

大肠杆菌基因组水平蛋白质-RNA 相互作用初步研究

徐 淞¹, 陈珪文¹, 应晓敏¹, 付汉江², 田宝磊², 宋 宜², 郑晓飞², 李伍举¹

[摘要] 目的 初步研究大肠杆菌中基因组水平的蛋白质-RNA 相互作用 (protein-RNA interactions, PRI)。方法 通过RNA 酶消化细菌裂解液, 提取与蛋白质相互作用的 RNA 片段, 构建 cDNA 文库, 进行高通量测序, 并通过生物信息学分析获得与蛋白质结合的转录本。结果 获得了与蛋白质结合的 3193 条转录本, 涉及 2234 个 mRNA、47 个 sRNA (small regulatory RNAs)、39 个 tRNA、11 个 rRNA 以及 862 个基因间区 (intergenic region, IGR)。结论 初步获得大肠杆菌中与蛋白质相互作用的转录本信息, 为进一步开展 PRI 研究提供了支持。

[关键词] 蛋白质-RNA 相互作用; PRI; sRNA; 大肠杆菌; MCA

[中图分类号] Q75;R341

[文献标志码] A

[文章编号] 1674-9960(2014)--

DOI: 10.7644/j.issn.1674-9960.2014.00.000

Protein-RNA interactions in *Escherichia coli*: a genome-wide study

XU Song¹, CHEN Yao-wen¹, YING Xiao-min¹, FU Han-jiang², TIAN Bao-lei², SONG Yi², ZHENG Xiao-fei^{2*}, LI Wu-ju^{1*}

(1. Institute of Basic Medical Sciences, Academy of Military Medical Sciences, Beijing 100850, China; 2. Institute of Radiation Medicine, Academy of Military Medical Sciences, Beijing 100850, China)

* Co-corresponding authors, LI Wu-ju, E-mail: liwj@nic.bmi.ac.cn; ZHENG Xiao-fei, E-mail: zhengxf@nic.bmi.ac.cn

[Abstract] **Objective** To conduct a pilot study on genome-wide *in vivo* protein-RNA interactions in *E. coli*. **Methods** Bacterial lysate was treated with RNase before the RNA fragments protected by proteins were extracted from treated lysate and used to construct cDNA library that was applied to high-throughput sequencing. Finally, the transcripts bound by proteins were obtained by bioinformatics analysis. **Results** A total of 3193 transcripts were obtained, including 2234 mRNAs, 47 sRNAs, 39 tRNAs, 11 rRNAs, and 862 intergenic regions. **Conclusion** Some information of transcripts interacting with proteins in *E. coli* is acquired, which will facilitate further studies of protein-RNA interactions.

[Key words] protein-RNA interactions; PRI; sRNA; *E. coli*; MCA

蛋白质-RNA 相互作用 (protein-RNA interactions, PRI) 在基因转录、转录后修饰和翻译等多个生物过程中发挥重要作用。例如, 大肠杆菌 sRNA SsrS, 通过折叠成特定的二级结构模仿 DNA 与 σ^{70} -RNA 聚合酶结合, 从而影响多种依赖 σ^{70} 的靶基因表达^[1]。此外, 细菌 sRNA CsrB 通过与 CsrA 等靶蛋白的结合调控碳的摄入、细胞运动、生物膜形成、群落感应与细菌致病性等^[2]。因此开展 PRI 研究对理解蛋白质和 RNA 功能具有重要意义。

目前研究 PRI 的实验方法主要有两类^[3]: 一类

是以 RNA 为中心, 利用合成的 RNA 钓取相关蛋白质, 得到与给定 RNA 相互作用的蛋白质信息; 另一类是以蛋白质为中心, 通过紫外交联细胞, 结合免疫共沉淀技术得到与蛋白质结合的 RNA, 并对 RNA 进行测序, 再利用生物信息学方法对 PRI 进行分析。当前最常用的方法是以蛋白质为中心, 主要有紫外交联免疫共沉淀方法 (crosslinking and immunoprecipitation, CLIP)^[4]、单个核苷酸分辨率 CLIP (individual-nucleotide resolution CLIP, iCLIP)^[5] 和光活性增强核糖核苷 CLIP (photoactivatable ribonucleoside-enhanced CLIP, PAR-CLIP)^[6]。这 3 种策略均是以特定的感兴趣蛋白质为基础, 通过抗体钓出被蛋白质保护的 RNA 片段, 最后通过高通量测序分析获得与特定蛋白质结合的转录本信息。然而, 一个细菌基因组通常可以表达几千种蛋白, 如果使用这种方法获得每个蛋白质的转录本, 则需要耗费大量人力和物力。为此, 我们以模式生物大肠杆菌为例,

[基金项目] 国家 973 计划资助项目 (2010CB912801); 国家自然科学基金资助项目 (31271404, 31071157)

[作者简介] 徐 淞, 男, 硕士研究生, 研究方向: 生物信息学

[作者单位] 1. 军事医学科学院基础医学研究所, 北京 100850; 2. 军事医学科学院放射与辐射医学研究所, 北京 100850

[通讯作者] 李伍举, E-mail: liwj@nic.bmi.ac.cn; 郑晓飞, E-mail: zhengxf@nic.bmi.ac.cn

结合高通量测序技术,探讨了基因组水平的 PRI 研究策略。首先,通过 RNA 酶消化细菌裂解液中未和蛋白质相互作用的 RNA,得到被蛋白质保护的 RNA 片段;然后以这些 RNA 为基础,构建 cDNA 文库并进行高通量测序,并通过综合已有的方法和自行开发的生物信息学软件对数据进行分析,初步获得了大肠杆菌中基因组水平的 PRI 信息。

1 材料与方法

1.1 材料

1.1.1 菌株 大肠杆菌 K-12 MG1655 由军事医学科学院生物工程研究所王恒樑教授惠赠。

1.1.2 主要试剂 溶菌酶购自 Amresco 公司; RNaseA、DNA 酶、T4 多核苷酸激酶(T4 PNK)购自 TaKaRa 公司;热敏感碱性磷酸酶(FastAP™)、RNaseT1 购自 Thermo Scientific 公司;RNAsafe、蛋白酶 K 购自北京天根生化有限公司;胶回收小 RNA 试剂盒购自 ZYMO Research 公司;TruSeq small RNA Sample pre Kit 购自 Illumina 公司;高通量测序由北京诺禾致源生物信息科技有限公司完成。

1.2 方法

1.2.1 紫外交联细菌 用 LB 培养基在 37℃ 过夜培养 *E. coli* K-12 MG1655 至 D_{600} 值为 1.5,即稳定期。离心后用 PBS 重悬细菌至浓度为 2.4×10^{10} 个/ml,将菌液平铺到平皿中,厚度不超过 5 mm,置于冰上,400 mJ/cm² 紫外交联 1 次,取出间隔 1 min,200 mJ/cm² 交联 1 次。照射完成后立即收集细菌。

1.2.2 裂解细菌及 RNA 酶消化 加入终浓度为 1 mg/ml 溶菌酶、0.5% SDS,64℃ 反应 2 min,冰上冷却,4℃ 离心 10 min,取上清,加入终浓度为 2000 U/ml RNaseT1,200 μg/ml RNaseA,37℃ 反应 1 h。加入 RNAsafe 60℃ 反应 20 min。

1.2.3 提取 RNA 加入终浓度为 400 μg/ml 蛋白酶 K、1 mmol/L 的 EDTA,55℃ 反应 1 h。冰上冷却后,加入等体积苯酚:氯仿(1:1)混合物,4℃ 12 000 r/min 离心 10 min,取上清。重复上一个步骤 1 次。加入等体积氯仿,混匀,4℃ 12 000 r/min 离心 10 min,取上清。加入 1/10 体积 3 mol/L 乙酸钠(pH 5.2),等体积异丙醇,-20℃ 沉淀 12 h,4℃ 12 000 r/min 离心 15 min。75% 乙醇洗 1 次,4℃ 7500 × g 离心 5 min。弃去乙醇,晾干 RNA,加 DEPC 处理过的水溶解。DNA 酶处理 1 h,水饱和酚抽提 RNA,乙醇沉淀。

1.2.4 聚丙烯酰胺凝胶电泳(PAGE)分离 RNA 片段 配置 7 mol/L 尿素、10% 聚丙烯酰胺凝胶,凝固后 200 V 预电泳 30 min。RNA 样品加入上样缓冲

液后 90℃ 变性 2 min,上样后 120 V 电泳 10 min,200 V 电泳 20 min。取出聚丙烯酰胺凝胶,用水漂洗 1 次,EB 染色 6 min,紫外灯下切下 20 ~ 50 nt 大小的 RNA 条带,按照小 RNA 回收试剂盒说明书回收 RNA 片段。

1.2.5 RNA 两端修饰处理 FastAP™ 处理 30 min,水饱和酚抽提 RNA。T4 PNK 处理 1 h,水饱和酚抽提 RNA,离心完成后可见白色沉淀,弃去液体,用 75% 乙醇洗 2 次,4℃ 7500 × g 离心 5 min。风干,加水溶解,可于 -80℃ 冻存。

1.2.6 构建 cDNA 文库以及测序 使用 TruSeq small RNA Sample pre Kit 建库。简言之,根据 RNA 结构上的特征(5'端有完整的磷酸基团,3'端有完整的羟基),对质量合格的 RNA 样品进行接头连接,完成两端连接后的产物经逆转录、PCR 扩增和 PAGE 胶回收后即得到 cDNA 文库。成功构建的文库通过 Illumina Hiseq2500 进行单端 SE50 测序。

1.2.7 数据分析 对数据运用两种不同的策略进行分析,一种是我们提出的基于突变的簇相关分析方法(mutation-based cluster association study, MCA),一种是 Wang 等^[7]开发的 dCLIP 软件。

MCA 分析:对测序获得的原始读段,去除序列标签获得干净读段(clean reads),采用 Langmead 等^[8]所述方法将干净读段比对到 *E. coli* 基因组上(NCBI code:NC_000913),在比对时,仅保留单一标注(unique mapping)读段,即如果一个读段标注到基因组 ≥ 2 个位置,这样的读段不予保留。另外,由于紫外交联易引起读段中某些碱基突变,所以在比对时,容许一个读段至多有 3 个碱基突变。在获得单一标注读段之后,根据每个读段在基因组上的位置,生成由位置重叠读段构成的簇。

为了获得与蛋白质强烈结合的转录本,对每一个来自实验组的簇,在对照组中寻找其基因组位置与之重叠的相关簇,然后,对这样的 2 个簇,计算每一列的碱基突变率,于是获得与 2 个簇对应的突变向量,如果这样的 2 个簇,其基因组位置没有完全一致,则对应的突变向量位置以零补齐,最终获得 2 个具有相同基因组位置的突变向量。在此基础上进行秩和检验,并基于 *P* 值计算相应的 FDR 值,这里我们以 FDR < 0.1 来选择具有显著性差异的簇作为与蛋白强烈结合的簇。分析流程如图 1B 所示。

dCLIP 分析:首先将 dCLIP 通过 <http://qbr.cswmed.edu/software/> 下载至本地,采用参数 $m1 = 5$, $m2 = 5$ (表示 2 个条件下均只考虑 ≥ 5 个读段形成的簇), $mut = 'all'$ (表示不筛选特定突变),其余参数均为默认,按照 dCLIP 流程对数据进行分析。

2 结果

2.1 cDNA 文库构建流程

为了开展基因组水平的 PRI 研究,我们建立了一个简单快速的实验流程(图 1A)。设置了一组进行紫外交联的实验组和未经紫外交联的对照组,除了紫外交联,后续步骤完全相同。在加入蛋白酶消化前,使用 RNA 酶灭活剂去除 RNA 酶活性,从而避免在使用蛋白酶 K 消化蛋白质的同时,残留的 RNA 酶继续消化被蛋白质保护的 RNA 片段。为了避免 DNA 污染 RNA 样品,提取的 RNA 用 DNA 酶处理,然后电泳回收 RNA 片段。由于使用的 RNA 酶消化产生的 RNA 片段 5'端是 OH,3'端是 PO_4 ,不能连接接头构建 cDNA 文库,所以需要先去除了 3'端 PO_4 ,5'端加上 PO_4 。PNK 催化活性受到缓冲液中的 pH 值影响,在给定的 PNK 缓冲液中(pH 8.0)通常只能处理 5'端,因此在 PNK 处理前需用 FastAP™ 去除 3'端 PO_4 ,然后才能按照文库构建试剂盒构建 cDNA 文库。

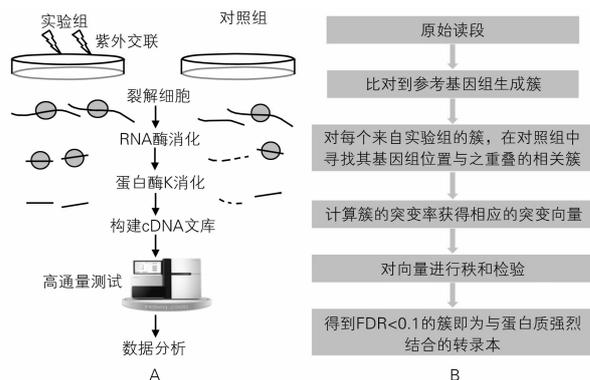


图 1 研究方法示意图

A. 实验流程; B. 分析流程

2.2 高通量测序结果概况

在文库质量检测合格后,对文库进行测序,通过去接头和质量控制等预处理后,实验组和对照组分别获得了 12 028 608 条和 11 106 770 条读段。运用 Langmead 等的方法将干净读段定位于参考基因组,在比对时,要求读段在参考基因组上具有唯一位置,并容许 3 个错配,主要原因是考虑到紫外交联易引起读段中碱基的突变,其次是拟通过突变分析找出实验组与对照组中具有显著差异的簇。如果容许突变数目较少,则通过突变分析找出的转录本数目就较少。事实上,如果在比对过程中不容许错配,则不可能利用突变分析找出与蛋白结合的转录本。结果表明,对实验组,可以比对到参考基因组上的读段为 11 896 126 条,占整个读段数的 98.9%,其中具有唯一定位的读段为 799 229 条,占整个读段数的

6.6%。对于对照组,可以比对到参考基因组上的读段为 10 962 375 条,占整个读段数的 98.7%,其中具有唯一定位的读段为 862 516 条,占整个读段数的 7.8%。由此可见,无论是实验组还是对照组,均有近 99.0%的读段可以比对到参考基因组上,表明测序样本的确是来自大肠杆菌 MG1655 株。其次,有大量读段由于可以比对到基因组上的多个位置而被丢弃,这种现象主要是由于读段长度较短(实验组与对照组均在 15 个核苷酸左右)而同时容许 3 个错配造成。最终获得的来自实验组的 799 229 条读段和来自对照组的 862 516 条读段,在基因组上的分布情况如图 2 所示。

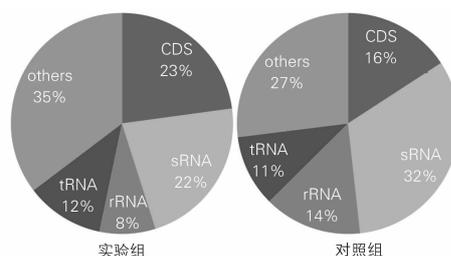


图 2 读段在基因组上的分布图

Others 包括未注释的读段、基因间区(IGR)以及与 2 个基因交叉的读段

2.3 基于 MCA 方法获得与蛋白质结合的转录本

经过高浓度 RNA 酶消化后,从理论上讲,经过紫外交联的实验组中留下的转录本是细胞内与蛋白质相互作用更为真实的转录本。对照组中虽然没有采用紫外交联将蛋白质与 RNA 固定,但是细菌体内本身就存在一部分相互作用比较强烈的蛋白质-RNA 复合物,即使在细菌被裂解后也不一定会分离,因此我们将同时出现在实验组和对照组中(在两组数据中序列有差异的读段数都 > 5)的转录本列为候选的与蛋白质相互作用的转录本。经过 MCA 分析得到了与蛋白质结合的 2421 条转录本,其中有 1787 个 mRNA、29 个 sRNA、33 个 tRNA、11 个 rRNA 以及 561 个 IGR。由于 sRNA 在细菌转录后调控过程中有重要作用,并且目前已有数据库收集了已发表的 sRNA 与蛋白质相互作用,因此这里我们将 MCA 结果中的 sRNA 与已有数据进行了比较,发现有 25 个 sRNA 已报道与蛋白质相互作用^[9-11],4 个 sRNA 目前还未见到与蛋白质相互作用的报道,如表 1 所示。

另外,我们对 MCA 中数据进行严格筛选,选择满足条件(FDR < 0.1,并且实验组碱基平均突变率大于对照组)的转录本,结果得到 221 个蛋白质结合的簇,涉及 60 条转录本,包括 20 个 mRNA、12 个 sRNA、9 个 tRNA、3 个 rRNA 以及 16 个 IGR。仍然

选择 sRNA 来考察结果,发现 12 个 sRNA 中 10 个是已报道与蛋白质存在相互作用,有 2 个未报道(RyD 和 SymR)(表 1)。这个结果表明,可能由于紫外交联效率比较低,导致严格筛选后结果数量减少,但是仍然有 2 个新的候选 sRNA,并且在 Kawano 等^[12]的报道中,RyD 在他们所有的培养条件下都高表达,SymR 同样也是在所有培养条件以及在 SOS 效应中都有表达,说明这 2 个 sRNA 对细菌生长有重要功能,因此可以用作后续实验验证。

表 1 MCA 分析结果中的 sRNA 与已知数据比较情况

报道情况	sRNA-蛋白质	参考文献
已报道	CsrB-CsrA, CsrC-CsrA, DicF-Hfq, DsrA-Hfq, Fls-Fli, MicF-Hfq, RnpB-RnpA, RprA-RprA, RyeA-Hfq, RybB-Hfq, SsrS-RNAP, Spf-Hfq, SsrA-SspB, GadY-Hfq, GlnY-Hfq, McaS-Hfq, PsoO-Hfq, RydB-Hfq, RyjA-Hfq, SgrS-Hfq, SibB-Hfq, Tzf-Hfq, SokE-RNase III, SokB-RNase III, SokC-RNase III	[9] [10] [11]
未报道	RyD, RyjB, SymR, SroH	

加粗部分为严格筛选结果;斜体部分为 MCA 与 dCLIP 比较后 MCA 中独有的 sRNA

2.4 通过 dCLIP 软件获得与蛋白质结合的转录本

为了充分挖掘由测序数据得到的与蛋白质结合的转录本信息,我们还使用 dCLIP 软件对数据进行分析。dCLIP 是一个用来比较不同 CLIP-seq 数据的计算方法,能够有效地确定蛋白质结合区域。它包括 2 个步骤:首先使用一个修改的 MA-plot 方法对来自不同数据集的 CLIP-seq 数据进行标准化,以获取更精细的结果;其次使用一个隐马尔可夫模型(Hidden Markov Model, HMM)来检测不同条件下的共同与不同的蛋白质-RNA 结合位点。在数据支持方面,dCLIP 可以输入多种 CLIP 类型的数据,包括 CLIP, PAR-CLIP 和 iCLIP。在精度方面,dCLIP 优于另一个蛋白质结合位点识别的软件 Piranha^[16]。我们使用 dCLIP 对我们的数据进行分析,最终得到了 2455 条转录本,含有 1763 个 mRNA、43 个 sRNA、10 个 rRNA、35 个 tRNA 以及 604 个 IGR,与 MCA 结果的比较情况如图 3 所示。这里仍然以 sRNA 为例来考察 dCLIP 的结果。在获得的 43 个 sRNA 中,其中有 25 个 sRNA 与 MCA 方法获得的结果相同(21 个已报道与蛋白质结合,4 个未报道)。在余下独特的 18 个 sRNA 中,其中有 13 个 sRNA 已知与蛋白质结合^[10,13,14],有 5 个 sRNA 目前还没有报道与蛋白质相互作用(表 2)。MCA 独有的 4 个 sRNA(表 1)都已报道与蛋白质存在相互作用。

最后,我们选择 MCA 和 dCLIP 两种方法结果的并集作为最终与蛋白质结合的转录本,共得到了 3193 条,包含 2234 个 mRNA、47 个 sRNA、11 个 rRNA、39 个 tRNA 以及 862 个 IGR。由于细菌 sRNA 常位于

IGR,所以,我们将此 862 个 IGR 与 Shinhara 等^[15]构建的大肠杆菌 sRNA 预测数据库进行比对,结果表明,有 178 个 IGR 可能是新的 sRNA。

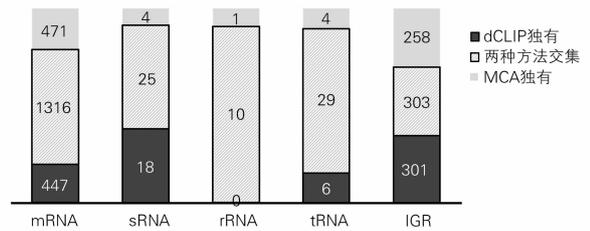


图 3 MCA 与 dCLIP 结果比较

表 2 dCLIP 独有结果中的 sRNA 与已知数据比较情况

报道情况	sRNA-蛋白质	参考文献
已报道	ArcZ-Hfq, CyaR-Hfq, GlnZ-Hfq, RyhB-Hfq, SibA-Hfq, SibD-Hfq, MicM-Hfq, CevB-Hfq, MicA-Hfq, SibC-Hfq, OmrA-Hfq, SdsR-Hfq, IstR-RNase III	[10] [13] [14]
未报道	ArrS, EsrE, RdlC, SibE, RdlA	

3 讨论

综上所述,本研究基于高通量测序初步探讨了大肠杆菌基因组水平的 PRI,所采用策略将为开展其他物种基因组水平的 PRI 提供借鉴。在进行本研究的同时,我们也注意到 Silverman 等^[17]最近在 HeLa 细胞中开展的基因组水平 PRI 研究。两者所采用策略基本相同。首先通过紫外交联和 RNA 酶消化裂解液等处理,获得与蛋白结合的 RNA 片段;其次是构建 cDNA 文库和高通量测序;最后利用生物信息学分析获得基因组水平上与蛋白相互作用的转录本。然而,据我们所知,这是首次原核系统基因组水平的 PRI 研究。

在进行高通量测序数据分析时,我们分别运用了本实验室提出的 MCA 方法和 Wang 等开发的 dCLIP 软件,分别获得了 2421 条和 2455 条转录本。dCLIP 方法是通过 HMM 比较分析实验组和对照组的差异区段,从而得到蛋白质结合位点,而 MCA 方法在考虑读段分布基础上还考虑了突变信息,两种方法虽然使用的策略不同,但是获得的结果有近 70.00% 重叠,表明 MCA 方法具有一定合理性。为了进一步验证 MCA 性能,我们将在更多数据集上测试 MCA 方法。另外,通过与已知的蛋白-sRNA 相互作用数据比较,发现在获得的 sRNA 列表中,与蛋白结合的 sRNA 分别占 86.2% (MCA) 和 79.1% (dCLIP),表明两种方法均具有较高的阳性检出率。最后,我们发现,经 MCA 严格筛选的转录本数量较少(FDR < 0.1),这可能是由于紫外交联效率较低引起。所以本文采用两种策略预测结果的并集,最终得到 3193

条与蛋白直接结合的转录本。通过与大肠杆菌 sRNA 预测数据库比较发现,有 178 条转录本可能是与蛋白质结合的新 sRNA。在后续工作中,我们将进一步证实其中部分 sRNA,并找出与之结合的蛋白。目前研究表明,细菌 sRNA 在与靶标 mRNA 相互作用时,通常需要 Hfq 等伴侣蛋白的参与。为此,我们将以新发现的与蛋白质结合的 sRNA 为基础,钓出与之结合的蛋白,将为发现新的伴侣蛋白提供可能。

最后,我们在基于 MCA 的分析过程中发现,有许多读段只有 1~2 个拷贝。例如,我们的测序数据共涉及 60 个 sRNA,然而由于许多 sRNA 拷贝数较少不能提供足够信息,最终只获得 29 个与蛋白结合的 sRNA,这也暗示我们,如果增加测序深度获得更多数据,或者进一步改进分析方法提高数据的利用率(本研究分析的唯一读段数占有比到基因组上的读段数比例 < 10%),我们将可能得到更多与蛋白质结合的转录本信息,从而为蛋白质-RNA 相互作用研究提供支持。

【参考文献】

[1] Wassarman KM. 6S RNA: a regulator of transcription[J]. *Mol Microbiol*, 2007, 65(6): 1425 - 1431.

[2] Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins[J]. *Curr Opin Microbiol*, 2007, 10(2): 156 - 163.

[3] McHugh CA, Russell P, Guttman M. Methods for comprehensive experimental identification of RNA-protein interactions[J]. *Genome Biol*, 2014, 15(1): 203.

[4] Wang Z, Tollervey J, Briese M, et al. CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins *in vivo*[J]. *Methods*, 2009, 48(3): 287 - 293.

[5] Konig J, Zarnack K, Rot G, et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution[J]. *Nat Struct Mol Biol*, 2010, 17(7): 909 - 915.

[6] Hafner M, Landthaler M, Burger L, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP[J]. *Cell*, 2010, 141(1): 129 - 141.

[7] Wang T, Xie Y, Xiao G. dCLIP: a computational approach for comparative CLIP-seq analyses[J]. *Genome Biol*, 2014, 15(1): R11.

[8] Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome[J]. *Genome Biol*, 2009, 10(3): R25.

[9] Yuan J, Wu W, Xie C, et al. NPInter v2.0: an updated database of ncRNA interactions[DB]. *Nucleic Acids Res*, 2014, 42(1): D104 - D108.

[10] Zhang A, Wassarman KM, Rosenow C, et al. Global analysis of small RNA and mRNA targets of Hfq[J]. *Mol Microbiol*, 2003, 50(4): 1111 - 1124.

[11] Franch T, Thisted T, Gerdes K. Ribonuclease III processing of coaxially stacked RNA helices[J]. *J Biol Chem*, 1999, 274(37): 26572 - 26578.

[12] Kawano M, Reynolds AA, Miranda-Rios J, et al. Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*[J]. *Nucleic Acids Res*, 2005, 33(3): 1040 - 1050.

[13] Wassarman KM, Repoila F, Rosenow C, et al. Identification of novel small RNAs using comparative genomics and microarrays[J]. *Genes Dev*, 2001, 15(13): 1637 - 1651.

[14] Vogel J, Argaman L, Wagner EGH, et al. The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide[J]. *Curr Biol*, 2004, 14(24): 2271 - 2276.

[15] Shinohara A, Matsui M, Hiraoka K, et al. Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*[J]. *BMC Genomics*, 2011, 12: 428.

[16] Uren PJ, Bahrami-Samani E, Burns SC, et al. Site identification in high-throughput RNA-protein interaction data[J]. *Bioinformatics*, 2012, 28(23): 3013 - 3020.

[17] Silverman IM, Li F, Alexander A, et al. RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome[J]. *Genome Biol*, 2014, 15(1): R3.

(孙承媛 编辑 2014-04-01 收稿)