

DOI: 10.13376/j.cblls/2014031

文章编号: 1004-0374(2014)03-0202-05



付向东 博士

付向东教授主要研究前体 RNA 剪接机制, 人类基因中的功能 RNA 单元, 以及发育和疾病中的 RNA 调节机制。

在哈佛大学博士后研究期间, 他大胆尝试了用部分纯化的 RNA 剪接免疫小鼠产生单克隆抗体来鉴定必需的剪接因子, 因此发现了对于哺乳动物细胞前体 mRNA 组成型剪接和可变剪接都至关重要的 SR 家族的创始成员 SC35。

在加州大学圣地亚哥分校建立了自己的实验室后, 付教授的研究小组不仅揭示了 SR 蛋白是作为第一类 RNA 结合蛋白呈递新生 RNA 进入剪接过程, 还发现了第一个 SR 蛋白特异性的激酶, 并证实它们是传递胞外信号到细胞核进行 RNA 剪接的关键信号分子。

付教授的团队一直处在发展基因组技术的研究前沿, 发明了第一个灵敏的 RNA 剪接异构体的检测平台。他们基于下一代测序技术的创新为基础和转化医学研究提供了新的概念和技术平台。

最近, 付教授的研究小组发现了一个关键的分子信号通路能够直接将人的成纤维细胞转变成神经细胞。这一新方法可能被直接用于神经组织退化疾病的治疗。

非编码RNA研究技术概述

苟兰涛¹, 刘默芳¹, 付向东^{2*}

(1中国科学院上海生命科学研究院生物化学与细胞生物学研究所, 分子生物学国家重点实验室, 上海 200031;
2 Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093-0651, USA)

摘要: 非编码 RNA 是当前生命科学研究领域的前沿热点。系统地研究生物体内非编码 RNA 的起源、结构、功能和作用机制, 有助于我们更加全面、深入地了解基因表达调控机理及生物体的生命活动, 而科学问题的解答依赖于实验技术和实验方法的创新和革命。本文按不同的实验目的和研究内容, 就当前非编码 RNA 领域内主要的研究策略和方法技术作一简述。

关键词: 非编码 RNA; 交联; 免疫沉淀; 高通量测序; RNA 结合蛋白

中国分类号: Q-33 **文献标志码:** A

Unique approaches to unique problems in ncRNA research

GOU Lan-Tao¹, LIU Mo-Fang¹, FU Xiang-Dong^{2*}

(1 State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; 2 Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093-0651, USA)

Abstract: Non-coding RNAs have been the hot frontier of life sciences in recent years. Systematic studies of the origins, structures, biological function and their mechanisms of non-coding RNAs will provide new avenues to help

收稿日期: 2014-02-17

*通信作者, E-mail: xdfu@ucsd.edu

us understand the complicated mechanisms of gene regulation as well as cellular activities. Since scientific breakthrough highly relies on technology innovations, here we summarize the main strategies and techniques in non-coding RNA research.

Key words: non-coding RNA; cross-linking; immunoprecipitation; high through sequencing; RNA binding protein

“人类基因组计划”研究表明, 大部分基因组都可以被转录生成 RNA, 但是绝大多数的 RNA 分子并不编码蛋白质。越来越多的证据显示, 这部分非编码 RNA 在基因表达的各个层面均发挥重要调控功能。当今, 研究领域内主要利用免疫沉淀、紫外或多聚甲醛交联等手段, 结合第二代高通量测序技术, 从生物体全基因组水平研究非编码 RNA 及其相互结合的 DNA、蛋白质。以下按照不同研究角度分类进行详述。

1 从整体水平检测基因组内的“结合事件”

目前已根据不同的实验目的发展了相应的研究策略和方法, 从整体水平研究细胞内的“结合事件”。应用较为普遍的主要有 ChIP-seq (chromatin immunoprecipitation-high throughput sequencing) 和 CLIP-seq (crosslinking- immunoprecipitation- high throughput sequencing) 技术; 利用这两种实验技术, 可以在基因组整体水平分别对“蛋白质与 DNA”、“蛋白质与 RNA”的相互作用进行研究。此外, CLASH (cross-linking, ligation and sequencing of hybrids) 技术可用于高通量鉴定“RNA 与 RNA”的相互作用, 而通过 ChIRP(chromatin isolation by RNA purifications) 技术能够获得“大分子 RNA 与染色质片段”的相互作用信息。

ChIP-seq 是将染色质免疫共沉淀技术 (ChIP) 与第二代高通量测序技术相结合的研究系统, 可高效地检测全基因组内与组蛋白、转录因子等特定蛋白因子相互结合的 DNA 区域^[1]。其主要实验步骤为: 首先通过甲醛 (formaldehyde)(有时使用 UV 或戊二醛) 处理细胞, 将目的蛋白质与 DNA 交联固定; 随后用超声将基因组 DNA 打断, 并利用目的蛋白质抗体进行免疫共沉淀; 最后通过蛋白酶处理降解蛋白质, 分离获得与目的蛋白交联的 DNA 片段, 然后在其两端加上接头 (linker), 进行高通量测序分析。

CLIP-seq 也被称为 HITS-CLIP, 是将紫外交联免疫共沉淀技术 (CLIP) 与高通量测序技术相结合的研究系统, 可在全基因组水平鉴定与特定 RNA 结合蛋白相互作用的 RNA 分子。其主要实验流程

为: 首先通过紫外照射将细胞内的 RNA 分子与 RNA 结合蛋白进行耦联; 随后将细胞裂解, 以目的蛋白质抗体进行免疫共沉淀, 利用核酸酶 (RNase) 将共沉淀的 RNA 切割, 以保留与目的蛋白质直接相互作用的 RNA 片段; 随后, 通过 5' 末端放射性标记 RNA 分子及 SDS-PAGE 电泳, 选择性地回收目的蛋白结合的 RNA 片段; 最后, 经过降解蛋白质、添加接头及 RT-PCR 等步骤后, 以高通量测序对目的蛋白质结合的 RNA 片段进行分析^[2-3]。此外, 通过对经典 CLIP-seq 的实验流程进行局部调整、修改或增加, 还衍生了许多变异的 CLIP-seq 方法, 例如增强了交联强度的 PAR-CLIP 和能够精确鉴定 RNA 分子中蛋白质结合位点的 iCLIP 等^[4-5]。

CLASH 是一种通过交联、连接和测序等步骤对 RNA 分子与 RNA 分子相互作用进行研究的实验系统, 主要被用于鉴定 miRNA 与其靶 mRNA 的相互作用。其实验流程为: 首先对细胞进行交联处理, 将 miRNA 及其靶 mRNA 与 RISC 复合物固定连接; 随后通过免疫共沉淀富集纯化 RISC 复合物, 并利用核酸酶 (RNaseA/T) 将靶 mRNA 裸露部分降解; 之后再通过分子内连接将 miRNA 的 3'-OH 末端与靶 mRNA 的 5'-PO₄ 末端连接, 形成一条长的单链 RNA; 最后, 在其两端添加接头并进行高通量测序分析^[6]。

ChIRP 是一项通过纯化 RNA 分子从而获得其结合的染色质片段 (包括 RNA 结合蛋白与基因组 DNA) 的实验方法, 主要用于鉴定长链非编码 RNA (lncRNA) 与染色质相互作用的状态和位点。其主要实验流程为: 首先对细胞进行交联处理, 使 lncRNA、染色质及与它们相互结合的蛋白被固定连接; 随后通过超声将染色质打断, 并使用带有生物素标记的 tiling oligos 与长链非编码 RNA 分子杂交; 之后利用亲和素磁珠纯化长链非编码 RNA 及其结合的染色质片段和蛋白因子; 最后, 经核酸酶消化获得与长链非编码 RNA 结合的染色质 DNA 片段及 RNA 结合蛋白, 可进行后续分析和测序^[7]。

2 功能性的测量

RNA-seq (RNA sequencing) 是一个后基因组时

代广泛应用的技术,也被称为转录组测序技术,即通过高通量测序技术分析包括 mRNA、小分子 RNA 及非编码 RNA 等 RNA 转录本种类,同时还可以从测序读数反映 RNA 的表达水平,可较为准确地分析组织细胞的基因表达谱。

RNA-seq 用于对 mRNA 测序,其主要实验流程为:首先提取样本总 RNA,用耦联磁珠的 Oligo (dT) 富集 mRNA 并同时去除 rRNA;随后将 mRNA 片断化,再以其为模板、使用随机引物依次合成 cDNA 链;最后,经过末端修复、添加测序接头以及 PCR 扩增等步骤,获得完整的测序文库,即可进行测序分析^[8]。RNA-seq 技术还可根据实验的具体需求,进行局部的调整 and 变化,衍生了许多相关的变种方法。其中一个主要的变种是只测 mRNA 的 3' 端^[9-10],该方法可大大降低成本,同时也得到 mRNA 末端的信息。

需要强调的是 mRNA 水平的变化并不等同于转录调控。GRO-seq (global nuclear run-on sequencing) 是一种专门测量新生 RNA 的方法,其原理是先提取细胞核,然后加入带标记的底物 (Br-UTP),让其整合到新生的 RNA 中。然后提取带标记的 RNA 用于高通量测序^[11]。新生 RNA 的变化直接反映了转录调控的多种机理。

3 RNA结合蛋白的系统性鉴定和研究策略

非编码 RNA 需通过 RNA 结合蛋白发挥功能,深入研究 RNA 结合蛋白对揭示非编码 RNA 功能与作用机制至关重要。现有多种方法对 RNA 结合蛋白进行系统鉴定和研究。

最近 Hentze 实验室报道了一种鉴定 mRNA 结合蛋白的方法。他们以 HeLa 细胞为材料:首先通过传统的紫外交联 (UV-CL) 或光激活核糖核苷酸增强交联 (PAR-CL),将 mRNA 与蛋白质的相互作用固定连接;随后使用耦联磁珠的 Oligo(dT) 富集纯化 mRNA 蛋白复合物,经过洗涤及降解 RNA 等步骤,最后将获得的蛋白质进行质谱分析^[12]。该研究显示细胞中约有上千种蛋白质直接结合 mRNA。

为了快速鉴定 RNA 结合蛋白特异性,Wickens 实验室最近报道了一种筛选 RNA 结合蛋白偏好结合序列的实验方法。他们首先在体外通过 T7 启动子表达一段含有 20 个核苷酸的随机 RNA 序列,随后将这一 RNA 文库与表达的已知蛋白质孵育,再利用结合蛋白将其纯化出来,最后经过 RT-PCR 扩增、两端加入测序接头等步骤后进行测序,对数据

进行统计分析后即可获得这一已知蛋白质的偏好结合序列^[13]。加拿大的 Hughes 实验室新近也报道了一种系统性鉴定 RNA 结合蛋白识别 RNA motifs 的分析方法。他们可以通过蛋白质的 RNA 结合结构域的序列,推测其识别 RNA 的偏好性,且体外实验结果与内源数据库较好地吻合。这项工作直观地研究了 RNA 结合蛋白与它们的 RNA 靶标,为真核生物转录后调控机制的研究提供了宝贵的资源^[14]。

4 研究调控机制的方法

对于 RNA 的功能及调控机制研究而言,RNA 的二级结构可能具有重要的指导意义。相比以往的方法,最近发展出一些高通量的方法来测定 RNA 的二级结构。

斯坦福的 Segal 和 Chang 实验室发展了 PARS (parallel analysis of RNA structure) 策略,它结合了高通量测序技术与结构特异性核酸酶消化方法,可在基因组水平以单核苷酸分辨率解析 RNA 的二级结构^[15]。首先,他们分别利用能够切割双链 RNA 的核酸酶 V1 以及识别切割单链 RNA 的 S1 处理酵母的总 mRNA,之后将其随机片段化并进行高通量测序,随后根据每一条片段序列在完整 mRNA 序列上的比对位置来判断此序列前一位碱基所处的结构状态,最后整合所有测序结果即可获得 RNA 分子中每一个碱基的结构信息(即处于单链或双链状态)^[16]。

Chang 和 Kools 实验室还设计了一种新的化学探针,可在活细胞中通过引物延伸分析,选择性地对 2' 羟基酰化作用,被称为 SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension),这种方法不仅可以直接在活细胞中探测 RNA 的结构,同时还能够在不同的细胞状态下评估 RNA 结构的动态变化^[17]。

Weissman 实验室发展的 Ribosome Profiling 应该是近年来发展建立的最重要的方法之一,该技术从全基因组水平展示了 mRNA 的翻译状态^[18]。他们首先通过核酸酶消化获得核糖体及其保护的 mRNA 片段,随后经过长度选择与加尾后,利用带有测序接头的引物进行反转录,再经过环化和再线性化,最后获得两端带有测序接头的 RNA 文库,即可进行高通量测序分析。利用此方法不仅可以鉴定正处于被翻译状态的蛋白序列,同时也能够筛查出决定蛋白质丰度及应对环境压力的翻译控制元件。

5 真核生物基因组编辑技术

最近在细菌中发现的 TALEs 编码系统和 CRISPR 免疫系统极大地促进了真核生物基因组编辑技术的发展, 在此基础上研发的 TALENs 技术和 CRISPR-Cas 技术使得对模式生物基因组 DNA 进行特定碱基的编辑修改变得方便快捷。

TALEN 靶向基因敲除技术利用 TAL 序列模块, 构建针对任意核苷酸靶序列的重组核酸酶, 可以在特定位点切开目标基因、敲除该基因^[18]。而 CRISPR RNA 是最近发现的原核生物中的调控 RNA, 用以抵御外来病毒和质粒入侵。在 II 型 CRISPR 系统中, CRISPR RNA (crRNA) 与转录激活 crRNA (Trans-activating crRNA) 退火形成的复合物, 能够特异识别基因组序列, 引导 Cas9 核酸内切酶在目的片段形成 DNA 双链断裂^[19]。最近的多项研究表明, 利用 CRISPR/Cas9 系统能够对小鼠和人类基因组特定基因位点进行精确编辑; 此外, 该系统还可以针对同一细胞中的多个位点实现多靶点同时酶切, 使得多个基因同时敲除或敲入成为可能^[20]。这些直接的基因组编辑技术可更有效地实现基因沉默而少有脱靶效应 (off-target effects), 因而将有非常广泛的应用, 包括研究非编码 RNA 的生物学功能。

[参 考 文 献]

- [1] Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein-DNA interactions and epigenomic states. *Nat Protoc*, 2013, 8(3): 539-54
- [2] Ule J, Jensen KB, Ruggiu M, et al. CLIP identified Nova-regulated RNA networks in the brain. *Science*, 2003, 302(5648): 1212-5
- [3] Yeo GW, Coufal NG, Liang TY, et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*, 2009, 16(2): 130-7
- [4] Hafner M, Burger LM, Hausser KM, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 2010, 141(1): 129-41
- [5] Konig J, Zarnack K, Rot G, et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*, 2010, 17(7): 909-15
- [6] Helwak A, Kudla G, Dudnakova T, et al. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 2013, 153(3): 654-65
- [7] Chu C, Qu K, Zhong FL, et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*, 2011, 44(4): 667-78
- [8] Katz Y, Wang ET, Airoidi EM, et al. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*, 2010, 7(12): 1009-15
- [9] Fox-Walsh K, Davis-Turak J, Zhou Y, et al. A multiplex RNA-seq strategy to profile poly(A+) RNA: Application to analysis of transcription response and 3' end formation. *Genomics*, 2011, 98(4): 266-71
- [10] Fu Y, Sun Y, Li Y, et al. Differential genome-wide profiling of tandem 3'UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res*, 2011, 21(5): 741-7
- [11] Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 2008, 322(5909): 1845-8
- [12] Castello A, Fischer B, Eichelbaum K, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 2012, 149(6): 1393-406
- [13] Campbell ZT, Bhimsaria D, Valley CT, et al. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep*, 2012, 1(5): 570-81
- [14] Ray D, Kazan H, Cook KB, et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, 2013, 499(7457): 172-7
- [15] Kertesz M, Wan Y, Mazor E, et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature*, 2010, 467(7311): 103-7
- [16] Wan Y, Qu K, Ouyang Z, et al. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat Protoc*, 2013, 8(5): 849-69
- [17] Spitale RC, Crisalli P, Flynn RA, et al. RNA SHAPE analysis in living cells. *Nat Chem Biol*, 2013, 9(1): 18-20
- [18] Ingolia NT, Ghaemmaghami S, Newman JR, et al. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, 2009, 324(5924): 218-23
- [19] Wei C, Liu J, Yu Z, et al. TALEN or Cas9 - rapid, efficient and specific choices for genome modifications. *J Genet Genomics*, 2013, 40(6): 281-9
- [20] Wang H, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 2013, 153(4): 910-8

付向东报告讨论

徐安龙 (中山大学生命科学学院)

Q: 功能是生物学研究的重点, 用什么 RNA 研究的新方法能够了解其功能?

A: 具体问题具体分析。举个例子, 我们做剪切 20 年, 最近发现所有的因子都和白血病有关。但是这些因子哪些是白血病产生的原因, 哪些是结果, 哪些是特定细胞系中特有的? 需要具体问题具体分析。

屈良鹄 (中山大学生命科学学院)

Q: 问一个概念问题。在报告中提到, mRNA 广义上其实也是一种非编码 RNA。您怎么理解这个问题, 或者说这个概念?

A: mRNA 90% 是非编码区, 因此说它是非编码 RNA 也不为过。非编码 RNA 主要起调控基因表达的作用。mRNA 除了翻译蛋白, 对基因的表达调控非常重要。传统观念认为, 基因的 promoter 和 enhancer 对基因表达调控至关重要。然而, 我们发现, 相同的 promoter, 后面接的基因不一样, 表达量高低就不一样。RNA 结合蛋白, 可以对 RNA 本身进行反馈调控, 即由这个基因自己转录出来的 RNA 又回头调节自己的基因。RNA 的表观遗传学, 现在也是热点。

张旭 (中国科学院上海生命科学研究院神经科学研究所)

Q: 你在报告中多次提到了 RNA 的 3'UTR 的功能, 很少提到 5'UTR 的功能?

A: 5' 端 UTR 现在也有人做, 研究基因的起始问题, 主要就是研究 5'UTR。例如, 有一种 Enhance RNA。有头没尾。即有 cap, 没有 poly A, 要研究它的功能, 就只能研究其 5'UTR。

施蕴渝 (中国科学技术大学生命科学学院)

Q: CRISPR 技术不光可以做 knockout, 还可以做 knockin, 它的原理是什么? 在小鼠中, 做了 knockout 以后, 有些时候看不到表型, 这是为什么呢?

A: CRISPR 技术会在基因组中产生切口。对于基因组中的切口, 细胞中有修复机制对其进行修复。DNA 酶在修复的过程中, 如果有外源的同源序列, 就会被修复酶整合到切口中去。因此, 只要提供一个外源的双链 DNA, 就可以进行 knockin 操作。

表型是生物学的基本问题。不管是动物还是植物, 都需要发展进化。如果一个基因非常重要, 一碰就死的话, 那对生物的延续是不利的。因此在进化过程中, 生物体往往通过基因倍增, 对重要基因复制出多份拷贝, 保证其能够正常生长。植物中的基因比动物中多一倍, 因为植物不能动, 对环境的依赖更大。大部分在动物中的必需基因, 在植物里都不是必需的。植物中某一个基因往往只管一个器官, 很少出现一个基因对整个植株都重要的情况。

因此, 对待功能研究, 我们一方面要做 knockout, 研究它独一无二的功能; 另一方面, knockout 还要跟过表达方法相辅相成。Knockout 方法和过表达方法研究出来的基因功能不一定会重合。

张旭 (中国科学院上海生命科学研究院神经科学研究所)

Q: 经常遇到这样的情况, 某个基因, 在人中有 isoform, 这些 isoform 很多情况下跟老鼠中的 RNA 不太一样, 不保守, 我们用老鼠来做体内实验的时候, 不知道该怎么办。

A: 这是一个生物学难题。做基因表达, 功能研究, 一般看保守区。但是 isoform 很多是不保守的。因此, 现在普遍有一个观念, 正是这些 isoform 之间不保守的差异形成了物种的特异性。isoform 是物种特有的。比如在果蝇中, 有些果蝇的性别决定就是靠 RNA 剪切完成, 有些就不是。这种性别决定上的差异就跟 isoform 有关。因此建议大家今后更多关注保守的 mRNA, 少关注差异大的 isoform。

曹晓风 (中国科学院遗传与发育生物学研究所)

Q: 在您的报告中提到, 你们实验室做的 miRACE, target active 互补配对特别好。在植物中是什么情况呢?

A: 植物中的 siRNA, 在动物中很少。siRNA 到了 11、12、13 核苷酸, Dicer 切的地方, 就不形成配对了。