

一种结合 SELEX 与二代测序技术筛选 RNA 结合蛋白特异识别序列新方法的建立

张璐, 徐晋晔, 麻锦彪

复旦大学 生命科学学院生物化学系 遗传工程国家重点实验室, 上海 200438

张璐, 徐晋晔, 麻锦彪. 一种结合 SELEX 与二代测序技术筛选 RNA 结合蛋白特异识别序列新方法的建立. 生物工程学报, 2016, 32(7): 966–974.

Zhang L, Xu JH, Ma JB. Screening specific recognition motif of RNA-binding proteins by SELEX in combination with next-generation sequencing technique. Chin J Biotech, 2016, 32(7): 966–974.

摘要: RNA 结合蛋白通过特异识别 RNA 底物发挥重要的生物学作用。指数富集的配体系统进化 (Systematic evolution of ligands by exponential enrichment, SELEX) 技术是一种体外筛选核酸底物的基本方法, SELEX 技术通过重复多轮筛选从随机核酸序列库中筛选出特异性与靶物质高度亲和的核酸底物, 本研究将利用该技术与二代高通量测序 (NGS) 相结合, 体外合成含有 20 个随机碱基的 RNA 文库, 将所要研究的蛋白构建到带有可被链亲和酶素磁珠捕获的 SBP 标记的载体上去, 显著提高筛选效率, 仅需 1 轮筛选即可获得所需 RNA 底物 motif。通过该方法获得了人的 hnRNP A1 的 UP1 结构域特异识别 AGG 和 AG 二种 RNA 序列, 并通过 EMSA 实验证实其可以与获得的 RNA motif 结合。这一方法的建立对于研究 RNA 结合蛋白识别底物的序列特异性, 并进一步了解其在生物体内的调控机制有重要意义。

关键词: SELEX, 高通量测序技术, hnRNP A1, RNA motif

Screening specific recognition motif of RNA-binding proteins by SELEX in combination with next-generation sequencing technique

Lu Zhang, Jinhao Xu, and Jinbiao Ma

State Key Laboratory of Genetics, Department of Biochemistry, School of Life Sciences, Fudan University, Shanghai 200438, China

Abstract: RNA-binding protein exerts important biological function by specifically recognizing RNA motif. SELEX

Received: October 12, 2015; **Accepted:** December 11, 2015

Corresponding author: Jinbiao Ma. Tel: +86-21-51630542; Fax: +86-21-51630541; E-mail: majb@fudan.edu.cn

(Systematic evolution of ligands by exponential enrichment), an *in vitro* selection method, can obtain consensus motif with high-affinity and specificity for many target molecules from DNA or RNA libraries. Here, we combined SELEX with next-generation sequencing to study the protein-RNA interaction *in vitro*. A pool of RNAs with 20 bp random sequences were transcribed by T7 promoter, and target protein was inserted into plasmid containing SBP-tag, which can be captured by streptavidin beads. Through only one cycle, the specific RNA motif can be obtained, which dramatically improved the selection efficiency. Using this method, we found that human hnRNP A1 RRM domain (UP1 domain) bound RNA motifs containing AGG and AG sequences. The EMSA experiment indicated that hnRNP A1 RRM could bind the obtained RNA motif. Taken together, this method provides a rapid and effective method to study the RNA binding specificity of proteins.

Keywords: SELEX, next-generation sequencing, hnRNP A1, RNA motif

SELEX 技术是 Systematic Evolution of Ligands by Exponential Enrichment 的缩写,即指数富集的配体系统进化技术。该技术根据 DNA/RNA 结合蛋白的能力,从随机的 DNA/RNA 文库中找到高亲和力的 DNA/RNA^[1-3]。1990 年, Tuerk 等^[4]应用该技术从人工构建的随机寡核苷酸文库中筛选到能特异性结合噬菌体 T4 DNA 聚合酶的寡核苷酸配体。这项技术自 1990 年发现至今,已经应用于多种不同靶物质的筛选^[5-6]。传统的 SELEX 技术的主要技术流程是:体外化学合成一个寡核苷酸文库,再与靶蛋白混合,形成蛋白-核酸复合物,洗去未与靶蛋白结合的核酸,分离与靶蛋白结合的核酸并扩增,重复多轮的筛选,最后构建克隆文库,挑取 50-100 个克隆测序,分析数据才能获得最适核酸配体。而二代测序技术的出现,因其测序数据通量远远高于一代测序^[7],将其应用于 SELEX 筛选核酸配体的新方法对蛋白-核酸底物的识别提供了更加深入的认识^[8-10]。

本研究主要是基于 SELEX 筛选方法,结合高通量测序技术,研究人的 hnRNP A1 的 RRM 结构域特异识别的 RNA motif, hnRNP A1 是一种核糖核蛋白,重要的转录后调控因子^[11]。其 N 端含有两个 RRM 结构域,RRM1 和 RRM2,也被称为 UP1 (1-196) 结构域,通过 UP1 参与

底物 RNA 的特异性识别^[12], hnRNP A1 参与调控不依赖于 mRNA 的 5'帽子结构的翻译过程^[13], microRNA 的形成^[14]以及可变剪切位点的选择^[15]。已有研究利用 SELEX 筛选技术找出 hnRNP A1 的高亲和力结合位点 UAGGGA/U,且该位点类似于脊椎动物 5'和 3'的剪切位点^[16]。我们将以这一研究成果作为对照,评价 SELEX 与高通量测序技术相结合发现 RNA 结合蛋白特异性识别的 RNA motif 的可行性。

我们在体外合成了带有 20 个随机碱基的 RNA 待筛选文库,将 UP1 蛋白构建到带有可筛选标记的载体中,构建了 pET28a-SUMO-SBP-UP1 的重组质粒,蛋白的纯化过程中切除 His-SUMO 标签,纯化得到带有 SBP-tag 的蛋白,SBP-tag 可以特异结合链亲和酶素,它含有 38 个氨基酸,且其亲和力强于 strep-tag^[17-18]。通过带有链亲和酶素的磁珠来筛选特异结合 SBP-UP1 的 RNA 文库,具有结合 UP1 能力的 RNA 被筛选出来并通过反转录成 cDNA, cDNA 经 10-11 轮的 PCR 扩增,在 PCR 的过程中将测序的 adapter 和 barcode 加在 cDNA 的两端,这样构建出可以用于高通量测序的文库。传统 SELEX 方法需要通过重复多轮的筛选,最后得到的只是亲和力最高的 motif,而且在 PCR 的过程中,会引入一些偏好性,而通过高通量测序方法,可以省去重

复筛选的过程,并且通过数据分析,我们可以得到更多关于转录后调控因子靶位点的信息,而不仅仅是一段 motif^[8-9]。相信这一方法的建立对于今后发现新的转录调控因子的靶作用位点会有重要的意义。

1 材料与方法

1.1 材料

1.1.1 菌株和质粒

大肠杆菌 DH5 α 和 BL21 (DE3) 由本实验室保存。pET28a-SUMO-SBP 载体由本实验室构建。

1.1.2 主要试剂和耗材

限制性内切酶、T4 DNA 连接酶和 ssRNA ladder 购自 New England Biolabs。DNA 凝胶回收试剂盒、PCR 清洁试剂盒购自 Axygen 公司。T7 RNA polymerase 由本实验室保存。5 mL HisTrap 镍柱、5 mL Heparin 肝素亲和层析柱、Superdex75 16/60 凝胶过滤层析柱购自 GE。蛋白分子量 marker 购自 Thermo Scientific。DNA marker 购自 TaKaRa 公司。链霉亲和素偶联的 Dynabeads 和反转录酶购自 Invitrogen。超滤管购自 Millipore。透析袋购自 Spectrum。引物由上海生工生物工程技术有限公司合成。

1.2 重组质粒的构建和蛋白的表达纯化

1.2.1 质粒构建

以 pET28a-SUMO-UP1 (1-196) 重组质粒为模板,设计引物 UP1 F 和 UP1 R,引物两端分

别加 *EcoR* I 和 *Sal* I 酶切位点,扩增 UP1 (1-196) 基因,引物序列见表 1。

将 UP1 (1-196) 亚克隆到 pET28a-SUMO-SBP 质粒,室温连接 3 h。将 pET28a-SUMO-SBP-UP1 (1-196) 重组质粒转化 *Escherichia coli* DH5 α , 挑选阳性克隆送上海睿迪生物公司测序。

1.2.2 蛋白的表达纯化

将测序正确的表达质粒转化 BL21 (DE3), 挑取克隆在 5 mL LB 培养基中,37 °C 培养 6-8 h,再将菌液转移至 1 L LB 培养基中,于 37 °C、200 r/min 振荡培养。当菌液 OD₆₀₀ 达到 0.6-0.8,降温至 18 °C,加入终浓度 0.2 mmol/L IPTG,18 °C、200 r/min 诱导 16 h 后收菌。

菌液重悬于 buffer 1 (20 mmol/L Tris pH 8.0, 500 mmol/L NaCl, 25 mmol/L 咪唑),经高压破菌,17 000 r/min 离心 1 h,留上清,上样至预先用 buffer 1 平衡的 HisTrap 柱,用 buffer 2 (20 mmol/L Tris, pH 8.0, 500 mmol/L NaCl, 500 mmol/L 咪唑) 以梯度洗脱,收集目的峰,加入 Ulp1 酶切 SUMO-tag,并在 20 mmol/L Tris, pH 8.0, 500 mmol/L NaCl 的溶液中透析 3 h,去除咪唑,再将样品上样至预先用 buffer 1 平衡的 HisTrap 柱,收集流穿液,Ulp1 和 6 \times His-SUMO-tag 则结合在 HisTrap 柱上。将蛋白液的盐浓度稀释至 100 mmol/L NaCl 上样到预先用 buffer A (20 mmol/L Tris, pH 8.0, 100 mmol/L NaCl) 平衡的 Heparin 柱,用 buffer B (20 mmol/L Tris,

表 1 PCR 引物序列

Table 1 Primer Sequences

Primer name	Primer sequence (5'-3')	Size (bp)
UP1 F	CGAT <u>G</u> AATTC*ATGTCTAAGTCAGAGTCTCCTAAAGAGCCCG	41
UP1 R	AGCGG <u>T</u> CGAC*TTATCGACCTCTTTGGCTGGATGAAGC	37

*Restriction enzyme cutting site was underlined.

pH 8.0, 1 mol/L NaCl) 梯度洗脱, 收集目的蛋白峰。目的蛋白经超滤浓缩后上样至预平衡的 Superdex 75 16/60, 对纯化获得的蛋白峰样品取样进行 SDS-PAGE 检测。平衡 Superdex 75 16/60 采用的缓冲液是 10 mmol/L Tris, pH 7.5, 100 mmol/L NaCl, 1 mmol/L DTT。

1.3 T7 转录 RNA template

将含有 20 个随机序列的 T7 template 和 T7 promoter (表 2) 均稀释至 100 $\mu\text{mol/L}$, 取等体积混匀, 95 $^{\circ}\text{C}$ 退火 5 min, 参照冷泉港的 T7 转录体系^[19], 37 $^{\circ}\text{C}$ 转录 12 h, 转录结束后加入 DNase I 消化 DNA 模板。利用 16% 的尿素 PAGE 胶纯化 RNA 转录产物。最终获得的随机 RNA 文库浓度用 Nanodrop2000 定量。

1.4 SBP-UP1-RRM 特异识别 RNA 文库的构建

取 250 μL 的结合缓冲液 (25 mmol/L Tris, pH 7.5, 150 mmol/L KCl, 3 mmol/L MgCl_2 , 0.01% Tween20, 1 mmol/L DTT, 30 $\mu\text{g/mL}$ polyI/C) 将 SBP-UP1 稀释至 250 nmol/L, 冰上孵育 30 min。然后加入终浓度为 1 $\mu\text{mol/L}$ 的 RNA 随机文库和 40 U RNase Inhibitor, 冰上孵育 1.5 h。同时, 吸取 100 μL Dynabeads 用清洗缓冲液 (25 mmol/L Tris, pH 7.5, 150 mmol/L

KCl, 0.5 mmol/L EDTA, 0.01% Tween20) 清洗 3 次, 每次 200 μL , 之后将 Dynabeads 平衡于 200 μL 的结合缓冲液中。将 SBP-UP1 和 RNA 文库的混合物, 与 Dynabeads 在室温充分混合 45 min, 再加入 1 mL 的清洗缓冲液, 洗去未结合 SBP-UP1 的 RNA, 然后将 Dynabeads 溶于洗脱缓冲液 (10 mmol/L Tris, pH 7.0, 1 mmol/L EDTA, 1% SDS), 70 $^{\circ}\text{C}$, 10 min。洗脱下来的 SBP-UP1 和 RNA 复合物, 通过苯酚/氯仿/异戊醇 (25:24:1) 抽提和乙醇沉淀, 去除蛋白, 得到被筛选出来 RNA 的文库。特异识别 UP1 蛋白的 RNA 经体外转录和 11 轮的 PCR 扩增, 最终获得可用于二代测序的文库。为了文库最终可用于二代测序, PCR 引物含有可用于二代测序的 adapter 和 barcode。PCR 产物利用 8% Native PAGE 胶纯化。整体实验流程如图 1 所示, RT-PCR 的引物序列见表 2。

1.5 二代测序及数据分析

二代测序文库送至上海晶能生物公司, 在 Illumina HiSeq2500 上测序, 预计每个文库测序得到 2 000 万条序列, 参照 Streaming Kmer Analysis (SKA) 的方法分析数据^[8], 该方法基于多轮读出序列数据, 持续更新相对识别参数,

表 2 SELEX 文库构建核酸序列

Table 2 Oligonucleotides used for SELEX

	Primer sequence (5'-3')	Size (bp)
T7 template	CCTTGACACCCGAGAATTCCA(N)20GATCGTTCGGACTGTAGA ACTCCCTATAGTGA GTCGTATTA	81
T7 promoter	TAATACGACTCACTATAGGG	20
RT primer	GCCTTGGCACCCGAGAATTCCA	22
PR1	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGACGATC	55
RPI5	CAAGCAGAAGACGGCATAACGAGATCACTGT*GTGACTGGAGTTCCTTGGCACCCG AGAATTCCA	63

*The barcode was underlined.

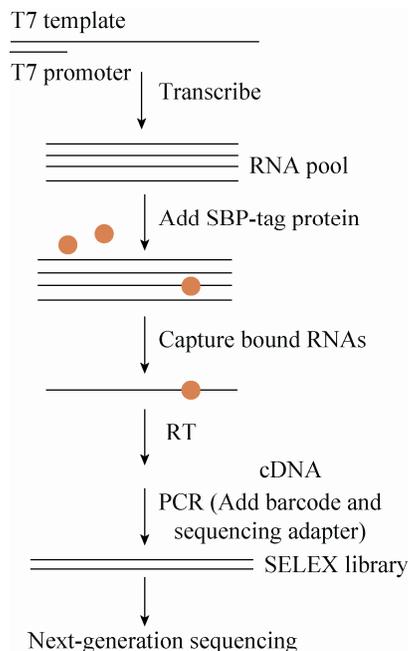


图1 SELEX 实验流程图

Fig. 1 Experimental strategy for SELEX.

将每一个序列概率分配识别到一个特定 kmer motif 上, $k=5-8$ 。最终将每一个 kmer 的权重排序, 得到优先识别的 RNA-motif。相对于传统的寻找 motif 的方法 MEME, SKA 更适合运用到 SELEX 的数据分析中, 因为 SKA 估算的是含有每一个 motif 的序列数, 而不是在一组序列中显著富集的 motif。在本实验中, 考虑到 UP1 含有 2 个 RRM domain, 因此计算 $k=8$ 的优先识别的 RNA-motif。

1.6 SBP-UP1 的体外结合试验

1.6.1 RNA oligo 的化学合成

以 3'-亚磷酰胺及 2'-叔丁基二甲基硅 (TBDMS) 保护的核糖核酸为单体, 以 CPG 为固相合成载体, 在 ABI 394 DNA/RNA synthesizer 化学合成经 SELEX 筛选得到的最优 RNA motif, 合成的 RNA oligo 经氨解和脱保护基, 利用 16%的尿素 PAGE 胶纯化, 最后将纯化得

到的 RNA oligo 用 Nanodrop2000 定量。

1.6.2 凝胶迁移实验 (EMSA)

在 10 μ L 的 RNA oligo 和 SBP-UP1 反应体系中, 底物 RNA 的终浓度固定为 10 μ mol/L, RNA 和 UP1 以 1:0, 1:0.5, 1:1, 1:2 及 1:4 的比例混合, 4 $^{\circ}$ C 反应 30 min, 反应 buffer 采用平衡 superdex75 16/60 的 buffer。反应结束后加入 1 μ L 100%甘油, 进行 Native PAGE 电泳。

2 结果与分析

2.1 SBP-UP1 (1-196) 的表达纯化

将表达 SBP-UP1 的重组质粒转化 BL21 (DE3), 经 IPTG 诱导表达, 高压破菌, 收集破菌上清液。蛋白经过 HisTrap、Heparin 和 Superdex75 纯化而获得, 凝胶阻滞筛峰图显示 SBP-UP1 的出峰体积在 59.9 mL, 根据本实验室 Superdex75 16/60 标准曲线公式, 计算得到分子量为 27 kDa (图 2A), SDS-PAGE (图 2B) 显示重组 SBP-UP1 位于 27 kDa 与理论分子量 (26 180 Da) 相符, 且纯度大于 95%。

2.2 T7 RNA 转录文库的鉴定

含有 20 个随机碱基的 RNA 文库理论大小在 62 bp, 经 16%尿素 PAGE 胶鉴定, T7 体外转录所获得 RNA 文库与理论值相符。

2.3 二代测序文库构建

经链亲和酶素磁珠筛选获得的特异识别 SBP-UP1 的 RNA 文库, 通过 RT-PCR 扩增, 所得 PCR 产物除了含有 20 个随机碱基序列, 随机序列两侧还包含用于二代测序的 adapter 和 barcode, 产物大小为 138 bp, Native PAGE 胶 (图 4) 显示 PCR 产物大小符合理论推测。经 PAGE 胶纯化所得 DNA 文库送上海晶能生物公司测序。

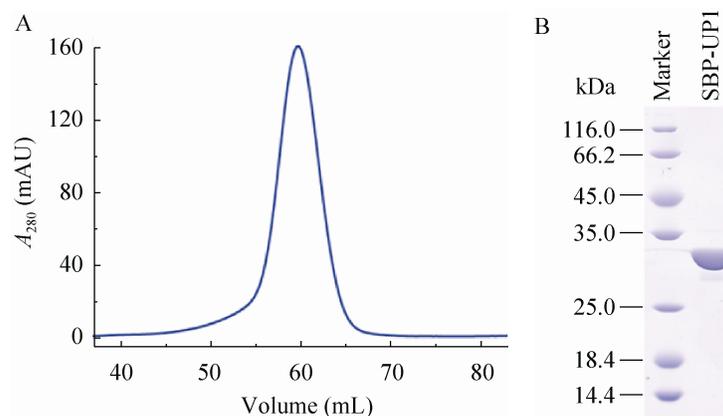


图 2 SBP-UP1 的 Superdex75 16/60 凝胶阻滞峰图 (A) 以及 15% SDS-PAGE (B) 分析

Fig. 2 Gel-filtration of SBP-UP1 performed by Superdex75 16/60(A) and 15% SDS-PAGE analysis of UP1(B).

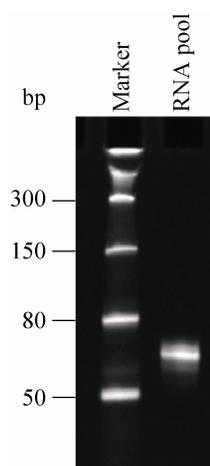


图 3 RNA 随机文库尿素 PAGE 胶分析

Fig. 3 Urea PAGE analysis of RNA random pool.

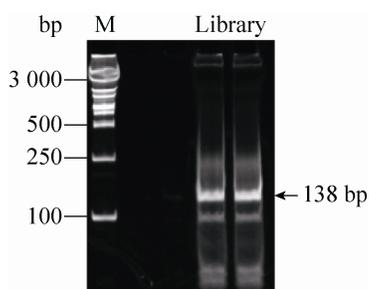


图 4 SBP-UP1 的 SELEX 文库

Fig. 4 SELEX library of SBP-UP1.

2.4 二代测序数据分析

测序数据经去掉两端 adapter, 得到 1 950 万条序列的数据量, 再利用 SKA 算法分析含有 20 个随机序列的测序数据, 设置 $k=8$, 将每一个 8 mer motif 的权重排序, 获得 UP1 结构域最优识别的 RNA motif, 5'-AGGCAGCA-3'。同时, 通过分析 SBP-UP1 优先识别的 8 条 RNA motif (表 3), 发现 AGG 和 AG 二种 RNA 序列出现的频率最高, 在这 8 条 RNA motif 中都存在, 但是这两种 RNA 序列两侧的序列是可变的、不确定的, 因此可以推断 UP1 特异识别 AGG 和 AG 两种保守 RNA 序列, 而其余的序列则有一定的变化。

表 3 SBP-UP1 特异识别 8 mer motif

Table 3 SBP-UP1 specifically binding of 8 mer motif

	8 mer motif	Weight
1	<u>AGGCAGCA</u>	7 281.95
2	<u>AGGUAGCA</u>	7 042.13
3	<u>AGGGAGCA</u>	6 655.34
4	<u>AGGCAGGG</u>	6 046.32
5	<u>AGGCAGUA</u>	5 518.36
6	<u>AGGCAGAA</u>	5 366.69
7	<u>AGGGAGUA</u>	4 567.51
8	<u>AGGAAGCA</u>	4 389.47

The consensus sequence was underlined.

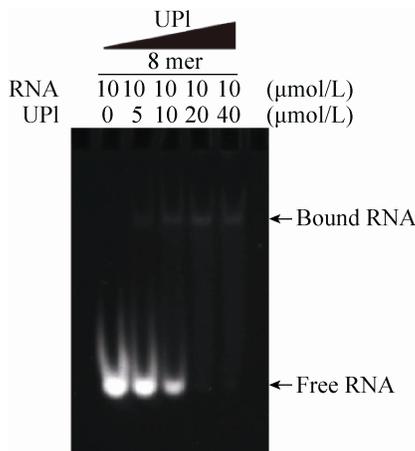


图5 SBP-UP1结合8 mer RNA的EMSA分析结果
Fig. 5 EMSA analysis of the binding between SBP-UP1 and 8 mer RNA.

2.5 EMSA验证SELEX结果

根据SELEX的结果,我们利用亚磷酸胺法化学合成了丰度最高的8 mer的RNA motif: 5'-AGGCAGCA-3', EMSA的结果显示UP1可以在体外结合5'-AGGCAGCA-3',且在UP1 RNA=2 1时, RNA都被UP1结合,几乎看不到 free-RNA, ,因此通过EMSA实验验证了经SELEX结合二代测序技术筛选得到的UP1特异识别RNA motif可以与UP1在体外结合,也说明结合SELEX技术与二代测序技术筛选UP1特异识别RNA motif的方法是可行的。

3 讨论

本研究将SELEX与二代高通量测序技术相结合,发现hnRNP A1的UP1结构域最优识别的RNA motif AGGCAGCA,并通过EMSA实验验证了该 motif与UP1在体外的特异性结合。经过SKA算法的数据分析,我们不仅获得了UP1的最优识别的RNA motif,还获得一些结合力稍弱的RNA motif,但在这些序列中AGG和

AG二种RNA序列出现的频率最高。由于UP1中含有两个连续的RRM结构域,所以我们推断UP1的两个RRM结构域有可能分别倾向于特异性识别AGG和AG,这与通过传统的SELEX方法获得的 motif UAGGGA/U中的核心序列相同^[16]。hnRNP A1蛋白通过识别不同的RNA底物参与microRNA的形成、mRNA的可变剪切等生物学过程^[14-15],其识别底物的序列也会有所差异,之前的研究从20 bp的随机核酸文库中,通过8轮筛选,获得了hnRNP A1最优识别的RNA motif^[16],而通过本研究我们可以发现结合力较弱的RNA motif,但是这些RNA motif的生物学意义还有待进一步的验证。

本研究将UP1蛋白片段插入表达载体pET28a中,构建了重组表达质粒pET28a-SUMO-SBP-UP1。SUMO-tag有利于蛋白的表达纯化,在纯化的过程中可以经Ulp1酶切除,通过HisTrap柱、Heparin柱以及Superdex75的纯化获得纯度在98%以上的SBP-UP1。SBP-tag则是用于SELEX体外筛选底物RNA,SBP-tag与链亲和酶素有着较强的亲和力, $K_d=2.5 \text{ nmol/L}$ ^[20],通过带有链亲和酶素标记的磁珠进行筛选,由于磁珠易于洗涤及浓缩富集的优点,可以提高筛选效率^[21]。传统的SELEX要经过大概10轮左右的筛选才能获得最适的RNA motif^[22],既耗费时间也耗费人力,而我们的方法只需经过1轮筛选,便可得到蛋白特异识别的RNA文库,再经反转录和PCR,在cDNA两端加上用于二代测序的adapter和barcode,即可完成SELEX文库的构建,经二代测序得到的高通量数据,用SKA算法加以分析,我们可得到蛋白特异识别的RNA motif。SKA算法被认为是比MEME更适合于分析SELEX筛选的文库数据,它是通

过累计计算每一条 kmer RNA 在测序数据中的权重,将这些 kmer RNA 的权重进行排序即可得到 RNA motif 的特异性的排序,SKA 可以准确分辨假阳性的结果,提高数据分析的可靠性^[8]。由于高通量测序数据的通量大,信息量丰富,我们不仅可以发现 RNA 结合蛋白的最适底物,还可以发现结合力次优的 RNA motif,因为在生物体内 RNA 结合蛋白可能会通过结合不同的 RNA 发挥不同的功能^[9-10],这对于我们更全面地了解 RNA 结合蛋白的调控机制提供了更多的信息。

SELEX 技术与高通测序技术的结合,使它具有优于其他研究蛋白-RNA 相互作用的研究方法的特点,可以通过测序数据的分析得到蛋白-RNA 相互作用的更多信息。传统的 SELEX 技术,通过多次循环筛选,最后可以得到亲和力最高的靶序列,但是却忽视了较弱的结合靶序列和蛋白 RNA 结合的特异性,且多轮筛选、PCR 扩增均会带入序列偏好性。CLIP,是一种体内筛选技术,它可以真实地反映蛋白与 RNA 相互作用在体内的情况,但是这种技术难度高,且重复的筛选步骤和交联的过程会引入序列的偏好性,后续的分析也很复杂,无法区别是单一蛋白的结合 RNA motif 还是蛋白复合物结合的 RNA motif^[23-24]。研究者通过比较 SELEX 技术和 CLIP 技术发现,SELEX 技术所发现的 motif,在 CLIP 中也有较高的信号。因此,新型的 SELEX 技术获得的结果能够一定程度反映体内的情况,也可以一定程度纠正 CLIP 偏好性,并互补于 CLIP 技术^[8]。

SELEX 与高通量测序技术的结合将为 RNA 结合蛋白识别底物的特异性研究提供快速有效的方法,除了基础生物学研究以外,该方

法还可以应用于更多核糖核酸结合分子识别特异性的研究,如生物分析、疾病诊断、治疗、药物筛选等方面。

REFERENCES

- [1] Ciesiolka J, Gorski J, Yarus M. Selection of an RNA domain that binds Zn^{2+} . *RNA*, 1995, 1(5): 538-550.
- [2] Nitsche A, Kurth A, Dunkhorst A, et al. One-step selection of vaccinia virus-binding DNA aptamers by MonoLEX. *BMC Biotechnol*, 2007, 7: 48.
- [3] Paige JS, Wu KY, Jaffrey SR. RNA mimics of green fluorescent protein. *Science*, 2011, 333(6042): 642-646.
- [4] Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 1990, 249(4968): 505-510.
- [5] Jolma A, Kivioja T, Toivonen J, et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res*, 2010, 20(6): 861-873.
- [6] Katz Y, Wang ET, Airoidi EM, et al. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*, 2010, 7(12): 1009-1015.
- [7] Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, 2008, 9: 387-402.
- [8] Lambert N, Robertson A, Jangi M, et al. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell*, 2014, 54(5): 887-900.
- [9] Campbell ZT, Bhimsaria D, Valley CT, et al. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep*, 2012, 1(5): 570-581.
- [10] Zykovich A, Korf I, Segal DJ. Bind-n-Seq: high-throughput analysis of *in vitro* protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res*, 2009, 37(22): e151.

- [11] Dreyfuss G, Matunis MJ, Pinol-Roma S, et al. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem*, 1993, 62: 289–321.
- [12] Shamoo Y, Krueger U, Rice LM, et al. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol*, 1997, 4(3): 215–222.
- [13] Lin JY, Shih SR, Pan MJ, et al. hnRNP A1 interacts with the 5' untranslated regions of enterovirus 71 and Sindbis virus RNA and is required for viral replication. *J Virol*, 2009, 83(12): 6106–6114.
- [14] Guil S, Cáceres JF. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol*, 2007, 14(7): 591–596.
- [15] Tange TØ, Damgaard CK, Guth S, et al. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J*, 2001, 20(20): 5748–5758.
- [16] Burd CG, Dreyfuss G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J*, 1994, 13(5): 1197–1204.
- [17] Barrette-Ng IH, Wu SC, Tjia WM, et al. The structure of the SBP-Tag-streptavidin complex reveals a novel helical scaffold bridging binding pockets on separate subunits. *Acta Crystallogr D Biol Crystallogr*, 2013, 69(Pt 5): 879–887.
- [18] Keefe AD, Wilson DS, Seelig B, et al. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. *Protein Expr Purif*, 2001, 23(3): 440–446.
- [19] Nilsen TW, Rio DC, Ares M Jr. High-yield synthesis of RNA using T7 RNA polymerase and plasmid DNA or oligonucleotide templates. *Cold Spring Harb Protoc*, 2013, 2013(11), doi: 10.1101/pdb.prot078535.
- [20] Wu SC, Wong SL. Structure-guided design of an engineered streptavidin with reusability to purify streptavidin-binding peptide tagged proteins or biotinylated proteins. *PLoS ONE*, 2013, 8(7): e69530.
- [21] Oh SS, Ahmad KM, Cho M, et al. Improving aptamer selection efficiency through volume dilution, magnetic concentration, and continuous washing in microfluidic channels. *Anal Chem*, 2011, 83(17): 6883–6889.
- [22] Nieuwlandt D, Wecker M, Gold L. *In vitro* selection of RNA ligands to substance P. *Biochemistry*, 1995, 34(16): 5651–5659.
- [23] Sugimoto Y, König J, Hussain S, et al. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol*, 2012, 13(8): R67.
- [24] Sugnet CW, Srinivasan K, Clark TA, et al. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput Biol*, 2006, 2(1): e4.

(本文责编 陈宏宇)