

• 生物技术与方法 •

肌肉特异性合成启动子文库构建及其高活性启动子元件组成与活性的关联性分析

王泽宁¹, 江明锋¹, 道杰日庆¹, 曲久², 李小伟³, 旦巴次仁², 刘益丽^{1*}

1 西南民族大学 畜牧兽医学院, 四川 成都 610041

2 那曲市农业农村局, 西藏 那曲 852000

3 四川省龙日种畜场, 四川 红原 624400

王泽宁, 江明锋, 道杰日庆, 曲久, 李小伟, 旦巴次仁, 刘益丽. 肌肉特异性合成启动子文库构建及其高活性启动子元件组成与活性的关联性分析[J]. 生物工程学报, 2024, 40(12): 4616-4627.

WANG Zening, JIANG Mingfeng, Daojieriqing, QU Jiu, LI Xiaowei, Danbaciren, LIU Yili. Construction of a muscle-specific synthetic promoter library and correlation analysis of the element composition and activity of highly active promoters[J]. Chinese Journal of Biotechnology, 2024, 40(12): 4616-4627.

摘要: 本研究旨在构建肌肉特异性合成启动子库, 筛选出活性较高的肌肉特异性启动子, 分析高活性启动子序列中启动子元件与其活性的关系, 为人工合成启动子提供一定理论依据。选取 19 个源于肌肉特异性元件、保守元件、病毒调控序列的元件, 随机连接构建肌肉特异性合成启动子库。构建荧光素酶质粒 pCMV-Luc 和 pSPs-Luc, 将以上质粒转染至成肌细胞系 C2C12, 通过荧光素酶活性检测评估合成启动子活性; 利用另外 2 种非肌肉源细胞系 HeLa 和 3T3 验证高活性启动子的肌肉特异性, 分析活性高、肌肉特异性好且测序比对正确的启动子序列, 探索元件组成与启动子活性之间的关系。本研究成功构建肌肉特异性启动子库, 筛选出有效合成启动子质粒 321 个, 其中 SP-301 启动子活性为 CMV 活性的 5.63 倍; 活性较高的 15 个启动子具有肌肉特异性; 分析高活性且测序正确的启动子, 发现元件组成与启动子活性之间存在一定关系。肌肉特异性元件在启动子中占比偏高, 但与启动子活性强弱相关性较小, 是组织特异性的决定元件; 病毒元件在强活性启动子中含量均不低于 20%, 可能是影响启动子活性的关键元件; 保守元件含量与启动子活性之间具有显著的相关性, 其含量与启动子活性成正比。本研究为合成组织特异性高效启动子提供了一定理论基础, 为原位基因传递体系的构建和应用提供了新思路。

关键词: 合成启动子; 筛选; 相关性分析; 启动子元件

资助项目: 那曲市科技局重点研发计划(NQKJ-2023-07); 西南民族大学中央高校基本科研业务费专项资金(ZYN2024062); 四川科技计划(2023YFQ0076); 四川省自然科学基金(25QNJJ1998)

This work was supported by the Naqu City Science and Technology Bureau Key Research and Development Project (NQKJ-2023-07), the Fundamental Research Funds for the Central Universities, the Southwest Minzu University (ZYN2024062), the Sichuan Science and Technology Program (2023YFQ0076), and the Natural Science Foundation of Sichuan Province (25QNJJ1998).

*Corresponding author. E-mail: yililium@163.com

Received: 2024-05-27; Accepted: 2024-11-07; Published online: 2024-11-08

Construction of a muscle-specific synthetic promoter library and correlation analysis of the element composition and activity of highly active promoters

WANG Zening¹, JIANG Mingfeng¹, Daojieriqing¹, QU Jiu², LI Xiaowei³, Danbaciren², LIU Yili^{1*}

1 College of Animal & Veterinary Sciences, Southwest Minzu University, Chengdu 610041, Sichuan, China

2 Naqu Bureau of Agriculture and Rural Affairs, Naqu 852000, Xizang, China

3 Sichuan Longri Breeding Farm, Hongyuan 624400, Sichuan, China

Abstract: The purpose of this study is to construct a muscle-specific synthetic promoter library, screen out muscle-specific promoters with high activity, analyze the relationship between element composition and activity of highly active promoters, and provide a theoretical basis for artificial synthesis of promoters. In this study, 19 promoter fragments derived from muscle-specific elements, conserved elements, and viral regulatory sequences were selected and randomly connected to construct a muscle-specific synthetic promoter library. The luciferase plasmids pCMV-Luc and pSPs-Luc were constructed and transfected into the myoblast cell line C2C12. The activities of the synthesized promoters were evaluated by the luciferase activity assay. Two non-muscle-derived cell lines HeLa and 3T3 were used to verify the muscle specificity of the highly active promoters. The sequences of promoters with high activity, good muscle specificity, and correct sequences were analyzed to explore the relationship between the element composition and activity of promoters. We successfully constructed a muscle-specific promoter library and screened out 321 effective synthetic promoter plasmids. Among them, the activity of SP-301 promoter was 5.63 times that of CMV. The 15 promoters with high activity were muscle-specific. In the promoters with high activity and correct sequences, there was a relationship between their element composition and activity. Muscle-specific elements accounted for a high proportion in the promoters, while they had weak correlations with the promoter activity, being tissue-specific determinants. Viral elements accounted for no less than 20% in highly active promoters, which may be the key elements for the promoter activity. The content of conserved elements was proportional to the promoter activity. This study lays a theoretical foundation for the synthesis of tissue-specific efficient promoters and provides a new idea for the construction and application of *in-situ* gene delivery systems.

Keywords: synthetic promoter; screening; correlation analysis; promoter elements

启动子位于基因 5'端上游的非编码区域, 是一段能被 RNA 聚合酶识别并结合的特异性 DNA 序列, 具有转录起始功能, 是调控基因转录的核心区域^[1]。启动子由多个重要的短核苷

酸序列组成, 主要分为核心区和上游调控区这 2 个部分。核心启动子区又称近端启动子区, 位于转录起点附近 35 bp 左右, 是一段利用 RNA 酶作用机制精确指导转录起始所必需的

最短连续 DNA 序列。此区域由一些非常重要的功能元件组成,能很准确地定位转录起点和方向,是基因调控过程的关键区域^[2]。核心启动子区主要包括几种最常见的核心区元件,即 TATA box、BRE 元件(TFIIB recognition element, BRE)^[3-4]、起始子(initiator, Inr)^[5-6]、下游启动子元件(downstream core promoter element, DPE)^[7-8]、DCE (downstream core element, DCE) 及 MTE 元件(motif ten element, MTE)^[9]。上游启动子区,又名远端启动子区,是指位于距离转录起始位点几百或上千个核苷酸不等的上游区域,该区域元件通常与相应的蛋白因子结合来改变转录效率^[10],由于每个基因的启动子元件和位置均不相同,故不同的基因表达调控不同。上游启动子区通常包括位于-70 bp 附近的 CAAT box、GC box 及上游增强子等^[2,11]。CAAT box 对基因转录有强激活作用,能控制转录起始的频率和强度。CAAT box 存在于 TATA box 上游,可使基因表达量显著提高^[12]。

随着合成生物学的快速发展,合成的功能元件在基因工程代谢领域表现出较大的应用潜力^[13]。启动子是基因表达过程中控制转录水平的重要元件,为显著提升基因表达量,可通过人工合成启动子文库对基因工程的代谢途径进行精细调控,使代谢途径最优化^[14-15]。本研究以来源于肌肉特异性元件、保守元件、病毒元件的 19 个启动子片段随机组合构建肌肉特异性合成启动子库,进一步构建荧光素酶质粒 pCMV-Luc 和 pSPs-Luc,将以上质粒转染至 C2C12 细胞,通过检测荧光素酶活性评估启动子活性高低,并进一步利用非肌肉源细胞系 HeLa 和 3T3 细胞验证高活性启动子的肌肉特异性。本研究将其中特异性好、活性高且测序比对正确的启动子进行分析,比对启动子元件组成,分析启动子元件组成与其活性之间的相关

性以期合成新型高效启动子提供理论基础,并为基因治疗中靶基因在肌肉组织的高效特异性表达提供更有效的方案。

1 材料与方法

1.1 材料

1.1.1 质粒和细胞

19 种启动子元件由北京擎科新业生物技术有限公司合成;pCR[®]-Blunt II-TOPO[®]质粒购自 Invitrogen 公司;pGL3-Basic 质粒系实验室保存;大肠杆菌(*Escherichia coli*) DH5 α 感受态细胞购自 TaKaRa 公司;C2C12 小鼠成肌细胞、HeLa 人宫颈癌细胞、NIH/3T3 小鼠胚胎成纤维细胞购自中国科学院上海生命科学研究院细胞资源中心。

1.1.2 主要试剂

Kpn I、*Xho* I、*Mlu* I 和 *Nhe* I 限制性内切酶购自 Fermentas 公司;Zero Blunt[®] TOPO PCR Cloning Kit、T4 polynucleotide kinase、T4 DNA ligase 购自 Invitrogen 公司;DL2000 DNA marker 购自 TaKaRa 公司;CHROMA SPIN-1000 Columns 购自 Clontech 公司;琼脂糖购自 Biowest 公司;E.Z.N.A.[™] Plasmid Mini Kit I 购自 Omega 公司;Pierce[™] BCA Protein Assay Kit 购自 Thermo 公司;Luciferase Assay Kit 购自 Promega 公司。

1.2 方法

1.2.1 启动子元件磷酸化与退火

根据文献[16-21]选取 19 个启动子元件,包括 8 个肌肉特异性元件、5 个保守元件和 6 个病毒元件,序列如表 1 所示。将合成的 19 对启动子元件单链进行磷酸化和退火,反应体系如下:T4 polynucleotide kinase buffer (10 \times) 3 μ L, 每对单链等比例混合(100 μ mol/L) 20 μ L, ATP 1 mmol/L, T4 polynucleotide kinase 30 U。反应条件为:37 $^{\circ}$ C 30 min, 94 $^{\circ}$ C 5 min。

表 1 启动子元件序列

Table 1 Promoter elements

Source	No.	Element	Sequence
Muscle specific element (ME)	1	SRE	F: 5'-GACACCCAAATATGGCGACGG-3' R: 3'-CTGTGGGTTTATAACCGCTGCC-5'
	2	MEF-1	F: 5'-CCAACACCTGCTGCCTGCC-3' R: 3'-GGTTGTGGACGACGGACGG-5'
	3	MEF-2	F: 5'-CGCTCTAAAATAACTCCC-3' R: 3'-GCGAGATTTTTATTGAGGG-5'
	4	MEF-3	F: 5'-GGCTTCAGGTTTCCCTA-3' R: 3'-CCGAAGTCCAAAGGGAT-5'
	5	TEF-1	F: 5'-CACCATTCTCAC-3' R: 3'-GTGGTAAGGAGTG-5'
	6	Trex	F: 5'-GGACACCCGAGATGCCTGGTTA-3' R: 3'-CCTGTGGGCTCTACGGACCAAT-5'
	7	MPEX	F: 5'-GGGCCCCTCCCTGGGGACAGCCCC-3' R: 3'-CCCGGGGAGGGACCCCTGTCGGGG-5'
	8	MAPF	F: 5'-ATATGGCGACGGCC-3' R: 3'-TATACCGCTGCCGG-5'
Conserved element (CE)	9	AP-1	F: 5'-TGCTGACTTGGT-3' R: 3'-ACGACTGAACCA-5'
	10	AP-2	F: 5'-CAAGGCCTGGGGACA-3' R: 3'-GTTCCGGACCCCTGT-5'
	11	GATA-2	F: 5'-ACCCGATATGCCTGG-3' R: 3'-TGGGCTCTACGGACC-5'
	12	TATA box	F: 5'-GGACTATATAAAAAACCTG-3' R: 3'-CCTGATATATTTTTGGAC-5'
	13	CAAT box	F: 5'-CCAGCCAATAGC-3' R: 3'-GGTCGGTTATCG-5'
Virus element (VE)	14	72 bp SV40	F: 5'-GGTGTGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCA AAGCATGCATCTCAATTAGTCAGCAACCA-3' R: 3'-CCACACCTTTCAGGGGTCCGAGGGGTCGTCCTTCATACGTT TCGTACGTAGAGTTAATCAGTCGTTGGT-5'
	15	16 bp in the CMV promoter	F: 5'-CTTGGCAGTACATCAA-3' R: 3'-GAACCGTCATGTAGTT-5'
	16	18 bp repeat of CMV promoter	F: 5'-ACTAACGGGACTTTCCAA-3' R: 3'-TGATTGCCCTGAAAGGTT-5'
	17	19 bp in the CMV promoter	F: 5'-CCCCATTGACGTCAATGGG-3' R: 3'-GGGGCTAACTGCAGTTACCC-5'
	18	19 bp repeat of CMV enhancer	F: 5'-GCCCATTGACGTCAATAAT-3' R: 3'-CGGGTAACTGCAGTTATTA-5'
	19	21 bp repeat of CMV promoter	F: 5'-ACGGTAAATGGCCCGCCTGGC-3' R: 3'-TGCCATTTACCGGGCGGACCG-5'

1.2.2 磷酸化启动子元件的连接

将上述 19 种磷酸化启动子片段等比例混合,以 T4 DNA 连接酶连接,16 °C 金属浴过夜,

72 °C 10 min 灭活连接酶,得到不同长度的启动子连接混合产物。使用 1% 琼脂凝胶电泳观察混合产物大小。

1.2.3 启动子连接产物柱纯化与电泳鉴定

颠倒纯化柱,完全悬浮凝胶混合物;常温700×g离心5 min,弃上清;100 μL连接产物悬空加至纯化柱凝胶,700×g离心5 min,收集纯化产物。测定纯化后产物的浓度和纯度,并通过1%琼脂凝胶电泳观察纯化后产物的长度。

1.2.4 启动子质粒构建

用Zero Blunt® TOPO PCR克隆试剂盒将不同大小的启动子纯化产物与pCR®-Blunt II-TOPO®质粒连接,构建Topo-SPs质粒,具体步骤如下:取纯化启动子片段3 μL、盐溶液1 μL、灭菌水1 μL和pCR®-Blunt II-TOPO®载体1 μL,依次加入到0.5 mL的EP管中,轻柔混匀,室温静置5 min,完成连接。将以上产物转化至*E. coli* DH5α感受态细胞,取50 μL菌液涂布于含卡那霉素的LB固体培养基37℃培养过夜。

1.2.5 Topo-SPs质粒鉴定

从培养过夜的LB固体培养基中随机挑取10个单菌落,接种于含卡那霉素的LB液体培养基中,编号1-10,37℃恒温培养箱中振荡培养24 h后,取100 μL菌液进行快速裂解鉴定。使用琼脂凝胶电泳初步观察质粒大小,构建成功的质粒送成都擎科梓熙生物技术有限公司进行测序。

1.2.6 合成启动子库构建

挑取LB固体培养基上所有单菌落(1 128个)摇菌,提取质粒并测序。在该平板上加10 mL含卡那霉素的LB液体培养基,移液枪吹下平板上的所有单菌落至液体培养基中,37℃振荡培养过夜,构成合成启动子库。

1.2.7 荧光素酶质粒构建

质粒pcDNA3.1(+)和pGL3-Basic进行*Mlu* I/*Nhe* I双酶切,将CMV promoter片段连接至pGL3-Basic构建pCMV-Luc质粒。质粒Topo-SPs和pGL3-Basic进行*Kpn* I/*Xho* I双酶切,将启动子片段连接至pGL3-Basic构建pSPs-Luc质粒。

1.2.8 启动子筛选及其肌肉特异性

将pCMV-Luc和pSPs-Luc质粒转染至C2C12细胞,48 h后检测细胞荧光素酶活性和蛋白含量,计算出样品相对荧光素酶活性。相对荧光素酶活性(RLU/mg protein)=荧光素酶活性值/蛋白含量。选取相对荧光素酶活性较高的启动子质粒和pCMV-Luc,分别转染至HeLa和3T3细胞,48 h后利用Pierce™ BCA Protein Assay Kit和Luciferase Assay Kit分别检测蛋白含量和细胞荧光素酶活性,通过上述公式计算出样品相对荧光素酶活性,通过比较合成启动子质粒在3种不同细胞中相对荧光素酶活性来分析其肌肉特异性。

1.2.9 启动子元件分析

对1.2.6中活性较高且测序成功的启动子元件进行组成分析,并利用皮尔逊相关系数分析各启动子元件之间的分布关系及启动子元件含量与启动子活性强度之间的关系,寻找启动子元件之间或启动子元件与启动子强度的潜在关系。

2 结果与分析

2.1 启动子元件连接产物

进行1%琼脂凝胶电泳,发现启动子连接产物大小分布较广(图1),从100 bp到2 000 bp均有片段存在,连接产物条带清晰且亮度强,证明连接效果好,可进行后续实验。

2.2 纯化后的合成启动子片段

测定纯化后的启动子元件,发现产物的平均浓度为102.1 ng/μL,且 A_{260}/A_{280} 和 A_{260}/A_{230} 均在正常范围内,纯度较好(表2)。进行1%琼脂凝胶电泳测定,发现条带较纯化前更加集中,100 bp到1 000 bp范围内亮度更高(图2)。

2.3 合成的启动子单菌落裂解产物

随机挑取10个菌落,其中9个菌落有电泳条带,但所含的质粒大小不一(图3),说明这些质粒是由大小不一的合成启动子片段与TOPO载体连接构成,表明初步连接成功。

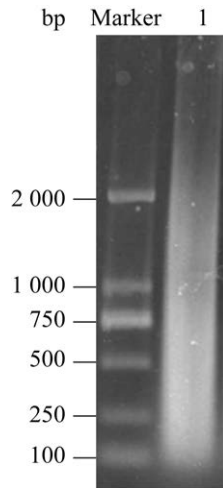


图 1 启动子连接产物电泳图 Marker: DL2000 DNA marker; Lane 1: 启动子连接产物

Figure 1 Electrophoretogram of promoter ligation products. Marker: DL2000 DNA marker; Lane 1: Promoter ligation product.

表 2 柱纯化合成启动子的浓度和纯度

Table 2 The concentration and purity of the synthetic promoter after column purification

Number	Concentration (ng/ μ L)	$A_{260/280}$	$A_{260/230}$
1	103.2	1.86	1.98
2	103.0	1.85	1.97
3	100.3	1.86	1.94
Mean value	102.1	1.86	1.96

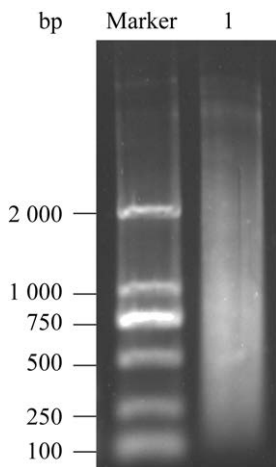


图 2 纯化后启动子连接产物电泳图 Marker: DL2000 DNA marker; Lane 1: 纯化后的启动子连接产物

Figure 2 Electrophoretogram of purified promoter ligation products. Marker: DL2000 DNA marker; Lane 1: The purified promoter ligation product.

2.4 启动子质粒活性

将从 1 128 个单菌落中提取的 pSPs-Luc 分别转染至 C2C12 细胞, 转染成功且具有荧光素酶活性的合成启动子质粒为 321 个, 其中 4 个质粒启动子活性为 CMV 的 2 倍以上, 分别是 SP-301、SP-985、SP-129 和 SP-182; 有 45 个启动子质粒活性是 CMV 的 0.5–2.0 倍; 其余 272 个启动子质粒活性为 CMV 活性的 50% 或以下(图 4)。

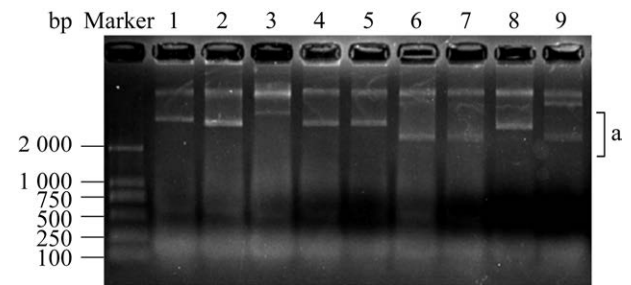


图 3 单菌落裂解产物电泳图 Marker: DL2000 DNA marker; Lane 1–9: 目的片段; a: Topo-SPs 质粒片段

Figure 3 Electrophoretogram of single colony lysis. Marker: DL2000 DNA marker; Lane 1–9: The target fragments; a: Topo-SPs plasmid fragment.

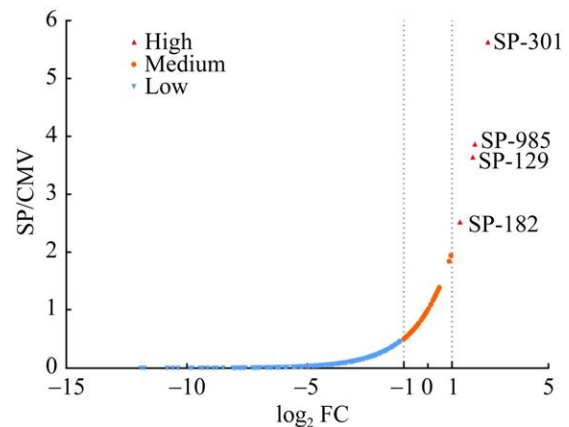


图 4 启动子质粒活性 横坐标: SP 启动子与 CMV 启动子差异倍数的对数值; 纵坐标: SP 启动子与 CMV 启动子活性的比值

Figure 4 Promoter screening map. Horizontal coordinate: Logarithmic value of difference between SP promoter and CMV promoter; Ordinate: The ratio of SP promoter activity to CMV promoter activity.

2.5 启动子特异性

将筛选得到的活性较好的启动子质粒 pSP-301、pSP-985、pSP-129、pSP-61、pSP-228、pSP-230、pSP-61、pSP-817、pSP-631、pSP-852、pSP-319、pSP-61、pSP-762、pSP-1069、pSP-504、pSP-849 和 pCMV-Luc 转染至 HeLa 和 3T3 细胞, 检测荧光素酶活性, 发现所有合成启动子质粒在 C2C12 细胞中的活性均极显著高于 HeLa 和 3T3 细胞($P<0.01$) (图 5)。

2.6 启动子元件组成分析

启动子活性高于 CMV 且测序比对正确的合成启动子序列如图 6A 所示, 可以看出这些合成启动子序列均由不同启动子元件序列拼接而成, 长短不一。分析启动子元件组成发现部分合成启动子所含的启动子元件分布较均匀, 如 SP-301 和 SP-817; 而部分合成启动子所含启动子元件具有偏向性, 如 SP-228 中 72 bp 的 SV40 元件占比达 40% (图 6B)。分析各个合成启动子的元件来源, 发现肌肉特异性元件在这些启动子组成中总体占比偏高(图 6C)。

2.7 启动子元件组成与酶活的相关性

分析启动子元件与活性的相关性(图 7A), 发现 MAPF、GATA-2、72 bp SV40 和 CMV 启动子 18 bp 与启动子活性相关性相对较高, 而且 MEF-1 与 MEF-2、MEF-1 与 CMV 增强子

19 bp、Trex 与 MEF-3、Trex 与 CMV 启动子 16 bp、CMV 启动子 16 bp 与 19 bp 之间相关性显著($P<0.05$)。进一步分析启动子来源与启动子的活性(图 7B), 发现保守元件与启动子活性之间相关性显著($P<0.05$)。而各元件之间, 肌肉特异性元件与病毒元件相关性显著($P<0.05$), 与保守元件相关性极显著($P<0.01$)。从图 8B 可看出保守元件含量与启动子活性的大致趋势, 即随着保守元件含量降低, 启动子活性也逐渐降低, 但并不是所有启动子都有此特点。由图 8A 和 8C 中并不能明显看出肌肉特异性元件和病毒元件含量与启动子活性之间的关系。

3 讨论与结论

通过分析启动子的结构组成, 发现启动子序列由不同大小的调控元件组成, 而启动子的启动强度和组织特异性对基因的表达和转录过程有一定影响。本研究在肌肉特异性元件、保守元件和病毒元件这 3 种来源的启动子中挑选了 19 个不同的 DNA 调控元件作为组装元件, 将不同启动子元件进行随机连接, 连接到载体上, 得到合成启动子库。在质粒构建过程中, 由于启动子片段大小不一且均为平末端连接, 普通 T4 连接酶难以连接成功^[22]。利用具有高效连接效率的 Zero Blunt® TOPO PCR Cloning Kit

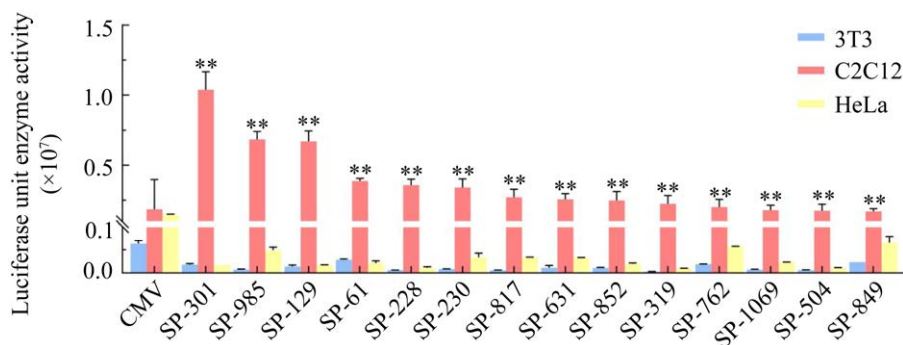


图 5 启动子质粒转染不同细胞的荧光素酶活性

Figure 5 Luciferase unit activity detection of promoter plasmid in different cells. **: $P<0.01$.

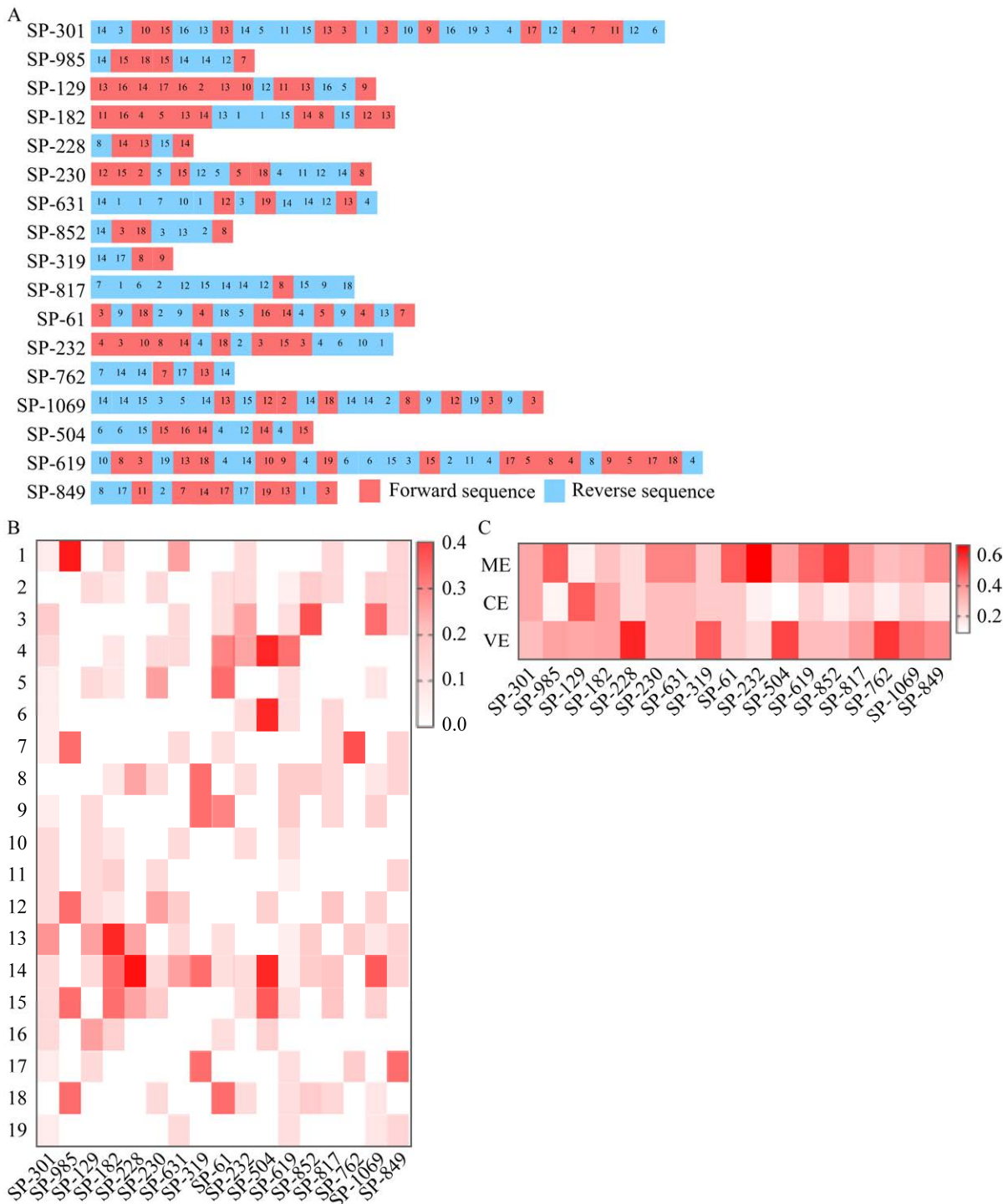


图 6 合成启动子序列元件组成与占比 A: 合成启动子序列元件组成. B: 各元件在合成启动子中的含量占比. C: 三种来源元件在合成启动子中的含量占比. 启动子元件序号如表 1 所示. ME: 肌肉特异性元件; CE: 保守元件; VE: 病毒元件

Figure 6 Component composition and proportion of synthetic promoter sequence. A: Synthetic promoter sequence components. B: The proportion of each element in the synthetic promoter. C: The content of three promoter elements in synthetic promoter. The promoter element number is shown in Table 1. ME: Muscle specific element; CE: Conserved element; VE: Virus element.

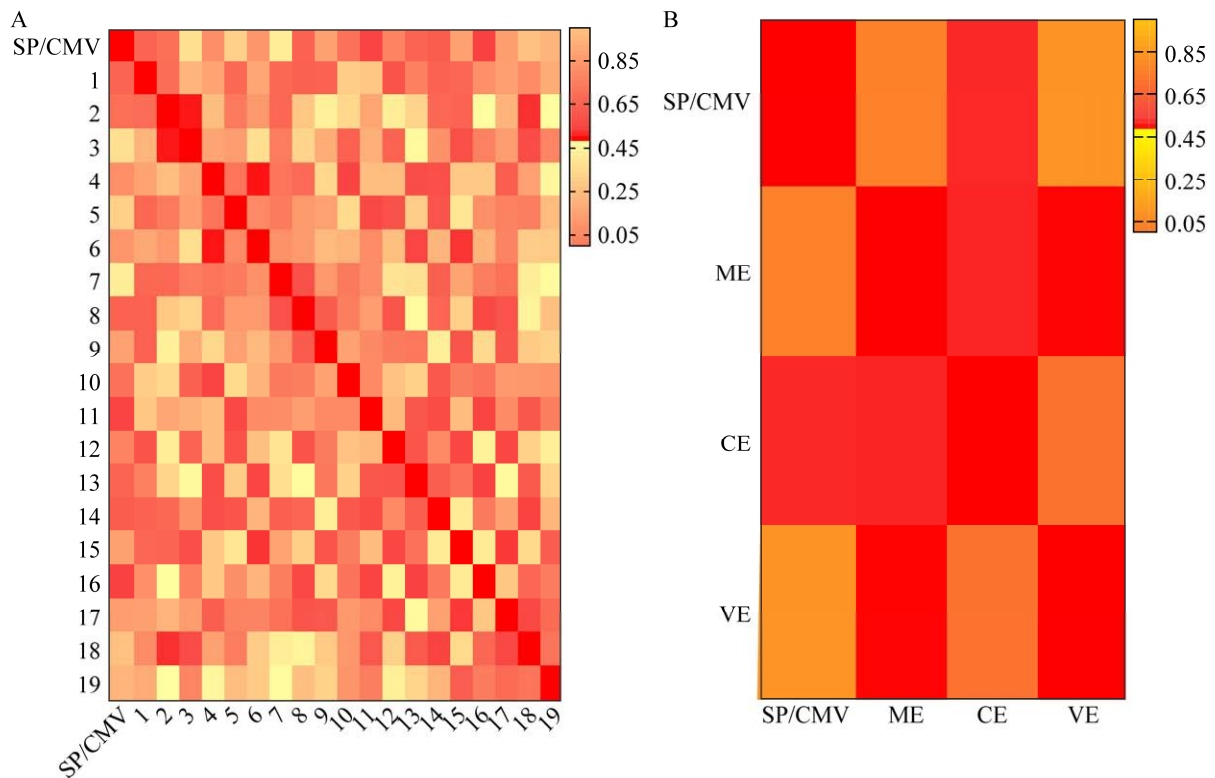


图7 各种合成启动子元件与活性的相关性 A: 各启动子元件与活性的相关性分析. B: 不同来源启动子与活性的相关性分析. SP/CMV: SP启动子与CMV启动子活性的比值; ME: 肌肉特异性元件; CE: 保守元件; VE: 病毒元件

Figure 7 Correlation analysis of composition and activity of synthetic promoter elements. A: Correlation analysis of promoter elements and activity. B: Correlation analysis of different promoters and activity. SP/CMV: The ratio of SP promoter activity to CMV promoter activity; ME: Muscle specific element; CE: Conserved element; VE: Virus element.

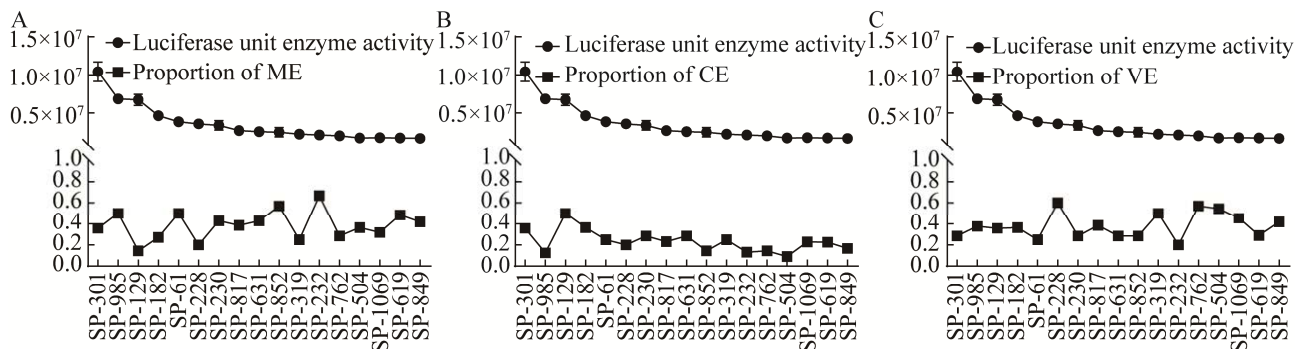


图8 不同来源启动子含量与启动子活性关系图 A: ME含量与各启动子活性. B: CE含量与各启动子活性. C: VE含量与各启动子活性

Figure 8 The relationship between promoter content and promoter activity from different sources. A: ME content and promoter activity. B: CE content and promoter activity. C: VE content and promoter activity.

盒,依赖于 A 和 T 碱基的互补配对,无需连接酶即可直接将平端 DNA 片段插入质粒载体,测序成功的质粒共同构成了合成启动子文库,可进行下一步筛选。

将 pSPs-Luc 和 pCMV-Luc 转染 C2C12 细胞后,用单位荧光素酶检测方法评估各合成启动子的活性,进行初步筛选。从合成启动子库中筛选了 1 128 个单菌落,质粒提取且转染成功 321 个 pSPs-Luc,将这些启动子的活性与 CMV 进行比较,发现大多数启动子活性较低,启动子片段随机组合得到高活性启动子的效率低。CMV 启动子是启动真核基因表达最有力的启动子之一^[23],而筛选出的活性高于 CMV 的启动子能使基因表达量更高,具有较好的应用前景,研究其序列组成有助于更高效地合成高活性启动子。为了进一步验证这些启动子是否具有肌肉特异性,将以上高活性启动子质粒 pSPs-Luc 分别转染到 2 种不同组织细胞 3T3 和 HeLa。结果发现,以上质粒在肌肉细胞内的单位荧光素酶活性极显著高于其他 2 种细胞,说明以上高活性启动子质粒同时也具有良好的肌肉特异性。

启动子作为基因表达载体的重要组成部分之一,在提高外源基因表达效率方面发挥着重要作用,在基因工程领域的开发和利用上具有重要的研究意义^[13]。因此,对启动子结构和功能的深入研究是相当重要的,一方面能更有利于阐明基因转录模式和调控机制,另一方面有助于对外源基因的表达进行合理调控^[24-25]。分析 17 个启动子活性高于 CMV 且测序比对正确的启动子序列,发现启动子各元件的相关性分析结果与不同来源启动子分析结果相符。从不同启动子来源占比结果可以看出,活性最高的 4 个启动子 SP-301、SP-985、SP-129 和 SP-182 中,SP-301 和 SP-182 中这 3 种来源的启动子元件含

量无较大差别,而 SP-985 中肌肉特异性启动子含量较高,SP-129 中保守元件含量较高。值得关注的是,病毒元件在这 4 种启动子内占据的比例相似,且在所有启动子中,病毒元件含量均不低于 20%。本研究中病毒元件为 CMV 或 SV40 的核心启动子元件,CMV 启动子是从人巨细胞病毒中发现的强启动子,主要包括 5 个部分:CMV 独特区域(-683 至-545 bp)、CMV 增强子区域(-544 至-663 bp)、核心聚合酶 II 启动子区域(-62 至-1 bp)、外显子 I 区域(+1 至+121 bp)和内含子 A 区域(+121 至+954 bp)^[26]。SV40 启动子是猿猴空泡病毒 40 启动子,常作为质粒构建的高活性启动子^[27],这可能是这些启动子具有高启动子活性的原因。肌肉特异性元件在这 17 个启动子中的比例均偏高,但与启动子活性关系弱,这可能是由于这些元件的主要功能为在特定的组织细胞中启动基因的转录与表达^[28]。分析启动子保守元件,发现该元件含量与启动子活性显著相关,且与启动子活性大体成正比。本研究中保守元件来源于启动子核心区 and 上游调控区元件,这些元件具有直接、准确和有效地指导起始转录、激活基因转录、控制转录起始的频率和强度以及促进相应基因转录等作用^[7,12,29-30],以上功能可能是保守元件含量会显著影响启动子活性的原因。在这些高活性肌肉特异性启动子中,肌肉特异性元件含量与保守元件、病毒元件含量显著相关,这可能是肌肉特异性元件在这些启动子中普遍占比偏高所致。尽管得到的高活性启动子数量较少,但基于本研究中的相关性结果以及不同启动子元件功能的区别,启动子元件含量组成与合成启动子活性之间的关系值得进一步探索。与此同时,由于缺乏两者的相关性研究,实验存在一定局限性,还需要更加深入地研究论证其之间可能存在的关联。

本研究首次构建了合成肌肉特异性启动子库, 验证了部分启动子的肌肉特异性, 并对启动子元件组成与启动子活性进行相关性分析, 有利于进一步阐述不同来源启动子元件含量高低对启动子活性的影响, 为合成新型高效启动子提供了理论基础, 并为疾病基因治疗中基因在肌肉组织的高效特异性表达提供了更有效的方案。

REFERENCES

- [1] ZHANG MQ. Identification of human gene core promoters *in silico*[J]. *Genome Research*, 1998, 8(3): 319-326.
- [2] SMALE ST, KADONAGA JT. The RNA polymerase II core promoter[J]. *Annual Review of Biochemistry*, 2003, 72: 449-479.
- [3] LITTLEFIELD O, KORKHIN Y, SIGLER PB. The structural basis for the oriented assembly of a TBP/TFB/promoter complex[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(24): 13668-13673.
- [4] DENG W, ROBERTS SGE. Core promoter elements recognized by transcription factor IIB[J]. *Biochemical Society Transactions*, 2006, 34(Pt 6): 1051-1053.
- [5] SMALE ST, BALTIMORE D. The 'initiator' as a transcription control element[J]. *Cell*, 1989, 57(1): 103-113.
- [6] LO K, SMALE ST. Generality of a functional initiator consensus sequence[J]. *Gene*, 1996, 182(1-2): 13-22.
- [7] KUTACH AK, KADONAGA JT. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters[J]. *Molecular and Cellular Biology*, 2000, 20(13): 4754-4764.
- [8] KADONAGA JT. The DPE, a core promoter element for transcription by RNA polymerase II[J]. *Experimental & Molecular Medicine*, 2002, 34(4): 259-264.
- [9] JUVEN-GERSHON T, KADONAGA JT. Regulation of gene expression *via* the core promoter and the basal transcriptional machinery[J]. *Developmental Biology*, 2010, 339(2): 225-229.
- [10] URIOSTEGUI-ARCOS M, MICK ST, SHI Z, RAHMAN R, FISZBEIN A. Splicing activates transcription from weak promoters upstream of alternative exons[J]. *Nature Communications*, 2023, 14(1): 3435.
- [11] SHAHMURADOV IA, GAMMERMAN AJ, HANCOCK JM, BRAMLEY PM, SOLOVYEV VV. PlantProm: a database of plant promoter sequences[J]. *Nucleic Acids Research*, 2003, 31(1): 114-117.
- [12] SREENIVASULU G, SENTHILKUMARAN B, SUDHAKUMARI CC, GUAN G, OBA Y, KAGAWA H, NAGAHAMA Y. 20 β -hydroxysteroid dehydrogenase gene promoter: potential role for cyclic AMP and xenobiotic responsive elements[J]. *Gene*, 2012, 509(1): 68-76.
- [13] KHALIL AS, COLLINS JJ. Synthetic biology: applications come of age[J]. *Nature Reviews Genetics*, 2010, 11(5): 367-379.
- [14] YANG Y, CHAFFIN TA, SHAO YH, BALASUBRAMANIAN VK, MARKILLIE M, MITCHELL H, RUBIO-WILHELMI MM, AHKAMI AH, BLUMWALD E, NEAL STEWART C Jr. Novel synthetic inducible promoters controlling gene expression during water-deficit stress with green tissue specificity in transgenic poplar[J]. *Plant Biotechnology Journal*, 2024, 22(6): 1596-1609.
- [15] ZHAO M, SHANG J, CHEN JJ, ZABED HM, QI XH. Fine-tuning the expression of the glycolate giosynthetic pathway in *Escherichia coli* using synthetic promoters[J]. *Fermentation*, 2024, 10(1): 67.
- [16] LEE TC, SHI Y, SCHWARTZ RJ. Displacement of BrdUrd-induced YY1 by serum response factor activates skeletal alpha-actin transcription in embryonic myoblasts[J]. *Proceedings of the National Academy of Sciences*, 1992, 89(20): 9814-9818.
- [17] BUSKIN JN, HAUSCHKA SD. Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene[J]. *Molecular and Cellular Biology*, 1989, 9(6): 2627-2640.
- [18] HIMEDA CL, TAI PW, HAUSCHKA SD. Analysis of muscle gene transcription in cultured skeletal muscle cells[J]. *Methods of Molecular Biology*, 2012, 798: 425-443.
- [19] WALSH K. Cross-binding of factors to functionally different promoter elements in *c-fos* and skeletal actin genes[J]. *Molecular and Cellular Biology*, 1989, 9(5): 2191-2201.
- [20] KADONAGA JT, JONES KA, TJIAN R. Promoter-specific activation of RNA polymerase II transcription by Sp1[J]. *Trends in Biochemical Sciences*, 1986, 11(1): 20-23.

- [21] STAMMINGER T, FICKENSCHER H, FLECKENSTEIN B. Cell type-specific induction of the major immediate early enhancer of human cytomegalovirus by cyclic AMP[J]. *Journal of General Virology*, 1990, 71(1): 105-113.
- [22] GREEN MR, SAMBROOK J. Cloning in plasmid vectors: blunt-end cloning[J]. *Cold Spring Harbor Protocols*, 2020, 2020(11): pdb.prot101246.
- [23] ZHANG JH, WANG TY, ZHANG CB, MI CL, GENG SL, TANG YY, WANG XY. CMV/AAT promoter of MAR-based episomal vector enhanced transgene expression in human hepatic cells[J]. *3 Biotech*, 2023, 13(11): 354.
- [24] YOUNG E, ALPER H. Synthetic biology: tools to design, build, and optimize cellular processes[J]. *Journal of Biomedicine & Biotechnology*, 2010, 2010: 130781.
- [25] KEASLING JD. Synthetic biology and the development of tools for metabolic engineering[J]. *Metabolic Engineering*, 2012, 14(3): 189-195.
- [26] 张华韵, 蔡欢嫦, 任静强. 基于巨细胞病毒启动子的三独立表达系统载体的构建和应用[J]. *中国兽医学报*, 2023, 43(6): 1257-1262.
- ZHANG HY, CAI HC, REN JQ. Design and application of the vector based on cytomegalovirus promoter in three independent expression systems[J]. *Chinese Journal of Veterinary Science*, 2023, 43(6): 1257-1262 (in Chinese).
- [27] WILDEMAN AG, ZENKE M, SCHATZ C, WINTZERITH M, GRUNDSTRÖM T, MATTHES H, TAKAHASHI K, CHAMBON P. Specific protein binding to the Simian virus 40 enhancer *in vitro*[J]. *Molecular and Cellular Biology*, 1986, 6(6): 2098-2105.
- [28] TAI PW, FISHER-AYLOR KI, HIMEDA CL, SMITH CL, MACKENZIE AP, HELTERLINE DL, ANGELLO JC, WELIKSON RE, WOLD BJ, HAUSCHKA SD. Differentiation and fiber type-specific activity of a muscle creatine kinase intronic enhancer[J]. *Skeletal Muscle*, 2011, 1: 25.
- [29] CAREY MF, PETERSON CL, SMALE AST. *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*[M]. 2nd Ed. New York: Cold Spring Harbor Laboratory Press, 2009.
- [30] JOSHI CP. An inspection of the domain between putative TATA box and translation start site in 79 plant genes[J]. *Nucleic Acids Research*, 1987, 15(16): 6643-6653.

(本文责编 郝丽芳)