海华舜生物有限公司

基因组抽提试剂盒提取细胞基因组 DNA。

1.7 基因克隆

按照参考文献 [4] 进行。

1.8 DNA 序列分析

由上海生工生物有限公司完成。

2 结 果

2.1 分泌蛋白特异性基因陷阱的设计

信号肽是分泌蛋白的共同特征, 因此选择是否具有信号肽作为区分分泌蛋白与细胞内蛋白的一个指标。为了使所设计的报告基因能够有效地指示与其融合的序列是否具有信号肽, 将绿脓杆菌外毒素 (*PE40*)、2A 序列、IL-2 受体 (*hTac*) 穿膜区以及 *Neo* 基因依次融合得到报告基因 *peo* (图 1), 其中, *PE40* 由绿脓杆菌外毒素基因 C 端的 3 个结构域 (II, I b, III) 组成, 能够修饰延长因子 2 (elongation factor 2, EF2), 使其失活, 从而杀死细胞^[5]。为了避免由于基因融合造成的活性丧失, 在 *PE40* 与 *hTac*-TM 之间插入一段 2A 序列, 该序列编码的短肽在翻译的过程中会发生断裂^[6], 使得 *peo* 表达后产生两种蛋白质产物: 毒素蛋白和带有 *hTac* 穿膜区的新霉素磷酸转移酶。后者可以赋予细胞 G418 抗性, 前者如果表达在细胞内就会杀死细胞 (没有信号肽), 如果通过分泌途径分泌到细胞外 (由信号肽介导), 细胞就会存活。这样, 用 *peo* 作为报告基因构建的基因陷阱载体, 在 G418 存在的情况下进行筛选, 得到的基因应该都能够编码信号肽 (将毒素蛋白分泌到细胞外), 否则就会被 G418 或毒素蛋白杀死。此外, 为了便于形态学上的观察, 在 *peo* 的下游用内部核糖体进入位点 (internal ribosome entry site, IRES) 序列共表达了一个人 *PLAP* (human placenta alkaline phosphatase) 基因, 用于指示筛选后的阳性克隆, 以及 ES 细胞胚胎注射后目的基因的组织表达分布。

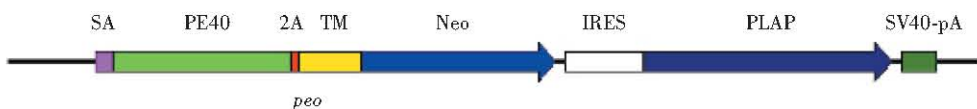


Fig. 1 The schematic representation of designed *peo* gene trap construction specific for secretory proteins

SA: Splicing acceptor; *PE40*: Domain II, I b and III of *Pseudomonas* exotoxin A; 2A: Sequence from foot-and-mouth disease virus; TM: The trans-membrane encoding sequence of *hTac* gene; *Neo*: Neomycin phosphotransferase; IRES: Internal ribosome entry site; *PLAP*: Placenta alkaline phosphatase; SV40-pA: The poly (A) sequence of SV40 virus gene; *peo*: The fused product of *PE40*, 2A, TM and *Neo*.

2.2 目的基因 PCR 扩增

用 PCR 方法分别以含有相应基因的质粒载体作为模板, 扩增 *PE40*、*hTac* 穿膜区 (*hTac*-TM) 以及 *Neo* 基因, 均得到特异性单一条带 (图 2),

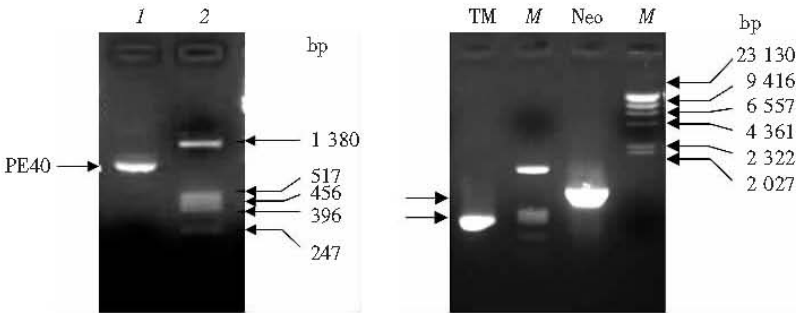


Fig. 2 Amplification of *PE40*, *hTac*-TM and *Neo* genes
Arrows indicate PCR products. Molecular mass marker (*M* and 2) are shown beside the graphs. 1: *PE40*; TM: the trans-membrane encoding sequence of *hTac* gene; Neo: neomycin phosphotransferase.

2.3 载体构建

限制性酶切去除 pCDNA3.0 中的 *Neo* 基因以及 *Bgl* II 位点 (切开补平), 得到 pCND. 依次将序列正确的 *PE40*、*hTac*-TM 以及 *Neo* 基因插入 pCND, 并在 *PE40* 与 *hTac*-TM 之间连入 2A linker, 得到载体 pPeo, 用 *Xba* I 从 pGT-pfs 上切下带有 IRES 序列的 *PLAP* 基因, 插入 pPeo 相应位点, 得到 pPeo-IP (图 3), 然后分别在 pPeo-IP 的多克隆位点处 (*PE40* 基因上游) 插入 23e *scfv* (提供信号肽编码序列) 和 *GFP* (提供起始密码子 ATG), 最后得到验证载体 pSS-Peo-IP 和 pATG-Peo-IP.

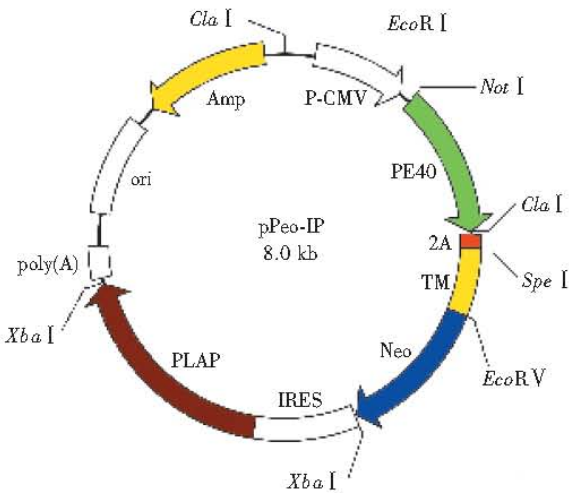


Fig. 3 The structure of plasmid vector —pPeo-IP

克隆入 pGEM-T (easy) 后进行序列分析, 通过基因克隆操作去除影响氨基酸编码的突变, 最后得到 3 种基因完全正确的克隆, 用于随后的载体构建.

2.4 细胞转染验证 *peo* 的有效性

分别将质粒 pEGFP-N1 (对照)、pPeo-IP、pSS-Peo-IP 和 pATG-Peo-IP 转染 HeLa 细胞, 转染后 48 h 将细胞按照 1:10 传代, 取一部分爬片后用 NBT/BCIP 检测碱性磷酸酶的表达. 其余的分别用不同浓度 (0、200、400、600 mg/L) G418 进行筛选, 两周后扩增阳性细胞克隆, 组织化学方法检测碱性磷酸酶的表达, RT-PCR 检测 *peo* 的表达, 同时提取细胞的基因组 DNA, PCR 检测载体在基因中的整合.

2.4.1 碱性磷酸酶表达分析: 转染后 48 h 检测各组细胞碱性磷酸酶的表达, 结果只有转染 pPeo-IP 和 pSS-Peo-IP 的细胞碱性磷酸酶的表达阳性, pEGFP-N1 和空白对照 (未作任何转染) 无表达, pATG-Peo-IP 转染的细胞只零星检测到一些信号, 但没有细胞形态, 存活的细胞不表达碱性磷酸酶, 说明 *peo* 能够表达而且具有 *PE40* 的细胞毒性. 筛选两周后只有转染 pSS-Peo-IP 的细胞有存活克隆, 而且所有存活细胞都表达碱性磷酸酶 (图 4), 说明表达的 *peo* 分子具有 Neo 活性, 能够有效的分解 G418, 使细胞存活. 以上结果与预期完全相符, 证明报告基因 *peo* 能够区分信号肽和非信号肽编码序列.

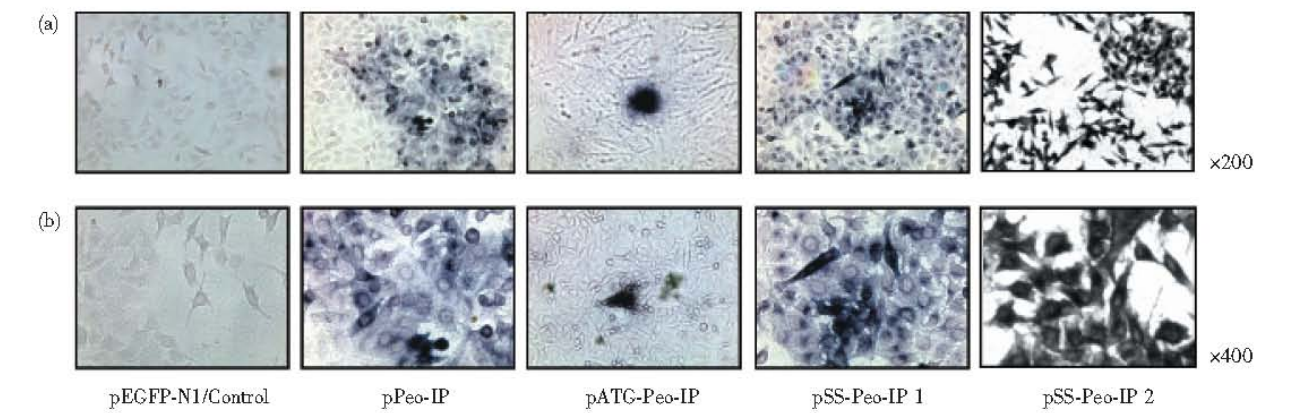


Fig. 4 Detection of hPLAP expression after transfection or selection

All the result were got 48 h after transfection with the plasmids indicated under the graphs except for pSS-Peo-IP 2 which is after 2 weeks selection with G418 (200 mg/L). (a) ×200, (b) ×400.

2.4.2 G418 筛选：为进一步确定 peo 的有效性，以及用于基因陷阱筛选的具体条件，分别将转染后的细胞用不同浓度的 G418 筛选两周，结果除阳性

Table 1 Selection with G418 of different concentration after transfection						
ρ (G418) / (mg · L ⁻¹)	0	200	300	400	600	
pEGFP-N1	+/-	+/-	+/-	+/-	+/-	
pPeo-IP	+/+	-/-	-/-	-/-	-/-	
pATG-Peo-IP	+/-	-/-	-/-	-/-	-/-	
pSS-Peo-IP	+/+	+/+	+/-	-/-	-/-	
Control	+/-	-/-	-/-	-/-	-/-	

Symbols (“ + ” and “ - ”) before and behind “ / ” represents cells surviving condition and hPLAP staining results respectively.

对照 pEGFP-N1 转染的细胞外，在 200 mg/L 浓度下，只有 pSS-Peo-IP 转染的细胞存活（表 1），提取细胞的基因组 DNA 作为模板，未转染细胞作对照，分别用两对引物（P1/TM2，TM1/Neo2）进行扩增，结果得到特异性扩增条带，大小与预期相符（图 5a），说明 pSS-Peo-IP 已经整合入细胞基因组。同时，提取细胞的总 RNA，反转录后用 RT-PCR（引物对：P1/Neo2）检测到 peo 的表达（图 5b）。用碱性磷酸酶检测显示所有细胞都有 PLAP 的表达（图 4），说明报告基因 peo 和 PLAP 都能够有效地表达。以上结果说明 peo 可以作为分泌蛋白特异性基因陷阱的报告基因，且用于筛选的 G418 浓度以 200 mg/L 为宜。

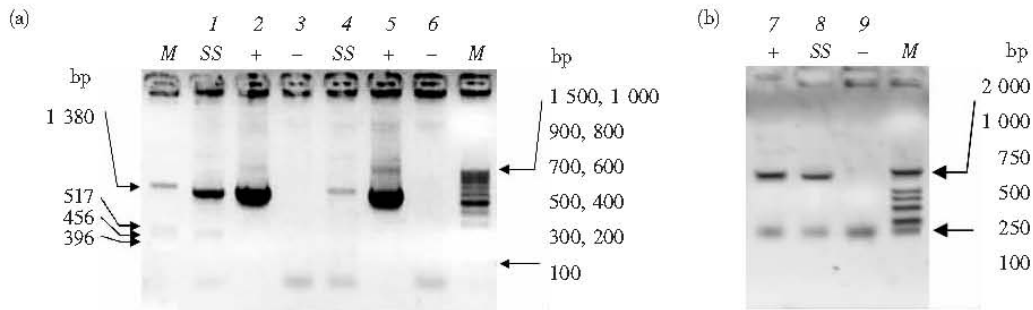


Fig. 5 Detecting the integration and expression of pSS-Peo-IP in HeLa cells

(a) Amplification of the integrated pSS-Peo-IP from genomic DNA of HeLa cells. (b) Detecting the expression of peo in HeLa cells stably transfected with pSS-Peo-IP. Molecular mass is shown beside the graph. M: molecular mass; SS: genomic DNA or cDNA from cells transfected with pSS-Peo-IP; +: pSS-Peo-IP plasmid; -: control genomic DNA or cDNA from untransfected cells. Primer pairs used in reaction 1, 2, 3 are P1/TM2; 4, 5, 6 are TM1/Neo2; and 7, 8, 9 are P1/Neo2.

3 讨 论

在以往的工作中,本实验室建立了 *suc2*-信号肽捕获系统 (*suc2*-SST)^[7],并用该系统对小鼠胚胎 11 天 cDNA 文库进行了筛选(待发表),筛选的结果证明用 *suc2*-SST 筛选 cDNA 文库的确非常高效、简便。但是由于筛选是在酵母细胞中进行的,筛选的结果存在假阳性率偏高,特异性不太好等问题。与酵母系统相比,哺乳动物细胞系统在这方面有明显的优越性^[8]。此外,还发现由于该系统的筛选对象是 cDNA 文库,所以该系统有两个问题难以解决: a. 是由于文库丰度差异导致的假阴性; b. 是由于特异性表达差异导致的假阴性。这两点使得低丰度 cDNA 以及时间和组织特异性表达的分子难以被克隆。

针对以上问题,本实验室对传统的基因陷阱载体进行改造,设计构建了可以用于对分泌蛋白基因进行筛选的 *peo* 基因陷阱报告系统。并以 *peo* 报告基因为核心构建了 3 种质粒载体进行模拟筛选,验证了 *peo* 报告系统用于分泌蛋白筛选的有效性。由于基因陷阱系统的核心是报告基因,报告基因被验证有效以后,只需要在报告基因上游添加一段拼接受体序列就可以作为有用的基因陷阱载体。本研究采用的拼接受体序列来源于 Skarnes 博士惠赠的基因陷阱载体 pGT1.8TM^[9],其有效性是可信的,而报告基因 *peo* 已被本研究证实是有效的,因此,理论上本研究构建的 *peo* 基因陷阱载体是可能有效地对分泌性蛋白分子编码基因进行筛选的。

用该系统在 ES 细胞中进行筛选,减少了酵母系统中由于不识别哺乳动物蛋白信号肽造成的假阳性。同时,与针对 cDNA 文库进行筛选的方法相比,使用基因陷阱方法对分泌蛋白进行筛选还有以下突出优点: a. 直接对基因组进行筛选,克服了表达丰度对筛选的影响; b. 可以方便地研究落入陷阱的基因对特定刺激的反应性; c. 由于筛选是在 ES 细胞进行的,而基因陷阱载体在基因组上插入对内源性基因是一个突变因素,所以可以用筛选得到的细胞克隆直接进行胚胎注射,制造基因剔除小鼠用于基因功能研究。

目前,已报道有两种基因陷阱方法可以用于分泌蛋白的筛选,他们分别采用了不同的报告基因: a. 1995 年,Skarnes 等^[9]将 CD4 穿膜区与 *geo* 融合作为报告基因; b. 2001 年,Gebauer 等^[10]将 CD2 分子与 Neo 融合得到报告基因 *Ceo* (CD2-Neo)。由

于这两种方法在报告基因的设计中没有引入负性筛选标志,所以筛选的假阳性率较高,尤其是前者,假阳性率高达 60% 以上^[11],给筛选工作带来很大困难。有鉴于此,我们在二者的基础上,引入负性筛选基因 *PE40*,由于 *PE40* 是绿脓杆菌外毒素的活性成分,表达在细胞浆中会杀死细胞,所以可以最大程度的减少假阳性。此外,我们在所设计的载体中,用 IRES 引入人 *PLAP* 基因,其原因有两点: a. *PLAP* 可以通过特定的底物显色,便于用组织化学的方法观察所克隆基因的表达式,组织分布; b. 人 *PLAP* 蛋白能够耐高温,所以可以用高温灭活内源性碱性磷酸酶,减少显色时的本底。

综上所述,分泌性蛋白在生物体个体发育、生理功能的发挥及各种病理过程的演进中起着重要作用。本研究以克隆分泌蛋白为出发点,借鉴以往克隆分泌蛋白的方法,设计并构建了一种新型基因陷阱报告系统,该系统能够对基因组中分泌蛋白的基因进行直接的筛选,为分泌蛋白基因的克隆和功能研究提供了新方法。

参 考 文 献

- 1 Tashiro K, Tada H, Heilker R, *et al.* Signal sequence trap: a cloning strategy for secreted proteins and type I membrane proteins. *Science*, 1993, **261** (5121): 600 ~ 603
- 2 Tashiro K, Nakamura T, Honjo T. The signal sequence trap method. *Methods Enzymol*, 1999, **303**: 479 ~ 495
- 3 Floss T, Wurst W. Functional genomics by gene-trapping in embryonic stem cells. *Methods Mol Biol*, 2002, **185**: 347 ~ 379
- 4 Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning. A Laboratory Manual*. 2nd. New York: Cold Spring Harbor Laboratory Press, 1989. 16 ~ 74
- 5 Pastan I, Chaudhary V, FitzGerald D J. Recombinant toxins as novel therapeutic agents. *Annu Rev Biochem*, 1992, **61**: 331 ~ 354
- 6 Donnelly M L, Hughes L E, Luke G, *et al.* The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J Gen Virol*, 2001, **82** (Pt 5): 1027 ~ 1041
- 7 孙 强,王冀姝,李 荣,等. *Suc2* 信号肽捕获系统的建立. 遗传学报, 2001, **28** (4): 379 ~ 384
- 8 Sun Q, Wang J S, Li R, *et al.* *Acta Genetica Sinica*, 2001, **28** (4): 379 ~ 384
- 8 Galliciotti G, Schneider H, Wyder L, *et al.* Signal-sequence trap in mammalian and yeast cells: a comparison. *J Membr Biol*, 2001, **183** (3): 175 ~ 182
- 9 Skarnes W C, Moss J E, Hurlley S M, *et al.* Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc Natl Acad Sci USA*, 1995, **92** (14): 6592 ~ 6596
- 10 Gebauer M, von Melchner H, Beckers T. Genomewide trapping of genes that encode secreted and transmembrane proteins repressed by oncogenic signaling. *Genome Res*, 2001, **11** (11): 1871 ~ 1877
- 11 Mitchell K J, Pinson K I, Kelly O G, *et al.* Functional analysis of secreted and trans- membrane proteins critical to mouse development. *Nat Genet*, 2001, **28** (3): 241 ~ 249

Design and Efficacy of a Gene Trap System for Secretory Proteins*

SUN Qiang, HAN Hua**

(Department of Medical Genetics and Developmental Biology, The Fourth Military Medical University,
Stem Cell Research Center of TangDu Hospital, Xi'an 710032, China)

Abstract The secretion of proteins depends on signal peptide. Gene trap is a powerful tool for functional gene identification and characterization. Classical gene trap employs *geo* as the reporter, and gives no discrimination on the types of trapped genes. A new reporter gene named *peo* was obtained by in turn fusing Exotoxin A of *Pseudomonas aeruginosa*, 2A sequence, transmembrane region of IL-2 receptor and neomycin phosphotransferase genes. *Peo* is able to report genes encoding secretory proteins specifically. To prove *peo*'s specificity for secretory proteins, three plasmid vectors, representing three types of sequences possibly trapped by gene trap vector during screening, were constructed based on *peo* reporter. After transfecting HeLa cells with these plasmids, followed by G418 selection and PLAP staining, it is shown that the new vector could distinguish secretory proteins from others, and the gene trap vector taking *peo* as the reporter can trap secretory protein genes efficiently.

Key words gene trap, signal peptide, *peo*, exotoxin A, neomycin phosphotransferase

* This work was supported by a grant from The Special Funds for Major State Basic Research of China (2001CB509900).

** Corresponding author. Tel: 86-29-3374490, E-mail: huahan@fmmu.edu.cn

Received: September 8, 2003 Accepted: October 31, 2003