

DOI: 10.13376/j.cblls/2016081

文章编号: 1004-0374(2016)06-0630-10



朱丹萌课题组主要从事中等长度及长链非编码 RNA 在高等植物生长发育及环境信号响应中的分子鉴定与作用机制研究。在植物研究体系中改进与完善 RNA 的特异性检测、亚细胞定位及生物学功能研究方法。以模式植物拟南芥长链非编码 RNA 在光控植物生长发育中的作用机理为研究重点,综合运用分子遗传学、生物化学、细胞生物学与生物信息学方法,阐明功能型新非编码 RNA 在转录水平和转录后水平发挥作用的分子机制。

高等植物中的较长非编码RNA: 从序列、功能到分子机理

王玉秋, 樊 德, 邓兴旺, 朱丹萌*

(北京大学生命科学学院与现代农学院, 北京大学-清华大学生命联合中心,
蛋白质与植物基因研究国家重点实验室, 北京 100871)

摘 要: 随着非编码 RNA 预测方法的推陈出新及高通量测序技术的发展, 先前由于研究技术所限未能被鉴定注释的较长非编码 RNA 逐渐成为研究热点。目前人们在拟南芥、水稻、玉米、苜蓿等高等模式植物中发现大量中等长度及长链非编码 RNA, 其中一些非编码 RNA 在植物生长发育及环境信号响应等关键生物学过程中发挥着重要的调控功能。将着重介绍长度大于 50 个核苷酸的非编码 RNA 在多种高等植物中的序列、分子特征与功能鉴定的方法与最新进展。同时, 对长链非编码 RNA 在高等植物中最新报道的作用模式与分子机理进行归纳总结, 并展开讨论。

关键词: 中等长度非编码 RNA; 长链非编码 RNA; 高等植物

中图分类号: Q522; Q943 **文献标志码:** A

Longer noncoding RNAs in higher plants: from the sequence, function to the molecular mechanism

WANG Yu-Qiu, FAN De, DENG Xing-Wang, ZHU Dan-Meng*

(State Key Laboratory of Protein and Plant Gene Research, Peking-Tsinghua Center for Life Sciences,
School of Life Sciences and School of Advanced Agriculture Sciences, Peking University, Beijing 100871, China)

Abstract: With the development and innovation of noncoding RNA prediction algorithms and high-throughput sequencing technologies, the research on previously uncharacterized longer noncoding RNAs gradually becomes a frontier in the field. At present, large amounts of intermediate-sized and long noncoding RNAs have been identified

收稿日期: 2016-04-12

基金项目: 中国博士后科学基金面上资助项目(2015M580013); 国家自然科学基金青年科学基金项目(31500974); 国家自然科学基金重大研究计划培育项目(9154010)

*通信作者: E-mail: zhudanmeng@pku.edu.cn; Tel: 010-62761093

in model plants including *Arabidopsis*, rice, maize, and *Medicago* etc. And those emerging genes accomplish variety of biological function during development and in response to environmental stimuli. Here, we reviewed the current and fast growing knowledge on the systematic identification and annotation of noncoding RNAs with the size of longer than 50 nt in higher plants. In addition, the recent advances on the regulatory mechanisms of plant long noncoding RNAs will be summarized and discussed.

Key words: im-ncRNA; long ncRNA; higher plants

1 概述

非编码 RNA (non-coding RNA, ncRNA) 是一类由基因组产生的不同于信使 RNA (messenger RNA, mRNA) 的核糖核酸分子。它们可不被进一步翻译为多肽或蛋白质, 直接以 RNA 的形式发挥功能。1953 年, 沃森和克里克建立的中心法则是人们发现 ncRNA 的第一个里程碑事件。编码基因终产物蛋白质的生成, 需要两类高丰度核糖核酸分子核糖体 RNA (ribosomal RNA, rRNA) 和转运 RNA (transfer RNA, tRNA) 的直接参与。它们成为最早为人们所发现的 ncRNA 的代表。在之后的近半个世纪, 人们主要通过生物化学纯化的方法, 零星地发现一些以核仁小分子 RNA (small nucleolus RNA, snoRNA) 和小核 RNA (small nuclear RNA, snRNA) 为代表的丰度较高的 ncRNA 类别。人类基因组计划的实施则是人们发现 ncRNA 的第二个里程碑事件。随着各个模式生物 (包括拟南芥) 基因组的破译, 人们逐渐开始关注占据着基因组绝大部分比例的非编码区域的作用与功能。此后, 第二代测序技术的建立与发展可谓是人们发现与鉴定 ncRNA 的第三个里程碑事件。与此同时, 生物信息分析与计算机预测能力的逐步提高进一步催化了各个物种 ncRNA 的系统发现与鉴定。

根据 ncRNA 的长度, 人们将其大致划分为三种类别: (1) 长度小于 50 nt 的小 RNA (small RNA), 研究得相对清楚的主要有 microRNA (miRNA), small interfering RNA (siRNA) 等亚类; (2) 长度大致在 50~300 nt 之间的中等长度非编码 RNA (intermediate-sized noncoding RNA, im-ncRNA), 包括以 snoRNA、snRNA 等为代表的已知类别的分子以及大量未知功能的 ncRNA; (3) 长链非编码 RNA (long noncoding RNA, lncRNA), 一般指长度大于 200 nt 的 ncRNA。在有关 small RNA 产生、加工与作用机制日渐清晰的同时, 较长非编码 RNA 逐渐成为新的研究热点。高通量测序技术的飞速发展以及针对 ncRNA 的分子生物学研究技术与方法的开发, 使得很多表达具

组织器官特异性、丰度低或二级结构更复杂的较长非编码 RNA 得以被发现和深入研究。本文将着重介绍高等植物研究领域大于 50 nt 的 ncRNA 的系统鉴定与功能解析, 并在文中将这些 ncRNA 简称为较长非编码 RNA (longer noncoding RNA, lrncRNA)。

2 高等植物lrncRNA的发现与系统鉴定

2.1 以snoRNA为im-ncRNA研究代表, 从个例发掘到系统鉴定的历程

植物中发现最早的 im-ncRNA 是参与 RNA 加工与修饰的 snRNA 和 snoRNA。1959 年, Stern 等^[1]在有关豌豆的研究中发现核仁区 RNA 呈现高度富集现象。随后于 1985 年, Kiss 等^[2]通过亚细胞分离、聚丙烯酰胺凝胶电泳及化学法测序, 首次确定植物多个物种中均存在的, 定位于核质的 snRNA U1、U2、U5、U6 以及定位于核仁内的 U3 (最初归为 snRNA, 现在划分为 snoRNA)。1994 年, Leader 等^[3]利用大鼠和酵母中已获得的序列, 通过同源序列扩增的方法获得了多种植物中 snoRNA U14 的序列。1997 年, Leader 等^[4]以玉米为主要材料, 进一步发现植物中以 U14 为代表的 snoRNA 以基因簇的形式存在, 并获得了多个 snoRNA 基因簇。2001 年, Barneche 等^[5]通过计算机预测与筛选, 发掘出拟南芥中 66 个 C/D Box 类型的 snoRNA, 实现了对特定类别 ncRNA 的系统鉴定。同年, Qu 等^[6]利用引物延伸 (primer extension) 等实验方法发现了 10 个可能具有不同于 rRNA 修饰功能的 snoRNA 基因簇。2010 年, Kim 等^[7]利用 RNA 组学的方法, 鉴定获得了以 U13 为代表的 31 个未知 snoRNA 以及 15 个已知 snoRNA 的新变体, 将拟南芥中的 snoRNA 数目增至 188 个。2013 年, Liu 等^[8]利用高通量测序法在单子叶模式植物栽培水稻中新发现 snoRNA 125 个。在水稻中, 90% 的 snoRNA 可作为前体产生 20~30 nt 的小 RNA, 并且这些小 RNA 特别富集于水稻 ARGONAUTE 1b 为核心的效应蛋白复合体中。2014 年, Wang 等^[9]进一步在双子叶模式植物拟南芥中鉴定获得 28 个新的 snoRNA, 其

中有 22 个是利用序列互补原则预测不到底物的新成员 (orphan snoRNA)。

与动物不同, 植物中大多数 snoRNA 以基因簇的形式存在于基因组中。有趣的是, 高等植物水稻中的 snoRNA 的种类与数量非常丰富, 串联重复 (tandem duplication) 是半数以上 snoRNA 家族产生的动力来源。在高等动物中, snoRNA 除以 RNA-蛋白质复合体的形式参与核酸加工调控与介导核酸修饰以外, 已有研究结果将它们的功能拓展到生长发育、代谢与重大疾病中的多个调控层面。其中包括, 作为小 RNA 的前体加工产生具有类似 miRNA 功能的 RNA、位于 lncRNA 的两端并使其转录产物稳定、参与 mRNA 的可变剪接调控、染色质结构调控等等^[10-11]。上述结果提示, 研究进化保守且种类丰富的 snoRNA 具有重要的生物学意义。关于单个 snoRNA 在植物领域中的功能机制方面目前报道较少, 相信不断进步完善的 snoRNA 系统鉴定将为在植物中研究 snoRNA 的生物学功能与基因组演化分析提供丰富的素材与资源。

2.2 依托于技术发展趋势的未知类型 lncRNA 系统鉴定

2001 年, MacIntosh 等^[12]通过对拟南芥 20 000 个表达序列标签克隆进行计算机过滤分析, 共鉴定到 19 个可能的未知类型 ncRNA 克隆。这个数目超过之前植物研究领域所有发现的新型 ncRNA 的总数, 并由此揭开了计算与实验组学方法系统分析植物 lncRNA 的热潮。2002 年, Alexander Hüttenhofer 及其同事建立了分离鉴定 50~500 nt 大小的 snmRNA (small non-messenger RNA) 的方法, 分析了多个模式生物的非编码序列, 其中共鉴定到 140 个拟南芥 snmRNA^[13-14]。该方法能够较为全面地获取一定大小范围内的 RNA 序列信息, 但测序方法主要采用 Sanger 测序法, 通量小且花费大。在此特定片段大小分离技术基础上, 人们进一步改进了建库策略, 即针对 RNA 5' 端结构进行链特异性建库, 并结合新一代高通量测序技术在全基因组范围鉴定 lncRNA, 进一步拓展了针对特定长度范围 ncRNA 的系统鉴定能力。利用上述方法, Liu 等^[8-9]和 Wang 等^[14]分别在水稻和拟南芥中发现了 754 个和 521 个新型 im-ncRNA, 这些基因中的绝大多数均为各自物种特有的序列。其中一些拟南芥编码基因 5'UTR 上 im-ncRNA 的产生与宿主基因高表达呈现正相关性, 而基因间区产生的 im-ncRNA 与编码基因表达受到相似的遗传与表观遗传调控。基于第二

代测序技术的高通量转录组测序分析已成为被最广泛应用于 lncRNA 系统鉴定与分析的手段之一^[15]。近几年, Pacific BioSciences 公司开发了 PacBio 单分子测序技术, 能够实现比第二代测序技术更长的测序读长, 提高对于一些高重复序列测定的准确性及进一步减少序列拼接引入的不确定性等弊病^[16]。更多新测序技术的出现与应用必将进一步促进植物 lncRNA 的系统鉴定与注释。

由微阵列芯片 (Microarray) 技术拓展而来的嵌合芯片技术 (Tilling array) 是与高通量测序技术并行发展, 并被广泛应用于 lncRNA 系统鉴定的另一项关键技术。2005 年, 人们首次采用嵌合芯片技术在全基因组水平分析了拟南芥的转录本^[17]。Li 等^[18]基于水稻的嵌合芯片转录谱数据, 系统分析了栽培水稻亚种粳稻与籼稻中的非外显子转录活性区域 (transcriptionally active region, TAR), 各获得 25 352 个和 27 744 个 TARs, 这些区域很可能会产生很多的 lncRNA。Liu 等^[19]通过分析已有的 200 个拟南芥转录组数据, 初步获得 6 480 个分布于基因间区的 lncRNA 候选基因, 并结合嵌合芯片技术系统进一步分析了它们的表达模式等分子特征。然而, 嵌合芯片技术有时因其探针杂交信号存在一定的假阳性率, 且无法确定转录本的 5' 端和 3' 端, 往往需要结合转录组测序或 RT-PCR 实验进一步确认其可靠性。

计算机预测与生物信息学分析在植物 lncRNA 的发掘与系统鉴定的历程中也发挥着不可或缺的作用。例如, 人们基于拟南芥的 cDNA 文库、微阵列芯片等基因表达数据, 开发了计算机算法, 在全基因组水平预测并分析了可能的以顺式 (*in cis*) 与反式 (*in trans*) 作用方式发挥功能的天然反义 RNA (natural antisense transcript, NAT)^[20-21]。至今, 人们已开发了大量的计算机预测与分析方法, 基于对已有表达数据的整合, 对多个植物物种的 lncRNA 进行了预测与分析^[22-28]。计算机预测与分析成本较低, 但是最终 ncRNA 的真实性鉴定需要进一步的实验验证与确认。

2.3 结合植物研究的重要问题多层次发掘 lncRNA

在以模式动物为研究对象的工作中, 人们逐渐认识到长链非编码 RNA 具有明显的组织特异性表达, 并参与重要的生长发育、代谢及重大疾病调控中。lncRNA 在植物中的研究也逐渐拓展至植物特有的生长发育及环境响应过程。

不同组织器官、细胞类型乃至细胞器中

lncRNA 的系统鉴定与分析逐渐为人们所关注。对于较为明确的组织器官, 如植物的根、茎、叶、花、果实、种子等, 人们可通过机械分离法直接获取特定组织器官材料, 并对其中的 lncRNA 进行鉴定、比较和功能初探。例如, Zhang 等^[29]采用链特异性 RNA 测序的方法, 系统分析获得了大量水稻有性生殖相关的 lncRNA, 并且通过实验证实了一些 lncRNA 的表达和功能与水稻生殖过程有关联。除此之外, 人们还建立了更加精细的分离方法, 分析特定组织器官中的转录组。例如, Jiao 等^[30]巧妙地通过具有组织特异性表达的启动子驱动标签化的核糖体大亚基蛋白基因, 成功捕获了特定组织中核糖体结合的 RNA, 并且意外地发现了核糖体结合 RNA 中亦存在可能的 ncRNA。在亚细胞层面, Zhelyazkova 等^[31]通过对大麦叶绿体初级转录本的系统分析, 初步确定了叶绿体可能产生 ncRNA 的候选基因。这些更为细致的研究进一步丰富了人们对于植物体内 lncRNA 的认识。

植物不能行走, 因此及时响应环境信号以适应环境变化对于其顺利完成生活史至关重要。来自不同实验室的科研人员报道了不同环境信号和胁迫信号条件下植物 lncRNA 的鉴定工作。例如, Wang 等^[32]分析获得了拟南芥中大量能够响应光信号变化的 NAT pair, 并且发现它们的表达变化大多与组蛋白乙酰化修饰相关联。又如, Di 等^[33]通过整合表达、表观遗传学修饰和结构特征分析, 分析了拟南芥响应多种非生物胁迫信号的 lncRNA, 并且发现它们自身含有的序列或结构元件可能与其响应信号能力相关联。此类工作的开展表明, lncRNA 可能作为响应因子参与了不同信号的传递, 而具体哪些因子直接参与了它们的表达调控以及它们的表达响应后如何执行下游功能尚待深入探究。

此外, 人们还将 lncRNA 的鉴定工作进一步拓展到多种具有重要农艺性状和经济价值的植物物种中^[34]。例如, Kang 和 Liu^[35]基于二倍体草莓花和果实的基因表达数据, 系统分析了可能产生的 lncRNA。通过序列比对, 初步推测它们大多进化上不具保守性。Zhu 等^[36]通过链特异性转录组数据分析和比较了野生型番茄和果实成熟突变型番茄中的 lncRNA, 并且发现其中一些 lncRNA 的表达与番茄果实成熟度相关联。随着 lncRNA 研究的延续, 更多植物物种中特异演化的 lncRNA 将被发现, 而且它们的表达与功能的研究会进一步深化人们对植物经济农艺性状的认识。

2.4 数据整合, lncRNA系统鉴定与分析的公共平台

在获得了大量有关 lncRNA 数据的同时, 人们也在不断对这些日益丰富的生物信息进行着归纳与整合。2011年, Mattick 及其同事建立的综合数据库 lncRNADB (lncRNAs Database) 囊括了真核生物中的 lncRNA 的表达情况与功能注释, 但其中仅包含 8 种植物中功能较为明确的 lncRNA 的注释^[37]。Zhao 等^[38]更新升级的 NONCODE 2016 数据库, 整合了 16 个模式生物的 ncRNA 的信息, 其中包含拟南芥中 3 853 条 lncRNA 的注释。植物研究的专业数据库中, TAIR (The *Arabidopsis* Information Resource) 数据库是拟南芥研究最为重要的基因信息来源之一, 目前包含了大约 478 个 lncRNA 的注释, 但是信息较为简略^[39]。PNRD (Plant Non-coding RNA Database) 数据库由专门研究 miRNA 的 PMRD (Plant microRNA Database) 数据库升级而来, 收集了包括 lncRNAs、tRNA、rRNA、snRNA 和 snoRNA 等共计 25 739 个 lncRNA 的信息, 是目前较为全面的植物 ncRNA 数据库^[40]。PLncDB 数据库是针对模式植物拟南芥已有 lncRNA 查询的专用数据库, 涵盖了拟南芥中通过嵌合芯片技术与转录组测序技术已发现的 16 227 个 lncRNA。该数据库较为详细地注释了 lncRNA 在拟南芥不同生长条件、不同发育时程和不同突变体中的表达特征与表观遗传修饰情况^[41]。PLNlncRbase 数据库通过手动搜集的方法整合了近 200 篇文献报道中植物 lncRNA 的信息, 记录了 43 种植物中 1 187 个 lncRNA 的信息^[42]。针对 NATs, Chen 等^[43]建立了专门用于查询植物中 NATs 的数据库 PlantNATsDB (Plant Natural Antisense Transcripts DataBase), 其中记录了 69 种植物中大约 200 万对 NATs。多种数据库的建立为人们从多层面、多角度研究植物 lncRNA 提供了必要的平台。

2.5 lncRNA真实性的界定方法

已被鉴定注释的 lncRNA 是否真的直接以 RNA 的形式发挥功能呢? lncRNA 真实性界定的核心问题是这些分子是否具有短肽或者蛋白质编码潜能。目前, 推断 lncRNA 的蛋白质编码潜能的方案大体可分为生物信息学分析和实验验证两种。

基于生物信息学分析, 人们开发了多种分析方法推断 lncRNA 是否具有蛋白质编码潜能^[44-49]。值得关注的是, 监督式机器学习的应用对于推断 lncRNA 的蛋白质编码潜能发挥着重要作用。例如, Liu 等^[50]开发的 CONC (coding or non-coding) 基于可能肽段长度、氨基酸组成、二级结构预测、组合熵、

同源序列比对等特征的机器学习程序,可综合推断 lncRNA 序列的蛋白质编码潜能。在此基础上, Kong 等^[51]进一步改善了 CONC 处理较大数据量等不足,开发了基于序列特征分析的转录本蛋白质编码能力的评估算法 CPC (Coding Potential Calculator),并且提供了易于使用的网页界面。Lu 等^[52]通过整合高通量测序数据、序列保守性分析以及二级结构稳定性等信息,建立了预测 lncRNA 的整合机器学习方法。

生物信息学分析对于 lncRNA 的编码潜能仅供参考,尚需要进一步的实验验证。归纳起来,可主要通过以下 4 种方法进一步验证候选 ncRNA 的蛋白质编码能力^[53]: (1) 通过体外翻译系统检测候选 ncRNA 基因能否产生与其含有的 ORF 相对应的肽段或蛋白质; (2) 检测候选 ncRNA 是否与核糖体存在相互作用,如果不存在互作,则可推测其可能不具备蛋白质编码能力; (3) 构建候选 ncRNA 基因与蛋白质标签基因的融合序列,转入生物体内表达,通过检测是否有融合蛋白的产生来判断其编码能力; (4) 通过破坏或干扰 ncRNA 编码序列中可能的 ORF,结合转基因互补实验判断其是否具有多肽或蛋白质编码能力。目前第 4 种方法已在拟南芥中初步验证了其有效性^[54]。

2.6 lncRNA 分子特征的多层次鉴定

基于 RNA 分子特性,人们开发了一系列的技术方法对植物中的 lncRNA 分子进行多层次的特征分析。从 lncRNA 基因序列与产生位置出发,基于注释位置可通过搜寻合适的 T-DNA 插入突变体获取特定 lncRNA 表达变化的株系;对于无合适 T-DNA 插入突变体的情况,除转基因过表达方法外,人们正在尝试用 artificial miRNAs (amiRNAs) 以及最新的 CRISPR/Cas9 系统影响特定 lncRNA 的表达^[55]。围绕 lncRNA 的分子表达特征,除了在组织器官层面结合转录组或者芯片数据分析外,人们可通过 RNA-FISH 和 RNA 成像等方法进一步探究 lncRNA 在细胞,甚至亚细胞水平的分布特征^[7,56];对于具有染色质结合能力的 lncRNA,还可采用 ChIRP-seq (chromatin isolated RNA purification-sequencing) 和 3C (chromosome conformation capture) 等技术开展特定 lncRNA 在基因组上的结合位点分析^[57]。围绕 RNA 的序列和结构特征,除了采用 mfold 等软件进行二级结构预测外,人们还研发了系统测定 RNA 体内二级结构的实验方法,这势必促进建立结构与功能之间的联系^[58]。围绕 lncRNA

可能互作效应因子的获得,在植物已有的研究中,人们主要通过结合特定生物学调控过程,推测特定 lncRNA 可能的互作蛋白质因子,进而通过 RIP (RNA immunoprecipitation) 等实验验证其可靠性^[59]。而动物体系应用广泛的直接从特定 lncRNA 序列出发,通过 RNA pull-down 等分析其互作蛋白质因子的方法,目前在植物个例研究中尚鲜有报道。该方法在植物体系中的应用对于深入解析 lncRNA 的功能机制将发挥重要的作用。

3 lncRNA 在植物体内的作用模式与分子机制

关于 lncRNA 在多种模式植物中的功能研究,近两年已有很好的综述进行了详细总结^[34,60-63]。人们发现 lncRNA 可与 DNA、RNA 或者蛋白质相结合,在转录、转录后等多个水平参与基因表达的调控。以下将根据 lncRNA 转录产生后是否作用于旁侧区域的基因表达调控,可将其在植物体内发挥作用的模式分为顺式和反式两类进行阐述。以下将选取典型 lncRNA 案例阐释以这两种不同作用模式参与植物重要生物学过程调控的研究进展。

3.1 顺式作用的 lncRNA 示例

3.1.1 lncRNAs 调控植物开花的分子机制

高等植物从营养期到生殖期(开花)的转变是非常重要的过程,其中参与这个过程调控的抑制基因 *FLC* (*FLOWERING LOCUS C*) 的表达调控对于开花时间的控制非常关键。目前,人们至少发现三类顺式作用的 lncRNA 参与了 *FLC* 基因的转录调控。2009 年和 2011 年,人们在 *FLC* 的正义链和反义链上发现两类 lncRNA 的产生,分别命名为 *COOLAIR* (*COLD INDUCED LONG ANTISENSE INTRAGENIC RNA*) 和 *COLDAIR* (*COLD ASSISTED INTRONIC NONCODING RNA*)^[59,64]。其中, *COOLAIR* 可间接通过抑制转录激活修饰 H3K36me3 的活性,在春化和 *FRIGIDA* 两个途径来抑制 *FLC* 基因的表达。*COLDAIR* 产生于 *FLC* 基因的第二个内含子区,在春化途径中可直接作为骨架分子招募基因沉默复合体 PRC2 (polycomb repressive complex 2),通过增加 H3K27me3 累积来维持 *FLC* 基因处于低表达水平。2014 年,人们发现了第三类参与 *FLC* 基因转录调控的 lncRNA *ASL* (*ANTISENSE LONG*),它们可能起到维持 *FLC* 基因区 H3K27me3 修饰状态的作用^[65]。三类 lncRNA 的共性为直接调控其产生位置的靶因子。对于 *COLDAIR*,其如何识别 PRC2 蛋白质复合体的核心组分是比较有意思的问题。而解析 *COOLAIR*

和 *ASL* 是通过一级序列还是二级结构, 如何调控 H3K36me3 和 H3K27me3 在 *FLC* 上的覆盖位置与水平, 将为深入阐明其作用机制提供新视点。

3.1.2 *APOLO* 参与生长素信号诱导的染色质结构改变

2014年, Ariel 等^[57]发现 lncRNA *APOLO* (*AUXIN REGULATED PROMOTER LOOP RNA*) 的表达能够影响其下游参与生长素极性运输的重要因子 *PINOID* (*PID*) 基因的转录。在响应生长素信号后, 起初转录 *APOLO* 的基因区发生 DNA 去甲基化现象, 染色质环逐渐打开且激活型组蛋白修饰 H3K9ac 上升, 抑制型修饰 H3K27me3 水平下降。*APOLO* 和 *PID* 的转录本会在 RNA 聚合酶 II (Pol II) 的介导下得以积累。随着 *APOLO* 转录的进行, 该分子可通过直接与抑制基因表达的基因沉默复合体 PRC1 组分 LHP1 (LIKE HETEROCHROMATIN PROTEIN 1) 相互作用, 逐渐介导 *PID* 启动子区域的染色质环形成; 同时, RNA 聚合酶 V (Pol V) 转录产生的 *APOLO* 会招募 ARGONAUTE 4 (AGO4) 复合体通过 RNA 指导的 DNA 甲基化 (RdDM) 途径来实现该区域 DNA 的甲基化重新积累, 而基因沉默复合体 PRC2 也会重新在 *PID* 与 *APOLO* 位点累积抑制型组蛋白修饰, 进而促使染色质环的再形成来下调 *PID* 的表达。这是植物中 lncRNA 通过动态影响染色质结构改变从而调控旁侧基因表达的重要案例, 深入研究它们与基因沉默复合体之间的互作方式与特征将为进一步阐明分子机理提供更有意思的线索。

3.2 反式作用的 lncRNA 示例

3.2.1 *HIDI* 在转录水平参与红光介导的光形态建成调控

光形态建成是植物响应环境的信号转导通路中研究得相对成熟且清楚的过程。近二十年来, 人们通过遗传学与生物化学手段筛选获得多个关键的负作用调控因子, 形成了较为完善的蛋白质调控构架。Wang 等^[54]在拟南芥中发现了一个植物特有的 lncRNA, 将其命名为 *HIDI* (*HIDDEN TREASURE 1*)。 *HIDI* 可参与多个光控植物发育过程, 其中在红光介导的光形态建成中 *HIDI* 能够直接参与抑制光形态建成关键抑制因子 *PIF3* (*PHYTOCHROME-INTERACTING FACTOR 3*) 基因的转录, 从而介导拟南芥红光信号中的下胚轴伸长抑制。该研究揭示了 lncRNA 参与植物光形态建成的精确调控过程。研究同时发现了两个对于 *HIDI* 发挥功能起重要作用的茎环结构, 且它们在陆地植物中的序列与二级预

测结构均是保守的。 *HIDI* 可装配进入 RNA- 蛋白质复合体, 深入研究复合体组分及其功能的关系将对于植物 lncRNA 如何以反式作用方式寻找靶因子并发挥调控作用具有重要指导意义。

3.2.2 lncRNA 参与水稻雄性不育的产生

植物生殖发育相关的研究中, 人们已发现许多 lncRNA 可能参与其中, 其中以水稻条件依赖的雄性不育中 lncRNA 参与调控的案例研究得较为深入。2012年, 两个研究组各自独立发现了同一个 ncRNA 基因的自发点突变是造成粳稻农垦 58S 光敏雄性不育系和籼稻培矮 64S 温敏不育系产生的遗传学基础。Ding 等^[66]的研究发现, 该点突变所在区域能够产生一个 1 236 nt 的 lncRNA, 并命名为 *LDMAR* (*LONG-DAY-SPECIFIC MALE-FERTILITY-ASSOCIATED RNA*)。农垦 58S 株系中 *LDMAR* 区域上述自发点突变类型, 可能影响了其二级结构的变化, 并且导致该 lncRNA 基因启动子区 DNA 甲基化水平显著上升, 进而通过某种未知途径造成 *LDMAR* 表达水平在长日照条件下明显下降。而 Zhou 等^[67]的研究则通过 RACE (rapid-amplification of cDNA ends) 的方法确定了一个 136 nt 的 ncRNA 中间产物, 并且发现该中间产物可进一步加工产生含有上述关键点突变位点的 21 nt 的小 RNA。无论是 *LDMAR* 自身还是可能由其加工而来的小 RNA, 在分子层面揭示它们如何作为反式作用因子执行下游功能, 有可能在分子层面进一步揭示水稻条件依赖的雄性不育产生的机制。

3.2.3 *IPSI* 在转录后水平参与植物体内磷酸盐的稳态调节

植物通过吸收土壤中的无机磷酸盐来实现自身体内的稳态。在这一过程中, *PHOSPHATE 2* (*PHO2*) 抑制植物根中磷的富集和活化过程, 且该基因的 mRNA 水平可受到 miRNA399 (miR399) 的切割抑制。2007年, 拟南芥的研究中报道了 *IPSI* (lncRNA *INDUCED BY PHOSPHATE STARVATION 1*) 能够作为 miR399 的诱饵 (decoy) 参与植物响应磷缺乏的调节过程^[50]。在磷缺乏胁迫条件下, *IPSI* 本身的表达可受到诱导。利用其与 miR399 匹配的特殊序列元件 (在预测的 miR399 效应复合体切割位置错配), *IPSI* 既能够与 miR399 结合又不被其所在的 RNA 沉默复合体切割, 从而实现对于 miR399 的靶基因 mRNA 的保护作用。 *IPSI* 是植物特有的一类非编码 RNA。它的作用机制被发现以后, 通过序列互补来负调控 miRNA 功能的 lncRNA 的例子在高等

植物和动物的研究中另有不少发现,它们可作为 miRNA 功能的一类重要调节因子^[27]。

3.2.4 ENOD40及其同源序列直接影响互作蛋白质功能

ENOD40 (*EARLY NODULING 40*) 是一个最早有文献记载的高等植物新型 lncRNA。1994 年, Crespi 等^[68] 在有关栽培苜蓿的研究中,通过比较分化中根瘤和完全分化根瘤中差异表达的 cDNA 发现了苜蓿中的 *ENOD40* 基因。2001 年, Sousa 等^[69] 进一步的实验发现,苜蓿 *ENOD40* 的 5' 端和 3' 端的两个短 ORF 对于其功能是必要的,同时两者间的结构性 RNA 序列对于其功能也不可或缺。2002 年, Rohrig 等^[70] 的研究发现,大豆中 *ENOD40* 的 5' 端区域可能会产生两个存在重叠的短肽,且与蔗糖合成酶的一个亚基存在直接相互作用,提示 *ENOD40* 可通过短肽行使生物学功能。2004 年, Campalans 等^[56] 通过酵母三杂交系统,筛选到了与 *ENOD40* RNA 直接互作的 RNA 结合蛋白 MtRBP1 (*Medicago truncatula* RNA binding protein 1), 并且实验结果表明在根瘤发生的过程中, *ENOD40* 能够通过互作将 MtRBP1 由核内聚集状态转变为细胞质弥散分布。最近, Crespi 研究组的实验结果显示,在侧根形成的过程中, *ENOD40* 在拟南芥中的同源基因编码的 lncRNA *ASCO-lncRNA* (*AS COMPETITOR LONG NONCODING RNA*) 能够作为竞争性靶标,阻抑其互作蛋白可变剪接调节因子 NSRs (nuclear speckle RNA-binding protein) 对于目的 mRNA 的结合,进而影响这些基因的可变剪接模式与表达^[71]。由此可见, lncRNA 既可直接以 RNA 形式,还可编码短肽在生物体内执行多个功能^[72-73]。这既反映了

核酸分子功能形式的多样性,也体现了生物系统的复杂性。

4 结语与展望

近二十年来,在真核生物中大量发现的功能型 lncRNA 已逐渐使人们意识到除中心法则外,遗传信息的传递还可以通过 RNA 形式的调控分子以从序列到结构到生物学功能的方式进行(图 1)。后基因组时代的组学研究方法快速更新以及基于 RNA 生物化学特性的新技术新方法的不断产生与改进,促使非编码 RNA 在动植物的生长发育、代谢等重大生命过程中研究飞速发展。虽然在植物生长发育的数个重要节点均已发现 lncRNA 的作用踪迹(图 1),但是,相对于蛋白质编码基因功能的了解来说,人们对于植物 lncRNA 的认识仍在初级阶段。结合模式植物特有的生命过程及其与环境的互作的调控过程(如光信号、生物及非生物胁迫等),如何有效地发现不同细胞类型中有功能的 lncRNA,如何快速准确地分析它们发挥功能的关键序列或结构,利用不断更新的新方法研究它们的作用方式,特别是它们效应复合体组分构成及其相互作用特征,如何结合 lncRNA 分子特征对它们进行功能分类等等都是既有意义又十分有挑战性的科学问题。此外,循着陆地植物演化规律,解析 lncRNA 序列上的自然遗传变异带来的其表达、功能和(或)作用机制上的改变,不仅将有助于从物种进化的角度深入了解 lncRNA 的多样性与功能演化,也将会促使 lncRNA 的研究从模式植物中的分子机理探索拓展为协助或指导农业应用的有力理论依据成为可能。

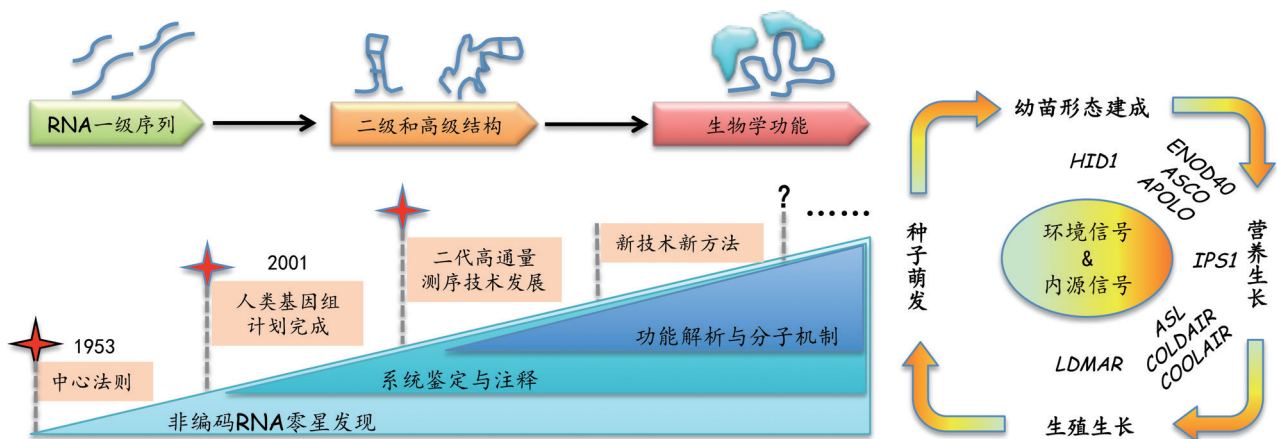


图1 植物中较长非编码RNA的发现与功能鉴定

[参 考 文 献]

- [1] Stern H, Johnston FB, Setterfield G. Some chemical properties of isolated pea nucleoli. *J Biophys Biochem Cytol*, 1959, 6: 57-60
- [2] Kiss T, Toth M, Solymosy F. Plant small nuclear RNAs. Nucleolar U3 snRNA is present in plants: partial characterization. *Eur J Biochem*, 1985, 152: 259-66
- [3] Leader DJ, Sanders JF, Waugh R, et al. Molecular characterisation of plant U14 small nucleolar RNA genes: closely linked genes are transcribed as polycistronic U14 transcripts. *Nucleic Acids Res*, 1994, 22: 5196-203
- [4] Leader DJ, Clark GP, Watters J, et al. Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J*, 1997, 16: 5742-51
- [5] Barneche F, Gaspin C, Guyot R, et al. Identification of 66 box C/D snoRNAs in *Arabidopsis thaliana*: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2'-O-methylation sites. *J Mol Biol*, 2001, 311: 57-73
- [6] Qu LH, Meng Q, Zhou H, et al. Identification of 10 novel snoRNA gene clusters from *Arabidopsis thaliana*. *Nucleic Acids Res*, 2001, 29: 1623-30
- [7] Kim SH, Spensley M, Choi SK, et al. Plant U13 orthologues and orphan snoRNAs identified by RNomics of RNA from *Arabidopsis* nucleoli. *Nucleic Acids Res*, 2010, 38: 3054-67
- [8] Liu TT, Zhu D, Chen W, et al. A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in *Oryza sativa*. *Mol Plant*, 2013, 6: 830-46
- [9] Wang Y, Wang X, Deng W, et al. Genomic features and regulatory roles of intermediate-sized non-coding RNAs in *Arabidopsis*. *Mol Plant*, 2014, 7: 514-27
- [10] Zhang Y, Yang L, Chen LL. Life without A tail: new formats of long noncoding RNAs. *Int J Biochem Cell Biol*, 2014, 54: 338-49
- [11] Dupuis-Sandoval F, Poirier M, Scott MS. The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdiscip Rev RNA*, 2015, 6: 381-97
- [12] MacIntosh GC, Wilkerson C, Green PJ. Identification and analysis of *Arabidopsis* expressed sequence tags characteristic of non-coding RNAs. *Plant Physiol*, 2001, 127: 765-76
- [13] Marker C, Zemann A, Terhorst T, et al. Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr Biol*, 2002, 12: 2002-13
- [14] Rymarquis LA, Kastenmayer JP, Huttenhofer AG, et al. Diamonds in the rough: mRNA-like non-coding RNAs. *Trends Plant Sci*, 2008, 13: 329-34
- [15] Dong Z, Chen Y. Transcriptomics: advances and approaches. *Sci Chn: Life Sci*, 2013, 56: 960-7
- [16] Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*, 2015, 13: 278-89
- [17] Stolc V, Samanta MP, Tongprasit W, et al. Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays. *Proc Natl Acad Sci USA*, 2005, 102: 4453-8
- [18] Li L, Wang X, Sasidharan R, et al. Global identification and characterization of transcriptionally active regions in the rice genome. *PLoS One*, 2007, 2: e294
- [19] Liu J, Jung C, Xu J, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell*, 2012, 24: 4333-45
- [20] Wang XJ, Gaasterland T, Chua NH. Genome-wide prediction and identification of *cis*-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol*, 2005, 6: R30
- [21] Wang H, Chua NH, Wang XJ. Prediction of *trans*-antisense transcripts in *Arabidopsis thaliana*. *Genome Biol*, 2006, 7: R92
- [22] Wen J, Parker BJ, Weiller GF. *In Silico* identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula*. *In Silico Biol*, 2007, 7: 485-505
- [23] Song D, Yang Y, Yu B, et al. Computational prediction of novel non-coding RNAs in *Arabidopsis thaliana*. *BMC Bioinformatics*, 2009, 10 suppl: S36
- [24] Chen CJ, Zhou H, Chen YQ, et al. Plant noncoding RNA gene discovery by "single-genome comparative genomics". *RNA*, 2011, 17: 390-400
- [25] Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One*, 2012, 7: e43047
- [26] Wu J, Okada T, Fukushima T, et al. A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in *Arabidopsis*. *RNA Biol*, 2012, 9: 302-13
- [27] Wu HJ, Wang ZM, Wang M, et al. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol*, 2013, 161: 1875-84
- [28] Li L, Eichten SR, Shimizu R, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*, 2014, 15: R40
- [29] Zhang YC, Liao JY, Li ZY, et al. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol*, 2014, 15: 512
- [30] Jiao Y, Meyerowitz EM. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol Syst Biol*, 2010, 6: 419
- [31] Zhelyazkova P, Sharma CM, Forstner KU, et al. The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell*, 2012, 24: 123-36
- [32] Wang H, Chung PJ, Liu J, et al. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res*, 2014, 24: 444-53
- [33] Di C, Yuan J, Wu Y, et al. Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J*, 2014, 80: 848-61
- [34] Shafiq S, Li J, Sun Q. Functions of plants long non-coding

- RNAs. *Biochim Biophys Acta*, 2016, 1859: 155-62
- [35] Kang C, Liu Z. Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genomics*, 2015, 16: 815
- [36] Zhu B, Yang Y, Li R, et al. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *J Exp Bot*, 2015, 66: 4483-95
- [37] Amaral PP, Clark MB, Gascoigne DK, et al. lncRNADB: a reference database for long noncoding RNAs. *Nucleic Acids Res*, 2011, 39: D146-51
- [38] Zhao Y, Li H, Fang S, et al. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res*, 2016, 44: D203-8
- [39] Swarbreck D, Wilks C, Lamesch P, et al. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res*, 2008, 36: D1009-14
- [40] Yi X, Zhang Z, Ling Y, et al. PNRD: a plant non-coding RNA database. *Nucleic Acids Res*, 2015, 43: D982-9
- [41] Jin J, Liu J, Wang H, et al. PLncDB: plant long non-coding RNA database. *Bioinformatics*, 2013, 29: 1068-71
- [42] Xuan H, Zhang L, Liu X, et al. PLNlncRbase: a resource for experimentally identified lncRNAs in plants. *Gene*, 2015, 573: 328-32
- [43] Chen D, Yuan C, Zhang J, et al. PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic Acids Res*, 2012, 40: D1187-93
- [44] Badger JH, Olsen GJ. CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol*, 1999, 16: 512-24
- [45] Hatziogeorgiou AG, Fiziev P, Reczko M. DIANA-EST: a statistical analysis. *Bioinformatics*, 2001, 17: 913-9
- [46] Rivas E, Eddy SR. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, 2001, 2: 8
- [47] Mignone F, Grillo G, Liuni S, et al. Computational identification of protein coding potential of conserved sequence tags through cross-species evolutionary analysis. *Nucleic Acids Res*, 2003, 31: 4639-45
- [48] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*, 2011, 39: W29-37
- [49] Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*, 2011, 27: i275-82
- [50] Franco-Zorrilla JM, Valli A, Todesco M, et al. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet*, 2007, 39: 1033-7
- [51] Kong L, Zhang Y, Ye ZQ, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*, 2007, 35: W345-9
- [52] Lu ZJ, Yip KY, Wang G, et al. Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome Res*, 2011, 21: 276-85
- [53] Ulitsky I, Shkumatava A, Jan CH, et al. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, 2011, 147: 1537-50
- [54] Wang Y, Fan X, Lin F, et al. *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. *Proc Natl Acad Sci USA*, 2014, 111: 10359-64
- [55] Teotia S, Singh D, Tang X, et al. Essential RNA-based technologies and their applications in plant functional genomics. *Trends Biotechnol*, 2016, 34: 106-23
- [56] Campalans A, Kondorosi A, Crespi M. *Enod40*, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell*, 2004, 16: 1047-59
- [57] Ariel F, Jegu T, Latrasse D, et al. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell*, 2014, 55: 383-96
- [58] Kwok CK, Tang Y, Assmann SM, et al. The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem Sci*, 2015, 40: 221-32
- [59] Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, 2011, 331: 76-9
- [60] Ariel F, Romero-Barrios N, Jegu T, et al. Battles and hijacks: noncoding transcription in plants. *Trends Plant Sci*, 2015, 20: 362-71
- [61] Chekanova JA. Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol*, 2015, 27: 207-16
- [62] Liu J, Wang H, Chua NH. Long noncoding RNA transcriptome of plants. *Plant Biotechnol J*, 2015, 13: 319-28
- [63] Liu X, Hao L, Li D, et al. Long non-coding RNAs and their biological roles in plants. *Genomics Proteomics Bioinformatics*, 2015, 13: 137-47
- [64] Swiezewski S, Liu F, Magusin A, et al. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature*, 2009, 462: 799-802
- [65] Shin JH, Chekanova JA. *Arabidopsis* RRP6L1 and RRP6L2 function in *FLOWERING LOCUS C* silencing via regulation of antisense RNA synthesis. *PLoS Genet*, 2014, 10: e1004612
- [66] Ding J, Lu Q, Ouyang Y, et al. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci USA*, 2012, 109: 2654-9
- [67] Zhou H, Liu Q, Li J, et al. Photoperiod- and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. *Cell Res*, 2012, 22: 649-60
- [68] Crespi MD, Jurkevitch E, Poiret M, et al. *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J*, 1994, 13: 5099-112
- [69] Sousa C, Johansson C, Charon C, et al. Translational and

- structural requirements of the early nodulin gene *enod40*, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol Cell Biol*, 2001, 21: 354-66
- [70] Rohrig H, Schmidt J, Miklashevichs E, et al. Soybean *ENOD40* encodes two peptides that bind to sucrose synthase. *Proc Natl Acad Sci USA*, 2002, 99: 1915-20
- [71] Bardou F, Ariel F, Simpson CG, et al. Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. *Dev Cell*, 2014, 30: 166-76
- [72] Ulveling D, Francastel C, Hube F. When one is better than two: RNA with dual functions. *Biochimie*, 2011, 93: 633-44
- [73] Bardou F, Merchan F, Ariel F, et al. Dual RNAs in plants. *Biochimie*, 2011, 93: 1950-4