

# 人 *STK11* 基因启动子区域的克隆\*

宁文锋\*\* 王永俊 姚茂金 汤建光 汪斐 周世权 李陈捷 施小六\*\*\*

中南大学湘雅二医院, 长沙 410011

**摘要** 通过生物信息学分析, 预测人 *STK11* 基因可能的转录起始位点和重要的调控区域. 以此为基础, 构建含人 *STK11* 基因上游序列缺失片段的 pGL3 重组载体, 瞬时转染 L-02 和 HeLa 两种细胞系, 通过检测荧光素活性分析缺失片段的启动活性. 结果显示各缺失片段在两株细胞中均有启动活性; 当 3' 端固定在 -918 时, 5' 端由 -1943—-1425 的移位对启动活性无明显影响, 提示这一区间内不存在重要的顺式作用元件; 将 5' 端固定在 -1943 时, 随着 3' 端向 5' 的缩短 (-918→-1039→-1160), 启动活性明显增强 (活性 pGL3-1943A > pGL3-1943B > pGL3-1943C), 提示 -1160 - -1039 和 -1039—-918 的区间分别存在负性顺式作用元件, 同时提示核心启动子的 3' 端位于 -1160 的 5' 末端. 在 pGL3-1322 (-1322—-918) 亦观察到启动活性, 提示核心启动子的 5' 端应该位于 -1322 的 3' 末端. 因此, 人 *STK11* 基因的核心启动子位于 -1322—-1160 之间.

**关键词** *STK11* 基因 生物信息学 启动子 顺式作用元件 转录起始位点

PJ 综合征 (Peutz-Jeghers syndrome) 是一种常染色体显性遗传疾病. 其疾病基因于 1998 年被克隆<sup>[1,2]</sup>. 基因所编码的蛋白质 STK11 是一种丝氨酸/苏氨酸蛋白激酶. 最近研究显示: STK11 参与了个体胚胎发育及细胞极化等重要生命活动的调控, 并与 P53, Wnt 及 AMPK 等多种信号系统相关<sup>[3]</sup>. 而关于其表达研究的结果表明其表达丰度在个体发育的不同阶段存在组织与时相差异性<sup>[4]</sup>, 提示 STK11 的表达受到复杂而又精密的调控. 在转录水平对基因进行调控在基因表达过程中有着重要的意义, 但目前尚无有关其转录调控研究的报道. 本实验以生物信息学分析为基础, 克隆人 *STK11* 基因上游序列的 5' 和 3' 缺失片段, 与 pGL3 荧光素酶报告载体构成重组载体, 分析双荧光素酶活性来评价各缺失片段的启动子功能.

## 1 材料和方法

### 1.1 材料

荧光素酶报告基因载体 pGL3, 海肾荧光素酶报告基因载体 pRL-TK 及双荧光素酶检测试剂均为 Promega 公司产品. 各种限制性内切酶、T4 连接酶为 TaKaRa 公司产品. Lipofectamine2000 购于 Invitrogen 公司. PRMI1640 培养基、胎牛血清为 GIBCO 公司产品. L-02 (正常人胎肝细胞)、HeLa 细胞系购自武汉大学中国典型培养物保藏中心. 抗 STK11 单克隆抗体 (UPSTATE, clone 5c10), 辣根过氧化物酶标记的抗小鼠 IgG 为 Sigma 公司产品, ECL 试剂盒购自 Pharmacia.

### 1.2 生物信息学分析

(1) 人 *STK11* 基因转录本信息的获得  
主要利用下列数据库:

2005-11-21 收稿, 2006-01-06 收修改稿

\* 国家自然科学基金资助项目 (批准号: 30400537)

\*\* 现工作单位: 南华大学附属第二医院, 湖南衡阳 421001

\*\*\* 通讯作者, E-mail: shixiaoliu61766@yahoo.com.cn

GenBank : <http://www.ncbi.nlm.nih.gov/BLAST/>

Ensemble : <http://www.ensembl.org/Homo-sapiens/>

UCSC <http://genome.ucsc.edu/>

### (2) 人 *STK11* 基因 CpG 岛分析和基因转录起始位点预测分析

以人 *STK11* 基因 5' 端 (-4000—+289, 定义翻译起始密码子 A 为 +1) 为处理序列进行 CpG 岛分析, 所用软件为 MethPrimer: (<http://itsa.ucsf.edu/urolab/methprimer/index1.htm>). 然后对此处理序列再进行转录起始位点预测, 所用预测软件为 Dragon Gene StartFinder (DGSF) ([http://sdmc.lit.org.sg/promoter/dragonGSF1\\_0/gene-start.htm](http://sdmc.lit.org.sg/promoter/dragonGSF1_0/gene-start.htm)). 该程序综合评价有关 CpG 岛、转录起始位点以及预计起始位点下游的信息, 对基因起始作出预测. DGSF 以 CpG 岛作为重要参数在预测过程中, 因此它对于含有 CpG 岛基因的起始预测十分

有效<sup>[5]</sup>.

### (3) *STK11* 启动子区域预测

以人 *STK11* 基因 (-4000--1) 为处理序列. 预测软件 1 为 CONSITE (<http://www.phylo-foot.org/>), 该软件以物种间基因序列和位置的保守性为基础进行启动子区域预测, 大大提高了预测的准确性<sup>[6]</sup>. 预测软件 2 为 PromoterInspector (<http://www.genomatix.de/cgi-bin/eldorado/main>), 特异性可高达 85%, 但敏感性为 50%<sup>[7]</sup>.

### 1.3 质粒构建

综合生物信息学分析结果, 构建 5' 和 3' 缺失片段作为候选启动子区域, 以人 *STK11* 基因 (NT\_086894.1) 序列为参照设计引物 (所用引物见表 1). 以正常人基因组 DNA 为模板, 用 PCR 法扩增各片段, 各片段用 *Kpn* I, *Xho* I 双酶切后连入到 pGL3-Basic 载体. 各重组质粒经酶切和测序证实. 各重组质粒中所插入的片段见图 1.

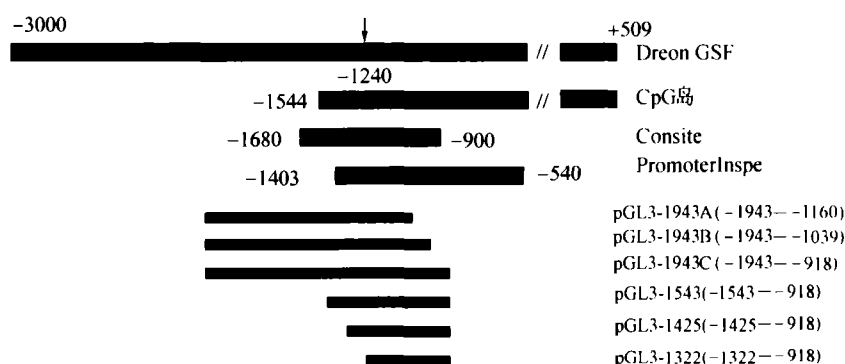


图 1 人 *STK11* 基因生物信息学分析及所构建的 5' 和 3' 缺失片段

箭头所指为预测的转录起始位点 -1240; CpG 岛预测区域位于 -1544—+509 区间; Consite 预测启动子区域为 -1680—-900; PromoterInspector 预测启动子区域位于 -1403—-540

### 1.4 细胞和细胞培养

L-02 和 HeLa 细胞采用含 10% 胎牛血清的 RPM1640 培养液, 置于 37°C 的 CO<sub>2</sub> 培养箱中培养.

### 1.5 转染细胞系内源性 *STK11* 蛋白表达检测 (Western blotting)

提取细胞总蛋白质. 取 20 μg 总蛋白行 10% SDS-PAGE 并将蛋白转至 PVDF 膜, 一抗为抗 *STK11* 单克隆抗体 (1:1000), 二抗为辣根过氧化物酶标记的羊抗小鼠抗体 (1:30000), 最后通过

ECL 试剂盒显色.

### 1.6 转染

L-02 及 HeLa 细胞转染前培养于 24 孔板, 生长密度约 90% 时进行 Lipofectamine2000 协助转染, 转染具体操作参考 Lipofectamine2000 操作说明. 分别以转染 pGL3-Basic, pGL3-Promoter 为阴性和阳性对照, 以转染海肾荧光素酶报告基因载体 pRL-TK 作为内对照. 转染 48 h 后进行双荧光素酶活性测定. 每种质粒均做复孔转染, 在每种细胞株

中至少重复3次转染及活性测定实验。

### 1.7 数据处理及活性分析

各质粒的相对荧光素酶活性(倍)的计算公式为:(实验片段荧光素酶活性/共转的海肾荧光素酶活性)/(pGL3-Basic 荧光素酶活性/共转染的海肾荧光素酶活性),相对荧光素酶活性数值均以均数±标准差( $X \pm S$ )表示。

表1 载体构建所用引物

引物	碱基序列(5'→3')
R-a(-1160)	CCGCTCGAG GGACACCGCACGCC
R-b(-1039)	CCGCTCGAG CACCCGCGCTCCCTTC
R-C(-918)	CCGCTCGAGGGTTCCCGAAGGTGCCG
F-1943	GG GTACC ACTGGGGAGGCTGAGGCAGGAG
F-1543	GGGGTACCGCTCCTTCCAGGTCCGCAAG
F-1425	GGGGTACCTCACAGCCGTGGCCCTCGTCTC
F-1322	GGGGTACCCCAAGCGCCGACCAATCGC

## 2 结果

### 2.1 生物信息学分析结果

#### (1) 人 *STK11* 基因转录本信息

在 GenBank, Emsemble 以及 UCSC 等数据库中进行检索以及核酸序列比对所获得的 *STK11* 的 cDNA 克隆 AF217978 其 5' 末端为最长, 其 5' UTR 区为 961 bp. 而获得的拼接序列 RefSeq Genes NM\_000455.4, 5' UTR 区有 1115 bp.

#### (2) CpG 岛分析结果

MethPrimer 分析结果显示: 人 *STK11* 第一外显子及上游富含 CpG 岛, 位于 -1544—+509 区间(图 1).

#### (3) 转录起始位点软件预测结果

DGSF 预测人 *STK11* 基因的转录起始位点在 -1240(图 1).

#### (4) 启动子区域预测结果

PromoterInspector 分析结果显示: 人 *STK11* 启动子区域位于 -1403—+540(图 1). Consite 分析结果为 -1680—+900(图 1).

### 2.2 pGL3 荧光素酶报告基因载体鉴定结果

构建的 pGL3-1322, pGL3-1425, pGL3-1543, pGL3-1943A, pGL3-1943B, pGL3-1943C 用 *Kpn*

I 和 *Xho* I 双酶切后 1% 琼脂糖凝胶电泳结果(图 2)显示: 各重组质粒中插入的目的片段大小与理想的片段大小完全符合, 进一步测序证明各重组质粒中插入的目的片段无突变发生, 表明载体构建成功。

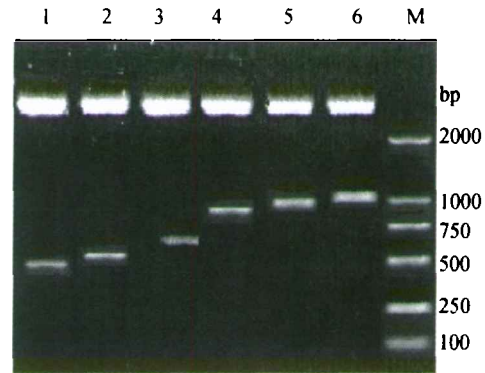


图2 pGL3 荧光素酶报告基因重组载体酶切鉴定结果  
M 为 DL2000 DNA 分子量标准; 1-6 分别代表 pGL3-1322, pGL3-1425, pGL3-1543, pGL3-1943A, pGL3-1943B, pGL3-1943C

### 2.3 转染细胞系内源性 *STK11* 蛋白表达检测 (Western blotting) 结果

L-02 和 HeLa 两细胞系内源性 *STK11* 表达的 Western blotting 结果(图 3)显示: L-02 细胞有内源性 *STK11* 表达; HeLa 细胞无内源性 *STK11* 表达。

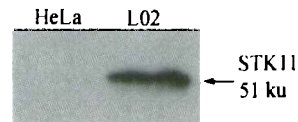


图3 L-02, HeLa 细胞中内源性 *STK11* 的表达

### 2.4 *STK11* 上游各片段双荧光素酶活性分析

各片段的荧光素酶相对活性结果显示(图 4): 各片段在 L-02 和 HeLa 细胞系中都具有荧光素酶活性. 在 L-02 细胞系中, pGL3-1322, pGL3-1425, pGL3-1543, pGL3-1943A, pGL3-1943B, pGL3-1943C 重组载体分别为 pGL3-Basic 的 25, 40, 50, 150, 80, 50 倍; 在 HeLa 细胞系中, 6 个重组载体分别为 pGL3-Basic 的 15, 20, 25, 140, 70, 25 倍。

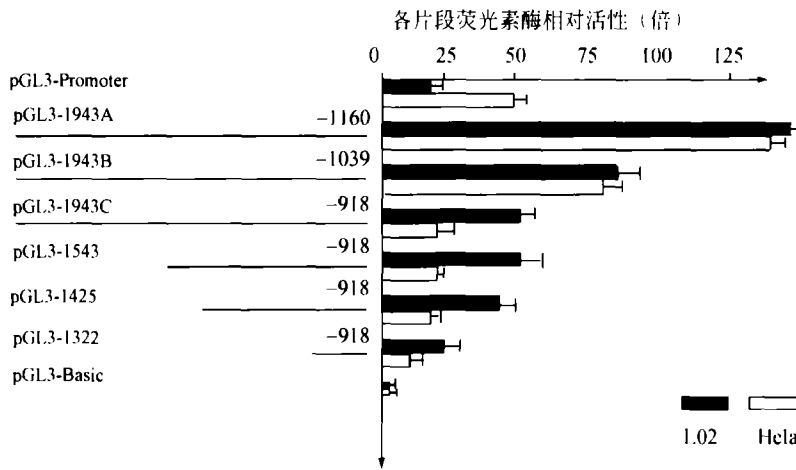


图 4 STK11 上游序列各片段的荧光素酶相对活性

### 3 讨论

人 *STK11* 基因位于 19p13.3, 总跨度约为 23 kb, 包含 10 个外显子, 其中 1 个外显子为可选择性内含子<sup>[2]</sup>. 基因所编码的蛋白质 STK11 是一种丝氨酸/苏氨酸蛋白激酶, 它含 433 个氨基酸残基. STK11 在染色质重构、细胞周期阻滞、细胞极性、Wnt 信号通道以及能量代谢中发挥重要作用<sup>[3]</sup>. STK11 的表达丰度在个体发育的不同阶段存在组织与时相差异性<sup>[4]</sup>, 提示 STK11 的表达受到复杂而精密的调控. 转录调控的研究对其功能的深入研究有着重要意义.

确定转录起始位点能为寻找启动子区段提供锚点, 故而对于基因的转录调控研究具有至关重要的意义. 目前用于确定转录起始位点的实验方法有引物延伸、RNA 酶保护以及 5'RACE (Rapid amplification of cDNA ends) 等, 但均耗时耗力, 生物信息学分析显示 *STK11* 启动子为富含 G/C 启动子, 现有的确定转录起始位点的实验方法难于定位其转录起始位点<sup>[8]</sup>. 可喜的是目前网上数据库资源丰富, 在获得基因的转录本后, 如能充分利用网上数据库资源寻找尽可能长的 5'端 cDNA 序列, 将大大缩减确定转录起始位点的实验工作量. 作者通过在 GenBank, Emsemble 以及 UCSC 等数据库中进行检索以及核酸序列比对所获得的 *STK11* 的 5'端最长的 cDNA 克隆为 AF217978, 其 5'末端位于 *STK11* 翻译起始位点的 -961 处. 但 RefSeq 数据资源查询结

果显示 *STK11* 的最长 mRNA 拼接序列的 5'末端位于 -1115 处. 富含 G/C 的不典型启动子其基因存在多个转录起始位点, 其间跨度可达数百个碱基<sup>[8]</sup>, 而 *STK11* 的启动子恰好属于这类不典型启动子, 所以数据库搜寻发现 5'末端最长 cDNA 克隆和 mRNA 拼接序列结果不符合的可能原因有: (1) 两者的 5'末端可能均已接近实际的转录起始位点 (即 *STK11* 在 -1115 和 -961 附近各存在一个转录起始位点); (2) -1115 附近是实际的转录起始位点. 软件 DGSF 预测人 *STK11* 基因的转录起始位点在 -1240 处. 综合数据库搜寻结果、软件预测结果以及 PCR 引物设计的合理性, 我们将 *STK11* 基因上游序列缺失片段 3'端定在 -918, -1039, -1160 3 处. 转录调控研究的一个常用策略是利用网上在线软件分析并获得候选启动子区段. 作者利用 PromoterInspector 分析结果显示: 人 *STK11* 启动子区域位于 -1403—-540, Consite 分析结果为 -1680—-900, 提示在 -1680 bp—-540 bp 区间可能存在启动子片段. 综合两软件预测结果以及 PCR 引物设计的合理性, 我们将 *STK11* 基因上游序列缺失片段 5'定在 -1943, -1534, -1425, 1322 4 处, 并结合前面确定的 3'端, 我们构建了 6 个候选启动子片段的 pGL3 重组载体.

在 3'缺失片的实验分析中, pGL3-1943A (-1943—-1160), pGL3-1943B (-1943—-1039), pGL3-1943C (-1943—-918) 的 5'端相同, 差异仅在 3'端, 它们的活性分别为 pGL3-Basic 的 150,

80, 50 倍(L-02 细胞系中)。3'端缺失最长的 pGL3-1943A 仍具有荧光素酶活性, 说明人 *STK11* 基因的转录起始位点在-1160 的 5'端。相应的, 其核心启动子区域亦在-1160 的 5'端。在 5'缺失片段的实验分析中, pGL3-1943C(-1943--918), pGL3-1543(-1543--918), pGL3-1425(-1425--918), pGL3-1322(-1322--918)的活性分别为 pGL3-Basic 的 50, 50, 40, 25 倍(L-02 细胞系中), 均具有一定的启动活性, 尤其是 5'端缺失最长的 pGL3-1322 仍具有荧光素酶活性, 说明 *STK11* 基因的核心启动子区域在-1322 的 3'端。结合 5'端和 3'端缺失片段的启动活性结果, 可以初步将 *STK11* 基因的核心启动子区域界定在-1322--1160 之间。

将 6 个片段的启动活性作相邻之间的比较, 可见 pGL3-1943C, pGL3-1543, pGL3-1425 3 者活性差异不明显, 提示(-1943--1425)区间内可能不存在重要的调控作用元件; pGL3-1943A(-1943--1160), pGL3-1943B(-1943--1039), pGL3-1943C(-1943--918)的活性分别为 pGL3-Basic 的 150, 80, 50 倍(L-02 细胞系中), 3 者活性差异明显, 在 HeLa 细胞系中存在相同的趋势, 提示-1160--1039 和-1039--918 的两个区间内都含有负性顺式作用元件。

用 Western blotting 方法在 L-02 细胞中可检测到 *STK11* 的表达, 在 HeLa 细胞中未见表达, 与以往文献报道结果一致<sup>[9]</sup>, 故我们选择 *STK11* 表达有差异的这两株细胞系进行转染实验, 以期发现细胞特异性转录因子。结果却发现不仅在 L-02 细胞中, 所有 6 个片段均具有不同程度的启动活性, 在 HeLa 细胞中, 所有 6 个片段均同样具有不同程度的启动活性, 其中原因推测可能如下: 基因的转录不但受核心启动子及其附近近端调控元件的影响,

还与远端调控元件的调控作用以及启动子区甲基化有关, 已有研究报道提到启动子区域的高甲基化就是导致 HeLa 细胞中 *STK11* 低表达的重要原因<sup>[9]</sup>。

综上所述, 在人 *STK11* 基因-1322--1160 区间内包含有核心启动子, 在-1160--1039 和-1039--918 的区间内存在有重要的顺式作用元件, 而且为负性顺式作用元件, 这一发现为 *STK11* 基因转录调控的深入研究奠定了基础。

### 参 考 文 献

- 1 Hemminki A, Markie D, Tomlinson I, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature*, 1998, 391: 184-187
- 2 Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nature Genetics*, 1998, 18: 38-43
- 3 Hardie DG. New roles for the LKB1 → AMPK pathway. *Curr Opin Cell Biol*, 2005, 17: 167-173
- 4 Luukko K, Ylikorkala A, Tiainen M, et al. Expression of LKB1 and PTEN tumor suppressor genes during mouse embryonic development. *Mech Dev*, 1999, 83: 187-190
- 5 Bajic V B, Seah S H. Dragon gene start finder: An advanced system for finding approximate locations of the start of gene transcriptional units. *Genome Res*, 2003, 13: 1923-1929
- 6 Sandelin A, Wasserman W W, Lenhard B. ConSite: Web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res*, 2004, 32(Web Server issue): W249-52
- 7 Scherf M, Klingenhoff A, Werner T. Highly specific localization of promoter regions in large genomic sequences by Promoter-Inspector: A novel context analysis approach. *J Mol Biol*, 2000, 297: 599-606
- 8 Carey M, Smale S T. *Transcriptional Regulation in Eukaryotes Concepts Strategies, and Techniques*. New York: Cold Spring Harbor Laboratory Press, 2000. 101
- 9 Tiainen M, Ylikorkala A, Makela T P. Growth suppression by Lkb1 is mediated by a G(1) cell cycle arrest. *Proc Natl Acad Sci USA*, 1999, 96: 9248-9251