



启动子克隆

基本概念

顺式作用元件(cis-acting element): 同一 DNA 分子中具有特殊功能的转录因子 DNA 结合位点和其它调控基序，如启动子、增强子及沉默子等。

启动子(Promoter): 能确保转录精确而有效起始的 DNA 序列，即 DNA 模板上专一地与 RNA 聚合酶结合并决定转录从何处起始的部位，也决定基因的转录效率。

反式作用因子(trans-acting factor): 参与基因表达调控的因子,它们与特异的靶基因的顺式元件结合起作用。编码反式作用因子的基因与被反式作用因子调控的靶序列(基因)不在同一染色体上。

转录因子(Transcription factors): 能够以序列特异性方式结合 DNA 并且调节转录的蛋白质。

启动子的分类

按生物种类分类

原核启动子 真核启动子

按转录模式分类

组成型启动子 组织特异性启动子
诱导型启动子

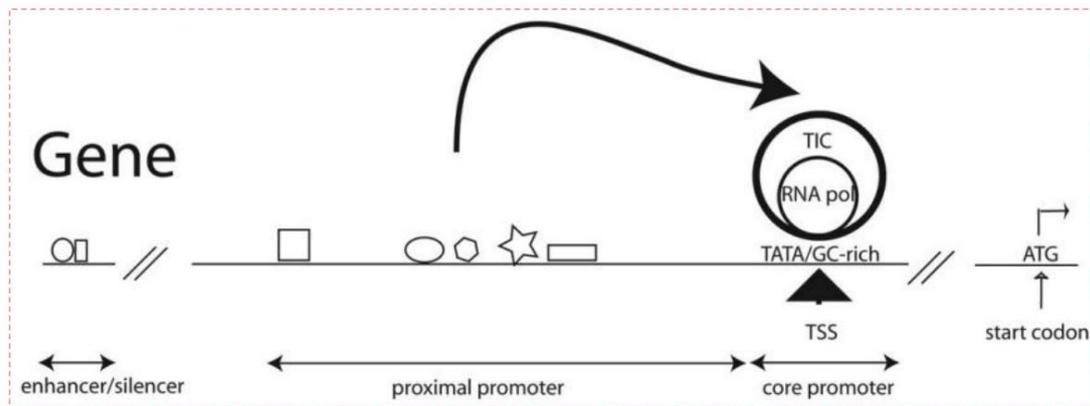
按启动效率分类

强启动子 弱启动子

按结构模型分类

polI 型启动子: rRNA
polIII 型启动子: mRNA
polIII 型启动子: tRNA

真核启动子的结构



enhancer: 增强子，顺势远端调控元件，促进远距离的靶基因转录。

silencer: 沉默子，顺势远端调控元件，阻遏远距离的靶基因转录。

core promoter: 包含转录起始位点以及该位点上下游的一部分序列，大约上下游各 50nt 左右。一般核心启动子只有大约 40nt。

proximal promoter: 近端启动子。紧接核心启动子上游数百个碱基对的区域，通常包含多个 TFs 结合位点。

TSS: 转录起始位点。即转录时，RNA 链第一个核苷酸相对应 DNA 链上的碱基，通常为一个嘌呤。

启动子的特点

- 1、结构复杂：通常含一些特定的元件，如 TATA box
- 2、具有方向性：启动下游的基因的表达启动子长度不确定性
- 3、一个基因通常含有多个启动子，启动不同转录本表达
- 4、某些启动子具有表达的时空性和组织特异性

The punctilious RNA polymerase II core promoter

Table 1. Consensus sequences of some core promoter elements

| Motif | Location | Consensus |
|------------------|--------------------------|--|
| TATA box | Upstream T at -32 to -28 | TATAWR |
| BRE ^u | Upstream of TATA box | SSRCGCC |
| BRE ^d | -23 to -17 | RTDKKKK |
| Inr | -2 to +4 | TCA ₁ GTY (<i>Drosophila</i>) |
| | -3 to +3 | BBCA ₁ BW (human) |
| TCT | -2 to +6 | YCA ₁ TTTTY (<i>Drosophila</i>) |
| | -1 to +6 | YC ₁ TYTTY (human) |
| XCPE1 | -8 to +2 | DSGYGGRAS ₁ M |
| XCPE2 | -9 to +2 | VCYCRTRRCM ₁ Y |
| MTE | +18 to +22 | CGANC |
| | +27 to +29 | CGG |
| DPE | +28 to +32 | RGWYV |
| DCE | Box I: +6 to +11 | CTTC |
| | Box II: +16 to +21 | CTGT |
| | Box III: +30 to +34 | AGC |
| DTIE | +23 to +31 | GSGRDNHGG |

[W] A or T; [R] A or G; [S] G or C; [D] A, G, or T (not C); [K] G or T; [Y] C or T; [B] C, G, or T (not A); [M] A or C; [V] A, C, or G (not T); [N] A, C, G, or T (any base); [H] A, C, or T (not G). (BRE) TFIIB recognition element upstream (u) or downstream (d); (Inr) initiator; (XCPE1) X core promoter element 1; (MTE) motif ten element; (DPE) downstream core promoter element; (DCE) downstream core element; (DTIE) downstream transcription initiation element.

启动子的研究思路

- 启动子结构研究，包括核心启动子区域、正调控区域、负调控区域及增强子的确定
- 组织特异性启动子筛选及确定
- 顺式作用元件与反式作用因子相互作用研究：转录因子结合位点，SNP（单核苷酸多态性）等
- 启动子甲基化作用研究
- 新顺式作用元件发现及功能鉴定

启动子的研究方法

- 生物信息学分析：启动子序列、转录因子结合位点等
- 启动子克隆及活性检测：荧光素酶报告分析
- 启动子与蛋白质相互作用分析
 - ① 酵母单杂交 (Y1H) 技术
 - ② 染色质免疫共沉淀 (ChIP) 技术
 - ③ 凝胶阻滞分析 (EMSA) 试验
 - ④ DNA 足迹 (DNase I footprinting) 分析法
 - ⑤ 启动子芯片 ...

启动子序列的查找

- 打开 <http://genome.ucsc.edu/>,
- 选择 genomes
- 选择 Human, 在 Position/Search Term 中输入所需查找的基因名称
- 点击 Tools 中 Table Browser
- 在 track 中选择 NCBI RefSeq, output format 选择 sequence
- 点击 get output, 选择 genomic 后 submit

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, and the [User's Guide](#) for general information and sample queries. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38)

group: Genes and Gene Predictions track: NCBI RefSeq add custom tracks track hubs

table: RefSeq All (ncbiRefSeq) describe table schema

region: genome position chr5:14,704,800-14,871,785 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

subtrack merge: create

intersection: create

correlation: create

output format: sequence Send output to Galaxy GREAT GenomeSpace

output file: out (leave blank to keep output in browser)

file type returned: plain text gzip compressed

get output summary/statistics

To reset all user cart settings (including custom tracks), [click here](#).

Using the Table Browser

选择

- Promoter/Up stream by 2000 bases
- 5'UTR
- Exons in upper cases, everything else in low case
- get sequence

ncbiRefSeq Genomic Sequence

Sequence Retrieval Region Options:

- Promoter/Upstream by 2000 bases
- 5' UTR Exons
- CDS Exons
- 3' UTR Exons
- Introns
- Downstream by 1000 bases

One FASTA record per gene.

One FASTA record per region (exon, intron, etc.) with 0 extra bases upstream (5') and 0 extra downstream (3')

Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

- Exons in upper case, everything else in lower case.
- CDS in upper case, UTR in lower case.
- All upper case.
- All lower case.
- Mask repeats: to lower case to N

get sequence cancel

在下载文件中，找到对应的转录本序列。

- 小写碱基为 promoter 上游 2000bp。
- 大写碱基是 5'UTR，第一个大写碱基是转录起始位点。
- 大写碱基取前面 4 行即为 promoter 下游 200bp。

启动子报告克隆

GLuc-ON™ 启动子报告克隆在 GLuc 报告基因上游插入一段 1.2~1.5kb 的序列,这段插入序列与基因转录起始位点 (TSS) 上游大约 1.5kb 到下游 200bp 之间的 5'侧翼序列一致, 可以通过观察荧光素酶的活性来研究启动子对基因的调节作用。

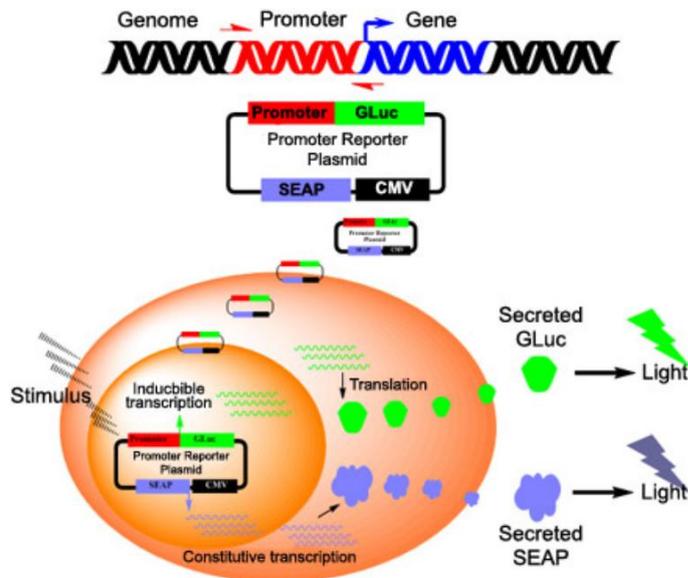


图1. GLuc-ON™ 启动子报告克隆的双报告基因检测原理图。

启动子报告克隆载体

载体类型

| 载体 | 报告基因 | 内参基因 | 筛选标记 | 载体类型 |
|-------------|---------------------------------|--------------------|-----------|------|
| pEZX-PG04 | <i>Gussia luciferase (GLuc)</i> | 分泌型碱性磷酸酶(SEAP) | Puromycin | 非病毒 |
| pEZX-PG02 | <i>Gussia luciferase (GLuc)</i> | N/A* | Puromycin | 非病毒 |
| pEZX-PF02 | <i>eGFP</i> | N/A* | Puromycin | 非病毒 |
| pEZX-PM02 | <i>mCherry</i> | N/A* | Puromycin | 非病毒 |
| pEZX-PL01 | Firefly luciferase | Renilla luciferase | Puromycin | 非病毒 |
| pEZX-LvPG04 | <i>Gussia luciferase (GLuc)</i> | 分泌型碱性磷酸酶((SEAP) | Puromycin | 慢病毒 |
| pEZX-LvPG02 | <i>Gussia luciferase (GLuc)</i> | N/A* | Puromycin | 慢病毒 |
| pEZX-LvPF02 | <i>eGFP</i> | N/A* | Puromycin | 慢病毒 |
| pEZX-LvPM02 | <i>mCherry</i> | N/A* | Puromycin | 慢病毒 |
| pEZX-LvPM03 | tdTomato | N/A* | Puromycin | 慢病毒 |
| pEZX-LvPL01 | Firefly luciferase | Renilla luciferase | Puromycin | 慢病毒 |

* SEAP可以在分泌型载体中表达。

启动子报告克隆 FAQs

Q: 我在你们公司购买了启动子克隆，但是测序的结果有几处突变，你们不保证碱基序列一致吗？会不会影响我的验证结果？

A: 我司的启动子克隆选择的是 Hela 细胞的基因组作为模板，由于 SNP 的存在，因此我们获得的启动子序列会出现不一致的情况，在大部分情况下 SNP 的存在如果是在非核心区域，并不影响启动子的功能验证。如果您需要碱基序列完全一致的可以选择基因合成服务，但在购买前应协商好。

Q: 启动子克隆的核心功能片段在哪？

A: 不同启动子的核心功能片段位置有不同，主要是 RNA 聚合酶结合位点，转录起始位点等，具体位置需要自行预测验证。我司也可提供有偿的预测服务。

Q: 我司如何预测基因的启动子序列？

A: 我们公司预测启动子的方法，是通过研究了多个公共的启动子数据库，并且进行了验证实验之后确定的。我们一般情况下，选定的启动子序列，从转录起始位点上游 1200bp 到下游 100-200bp。当然也有部分特殊的启动子序列采用文献报道。