

靶向小核RNA基因构建PCR教学实验的生物信息学分析

刘妍¹, 钟柳英¹, 徐嘉琪², 何震宇^{2*}

¹广东药科大学药学院, 广东 广州

²广东药科大学生命科学与生物制药学院, 广东 广州

Email: *jdbshzy@163.com, 751425677@qq.com, 1144744278@qq.com, 863462675@qq.com

收稿日期: 2020年8月27日; 录用日期: 2020年9月4日; 发布日期: 2020年9月11日

摘要

目的: 探讨针对小核RNA基因, 构建本科聚合酶链反应(PCR)教学实验的可行性。方法: 调取NCBI数据库中小核RNA基因U1、U2、U4、U5及U6的序列, 分别截取5'端和3'端部分序列作为上、下游引物, 进行电子PCR, 分析其靶点情况和引物效率。结果: 电子PCR能清楚显示每对引物潜在的靶位点及相应PCR产物的大小。针对U1、U2、U4、U5、U6基因的引物, 对应的靶位点数目分别为18、16、4、1、61, 对应的PCR产物大小分别为164~166、187~188、144、117、95~108 bp。结论: 针对U1、U2、U6基因的引物, 其潜在的靶位点数量较多, 且绝大部分靶位点的引物结合效率较好, 相应的PCR扩增子长度较短, PCR容易实现, PCR连同电泳检测有望在2个小时内完成, 比较适合构建本科PCR教学实验。

关键词

小核RNA, 聚合酶链反应, 教学实验

Bioinformatics Analysis of Establishment of PCR Teaching Experiments Targeting Small Nuclear RNA Genes

Yan Liu¹, Liuying Zhong¹, Jiaqi Xu², Zhenyu He^{2*}

¹School of Pharmacy, Guangdong Pharmaceutical University, Guangzhou Guangdong

²School of Life Science and Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou Guangdong

Email: *jdbshzy@163.com, 751425677@qq.com, 1144744278@qq.com, 863462675@qq.com

Received: Aug. 27th, 2020; accepted: Sep. 4th, 2020; published: Sep. 11th, 2020

*通讯作者。

文章引用: 刘妍, 钟柳英, 徐嘉琪, 何震宇. 靶向小核 RNA 基因构建 PCR 教学实验的生物信息学分析[J]. 计算生物学, 2020, 10(3): 41-47. DOI: 10.12677/hjcb.2020.103005

Abstract

Objective: To establish PCR experiments targeting small nuclear RNA genes, which is expected to be used in undergraduate teaching. **Methods:** The sequences of snRNA coding genes U1, U2, U4, U5 and U6 in NCBI database were obtained, and the 5' and 3' end partial sequences were intercepted as upstream and downstream primers for In-Silico PCR. The targets and primer efficiency were analyzed. **Results:** The potential target sites of each pair of primers and the size of corresponding PCR products could be clearly displayed by In-Silico PCR. The corresponding target number of primers for U1, U2, U4, U5 and U6 genes were 18, 16, 4, 1, 61, and the corresponding PCR product sizes were 164 - 166, 187 - 188, 144, 117 and 95 - 108 bp. **Conclusion:** For the primers of U1, U2, U6 gene, there are many potential target binding sites, and most of them have good binding efficiency. The corresponding length of PCR amplicons is short, and the PCR is easy to realize. PCR together with electrophoresis detection is expected to be completed in two hours, which is suitable for the establishment of undergraduate PCR teaching experiment.

Keywords

Small Nuclear RNA, Polymerase Chain Reaction, Teaching Experiment

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

PCR 作为经典的分子生物学技术, 很有必要纳入相关专业的本科实验教学体系, 但目前使用市面上的教学用 PCR 试剂盒开展实验教学, 时程长、成本高, 且试剂盒提供的模板、引物背景不清楚, 不太利于教学, 亟需加以改进。

本研究旨在自主构建一个更贴近教学实际的 PCR 实验体系, 其核心要素即模板(靶标)、引物的背景必须清晰, 且能在较短时间内完成实验流程。我们选择小核 RNA (snRNA) 基因作为靶标, 设计引物, 通过电子 PCR 分析其可行性。

2. 方法

从 GenBank 中调取 snRNA 基因组 DNA 序列, 找出编码区序列, 再根据实际情况分别截取 5' 端和 3' 端部分序列作为上、下游引物进行电子 PCR (见表 1), 使用的在线分析程序为 UCSC Genome Browser 的 In-Silico PCR [1] 和 NCBI 的 primer-blast [2]。使用 Primer Premier 5.0 软件分析引物与模板的结合效率。

Table 1. Sequence information of snRNA coding genes and primer sequences for In-Silico PCR

表 1. snRNA 编码基因序列信息及电子 PCR 所用引物序列

基因	GenBank 登录号	编码区域	引物序列(5'→3')
U1	J00318.1	433~596	F1: ATACTTACCTGGCAGGGGAGATACCATGA R1: CAGGGGAAAGCGCGAACGC
U2	U57614.1	4882~5069	F2: ATCGCTTCTCGGCCTTTGGCTAAG R2: GGGTGCACCGTTCCTGGAGGTACTG

Continued

U4	NR_003925.1	1~144	F4: AGCTTTGCGCAGTGGCAGTATCGT R4: CAGTCTCCGTAGAGACTGTCAAAAATTGCC
U5	NR_002756.2	1~117	F5: ATACTCTGGTTTCTTTCAGATCGCATAAATC R5: TAGCCTTGCCAAAGCAAGGCC
U6	M14486.1	329~434	F6: GTGCTCGCTTCGGCAGCACATATAC R6: AAAATATGGAACGCTTCACGAATTGCG

3. 结果

针对不同 snRNA 编码基因相应引物的配对情况、结合效率、潜在的靶位点数量及 PCR 产物的大小分别见表 2~5。

Table 2. In-Silico PCR results of U1 gene primers F1, R1

表 2. U1 基因引物 F1、R1 电子 PCR 结果

序号	靶序列所在染色体	所在染色体的区域	PCR 产物大小(bp)	与参考序列的碱基差异数目(bp)	引物 F1 错配碱基数	引物 R1 错配碱基数	引物结合效率
1	1	16514122~16514285	164	0	0	0	72
2	1	16666785~16666948	164	0	0	0	72
3	1	16740516~16740679	164	0	0	0	72
4	1	16895980~16896143	164	0	0	0	72
5	1	143729407~143729570	164	0	0	0	72
6	1	146376807~146376970	164	0	0	0	72
7	1	144412576~144412740	165	18	2	0	56
8	1	144560666~144560829	164	1	0	0	72
9	1	145465617~145465780	164	2	0	0	72
10	1	148038753~148038916	164	1	0	1	72
11	1	148522601~148522765	165	15	0	1	53
12	4	135995929~135996092	164	5	0	1	72
13	7	141727984~141728148	165	13	1	1	72
14	14	34546714~34546877	164	0	0	0	72
15	14	34556226~34556389	164	0	0	0	72
16	15	44884616~44884781	166	20	2	2	72
17	15	45002863~45003028	166	18	2	2	72
18	X	119423742~119423905	164	8	2	2	72

Table 3. In-Silico PCR results of U2 gene primers F2, R2

表 3. U2 基因引物 F2、R2 电子 PCR 结果

序号	靶序列所在染色体	所在染色体的区域	PCR 产物大小(bp)	与参考序列的碱基差异数目(bp)	引物 F2 错配碱基数	引物 R2 错配碱基数	引物结合效率
1	1	150236967~150237153	187	21	1	1	76
2	10	101364848~101365035	188	8	1	1	76
3	11	62841622~62841809	188	6	1	2	76

Continued

4	17	43296428~43296615	188	0	0	0	76
5	17	43300044~43300231	188	0	0	0	76
6	17	43284142~43284329	188	0	0	0	76
7	17	43277981~43278168	188	0	0	0	76
8	17	43271372~43271559	188	0	0	0	76
9	17	43266636~43266823	188	0	0	0	76
10	17	43251832~43252019	188	0	0	0	76
11	17	43245677~43245864	188	0	0	0	76
12	17	43239924~43240111	188	0	0	0	76
13	17	43241183~43241370	188	0	0	0	76
14	17	43233790~43233977	188	0	0	0	76
15	17	43290294~43290481	188	1	0	1	76
16	18	41467956~41468143	188	19	2	2	76

Table 4. In-Silico PCR results of U4 gene primers F4, R4 and U5 gene primers F5, R5**表 4.** U4 基因引物 F4、R4 及 U5 基因引物 F5、R5 电子 PCR 结果

序号	使用引物	靶序列所在染色体	所在染色体的区域	PCR 产物大小(bp)	与参考序列的碱基差异数目(bp)	引物错配碱基数	引物错配碱基数	引物结合效率
1	F4、R4	12	120293094~120293237	144	0	0	0	65
2	F4、R4	12	120291760~120291903	144	4	0	0	65
3	F4、R4	12	25404285~25404428	144	14	2	1	65
4	F4、R4	12	59537356~59537499	144	13	1	1	65
5	F5、R5	15	65296051~65296167	117	0	0	0	66

Table 5. In-Silico PCR results of U6 gene primers F6, R6**表 5.** U6 基因引物 F6、R6 电子 PCR 结果

序号	靶序列所在染色体	所在染色体的区域	PCR 产物大小(bp)	与参考序列的碱基差异数目(bp)	引物 F6 错配碱基数	引物 R6 错配碱基数	引物结合效率
1	1	31497578~31497683	106	4	1	0	67
2	1	10298966~10299071	106	1	0	1	67
3	1	182982213~182982318	106	3	0	2	67
4	1	180758722~180758816	95	17	0	2	67
5	1	235915415~235915520	106	8	2	1	67
6	1	27675604~27675709	106	8	2	2	67
7	2	174557967~174558072	106	0	0	0	67
8	2	200830009~200830115	107	4	1	0	67
9	2	32214457~32214562	106	9	2	2	67
10	3	181231737~181231842	106	0	0	0	67
11	3	195214787~195214892	106	1	0	1	67
12	3	196784981~196785086	106	3	1	0	67
13	3	98804979~98805084	106	5	1	1	67
14	3	169720188~169720293	106	7	1	1	67
15	3	57512450~57512554	105	11	2	2	67

Continued

16	3	100952627~100952732	106	14	2	2	67
17	4	108652150~108652255	106	7	0	0	67
18	4	39297606~39297711	106	1	0	0	67
19	4	76532223~76532327	105	8	1	0	67
20	4	88684849~88684954	106	1	1	0	67
21	4	109992325~109992430	106	4	1	1	67
22	4	96152298~96152403	106	5	1	1	67
23	5	51055354~51055459	106	7	1	2	67
24	5	32309662~32309767	106	6	2	2	67
25	6	23124982~23125087	106	6	2	1	67
26	7	123790606~123790711	106	0	0	1	67
27	7	92701709~92701814	106	3	1	1	67
28	7	98794719~98794824	106	6	2	2	67
29	8	56917084~56917189	106	1	0	0	67
30	8	118976513~118976618	106	1	0	0	67
31	8	13044350~13044455	106	7	2	2	67
32	8	124260646~124260751	106	11	2	2	67
33	8	137105352~137105457	106	6	2	2	67
34	10	13217269~13217374	106	0	0	0	67
35	10	21321961~21322066	106	2	1	1	67
36	10	30299570~30299673	104	18	2	2	67
37	11	63970470~63970575	106	4	2	0	67
38	11	67895631~67895736	106	5	2	0	67
39	12	97721716~97721821	106	1	1	0	67
40	13	75109587~75109692	106	5	1	2	67
41	13	27828763~27828868	106	7	2	1	67
42	14	32202045~32202150	106	0	0	0	67
43	14	32203164~32203269	106	0	0	0	67
44	15	32284152~32284257	106	1	0	0	67
45	15	30254403~30254508	106	1	0	0	67
46	15	67839940~67840045	106	0	0	0	67
47	15	65553082~65553187	106	4	1	0	67
48	16	69175672~69175777	106	4	1	0	67
49	16	62187541~62187646	106	5	1	2	67
50	16	57830537~57830642	106	4	2	1	67
51	18	68858934~68859039	106	2	1	0	67
52	18	61391595~61391702	108	11	2	1	52
53	19	893484~893589	106	0	0	0	67
54	19	1021522~1021627	106	0	0	0	67
55	19	53653685~53653790	106	6	1	2	67
56	22	20416798~20416899	102	14	0	2	67
57	X	141118031~141118136	106	0	0	0	67
58	X	101634007~101634112	106	4	0	2	67
59	X	48776966~48777071	106	3	2	1	67
60	X	37870109~37870214	106	5	1	2	67
61	X	42174280~42174385	106	7	2	2	67

4. 讨论

PCR 是聚合酶链反应(Polymerase Chain Reaction)的简称,用于体外扩增特定的 DNA 片段,可看作生物体外的一种特殊的 DNA 复制,其最大特点是能使微量的 DNA 大量增加。PCR 技术堪称分子生物学的支撑技术,在生命科学领域应用广泛,比如,在基因工程中,用于目的基因的制备;在临床上,用于遗传病的基因诊断、感染性疾病的病原体检测、肿瘤基因检测及器官移植的配型[3];在法医学鉴定中,用于亲权鉴定、个体识别[4] [5];在食品领域,用于食源性致病菌的鉴定、食品中成分种类的检测、食品中有效成分的检测、转基因食品的鉴定等[6];在药学领域,可用于中药的鉴定[7]、个体化用药相关基因的检测[8];在环境科学领域,用于环境监测[9]。因此,在《生物化学与分子生物学》以及相关的课程中开设 PCR 实验的意义不言而喻,它不仅能加深同学们对 PCR 原理的理解,而且能让同学们通过实践认识到该项技术的特点和价值。

目前有多款教学用 PCR 试剂盒,大体上分为两类,第一类带有模板、上游引物、下游引物、dNTPs、Taq DNA 聚合酶、10× PCR 缓冲液、超纯水;第二类带有模板、上游引物、下游引物、2× Taq PCR MasterMix、超纯水,其中的 2× Taq PCR MasterMix 包含 Taq DNA 聚合酶、dNTPs、MgCl₂、反应缓冲液、PCR 反应的增强剂和优化剂以及稳定剂,使用时,只需补加实验所需的引物和模板后,即可进行 PCR 反应,它比第一类要简便、快速。但无论是哪一类的试剂盒,公司出于商业利益考虑,均未公布模板和引物序列,由于无法知晓具体的靶标序列和引物序列,教师在讲解引物和模板的关系时只能泛泛而谈,在介绍引物的设计原则和 PCR 循环参数的设置时更是无法条分缕析,学生只能一知半解。

本研究旨在构建一个靶标、引物背景明确且简便、快速的教學用 PCR 体系,由于 PCR 反应体系包含五个基本要素:引物、耐热 DNA 聚合酶、dNTPs、模板和缓冲液(含 Mg²⁺),其中耐热 DNA 聚合酶、dNTPs、缓冲液三个基本要素可以三合一,目前有现成的含有三者成分的混合物(Mix)可购买使用,可以大大节省配制 PCR 体系的时间,并且有一些 Mix 中含有的 DNA 聚合酶具有极强的抗抑制能力,对模板的纯度要求低,能够直接用未纯化的样本进行 PCR 扩增(所谓的直接 PCR),无需核酸纯化步骤,只需微量的原始样本作为 PCR 反应的模板,从而节约了实验时间和成本,为 DNA 扩增带来前所未有的便捷。余下的二个要素即模板、引物成为关键要素,本研究拟使用人的口腔上皮细胞为样本进行直接 PCR,可以在实验课现场取样,取样几分钟内就可完成,且由于材料来自同学自身,能提高其学习兴趣。要使 PCR 达到良好的效果,模板(靶标)非常关键,如果靶标丰度高,意味着起始模板浓度高,不仅 PCR 的成功率会大大增加,而且运行的循环次数可以相应减少,从而缩短扩增时间。本研究拟选择 snRNA 编码基因为靶标,snRNA 编码基因产物为 snRNA (small nuclear RNA,小核 RNA),包括 U1、U2、U4、U5、U6 等,它们是真核生物转录后加工过程中 RNA 剪接体(spliceosome)的主要成分,参与 mRNA 前体的加工过程[10] [11]。理论上,适应机体生理功能的需要,这类基因应该有较高丰度,事实上,本研究的生物信息学分析显示 U1、U2、U6 基因确实丰度较高,与文献[12] [13]报道一致,它们存在基因家族,不仅有多拷贝的真基因,还有很多序列同源的假基因,因此引物结合的靶位点较多,PCR 成功的概率很大,且每对引物扩增的数个靶位点的 PCR 产物大小一致或接近,容易形成集中的便于观察的产物条带,这些产物长度都在 200 bp 以下,分子量较小,电泳鉴定所需时间较少,能进一步缩短整个实验的时长。

根据生物信息学分析结果,使用口腔上皮细胞进行直接 PCR,PCR 反应条件可以如下设置:94℃ 预变性 8 min;94℃ 30 s,55℃ 30 s,72℃ 10 s,30 个循环;72℃ 再延伸 5 min。考虑到 PCR 仪的升温、降温,PCR 运行的时间可以控制在 1 h 以内,PCR 前的样本采集及体系配制可以在 10 min 内完成,在 PCR 运行的过程中,可以介绍 PCR 基本原理和配制琼脂糖凝胶,PCR 结束后,点样、电泳及结果观察可以在 40 min 内完成,整个实验流程不超过 2 h。

综上, 根据我们自行设计的引物, 靶向 snRNA 编码基因 U1、U2、U6 基因中的任何一个, 都有望在 2 h 内获得良好的 PCR 结果, 从而构建一个贴近本科教学实际的 PCR 实验。

致 谢

感谢广东药科大学 2018 年教育教学改革项目的支持。

参考文献

- [1] Rhead, B., Karolchik, D., Kuhn, R.M., *et al.* (2010) The UCSC Genome Browser Database: Update 2010. *Nucleic Acids Research*, **38**, D613-D619. <https://doi.org/10.1093/nar/gkp939>
- [2] Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T.L. (2012) Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction. *BMC Bioinformatics*, **13**, Article No. 134. <https://doi.org/10.1186/1471-2105-13-134>
- [3] 胥振国, 蔡玉华. PCR 技术在疾病基因检测方面应用进展[J]. 齐齐哈尔医学院学报, 2018, 39(21): 2539-2542.
- [4] 韩俊萍, 李洋, 马原, 尚蕾, 李彩霞, 孙敬. 快速 PCR 方法在法医 DNA 检验中的研究进展[J]. 生命科学研究, 2017, 21(5): 442-449.
- [5] 秦文钊, 周强, 朱为筑, 喻芳. PCR 技术在亲子鉴定中的应用[J]. 贵州医药, 2000, 24(4): 203-204.
- [6] 江飞, 张旭伟. PCR 技术在食品检测中的应用和发展[J]. 河南农业, 2017(6): 56-58.
- [7] 罗达龙, 黄琳. PCR 技术在中药鉴定中的应用[J]. 临床医药文献电子杂志, 2017, 4(24): 4731.
- [8] 刘志艳, 杨兵, 赵荣生. 基因导向的个体化治疗[J]. 临床药物治疗杂志, 2017, 15(1): 14-20.
- [9] 高琼. 聚合酶链式反应(PCR)技术在环境监测中的应用[J]. 安徽农业科学, 2014, 42(36): 12825-12828.
- [10] Mattaj, I.W., Tollervey, D. and Séraphin, B. (1993) Small Nuclear RNAs in Messenger RNA and Ribosomal RNA Processing. *The FASEB Journal*, **7**, 47-53. <https://doi.org/10.1096/fasebj.7.1.8422974>
- [11] Newman, A. (1994) Small Nuclear RNAs and Pre-mRNA Splicing. *Current Opinion in Cell Biology*, **6**, 360-367. [https://doi.org/10.1016/0955-0674\(94\)90027-2](https://doi.org/10.1016/0955-0674(94)90027-2)
- [12] Domitrovich, A.M. and Kunkel, G.R. (2003) Multiple, Dispersed Human U6 Small Nuclear RNA Genes with Varied Transcriptional Efficiencies. *Nucleic Acids Research*, **31**, 2344-2352. <https://doi.org/10.1093/nar/gkg331>
- [13] Yuan, Y. and Reddy, R. (1988) Organization of Spliceosomal U6 snRNA Genes in the Mouse Genome. *Molecular Biology Reports*, **13**, 159-164. <https://doi.org/10.1007/BF00444312>